AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Solid Earth

RESEARCH ARTICLE

10.1002/2014JB011077

Key Points:
- Efficient transient detection using linear sparse estimation techniques
- Bayesian sampling provides uncertainties for transient amplitudes and durations
- Spatial weighting filters for spatially coherent signals

Correspondence to:
B. Riel,
briel@caltech.edu

Citation:
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.

Received 26 FEB 2014
Accepted 22 MAY 2014
Accepted article online 27 MAY 2014
Published online 10 JUN 2014

Detecting transient signals in geodetic time series using sparse estimation techniques

Bryan Riel[1], Mark Simons[1], Piyush Agram[1,2], and Zhongwhen Zhan[1,3]

[1]Seismological Laboratory, California Institute of Technology, Pasadena, California, USA, [2]Now at Jet Propulsion Laboratory, Pasadena, California, USA, [3]Now at Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California, USA

Abstract We present a new method for automatically detecting transient deformation signals from geodetic time series. We cast the detection problem as a least squares procedure where the design matrix corresponds to a highly overcomplete, nonorthogonal dictionary of displacement functions in time that resemble transient signals of various timescales. The addition of a sparsity-inducing regularization term to the cost function limits the total number of dictionary elements needed to reconstruct the signal. Sparsity-inducing regularization enhances interpretability of the resultant time-dependent model by localizing the dominant timescales and onset times of the transient signals. Transient detection can then be performed using convex optimization software where detection sensitivity is dependent on the strength of the applied sparsity-inducing regularization. To assess uncertainties associated with estimation of the dictionary coefficients, we compare solutions with those found through a Bayesian inference approach to sample the full model space for each dictionary element. In addition to providing uncertainty bounds on the coefficients and confirming the optimization results, Bayesian sampling reveals trade-offs between dictionary elements that have nearly equal probability in modeling a transient signal. Thus, we can rigorously assess the probabilities of the occurrence of transient signals and their characteristic temporal evolution. The detection algorithm is applied on several synthetic time series and real observed GPS time series for the Cascadia region. For the latter data set, we incorporate a spatial weighting scheme that self-adjusts to the local network density and filters for spatially coherent signals. The weighting allows for the automatic detection of repeating slow slip events.

1. Introduction

We define transient deformation signals as nonperiodic, nonsecular accumulation of strain in the crust. Over seismically active regions, transients are often the surface manifestations of slow slip events that are difficult to measure directly with traditional seismological instruments [e.g., *Rogers and Dragert*, 2003; *McGuire and Segall*, 2003; *Gomberg et al.*, 2010; *Kato et al.*, 2012; *Szeliga et al.*, 2008]. In volcanically active regions, transients frequently correspond to periods of ground deformation caused by underlying magmatic activity [e.g., *Ji and Herring*, 2011; *Langbein*, 2003; *Masterlark and Lu*, 2004; *Pritchard and Simons*, 2004]. Previously studied transient events vary widely in magnitude, ranging from spatially coherent surface motions of several centimeters to more subtle motions of only a few millimeters. These signals have also varied widely in duration from year-long signals in subduction zone areas [e.g., *Miyazaki et al.*, 2003] to very short episodes lasting only a few days [e.g., *Rogers and Dragert*, 2003]. Despite the highly nonuniform properties associated with transients, their detection has relied on their combined temporal and spatial coherency, i.e., they are defined by a measurable temporal evolution and systematic spatial structures [e.g., *Ji and Herring*, 2013].

Detection of transient events with unknown magnitudes and durations requires precise measurements of surface displacements over sufficiently large regions. Over the past two decades, the availability of such measurements for monitoring crustal deformation has rapidly increased. Large-scale continuously operating GPS networks, such as the Plate Boundary Observatory (PBO) network in the western United States (http://pboweb.unavco.org), are used to derive station positions with typical daily repeatabilities of 2–3 mm for horizontal positions and 7–8 mm for vertical positions [*Williams et al.*, 2004]. In addition to regional coverage, many GPS networks are also very dense, with 1100 permanent GPS stations for PBO, 1200 for Japan's GEONET network [*Sagiya*, 2004], over 200 for Taiwan [*Hsu et al.*, 2009; *Hung and Rau*, 2013], over 200 for New Zealand [*Wallace and Beavan*, 2010], etc. Geodesy based on repeat imagery (e.g., interferometric

synthetic aperture radar (InSAR)) naturally provides spatially dense observations of surface motion but typically suffers from poor temporal sampling. However, the advent of long time span InSAR time series, new methods for analyzing the temporal evolution of signals contained in interferograms, and future InSAR missions with approximately weekly repeat times will enable large-scale, high-resolution studies of crustal deformation with sufficiently high temporal resolution to capture many transient processes [*Covello et al.*, 2010; *Hetland et al.*, 2012; *Agram et al.*, 2013].

Each geodetic data type has its own unique set of error characteristics which complicate transient detection. GPS time series typically contain Gaussian white noise plus time-correlated random walk components which can resemble transient signals [*Zhang et al.*, 1997; *Langbein*, 2004]. GPS networks exhibit spatially correlated common mode errors which must be estimated and removed as part of any analysis [*Dong et al.*, 2006]. Many of these errors can be mitigated by analyzing an ensemble of data sets. For a given GPS network density, time-correlated signals in GPS data that are evident only at individual stations can be classified as either colored noise or more local processes. Similarly, for InSAR time series, phase delays induced by heterogeneous propagation velocities in the atmosphere can be modeled out or mitigated by ensemble averaging of line-of-sight velocities or other time series techniques [e.g., *Williams et al.*, 1998; *Berardino et al.*, 2002; *Jolivet et al.*, 2011; *Hetland et al.*, 2012]. In addition to noise characteristics, the presence of other confounding signals, such as secular and seasonal effects, can complicate detection of transients. When data volume is large, properly handling sources of errors and nontransient signals makes manual inspection of the data infeasible and requires a sufficiently automated detection algorithm.

We propose a new method for estimating the time and duration of anomalous transient signals in geodetic time series by employing sparse estimation techniques. This method makes use of a dictionary of nonorthogonal time evolution functions that resemble temporally correlated transient events. Estimating the coefficients of the dictionary allows for the reconstruction of transient signals of varying durations and start times. The flexibility of the dictionary also allows for inclusion of known signals, such as seasonal, secular, coseismic/instruments offsets, and postseismic deformation. The nonorthogonality of the dictionary requires regularization during least squares estimation of the dictionary coefficients. We employ a sparsity-promoting regularization approach to compactly reconstruct the underlying transient signal. Additionally, we introduce a Bayesian sampling scheme for the estimation problem to rigorously assess the uncertainties associated with coupling a nonorthogonal, overcomplete dictionary with sparsity-promoting techniques. The final reconstruction inherently includes information about the dominant timescales and likely start times of any transient signals. When time series are available from multiple locations within a certain region, we can exploit the expected spatial coherency of transient signals with a straightforward spatial weighting scheme that encourages selection of dictionary elements that are common to stations within a given length scale. Application of this method to both synthetic and real GPS time series from the Cascadia region demonstrates the successful recovery of signals of different timescales and magnitudes while providing a direct estimate of the long-term tectonic signal.

2. Transient Detection

Here we assume no a priori information about the underlying physical mechanisms responsible for a given transient signal. This assumption prevents us from imposing time functions corresponding to a specific physical description. Instead, we use a flexible approach that parameterizes time-dependent deformation with an *overcomplete* set, or dictionary, of functions that describe the full suite of behaviors we would expect to be present in a given time series. This dictionary can in principle include sinusoidal functions to model seasonal signals, linear terms for secular velocities, heaviside functions for coseismic offsets, etc. For transient signals of unknown initiation times and durations, we populate the dictionary with third-order time-integrated B-splines, hereafter referred to as B^i-splines, which exhibit one-sided behavior of a particular timescale (Figure 1) [*Hetland et al.*, 2012]. By dividing the time span of a time series into uniformly spaced intervals, we can generate a series of B^i-splines centered at the endpoints of the intervals with durations proportional to the interval durations. In this work, a dyadic spacing scheme (i.e., B^i-splines of duration $T/4$, $T/8$, $T/16$, etc., for a given time span T) is used to populate the dictionary. A similar approach is used in wavelet analysis to efficiently cover the frequency spectrum of a signal [e.g., *Mallat*, 1989]. At this stage, modeling time-dependent deformation for time series data, $\mathbf{d}$, is reduced to estimating the coefficients, $\mathbf{m}$, of the dictionary elements in $\mathbf{G}$, while imposing a linear relation between the model parameters and data, $\mathbf{Gm} = \mathbf{d}$ (here boldface indicates matrix or vector quantities).

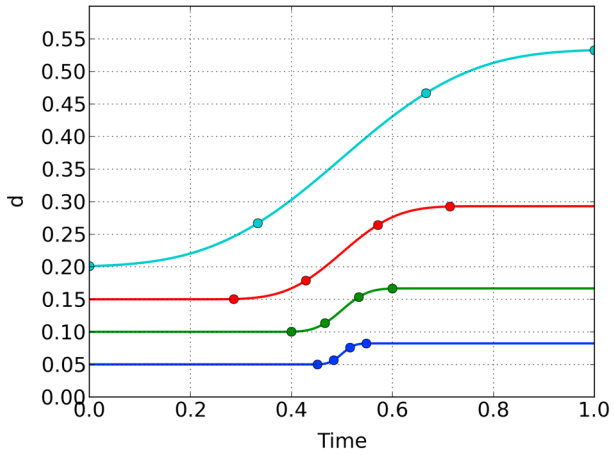

Figure 1. Uniform integrated B-splines (B^i-splines) of various timescales used as candidate temporal displacement functions. The characteristic timescales for the B^i-splines are determined by dividing the time span into uniformly spaced intervals. The filled circles denote the interval spacings for each B^i-spline and differ by factors of 2.

2.1. Regularized Least Squares

Due to the nonorthogonality of the B^i-splines in $\mathbf{G}$, any estimate of $\mathbf{m}$ derived using ordinary least squares methods will be particularly sensitive to the data noise and will exhibit large variances for the estimated parameters. Regularization techniques aim to reduce this sensitivity by jointly minimizing a measure of the residual $\|\mathbf{Gm}-\mathbf{d}\|_2^2$, where $\|\cdot\|_2$ denotes the Euclidean or ℓ_2-norm, and a regularizing function that incorporates a priori information about the solution. Typically, regularized least squares optimization minimizes the unconstrained cost function, $\varphi(\mathbf{m})$:

$$\varphi(\mathbf{m}) = \|\mathbf{Gm}-\mathbf{d}\|_2^2 + \lambda F(\mathbf{m}), \tag{1}$$

where $\lambda > 0$ controls the degree of regularization and $F(\mathbf{m})$ is the regularizing function. The above formulation can be modified to incorporate uncertainties on the observed data in $\mathbf{d}$. Traditional zeroth-order Tikhonov regularization, where $F(\mathbf{m}) = \|\mathbf{m}\|_2^2$, minimizes the size or energy of the solution $\mathbf{m}$. For a mixed dictionary of B^i-splines and steady state functions, such as seasonal and secular terms, we generally only penalize the B^i-splines and allow the steady signals to compensate for the rest of the displacement provided by the data. Thus, in equation (1), $F(\mathbf{m}) \rightarrow F\left(\mathbf{m}^{B^i}\right)$, where $\mathbf{m}^{B^i} \subseteq \mathbf{m}$ denotes the B^i-spline coefficients. For brevity in the following discussion, we assume $\mathbf{m} = \mathbf{m}^{B^i}$. In a Bayesian framework, Tikhonov regularization implies an uncorrelated zero-mean Gaussian prior for the coefficients of the B^i-splines and uniform priors for everything else.

However, transient events are not well described by Gaussian statistics; rather, transients are sparse and irregular in nature. To promote sparsity, we would like to use $F(\mathbf{m}) = \|\mathbf{m}\|_0$, where $\|\cdot\|_0$ denotes the counting pseudo-norm and measures the number of nonzero elements in $\mathbf{m}$ [*Candés and Wakin*, 2008; *Donoho*, 2006]. Since this formulation results in an intractable combinatorial problem, it is common to use an approximation with $F(\mathbf{m}) = \|\mathbf{m}\|_1$, where $\|\cdot\|_1$ is the ℓ_1-norm or the sum of the absolute values of $\mathbf{m}$, leading to a *convex* cost function [*Tibshirani*, 1996; *Chen et al.*, 1998; *Donoho*, 2006]:

$$\varphi(\mathbf{m})_{\ell_1} = \|\mathbf{Gm}-\mathbf{d}\|_2^2 + \lambda\|\mathbf{m}\|_1. \tag{2}$$

Using ℓ_1-norm regularization, the solution still maintains sparsity, i.e., many components of $\mathbf{m}$ are very close to zero and the remaining components can still effectively describe the data. Furthermore, a convex cost function results in a solution that is guaranteed to be globally optimal [*Boyd and Vandenberghe*, 2004]. We cast the sparse regularization problem as a quadratic program and solve for $\mathbf{m}$ using the CVXOPT software (http://abel.ee.ucla.edu/cvxopt/index.html).

Sparsity-promoting regularization with the ℓ_1-norm has been effectively used for recovering isolated spikes in seismic data [*Taylor et al.*, 1979], detecting sharp discontinuities in tomography studies [*Gholami and Siahkoohi*, 2010], and estimating compact distributions of fault slip for large earthquakes [*Evans and Meade*, 2012]. In the context of transient detection, this method automatically determines the B^i-splines which best

model the deformation while zeroing out the others. The advantage of enforcing a sparse set of B^i-splines is that we automatically place higher importance to B^i-splines that have nearly the same timescales and onset times as any transient signals present in the data and heavily penalize those that do not significantly improve our data fit. Steady signals, such as those from seasonal and secular processes, are also required to be consistent with a sparse set of B^i-splines. This requirement is beneficial for ensuring that any estimated steady signals do not overly accommodate the observed displacement which can lead to false positives in transient signal detection (section 5).

The effectiveness of the ℓ_1-norm for recovering sparse solutions can be enhanced by adaptive reweighting techniques. In *Candés et al.* [2008], an iterative reweighting algorithm was introduced where each coefficient, m_i, is assigned a different penalty parameter, λ_i, which is inversely proportional to $|m_i|$ at the current iteration. By initializing the algorithm with uniform values for λ_i, each successive iteration causes larger coefficients to be penalized less heavily than smaller coefficients, leading to a solution that enhances the most dominant B^i-splines. *Candés et al.* [2008] demonstrated that reweighting brings the ℓ_1-norm closer to the ℓ_0-pseudo-norm by increasing the strength of the regularizing function $F(\mathbf{m})$ near the origin. Theoretically, one could choose from a multitude of functional forms relating λ_i to $|m_i|$, e.g., $\lambda_i \propto |m_i|^{-1}$, $\lambda_i \propto m_i^{-2}$, $\lambda_i \propto \log\left(|m_i|^{-1}\right)$, etc. Larger negative powers for m_i will increase the strength of the sparsity constraint and bring the ℓ_1-norm very close to the ℓ_0-norm. In practice, we have found that using the stronger reweighting functions favors selection of short timescale B^i-splines while the logarithmic functions favor longer timescales. Thus, selecting the appropriate reweighting function can depend on the expected timescales of the transient signals in a data set. Typically, for all reweighting functions, 5–10 reweighting iterations are required for convergence.

2.2. Posterior Uncertainties

In ordinary least squares problems where all variables are assumed to be Gaussian (unregularized or Tikhonov regularization), analytic relations exist to estimate model and predicted data uncertainties [*Tarantola*, 2005]. While the ℓ_1-norm regularization prevents us from directly using those relations, we can interpret the minimization of the cost function in equation (2) as choosing the optimum subset of the elements of $\mathbf{G}$ that minimize the data misfit as well as the number of elements used for the solution and determining the coefficients of those elements. Thus, for a given solution vector $\mathbf{m}$, we can construct a compact dictionary $\tilde{\mathbf{G}}$ populated with the elements corresponding to the largest coefficient absolute values in $\mathbf{m}$. We can construct a diagonal prior covariance matrix, $\mathbf{C}_m$, where the values along the diagonal correspond to the squared coefficients in $\mathbf{m}$. Then, for a given data covariance matrix $\mathbf{C}_d$, we can apply the standard least squares formulation to obtain a solution $\tilde{\mathbf{m}}$:

$$\tilde{\mathbf{m}} = \left(\tilde{\mathbf{G}}^{\intercal}\mathbf{C}_d^{-1}\tilde{\mathbf{G}} + \mathbf{C}_m^{-1}\right)^{-1}\tilde{\mathbf{G}}^{\intercal}\mathbf{C}_d^{-1}\mathbf{d}. \tag{3}$$

Various stopping criteria can be applied for determining the number of elements to include in $\tilde{\mathbf{G}}$. Here we apply a variance reduction criterion where we first remove the estimated steady state signals from the data and iteratively remove modeled transient displacements corresponding to the largest values of $\mathbf{m}$ from the data until the variance reduction reaches a prescribed threshold. This approach is similar to matching pursuit methods that iteratively search through nonorthogonal bases and add those to the dictionary that are most correlated with the data residual at each iteration [e.g., *Mallat*, 1989]. In our case, the "best basis" is determined in one step through the convex optimization and is less susceptible to high data noise or initial errors in basis selection [*Chen et al.*, 1998].

Using knowledge about uncertainties for the observed data in $\mathbf{C}_d$, the posterior model covariance matrix can be computed as follows:

$$\tilde{\mathbf{C}}_m = \left(\tilde{\mathbf{G}}^{\intercal}\mathbf{C}_d^{-1}\tilde{\mathbf{G}} + \mathbf{C}_m^{-1}\right)^{-1}. \tag{4}$$

Since the resultant compact dictionary $\tilde{\mathbf{G}}$ is in most cases composed of nonorthogonal elements and will be ill posed, we can expect some large off-diagonal components in $\tilde{\mathbf{C}}_m$. The probability density function of the predicted data is then a multivariate Gaussian distribution with a mean of $\tilde{\mathbf{G}}\tilde{\mathbf{m}}$ and a covariance matrix, $\tilde{\mathbf{C}}_d$, given by the following:

$$\tilde{\mathbf{C}}_d = \tilde{\mathbf{G}}\tilde{\mathbf{C}}_m\tilde{\mathbf{G}}^{\intercal}. \tag{5}$$

We note that while this approach is useful for assessing the uncertainties of the coefficients of $\tilde{\mathbf{G}}$ and the predicted data, it does not address the uncertainties associated with the subsetting of $\mathbf{G}$ to form $\tilde{\mathbf{G}}$. Furthermore, this approach relies on the assumption that the model parameters are normally distributed, which is contrary to our assumption of transients as temporally sparse.

2.3. Bayesian Sampling

As was previously mentioned, the commonly implemented form of Tikhonov regularization is equivalent to enforcing a Gaussian prior on the elements of $\mathbf{m}$. Analogously, sparsity-promoting regularization in its most basic form can be achieved by enforcing a Laplace prior for $\mathbf{m}$, which has the form $p(m_i) \propto \exp\{-\lambda|m_i|\}$ [*Tibshirani*, 1996]. We can see this result by considering Bayes' theorem, $P(\mathbf{m}|\mathbf{d}) \propto P(\mathbf{d}|\mathbf{m})\,P(\mathbf{m})$, where $P(\mathbf{m}|\mathbf{d})$ is the posterior distribution of our model coefficients (i.e., the distribution of values for $\mathbf{m}$ that explain the data), $P(\mathbf{d}|\mathbf{m})$ is the data likelihood, and $P(\mathbf{m})$ is the prior distribution of the coefficients. By maximizing $\exp\left\{-\varphi(\mathbf{m})_{\ell_1}\right\}$, it can be shown that the regularized least squares solution is equivalent to maximizing the posterior distribution with a Gaussian misfit between the data and model prediction and a Laplace prior on the model coefficients.

Laplace priors are characterized by high probabilities near the origin with long-tails to allow for an increased likelihood of arbitrarily large values relative to a Gaussian prior. The penalty term λ acts as a scale factor that controls the width of the distribution and the probability that the elements of $\mathbf{m}$ will be sparse. Since there is no convenient conjugate relation between a Gaussian likelihood and Laplace prior, we cannot derive a closed-form solution for the posterior distribution $p(\mathbf{m})$. Instead, we employ a Gibbs sampler to draw samples from the posterior distribution [*Gelman et al.*, 2004]. The Gibbs sampler explores the posterior distribution of each variable in the model using distributions conditional on the current values for all other variables. Following the approach of *Park and Casella* [2008], we group the coefficient amplitudes, m_i, as a single variable and the coefficient *precisions*, τ_i, as another group of variables. The conditional distributions relating the coefficient amplitudes and precisions are obtained by expressing the Laplace prior as a scale mixture of normals with an exponential mixing density:

$$\frac{\lambda}{2}e^{-\lambda m_i} = \int_0^\infty \frac{1}{\sqrt{2\pi\tau_i}} e^{-m_i^2/(2\tau_i)} \frac{\lambda^2}{2} e^{-\lambda^2\tau_i/2} \mathrm{d}\tau_i. \tag{6}$$

The posterior distribution can now be expressed as a product of a Gaussian data likelihood, Gaussian priors for the dictionary coefficients, and exponential hyperpriors for the coefficient precisions. This hierarchical representation where each τ_i is treated as a hyperparameter is analogous to the reweighting scheme described in section 2.1. The conditional distributions in this hierarchy are straightforward to sample from with block updates of $\mathbf{m}$ and $(\tau_1^2, \ldots, \tau_P^2)$.

There are several advantages that favor a Bayesian sampling method over regularized least squares. For one, optimization-based approaches for non-Gaussian priors do not permit us to readily compute uncertainties associated with our estimate for $\mathbf{m}$. While we could apply least squares theory using a sparse subset of $\mathbf{G}$ that best explain the data (see section 2.2), we would still require the assumption that the model parameters are normally distributed. We also do not obtain much information about the full solution space, which is necessary if different families of solutions exist with nearly the same predictive power as the optimal solution. For example, consider the inherent trade-off between a single B^i-spline to model a transient signal and two shorter timescale B^i-splines located at the same time. While selecting the single B^i-spline would be the sparser solution, we can imagine a situation where the combination of the two shorter B^i-splines provided a better fit to the data. In this case, the single, longer timescale B^i-spline would trade off with the shorter ones where the strength of the trade-off would be dependent on the value of the penalty λ. Bayesian sampling allows for sampling from the full solution space where models are produced in numbers proportional to their probability given the data [e.g., *Gelman et al.*, 2004; *Tarantola*, 2005].

2.4. Selecting the Penalty Parameter

The parameter λ in the Laplace prior controls the relative strengths of the steady state terms and the B^i-spline coefficients. Larger values of λ will minimize the contributions of the B^i-splines, leading to a smoother solution. Smaller values of λ will distribute more weight across the B^i-splines, leading to a rougher solution. Thus, the amplitude of the steady state signals is also sensitive to the value of λ and may vary as λ changes. For sparse regularization optimization problems, we use K-fold cross validation to select the optimal values for λ, where K depends on the number of data points available for partitioning into training and

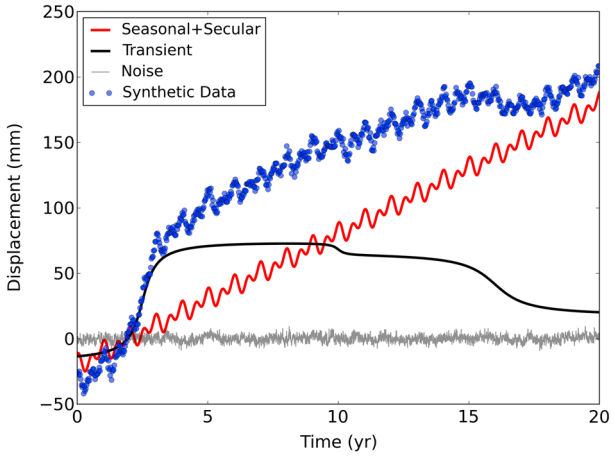

Figure 2. Individual model components input into the synthetic GPS time series. The seasonal signals are constructed using annual and semi-annual periods while the transients are formed using arctangent functions of various amplitudes and durations. Additionally, white and colored noise are added to the model to simulate measurement noise typically found in geodetic time series.

testing subsets. Due to the sparsity-enforcing regularization, we can directly quantify the effective data resolution of $\mathbf{G}$ (i.e., the shortest-duration resolvable signal) as the duration of the shortest timescale B^i-spline included in $\mathbf{G}$. A single B^i-spline is only able to predict four independent observations spaced $T_k/2$ time units apart, where T_k is the effective timescale of the kth B^i-spline. For daily observations, we first partition the data into S random subsets where S is the number of days spanned by the time series divided by the number of observations predictable by $\mathbf{G}$. Within each subset, we further divide the data into K random partitions where one of the partitions is used to compute the data misfit for the proposed model $\mathbf{m}$ trained by the other $K - 1$ partitions. We then average over $S \cdot K$ cross-validation experiments to obtain the average data misfit for the current value of the penalty parameter, λ.

Since cross validation would be computationally expensive for a high number of Gibbs sampling runs, alternative methods are required. A variety of model class selection methods are available that allow for estimation of the *evidence* of a model class, where a model class is defined by the value of λ [*Beck and Yuen*, 2004; *Ching et al.*, 2006]. The evidence measures the average data fit for a model class and the amount of information the model class extracts from the data, i.e., some metric of distance between the posterior and prior distributions [*Beck and Yuen*, 2004]. Unfortunately, model class selection is highly influenced by the choice of the prior distribution, which could lead to significant biases for values of λ that maximize the evidence [*Ching et al.*, 2006]. Here we sample for λ by assigning it a diffuse hyperprior, allowing for a wide range of possible widths for the corresponding Laplace priors [*Park and Casella*, 2008]. We consider a gamma hyperprior on λ^2 such that the prior density is relatively flat up to $\lambda \approx 10^3$ and then decreases steeply to penalize very large values.

3. Synthetic Example

To test the temporal transient detection capabilities of the proposed method, we generated a 20 year synthetic daily GPS time series consisting of seasonal, secular, and transient deformation. The seasonal signals are a linear combination of annual and semi-annual sinusoids. A transient signal is constructed using arctangent functions of three different amplitudes, timescales, and centroids to simulate slow deformation events with various properties (Figure 2). We add white noise plus colored noise using a power law model to mimic errors commonly found in geodetic data [*Langbein*, 2004]. Robust detection of transients is traditionally difficult when the amplitude of the temporally correlated colored noise is on the same order as the signal of interest (as is the case for the weakest synthetic transient). In practice, this problem can be mitigated by exploiting coherency within a geodetic network, but we will demonstrate that successful detection is still possible with a single time series.

We construct $\mathbf{G}$ using a dictionary of reference functions that include simple sinusoidal and linear functions to capture seasonal and secular terms, respectively. Additionally, the dictionary includes B^i-splines with

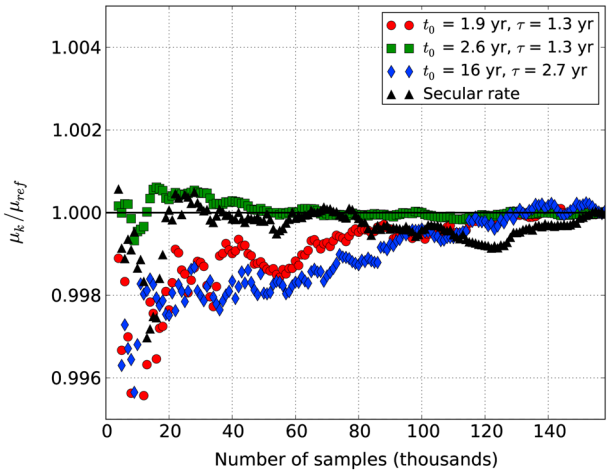

Figure 3. Convergence test of the Gibbs sampler for coefficients corresponding to several B^i-splines. The mean of the samples drawn from the posterior distribution for each B^i-spline is monitored for increasing number of samples. The means are normalized by the mean computed using $2e^5$ samples. After $\mathcal{O}(7e^4)$ samples, the means are within 0.2% of their final values.

effective timescales of 0.3, 0.6, 1.3, 2.7, 5.7, and 13.3 years (chosen by dividing the 20 year time period into 128, 64, 32, 16, 8, and 4 uniformly spaced intervals, respectively). The coefficients of the dictionary terms are estimated using two different methods: least squares optimization with sparsity-inducing regularization on the B^i-spline coefficients and Gibbs sampling of the posterior distribution. As described earlier, we prescribe Gaussian priors on the seasonal and secular terms and Laplace priors on the B^i-splines, although the sampler is initialized with random variates from a wide Gaussian distribution for all coefficients. For this example, we run the Gibbs sampler for 10^5 samples, which is far more than required for convergence but allows the posterior means to evolve to within $< 1\%$ of their final values (Figure 3). After cross validation of 14 independent data subsets, the average optimal penalty parameter was $\lambda = 0.66$, which agreed fairly well with the mean of the posterior distribution of λ constructed with the Gibbs sampler (Figure 4). Prescribing $\lambda = 0.66$ results in a reconstructed transient signal that is rougher than the input transient (Figure 5a). This behavior is primarily due to the temporally correlated colored noise causing several false detections throughout the time series. The random walk characteristics of the noise resemble small transient events that are indistinguishable from true transients (for data from a single station). Colored noise also has the effect of biasing the estimate of the long-term secular rate as demonstrated by the underestimation and

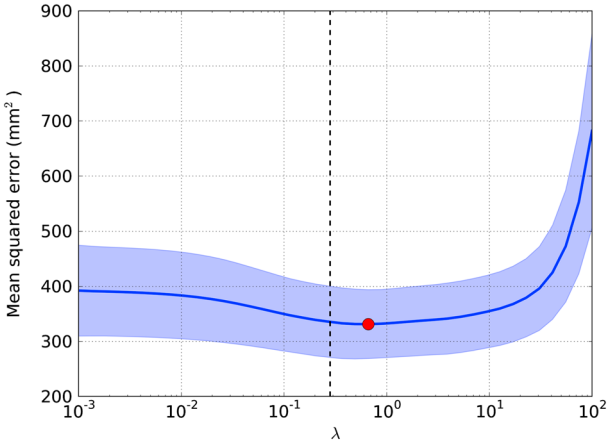

Figure 4. *K*-fold cross-validation results for selection of the penalty parameter, λ, for the synthetic time series. Fourteen separate cross-validation experiments were performed for 14 independent subsets of the time series. The solid blue line shows the mean testing error for all experiments while the shaded region denotes the standard deviation. The vertical black dashed line marks the mean λ as determined by the Gibbs sampler which shows relatively good agreement with the optimal λ obtained from the cross validation (red circle).

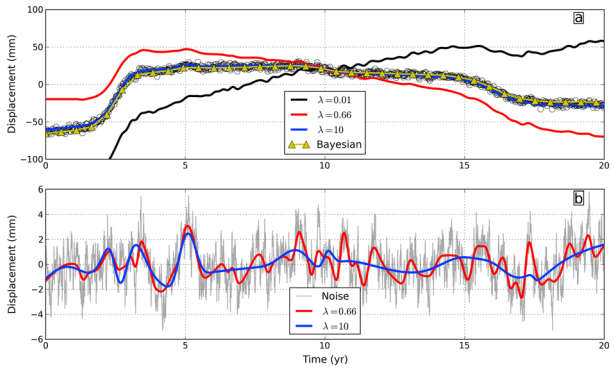

Figure 5. Reconstructed time series for the synthetic data with the simultaneously estimated seasonal and secular signal removed. (a) Full transient signal for models corresponding to different values of the penalty parameter λ. The black circles show the input data after removing the true secular and seasonal signals. Lower values of λ correspond to rougher models. The model for $\lambda = 10$ and the point estimate derived from the Gibbs samples for the same λ are able to nearly exactly reproduce the input signal. (b) The high-frequency component of the reconstructed transient signal compared with the input colored noise. The $\lambda = 0.66$ model (chosen through cross validation) is able to capture nearly all of the temporally correlated noise structure while higher penalties result in smoothing over the higher-frequency variations.

overestimation of the secular rate for the models constructed with $\lambda = 0.01$ and $\lambda = 0.66$, respectively. Low values of λ allow selection of more B^i-splines from the dictionary to fit the smaller signals. By plotting the reconstructed high-frequency signal against the input colored noise (Figure 5b), we observe that enforcing $\lambda = 0.66$ allows us to reconstruct a majority of the structure of the input noise. The combined dictionary plus sparsity-inducing regularization approach thus acts as a smoothing method where λ controls the degree of smoothness of the reconstructed signal. Selecting a much higher value of $\lambda = 10$ still allows us to reconstruct the largest signals of the input noise while smoothing over the higher-frequency variations.

Without spatial information, it is impossible to distinguish between colored noise and true transient signals for tuning λ to recover the correct secular rate. Data from multiple stations must be used to determine whether a temporally coherent signal persists over a finite region (section 4). In a separate cross-validation experiment performed on synthetic data with white noise only resulted in optimal values of $\lambda \approx 10$, which successfully isolates the transient signals. Since the focus of this paper is on transient detection, we proceed with the model resulting from $\lambda = 10$ to smooth over the noise signals, but we emphasize that in general applications, the issue of colored noise must not be neglected.

The reconstructed time series with $\lambda = 10$ successfully models all three input transient signals (Figure 5a). A nearly identical model is achieved by deriving a point estimate from the means of the posterior distribution constructed from the Gibbs sampler with the same value for λ. Even with this higher value of λ, both the optimization and Bayesian solutions are corrupted by a relatively long-duration random walk process that started around year 4.5. The duration of this colored noise was of sufficient length to be modeled by one of the B^i-splines in the dictionary. Increasing the penalty parameter further would smooth over this noise signal at the cost of losing recovery of the smallest transient signal at year 10.

3.1. Coefficient Scalograms

The limited data resolution (as opposed to model resolution) of the dictionary $\mathbf{G}$ will limit the precision of our estimates for transient durations. One can also expect significant covariances between dictionary elements that may have different timescales but share common centroid times. Viewing the estimated B^i-spline coefficients in a scalogram-type fashion reveals the effectiveness of the different estimators in limiting the intradictionary covariances (Figure 6). For comparative purposes, we also show the scalogram for a model using a traditional zeroth-order Tikhonov regularization scheme. The Tikhonov estimator locates the onset times of the largest transient signal relatively well but tends to spread the energy across the timescales, much the same way that a wavelet transform would spread the energy across wavelet scales. In fact, previous methods using wavelet transforms to pick onset times of transients in GPS data have presented scalograms that suggest permanent deformation across all temporal scales [e.g., *Szeliga et al.*, 2008]. The advantage of the sparse estimation techniques is a much stronger localization of energy to very few scales and times, as seen on the bottom two plots in Figure 6. The majority of each input transient signal can be

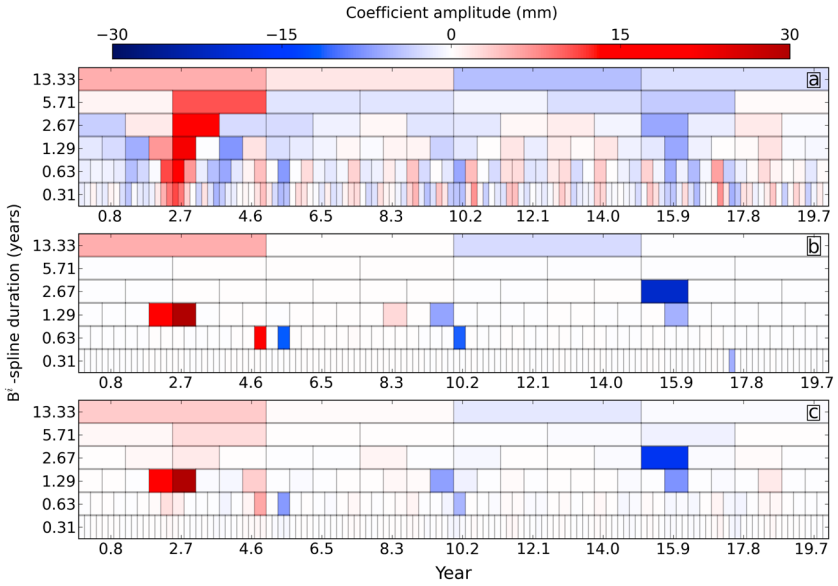

Figure 6. Synthetic scalograms showing the amplitudes of the B^i-splines estimated using three different methods: (a) Tikhonov (ℓ_2-norm) regularization, (b) Sparse (ℓ_1-norm) regularization with $\lambda = 10$, and (c) posterior distribution means derived from the Gibbs samples. Each row of the scalograms corresponds to B^i-splines of a given timescale. Tikhonov regularization results in nonzero amplitudes for nearly all B^i-splines and tends to spread energy across timescales. The sparsity-promoting regularization methods zero-out nearly all of the B^i-splines, leaving only those that describe transients in the data. The estimated secular rates for the Tikhonov, sparse regularization, and Gibbs solutions are 12.0, 9.7, and 8.6 mm/yr, respectively (for an input secular rate of 10 mm/yr).

recovered by just two distinct B^i-splines, allowing for more precise estimates of the transient start times and durations. We reiterate that these estimates have been obtained almost completely automatically, with some minor supervision for selection of the penalty parameter (i.e., choosing the correct data subset size for cross validation such that the optimal estimate of λ does not change significantly for slightly different subset sizes).

3.2. Covariances Between Dictionary Elements

Estimating the uncertainties of the B^i-spline coefficients is straightforward since samples have been drawn from the posterior distribution via the Gibbs sampler. Furthermore, we can directly investigate trade-offs between different parameters and gain insight into the nature of sparse transient detection with a nonorthogonal dictionary. For example, the longer duration transient centered around the 16 year mark can be described well by both a 2.67 year and a 1.29 year B^i-spline, leading to large standard deviations for both coefficients accompanied by a strong trade-off (Figure 7a). The longer duration B^i-spline more closely matches the duration of the input transient but is slightly time shifted from the true centroid time, whereas the 1.29 year B^i-spline is more closely aligned with the centroid time but under-represents the signal duration. The negative slope in the covariance plot is a direct result of the Laplace prior placed on m_i, which tries to drive the coefficients closer to zero. In the case when two coefficients have nearly equal probability in matching the data, many samples will be drawn from the model space where both coefficients are nonzero. However, this behavior is entirely dependent on the value of the penalty parameter λ. Higher values of λ would draw more samples for the more probable coefficient and less for the less probable coefficient (see section 5.1).

We can observe a similar trade-off between a long-timescale B^i-spline with a time centroid of $t = 0$ and the secular rate, implying that long-timescale B^i-splines are nearly as effective in modeling long-term, steady displacement signals (Figure 7c). Trade-offs between B^i-splines adjacent in time with identical durations also show large standard deviations but with slightly weaker trade-offs (Figure 7b). In this case, we can infer that the centroid of the true signal is most likely between the two and has a time duration of $\approx (3/2)T_k$. At the other extreme, when two B^i-splines are centered in a period when no transient signal occurs, nearly all of the samples lie very close to the origin (Figure 7d), replicating the expected probability distributions of two sparse parameters [*Tipping*, 2004]. Another feature evident in many coefficient posterior distributions is the

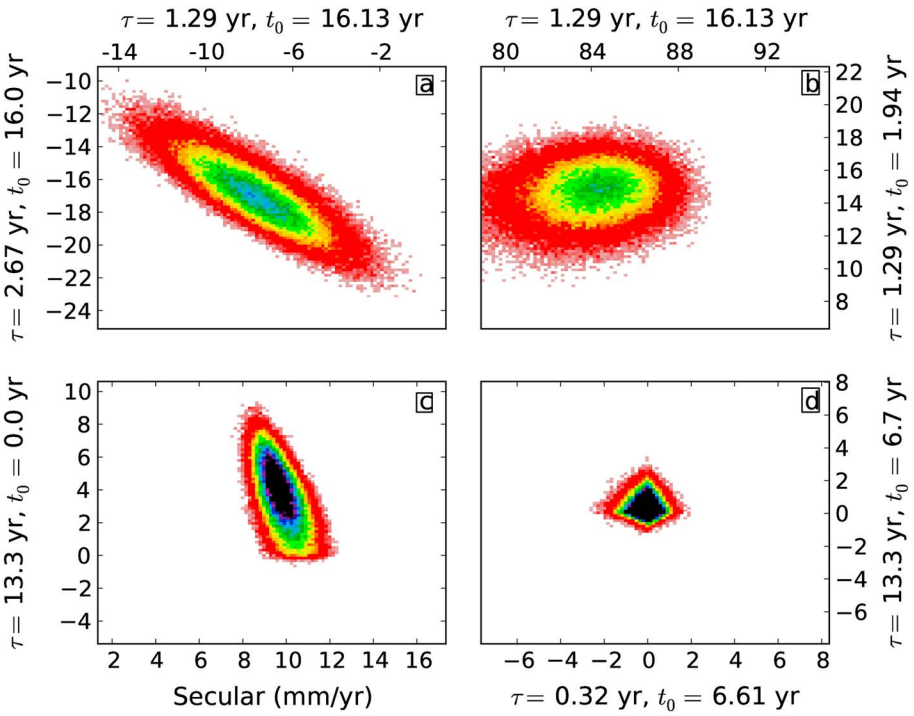

Figure 7. Two-dimensional histograms of samples drawn from posterior distributions for different pairs of B^i-splines. (a) When two B^i-splines have nearly coincident time centroids and are both able to reconstruct a transient signal, their joint posterior exhibits a strong negative covariance. (b) For adjacent B^i-splines of the same timescale where the centroid of the true signal lies between the centroids of the B^i-splines, their amplitudes will covary in a positive manner. (c) Long-timescale B^i-splines and the secular rate will exhibit covariances similar to Figure 7a. Note the sharp corner in the joint distribution due to the ℓ_1-norm penalty. (d) Two B^i-splines located in a time window where no transient signals occur will have samples located very close to the origin with a structure resembling the prior (bivariate Laplace distribution).

sharp corner seen in Figure 7c which is a consequence of the ℓ_1-norm strongly penalizing coefficients of the wrong sign [*Park and Casella*, 2008]. Thus, a simple viewing of the posterior samples can provide a strong indication of the most likely sign of displacement of a detected transient signal.

3.3. Data Subsampling

As discussed previously, when geodetic time series have very high temporal sampling rates, the limited data resolution of $\mathbf{G}$ requires a certain level of data subsampling to obtain reliable estimates of the penalty λ during cross validation. On the other hand, for studies where time series are expected to have poorer temporal sampling, we can estimate the minimum amount of data required to detect a transient signal of a given duration since a B^i-spline of duration T_k is expected to predict four observations spaced $Tk/2$ time units apart. While daily GPS solutions provide adequate sampling rates for capturing many transient processes, other geodetic time series, such as InSAR, provide observations that are typically sparser in time. If the duration of a transient process is appreciably shorter than the time interval between observations, there is a risk of severely mis-estimating the timescale of the signal or not detecting the signal at all. To investigate the effect of sampling rate on transient detection, we repeated the least squares optimization with sparsity-promoting regularization on the synthetic time series with increasing data decimation factors. We varied λ for each decimation factor in order to keep the results consistent. Remarkably, the three input transient events were successfully recovered up to a decimation factor of 256 (Figure 8a). For this highly ill posed case where the number of candidate features (252 B^i-splines) is significantly greater than the number of data points (29 points), sparse regularization was able to recover a stable solution. Reconstruction of the input transients becomes less accurate with increasing time intervals between observations, but even for the smallest signal, only 2–3 data points are required to register a positive detection, in agreement with the expected data resolution of the B^i-splines (Figure 8b).

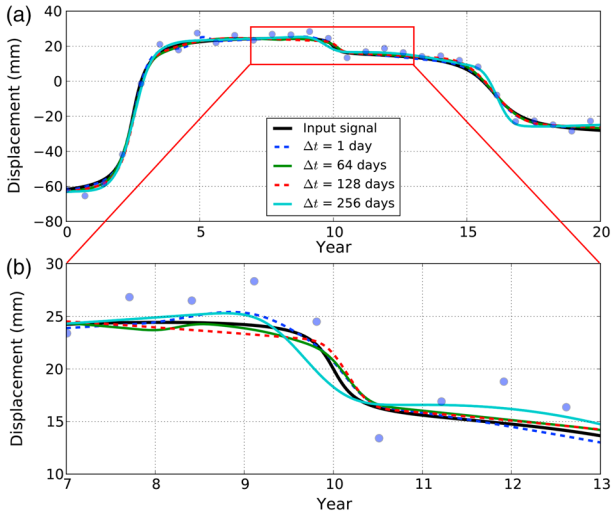

Figure 8. (a) Transient signal recovery using subsampled data with increasing subsampling factors. The limited data resolution of the smallest timescale B^i-splines requires three to four observations to resolve a transient signal. (b) In the case of the smallest synthetic transient around year 10, the temporal spacing of $\Delta t = 256$ days is too large to resolve the short timescale duration, and the reconstructed signal is smoothed. However, this result still demonstrates that a positive detection is still possible with temporally coarse time series.

Decimation also reduces the influence of higher-frequency time-correlated noise by effectively low-pass filtering the data before estimation of $\mathbf{m}$. We can observe this effect from the increasing smoothness of the reconstructed signal for larger decimation factors. This result suggests that a data cascading approach could be beneficial in recovering the strongest signals. For cascading, we would estimate the B^i-spline coefficients using only a subset of the data and use those results as an *a priori* estimate of the coefficients for use with a larger subset of data [*Minson et al.*, 2013]. The a priori estimate could then be integrated with the reweighting approach in section 2.1 to impose smaller penalties on B^i-splines that have larger values in the initial estimate. Subsequent estimates of $\mathbf{m}$ using more data would enhance the stronger signals and reduce the overall effect of high-frequency colored noise. This approach would be useful for combining coincident geodetic time series with different temporal sampling rates and noise characteristics, e.g., InSAR and GPS. Transient detection would first be performed with an InSAR time series. Feeding this initial result into a detection procedure with the GPS data would then enhance the transient signals that are consistent between the two data types.

4. Spatial Sparsity Weighting

Much of the ambiguity over the correct choice of λ and the influence of local noise can be mitigated by using data from multiple surrounding stations. By the adopted definition of what constitutes a transient signal, displacements should be coherent over a finite region and would lead to common nonzero B^i-splines over multiple stations. Thus, the reweighting scheme discussed in section 2.1 can be performed in a spatial sense where the coefficient-dependent penalties are enforced to be consistent over stations within a certain length scale. Using this approach, the spatial reweighting is initialized by performing a single iteration of the minimization of equation (2) independently for each GPS station. After this iteration, candidate coefficient-dependent penalty parameters λ_i are computed for each station based on that station's current initial estimate for $\mathbf{m}$. To enforce spatial consistency for n stations, we select $\lambda_i = f\left(\lambda_i^1, \ldots, \lambda_i^n, w^1, \ldots, w^n\right)$, where $f()$ is a weighted median and w^j are the weights assigned to each station. The weights are recomputed at every kth station using the spatial weighting function:

$$w^j = \exp\left(-\frac{d(j,k)}{L_0^j}\right), \tag{7}$$

where $d(j,k)$ is the distance between station j and k and L_0^j is a prescribed correlation length for the jth station. To account for variable station densities within GPS networks, we choose L_0^j to be the average distance

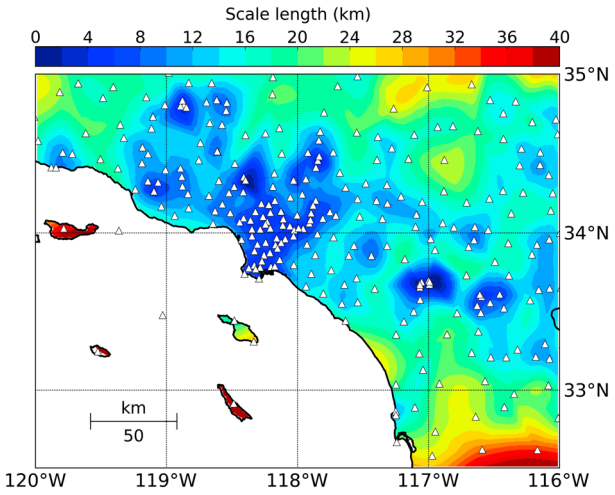

Figure 9. Computed scale length (correlation length) of the GPS network used in the SCEC Phase III validation exercise. Scale length is computed using the average distance from each station to the nearest three stations.

from station *j* to the nearest three or four stations, allowing us to detect spatially correlated transients with length scales proportional to the resolving power of a given network.

Since the spatial sparsity weighting method relies on spatial consistencies of scalograms for neighboring stations, we must ensure that the B^i-spline decomposition of transient signals is translation invariant. Translation invariance states that a time shift of the input signal (the data) will only result in an equivalent time shift of the selected B^i-splines without modification of the amplitudes [*Mallat*, 1989]. In wavelet analysis, wavelet transforms that are not translation invariant can result in vastly different decompositions for small time shifts in the data. For ℓ_1-regularized least squares problems, we can achieve translation invariance by constructing a translation-invariant dictionary $\mathbf{G}$. A given dictionary $\mathbf{G}$ is translation invariant if for any temporal function $\mathbf{g}_i(t) \in \mathbf{G}$ and $t_0 \in \Delta t * [0, N-1]$, where N is the number of data points and Δt is the time duration between observations, then $\mathbf{g}_i[t - t_0] \in \mathbf{G}$ [*Mallat*, 1989]. In other words, we construct $\mathbf{G}$ such that every observation epoch in the time series is associated with a B^i-spline of all valid temporal scales, which would result in a $\mathbf{G}$ matrix with N rows and $N \log_2 N$ columns if a dyadic scale approach is used for the B^i-splines. Due to the large number of parameters associated with translation-invariant dictionaries, we generally only enforce translation-invariance when applying the spatial sparsity weighting to time series that potentially contain rapidly propagating transients, such as in Cascadia. In practice, we have found that longer-duration transients can be effectively isolated using the spatial sparsity weighting with the standard $\mathbf{G}$ construction, i.e., four B^i-splines of duration $T/4$, eight B^i-splines of duration $T/8$, etc.

4.1. Example: Southern California Earthquake Center Validation Exercises

Since 2009, the Southern California Earthquake Center (SCEC) community has coordinated transient detection validation workshops where participants are able to test their detection methods on several synthetic time series resembling data from southern California GPS stations [*Lohman and Murray*, 2013]. The data are generated by the Fakenet package which simulates transient processes of varying complexity while including additional signals from seasonal and secular processes, random and common mode noise, and data gaps [*Agnew*, 2013]. Four phases of testing from 2009 to 2012 were performed with transient sources ranging from slow slip events with strike slip and thrust motions to small- and large-scale aquifer inflations. We apply the spatial sparsity weighting approach to a 10 year synthetic data set from Phase 3 (set D) which contains signals from a simulated thrust event on the Santa Monica fault. As before, we populate a global temporal dictionary with B^i-splines of timescales of $\approx$ 0.16, 0.32, 0.65, 1.33, 2.86, and 6.67 years, as well as functions for seasonal and secular processes. The correlation lengths L_0^j are computed for each station using the average distance to the nearest three stations, resulting in strong resolution power over the Los Angeles basin and weaker resolution near the California-Mexico border and islands (Figure 9).

After about 20 iterations of the spatial sparsity weighting, the thrust event is strongly localized to the stations nearest to the Santa Monica fault with good agreement with the true signal (Figure 10). While there are a few errant nonzero signals for stations outside of the deforming zone, their spatial characteristics suggest

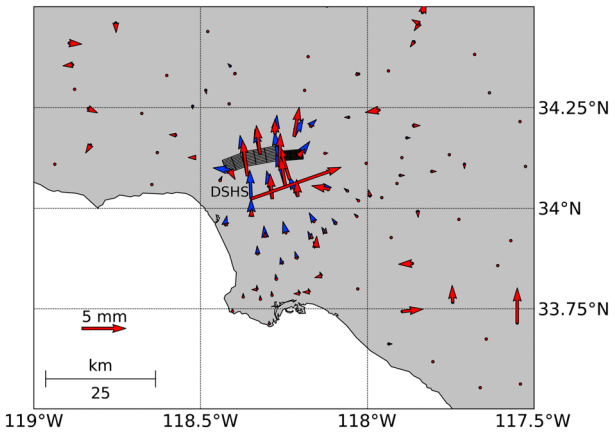

Figure 10. Reconstructed transient signal (red arrows) corresponding to simulated thrust event (blue arrows). Most of the signal is isolated close to the fault patches, although a few stations outside of the basin region show extraneous transients. The large east error at station DSHS is due to a large time-correlated noise signal.

these signals are primarily from local noise processes that are not removed in the weighting. Overall, the weighting greatly improves the spatial consistency of the modeled transient signal. The reconstructed transient time series corresponding to timescales of approximately 4 months show that most time-correlated signals that are not persistent over multiple stations are smoothed over in the spatially weighted solution, thus isolating the signal due to the thrust event (Figure 11).

5. Slow Slip Events in Cascadia

Continuous GPS measurements above the Cascadia subduction zone have revealed episodic slow slip events located deep on the plate interface that are accompanied by subduction-related tremor signals [*Rogers and Dragert*, 2003; *Szeliga et al.*, 2008]. These slow slip events exhibit a fairly persistent quasi-periodicity in this region ($\approx$ 14 months) and inform our understanding of the fault physics and frictional properties through the slow slip location, amplitude, and timing. However, the periodicity and amount of slip for each transient event are both spatially and temporally variable and a priori unknown. Inference of these slow slip properties can be obtained with precise measurements of the surface deformation field over time.

To test the temporal transient detection capabilities of our proposed method, we use daily GPS solutions for the east component of station ALBH located within the Pacific Northwest Geodetic Array (PANGA). The data cover the time span from 2005 to mid-2012 and were processed by the Scripps Orbit and Permanent Array Center (SOPAC) with regional filtering applied to remove common mode errors [*Williams et al.*, 2004]. Known offsets due to hardware changes were removed before analysis. For this work, we examine the east component of the data since the surface deformation in this area occurs primarily in an east-west fashion. Manual inspection of the time series reveals at least six distinct slow slip events. As in the synthetic example, the dictionary $\mathbf{G}$ consists of seasonal, secular, and transient displacement functions. We uniformly subdivide the 7.5 year time series into 256, 128, 64, 32, 16, 8, and 4 intervals to construct B^i-splines of timescales of $\approx$ 3, 6, 12, 25, 52, 113, and 263 weeks, respectively. Slow slip events typically have recorded durations of 3 weeks [*Rogers and Dragert*, 2003]. Both the regularized least squares approach and the Gibbs sampler are used to estimate the dictionary coefficients, where the latter is used to construct the full posterior distribution. Cross validation was performed to select the optimal penalty parameter λ (Figure 12). The increased number of transient events recorded in the ALBH time series favored a smaller penalty parameter than the synthetic data.

As was seen with the synthetic time series case, a challenge for transient detection is correctly estimating the contribution from seasonal and secular processes. Our estimates of these steady processes ultimately govern the magnitude of the detected transient events and their overall interpretation in terms of onset times and durations. Using our approach, removal of the estimated seasonal and secular displacements results in a modeled transient evolution with distinct, step-like motions corresponding to the slow slip events (Figure 13a). By comparing the structure of these transients with those presented in previous studies

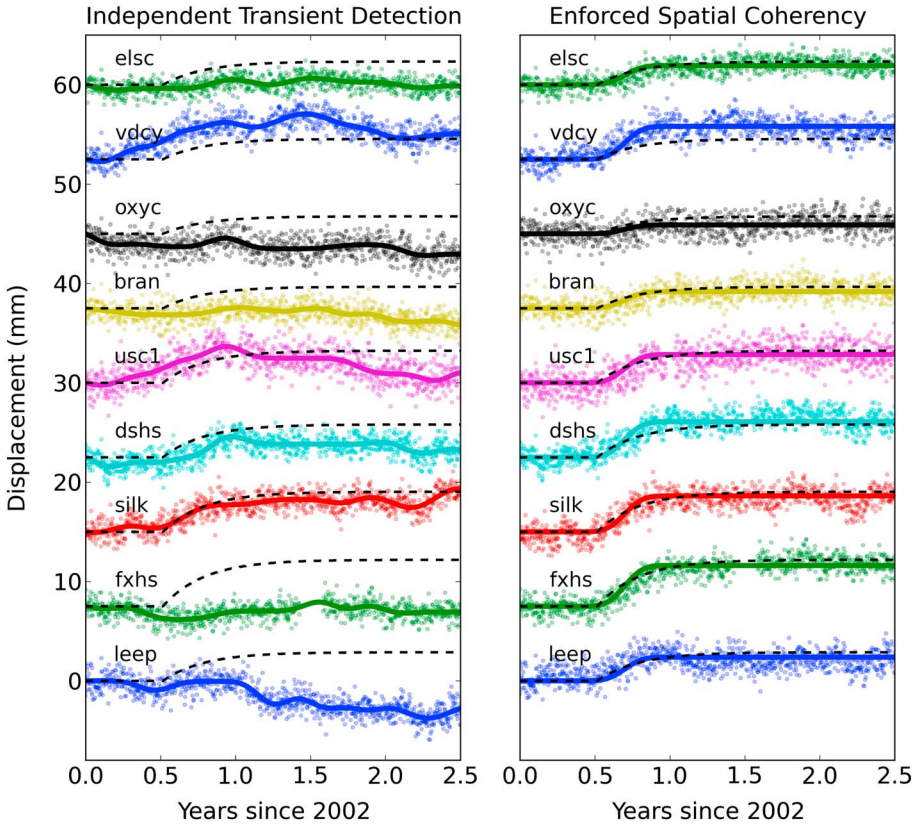

Figure 11. Reconstructed north component transient time series for SCEC Phase III Set D stations near the Santa Monica fault. The data (dots) have the estimated secular and seasonal signals removed to match the reconstructed signal (bold line). Dashed black lines show the true ground signal for each station. (left) Without spatial weighting, the transient signal is corrupted by time-correlated colored noise. (right) With spatial weighting, the signal from the thrust event is isolated.

[e.g., *Szeliga et al.*, 2008], we can observe a distinctly different interpretation of the relative contributions from the secular rate and transient events. Here removal of the estimated secular rate results in slow slip events that are modeled as displacement phenomena that occur in an otherwise quiescent field. Other studies have presented the transient time series in a more sawtooth-like fashion in which the station moves slowly eastward in the interevent period (superposed on the long-term plate rate) before moving rapidly westward. In our framework, such an interpretation would require a positive, longer duration B^i-spline before each event, followed by a short duration, negative B^i-spline to model the actual event. This model would be unfavorable due to the sparse regularization which penalizes the total number of nonzero B^i-splines. The optimal model is the one with the most compact representation, which we can confirm from the scalogram of the B^i-spline coefficients (Figure 13c). In other words, the sparsity-promoting regularization allows us to automatically estimate the interevent secular rate rather than an average rate that mixes the contributions from the secular rate and transient processes.

Here all of the slow slip events are modeled either by 3 or 7 week B^i-splines, although increasing the penalty λ would start to oversmooth the data by selecting longer 14 week B^i-splines. The posterior data covariance, obtained using the procedure outlined in section 2.2, shows stronger covariances between modeled displacements in the interevent period, which is a consequence of the finite support of the B^i-splines (Figure 13b). Similarly, the striping in the off-diagonal terms during the interevent period is also due to the construction of the temporal dictionary since the modeled value at one observation epoch will covary with the other observations depending on the B^i-spline coefficients.

5.1. Dictionary Covariances

For all slow slip events, there are strong covariances between B^i-splines that have time-coincident centroids, and the strongest covariances are between 3-week and 7-week B^i-splines. As with the synthetic data, the

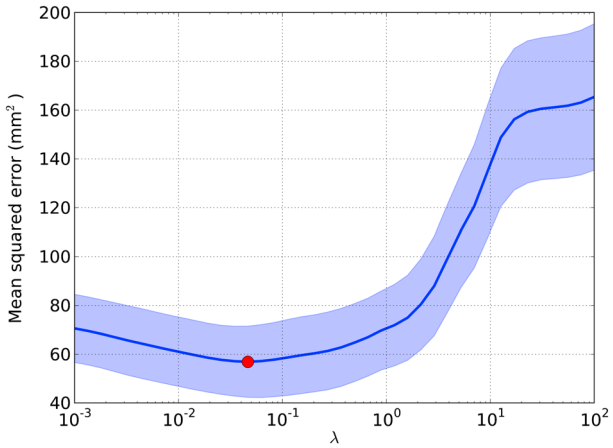

Figure 12. *K*-fold cross-validation results for selection of the penalty parameter λ for the east-component ALBH time series. Cross validation was performed on three independent, equally sized subsets. The blue line shows the mean testing error and the shaded region denotes the standard deviation.

strength of the covariance is primarily determined by the value of λ which moves the areas of high posterior probability along a fairly defined trajectory. The covariance behavior between 3 week and 7 week B^i-splines with time centroids corresponding to a slow slip event in January 2007 shows that for low values of λ, the joint posterior distribution is fairly diffuse and exhibits large variances with the probability peaking over the shorter timescale B^i-spline (Figure 14). However, we can still observe a tail in the joint distribution that points toward a nonzero value for the longer-timescale B^i-spline. Increasing λ is equivalent to shrinking the prior densities for all B^i-splines, causing the joint distribution to move along the vertical axis defined by small values for the longer-timescale B^i-spline and decreasing values for the shorter timescale B^i-spline. Once λ exceeds a certain value, the high-probability areas of the joint distribution transition toward the longer-timescale B^i-spline via the dominant covariance direction. Eventually, for high enough λ, the posterior distribution is forced to be identical to the prior distribution with a peak at the origin which indicates that the signal has been completely smoothed over.

5.2. Spatial Sparsity Weighting

As with the SCEC synthetic data set, we apply the spatial sparsity weighting approach in section 4 to station ALBH and 32 nearby GPS stations located within the Cascadia region (Figure 15). We use raw daily positions from the PANGA network processed by the Central Washington University Geodesy Lab

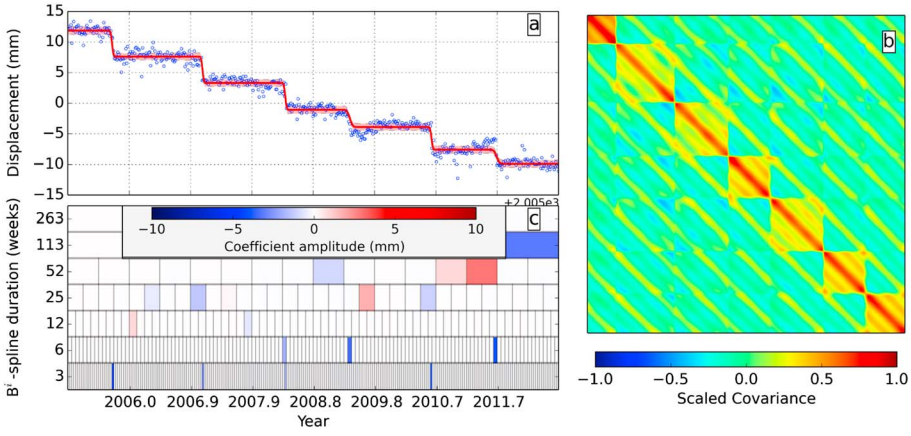

Figure 13. Transient detection results for the east-component ALBH time series. (a) GPS data with estimated seasonal and secular signals removed (black circles) and estimated transient signal from a reduced dictionary corresponding to a 99% variance reduction (red line). Shaded area denotes 5σ uncertainties on predicted displacement using the method in section 2.2. (b) Posterior data covariance matrix of the time series fit scaled to unity. (c) Scalogram for all B^i-spline coefficients. Comparison with the time series shows the direct correspondence between the slow slip events and the nonzero 3 and 7 week B^i-splines. No strong long-term transient signals are present.

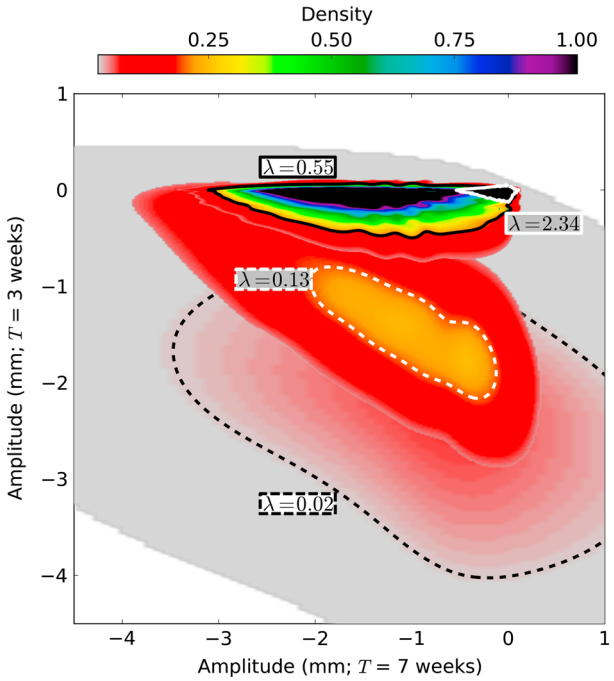

Figure 14. Evolution of the posterior covariance with penalty λ between two centroid-coincident B^i-splines corresponding to the January 2007 slow slip event. Solid and dashed lines mark lines of constant posterior density in the following manner: black dashed → 0.03, white dashed → 0.2, black solid → 0.2, and white solid → 0.95. For small λ, the joint distribution is diffuse with larger amplitude samples drawn for the shorter timescale B^i-spline. As λ increases, the distribution shrinks and moves along the vertical axis. Eventually, the distribution transitions to larger values for the longer timescale B^i-spline and small values for the shorter timescale B^i-spline. For the largest λ, the samples are tightly clustered around the origin.

(http://www.geodesy.cwu.edu/). To remove common mode network errors, we use the spatiotemporal filtering method of *Dong et al.* [2006]. We first apply the temporal transient detection method independently for each station and displacement component and remove the total modeled displacements from the data. From the detrended and demeaned residuals for all stations, principal component analysis is performed to estimate the largest principal component corresponding to common mode error. The residuals

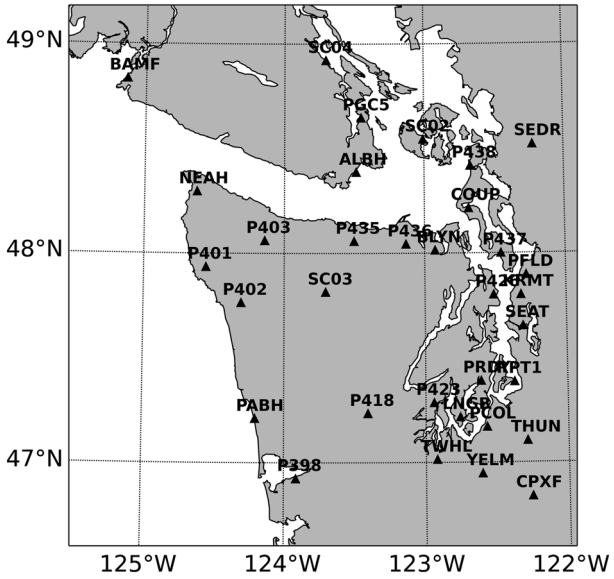

Figure 15. Map of select GPS stations from the PANGA network used in the spatial sparsity weighting.

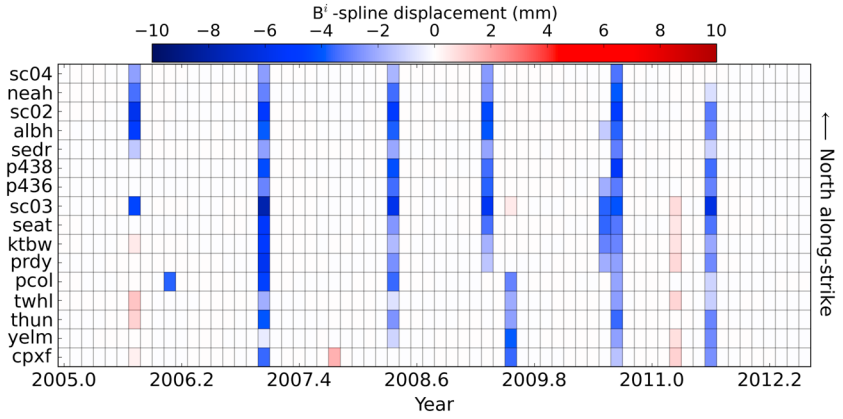

Figure 16. Amplitudes corresponding to 3 month B^i-splines used to fit east-component time series data from Cascadia GPS stations. Stations are arranged by increasing latitude, and spatial sparsity weighting was applied to isolate spatially coherent B^i-splines. The episodic slow slip events are easily visualized, and longer timescale propagation effects can be seen in two of the events (mid-2009 and late-2010).

with the common mode signal removed are then added back to the modeled displacements to obtain the filtered data.

We divide the detection procedure into two stages: (1) an initial detection phase where the dictionary is populated with longer timescale B^i-splines ($\tau_k > 1$ month) and seasonal and secular processes and (2) an

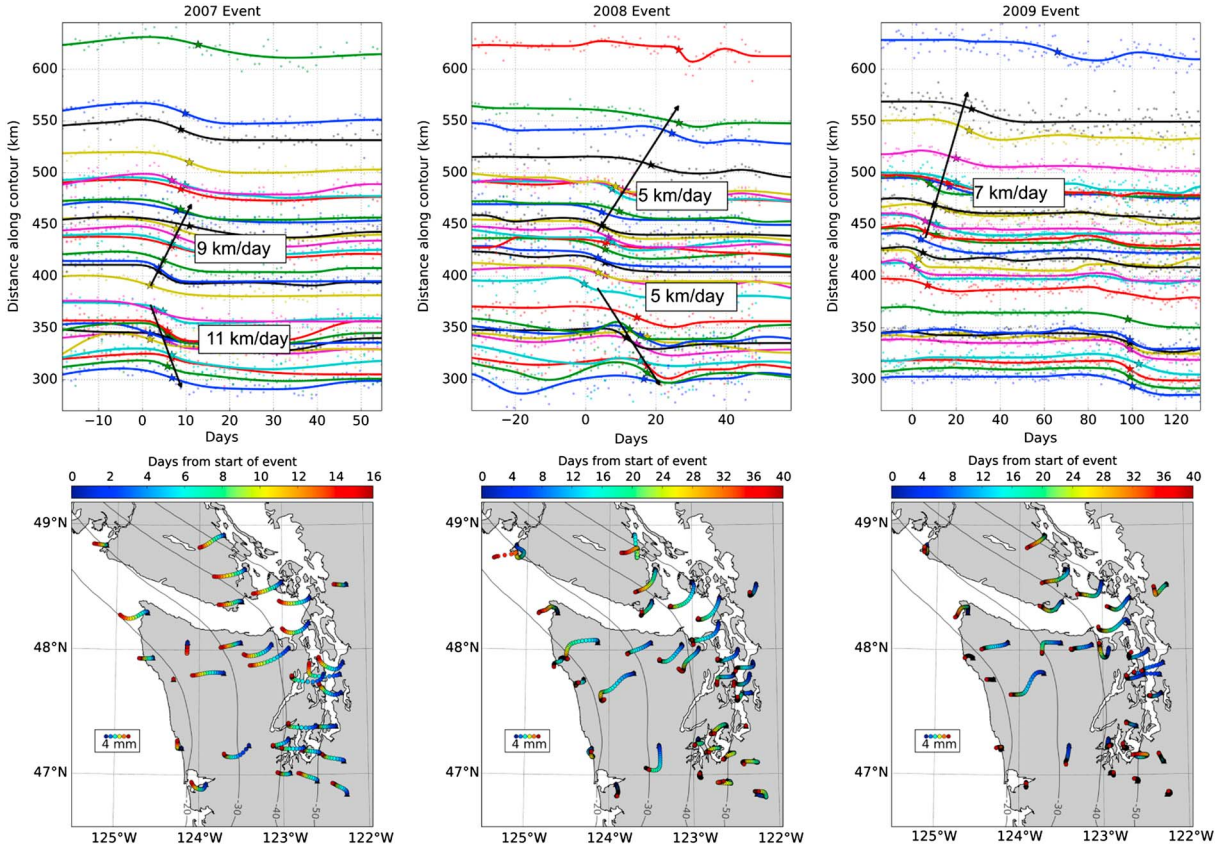

Figure 17. Transient detection with spatial sparsity weighting for 33 GPS stations within the Cascadia region. Three separate slow slip events are shown: (left) January 2007, (middle) mid-2008, and (right) mid-2009. (top) The normalized east-component displacements corresponding to the slow slip events. The displacement time series are ordered in distance along the 40 km depth contour of the downgoing slab model from *McCrory et al.* [2004]. (bottom) The transient ground motion during each slow slip event where the marker color indicates days from the start of the event.

analysis phase where we select only a year of data encompassing the detected slow slip events and the dictionary is populated exclusively with B^i-splines ($\tau_k > 5$ days). Prior to the analysis phase, we remove the estimated signals from seasonal and secular processes, as well as detected transients with timescales longer than 1 year. While the ≈ 1–3 month B^i-splines used in the detection phase will oversmooth many slow slip events, we still gain information about the dominant event durations and onset times (Figure 16). The detailed analysis phase can then reveal any subtle propagation behaviors of detected slow slip events and constrain the spatial extent of the transient surface strain. In cases where the data size of the time series are relatively small, the detection phase may be skipped. We utilize it for the Cascadia data because the timescales of the slow slip events are considerably shorter than the time extent of the data.

For the analysis phase, we focus on three slow slip events: (1) January 2007, (2) mid-2008, and (3) mid-2009 (Figure 17). For all three events, we can observe several characteristics common to all events. First, slow slip tends to nucleate on the eastern Olympic peninsula near the Seattle area. The slow slip front (determined by the GPS stations' peak ground velocities) propagates bilaterally with total ground motion larger at stations higher than ≈ 47.5°N. The southern stations tend to show transient motion near the cessations of the slow slip events, which may suggest a change in frictional properties or fault geometry around 47°N. For the 2007 event, the GPS stations show distinctly different displacement azimuths with larger overall velocities (larger displacements in a shorter time period), which may indicate a different slipping area in the underlying thrust fault [*Wech et al.*, 2009]. Additionally, the propagation speed is significantly faster than either the 2008 or 2009 events. Ground motions for the 2008 and 2009 events initiate at a higher latitude, and the northern stations exhibit stronger southward motion in the first few days of each event. Similarly, station P418 in the southern section shows strong southward motion in the first 10 days of each event. Comparison with previous studies estimating static slip for the 2007 and 2008 events [e.g., *Aguiar et al.*, 2009; *Wech et al.*, 2009] reveals that the areas of peak static slip on the underlying fault are closely located to GPS stations with the first ground motions for each event. This relationship suggests that the nucleation zones on the fault experience the highest cumulative slip. For all three events, we can also observe that many of the stations cease their north-south motions in the second half of the event and move primarily in an east-west fashion.

6. Discussion

Generally, transient signal reconstructions are remarkably consistent between the least squares optimization with sparsity-inducing regularization and the Bayesian sampling approach with a Laplace prior. Moreover, the Bayesian approach does not explicitly implement reweighting for enhancing sparsity as mentioned in section 2.1 and in *Candés et al.* [2008] and can be viewed as solving a single iteration of the sparsity-regularized least squares problem. Even so, the scalogram comparison between the two different approaches (Figure 6) confirms that Gibbs sampling can recover the same B^i-splines corresponding to transient signals while zeroing out the contributions from other B^i-splines. The fact that we use the mean of the posterior distribution rather than the mode as a point estimate is because we have assigned a squared-error "loss-function" to the data misfit [*Tarantola*, 2005]. In a study by *Hans* [2009], it was shown that the posterior mode for a Laplace prior can be interpreted as a limiting case corresponding to a zero-one loss function, which resembles the ℓ_0-norm penalty. However, using the mode does not capture the best point estimate when the marginal posteriors are skewed, as was observed in Figure 7. While the penalty λ controls the degree of skewness of the marginal posteriors, it is always optimal to use the mean as a point estimate when a Gaussian data misfit is used.

In section 3.3, we demonstrated the successful detection of multiple transient signals with a relatively sparse data set. In practical applications, these results suggest that a minimum of ≈ 3–4 data points spanning the transient process would be required to recover its signal. For daily GPS solutions, we could therefore potentially recover very rapid processes, provided that the signal-to-noise ratio was high enough and the dictionary contained B^i-splines with equally short timescales. On the opposite end of the spectrum, time series with much coarser temporal resolution, such as an InSAR time series, could still detect transient signals of durations comparable to the time spacing between data acquisitions. For both of these cases, the penalty parameter could be chosen using *K*-fold cross validation, which was demonstrated to favor lower penalties to allow for robust reconstruction of very small signals.

When two B^i-splines are centered over the same time epoch and can both fit the data reasonably well (as in section 5.1), lower values of the penalty parameter λ will tend to allow larger amplitudes for the shorter

timescale B^i-spline, allowing the reconstructed transient signal to contain higher-frequency components. Increasing λ is equivalent to shrinking the prior probabilities of the B^i-splines, forcing the solution to favor the longer-timescale B^i-spline and smoothing the reconstructed signal. For real geodetic time series, this behavior implies that successful detection of very short duration transient signals would most likely require a fairly low value for λ which could be chosen through K-fold cross validation. The side effect of using a small λ is that the reconstructed signal would have a higher probability of being corrupted by colored noise processes in the data. If the signal of interest is known to have a longer duration than typical colored noise, then one could safely choose a higher value of λ to favor longer-timescale B^i-splines. An alternative approach would be to construct the dictionary such that it included only longer-timescale B^i-splines, limiting the effective data resolution of the model.

The spatial sparsity weighting discussed in section 4 was effective for minimizing the effects of local ground motion and colored noise. However, the performance of the spatial weighting is inherently dependent on the density of the network and the value of the correlation length used in the distance weighting scheme. Larger correlation lengths will tend to reconstruct long-wavelength deformation fields while smaller correlation lengths will reconstruct more local ground motions. The variable correlation length approach used in this work has the advantage of only reconstructing signals that are resolvable by the geodetic network and minimizing false detections of spurious transients due to data noise. Our approach of prescribing the correlation length, L_0^j, to be the average distance from station j to the nearest three or four stations will be affected by the addition of a new station in the vicinity of station j. However, the addition of stations can only decrease the correlation length. From equation (7), the weighting function would decrease in strength at a given distance, corresponding to an increase in the effective spatial resolution at the current location. Likewise, station removals would increase the correlation length and decrease the effective spatial resolution. For both cases, changes in the station distribution would only affect the spatial reconstructions in the vicinity of the station addition/removal, and the solutions for the remaining stations in the network will be unaffected. In terms of solution stability, the weighting scheme is thus stable and adaptive to changes in network geometry.

The spatial weighting can also be formulated in a Bayesian sense by using the cascading approach discussed earlier and in *Minson et al.* [2013]. We can write the posterior distribution of the B^i-spline coefficients for the kth GPS station as $p\left(\mathbf{m}_k|\mathbf{d}_k\right) \propto p\left(\mathbf{d}_k|\mathbf{m}_k\right) p\left(\mathbf{m}_k\right)$. By assuming that the terms in $\mathbf{m}_k$ are correlated with the same terms in N surrounding stations, we can setup a joint estimation problem, where $\mathbf{M} = \left[\mathbf{m}_1, \ldots, \mathbf{m}_N\right]$ and $\mathbf{D} = \left[\mathbf{d}_1, \ldots, \mathbf{d}_N\right]$. Then, for the kth station, where $k \leq N$, the joint posterior distribution would be as follows:

$$p\left(\mathbf{m}_k|\mathbf{D}\right) \propto \left[\prod_{j=1, j\neq k}^{N} p\left(\mathbf{d}_j|\mathbf{m}_j\right) p\left(\mathbf{m}_j\right)\right] p\left(\mathbf{d}_k|\mathbf{M}\right) p\left(\mathbf{M}\right)$$
$$\propto \left[\prod_{j=1, j\neq k}^{N} p\left(\mathbf{m}_j|\mathbf{d}_j\right)\right] p\left(\mathbf{d}_k|\mathbf{M}\right) p\left(\mathbf{M}\right). \quad (8)$$

The prior distribution $p\left(\mathbf{M}\right)$ would account for spatial coherency between B^i-spline amplitudes by incorporating a prior covariance matrix with nonzero off-diagonal components for the elements in $\mathbf{M}$. As before, the structure of the prior covariance matrix would depend on some form of distance weighting between stations and would be recomputed for each kth station. Then, independent Bayesian sampling runs performed for each station would be combined to form the product in brackets in equation (8), and the final posterior distribution $p\left(\mathbf{m}_k|\mathbf{D}\right)$ would be sampled by constructing an appropriate likelihood function $p\left(\mathbf{d}_k|\mathbf{M}\right)$.

The reconstructed transient motions and propagation characteristics of the Cascadia slow slip events agree well with independent studies of tremor space-time propagation [e.g., *Houston et al.*, 2011]. The geodetic propagation speeds estimated here fall within the range of the tremor propagation speeds, and the nucleation zones for tremor activity correspond well to the first ground motions for each event. Inspection of the reconstructed transient time series for each station (Figures 13 and 17) indicate that the reconstructions do not suffer from significant oversmoothing. The inclusion of short timescale B^i-splines in the temporal dictionary $\mathbf{G}$ permit us to detect and model very subtle ground motions due to slow slip while the spatial sparsity weighting prevents us from overfitting local ground motions.

Finally, the construction of the transient detection method as a linear model results in very efficient time series processing which can be easily scaled for large geodetic networks. As with any linear model, the effi-

ciency of the linear algebra routines used for performing matrix-vector operations will greatly impact the detection speed and can be enhanced with straightforward parallelization. An alternative to batch estimation of the coefficients is a recursive approach where estimation is performed for each data point in a sequential fashion. The ℓ_1-norm penalty would be enforced as a pseudo-measurement with an associated Kalman gain that tracks the ℓ_1-norm of the current state of coefficients [*Carmi et al.*, 2010]. A recursive approach would require storing only a single row of the dictionary $\mathbf{G}$ into computer memory, limiting the computational cost of the estimation and allowing an arbitrarily large number of dictionary elements. Furthermore, the update would be very fast for a single observation and would provide a real-time transient detection capability.

7. Conclusions

We demonstrated successful transient detection for a single geodetic time series by estimating the coefficients corresponding to a highly overcomplete dictionary (design matrix) of integral B-splines that resemble transient events of various timescales and start times. Regularizing the estimation procedure with an ℓ_1-norm on the coefficients favors *sparse* solutions, limiting the number of B^i-splines needed to describe transient events while still providing a good fit to the data. For GPS networks with sufficient station density, we can perform the regularization simultaneously across the whole network with a distance weighting procedure to enhance signals that are spatially coherent over a given length scale. The reconstructed time series essentially resemble smoothed versions of the input data but with additional critical information regarding transient event start times and durations. The temporal resolution of the detection method is only limited by the signal-to-noise ratio of the data and the smallest timescale B^i-splines included in the dictionary. The spatial resolution is then limited by the density of the geodetic network. Detection sensitivity is controlled by the penalty parameter on the ℓ_1-norm, which can be robustly and automatically chosen with cross validation. We also presented a Gibbs sampling approach to construct the full posterior distribution of each element in the dictionary which, in addition to stand-alone point estimates of the coefficients via the posterior means, provides quantifiable uncertainties on the coefficients and valuable insight into trade-offs between dictionary elements. In the absence of a priori knowledge about transient event start times, durations, and physical sources, this method automatically and efficiently determines the most dominant signals in a time series in a compact and interpretable manner.

Acknowledgments

We thank two anonymous reviewers for improving the quality of this paper. Bryan Riel is supported by a NASA Earth and Space Science Fellowship. GPS data used in this paper can be found at the SOPAC data archive, http://sopac.ucsd.edu/dataArchive/, and the Central Washington University archive, http://www.geodesy.cwu.edu/.

References

Agnew, D. C. (2013), Realistic simulations of geodetic network data: The Fakenet package, *Seismol. Res. Lett.*, *84*(3), 426–432.

Agram, P., R. Jolivet, B. Riel, Y. Lin, M. Simons, E. Hetland, M.-P. Doin, and C. Lasserre (2013), New radar interferometric time series analysis toolbox released, *Eos Trans. AGU*, *94*(7), 69–70.

Aguiar, A. C., T. I. Melbourne, and C. W. Scrivner (2009), Moment release rate of Cascadia tremor constrained by GPS, *J. Geophys. Res.*, *114*(B7), B00A05, doi:10.1029/2008JB005909.

Beck, J. L., and K.-V. Yuen (2004), Model selection using response measurements: Bayesian probabilistic approach, *J. Eng. Mech.*, *130*(2), 192–203.

Berardino, P., G. Fornaro, R. Lanari, and E. Sansosti (2002), A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms, *IEEE Trans. Geosci. Remote Sens.*, *40*(11), 2375–2383.

Boyd, S., and L. Vandenberghe (2004), *Convex Optimization*, Cambridge Univ. Press, New York.

Candés, E. J., and M. B. Wakin (2008), An introduction to compressive sampling, *IEEE Signal Process Mag.*, *21*, 21–30.

Candés, E. J., M. B. Wakin, and S. P. Boyd (2008), Enhancing sparsity by reweighted ℓ_1 minimization, *J. Fourier Anal. Appl.*, *14*, 877–905.

Carmi, A., P. Gurfil, and D. Kanevsky (2010), Methods for sparse signal recovery using Kalman filtering with embedded pseudo-measurement norms and quasi-norms, *IEEE Trans. Signal Process.*, *58*(4), 2405–2409.

Chen, S. S., D. L. Donoho, and M. A. Saunders (1998), Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.*, *20*(1), 33–61.

Ching, J., M. Muto, and J. Beck (2006), Bayesian linear structural model updating using Gibbs sampler with modal data, *Comput.-Aided Civ. Infrastruct. Eng.*, *21*(4), 242–257.

Covello, F., F. Battazza, A. Coletta, E. Lopinto, C. Fiorentino, L. Pietranera, G. Valentini, and S. Zoffoli (2010), COSMO-SkyMed an existing opportunity for observing the Earth, *J. Geodyn.*, *49*(3), 171–180.

Dong, D., P. Fang, Y. Bock, F. Webb, L. Prawirodirdjo, S. Kedar, and P. Jamason (2006), Spatiotemporal filtering using principle component analysis and Karhunen-Loeve expansion approaches for regional GPS network analysis, *J. Geophys. Res.*, *111*, B03405, doi:10.1029/2005JB003806.

Donoho, D. L. (2006), For most large underdetermined systems of linear equations the minimal ℓ^1-norm solution is also the sparsest solution, *Commun. Pure Appl. Math.*, *59*(6), 797–829.

Evans, E. L., and B. J. Meade (2012), Geodetic imaging of coseismic slip and postseismic afterslip: Sparsity promoting methods applied to the great Tohoku earthquake, *Geophys. Res. Lett.*, *39*(11), L11314, doi:10.1029/2012GL051990.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004), *Bayesian Data Analysis*, Chapman & Hall/CRC, Boca Raton, Fla.

Gholami, A., and H. Siahkoohi (2010), Regularization of linear and non-linear geophysical ill-posed problems with joint sparsity constraints, *Geophys. J. Int.*, *180*, 871–882.

Gomberg, J., et al. (2010), Slow-slip phenomena in Cascadia from 2007 and beyond: A review, *Geol. Soc. Am. Bull.*, *122*(7–8), 963–978.

Hans, C. (2009), Bayesian lasso regression, *Biometrika*, *96*(4), 835–845.

Hetland, E., P. Musé, M. Simons, Y. Lin, P. Agram, and C. DiCaprio (2012), Multiscale InSAR time series (MInTS) analysis of surface deformation, *J. Geophys. Res.*, *117*, B02404, doi:10.1029/2011JB008731.

Houston, H., B. G. Delbridge, A. G. Wech, and K. C. Creager (2011), Rapid tremor reversals in Cascadia generated by a weakened plate interface, *Nat. Geosci.*, *4*(6), 404–409.

Hsu, Y.-J., S.-B. Yu, M. Simons, L.-C. Kuo, and H.-Y. Chen (2009), Interseismic crustal deformation in the Taiwan plate boundary zone revealed by GPS observations, seismicity, and earthquake focal mechanisms, *Tectonophysics*, *479*(1), 4–18.

Hung, H.-K., and R.-J. Rau (2013), Surface waves of the 2011 Tohoku earthquake: Observations of Taiwan's dense high-rate GPS network, *J. Geophys. Res. Solid Earth*, *118*, 332–345, doi:10.1029/2012JB009689.

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.

Jolivet, R., R. Grandin, C. Lasserre, M.-P. Doin, and G. Peltzer (2011), Systematic InSAR tropospheric phase delay corrections from global meteorological reanalysis data, *Geophys. Res. Lett.*, *38*(17), L17311, doi:10.1029/2011GL048757.

Kato, A., K. Obara, T. Igarashi, H. Tsuruoka, S. Nakagawa, and N. Hirata (2012), Propagation of slow slip leading up to the 2011 M_w 9.0 Tohoku-Oki earthquake, *Science*, *335*, 705–708.

Langbein, J. (2004), Noise in two-color electronic distance meter measurements revisited, *J. Geophys. Res.*, *109*, B04406, doi:10.1029/2003JB002819.

Langbein, J. O. (2003), Deformation of the Long Valley Caldera, California: Inferences from measurements from 1988 to 2001, *J. Volcanol. Geotherm. Res.*, *127*, 247–267.

Lohman, R. B., and J. R. Murray (2013), The SCEC geodetic transient-detection validation exercise, *Seismol. Res. Lett.*, *84*(3), 419–425.

Mallat, S. (1989), A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intell.*, *11*(7), 674–693.

Masterlark, T., and Z. Lu (2004), Transient volcano deformation sources imaged with interferometric synthetic aperture radar: Application to Seguam Island, Alaska, *J. Geophys. Res.*, *109*, B01401, doi:10.1029/2003JB002568.

McCrory, P. A., J. L. Blair, D. H. Oppenheimer, and S. R. Walter (2004), *Depth to the Juan de Fuca Slab Beneath the Cascadia Subduction Margin: A 3-D Model for Sorting Earthquakes*, U.S. Department of the Interior, U.S. Geological Survey, Denver, Colo.

McGuire, J. J., and P. Segall (2003), Imaging of aseismic fault slip transients recorded by dense geodetic networks, *Geophys. J. Int.*, *155*, 778–788.

Minson, S., M. Simons, and J. Beck (2013), Bayesian inversion for finite fault earthquake source models I—Theory and algorithm, *Geophys. J. Int.*, *194*, 1701–1726.

Miyazaki, S., J. J. McGuire, and P. Segall (2003), A transient subduction zone slip episode in southwest Japan observed by the nationwide GPS array, *J. Geophys. Res.*, *108*(B2), 2087, doi:10.1029/2001JB000456.

Park, T., and G. Casella (2008), The Bayesian lasso, *J. Am. Stat. Assoc.*, *103*(482), 681–686.

Pritchard, M., and M. Simons (2004), An InSAR-based survey of volcanic deformation in the Central Andes, *Geochem. Geophys. Geosyst.*, *5*(2), Q02002, doi:10.1029/2003GC000610.

Rogers, G., and H. Dragert (2003), Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip, *Science*, *300*, 1942–1943.

Sagiya, T. (2004), A decade of GEONET: 1994-2003-the continuous GPS observation in Japan and its impact on earthquake studies, *Earth Planets Space*, *56*(8), xxix–xli.

Szeliga, W., T. Melbourne, M. Santillan, and M. Miller (2008), GPS constraints on 34 slow slip events within the Cascadia subduction zone, 1997–2005, *J. Geophys. Res.*, *113*, B04404, doi:10.1029/2007JB004948.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial and Applied Mathematics, Philadelphia, Pa.

Taylor, H., S. Banks, and J. McCoy (1979), Deconvolution with the ℓ_1-norm, *Geophysics*, *44*(1), 39–52.

Tibshirani, R. (1996), Regression shrinkage and selection via the lasso, *J. Roy. Stat. Soc. B*, *58*(1), 267–288.

Tipping, M. E. (2004), Bayesian inference: An introduction to principles and practice in machine learning, in *Advanced Lectures on Machine Learning*, vol. 3176, edited by O. Bousquet, U. Luxburg, and G. Rätsch, pp. 41–62, Springer, Berlin, Heidelberg.

Wallace, L. M., and J. Beavan (2010), Diverse slow slip behavior at the Hikurangi subduction margin, New Zealand, *J. Geophys. Res.*, *115*(B12), B12402, doi:10.1029/2010JB007717.

Wech, A. G., K. C. Creager, and T. I. Melbourne (2009), Seismic and geodetic constraints on Cascadia slow slip, *J. Geophys. Res.*, *114*(B10), B10316, doi:10.1029/2008JB006090.

Williams, S., Y. Bock, and P. Fang (1998), Integrated satellite interferometry: Tropospheric noise, GPS estimates and implications for interferometric synthetic aperture radar products, *J. Geophys. Res.*, *103*(B11), 27,051–27,067.

Williams, S. D., 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.