submitted to *Geophys. J. Int.*

Revealing transient strain in geodetic data with Gaussian process regression

T. T. Hines[1] and E. A. Hetland[1]

[1] *Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI, USA*

SUMMARY

Transient strain rates derived from GNSS data can be used to detect and understand geophysical ~~phenomena~~ processes such as slow slip events and postseismic deformation. Here we propose using Gaussian process regression (GPR) as a tool for estimating transient strain rates from GNSS data. GPR is a non-parametric, Bayesian method for interpolating scattered data. ~~Transient strain rates estimated with GPR have meaningful uncertainties, allowing geophysical signal to be easily discerned from noise.~~ In our approach, we assume a stochastic prior model for transient displacements. The prior describes how much ~~one expects~~ we expect transient displacements to covary spatially and temporally. A posterior estimate of transient strain rates is obtained by differentiating the posterior displacements~~. One limitation with GPR is that it is not robust against outliers, so we introduce a pre-processing method for detecting and removing outliers in~~ , which are formed by conditioning the prior with the GNSS data. As a demonstration, we use GPR to detect transient strain resulting from slow slip events in ~~Cascadia~~the Pacific Northwest. Maximum likelihood methods are used to constrain a prior model for transient displacements in this region. The temporal covariance of our prior model is described by a compact Wendland covariance function, which significantly reduces the computational burden that can be associated with GPR. Our results reveal the spatial and temporal evolution of strain from slow slip events. We verify that the transient strain estimated with GPR is in fact geophysical signal by comparing it to the seismic tremor that is associated with ~~Cascadia~~ Pacific Northwest slow slip events.

Key words: XXX – XXX – XXX – XXX.

1 INTRODUCTION

Crustal strain rates are fundamentally important quantities for assessing seismic hazard. Knowing where and how quickly strain is accumulating gives insight into where we can expect stored elastic energy to be released seismically. Consequently, secular crustal strain rates estimated from GNSS data have been used to constrain seismic hazard models such as UCERF3 (Field et al. 2014). Transient crustal strain, which is caused by geophysical phenomena such as slow slip events (SSEs) or postseismic deformation, is also relevant for assessing seismic hazard. While transient strain itself is not damaging, there is a risk that it can trigger major earthquakes (Roeloffs 2006; Freed & Lin 2001). Dense networks of continuous GNSS stations, such as the Plate Boundary Observatory (PBO), make it possible to identify transient strain with high fidelity. Developing and improving upon methods for deriving secular and transient strain from GNSS data is an active area of research.

Most methods for estimating strain rates from GNSS data assume some parametric form ~~of~~ for the deformation signal. The simplest method for estimating secular strain rates assumes that GNSS derived velocities can be described with a first-order polynomial (i.e~~., having constant deformation gradients~~ , the deformation gradients are constant) over some subnetwork of the GNSS stations (e.g., Feigl et al. 1990; Murray & Lisowski 2000). The components of the strain rate tensor for each subnetwork are then determined ~~through a~~ from the least squares fit to the observations. The assumption that deformation gradients are spatially uniform is not appropriate when subnetworks span too large of an area. To help overcome this deficiency, Shen et al. (1996, 2015) used an inverse distance weighting scheme, in which the estimated strain rate at ~~some~~ a point is primarily controlled by observations at nearby stations. ~~However, the methods described in Shen et al. (1996, 2015) are still formulated by assuming that the deformation gradients are uniform over the entire network. The errors in this assumption manifest as implausibly low formal uncertainties for the estimated strain rates~~The method of Shen et al. (1996, 2015) can be viewed as a form of local least squares regression with a first-order polynomial (e.g., Hastie et al. 2009, sec. 6). Other methods for estimating secular strain rates have parameterized GNSS derived velocities with bi-cubic splines (Beavan & Haines 2001), spherical wavelets (Tape et al. 2009), and elastostatic Green's functions (Sandwell & Wessel 2016). The type of basis functions and the number of degrees of freedom for a parameterization~~can be subjective~~, which are often chosen subjectively, have a strong influence on the strain solution. If there are too few degrees of freedom in the parameterization, then estimated strain rates will be biased and the uncertainties will be underestimated. On the other hand, if there are too many degrees of freedom, then there will not be any coherent features in the estimated strain rates. The methods described by Beavan & Haines (2001) and Tape et al. (2009) also require the user to specify penalty parameters that control a similar trade-off between bias and variance in the solution. One could parameterize deformation with a physically motivated model of interseismic deformation (e.g., Meade & Hager 2005; McCaffrey et al. 2007). In such models the lithospheric rheology and fault geometries are assumed to be known. Any errors in the assumed physical model could result in biased strain estimates and underestimated formal uncertainties.

The aforementioned studies are concerned with estimating secular strain rates. In recent years the Southern California Earthquake Center (SCEC) community has shown interest in developing methods for detecting transient strain. SCEC supported a transient detection exercise (Lohman & Murray 2013), where several research groups tested their methods for detecting transient geophysical signal with a synthetic GNSS dataset. Among the methods tested were the Network Strain Filter (NSF) (Ohtani et al. 2010) and the Network Inversion Filter (NIF) (Segall & Mathews 1997). The NSF uses a wavelet parameterization to describe the spatial component of geophysical signal. The NIF, which is intended for imaging slow fault slip from geodetic data, uses the elastic dislocation Green's functions from Okada (1992). For the NSF and NIF, the time dependence of the geophysical signal is modeled as integrated Brownian motion. The method described in Holt & Shcherbenko (2013) was also tested in the SCEC transient detection exercise, which calculates strain rates using a bi-cubic spatial parameterization of displacements between time epochs. Holt & Shcherbenko (2013) defined a detection threshold based on the strain rate magnitude, and ~~below~~ we demonstrate that this is indeed an effective criterion for identifying geophysical signal. For the same reasons descibed above, the transient deformation and corresponding uncertainties estimated by these methods can be biased by the chosen spatial parameterization. It is then difficult to distinguish signal from noise with these methods, which limits their utility for transient detection.

Here we propose using Gaussian process regression (GPR) (Rasmussen & Williams 2006) to estimate transient strain from GNSS data. GPR is closely related to kriging (Cressie 1993) and least squares collocation (Moritz 1978). The latter has been used by Kato et al. (1998) and El-Fiky & Kato (1998) to estimate secular strain rates from GNSS data. GPR is a Bayesian, non-parametric method for inferring a continuous signal from scattered data. Since GNSS stations are irregularly spaced and observation times may differ between stations, GPR is an ideal tool for synthesizing discrete GNSS data into a spatially and temporally continuous representation of ~~surface deformation. GPR is closely related~~

~~to kriging (Cressie 1993) and least squares collocation (Moritz 1978). The latter has been used by Kato et al. (1998) and El-Fiky & Kato (1998) to estimate secular strainrates from GNSS data~~transient strain. GPR is Bayesian in that we ~~describe our prior understanding of the geophysical signal with a Gaussian process. A Gaussian process is a normally distributed stochastic processthat is fully defined in terms of a mean function and a positive-definite covariance function. For example, Brownian motion, $B(t)$, is a well known Gaussian processin $\mathbb{R}^1$ which has zero mean and covariance function $\mathrm{cov}(B(t), B(t')) = \min(t, t')$, where $t, t' \geq 0$. If no prior information is available for the geophysical signal~~use a prior model to control the spatial and temporal roughness of the inferred transient strain. The prior is specified as a stochastic process, namely a Gaussian process. If there is no information available to help choose an appropriate prior Gaussian process, then maximum likelihood methods can be used to ~~choose a prior Gaussian process~~ objectively choose one that is most consistent with the ~~observations. We incorporate GNSS observations with the prior to form a posterior estimate of transient strain. The posterior transient strain is also a Gaussian process, and we can use its distribution to confidently discern geophysical signal from noise~~data. We use GPR to infer transient strain resulting from SSEs in ~~Cascadia. Our results demonstrate~~ the Pacific Northwest, demonstrating that GPR is an effective tool for detecting ~~SSEs and revealing how strain evolves during SSEs~~transient geophysical processes.

2 ESTIMATING TRANSIENT STRAIN RATES

We seek a spatially and temporally dependent estimate of transient ~~crustal~~ strain rates. We consider transient strain rates to be any deviation from the secular strain rates~~, and our attention is limited to horizontal strain rates in this study~~. We denote transient ~~crustal~~ strain rates as ~~$\dot{\varepsilon}(p)$~~$\dot{\varepsilon}(p)$, where p represents the ordered pair ~~$(\vec{x}, t)$, $\vec{x}$~~ $(\vec{x}, t)$, $\vec{x} = (x_\mathrm{e}, x_\mathrm{n})$ are spatial coordinates~~in $\mathbb{R}^2$~~, and t is time. ~~We determine $\dot{\varepsilon}(p)$~~ The subscripts "e" and "n" indicate the east and north component, and we assume that the study region is sufficiently small that $\vec{x}$ can be considered a point in a 2-D Cartesian map projection that is aligned with the cardinal directions. We determine $\dot{\varepsilon}$ by spatially and temporally differentiating estimates of transient displacements, ~~$\vec{u}(p)$. We make a prior assumption that~~ which we constrain with GNSS data.

We let $\vec{u}(p) = (u_\mathrm{e}(p), u_\mathrm{n}(p))$ be our prior understanding of transient displacements. Since transient displacements are not precisely known, we cannot consider $\vec{u}$ to be a deterministic function. Instead, $\vec{u}$ is considered to be a stochastic process containing a distribution of functions that could potentially describe transient displacements. Specifically, we let each component of $\vec{u}$ ~~is~~ be a Gaussian process~~,~~

$$u_i(p) \sim \mathcal{N}\left(0, C_{u_i}\right),$$

~~where $C_{u_i}(p,p')$ is~~ . A Gaussian process is a stochastic process whose value at any collection of points can be described with a multivariate normal distribution. That is to say, the random vector $[u_i(p)]_{p\in\boldsymbol{P}}$ has a multivariate normal distribution for any collection of points $\boldsymbol{P}$. A realization of the random vector $[u_i(p)]_{p\in\boldsymbol{P}}$ can be interpretted as a realization of u_i (i.e., a sample function) that is evaluated at $\boldsymbol{P}$. Just as a multivariate normal distribution is fully determined by a mean vector and a covariance matrix, a Gaussian process is fully determined by a mean function and a covariance function~~indicating how we expect $u_i(p)$ to covary with $u_i(p')$. For simplicity, we treat~~ . For example, Brownian motion, $B(t)$, is a well known Gaussian process in $\mathbb{R}^1$ which has the mean function $\mathrm{E}[B(t)] = 0$ and the covariance function $\mathrm{Cov}[B(t), B(t')] = \min(t,t')$, where $t, t' \geq 0$. We let each component of ~~displacements independently and ignore any potential covariance . Hence, we drop the component subscripts with the understanding that the same analysis is being repeated to estimate the easting and northing components of~~ $\vec{u}$ have zero mean and a generic covariance function $\mathrm{Cov}[u_i(p), u_i(p')] = C_{u_i}(p,p')$. Using a more concise notation, we write our prior on each component of $\vec{u}$ as $u_i \sim \mathcal{GP}(0, C_{u_i})$.

The function C_{u_i} must be positive definite in order to be a valid covariance function. By definition, C_{u_i} is a positive definite function if the matrix $[C_{u_i}(p,p')]_{(p,p')\in\boldsymbol{P}\times\boldsymbol{P}}$ is positive definite for any set of points $\boldsymbol{P}$ (Cressie 1993, sec. 2.5). We assume that ~~C_u~~ C_{u_i} can be separated into ~~positive definite~~ spatial and temporal functions as

$$C_{\cancel{u}u_i}(p,p') = C_{u_i}\left((\vec{x},t),(\vec{x}',t')\right) = X_i(\vec{x},\vec{x}')T_i(t,t'). \tag{1}$$

As long as the functions X_i and T_i are positive definite, C_{u_i} is guaranteed to also be positive definite (Rasmussen & Williams 2006, sec. 4.2.4). We also require that the derivatives $\partial^2 X_i(\vec{x},\vec{x}')/\partial x_j \partial x'_j$ and $\partial^2 T_i(t,t')/\partial t \partial t'$ exist. This ensures that u_i is spatially and temporally differentiable, allowing us to compute transient strain rates (See Adler (1981, sec. 2.2) or Papoulis (1991, sec. 10A) for a definition of stochastic differentiation and the conditions for differentiability).

We provide an example to give the prior on transient displacements a more tangible meaning. we can use a squared exponential function for X_i and T_i,

$$X_i(\vec{x},\vec{x}') = \exp\left(\frac{-||\vec{x}-\vec{x}'||_2^2}{2\ell^2}\right), \quad T_i(t,t') = \phi^2 \exp\left(\frac{-|t-t'|^2}{2\tau^2}\right), \tag{2}$$

which satisfies our requirements for positive definiteness and differentiability. The parameters ℓ and τ control the length-scale and time-scale, respectively, of realizations of u_i. The parameter ϕ, which we have arbitrarily chosen to incorporate into T_i rather than X_i, controls the amplitude of realizations of u_i. Ideally, we want realizations of u_i to have a length-scale, time-scale, and amplitude that resemble what we expect for the true transient displacements. In Figure 1A, we show C_{u_i} using the squared exponential function for X_i and T_i and the parameters $\ell = 100$ km, $\tau = 10$ days, and $\phi = 1$ mm. A

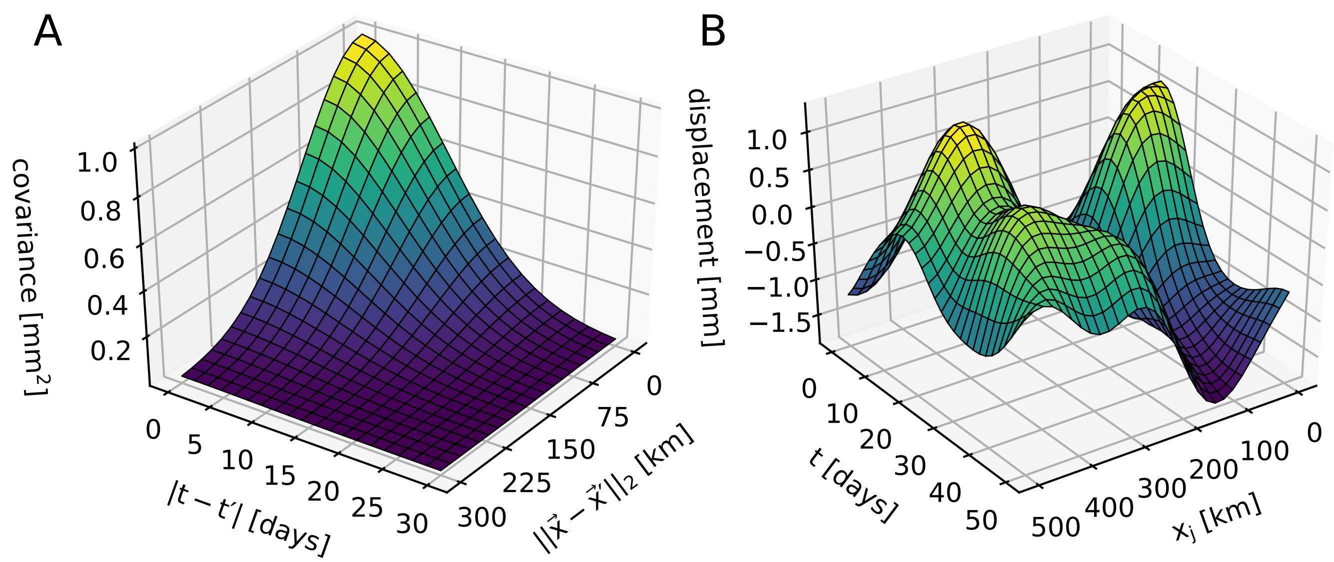

Figure 1. (A) An example covariance function for transient displacements, $C_{u_i}(p, p')$, shown as a function of the spatial and temporal distance between p and p'. (B) A single realization of transient displacements, $u_i(p)$, corresponding to the covariance function from Panel A. The realization is a function of three variables, t, x_{e}, and x_{n}, although we only show its dependence on t and one of the spatial dimensions.

single realization of u_i corresponsing to that choice of covariance function is shown in Figure 1B (See Rasmussen & Williams (2006, sec. A3) for details on drawing realization from Gaussian processes). While the squared exponential function is a commonly used covariance function for GPR, it is not appropriate for every application. The appropriate choice for ~~X and T~~ X_i and T_i may vary depending on the geophysical signal we are trying to describe~~, and we discuss this matter in Section 3.2~~. To keep this section sufficiently general, we hold off on specifying X_i and T_i until Section 3.2, where we estimate transient strain from slow slip events in the Pacific Northwest.

~~We constrain u with GNSS data , which records u~~

GNSS data records transient displacements as well as other physical and non-physical processes that we are not interested in. We ~~describe GNSS observations at position $\vec{x}_i$ and time t_j~~ first consider component i of a single GNSS displacement observation made at station j, which is located at $\vec{x}^{(j)}$, and time $t^{(k)}$. We describe this observation, $d_i^{*(jk)}$, as a realization of the random variable

$$
\begin{aligned}
d_i^{(jk)} =& u_i\left(\vec{x}^{(j)}, t^{(k)}\right) + \eta_i^{(jk)} + a_i^{(j)} + b_i^{(j)} t^{(k)} + \\
& c_i^{(j)} \sin\left(2\pi t^{(k)}\right) + e_i^{(j)} \cos\left(2\pi t^{(k)}\right) + \\
& f_i^{(j)} \sin\left(4\pi t^{(k)}\right) + g_i^{(j)} \cos\left(4\pi t^{(k)}\right),
\end{aligned}
\tag{3}
$$

where ~~$a_i^{(1)}$~~ $\eta_i^{(jk)}$ describes noise, $a_i^{(j)}$ is an offset that is unique for each ~~GNSS monument, $a_i^{(2)}$~~ station, $b_i^{(j)}$ is the secular velocity at ~~$\vec{x}_i$~~ $\vec{x}^{(j)}$, and the sinusoids describe seasonal deformation (using units of years for ~~t_j). We use w_{ij} to denote normally distributed, uncorrelated noise. Correlated noise which does not have a parametric representation is denoted by η. For example, η can consist of temporally correlated noise describing benchmark wobble (e.g., Wyatt 1982, 1989), and/or spatially correlated noise describing common mode error (e.g., Wdowinski et al. 1997). For now, we will only assume that $\eta \sim \mathcal{N}(0, C_\eta)$. We consider the six coefficients in~~ $t^{(k)}$).

We then consider the column vector of n GNSS displacement observations, $\boldsymbol{d}_i^*$, where the subscript indicates that we are still only considering component i of displacements. The observations are made at m stations, and the times and positions for each observation are described by the set $\boldsymbol{P}$. The data vector $\boldsymbol{d}_i^*$ can be considered a realization of the random vector $\boldsymbol{d}_i$, which is formed by evaluating eq. (3) at each point in $\boldsymbol{P}$. To write out $\boldsymbol{d}_i$ explicitly, we let $\boldsymbol{G}$ be an $n \times 6m$ matrix consisting of the basis functions from eq. (3) ~~to be uncorrelated random variables distributed as $\mathcal{N}(0, \kappa^2)$ in the limit as $\kappa \to \infty$~~ (i.e., the ~~coefficients have diffuse priors)~~linear trends and sinusoids for each station) evaluated at each point in $\boldsymbol{P}$. The coefficients corresponding to each basis function are collected into the column vector $\boldsymbol{m}_i$. The noise for all the observations are described by the column vector $\boldsymbol{\eta}_i$. We can then write $\boldsymbol{d}_i$ as

$$\boldsymbol{d}_i = u_i(\boldsymbol{P}) + \boldsymbol{\eta}_i + \boldsymbol{G}\boldsymbol{m}_i, \tag{4}$$

where the notation $u_i(\boldsymbol{P})$ represents the column vector $[u_i(p)]_{p \in \boldsymbol{P}}$.

We assume a diffuse prior for the components of $\boldsymbol{m}_i$, that is to say $\boldsymbol{m}_i \sim \mathcal{N}(\boldsymbol{0}, \kappa^2 \boldsymbol{I})$ in the limit as $\kappa \to \infty$. Of course, the secular velocities, ~~$a_i^{(2)}$~~ $b_i^{(j)}$, are spatially correlated and we could invoke a tectonic model to form a prior on ~~$a_i^{(2)}$~~ $b_i^{(j)}$. However, in our application to ~~Cascadia~~the Pacific Northwest, we will be using displacement time series which are long enough to sufficiently constrain ~~$a_i^{(2)}$~~ $b_i^{(j)}$ for each station, avoiding the need to incorporate a prior. Likewise, seasonal deformation is spatially correlated (Dong et al. 2002; Langbein 2008), and it may be worth exploring and exploiting such a correlation in a future study.

~~We now consider the column vector of n GNSS observations made at m stations, $\boldsymbol{d}_*$. Let $\boldsymbol{P}$ be the set of $(\vec{x}_i, t_j)$ pairs describing where andwhen each of the GNSS observations have been made. Let $\boldsymbol{a}$ be the vector of coefficients from eq. (3) for each of the m GNSS stations. We use $\boldsymbol{G}$ to represent the $n \times 6m$ matrixof corresponding basis functions evaluated at each point in $\boldsymbol{P}$. We also denote the vector of uncorrelated noise for each observation as $\boldsymbol{w}$, whose standard deviations are given by the formal data uncertainty, $\boldsymbol{\sigma}$. The observations can then be viewed as a realization of the random vector~~

$$\cancel{\boldsymbol{d} = u(\boldsymbol{P}) + \eta(\boldsymbol{P}) + \boldsymbol{w} + \boldsymbol{G}\boldsymbol{a},}$$

~~which is distributed as $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma} + \kappa^2 \boldsymbol{G}\boldsymbol{G}^T)$, where~~

$$\boldsymbol{\Sigma} = C_u(\boldsymbol{P}, \boldsymbol{P}) + C_\eta(\boldsymbol{P}, \boldsymbol{P}) + \mathrm{diag}\left(\boldsymbol{\sigma}^2\right).$$

~~It should be understood that notation such as $u(\boldsymbol{P})$ and $C_u(\boldsymbol{P}, \boldsymbol{P})$ represents the column vector $[u(P_i)]_{P_i \in \boldsymbol{P}}$ and the matrix $[C_u(P_i, P_j)]_{(P_i,P_j)\in \boldsymbol{P}\times\boldsymbol{P}}$, respectively~~We assume that $\boldsymbol{\eta}_i$ is a spatially and/or temporally correlated random vector distributed as $\mathcal{N}(\mathbf{0}, \boldsymbol{C}_{\eta_i})$. For example, $\boldsymbol{\eta}_i$ can be uncorrelated white noise, temporally correlated noise describing benchmark wobble (e.g., Wyatt 1982, 1989), and/or spatially correlated noise describing common mode error (e.g., Wdowinski et al. 1997). The appropriate noise model may vary depending on the application, and we hold off on specifying the covariance matrix, $\boldsymbol{C}_{\eta_i}$, until Section 3.1. We are now able to write the distribution of $\boldsymbol{d}_i$ as

$$\boldsymbol{d}_i \sim \mathcal{N}\left(\mathbf{0}, C_{u_i}(\boldsymbol{P}, \boldsymbol{P}) + \boldsymbol{C}_{\eta_i} + \kappa^2 \boldsymbol{G}\boldsymbol{G}^T\right), \tag{5}$$

where $C_{u_i}(\boldsymbol{P}, \boldsymbol{P})$ represents the matrix $[C_{u_i}(p, p')]_{(p,p')\in\boldsymbol{P}\times\boldsymbol{P}}$.

~~The prior for transient displacements is then conditioned with $\boldsymbol{d}_*$ to~~

We form a posterior estimate of transient displacements, ~~$\hat{u} = u|\boldsymbol{d}_*$. For now, we will assume that appropriate covariance functions and corresponding hyperparameters for X, T, and C_η have already been chosen. In Section 3.1 and 3.2, we discuss how the covariance functions are chosen for our application to GNSS data from Cascadia. If κ is kept finite then, following Rasmussen & Williams (2006),~~ denoted as $\hat{u}_i$, by updating u_i with the fact that $\boldsymbol{d}_i^*$ was realized from the random vector $\boldsymbol{d}_i$, that is to say $\hat{u}_i = u_i|(\boldsymbol{d}_i = \boldsymbol{d}_i^*)$. A general solution for $\hat{u}_i$ is derived in von Mises (1964, sec. 8.9), where we find that ~~$\hat{u}$~~ $\hat{u}_i$ is distributed as ~~$\mathcal{N}(\mu_{\hat{u}}, C_{\hat{u}})$, where~~

$$\mu_{\hat{u}}(p) = C_u(p, \boldsymbol{P})\left(\boldsymbol{\Sigma} + \kappa^2 \boldsymbol{G}\boldsymbol{G}^T\right)^{-1} \boldsymbol{d}_*$$

~~and~~

$$C_{\hat{u}}(p, p') = C_u(p, p') - C_u(p, \boldsymbol{P})\left(\boldsymbol{\Sigma} + \kappa^2 \boldsymbol{G}\boldsymbol{G}^T\right)^{-1} C_u(\boldsymbol{P}, p').$$

$\mathcal{GP}(\mu_{\hat{u}_i}, C_{\hat{u}_i})$ with mean function

$$\begin{aligned}\mu_{\hat{u}_i}(p) &= \mathrm{E}\left[u_i(p)\right] + \mathrm{Cov}\left[u_i(p), \boldsymbol{d}_i\right] \mathrm{Cov}\left[\boldsymbol{d}_i\right]^{-1}\left(\boldsymbol{d}_i^* - \mathrm{E}\left[\boldsymbol{d}_i\right]\right) \\ &= C_{u_i}(p, \boldsymbol{P})\left(C_{u_i}(\boldsymbol{P}, \boldsymbol{P}) + \boldsymbol{C}_{\eta_i} + \kappa^2 \boldsymbol{G}\boldsymbol{G}^T\right)^{-1} \boldsymbol{d}_i^*\end{aligned} \tag{6}$$

and covariance function

$$\begin{aligned}C_{\hat{u}_i}(p, p') &= \mathrm{Cov}\left[u_i(p), u_i(p')\right] - \mathrm{Cov}\left[u_i(p), \boldsymbol{d}_i\right] \mathrm{Cov}\left[\boldsymbol{d}_i\right]^{-1} \mathrm{Cov}\left[\boldsymbol{d}_i, u_i(p')\right] \\ &= C_{u_i}(p, p') - C_{u_i}(p, \boldsymbol{P})\left(C_{u_i}(\boldsymbol{P}, \boldsymbol{P}) + \boldsymbol{C}_{\eta_i} + \kappa^2 \boldsymbol{G}\boldsymbol{G}^T\right)^{-1} C_{u_i}(\boldsymbol{P}, p').\end{aligned} \tag{7}$$

However, we are interested in the limit as $\kappa \to \infty$, and the form for ~~eq~~eqs. (6) and ~~eq.~~ (7) is not suitable for evaluating this limit. We use ~~the~~ a partitioned matrix inversion identity ~~(e.g., Press et al. 2007) to rewrite eq~~(Press et al. 2007, sec. 2.7.4) to rewrite eqs. (6) and ~~eq.~~ (7) as

$$\mu_{\hat{u}\hat{u}_i}(p) = \begin{bmatrix} C_{u_i}(p,\boldsymbol{P}) & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} C_{u_i}(\boldsymbol{P},\boldsymbol{P}) + \boldsymbol{C}_{\eta_i} & \boldsymbol{G} \\ \boldsymbol{G}^T & -\kappa^{-2}\boldsymbol{I} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{d}_i^* \\ \boldsymbol{0} \end{bmatrix} \tag{8}$$

and

$$C_{\hat{u}}(p,p') = C_u(p,p') - \begin{bmatrix} C_u(p,\boldsymbol{P}) & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{G} \\ \boldsymbol{G}^T & -\kappa^{-2}\boldsymbol{I} \end{bmatrix}^{-1} \begin{bmatrix} C_u(\boldsymbol{P},p') \\ \boldsymbol{0} \end{bmatrix}.$$

$$C_{\hat{u}_i}(p,p') = C_{u_i}(p,p') - \begin{bmatrix} C_{u_i}(p,\boldsymbol{P}) & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} C_{u_i}(\boldsymbol{P},\boldsymbol{P}) + \boldsymbol{C}_{\eta_i} & \boldsymbol{G} \\ \boldsymbol{G}^T & -\kappa^{-2}\boldsymbol{I} \end{bmatrix}^{-1} \begin{bmatrix} C_{u_i}(\boldsymbol{P},p') \\ \boldsymbol{0} \end{bmatrix}. \tag{9}$$

Taking the limit as $\kappa \to \infty$, we get the solution for the mean and covariance of ~~$\hat{u}$,~~ $\hat{u}_i$,

$$\mu_{\hat{u}\hat{u}_i}(p) = \begin{bmatrix} C_{u_i}(p,\boldsymbol{P}) & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} C_{u_i}(\boldsymbol{P},\boldsymbol{P}) + \boldsymbol{C}_{\eta_i} & \boldsymbol{G} \\ \boldsymbol{G}^T & \boldsymbol{0} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{d}_i^* \\ \boldsymbol{0} \end{bmatrix} \tag{10}$$

and

$$C_{\hat{u}\hat{u}_i}(p,p') = C_{uu_i}(p,p') - \begin{bmatrix} C_{u_i}(p,\boldsymbol{P}) & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} C_{u_i}(\boldsymbol{P},\boldsymbol{P}) + \boldsymbol{C}_{\eta_i} & \boldsymbol{G} \\ \boldsymbol{G}^T & \boldsymbol{0} \end{bmatrix}^{-1} \begin{bmatrix} C_{u_i}(\boldsymbol{P},p') \\ \boldsymbol{0} \end{bmatrix}. \tag{11}$$

~~We use eq~~To ensure that the inverse matrices in eqs. (10) and (11) ~~to find the posterior~~ exist, each column in $\boldsymbol{G}$ must be linearly independent. This condition tends to be violated when there are too few observations at a station. In that case, a singular value decomposition can be used to remove linearly dependent components from $\boldsymbol{G}$.

It should be noted that we have ignored any covariances between the easting and northing components of ~~transient displacements. Using $\hat{u}_i$ to denote the~~ $\vec{u}$ and $\vec{\boldsymbol{d}}$. This simplification reduces the computational complexity of evaluating the posterior transient displacements because each component can be evaluated independently. However, we are inherently assuming that the principle axes describing the distribution of $\vec{u}(p)$ and $\vec{d}^{(jk)}$ are aligned with the cardinal directions, which is admittedly an arbitrary assumption.

The posterior transient displacements ~~along direction i and x_i to represent the components of $\vec{x}$, we can write the components of $\dot{\varepsilon}$ as~~ are spatially and temporally continuous, and we can use eqs. (10)

and (11) to evaluate $\hat{u}_i$ at any p. Furthermore, $\hat{u}_i$ is spatially and temporally differentiable, allowing us to formulate $\dot{\varepsilon}$ at any p that we may be interested in. The components of $\dot{\varepsilon}$ can be written as

$$\varepsilon\dot{\varepsilon}_{ij}(p) = \frac{1}{2}\frac{\partial}{\partial t}\left(\frac{\partial \hat{u}_i(p)}{\partial x_j}\frac{\partial \hat{u}_i(p)}{\partial x_j} + \frac{\partial \hat{u}_j(p)}{\partial x_i}\frac{\partial \hat{u}_j(p)}{\partial x_i}\right). \tag{12}$$

~~The transient strain rate components are Gaussian processes with mean functions~~

$$\mu_{\dot{\varepsilon}_{ij}}(p) = \frac{1}{2}\frac{\partial}{\partial t}\left(\frac{\partial \mu_{\hat{u}_i}(p)}{\partial x_j} + \frac{\partial \mu_{\hat{u}_j}(p)}{\partial x_i}\right)$$

Since eq. (12) is a linear operation on the Gaussian processes $\hat{u}_i$ and $\hat{u}_j$, we know that $\dot{\varepsilon}_{ij}$ is also a Gaussian process. From Papoulis (1991, sec. 10), we find that the mean and covariance functions for $\dot{\varepsilon}_{ij}$ are

$$\mu_{\dot{\varepsilon}_{\mathrm{ee}}}(p) = \frac{\partial^2 \mu_{\hat{u}_{\mathrm{e}}}(p)}{\partial t\,\partial x_{\mathrm{e}}} \tag{13}$$

$$\mu_{\dot{\varepsilon}_{\mathrm{nn}}}(p) = \frac{\partial^2 \mu_{\hat{u}_{\mathrm{n}}}(p)}{\partial t\,\partial x_{\mathrm{n}}} \tag{14}$$

$$\mu_{\dot{\varepsilon}_{\mathrm{en}}}(p) = \frac{1}{2}\frac{\partial}{\partial t}\left(\frac{\partial \mu_{\hat{u}_{\mathrm{e}}}(p)}{\partial x_{\mathrm{n}}} + \frac{\partial \mu_{\hat{u}_{\mathrm{n}}}(p)}{\partial x_{\mathrm{e}}}\right) \tag{15}$$

and

$$C_{\dot{\varepsilon}_{\mathrm{ee}}}(p,p') = \frac{\partial^4 C_{\hat{u}_{\mathrm{e}}}(p,p')}{\partial t\,\partial t'\,\partial x_{\mathrm{e}}\,\partial x'_{\mathrm{e}}} \tag{16}$$

$$C_{\dot{\varepsilon}_{\mathrm{nn}}}(p,p') = \frac{\partial^4 C_{\hat{u}_{\mathrm{n}}}(p,p')}{\partial t\,\partial t'\,\partial x_{\mathrm{n}}\,\partial x'_{\mathrm{n}}} \tag{17}$$

$$C_{\dot{\varepsilon}_{\mathrm{en}}}(p,p') = \frac{1}{4}\frac{\partial^2}{\partial t\,\partial t'}\left(\frac{\partial^2 C_{\hat{u}_{\mathrm{e}}}(p,p')}{\partial x_{\mathrm{n}}\,\partial x'_{\mathrm{n}}} + \frac{\partial^2 C_{\hat{u}_{\mathrm{n}}}(p,p')}{\partial x_{\mathrm{e}}\,\partial x'_{\mathrm{e}}}\right), \tag{18}$$

respectively. The above derivatives of $\mu_{\hat{u}_i}$ and $C_{\hat{u}_i}$ are computed analytically by replacing the terms $C_{u_i}(p,\boldsymbol{P})$, $C_{u_i}(\boldsymbol{P},p')$, and $C_{u_i}(p,p')$ in eqs. (10) and (11) with their appropriate derivatives. For example, the mean and covariance functions for $\dot{\varepsilon}_{\mathrm{ee}}$ can be written more verbosely as

$$C_{\dot{\varepsilon}_{ij}}\mu_{\dot{\varepsilon}_{\mathrm{ee}}}(p,p') = \frac{1}{4}\frac{\partial^2}{\partial t\,\partial t'}\frac{\partial^2 C_{\hat{u}_i}(p,p')}{\partial x_j\,\partial x'_j} + \frac{\partial^2 C_{\hat{u}_j}(p,p')}{\partial x_i\,\partial x'_i}. \begin{bmatrix} \frac{\partial^2 C_{u_{\mathrm{e}}}(p,\boldsymbol{P})}{\partial t\,\partial x_{\mathrm{e}}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} C_{u_{\mathrm{e}}}(\boldsymbol{P},\boldsymbol{P}) + \boldsymbol{C}_{\eta_{\mathrm{e}}} & \boldsymbol{G} \\ \boldsymbol{G}^T & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{d}^*_{\mathrm{e}} \\ \mathbf{0} \end{bmatrix} \tag{19}$$

and

$$
C_{\dot{\varepsilon}_{\mathrm{ee}}}(p,p') = \frac{\partial^4 C_{u_{\mathrm{e}}}(p,p')}{\partial t\,\partial t'\,\partial x_{\mathrm{e}}\,\partial x'_{\mathrm{e}}} -
\begin{bmatrix} \frac{\partial^2 C_{u_{\mathrm{e}}}(p,\boldsymbol{P})}{\partial t\,\partial x_{\mathrm{e}}} & \boldsymbol{0} \end{bmatrix}
\begin{bmatrix} C_{u_{\mathrm{e}}}(\boldsymbol{P},\boldsymbol{P}) + \boldsymbol{C}_{\eta_{\mathrm{e}}} & \boldsymbol{G} \\ \boldsymbol{G}^T & \boldsymbol{0} \end{bmatrix}^{-1}
\begin{bmatrix} \frac{\partial^2 C_{u_{\mathrm{e}}}(\boldsymbol{P},p')}{\partial t'\,\partial x'_{\mathrm{e}}} \\ \boldsymbol{0} \end{bmatrix}. \tag{20}
$$

2.1 Transient detection criterion

Our motivation for estimating transient strain rates is, in part, to detect ~~geophysical phenomena. We propose using~~ transient geophysical processes. As we will see, geophysical signal can be easily identified by visually inspecting the solution for $\dot{\varepsilon}$ from eqs. (13) and (16). However, if we want to detect geophysical signal automatically, then we need to define a detection criterion. We use a signal-to-noise ratio, SNR, ~~based on $\dot{\varepsilon}$ for our detection criterion. The~~ that is based on the Frobenius norm of ~~$\dot{\varepsilon}$, $||\dot{\varepsilon}||_{\mathrm{F}}$, which is sometimes referred to as the second invariant of strain rate in~~ $\dot{\varepsilon}$, $||\dot{\varepsilon}||_{\mathrm{F}} = \left(\dot{\varepsilon}_{\mathrm{ee}}^2 + \dot{\varepsilon}_{\mathrm{nn}}^2 + 2\dot{\varepsilon}_{\mathrm{en}}^2\right)^{\frac{1}{2}}$, for our detection criterion. In the geodetic literature, $||\dot{\varepsilon}||_{\mathrm{F}}$ is often used as a metric for the strain rate "magnitude~~"~~", and it is sometimes referred to as the second invariant of strain rate. Noting that ~~$||\dot{\varepsilon}||_{\mathrm{F}}$~~ $||\dot{\varepsilon}||_{\mathrm{F}}$ is a random variable, ~~SNR can be taken as~~ we take SNR to be the ratio of the estimated mean and standard deviation of ~~$||\dot{\varepsilon}||_{\mathrm{F}}$. Using~~ $||\dot{\varepsilon}||_{\mathrm{F}}$. An estimate of the mean is found by evaluating $||\dot{\varepsilon}||_{\mathrm{F}}$ at the mean of $\dot{\varepsilon}$,

$$
\begin{aligned}
\mu_{||\dot{\varepsilon}||_{\mathrm{F}}} &\approx \left. ||\dot{\varepsilon}||_F \right|_{\dot{\varepsilon}=\mu_{\dot{\varepsilon}}} \\
&= \left(\mu_{\dot{\varepsilon}_{\mathrm{ee}}}^2 + \mu_{\dot{\varepsilon}_{\mathrm{nn}}}^2 + 2\mu_{\dot{\varepsilon}_{\mathrm{en}}}^2\right)^{\frac{1}{2}},
\end{aligned} \tag{21}
$$

and we use nonlinear uncertainty propagation to estimate the standard deviation,

$$
\sigma_{||\dot{\varepsilon}||_{\mathrm{F}}} \approx \left(\left(\left. \frac{\partial ||\dot{\varepsilon}||_{\mathrm{F}}}{\partial \dot{\varepsilon}_{\mathrm{ee}}} \right|_{\dot{\varepsilon}=\mu_{\dot{\varepsilon}}} \right)^2 \sigma_{\dot{\varepsilon}_{\mathrm{ee}}}^2 + \left(\left. \frac{\partial ||\dot{\varepsilon}||_{\mathrm{F}}}{\partial \dot{\varepsilon}_{\mathrm{nn}}} \right|_{\dot{\varepsilon}=\mu_{\dot{\varepsilon}}} \right)^2 \sigma_{\dot{\varepsilon}_{\mathrm{nn}}}^2 + \left(\left. \frac{\partial ||\dot{\varepsilon}||_{\mathrm{F}}}{\partial \dot{\varepsilon}_{\mathrm{en}}} \right|_{\dot{\varepsilon}=\mu_{\dot{\varepsilon}}} \right)^2 \sigma_{\dot{\varepsilon}_{\mathrm{en}}}^2 \right)^{\frac{1}{2}}, \tag{22}
$$

where $\sigma_{\dot{\varepsilon}_{ij}}^2(p) = C_{\dot{\varepsilon}_{ij}}(p,p)$. After some calculations, we find SNR to be

$$
\mathrm{SNR}(p) = \frac{\mu_{\dot{\varepsilon}_{\mathrm{nn}}}(p)^2 + \mu_{\dot{\varepsilon}_{\mathrm{ee}}}(p)^2 + 2\mu_{\dot{\varepsilon}_{\mathrm{en}}}(p)^2}{\left(C_{\dot{\varepsilon}_{\mathrm{nn}}}(p,p)\mu_{\dot{\varepsilon}_{\mathrm{nn}}}(p)^2 + C_{\dot{\varepsilon}_{\mathrm{ee}}}(p,p)\mu_{\dot{\varepsilon}_{\mathrm{ee}}}(p)^2 + 4C_{\dot{\varepsilon}_{\mathrm{en}}}(p,p)\mu_{\dot{\varepsilon}_{\mathrm{en}}}(p)^2\right)^{\frac{1}{2}}},
$$

~~where the subscripts "n" and "e" denote north and east, respectively. For simplicity, we have ignored covariances between the transient strain rate components in eq. (23), even though they are non-zero~~

$$\mathrm{SNR}(p) = \frac{\mu_{||\dot{\varepsilon}||_\mathrm{F}}(p)}{\sigma_{||\dot{\varepsilon}||_\mathrm{F}}(p)} \tag{23}$$

$$= \frac{\mu_{\dot{\varepsilon}_\mathrm{ee}}(p)^2 + \mu_{\dot{\varepsilon}_\mathrm{nn}}(p)^2 + 2\mu_{\dot{\varepsilon}_\mathrm{en}}(p)^2}{\left(\sigma^2_{\dot{\varepsilon}_\mathrm{ee}}(p)\mu_{\dot{\varepsilon}_\mathrm{ee}}(p)^2 + \sigma^2_{\dot{\varepsilon}_\mathrm{nn}}(p)\mu_{\dot{\varepsilon}_\mathrm{nn}}(p)^2 + 4\sigma^2_{\dot{\varepsilon}_\mathrm{en}}(p)\mu_{\dot{\varepsilon}_\mathrm{en}}(p)^2\right)^{\frac{1}{2}}}. \tag{24}$$

We explicitly show that SNR is a function of p to emphasize that it identifies the position and time of anomalous deformation. We can reasonably suspect that some transient geophysical phenomena is occurring wherever and whenever SNR is larger than ∼3.

3 ~~OUTLIER DETECTION~~

2.1 Outlier detection

~~In~~

In deriving our formulation for ~~estimating~~ transient strain rates, we have assumed that noise in the data vector is normally distributed. This is not an appropriate assumption for GNSS data, which often have more outliers than would be ~~predicted~~ expected for normally distributed noise. ~~It follows that proposed methods~~ Methods for analyzing GNSS data should either be robust against outliers ~~(e.g., Blewitt et al. 2016). In order to make our estimates of transient strain more robust, we automatically~~ or should involve a preprocessing step in which outliers are detected and removed. Examples of the former include the MIDAS method for estimating secular velocities (Blewitt et al. 2016) and the GPS Imaging method for detecting spatially coherent features (Hammond et al. 2016). In this study, we identify and remove outliers ~~in the GNSS data as a pre-processing step .~~

~~Our method for detecting outliers is~~ as a preprocessing step before estimating $\dot{\varepsilon}$. Outliers are identified based on the ~~data editing algorithm described in Gibbs (2011). We calculate the residuals between the observations and a best fitting model . Data~~ residuals for a model that best fits the observed data. Observations with residuals that exceed some threshold are removed. This strategy for detecting outliers is commonly used for GNSS data, where the model being fit to the data typically consists of a linear trend and seasonal terms for each GNSS station (e.g., Johansson et al. 2002; Dong et al. 2006; Bos et al. 2013). To prevent transient geophysical signal from being erroneously identified as outliers, the model used in our outlier detection algorithm additionally consists of a temporally correlated Gaussian process. The details of our algorithm are ~~anomalously large are then identified as outliers. We treat $\boldsymbol{d}_*$ as a sample of $\boldsymbol{d}$ and assume that there is no correlated noise (i.e., $\eta(p) = 0$). The best fitting model for $\boldsymbol{d}_*$ is considered to be the expected value of the random vector $u(\boldsymbol{P}) + \boldsymbol{Ga}$ after conditioning it with~~

~~non-outlier observations. We still consider u to have a separable covariance function as in eq. (1), and the choice for X and T does not need to be the same as that used in Section 2. Since outliersare determined based on how well a spatially and temporally dependent model fits the data, we are able to identify anomalous observations which may not be immediately apparent from inspecting individual time series.~~

~~To begin the algorithm, we let $\boldsymbol{\Omega}$ be the index set of non-outliers in $\boldsymbol{d}_*$ and initiate it with all n indices. This algorithm is iterative, and for each iteration we calculate the residual vector~~

$$\boldsymbol{r} = \frac{\boldsymbol{d}_* - \mathrm{E}\left[(u(\boldsymbol{P}) + \boldsymbol{G}\boldsymbol{a})|\tilde{\boldsymbol{d}}_*\right]}{\boldsymbol{\sigma}}$$

$$= \frac{1}{\boldsymbol{\sigma}}\left(\boldsymbol{d}_* - \left[\begin{array}{cc} C_u(\boldsymbol{P},\tilde{\boldsymbol{P}}) & \boldsymbol{G} \end{array}\right]\left[\begin{array}{cc} C_u(\tilde{\boldsymbol{P}},\tilde{\boldsymbol{P}}) + \mathrm{diag}(\tilde{\boldsymbol{\sigma}}^2) & \tilde{\boldsymbol{G}} \\ \tilde{\boldsymbol{G}}^T & \boldsymbol{0} \end{array}\right]^{-1}\left[\begin{array}{c} \tilde{\boldsymbol{d}}_* \\ \boldsymbol{0} \end{array}\right]\right),$$

~~where the tilde indicates that only elements corresponding to indices in $\boldsymbol{\Omega}$ are retained (e.g., $\tilde{\boldsymbol{P}} = \{P_i\}_{i\in\boldsymbol{\Omega}}$). We then update $\boldsymbol{\Omega}$ to be~~

$$\boldsymbol{\Omega} = \{i : |r_i| < \lambda \cdot \mathrm{RMS}\}, \quad r_i \in \boldsymbol{r},$$

~~where RMS is the root-mean-square of $\tilde{\boldsymbol{r}}$ and λ is an outlier tolerance. We use $\lambda = 4$ in this study, which in our experience accurately identifies outliers without unnecessarily decimating the data. Iterations continue until the new $\boldsymbol{\Omega}$ is equal to the previous $\boldsymbol{\Omega}$~~given in Appendix A.

It should be noted that ~~this~~ our algorithm does not identify jumps in GNSS time series, which are another common issue. Some, but not all, jumps can be automatically removed by looking up the dates of equipment changes and earthquakes (Gazeaux et al. 2013). However, it is still necessary to manually find and remove jumps of unknown origin. ~~That being said, this outlier detection algorithm significantly reduces the effort needed to manually clean GNSS data.~~

3 APPLICATION TO ~~CASCADIA~~ PACIFIC NORTHWEST SLOW SLIP EVENTS

~~We use our method to~~

In this section we estimate transient strain rates in ~~Cascadia~~the Pacific Northwest, and we are specifically interested in identifying transient strain resulting from SSEs (e.g., Dragert et al. 2001). ~~In Cascadia, SSEs~~ Before estimating transient strain rates, we establish a noise model for GNSS stations in this region, and we establish a prior Gaussian process to describe displacements from SSEs. SSEs in the Pacific Northwest can be detected by monitoring for associated seismic tremor (Rogers & Dragert 2003), which is actively being done by the Pacific Northwest Seismic Network ~~(PNSN)~~ (Wech 2010).

We can ~~thus~~ then compare the tremor records to ~~the~~ our estimated transient strain rates ~~estimated with GPR~~ to verify that we are indeed identifying strain from SSEs.

~~We use publicly available continuous GNSS data from UNAVCO (Herring et al. 2016), which can be found at www.unavco.org~~

We use the daily displacement solutions for continuous GNSS stations generated by the Geodesy Advancing Geosciences and EarthScope (GAGE) Facility (Herring et al. 2016). We limit the dataset to the stations and ~~time ranges which~~ times that are pertinent to the seven most recent SSEs in the Puget Sound region. The earliest SSE considered in this study began in August 2010, and the most recent SSE began in February 2017. ~~We use these most recent SSEs because the station coverage is sufficiently dense for us to use maximum likelihood methods to constrain prior models.~~ The positions of GNSS stations used to estimate transient strain rates are shown in Figure 2.

3.1 Noise model

~~Before we determine the transient strain rates, we must establish a prior for the transient displacements,~~ ~~u, and the noise, η. In this section we discuss our choice for the noise covariance function C_η. There have been numerous studies on~~

We consider the noise vector, $\boldsymbol{\eta}_i$, to be composed of a temporally correlated Gaussian process, $z_i \sim \mathcal{GP}(0, C_{z_i})$, and a vector of uncorrelated Gaussian noise, $\boldsymbol{w}_i$, so that

$$\boldsymbol{\eta}_i = z_i(\boldsymbol{P}) + \boldsymbol{w}_i. \tag{25}$$

The standard deviations for $\boldsymbol{w}_i$ are taken to be the uncertainties derived for the GNSS displacement solutions, $\boldsymbol{\sigma}_i$. The noise vector then has zero mean and the covariance matrix

$$\boldsymbol{C}_{\eta_i} = C_{z_i}(\boldsymbol{P}, \boldsymbol{P}) + \mathrm{diag}(\boldsymbol{\sigma}_i^2). \tag{26}$$

The temporally correlated noise in GNSS data ~~(e.g., Zhang et al. 1997; Mao et al. 1999; Williams et al. 2004; Langbein 2~~ been thoroughly studied over the past two decades (e.g., Zhang et al. 1997; Mao et al. 1999; Williams et al. 2004; Langbein 2 In these studies, temporally correlated noise ~~was~~ tends to be described with some combination of Brownian motion (also known as random walk noise or a Weiner process), a first-order Gauss-Markov (FOGM) process, and/or flicker noise. There is some physical justification for using Brownian motion as a noise model because it accurately describes the power spectrum of motion resulting from instability in ~~geodetic monuments (e.g., Wyatt 1982, 1989). Here we describe the time dependence of η as a FOGM process and consider η to be spatially uncorrelated. A FOGM process is a solution to the stochastic differential equation~~

$$\dot{\eta}(t) + \alpha\eta(t) = \beta w(t),$$

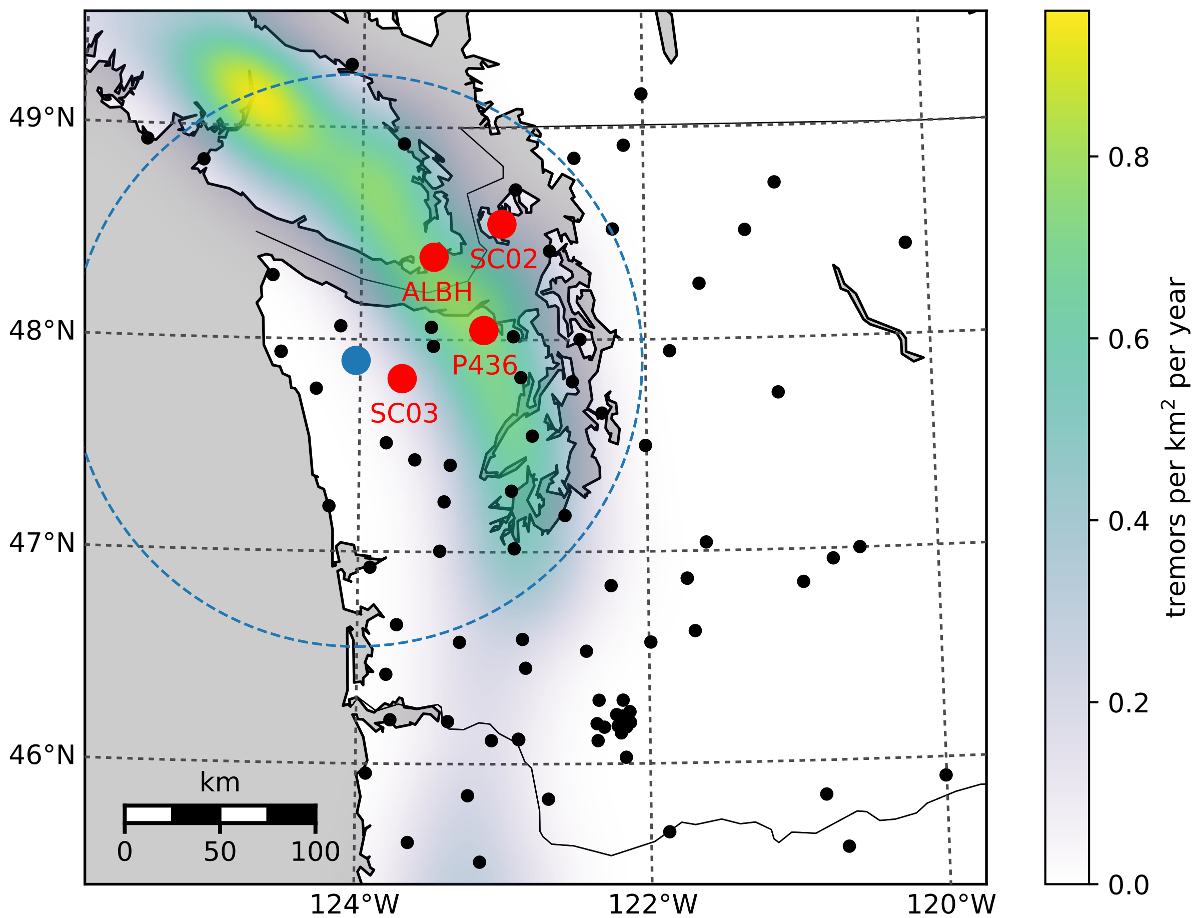

Figure 2. Positions of ~~continuous~~ the GNSS stations used to estimate transient strain rates. The colored regions indicate the distribution of seismic tremor as determined by Wech (2010). The red dots show the positions of GNSS stations mentioned in this paper. The blue dot indicates the location of the transient strain rates shown in Figure 7 and the signal-to-noise ratio shown in Figure 8. The blue dashed circle demarcates the spatial extent of the tremors shown in Figure 8.

~~where $w(t)$ is white noise with unit variance. The FOGM process degenerates to the commonly used Brownian motion noise model under the condition that $\alpha = 0$ and $\eta(0) = 0$.Our noise model that satisfies eq.(**??**) is a Gaussian processwith zero mean and~~ a geodetic monument (e.g., Wyatt 1982, 1989). In particular, the power spectrums for Brownian motion and monument motion both decay proportionally to f^{-2}, where f is frequency. However, Brownian motion necessarily contains a reference time at which the process begins. Since there is no notion of when noise "begins" in GNSS data, we do not find Brownian motion to be an appropriate noise model. On the other hand, a FOGM process has no reference time (i.e., it is stationary), and its power spectrum,

$$S(f) = \frac{\beta^2}{(2\pi f)^2 + \alpha^2}, \tag{27}$$

decays proportionally to f^{-2} above the cutoff frequency $\alpha/(2\pi)$. We then choose z_i to be a spatially uncorrelated FOGM process, which has the covariance function

$$C_{\cancel{\eta} z_i}\left((\vec{x}, t), (\vec{x}', t')\right) = \frac{\beta^2}{2\alpha} \exp\left(-\alpha|t - t'|\right) \delta \cancel{(||-'||_2)}_{\vec{x},\vec{x}'}, \tag{28}$$

where $\delta_{\vec{x},\vec{x}'}$ is 1 if $\vec{x} = \vec{x}'$ and 0 otherwise. See Rasmussen & Williams (2006, sec. B2) for more information on the FOGM process.

~~We constrain the hyperparameters for η~~

By choosing a FOGM process for z_i, we have introduced two parameters that need to be constrained, α and β~~, with a set of~~. We constrain these parameters, which we collectively refer to as $\boldsymbol{\theta}$, with the Restricted Maximum Likelihood (REML) method (Harville 1974). The REML method is conceptually similar to the Maximum Likelihood Estimation (MLE) method from Langbein & Johnson (1997), where we pick $\boldsymbol{\theta}$ to maximize the probability of drawing the observed data, $\boldsymbol{d}_i^*$, from the random vector $\boldsymbol{d}_i$. However, unlike the MLE method, the REML method produces unbiased estimates of $\boldsymbol{\theta}$ (Cressie 1993, sec. 2.6). We use the REML method to estimate $\boldsymbol{\theta}$ at 38 continuous GNSS stations in ~~Cascadia~~ the Pacific Northwest that are east of 121°W. These stations are sufficiently far from the subduction zone that they are unlikely to ~~contain transient signal associated with SSEs. We clean the data for these stations by removing jumps at times of equipment changes, and we remove outliers that have been detected with the algorithm described in Section 2.1. We then find α and β for each station time series with the Restricted Maximum Likelihood (REML) method (e.g., Harville 1974; Cressie 1993; Hines & Hetland 2017). The REML method finds the hyperparameters, which we collectively refer to as~~ record transient deformation from SSEs, allowing us to ingore the term $u_i(\boldsymbol{P})$ in $\boldsymbol{d}_i$. We assume the noise at these inland stations is representative of the noise at all the stations considered in this study, which is probably a poor assumption since the inland stations are subject to distinctly different climatic conditions. Nonetheless, we find $\boldsymbol{\theta}$ for each of the inland stations and for each displacement component by maximizing the REML likelihood function

$$\mathcal{L}(\boldsymbol{\theta}) = \left(\frac{\left|\boldsymbol{G}^T \boldsymbol{G}\right|}{(2\pi)^{n-6m} \left|\boldsymbol{C}_{\eta_i}(\boldsymbol{\theta})\right| \left|\boldsymbol{G}^T \boldsymbol{C}_{\eta_i}(\boldsymbol{\theta})^{-1} \boldsymbol{G}\right|} \right)^{\frac{1}{2}} e^{-\frac{1}{2} \boldsymbol{d}_i^{*T} \boldsymbol{K}(\boldsymbol{\theta}) \boldsymbol{d}_i^*} \tag{29}$$

with respect to $\boldsymbol{\theta}$, ~~that maximize the likelihood function~~

$$\mathcal{L}(\boldsymbol{\theta}) = \left(\frac{\left|\boldsymbol{G}^T\boldsymbol{G}\right|}{(2\pi)^{n-6m}\left|\boldsymbol{\Sigma}(\boldsymbol{\theta})\right|\left|\boldsymbol{G}^T\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{G}\right|}\right)^{\frac{1}{2}} e^{-\frac{1}{2}\boldsymbol{d}_*^T\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{d}_*},$$

~~where~~ where

$$\boldsymbol{K}(\boldsymbol{\theta}) = \boldsymbol{C}_{\eta_i}(\boldsymbol{\theta})^{-1} - \boldsymbol{C}_{\eta_i}(\boldsymbol{\theta})^{-1}\boldsymbol{G}\left(\boldsymbol{G}^T\boldsymbol{C}_{\eta_i}(\boldsymbol{\theta})^{-1}\boldsymbol{G}\right)^{-1}\boldsymbol{G}^T\boldsymbol{C}_{\eta_i}(\boldsymbol{\theta})^{-1}. \tag{30}$$

~~Harville (1974) showed that choosing the hyperparameters which maximize eq. (??) is equivalent to choosing the hyperparameters which maximize the probability of drawing $\boldsymbol{d}_*$ from $\boldsymbol{d}$. We independently estimate $\boldsymbol{\theta}$ for each station, and so $\boldsymbol{d}_*$ consists of displacements for an individual station. We are assuming $u(p) = 0$ when estimating the noise hyperparameters for this section. We use the REML method over the maximum likelihood (ML) method (e.g., Langbein & Johnson 1997) because the REML method accounts for the improper prior that we assigned to $\boldsymbol{\alpha}$ (Hines & Hetland 2017).~~ In the above equation, $\boldsymbol{d}_i^*$ consists of the data for a single station.

The distribution of inferred values for α and β are shown in Figure 3. The ~~amplitude of FOGM noise,~~ estimates of β ~~,~~ for the easting and northing components ~~is notable low and~~ are clustered around 0.5 mm/yr$^{0.5}$. The corresponding estimates of α tend to cluster around 0 yr^{-1}, ~~suggesting that noise can be described with Brownian motion~~indicating that the power spectrum of noise indeed decays proportionally to f^{-2} over a wide band of frequencies. We also estimate ~~hyperparameters~~ α and β for the vertical component of displacements, ~~under~~ with the hope that vertical deformation gradients could reveal some geophysical signal. ~~The amplitude of FOGM noise for~~ For the vertical component, β is relatively large with a median value of 13.5 mm/yr$^{0.5}$. The inferred values for α are also higher for the vertical component with a median value of 8.21 yr^{-1}. In Figure 4, we use the median values of α and β to generate two ~~random samples~~ realizations of FOGM noise for each component. The ~~samples~~ realizations span five years~~and over these five years~~ ~~,~~, and the easting and northing ~~samples~~ realizations drift by about 1 mm over these five years. In the context of detecting SSEs, which produce several mm's of surface displacement on the time-scale of weeks, the estimated FOGM noise for the easting and northing component is negligible. In contrast, the estimated FOGM noise for the vertical component is larger than the signal we would expect from SSEs. We suspect that the higher amplitude for the FOGM noise in the vertical component is accommodating for deficiencies in our rather simple seasonal model. Based on this analysis, we henceforth ignore temporally correlated noise in the easting and northing component because of its low amplitude. We also do not use vertical displacements because of the presumably low signal-to-noise ratio.

Another significant source of noise in GNSS data is common mode error (e.g., Wdowinski et al.

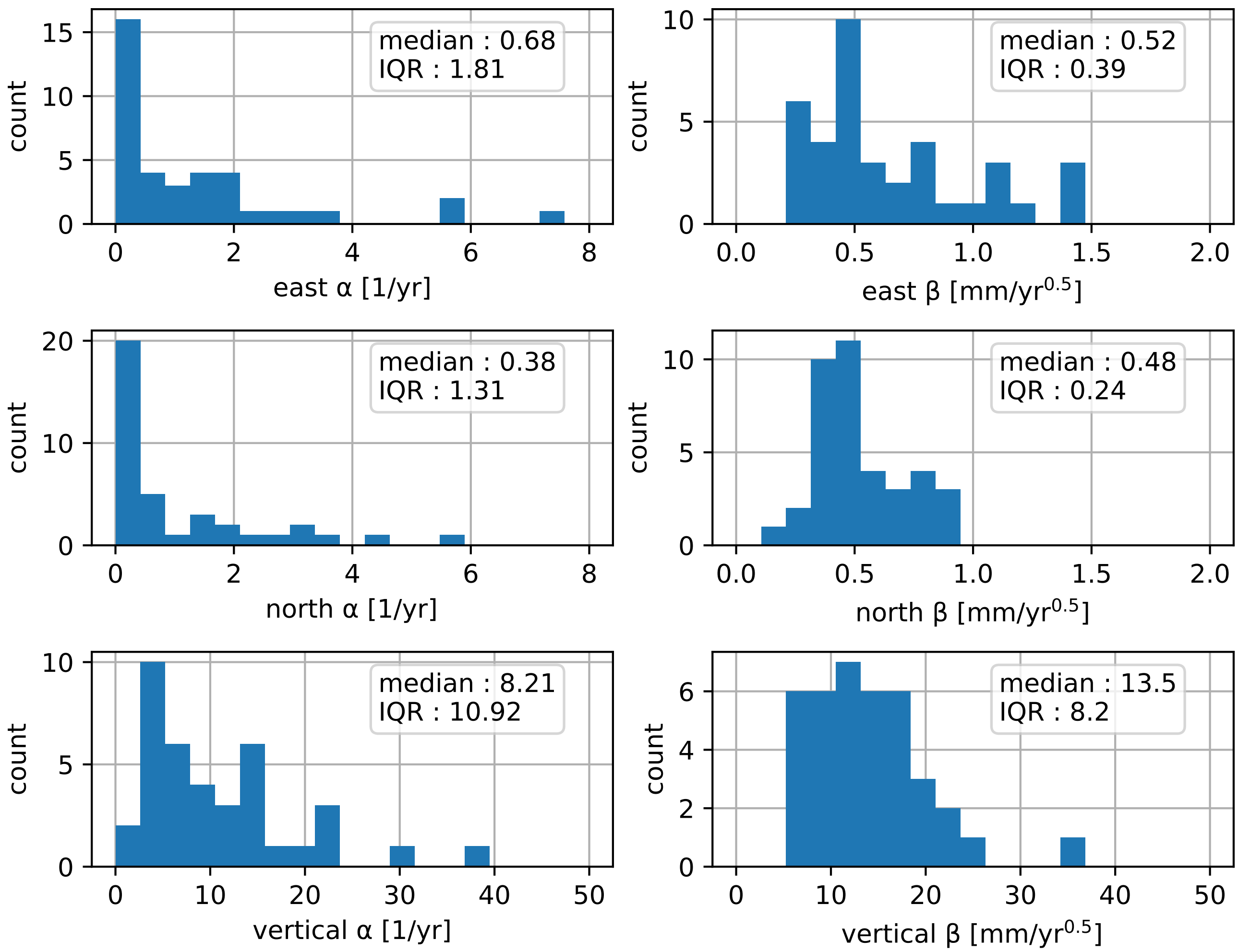

Figure 3. Distribution of ~~estimated~~ inferred values for the parameters in the FOGM ~~hyperparameters~~ noise model (eq. 28). ~~Hyperparameters are estimated with the REML method for 38 stations in Cascadia that are east of 121°W.~~ "IQR" is the interquartile range of inferred values.

1997; Dong et al. 2006), which is noise that is highly spatially correlated. When not accounted for, common mode error manifests as spatially uniform undulations in our estimated transient displacements. ~~However, estimated~~ These undulations have no effect on the transient strain rates~~are insensitive to common mode error. We therefore do not~~ , and so we do not need to include common mode error in our noise model. ~~We then make the simplifying assumption that $\eta(p) = 0$ for the easting and northing component of GNSS data.~~

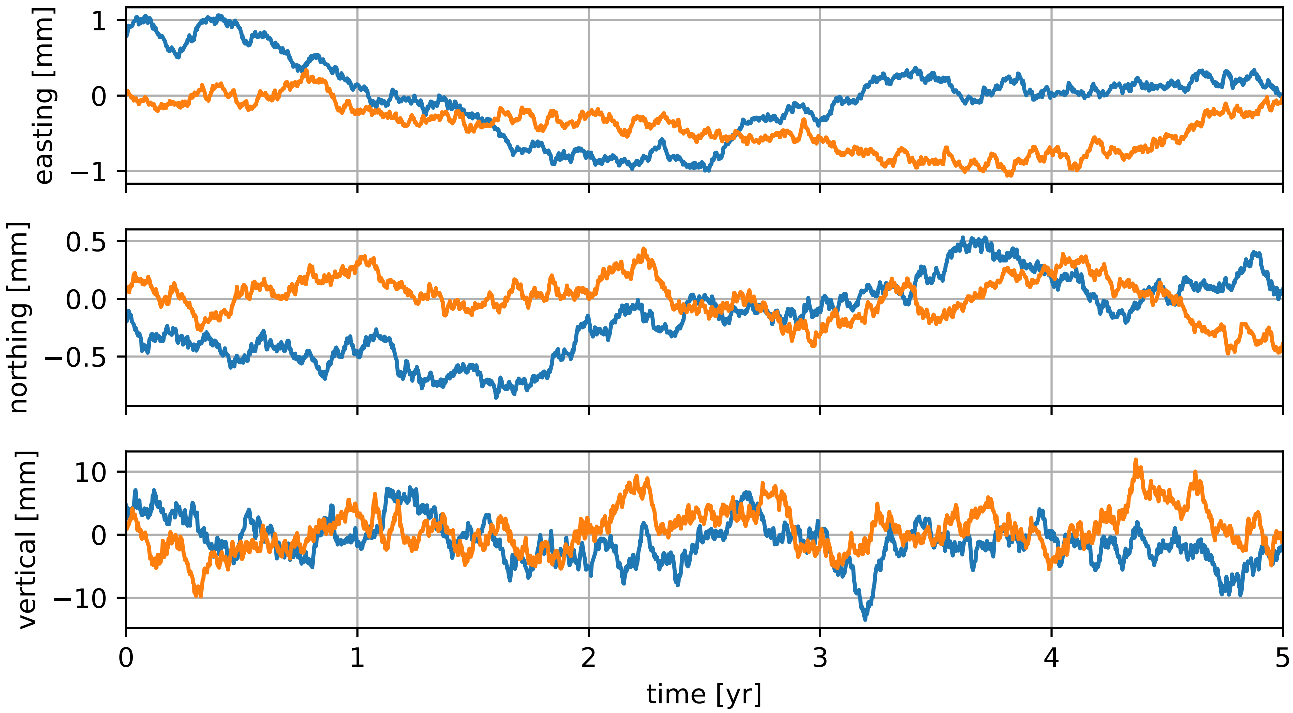

Figure 4. Two samples of FOGM noise ~~samples~~ for each displacement component. The parameters for the FOGM ~~hyperparameters~~ noise model have been set to the median values from Figure 3.

3.2 Prior model

We next establish ~~our~~ a prior model for transient displacements. Specifically, we discuss our choice for the ~~covariance functions $X(\vec{x},\vec{x}')$ and $T(t,t')$. For X~~ spatial and temporal functions making up C_{u_i}, X_i and T_i. For X_i, we use the squared exponential (SE) covariance function,

$$X_i(\vec{x},\vec{x}') = \exp\left(\cancel{\frac{-||\vec{x}-\vec{x}'||_2^2}{2\ell^2}}\,\frac{-||\vec{x}-\vec{x}'||_2^2}{2\ell^2}\right). \tag{31}$$

The SE covariance function is commonly used ~~in~~ for kriging (e.g, Cressie 1993) and Gaussian process regression (e.g., Rasmussen & Williams 2006). ~~The SE is a positive definite covariance function for any number of spatial dimensions. A Gaussian process with an SE covariance function is isotropic and has realizations that are infinitely differentiable.~~ In terms of geodetic applications, Kato et al. (1998) and El-Fiky & Kato (1998) demonstrated that the SE covariance function accurately describes the spatial covariance of secular GNSS derived velocities in Japan.

We consider three potential ~~models for the temporal covariance of u. First, we consider~~ choices for T_i. The first option is the one-dimensional SE covariance function,

$$T_i(t,t') = \phi^2 \exp\left(\frac{-|t-t'|^2}{2\tau^2}\right). \tag{32}$$

Note that ~~T includes the hyperparameter~~ T_i includes the parameter ϕ, which serves to scale the covariance function ~~C_u. Second, we consider~~ C_{u_i}. The second option is a member of the Wendland class of covariance functions (Wendland 2005). Wendland covariance functions have compact support and hence their corresponding covariance matrices are sparse. In our analysis, we exploit this sparsity with the CHOLMOD software package (Chen et al. 2008). Wendland covariance functions can be constructed such that they are positive definite in $\mathbb{R}^d$ and their corresponding Gaussian process can be differentiated k times. We use $d = 1$, because we are describing the temporal covariance of u_i, and we use $k = 2$, giving samples of u_i continuous velocities and accelerations. The corresponding Wendland covariance function is

$$T_i(t,t') = \phi^2 \left(1 - \frac{|t-t'|}{\tau}\right)_+^5 \left(\frac{8|t-t'|^2}{\tau^2} + \frac{5|t-t'|}{\tau} + 1\right), \tag{33}$$

where $(t)_+ = \max(0,t)$. The third option for T_i is the covariance function for integrated Brownian motion (IBM). IBM ~~has~~ is a Gaussian process with zero mean and its covariance function can be found by integrating the covariance function for Brownian motion~~as~~,

$$T_i(t,t') = \int_0^t \int_0^{t'} \phi^2 \min(s,s')\, ds'\, ds \tag{34}$$

$$= \frac{\phi^2}{2} \min(t,t')^2 \left(\max(t,t') - \frac{1}{3}\min(t,t')\right), \quad t,t' \geq 0. \tag{35}$$

IBM has been used in the context of Kalman filtering as a non-parametric model for the time dependence of geophysical signals ~~(e.g., Segall & Mathews 1997; McGuire & Segall 2003; Ohtani et al. 2010; Hines & Hetland 20~~ ~~It should be emphasized $t = 0$ is~~ (e.g., Segall & Mathews 1997; McGuire & Segall 2003; Ohtani et al. 2010; Hines & Hetlan Similar to Brownian motion, IBM has a reference time, $t = 0$, at which the ~~Gaussian process is exactly zero~~process begins. For some geophysical signals, it is appropriate to have this reference time. For example, if we are trying to identify postseismic deformation then t should be zero at the time of the earthquake. However, if we are ~~interesting~~ interested in detecting transient events, where there is no known start time, then IBM may not be an appropriate prior ~~, and an isotropic Gaussian process should be preferred~~model. Instead, one may prefer to use the SE or Wendland covariance functions because they are stationary. In the following analysis, we make the ~~quite arbitrary~~ choice that t is zero on the first epoch of $\cancel{\boldsymbol{d}_*}\boldsymbol{d}_i^*$. Using an earlier reference time does not change the results discussed in this section. ~~Our third option for T is the Wendland class of covariance functions (Wendland 2005). Wendland covariance functions have compact support and hence their~~

~~corresponding covariance matrices will be sparse. In our analysis, we exploit this sparsity with the CHOLMOD software package (Chen et al. 2008). Wendland functions are positive definite in $\mathbb{R}^d$, and they describes an isotropic Gaussian process with realizations that can be differentiated k times. The form of the covariance function depends on the choice of d and k. We use $d = 1$ since we are describing the temporal covariance of u. We use $k = 2$, giving samples of u continuous velocities and accelerations. The corresponding Wendland covariance function is~~

$$T(t,t') = \phi^2 \left(1 - \frac{|t-t'|}{\tau}\right)_+^5 \left(\frac{8|t-t'|^2}{\tau^2} + \frac{5|t-t'|}{\tau} + 1\right),$$

~~where~~

$$(t)_+ = \begin{cases} t, & t > 0 \\ 0, & \text{otherwise.} \end{cases}$$

~~We next determine appropriate hyperparameters for X and each of the three candidate covariance functions for T. First, we clean the GNSS datasets by removing offsets at times of equipment changes and removing outliers with the method describe in Section 2.1. For the outlier detection algorithm, our prior model, u, is chosen to have a length-scale and time-scale which is able to approximately describe SSE displacements. We use the SE covariance function for X with length-scale $\ell = 100$ km, and we use the Wendland covariance function for T, due to its computational efficiency, with time-scale $\tau = 0.1$ yr and $\phi = 1$ mm. The outlier detection algorithm is particularly effective at removing outliers for stations at high elevation (Figure A1), which can be adversely affected by ice or snow during the winter (Lisowski et al. 2008). After cleaning the dataset, we divide it~~

We use the REML method to determine the appropriate values for the parameters in C_{u_i}, which are ℓ, ϕ, and τ. We refer to these parameters collectively as $\boldsymbol{\theta}$. Just as in Section 3.1, we pick $\boldsymbol{\theta}$ to maximize the probability of drawing $\boldsymbol{d}_i^*$ from $\boldsymbol{d}_i$, but now we are including $u_i(\boldsymbol{P})$ in $\boldsymbol{d}_i$, and we are optimizing the parameters for C_{u_i} rather than $\boldsymbol{C}_{\eta_i}$. We divide the GNSS data into seven subsets ~~which~~ that are four months long and each centered on the time of a SSE. The times of the seven SSEs are determined with tremor records from Wech (2010). We ~~use the REML method to~~ find the optimal ~~hyperparameters for T and X~~ $\boldsymbol{\theta}$ for each subset of data~~. We choose to make each data subsets four months long because it is long enough to encompass a SSE in Cascadia, while it is short enough to still be computationally tractable. However, four months is too short to resolve the sinusoids in $\boldsymbol{d}$, and they are omitted from $\boldsymbol{d}$ in this REML analysis for Cascadia SSEs~~, for each displacement component, and for each choice of T_i. The REML likelihood function that we are maximizing with respect to $\boldsymbol{\theta}$ is

Table 1. Optimal ~~hyperparameters~~ values for the ~~prior on transient displacements~~ parameters in X_i and T_i determined with the REML method. The ~~temporal~~ covariance function for T_i is indicated by the "~~T~~ T_i" column. The SE, ~~IBM~~Wendland, and ~~Wendland~~ IBM covariance functions are defined in eqs. (32), (~~34~~33), and (~~33~~34), respectively. ~~The spatial covariance function, X, is the squared exponential (~~We use eq. (31) ~~in all cases~~for X_i. The ~~hyperparameters~~ parameters are estimated for each of the seven SSEs considered in this study, and the tabulated values indicate the median and interquartile ranges of ~~estimates~~estimated values. The "diff log(REML)" column compares the optimal log REML ~~likelihood~~ likelihoods to ~~the log REML likelihood~~ those for when ~~using~~ T_i is the SE covariance function~~for T~~. Negative values indicate that observations are more consistent with the SE covariance function.

~~T~~ T_i	~~direction~~ component	ℓ	ϕ	τ	diff. log(REML)
SE	east	92 ± 25 km	0.62 ± 0.11 mm	0.026 ± 0.011 yr	-
SE	north	91 ± 53 km	0.43 ± 0.05 mm	0.030 ± 0.017 yr	-
Wendland	east	95 ± 30 km	0.66 ± 0.15 mm	0.093 ± 0.044 yr	0.78 ± 0.87
Wendland	north	92 ± 57 km	0.46 ± 0.10 mm	0.116 ± 0.057 yr	0.08 ± 0.58
IBM	east	110 ± 130 km	290 ± 420 mm/yr$^{1.5}$	-	-16.4 ± 7.8
IBM	north	150 ± 560 km	110 ± 250 mm/yr$^{1.5}$	-	-10.1 ± 2.3

now

$$\mathcal{L}(\boldsymbol{\theta}) = \left(\frac{\left|\boldsymbol{G}^T\boldsymbol{G}\right|}{(2\pi)^{n-6m}\left|\boldsymbol{\Sigma}(\boldsymbol{\theta})\right|\left|\boldsymbol{G}^T\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{G}\right|} \right)^{\frac{1}{2}} e^{-\frac{1}{2}\boldsymbol{d}_i^{*T}\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{d}_i^*}, \tag{36}$$

where

$$\boldsymbol{K}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} - \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{G}\left(\boldsymbol{G}^T\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{G}\right)^{-1}\boldsymbol{G}^T\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \tag{37}$$

and $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{C}_{\eta_i} + C_{u_i}(\boldsymbol{P},\boldsymbol{P};\boldsymbol{\theta})$. In the above equation, $\boldsymbol{d}_i^*$ consists of a single subset of data. The estimated ~~hyperparameters for u~~ parameters are summarized in Table 1. Based on the interquartile ranges, the estimated ~~hyperparameters~~ parameters for the SE and Wendland covariance functions do not vary significantly between SSEs. This suggests that the median of ~~estimated hyperparameters should be an appropriate prior model for all Cascadia~~ the estimated parameters should be appropriate for all the Pacific Northwest SSEs. For the IBM model, there are several anomalously large values for ℓ and ϕ, which ~~results in large interquartile ranges~~ is why the interquartile ranges are large.

~~Detrended easting component of displacements at station SC03, which is located on Mount Olympus in Washington. The orange markers indicate outliers that were automatically detected using the algorithm from Section 2.1. The error bars show one standard deviation uncertainties. Note that outliers tend be observed in the winter, suggesting that they were caused by snow or ice.~~

~~Next~~

Next, we identify which covariance function for ~~T~~ T_i best describes the time dependence of deformation from SSEs. One approach is to ~~compare the REML likelihoods for each covariance function~~ choose the covariance function that produces the largest optimal REML likelihoods, similar to the analysis in Langbein (2004). In Table 1, we summarize how the ~~log~~ optimal REML likelihoods for the Wendland and IBM covariance functions compare to those for the SE covariance function. Based on the differences in ~~log~~ optimal REML likelihoods, the data is substantially more likely to come from a Gaussian process with a SE or Wendland covariance function than an IBM covariance function. The REML likelihoods do not definitively indicate whether the SE or Wendland covariance function is preferable.

~~To further explore which covariance function for T best describes SSEs, we compare the observations to the predicted displacementsfor each covariance function. We consider the data prediction vector to be $\hat{\boldsymbol{d}} = (u(\boldsymbol{P}) + \boldsymbol{Ga})\,|\boldsymbol{d}_*$. It~~

A more intuitive way of deciding which function to use for T_i is to compare the posterior displacements and the observed displacements. The posterior displacements consist of our estimate of transient displacements, secular trends, and seasonal deformation. Specifically, the posterior displacement vector is $\hat{\boldsymbol{d}}_i = (u_i(\boldsymbol{P}) + \boldsymbol{Gm}_i)\,|\,(\boldsymbol{d}_i = \boldsymbol{d}_i^*)$. Following a similar procedure as in Section 2, it can be shown that ~~$\hat{\boldsymbol{d}}$ is normally distributed with mean~~ $\hat{\boldsymbol{d}}_i$ has a multivariate normal distribution with mean vector

$$\boldsymbol{\mu}_{\cancel{\hat{d}}\hat{d}_i} = \begin{bmatrix} C_{u_i}(\boldsymbol{P},\boldsymbol{P}) & \boldsymbol{G} \end{bmatrix} \begin{bmatrix} \boldsymbol{C}_{\eta_i} + C_{u_i}(\boldsymbol{P},\boldsymbol{P}) & \boldsymbol{G} \\ \boldsymbol{G}^T & \boldsymbol{0} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{d}_i^* \\ \boldsymbol{0} \end{bmatrix} \tag{38}$$

and covariance matrix

$$\boldsymbol{C}_{\cancel{\hat{\boldsymbol{d}}}\hat{d}_i} = C_{\cancel{u}u_i}(\boldsymbol{P},\boldsymbol{P}) - \begin{bmatrix} C_{u_i}(\boldsymbol{P},\boldsymbol{P}) & \boldsymbol{G} \end{bmatrix} \begin{bmatrix} \boldsymbol{C}_{\eta_i} + C_{u_i}(\boldsymbol{P},\boldsymbol{P}) & \boldsymbol{G} \\ \boldsymbol{G}^T & \boldsymbol{0} \end{bmatrix}^{-1} \begin{bmatrix} C_{u_i}(\boldsymbol{P},\boldsymbol{P}) \\ \boldsymbol{G}^T \end{bmatrix}. \tag{39}$$

We compute ~~$\hat{\boldsymbol{d}}$ using~~ $\hat{\boldsymbol{d}}_i$ using the SE, Wendland, and IBM covariance functions for ~~T and the median hyperparameters from Table 1.~~ T_i, and we use the median values from Table 1 for $\boldsymbol{\theta}$. Figure 5 compares ~~the easting component of $\boldsymbol{d}_*$ to $\hat{\boldsymbol{d}}$ for the Winter~~ $\boldsymbol{d}_i^*$ to $\hat{\boldsymbol{d}}_i$ during the winter 2015-2016 SSE at ~~Station~~ stations ALBH, P436~~. The data prediction vector appears to accurately describe displacements throughout the SSE, regardless of the choice of T. Since the SSE signal is strongest at Station P436, it can be assumed that T adequately describes the signal elsewhere. The prediction~~ , and SC02, which are the stations that recorded the strongest signal from this SSE. Based on Figure 5, the chosen covariance function for T_i has almost no effect on $\hat{\boldsymbol{d}}_i$. The posterior displacements for the IBM covariance function ~~contains~~ contain slightly more high frequency, and perhaps spurious, features. ~~The predictions for the Wendland and SE covariance functions are nearly indistinguishable. In~~ Regardless of the chosen covariance function, $\hat{\boldsymbol{d}}_i$ appears to underestimate the rate of deformation during the SSE at stations ALBH and SC02. The deformation at these two stations is particularly rapid compared to the surrounding stations, and the misfit between $\boldsymbol{d}_i^*$ and $\hat{\boldsymbol{d}}_i$ is likely due to over-smoothing. This over-smoothing could indicate that the chosen values for τ or ℓ are too large. However, $\hat{\boldsymbol{d}}_i$ does faithfully fit $\boldsymbol{d}_i^*$ at the remaining stations, and so we do not attempt to further refine the parameters.

For our estimates of transient strain discussed in the next section, we ultimately settle on the Wendland covariance function for ~~T and~~ T_i, and we use the median values from Table 1 for the ~~hyperparameters~~parameters. We choose the Wendland covariance function over the SE covariance function because of its computational advantages.

3.3 Transient Strain Rates

Having established a noise model and a prior for transient displacements, we ~~use the cleaned GNSS dataset to~~ can now calculate transient strain rates, $\dot{\varepsilon}$, in the Puget Sound region ~~. We calculate transient strain rates~~ from the GNSS data. We evaluate $\dot{\varepsilon}$ at a grid of points spanning the study area for each day from January 1, 2010 to May 15, 2017. ~~The strain rates are estimates at a grid of points spanning the study area.~~ In Figure 6~~we show the transient strain rates~~ , we show a map view of $\dot{\varepsilon}$ on January 1, 2016, which coincides with the height of ~~an SSE. We have included a video showing the~~ the winter 2015-2016 SSE. In Figure 7, we show a time series of $\dot{\varepsilon}$ at a position on the Olympic Peninsula, which is where $\dot{\varepsilon}$ tends to be the largest during SSEs. We also include a supplementary animation showing a map view of ~~strain rates through timeas supplementary material. The~~ $\dot{\varepsilon}$ over time. In each figure, we show the mean and standard deviation of $\dot{\varepsilon}$, making it easy to identify which features are statistically significant.

The transient strain rates shown for the winter 2015-2016 SSE in Figure 6 are ~~generally similar to the strain rates for the other six SSEs considered in this study~~characteristic of most SSEs in the Puget Sound region. The SSEs cause trench perpendicular compression in the Olympic Peninsula and extension east of Puget Sound. ~~For comparison, estimated secular strain rates indicate~~ The strain transitions from compression to extension around the southern tip of Vancouver Island, coinciding with where fault slip tends to be inferred (e.g., Dragert et al. 2001; Wech et al. 2009; Schmidt & Gao 2010). Thus, this pattern of strain is to be expected. Over decadal timescales, there is trench perpendicular compression throughout this study region ~~(Murray & Lisowski 2000; McCaffrey et al. 2007, 2013). The SSEs are thus~~ caused by steady tectonic plate motion (Murray & Lisowski 2000; McCaffrey et al. 2007, 2013). When comparing the transient strain rates caused by SSEs to the secular strain rates, we see that SSEs are concentrating tectonically accumulated strain energy ~~trench-ward~~towards the trench, and presumably pushing the subduction zone closer to failure. ~~Similar conclusions have been drawn based on fault slip models (e.g., Dragert et al. 2001; Wech et al. 2009; Schmidt & Gao 2010), which reveal that~~

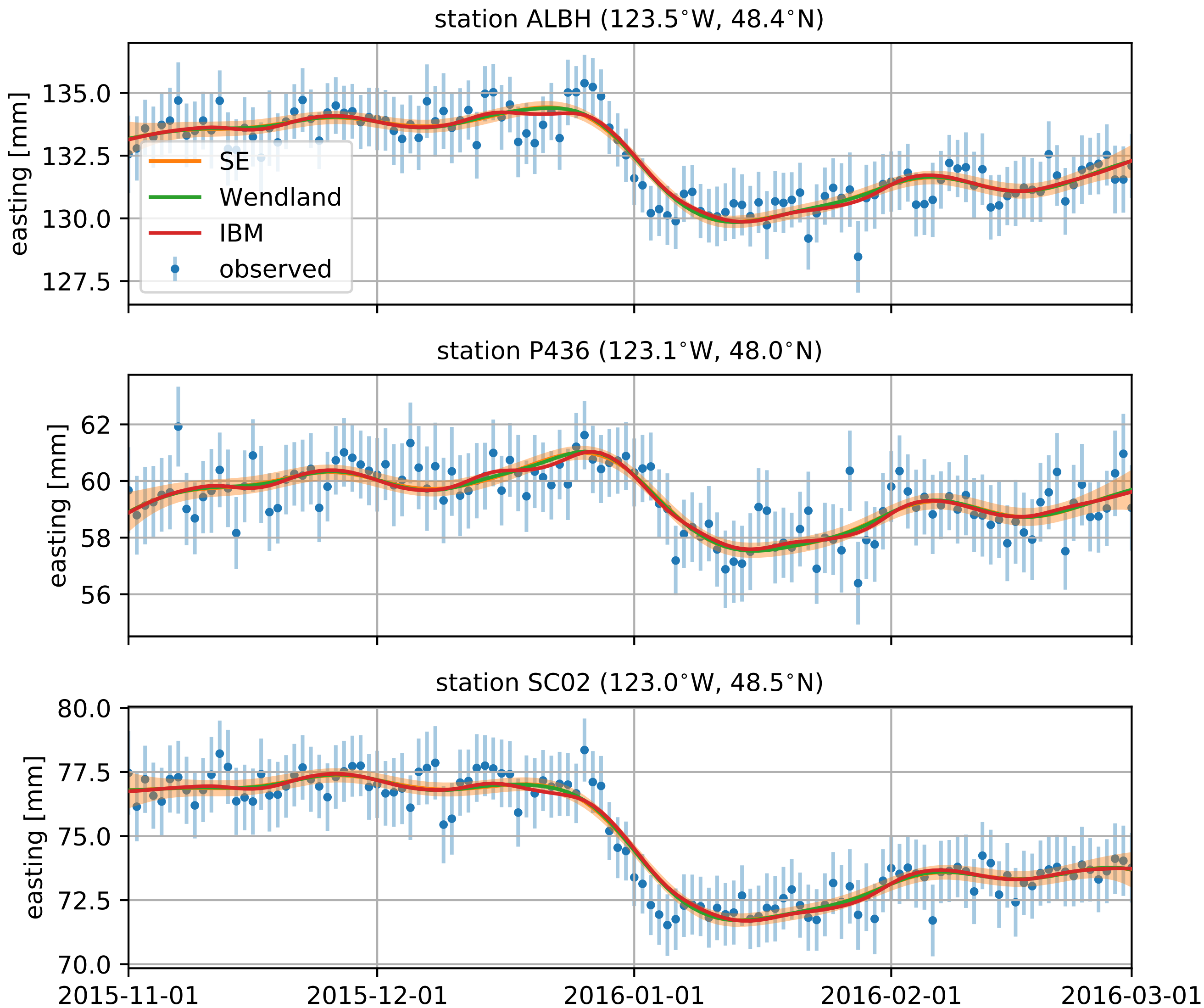

Figure 5. ~~Observed easting~~ Easting component of the observed displacements~~at station P436~~, $\boldsymbol{d}^*_{\mathrm{e}}$, and ~~predicted~~ posterior displacements, $\hat{\boldsymbol{d}}_{\mathrm{e}}$, during the winter 2015-2016 SSE. We show $\hat{\boldsymbol{d}}_{\mathrm{e}}$ for when ~~using different~~ T_i is a SE, Wendland, and IBM covariance ~~functions for T~~function. The one standard deviation uncertainties are shown for ~~the observations~~ $\boldsymbol{d}^*_{\mathrm{e}}$ and the ~~predicted displacements when using the~~ $\hat{\boldsymbol{d}}_{\mathrm{e}}$ that has a SE covariance function for T_i. For clarity, uncertainties are not shown for the $\hat{\boldsymbol{d}}_{\mathrm{e}}$ that have an IBM ~~and~~ or Wendland covariance ~~functions, but they are nearly equivalent to the uncertainties~~ function for T_i. The posterior displacements for the ~~SE~~ different covariance ~~function~~functions are all practically indistinguishable.

~~SSEs are occurring down-dip of the seismogenic zone and migrating stress upward. A key difference between the strain inferred here and strain that can be derived from fault slip models is that our estimated strain rates are not based on an assumed physical model. In contrast, fault slip models can be biased by errors in the assumed fault geometry or lithospheric rheology. Moreover, the degrees of~~

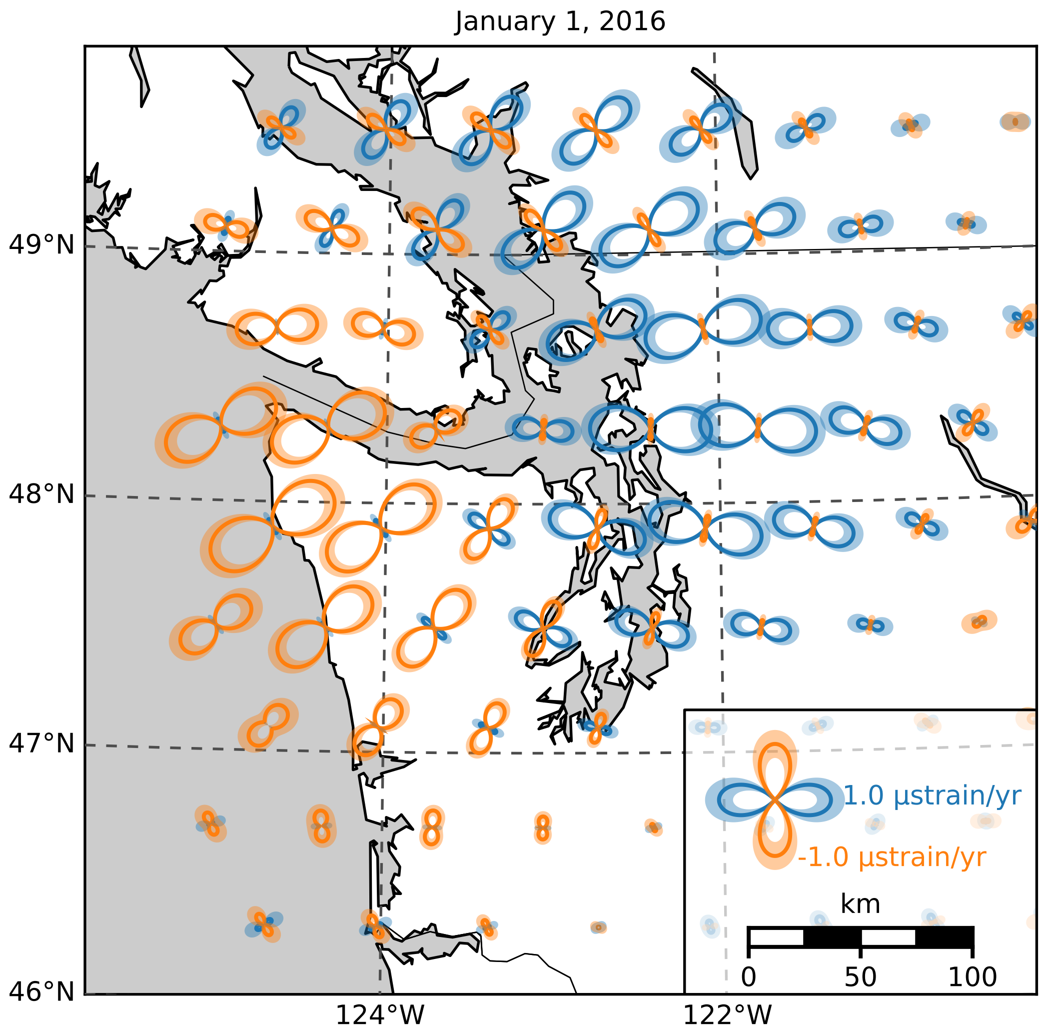

Figure 6. ~~Estimated transient~~ Transient strain rates during the Winter 2015-2016 SSE. ~~Strain~~ The glyphs show the normal strain ~~rate along each~~ rates as a function of azimuth, where orange indicates compression and blue indicates extension. The shaded regions indicate the one standard deviation uncertainties ~~in~~ for the normal strain rates.

~~freedom in fault slip models usually cannot be constrained by GNSS data alone, and it is necessary to impose regularization which further biases the results. Since our estimated strain rates lack such systematic errors, we can be more confident that our solution is unbiased and has meaningful uncertainties.~~

~~In Figure 7 we show the time dependence of estimated transient strain rates at a position on the Olympic Peninsula, where transient strain rates from SSEs are largest.~~

To verify that the estimated transient strain rates are accurately identifying geophysical signal, we

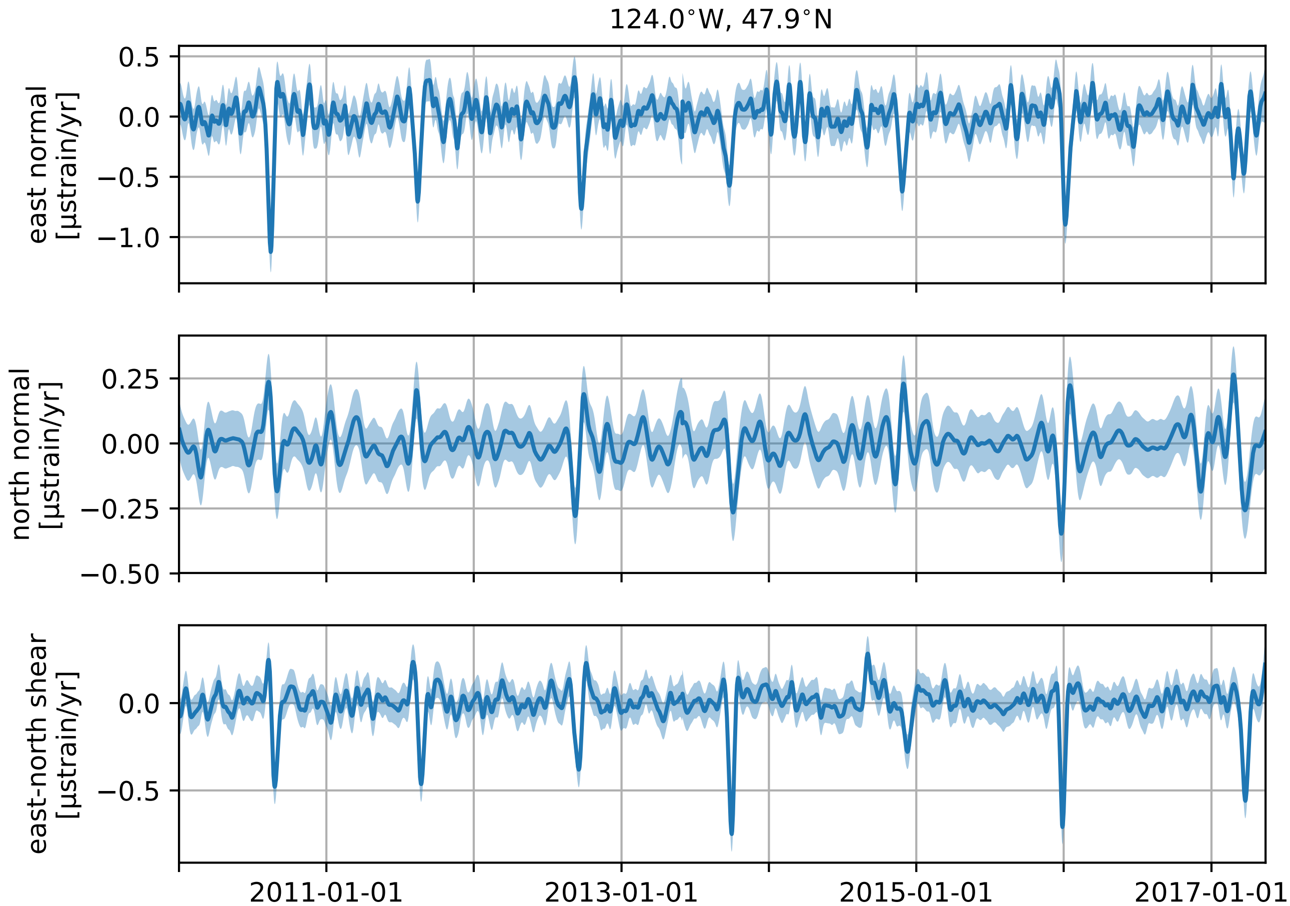

Figure 7. ~~Three components~~ Components of the transient ~~horizontal~~ strain ~~rate tensor estimated~~ rates at the position shown in Figure 2. The shaded regions indicate the one standard deviation ~~uncertainty~~uncertainties.

compare the signal-to-noise ratio from eq. (23) at a point on the Olympic Peninsula to the frequency of seismic tremor (Figure 8). A signal-to-noise ratio greater than $\sim$3 can be interpreted as a detected geophysical signal. ~~For each detected event there is a corresponding peak~~ We detect nine distinct events, which each correspond to peaks in seismic tremor. ~~We are also able to clearly identify transient strain associated with a more subtle SSE in August 2014. In between SSEs~~The smaller events detected in August 2014 and February 2017 can be considered inter-SSE events. They were not among the SSEs used to constrain the prior covariance function. In between peaks in seismic tremor, the signal-to-noise ratio is consistently between 0 and 2, ~~indicating that no anomalous non-SSE events are being detected, at least at this location~~suggesting that all the transient strain detected at this point on the Olympic Peninsula is associated with SSEs and inter-SSE events.

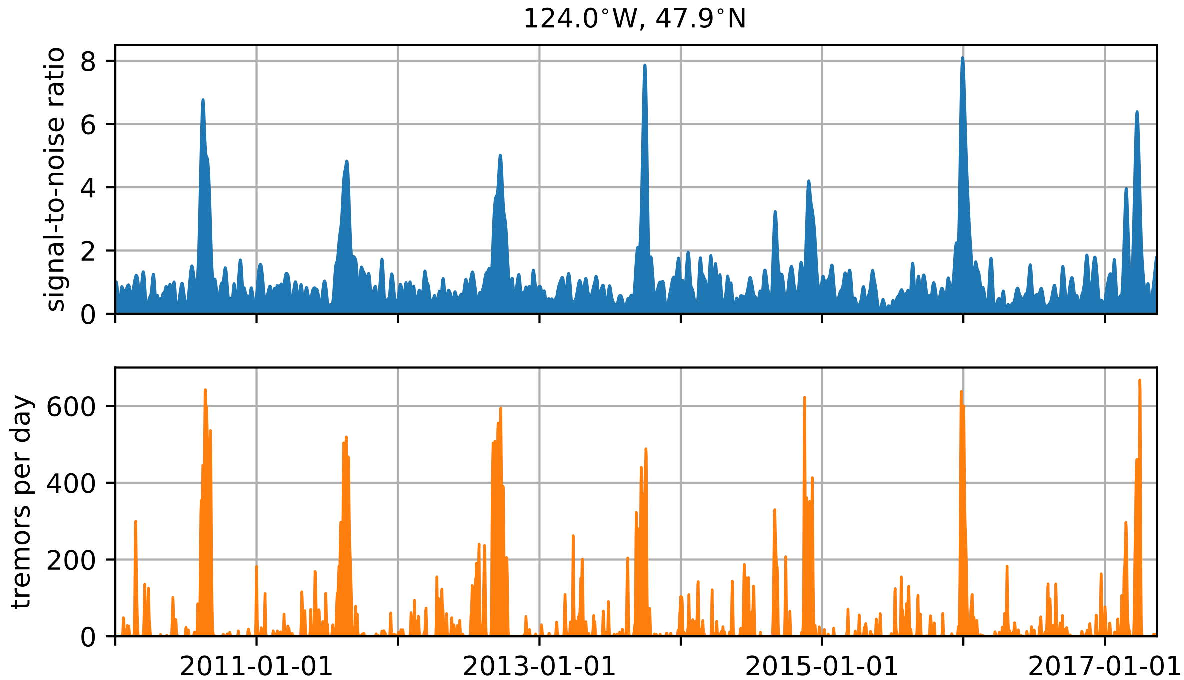

Figure 8. (top) Signal-to-noise ratio (eq. 23) at the position shown in Figure 2. (bottom) Frequency of seismic tremors in the region shown in Figure 2.

4 ~~DISCUSSION~~

~~We have demonstrated that transient strain rates estimated with the method described in Section 2 can be used to detect SSEs and will be robust to detect other transient geophysical phenomena. Another potential application would be to use GNSS derived transient strain rates to develop noise models for borehole strain meters (BSMs). The Plate Boundary Observatory maintains 43 BSMs in Cascadia, and it has been demonstrated that BSMs are able to record transient geophysical events such as SSEs (e.g., Dragert & Wang 2011)~~

The results we have presented indicate that we are identifying the transient strain that we should expect to see. This is not to say that there are no unexpected features in $\dot{\varepsilon}$ that are worth further exploration. However, ~~noise in BSM data is not well understood, which complicates the use of BSM data for detecting geophysical events. The noise consists, in part, of a long-term decay resulting from the instrument equilibrating with the surrounding rock (Gladwin et al. 1987). Typically, this noise is dealt with in an ad-hoc manner by fitting and removing exponentials and low-order polynomials. A more rigorous quantification of BSM noise could be performed by analyzing the residuals between the the BSM data and GNSS derived strain rates.~~ a discussion on features in $\dot{\varepsilon}$ that have unknown origin would be outside the scope of this paper.

~~There~~

4 DISCUSSION

Our results demonstrate that GPR is an effective tool for estimating transient strain rates and detecting geophysical processes from GNSS data. However, there are some aspects of our method that warrant further research. For example, there is potential for a more thorough analysis of the spatio-temporal noise in ~~GNSS data, η,~~ the GNSS data than what was performed in Section 3.1. ~~We~~ Specifically, we did not explore ~~the spatial covariance of η. Spatially correlated common mode noise, which results from factors such as reference frame error, is a non-negligible component of GNSS data. We ignore~~ spatially correlated noise, such as common mode error~~in this study because transients strain rates are insensitive to it. For other geophysical studies based on GNSS data, such as fault slip inversions, it may be necessary to incorporate a spatially covarying noise model (e.g., Miyazaki et al. 2003)~~. We have assumed that any spatially correlated noise has a sufficiently long wavelength that it has a negligible effect on transient strain rates. We can also improve upon the seasonal noise model used in this study, which consists of four ~~spatially uncorrelated~~ sinusoids for each station. We did not explore the ~~spatial covariance of seasonal deformation or the temporal roughness~~ roughness of the seasonal deformation (i.e., the number of sinusoids needed to describe the ~~observations~~deformation). The periodic Gaussian process (Mackay 1998) is an alternative model for seasonal deformation and is well suited for exploring the roughness of seasonal deformation. The periodic Gaussian process has zero mean and the covariance function

$$T(t,t') = \phi^2 \exp\left(\frac{-\sin(\pi|t-t'|)^2}{2\tau^2}\right). \tag{40}$$

Realizations have annual periodicity and the roughness is controlled by τ. Decreasing τ has the same effect as including higher frequency sinusoids in the seasonal model. The optimal value for τ can be found with the REML method~~as described in Section 3.1~~.

Another potential research direction would be to reduce the computational cost of our method for estimating transient strain rates. GPR is generally computationally expensive when there are many observations. The transient strain rates estimated in this study are constrained by ~~about~~ seven years of daily displacement observations from 94 GNSS stations~~. It can be computationally intensive to evaluate~~ , which amounts to about 240,000 observations for each displacement component. For a dataset with this size, it is difficult to evaluate the matrix inverses in eqs. (10) and (11)~~for a dataset with this size. We significantly reduce the amount of memory needed to estimate transient strain rates by describing the temporal covariance of displacements with~~. We alleviate this computational cost by using a compact Wendland covariance function ~~. Using~~ for our prior. By using a compact covariance function for our prior~~turns~~ , eqs. (10) and (11) ~~into~~ become sparse systems of equations ~~, which we then solve with CHOLMOD. CHOLMOD is designed for solving sparse, positive definite systems of equations. The matrix being inverted in eqs. (10) and (11) is not positive definite; however, we can use another partitioned matrix inversion identity from Press et al. (2007) to partition it into positive definite submatrices to be inverted. Even when using a compact covariance function, it may still be necessary to reduce the computational burden by dividing the data into subsets and evaluating transient strain rates for each subset~~that are more tractable to evaluate. However, the sparsity decreases when the Wendland covariance function has a larger time-scale parameter. The Wendland covariance then offers less of a computational advantage when we were interested in geophysical processes that occur over longer times-scales. An alternative approach to dealing with large datasets would be to explore approximation methods for GPR (Rasmussen & Williams 2006, sec. 8).

In addition to detecting geophysical processes, the GNSS derived transient strain rates can be used to better understand the data from borehole strain meters (BSMs). The Plate Boundary Observatory contains about forty BSMs in the Pacific Northwest, and it has been demonstrated that BSMs are able to record transient geophysical events such as SSEs (e.g., Dragert & Wang 2011). However, there are complications that prevent BSM data from being used quantitatively in geophysical studies. One difficulty is that BSM data should be calibrated with a well known strain source, such as diurnal and semidiurnal tides (Hart et al. 1996; Roeloffs 2010; Hodgkinson et al. 2013). Unfortunately, the tidal forces at BSMs which record SSEs are strongly influenced by local bodies of water such as the Straight of Juan de Fuca, making it difficult to form a theoretical prediction of tidal strains (Roeloffs 2010). Another complication is that noise in BSM data is not well understood. The noise consists, in part, of a long-term decay resulting from the instrument equilibrating with the surrounding rock (Gladwin et al. 1987). Typically, this noise is dealt with in an ad-hoc manner by fitting and removing exponentials and low-order polynomials. We envision that the GNSS derived strain rates from this paper can be used as a reference strain for calibrating BSM data and quantify its noise.

5 CONCLUSION

~~In this paper we~~

We propose using Gaussian process regression (GPR) to estimate transient strain rates from GNSS data. Most other methods for estimating strain rates assume a parametric representation of deformation, which can bias the results if the parameterization is not chosen carefully. Here we assume a stochastic, rather than parametric, prior model for displacements. Our prior model describes how much we expect transient displacements to covary spatially and temporally. If we know nothing about the underlying signal that we are trying to recover, then the prior model can be chosen objectively with maximum likelihood methods. ~~Because GPR is a Bayesian method, the uncertainties on our estimated transient strain rates are well quantified, allowing one to discern geophysical signal from noise.~~ We demonstrate that GPR is an effective tool for detecting geophysical ~~phenomena~~processes, such as slow slip events, in our application to ~~GNSS data from Cascadia. One limitation with GPR is that it is not robust against outliers. To overcome this limitation, we have introduced an effective pre-processing method for identifying and removing outliers from GNSS datasets. Another complication with GPR is that it usually involves inverting a dense matrix where the number of rows and columns is equal to the number of observations. This is prohibitive when using several years of daily GNSS observations from a network of several hundred stations. We significantly reduce the computational burden of GPR by using compact Wendland covariance function to describe our prior model~~the Pacific Northwest. While this paper just focuses on ~~estimating~~ using GPR to estimate transient strain rates and detect geophysical processes, we believe that GPR is a powerful tool that can be applied to a wide range of geophysical problems.

6 ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under grant EAR 1245263. The EarthScope Plate Boundary Observatory data is provided by UNAVCO through the GAGE Facility with support from the National Science Foundation (NSF) and National Aeronautics and Space Administration (NASA) under NSF Cooperative Agreement EAR-1261833. An implementation of the method described in this paper is named Python-based Geodetic Network Strain software (PyGeoNS). PyGeoNS is distributed under the MIT License and can be found at www.github.com/treverhines/PyGeoNS.

APPENDIX A: OUTLIER DETECTION ALGORITHM

Our outlier detection algorithm is loosely based on the data editing algorithm from Acheson (1975). Let $\boldsymbol{d}^*$ denote all n GNSS displacement observations for a single directional component, which have been made at positions and times $\boldsymbol{P}$. We describe $\boldsymbol{d}^*$ as a realization of the random vector

$$\boldsymbol{d} = \boldsymbol{G}\boldsymbol{m} + v(\boldsymbol{P}) + \boldsymbol{w}, \tag{A.1}$$

where $\boldsymbol{G}$ and $\boldsymbol{m}$ are the same as in eq. (4), v is a Gaussian process distributed as $\mathcal{GP}(0, C_v)$, and $\boldsymbol{w}$ is a vector of uncorrelated Gaussian noise with known standard deviations $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, ..., \sigma_n]$. The Gaussian process v is intended to describe transient features in the data that cannot be explained by the linear trend or seasonal terms in $\boldsymbol{G}$. We let the temporal covariance of v be a squared exponential, and we let v be spatially uncorrelated so that,

$$C_v(p, p') = \phi^2 \exp\left(\frac{-|t - t'|^2}{2\tau^2}\right) \delta_{\vec{x},\vec{x}'}, \tag{A.2}$$

where $\delta_{\vec{x},\vec{x}'} = 1$ if $\vec{x} = \vec{x}'$ and 0 otherwise. The spatial covariance of v has little effect on the detected outliers, and so we have assumed that v is spatially uncorrelated for simplicity. Based on our experience, v can reasonably describe most transient features in the data when we set $\phi = 1$ mm and $\tau = 10$ days.

Our goal is to find the index set of non-outliers in $\boldsymbol{d}^*$, which we denote as $\boldsymbol{\Omega}$. We use a tilde to indicate that an array only contains elements corresponding to $\boldsymbol{\Omega}$ (e.g., the vector of non-outlier observations is denoted $\tilde{\boldsymbol{d}}^* = [d_i^*]_{i\in\boldsymbol{\Omega}}$). The outliers are identified iteratively, and we initiate $\boldsymbol{\Omega}$ with all n indices. We consider outliers to be data that are poorly explained by the model $\boldsymbol{Gm} + v(\boldsymbol{P})$, which is determine by the residual vector

$$\begin{aligned} \boldsymbol{r} &= \boldsymbol{d}^* - \mathrm{E}\left[\left(\boldsymbol{Gm} + v(\boldsymbol{P})\right)\middle|\left(\tilde{\boldsymbol{d}} = \tilde{\boldsymbol{d}}^*\right)\right] \\ &= \boldsymbol{d}^* - \begin{bmatrix} C_v(\boldsymbol{P}, \tilde{\boldsymbol{P}}) & \boldsymbol{G} \end{bmatrix} \begin{bmatrix} C_v(\tilde{\boldsymbol{P}}, \tilde{\boldsymbol{P}}) + \mathrm{diag}(\tilde{\boldsymbol{\sigma}}^2) & \tilde{\boldsymbol{G}} \\ \tilde{\boldsymbol{G}}^T & \boldsymbol{0} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\boldsymbol{d}}^* \\ \boldsymbol{0} \end{bmatrix}. \end{aligned} \tag{A.3}$$

Data with abnormally large residuals are identified as outliers. For each iteration, we compute $\boldsymbol{r}$ and then update $\boldsymbol{\Omega}$ so that it contains the indices of $\boldsymbol{r}$ whose weighted values are less than λ times the weighted root mean square of $\tilde{\boldsymbol{r}}$,

$$\boldsymbol{\Omega} \leftarrow \left\{ i : \left|\frac{r_i}{\sigma_i}\right| < \lambda \cdot \sqrt{\frac{1}{|\boldsymbol{\Omega}|} \sum_{j\in\boldsymbol{\Omega}} \frac{r_j^2}{\sigma_j^2}} \right\}. \tag{A.4}$$

Iterations continue until the new $\boldsymbol{\Omega}$ is the same as the previous $\boldsymbol{\Omega}$.

The outlier detection algorithm is demonstrated in Figure A1. For the demonstration, we use the easting component of displacements at a single station, SC03, which is located on Mt. Olympus in Washington state. Station SC03 records anomalous observations during the winter, presumably because of snow and ice accumulation, and we want to remove these observations. The station also records periodic westward motion from slow slip events, and we want to keep this deformation intact. The detected outliers are shown in Panel B. For comparison, we also show the detected outliers when we do not include the Gaussian process v in our model for the data (Panel A). When v is not included, real transient deformation resulting from slow slip events is erroneously identified as outliers. When v is included, the identified outliers only consist of the anomalous deformation that lacks temporal continuity. It should be noted that we use $\lambda = 2.5$ for this demonstration, which causes the outlier

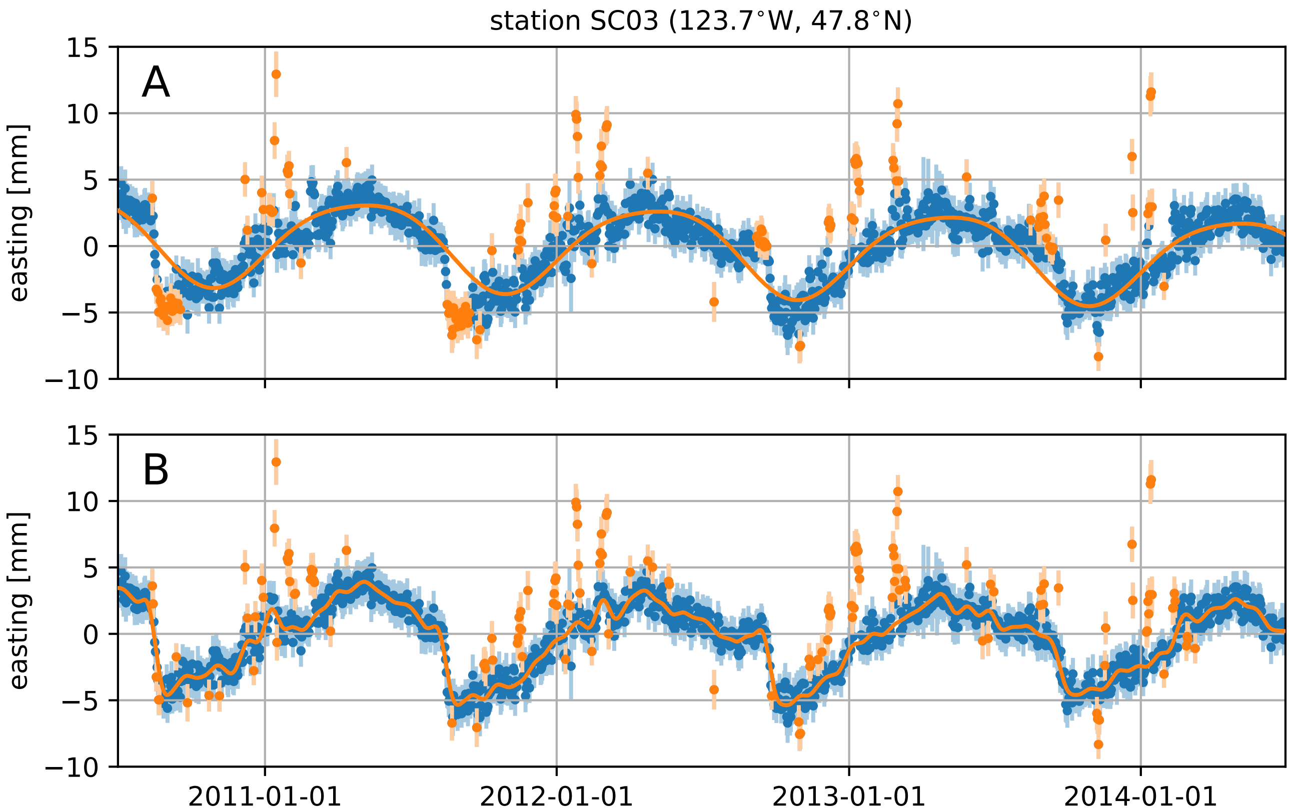

Figure A1. Outliers detected in the easting component of displacements at station SC03. The orange markers indicate detected outliers. The orange line is the best fit model to the data, which is used to compute the residual vector $\boldsymbol{r}$. The model being fit to the data in Panel A is $\boldsymbol{Gm}$, and the model in Panel B is $\boldsymbol{Gm} + v(\boldsymbol{P})$.

detection algorithm to be particularly aggressive. In Section 3, we clean the data using a more tolerant $\lambda = 4.0$.

REFERENCES

Acheson, D. T., 1975. Data editing - subroutine editq, Tech. rep., US Department of Commerce, National Oceanic and Atmospheric Administration, Environmental Data Service.

Adler, R. J., 1981. *The Geometry of Random Fields*, John Wiley & Sons, Chichester.

Beavan, J. & Haines, J., 2001. Contemporary horizontal velocity and strain rate fields of the Pacific-Australian plate boundary zone through New Zealand, *Journal of Geophysical Research*, **106**(B1), 741–770.

Blewitt, G., Kreemer, C., Hammond, W. C., & Gazeaux, J., 2016. MIDAS robust trend estimator for accurate GPS station velocities without step detection, *Journal of Geophysical Research : Solid Earth*, **121**, 2054–2068.

Bos, M. S., Fernandes, R. M. S., Williams, S. D. P., & Bastos, L., 2013. Fast error analysis of continuous GNSS observations with missing data, *Journal of Geodesy*, **87**(4), 351–360.

Chen, Y., Davis, T. a., & Hager, W. W., 2008. Algorithm 887 : CHOLMOD , Supernodal Sparse Cholesky Factorization and Update/downdate, *ACM Transactions on Mathematical Software*, **35**(3), 1–12.

Cressie, N., 1993. *Statistics for Spatial Data*, John Wiley & Sons, New York, rev. edn.

Dong, D., Fang, P., Bock, Y., Cheng, M. K., & Miyazaki, S., 2002. Anatomy of apparent seasonal variations from GPS-derived site position time series, *J. Geophys. Res.*, **107**(B4), 2075.

Dong, D., Fang, P., Bock, Y., Webb, F., Prawirodirdjo, L., Kedar, S., & Jamason, P., 2006. Spatiotemporal filtering using principal component analysis and Karhunen-Loeve expansion approaches for regional GPS network analysis, *Journal of Geophysical Research: Solid Earth*, **111**(3), 1–16.

Dragert, H. & Wang, K., 2011. Temporal evolution of an episodic tremor and slip event along the northern Cascadia margin, *Journal of Geophysical Research: Solid Earth*, **116**(12), 1–12.

Dragert, H., Wang, K., & James, T. S., 2001. A silent slip event on the deeper Cascadia subduction interface., *Science*, **292**, 1525–1528.

El-Fiky, G. S. & Kato, T., 1998. Continuous distribution of the horizontal strain in the Tohoku district, Japan, predicted by least-squares collocation, *Journal of Geodynamics*, **27**(2), 213–236.

Feigl, K. L., King, R. W., & Jordan, T. H., 1990. Geodetic measurement of tectonic deformation in the Santa Maria Fold and Thrust Belt, California, *Journal of Geophysical Research: Solid Earth*, **95**(B3), 2679–2699.

Field, E. H., Arrowsmith, R. J., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., Jackson, D. D., Johnson, K. M., Jordan, T. H., Madden, C., Michael, A. J., Milner, K. R., Page, M. T., Parsons, T., Powers, P. M., Shaw, B. E., Thatcher, W. R., Weldon, R. J., & Zeng, Y., 2014. Uniform California Earthquake Rupture Forecast, version 3 (UCERF3) -The time-independent model, *Bulletin of the Seismological Society of America*, **104**(3), 1122–1180.

Freed, a. M. & Lin, J., 2001. Delayed triggering of the 1999 Hector Mine earthquake by viscoelastic stress transfer., *Nature*.

Gazeaux, J., Williams, S., King, M., Bos, M., Dach, R., Deo, M., Moore, A. W., Ostini, L., Petrie, E., Roggero, M., Teferle, F. N., Olivares, G., & Webb, F. H., 2013. Detecting offsets in GPS time series: First results from the detection of offsets in GPS experiment, *Journal of Geophysical Research: Solid Earth*, **118**(5), 2397–2407.

Gibbs, B. P., 2011. *Advanced Kalman Filtering, Least-Squares and Modeling: A Practical Handbook*, John Wiley & Sons, Hoboken, NJ.

Gladwin, M. T., Gwyther, R. L., Hart, R., Francis, M., & Johnston, M. J. S., 1987. Borehole tensor strain measurements in California, *Journal of Geophysical Research: Solid Earth*, **92**(B8), 7981–7988.

Hammond, W. C., Blewitt, G., & Kreemer, C., 2016. GPS imaging of vertical land motion in California and Nevada: Implications for Sierra Nevada uplift, *Journal of Geophysical Research : Solid Earth*.

Hart, R. H. G., Gladwin, M. T., Gwyther, R. L., Agnew, D. C., & Wyatt, F. K., 1996. Tidal calibration of borehole strain meters: Removing the effects of small-scale inhomogeneity, *Journal of Geophysical Research*,

101(96).

Harville, D. A., 1974. Bayesian Inference for Variance Components Using Only Error Contrasts, *Biometrika*, **61**(2), 383–385.

Hastie, T., Tibshirani, R., & Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, 2nd edn.

Herring, T. A., Melbourne, T. I., Murray, M. H., Floyd, M. A., Szeliga, W. M., King, R. W., Phillips, D. A., Puskas, C. M., Santillan, M., & Wang, L., 2016. Plate Boundary Observatory and related networks: GPS data analysis methods and geodetic product, *Reviews of Geophysics*, pp. 1–50.

Hines, T. T. & Hetland, E. A., 2016. Rheologic constraints on the upper mantle from five years of postseismic deformation following the El Mayor-Cucapah earthquake, *Journal of Geophysical Research: Solid Earth*, **121**.

Hines, T. T. & Hetland, E. A., 2017. Unbiased characterization of noise in geodetic data, *submitted to Journal of Geodesy*.

Hodgkinson, K., Agnew, D., & Roeloffs, E., 2013. Working With Strainmeter Data, *Eos, Transactions American Geophysical Union*, **94**(9), 91–91.

Holt, W. E. & Shcherbenko, G., 2013. Toward a Continuous Monitoring of the Horizontal Displacement Gradient Tensor Field in Southern California Using cGPS Observations from Plate Boundary Observatory (PBO), *Seismological Research Letters*, **84**(3), 455–467.

Johansson, J. M., Davis, J. L., Scherneck, H., Milne, G. A., Vermeer, M., Mitrovica, J. X., Bennett, R. A., Jonsson, B., Elgered, G., Elo, P., & Koivula, H., 2002. Continuous GPS measurements of postglacial adjustment in Fennoscandia 1. Geodetic results, *Journal of Geophysical Research*, **107**.

Kato, T., El-Fiky, G. S., Oware, E. N., & Miyazaki, S., 1998. Crustal strains in the Japanese islands as deduced from dense GPS array, *Geophysical Research Letters*, **25**(18), 3445–3448.

Langbein, J., 2004. Noise in two-color electronic distance meter measurements revisited, *Journal of Geophysical Research: Solid Earth*, **109**(4), 1–16.

Langbein, J., 2008. Noise in GPS displacement measurements from Southern California and Southern Nevada, *Journal of Geophysical Research: Solid Earth*, **113**(5), 1–12.

Langbein, J. & Johnson, H., 1997. Correlated errors in geodetic time series: Implications for time-dependent deformation, *Journal of Geophysical Research*, **102**(B1), 591–603.

Lisowski, M., Dzurisin, D., Denlinger, R. P., & Iwatsubo, E. Y., 2008. Analysis of GPS-Measured Deformation Associated with the 2004 2006 Dome-Building Eruption of Mout St. Helens, Washington, Tech. Rep. September 1984.

Lohman, R. B. & Murray, J. R., 2013. The SCEC Geodetic Transient-Detection Validation Exercise, *Seismological Research Letters*, **84**(3), 419–425.

Mackay, D. J. C., 1998. Introduction to Gaussian processes, *Neural Networks and Machine Learning*, **168**(1996), 133–165.

Mao, A., Harrison, G. A., & Dixon, H., 1999. Noise in GPS coordinate time series, *Journal of Geophysical*

Research, **104**(B2), 2797–2816.

McCaffrey, R., Qamar, A. I., King, R. W., Wells, R., Khazaradze, G., Williams, C. A., Stevens, C. W., Vollick, J. J., & Zwick, P. C., 2007. Fault locking, block rotation and crustal deformation in the Pacific Northwest, *Geophysical Journal International*, **169**(3), 1315–1340.

McCaffrey, R., King, R. W., Payne, S. J., & Lancaster, M., 2013. Active tectonics of northwestern U.S. inferred from GPS-derived surface velocities, *Journal of Geophysical Research: Solid Earth*, **118**, 709–723.

McGuire, J. J. & Segall, P., 2003. Imaging of aseismic fault slip transients recorded by dense geodetic networks, *Geophysical Journal International*, **155**, 778–788.

Meade, B. J. & Hager, B. H., 2005. Block models of crustal motion in southern California constrained by GPS measurements, *Journal of Geophysical Research: Solid Earth*, **110**, 1–19.

Miyazaki, S., McGuire, J. J., & Segall, P., 2003. A transient subduction zone slip episode in southwest Japan observed by the nationwide GPS array, *Journal of Geophysical Research*, **108**(B2), 1–15.

Moritz, H., 1978. Least-Squares Collocation, *Reviews of Geophysics*, **16**(3), 421–430.

Murray, M. H. & Lisowski, M., 2000. Strain accumulation along the Cascadia subduction zone in western Washington, *Geophysical Research Letters*, **27**(22), 3631–3634.

Ohtani, R., McGuire, J. J., & Segall, P., 2010. Network strain filter: A new tool for monitoring and detecting transient deformation signals in GPS arrays, *Journal of Geophysical Research: Solid Earth*, **115**(12), 1–17.

Okada, Y., 1992. Internal deformation due to shear and tensile faults in a half space, *Bulletin of the Seismological Society of America*, **82**(2), 1018–1040.

Papoulis, A., 1991. *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 3rd edn.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T., 2007. *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 3rd edn.

Rasmussen, C. E. & Williams, C. K. I., 2006. *Gaussian processes for machine learning*, The MIT Press.

Roeloffs, E., 2010. Tidal calibration of Plate Boundary Observatory borehole strainmeters: Roles of vertical and shear coupling, *Journal of Geophysical Research: Solid Earth*, **115**(6), 1–25.

Roeloffs, E. A., 2006. Evidence for Aseismic Deformation Rate Changes Prior To Earthquakes, *Annual Review of Earth and Planetary Sciences*, **34**(1), 591–627.

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

Sandwell, D. T. & Wessel, P., 2016. Interpolation of 2-D vector data using constraints from elasticity, *Geophysical Research Letters*, pp. 1–7.

Schmidt, D. A. & Gao, H., 2010. Source parameters and time-dependent slip distributions of slow slip events on the Cascadia subduction zone from 1998 to 2008, *Journal of Geophysical Research: Solid Earth*, **115**(4), 1–13.

Segall, P. & Mathews, M., 1997. Time dependent inversion of geodetic data, *Journal of Geophysical Research*, **102**(B10), 22391–22409.

Shen, Z., Wang, M., Zeng, Y., & Wang, F., 2015. Optimal Interpolation of Spatially Discretized Geodetic Data, *Bulletin of the Seismological Society of America*, **105**(4), 2117–2127.

Shen, Z. K., Jackson, D. D., Ge, B. X., & Bob, X. G., 1996. Crustal deformation across and beyond the Los Angeles basin from geodetic measurements, *Journal of Geophysical Research*, **101**(B12), 27927–27957.

Tape, C., Musé, P., Simons, M., Dong, D., & Webb, F., 2009. Multiscale estimation of GPS velocity fields, *Geophysical Journal International*, **179**(2), 945–971.

von Mises, R., 1964. *Mathematical Theory of Probability and Statistics*, Academic Press, New York.

Wdowinski, S., Zhang, J., Fang, P., & Genrich, J., 1997. Southern California Permanent GPS Geodetic Array: Spatial filtering of daily positions for estimating coseismic and postseismic displacements induced by the 1992 Landers earthquake, **102**(97), 57–70.

Wech, A. G., 2010. Interactive Tremor Monitoring, *Seismological Research Letters*, **81**(4), 664 – 669.

Wech, A. G., Creager, K. C., & Melbourne, T. I., 2009. Seismic and geodetic constraints on Cascadia slow slip, *Journal of Geophysical Research: Solid Earth*, **114**(10), 1–9.

Wendland, H., 2005. *Scattered data approximation*.

Williams, S. D. P., Bock, Y., Fang, P., Jamason, P., Nikolaidis, R. M., Prawirodirdjo, L., Miller, M., & Johnson, D. J., 2004. Error analysis of continuous GPS position time series, *Journal of Geophysical Research: Solid Earth*, **109**(B3).

Wyatt, F., 1982. Displacement of Surface Monuments: Horizontal Motion, *Journal of Geophysical Research*, **87**(B2), 979–989.

Wyatt, F. K., 1989. Displacement of surface monuments: Vertical motion, *Journal of Geophysical Research*, **94**(B2), 1655–1664.

Zhang, J., Bock, Y., Johnson, H., Fang, P., Williams, S., Genrich, J., Wdowinski, S., & Behr, J., 1997. Southern California Permanent GPS Geodetic Array: Error analysis of daily position estimates and site velocities, *Journal of Geophysical Research*, **102**(B8), 18035–18055.