

**Transient ground deformation in tectonically active regions and
implications for the mechanical behavior of the crust and upper
mantle**

by

Trever T. Hines

A dissertation submitted in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
(Earth and Environmental Sciences)
in the University of Michigan
2017

Doctoral Committee:

Associate Professor Eric Hetland
Professor Jeroen Ritsema
Associate Professor Jeremy Bassis
Associate Professor Nathan Niemi
Professor John Boyd

Trever T. Hines
hinest@umich.edu
ORCID iD: 9999-9999-9999-9999

ACKNOWLEDGEMENTS

fuck your face

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	vii
LIST OF TABLES	viii
ABSTRACT	ix
CHAPTER 1: Bias in estimates of lithosphere viscosity from interseismic deformation	1
CHAPTER 2: Rapid and simultaneous estimation of fault slip and heterogeneous lithospheric viscosity from postseismic deformation	2
CHAPTER 3: Rheologic constraints on the upper mantle from 5 years of postseismic deformation following the El Mayor-Cuapah earthquake	3
CHAPTER 4: Unbiased characterization of noise in geodetic data	4
4.1 abstract	5
4.2 Introduction	5
4.3 Maximum likelihood methods	6
4.4 Synthetic demonstration	8
4.5 Discussion and conclusion	10
4.6 acknowledgements	10
4.7 References	11
CHAPTER 5: Revealing transient strain in geodetic data with Gaussian process regression	13
5.1 Summary	14
5.2 Introduction	14
5.3 Estimating Transient Strain Rates	16

5.4	Outlier detection	19
5.5	Application to Cascadia Slow Slip Events	19
5.5.1	Noise model	20
5.5.2	Prior model	22
5.5.3	Transient Strain Rates	28
5.6	Discussion	29
5.7	Conclusion	34
5.8	Acknowledgements	34
5.9	References	34

LIST OF FIGURES

4.1	Random walk amplitudes, σ_{rw} , estimated by the MLE and REML methods from synthetic data. The length of the synthetic time series used to estimate σ_{rw} is varied from 0.1 yr to 2.5 yr. The black line indicates the true random walk amplitude ($\sigma_{rw} = 1.3$), the light blue region shows the 10-90 percentile of estimates, the dark blue region shows the 25-75 percentile of estimates, the solid blue line indicates the median, and the dashed blue line indicates the mean.	9
5.1	Positions of continuous GNSS stations used to estimate transient strain rates. The colored regions indicate the distribution of seismic tremor as determined by <i>Wech</i> (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 5.7 and the signal-to-noise ratio shown in Figure 5.8. The blue dashed circle demarcates the spatial extent of the tremors shown in Figure 5.8.	21
5.2	Distribution of estimated FOGM hyperparameters (eq. 5.20). Hyperparameters are estimated with the REML method for 38 stations in Cascadia that are east of 121°W. “IQR” is the interquartile range.	23
5.3	Two FOGM noise samples for each component. The FOGM hyperparameters have been set to the median values from Figure 5.2.	24
5.4	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 5.4. The error bars show one standard deviation uncertainties. Note that outliers tend to be observed in the winter, suggesting that they were caused by snow or ice. .	27
5.5	Observed easting component of displacements at station P436 and predicted displacements when using different covariance functions for T . The one standard deviation uncertainties are shown for the observations and the predicted displacements when using the SE covariance function. For clarity, uncertainties are not shown for the IBM and Wendland covariance functions, but they are nearly equivalent to the uncertainties for the SE covariance function. . .	28
5.6	Estimated transient strain rates during the Winter 2015-2016 SSE. Strain glyphs show the normal strain rate along each azimuth, where orange indicates compression and blue indicates extension. The shaded regions indicate one standard deviation uncertainties in the normal strain rates.	30

5.7	Three components of the transient horizontal strain rate tensor estimated at the position shown in Figure 5.1. The shaded regions indicate one standard deviation uncertainty.	31
5.8	(top) Signal-to-noise ratio (eq. 5.15) at the position shown in Figure 5.1. (bottom) Frequency of tremors in the region shown in Figure 5.1.	32

LIST OF TABLES

5.1	Optimal hyperparameters for the prior on transient displacements determined with the REML method. The temporal covariance function is indicated by the “ T ” column. The SE, IBM, and Wendland covariance functions are defined in eqs. (5.24), (5.25), and (5.27), respectively. The spatial covariance function, X , is the squared exponential (eq. 5.23) in all cases. The hyperparameters are estimated for each of the seven SSEs considered in this study, and the tabulated values indicate the median and interquartile ranges of estimates. The “diff log(REML)” column compares the log REML likelihood to the log REML likelihood when using the SE covariance function for T . Negative values indicate that observations are more consistent with the SE covariance function.	26
-----	---	----

ABSTRACT

CHAPTER 1

Bias in estimates of lithosphere viscosity from interseismic deformation

CHAPTER 2

Rapid and simultaneous estimation of fault slip and
heterogeneous lithospheric viscosity from postseismic deformation

CHAPTER 3

Rheologic constrains on the upper mantle from 5 years of
postseismic deformation following the El Mayor-Cucapah
earthquake

CHAPTER 4

Unbiased characterization of noise in geodetic data

4.1 abstract

Geodetic time series contain temporally correlated noise that must be quantified before the data can be used to make geophysical inferences. If the noise is not accurately quantified then there is a risk of underestimating the uncertainties on inferred geophysical parameters. The maximum likelihood estimation (MLE) method is commonly used to characterize noise in geodetic time series; however, this method is known to be biased. Specifically, the MLE method has a tendency to underestimate the amplitude of random walk noise. This bias is most pronounced when estimating the noise in shorter time series. We discuss an unbiased alternative to the MLE method, which is known as the restricted maximum likelihood (REML) method. We use synthetic tests to demonstrate that the REML method does not suffer from the bias inherent in the MLE method. Considering that the computational costs of the REML and MLE methods are nearly equivalent, there is no reason to prefer the MLE method over the REML method for quantifying noise in geodetic data.

4.2 Introduction

Before geodetic data can be used to make geophysical inferences, it is necessary to have an accurate noise model. Here we consider noise to be any observed deformation that is not representative of tectonic processes. The noise in geodetic time series is temporally correlated and its power spectrum is often described by the power law relationship (*Agnew, 1992*)

$$P(f) = P_o f^{-k}, \quad (4.1)$$

where f is frequency, k is the spectral index and P_o is the noise amplitude. For data recorded by strain and tilt meters (*Wyatt, 1982, 1989*) and electronic distance measuring (EDM) instruments (*Langbein and Johnson, 1997*), the temporally correlated noise can be modeled as a random walk ($k = 2$). In those studies, the random walk noise was attributed to unstable geodetic monuments. Global Navigation Satellite System (GNSS) data, which is prone to additional non-physical sources of error, often has temporally correlated noise that is best modeled as flicker noise ($k = 1$) (e.g., *Zhang et al., 1997; Mao et al., 1999; Williams et al., 2004*). However, the most appropriate model can vary between stations. Generally, the noise in GNSS data is best described as white noise plus some combination of random walk and flicker noise (*Langbein, 2008*).

If temporally correlated noise is mismodeled or ignored then the uncertainties in geophysical parameters inferred from geodetic data may be underestimated (e.g., *Zhang et al., 1997; Langbein, 2012*). Since no single noise model is universally appropriate, it may be preferable to determine a noise model for each station before attempting to study any underlying signal. *Langbein and Johnson (1997)* introduced a maximum likelihood estimation (MLE) method to determine the hyperparameters (e.g., P_o and k) that best characterize the noise in geodetic time series. Furthermore, the MLE method can be used to discern which type of stochastic process (e.g., power law or Gauss-Markov) is most appropriate (*Langbein, 2004*). There are other methods for determining noise models, such as the least squares variance component estimation method (*Amiri-Simkooei et al., 2007*) and the network noise estimator (*Dmitrieva et al., 2015*). However, the MLE method from *Langbein and Johnson*

(1997) is the most widely used (e.g., *Langbein*, 2004, 2008; *Zhang et al.*, 1997; *Mao et al.*, 1999; *Williams et al.*, 2004; *Hill et al.*, 2009; *King and Williams*, 2009; *Murray and Svarc*, 2017).

One deficiency with the MLE method, which was recognized by *Langbein and Johnson* (1997), is that it can be biased towards underestimating the amplitude of random walk noise. The MLE method is biased because it assumes that residual geodetic time series, with geophysical signals estimated and removed, are representative samples of noise. This is not always a fair assumption because estimating and removing geophysical signals will inevitably also remove low frequency components of noise. *Langbein* (2012) further explored the bias in the MLE method and how it propagates into the uncertainties for estimated tectonic rates of deformation. They demonstrated with synthetic data, consisting of white noise and random walk noise, that the bias is stronger for shorter time series. *Langbein* (2012) emphasized the role of the crossover frequency, f_c , which is the frequency where the power of the white noise is equal to the power of the random walk noise. They suggested that a time series should be several times longer than f_c^{-1} in order to accurately quantify its random walk noise.

In this paper we discuss an alternative to the MLE method, which is known as the restricted maximum likelihood (REML) method (e.g., *Cressie*, 1993). We use synthetic tests to demonstrate that the REML method produces unbiased estimates of random walk noise. With the REML method, we can accurately quantify the random walk noise in time series that are as short as f_c^{-1} . Furthermore, the REML and MLE methods have practically equivalent computational costs. For these reasons, we argue that there is no reason to prefer using the MLE method over the unbiased REML method.

4.3 Maximum likelihood methods

In this section we briefly describe the MLE method and explain why it is biased. We then provide a description of the REML method. Let \mathbf{d}_* denote a column vector of n observations. We treat \mathbf{d}_* as a realization of the random vector

$$\mathbf{d} = \mathbf{G}\mathbf{m} + \epsilon, \quad (4.2)$$

where ϵ is the data noise vector, \mathbf{G} is an $n \times m$ matrix with linearly independent columns that are used to describe geophysical signal in \mathbf{d} (e.g., secular rates, coseismic offsets, postseismic transience), and \mathbf{m} is a column vector of m model parameters which have uninformative priors (i.e. $\mathbf{m} \sim \mathcal{N}(\mathbf{0}, \lambda\mathbf{I})$ in the limit as $\lambda \rightarrow \infty$). We assume that the data noise can be described as $\epsilon \sim \mathcal{N}(\mathbf{0}, \Sigma(\theta))$, where θ are the hyperparameters which we want to estimate appropriate values for. If we had chosen an informed prior for \mathbf{m} , we would select θ such that the probability of drawing \mathbf{d}_* from \mathbf{d} , $p_{\mathbf{d}}(\mathbf{d}_*|\theta)$, is maximized. However, the uninformed prior on \mathbf{m} makes \mathbf{d} improper and $p_{\mathbf{d}}$ is infinitesimally small for all choices of θ . Consequently, it is not possible to numerically maximize $p_{\mathbf{d}}$ and we must seek an alternative likelihood function to maximize.

The MLE method chooses θ such that the probability of sampling the least squares residual vector,

$$\mathbf{r}_* = \left(\mathbf{I} - \mathbf{G} (\mathbf{G}^T \Sigma(\theta)^{-1} \mathbf{G})^{-1} \mathbf{G}^T \Sigma(\theta)^{-1} \right) \mathbf{d}_*, \quad (4.3)$$

from ϵ is maximized. To put it explicitly, The MLE method maximizes the probability density function

$$p_\epsilon(\mathbf{r}_*|\theta) = \left(\frac{1}{(2\pi)^n |\Sigma(\theta)|} \right)^{\frac{1}{2}} e^{-\frac{1}{2} \mathbf{d}_*^T \mathbf{K}(\theta) \mathbf{d}_*} \quad (4.4)$$

with respect to θ , where

$$\mathbf{K}(\theta) = \Sigma(\theta)^{-1} - \Sigma(\theta)^{-1} \mathbf{G} (\mathbf{G}^T \Sigma(\theta)^{-1} \mathbf{G})^{-1} \mathbf{G}^T \Sigma(\theta)^{-1}. \quad (4.5)$$

Implementations of the MLE method typically maximize the logarithm of eq. (4.4) with the downhill simplex method (*Press et al.*, 2007). It is important to recognize that the MLE method assumes that \mathbf{r}_* is a representative sample of ϵ . This assumption is only valid when n is sufficiently large. To elaborate, we note that \mathbf{r}_* is a sample of the random variable

$$\mathbf{r} = \left(\mathbf{I} - \mathbf{G} (\mathbf{G}^T \Sigma(\theta)^{-1} \mathbf{G})^{-1} \mathbf{G}^T \Sigma(\theta)^{-1} \right) \mathbf{d}, \quad (4.6)$$

which is distributed as

$$\mathbf{r} \sim \mathcal{N} \left(\mathbf{0}, \Sigma(\theta) - \mathbf{G} (\mathbf{G}^T \Sigma(\theta)^{-1} \mathbf{G})^{-1} \mathbf{G}^T \right). \quad (4.7)$$

The term being subtracted in eq. (4.7) is the covariance of the least squares prediction vector, which will typically get smaller as n increases. The distribution of \mathbf{r} will then tend towards that of ϵ as n increases. Hence, we can only assume that \mathbf{r}_* is a representative sample of ϵ when n is sufficiently large. We can also observe from eq. (4.7) that the variance of \mathbf{r} will always be less than the variance of ϵ . This is the reason why the MLE method is biased towards underestimating the noise in short time series.

Having demonstrated that the MLE method is biased, we move on to discuss the REML method for selecting θ . The REML method was introduced by *Patterson and Thompson* (1971), and is now established in the Kriging literature as an unbiased method for estimating covariance functions (e.g., *Cressie*, 1993). The REML method can be understood by first considering an $(n - m) \times n$ matrix \mathbf{R} which satisfies $\mathbf{R}\mathbf{G} = \mathbf{0}$. We then consider the random variable $\mathbf{x} = \mathbf{R}\mathbf{d}$, which is distributed as $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}\Sigma(\theta)\mathbf{R}^T)$. As opposed to \mathbf{d} , \mathbf{x} is a proper random variable since it is independent of the prior on \mathbf{m} . The REML method chooses θ such that the probability of drawing $\mathbf{x}_* = \mathbf{R}\mathbf{d}_*$ from \mathbf{x} , $p_{\mathbf{x}}(\mathbf{x}_*|\theta)$, is maximized. As shown by *Harville* (1974), the θ which maximizes $p_{\mathbf{x}}(\mathbf{x}_*|\theta)$ also maximizes $p_{\mathbf{d}}(\mathbf{d}_*|\theta)$ because the two likelihood functions are proportional. The REML method thus circumvents the numerical difficulty that $p_{\mathbf{d}}$ is infinitesimally small and identifies the θ which we initially sought to find. The particular choice for \mathbf{R} does not matter because it will only change the likelihood function which we are maximizing by a scale factor. Following *Harville* (1974), we then let \mathbf{R} have the properties $\mathbf{R}^T \mathbf{R} = \mathbf{I} - \mathbf{G}(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T$ and $\mathbf{R}\mathbf{R}^T = \mathbf{I}$. The probability density function for \mathbf{x} can then be written as

$$p_{\mathbf{x}}(\mathbf{x}_*|\theta) = \left(\frac{|\mathbf{G}^T \mathbf{G}|}{(2\pi)^{n-m} |\Sigma(\theta)| |\mathbf{G}^T \Sigma(\theta)^{-1} \mathbf{G}|} \right)^{\frac{1}{2}} e^{-\frac{1}{2} \mathbf{d}_*^T \mathbf{K}(\theta) \mathbf{d}_*}. \quad (4.8)$$

Note the similarity between eq. (4.8) and eq. (4.4). If programmed efficiently (see Appendix 1) and if $m \ll n$, the computational cost of the REML method is practically equivalent to

that of the MLE method. What remains to be determined is whether the REML method remedies the bias in the MLE method. We demonstrate that this is indeed the case with a numerical test.

4.4 Synthetic demonstration

We compare the REML and MLE methods by using them to estimate hyperparameters from synthetic data. This demonstration is modeled after the demonstration from *Langbein* (2012) which highlights bias in the MLE method. Our synthetic noise is a combination of white and random walk noise, which has a power spectrum described by

$$P(f) = \frac{\sigma_{rw}^2}{2\pi^2 f^2} + 2\sigma_w^2 \Delta t, \quad (4.9)$$

where σ_{rw} and σ_w are hyperparameters for the random walk and white noise components, respectively. Δt is the sampling period, which is set at one day. The crossover frequency for the synthetic noise is then

$$f_c = \frac{1}{2\pi\sqrt{\Delta t}} \frac{\sigma_{rw}}{\sigma_w}. \quad (4.10)$$

In order to use the MLE or REML method, we must express the power law relationship in the frequency domain as a covariance matrix in the time domain. A general procedure for doing so can be found in *Langbein* (2004). The components of the covariance matrix corresponding to eq. (4.9) can be concisely written as

$$\Sigma_{ij} = \sigma_{rw}^2 \min(i\Delta t, j\Delta t) + \sigma_w^2 \delta_{ij}, \quad (4.11)$$

where δ_{ij} is the Kronecker delta function. Similar to *Langbein* (2012), we set $\sigma_{rw} = 1.3$ mm/yr^{0.5} and $\sigma_w = 1.1$ mm. We generate 5,000 synthetic noise time series, which each have a length of 2.5 yr.

We consider σ_w to be known, and we want to estimate σ_{rw} from the synthetic data. Although our synthetic data just consists of noise, we assume that the unknown underlying geophysical signal (i.e. **Gm**) consists of an offset plus a linear trend. We estimate σ_{rw} with the MLE and REML methods using varying lengths of the synthetic time series. The time series lengths range from 0.1 yr to 2.5 yr at 0.1 yr increments. The distribution of estimated σ_{rw} is shown in Figure 4.1.

The distribution of σ_{rw} estimated by the MLE method indicates that there is a bias towards underestimating σ_{rw} when the length of the time series is comparable to f_c^{-1} , which is 0.3 yr in this demonstration. The bias is appreciable when the time series is shorter than ~ 1 yr, and estimates of σ_{rw} cluster around 0.0 when the time series is shorter than f_c^{-1} . The distribution tightens up around the true value and remains relatively constant for time series with length greater than ~ 1 yr. This is consistent with *Langbein and Johnson* (1997) who said that the time series should be at least 5 times greater than f_c^{-1} to get a good estimate of the random walk component. Even when the full length of the time series is used, the mean and median of the distribution tend to be slightly less than the true value.

In contrast, the REML method does significantly better at estimating σ_{rw} . For every time series length considered, the true value for σ_{rw} is within the 25-75 percentile of estimated σ_{rw} .

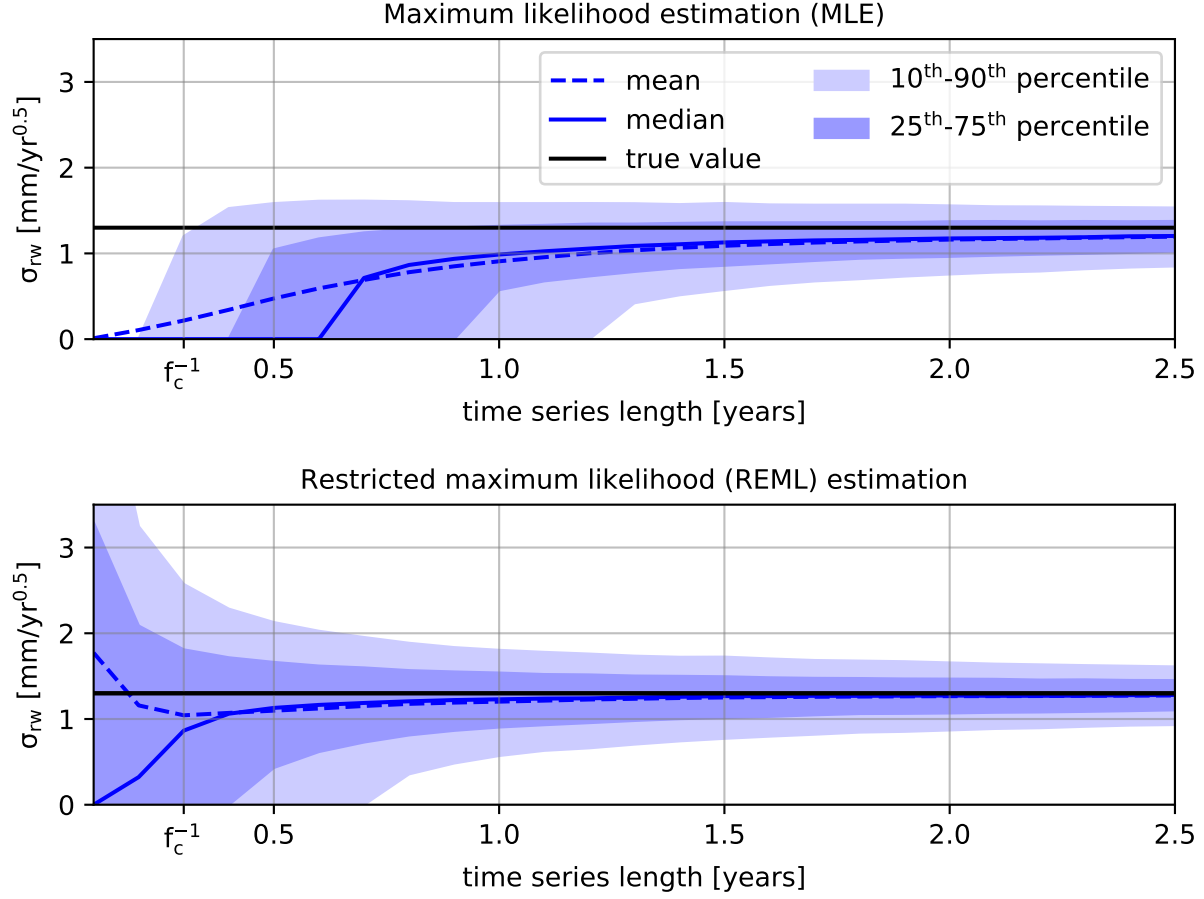


Figure 4.1: Random walk amplitudes, σ_{rw} , estimated by the MLE and REML methods from synthetic data. The length of the synthetic time series used to estimate σ_{rw} is varied from 0.1 yr to 2.5 yr. The black line indicates the true random walk amplitude ($\sigma_{rw} = 1.3$), the light blue region shows the 10-90 percentile of estimates, the dark blue region shows the 25-75 percentile of estimates, the solid blue line indicates the median, and the dashed blue line indicates the mean.

For time series longer than f_c^{-1} , the mean and median of estimated σ_{rw} closely resembles the true value, indicating that the REML method is indeed unbiased. When the length of the time series is less than f_c^{-1} , the mean and median deviate from the true value and the variance of estimated σ_{rw} sharply increases. For such short time series, the random walk component cannot be resolved because it is being masked by the white noise.

4.5 Discussion and conclusion

The MLE method is the most commonly used method for quantifying noise in geodetic time series, despite the fact that it is known to be biased. The bias in the MLE method can result in underestimated uncertainties in geophysical parameters derived from geodetic time series (Langbein, 2012). The intention of this paper is to bring the REML method to light in the geodetic community as an unbiased alternative. Since the MLE method is well established, it may not be reasonable to suggest that researchers abandon it in favor of the REML method. However, the REML method is nearly identical to the MLE method in terms of its computational cost and in terms of its implementation. Indeed, the only difference between the log likelihood functions being maximized by the REML and MLE methods is that the REML method includes two additional, easily computed, terms (See Appendix 1). We can therefore view the REML method as merely an unbiased correction to the MLE method.

In this paper, we have used synthetic tests to demonstrate that the REML method does not suffer from the bias inherent in the MLE method from Langbein and Johnson (1997). We show that the REML method is able to characterize random walk noise in geodetic time series that are as short as f_c^{-1} . In contrast, the MLE method can only accurately quantify random walk noise for time series that are several times longer than f_c^{-1} . We believe that the REML method should always be preferred over the MLE method for quantifying noise in geodetic time series.

4.6 acknowledgements

This material is based upon work supported by the National Science Foundation under grant EAR 1245263.

Appendix 1: REML algorithm

Algorithm 1 demonstrates how to efficiently compute the log of the REML likelihood function. For comparison, we include Algorithm 2, which evaluates the log of the MLE likelihood function. If we assume that $m \ll n$, then the main computational burden in both algorithms is computing the Cholesky decomposition of Σ . Since we are just interested in finding the Σ that maximizes these functions, we can omit the terms in the log likelihood functions that are independent of Σ . In that case, the only difference between the two algorithms is that Algorithm 1 includes a summation along the diagonals of \mathbf{C} in the log likelihood function.

Algorithm 1 Function that takes \mathbf{d} , Σ , and \mathbf{G} as input and returns the logarithm of eq. (4.8). We use the notation $\mathbf{X} \backslash \mathbf{Z}$ to denote solving the system of equations $\mathbf{X}\mathbf{Y} = \mathbf{Z}$ for \mathbf{Y} .

```

function reml_log_likelihood( $\mathbf{d}, \Sigma, \mathbf{G}$ )
   $\mathbf{A} \leftarrow \text{cholesky}(\Sigma)$ 
   $\mathbf{B} \leftarrow \mathbf{A} \backslash \mathbf{G}$ 
   $\mathbf{C} \leftarrow \text{cholesky}(\mathbf{B}^T \mathbf{B})$ 
   $\mathbf{D} \leftarrow \text{cholesky}(\mathbf{G}^T \mathbf{G})$ 
   $\mathbf{a} \leftarrow \mathbf{A} \backslash \mathbf{d}$ 
   $\mathbf{b} \leftarrow \mathbf{C} \backslash (\mathbf{B}^T \mathbf{a})$ 
  return  $\sum_i^m \log(D_{ii}) - \sum_i^n \log(A_{ii}) - \sum_i^m \log(C_{ii}) - \frac{1}{2} \mathbf{a}^T \mathbf{a} + \frac{1}{2} \mathbf{b}^T \mathbf{b} - \frac{n-m}{2} \log(2\pi)$ 
end function

```

Algorithm 2 Function that takes \mathbf{d} , Σ , and \mathbf{G} as input and returns the logarithm of eq. (4.4).

```

function mle_log_likelihood( $\mathbf{d}, \Sigma, \mathbf{G}$ )
   $\mathbf{A} \leftarrow \text{cholesky}(\Sigma)$ 
   $\mathbf{B} \leftarrow \mathbf{A} \backslash \mathbf{G}$ 
   $\mathbf{C} \leftarrow \text{cholesky}(\mathbf{B}^T \mathbf{B})$ 
   $\mathbf{a} \leftarrow \mathbf{A} \backslash \mathbf{d}$ 
   $\mathbf{b} \leftarrow \mathbf{C} \backslash (\mathbf{B}^T \mathbf{a})$ 
  return  $-\sum_i^n \log(A_{ii}) - \frac{1}{2} \mathbf{a}^T \mathbf{a} + \frac{1}{2} \mathbf{b}^T \mathbf{b} - \frac{n}{2} \log(2\pi)$ 
end function

```

4.7 References

- Agnew, D. C. (1992), The time-domain behavior of power-law noises, *Geophysical Research Letters*, 19(4), 333–336.
- Amiri-Simkooei, A. R., C. C. J. M. Tiberius, and P. J. G. Teunissen (2007), Assessment of noise in GPS coordinate time series: Methodology and results, *Journal of Geophysical Research: Solid Earth*, 112(7), 1–19, doi:10.1029/2006JB004913.
- Cressie, N. (1993), *Statistics for Spatial Data*, rev. ed., John Wiley & Sons, New York, doi:10.1111/j.1365-3121.1992.tb00605.x.
- Dmitrieva, K., P. Segall, and C. DeMets (2015), Network-based estimation of time-dependent noise in GPS position time series, *Journal of Geodesy*, 89(6), 591–606, doi:10.1007/s00190-015-0801-9.
- Harville, D. A. (1974), Bayesian Inference for Variance Components Using Only Error Contrasts, *Biometrika*, 61(2), 383–385.
- Hill, E. M., J. L. Davis, P. Elósegui, B. P. Wernicke, E. Malikowski, and N. A. Niemi (2009), Characterization of site-specific GPS errors using a short-baseline network of braced monuments at Yucca Mountain, southern Nevada, *Journal of Geophysical Research: Solid Earth*, 114(11), 1–13, doi:10.1029/2008JB006027.

- King, M. A., and S. D. P. Williams (2009), Apparent stability of GPS monumentation from short-baseline time series, *Journal of Geophysical Research: Solid Earth*, *114*(10), 1–21, doi:10.1029/2009JB006319.
- Langbein, J. (2004), Noise in two-color electronic distance meter measurements revisited, *Journal of Geophysical Research: Solid Earth*, *109*(4), 1–16, doi:10.1029/2003JB002819.
- Langbein, J. (2008), Noise in GPS displacement measurements from Southern California and Southern Nevada, *Journal of Geophysical Research: Solid Earth*, *113*(5), 1–12, doi:10.1029/2007JB005247.
- Langbein, J. (2012), Estimating rate uncertainty with maximum likelihood: Differences between power-law and flicker-random-walk models, *Journal of Geodesy*, *86*(9), 775–783, doi:10.1007/s00190-012-0556-5.
- Langbein, J., and H. Johnson (1997), Correlated errors in geodetic time series: Implications for time-dependent deformation, *Journal of Geophysical Research*, *102*(B1), 591–603, doi:10.1029/96JB02945.
- Mao, A., G. A. Harrison, and H. Dixon (1999), Noise in GPS coordinate time series, *Journal of Geophysical Research*, *104*(B2), 2797–2816.
- Murray, J. R., and J. Svarc (2017), Global Positioning System Data Collection, Processing, and Analysis Conducted by the U.S. Geological Survey Earthquake Hazards Program, *Seismological Research Letters*, *88*(3), 1 – 10, doi:10.1785/0220160204.
- Patterson, H. D., and R. Thompson (1971), Recovery of Inter-Block Information when Block Sizes are Unequal, *Biometrika*, *58*(3), 545–554.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (2007), *Numerical Recipes: The Art of Scientific Computing*, 3 ed., Cambridge University Press, Cambridge, doi:10.1137/1031025.
- Williams, S. D. P., Y. Bock, P. Fang, P. Jamason, R. M. Nikolaidis, L. Prawirodirdjo, M. Miller, and D. J. Johnson (2004), Error analysis of continuous GPS position time series, *Journal of Geophysical Research: Solid Earth*, *109*(B3), doi:10.1029/2003JB002741.
- 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., Y. Bock, H. Johnson, P. Fang, S. Williams, J. Genrich, S. Wdowinski, and J. Behr (1997), Southern California Permanent GPS Geodetic Array: Error analysis of daily position estimates and site velocities, *Journal of Geophysical Research*, *102*(B8), 18,035–18,055.

CHAPTER 5

Revealing transient strain in geodetic data with Gaussian process regression

5.1 Summary

Transient strain rates derived from GNSS data can be used to detect and understand geophysical phenomena 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 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 GNSS data. As a demonstration, we use GPR to detect transient strain resulting from slow slip events in Cascadia. 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 slow slip events.

5.2 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 and 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 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) over some subnetwork of the GNSS stations (e.g., *Feigl et al.*, 1990; *Murray and Lisowski*, 2000). The components of the strain rate tensor for each subnetwork are then determined through a 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 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. Other methods for estimating secular strain rates have parameterized GNSS derived velocities with bi-cubic splines (*Beavan and Haines*, 2001), spherical wavelets (*Tape et al.*, 2009), and elastostatic Green’s functions (*Sandwell and Wessel*, 2016). The type of basis functions and the number of degrees of freedom for a parameterization can be subjective. 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 and 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 and 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 and 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 and 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 and 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 and 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 described 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 and Williams*, 2006) to estimate transient strain 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 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 and Kato* (1998) to estimate secular strain rates from GNSS data. 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 process that 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 process in \mathbb{R}^1 which has zero mean and covariance function $\text{cov}(B(t), B(t')) = \min(t, t')$, where $t, t' \geq 0$. If no prior information is available for the geophysical signal, then maximum likelihood methods can be used to objectively choose a prior Gaussian process 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. We use GPR to infer strain resulting from SSEs in Cascadia, demonstrating that GPR is an effective tool for detecting transient geophysical processes.

5.3 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 secular strain rates, and our attention is limited to horizontal strain rates in this study. We denote transient crustal strain rates as $\dot{\epsilon}(p)$, where p represents the ordered pair (\vec{x}, t) , \vec{x} are spatial coordinates in \mathbb{R}^2 , and t is time. We determine $\dot{\epsilon}(p)$ by spatially and temporally differentiating estimates of transient displacements, $\vec{u}(p)$. We make a prior assumption that each component of \vec{u} is a Gaussian process,

$$u_i(p) \sim \mathcal{N}(0, C_{u_i}), \quad (5.1)$$

where $C_{u_i}(p, p')$ is a covariance function indicating how we expect $u_i(p)$ to covary with $u_i(p')$. For simplicity, we treat each component of displacements independently and ignore any potential covariance between them. 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} . We assume that C_u can be separated into spatial and temporal functions as

$$C_u((\vec{x}, t), (\vec{x}', t')) = X(\vec{x}, \vec{x}')T(t, t'). \quad (5.2)$$

As long as the functions X and T are positive definite, C_u is guaranteed to also be positive definite and thus a valid covariance function (*Rasmussen and Williams*, 2006, sec. 4.2.4). The appropriate choice for X and T may vary depending on the geophysical signal we are trying to describe, and we discuss this matter in Section 5.5.2.

We constrain u with GNSS data, which records u 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 as a realization of the random variable

$$\begin{aligned} d_{ij} = & u(\vec{x}_i, t_j) + \eta(\vec{x}_i, t_j) + w_{ij} + a_i^{(1)} + a_i^{(2)}t_j + \\ & a_i^{(3)} \sin(2\pi t_j) + a_i^{(4)} \cos(2\pi t_j) + a_i^{(5)} \sin(4\pi t_j) + a_i^{(6)} \cos(4\pi t_j), \end{aligned} \quad (5.3)$$

where $a_i^{(1)}$ is an offset that is unique for each GNSS monument, $a_i^{(2)}$ is the secular velocity at \vec{x}_i , 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., *Wdowski et al., 1997*). For now, we will only assume that $\eta \sim \mathcal{N}(0, C_\eta)$. We consider the six coefficients in eq. (5.3) to be uncorrelated random variables distributed as $\mathcal{N}(0, \kappa^2)$ in the limit as $\kappa \rightarrow \infty$ (i.e., the coefficients have diffuse priors). Of course, the secular velocities, $a_i^{(2)}$, are spatially correlated and we could invoke a tectonic model to form a prior on $a_i^{(2)}$. However, in our application to Cascadia, we will be using displacement time series which are long enough to sufficiently constrain $a_i^{(2)}$ 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, \mathbf{d}_* . Let \mathbf{P} be the set of (\vec{x}_i, t_j) pairs describing where and when each of the GNSS observations have been made. Let \mathbf{a} be the vector of coefficients from eq. (5.3) for each of the m GNSS stations. We use \mathbf{G} to represent the $n \times 6m$ matrix of corresponding basis functions evaluated at each point in \mathbf{P} . We also denote the vector of uncorrelated noise for each observation as \mathbf{w} , whose standard deviations are given by the formal data uncertainty, σ . The observations can then be viewed as a realization of the random vector

$$\mathbf{d} = u(\mathbf{P}) + \eta(\mathbf{P}) + \mathbf{w} + \mathbf{G}\mathbf{a}, \quad (5.4)$$

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

$$\mathbf{\Sigma} = C_u(\mathbf{P}, \mathbf{P}) + C_\eta(\mathbf{P}, \mathbf{P}) + \text{diag}(\sigma^2). \quad (5.5)$$

It should be understood that notation such as $u(\mathbf{P})$ and $C_u(\mathbf{P}, \mathbf{P})$ represents the column vector $[u(P_i)]_{P_i \in \mathbf{P}}$ and the matrix $[C_u(P_i, P_j)]_{(P_i, P_j) \in \mathbf{P} \times \mathbf{P}}$, respectively.

The prior for transient displacements is then conditioned with \mathbf{d}_* to form a posterior estimate of transient displacements, $\hat{u} = u|\mathbf{d}_*$. For now, we will assume that appropriate covariance functions and corresponding hyperparameters for C_u and C_η have already been chosen. In Section 5.5.1 and 5.5.2, we discuss how the covariance functions are chosen for our application to GNSS data from Cascadia. If κ is kept finite then, from *Rasmussen and Williams* (2006, sec. 2.2), we find that \hat{u} is distributed as $\mathcal{N}(\mu_{\hat{u}}, C_{\hat{u}})$, where

$$\mu_{\hat{u}}(p) = C_u(p, \mathbf{P}) (\mathbf{\Sigma} + \kappa^2 \mathbf{G}\mathbf{G}^T)^{-1} \mathbf{d}_* \quad (5.6)$$

and

$$C_{\hat{u}}(p, p') = C_u(p, p') - C_u(p, \mathbf{P}) (\mathbf{\Sigma} + \kappa^2 \mathbf{G}\mathbf{G}^T)^{-1} C_u(\mathbf{P}, p'). \quad (5.7)$$

However, we are interested in the limit as $\kappa \rightarrow \infty$, and the form for eq. (5.6) and eq. (5.7) is not suitable for evaluating this limit. We use a partitioned matrix inversion identity (*Press et al., 2007, sec. 2.7.4*) to rewrite eq. (5.6) and eq. (5.7) as

$$\mu_{\hat{u}}(p) = \begin{bmatrix} C_u(p, \mathbf{P}) & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma} & \mathbf{G} \\ \mathbf{G}^T & -\kappa^{-2} \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{d}_* \\ \mathbf{0} \end{bmatrix} \quad (5.8)$$

and

$$C_{\hat{u}}(p, p') = C_u(p, p') - \begin{bmatrix} C_u(p, \mathbf{P}) & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma} & \mathbf{G} \\ \mathbf{G}^T & -\kappa^{-2} \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} C_u(\mathbf{P}, p') \\ \mathbf{0} \end{bmatrix}. \quad (5.9)$$

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

$$\mu_{\hat{u}}(p) = \begin{bmatrix} C_u(p, \mathbf{P}) & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Sigma & \mathbf{G} \\ \mathbf{G}^T & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{d}_* \\ \mathbf{0} \end{bmatrix} \quad (5.10)$$

and

$$C_{\hat{u}}(p, p') = C_u(p, p') - \begin{bmatrix} C_u(p, \mathbf{P}) & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Sigma & \mathbf{G} \\ \mathbf{G}^T & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} C_u(\mathbf{P}, p') \\ \mathbf{0} \end{bmatrix}. \quad (5.11)$$

We use eq. (5.10) and (5.11) to find the posterior easting and northing components of transient displacements. Using \hat{u}_i to denote the posterior transient displacements along direction i and x_i to represent the components of \vec{x} , we can write the components of $\dot{\epsilon}$ as

$$\dot{\epsilon}_{ij}(p) = \frac{1}{2} \frac{\partial}{\partial t} \left(\frac{\partial \hat{u}_i(p)}{\partial x_j} + \frac{\partial \hat{u}_j(p)}{\partial x_i} \right). \quad (5.12)$$

The transient strain rate components are stochastic processes whose distributions can be interpreted as the distributions of samples of \hat{u} that have been differentiated by the operator from eq. (5.12). Since differentiation is a linear operation, the transient strain rate components are Gaussian processes. From *Papoulis* (1991, sec. 10.2), we find that the components of $\dot{\epsilon}$ have mean functions

$$\mu_{\dot{\epsilon}_{ij}}(p) = \frac{1}{2} \frac{\partial}{\partial t} \left(\frac{\partial \mu_{\hat{u}_i}(p)}{\partial x_j} + \frac{\partial \mu_{\hat{u}_j}(p)}{\partial x_i} \right) \quad (5.13)$$

and covariance functions

$$C_{\dot{\epsilon}_{ij}}(p, p') = \frac{1}{4} \frac{\partial^2}{\partial t \partial t'} \left(\frac{\partial^2 C_{\hat{u}_i}(p, p')}{\partial x_j \partial x'_j} + \frac{\partial^2 C_{\hat{u}_j}(p, p')}{\partial x_i \partial x'_i} \right). \quad (5.14)$$

Our motivation for estimating transient strain rates is, in part, to detect geophysical phenomena. As we will see, geophysical signal can be easily identified by visually inspecting the solution for transient strain rates. However, if we want to detect geophysical phenomena automatically, then we need to define a detection criterion. We propose using a signal-to-noise ratio, SNR, based on $\dot{\epsilon}$ for our detection criterion. The Frobenius norm of $\dot{\epsilon}$, $\|\dot{\epsilon}\|_F$, which is sometimes referred to as the second invariant of strain rate in the geodetic literature, is often used as a metric for the strain rate “magnitude”. Noting that $\|\dot{\epsilon}\|_F$ is a random variable, SNR can be taken as the ratio of the estimated mean and standard deviation of $\|\dot{\epsilon}\|_F$. Using nonlinear uncertainty propagation, we find SNR to be

$$\text{SNR}(p) = \frac{\mu_{\dot{\epsilon}_{nn}}(p)^2 + \mu_{\dot{\epsilon}_{ee}}(p)^2 + 2\mu_{\dot{\epsilon}_{en}}(p)^2}{\left(C_{\dot{\epsilon}_{nn}}(p, p)\mu_{\dot{\epsilon}_{nn}}(p)^2 + C_{\dot{\epsilon}_{ee}}(p, p)\mu_{\dot{\epsilon}_{ee}}(p)^2 + 4C_{\dot{\epsilon}_{en}}(p, p)\mu_{\dot{\epsilon}_{en}}(p)^2 \right)^{\frac{1}{2}}}, \quad (5.15)$$

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. (5.15), even though they are non-zero.

5.4 Outlier detection

In 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 for normally distributed noise. It follows that proposed methods for analyzing GNSS data should be robust against outliers (e.g., *Blewitt et al.*, 2016). In order to make our estimates of transient strain more robust, we automatically identify and remove outliers in the GNSS data as a pre-processing step.

Our method for detecting outliers is based on the data editing algorithm described in *Gibbs* (2011). We calculate the residuals between the observations and a best fitting model. Data with residuals that are anomalously large are then identified as outliers. We treat \mathbf{d}_* as a sample of \mathbf{d} and assume that there is no correlated noise (i.e., $\eta(p) = 0$). The best fitting model for \mathbf{d}_* is considered to be the expected value of the random vector $u(\mathbf{P}) + \mathbf{G}\mathbf{a}$ after conditioning it with non-outlier observations. We still consider u to have a separable covariance function as in eq. (5.2), and the choice for X and T does not need to be the same as that used in Section 5.3. Since outliers are 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 Ω be the index set of non-outliers in \mathbf{d}_* and initiate it with all n indices. This algorithm is iterative, and for each iteration we calculate the residual vector

$$\mathbf{r} = \frac{\mathbf{d}_* - \mathbb{E} \left[(u(\mathbf{P}) + \mathbf{G}\mathbf{a}) | \tilde{\mathbf{d}}_* \right]}{\sigma} \quad (5.16)$$

$$= \frac{1}{\sigma} \left(\mathbf{d}_* - \begin{bmatrix} C_u(\mathbf{P}, \tilde{\mathbf{P}}) & \mathbf{G} \end{bmatrix} \begin{bmatrix} C_u(\tilde{\mathbf{P}}, \tilde{\mathbf{P}}) + \text{diag}(\tilde{\sigma}^2) & \tilde{\mathbf{G}} \\ \tilde{\mathbf{G}}^T & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\mathbf{d}}_* \\ \mathbf{0} \end{bmatrix} \right), \quad (5.17)$$

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

$$\Omega = \{i : |r_i| < \lambda \cdot \text{RMS}\}, \quad r_i \in \mathbf{r}, \quad (5.18)$$

where RMS is the root-mean-square of $\tilde{\mathbf{r}}$ 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 Ω is equal to the previous Ω .

It should be noted that this 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. 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.

5.5 Application to Cascadia Slow Slip Events

We use our method to estimate transient strain rates in Cascadia, and we are specifically interested in identifying strain resulting from SSEs (e.g., *Dragert et al.*, 2001). In Cascadia,

SSEs can be detected by monitoring for associated seismic tremor (*Rogers and Dragert, 2003*), which is actively being done by the Pacific Northwest Seismic Network (*Wech, 2010*). We can thus compare the tremor records to the transient strain rates estimated with GPR to verify that we are indeed identifying strain from SSEs.

We use the daily displacement solutions generated by the Geodesy Advancing Geosciences and EarthScope (GAGE) facility for continuous GNSS stations (*Herring et al., 2016*). This data is publicly available and can be found at www.unavco.org. We limit the dataset to the stations and time ranges which 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 5.1.

5.5.1 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 temporally correlated noise in GNSS data (e.g., *Zhang et al., 1997; Mao et al., 1999; Williams et al., 2004; Langbein, 2008*). In these studies, temporally correlated noise was described with some combination of Brownian motion, 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

$$\dot{\eta}(t) + \alpha\eta(t) = \beta w(t), \quad (5.19)$$

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. (5.19) is a Gaussian process with zero mean and the covariance function

$$C_\eta((\vec{x}, t), (\vec{x}', t')) = \frac{\beta^2}{2\alpha} \exp(-\alpha|t - t'|) \delta(\|\vec{x} - \vec{x}'\|_2). \quad (5.20)$$

We constrain the hyperparameters for η , α and β , with a set of 38 continuous GNSS stations in Cascadia 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 5.4. 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 and Hetland, 2017*). The REML method finds the hyperparameters, which we collectively refer to as θ , that maximize the

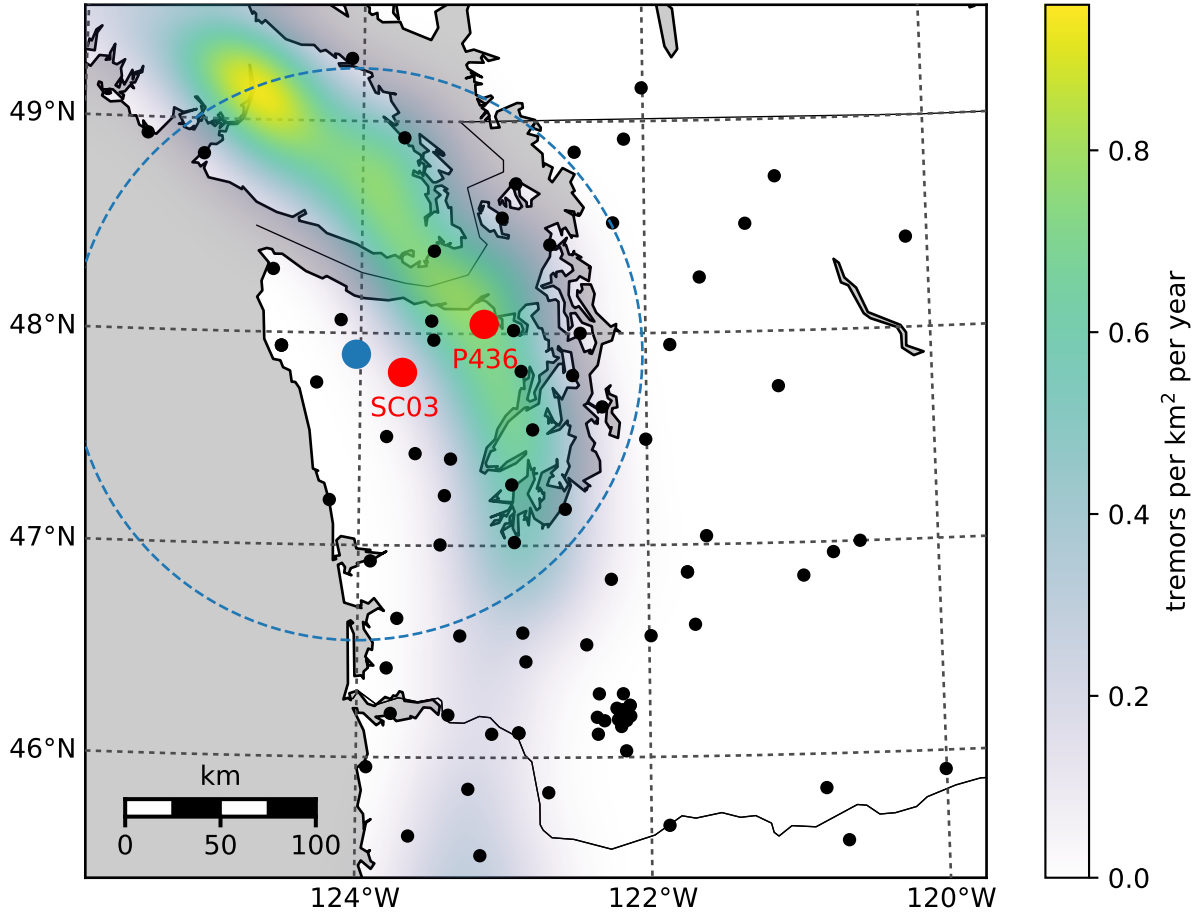


Figure 5.1: Positions of continuous GNSS stations used to estimate transient strain rates. The colored regions indicate the distribution of seismic tremor as determined by *Wech* (2010). The red dots show the positions of GNSS stations mentioned in this paper. The blue dot indicates the location of the transient strain rates shown in Figure 5.7 and the signal-to-noise ratio shown in Figure 5.8. The blue dashed circle demarcates the spatial extent of the tremors shown in Figure 5.8.

likelihood function

$$\mathcal{L}(\theta) = \left(\frac{|\mathbf{G}^T \mathbf{G}|}{(2\pi)^{n-6m} |\boldsymbol{\Sigma}(\theta)| |\mathbf{G}^T \boldsymbol{\Sigma}(\theta)^{-1} \mathbf{G}|} \right)^{\frac{1}{2}} e^{-\frac{1}{2} \mathbf{d}_*^T \mathbf{K}(\theta) \mathbf{d}_*}, \quad (5.21)$$

where

$$\mathbf{K}(\theta) = \boldsymbol{\Sigma}(\theta)^{-1} - \boldsymbol{\Sigma}(\theta)^{-1} \mathbf{G} (\mathbf{G}^T \boldsymbol{\Sigma}(\theta)^{-1} \mathbf{G})^{-1} \mathbf{G}^T \boldsymbol{\Sigma}(\theta)^{-1}. \quad (5.22)$$

Harville (1974) showed that choosing the hyperparameters which maximize eq. (5.21) is equivalent to choosing the hyperparameters which maximize the probability of drawing \mathbf{d}_* from \mathbf{d} . We use the REML method over the maximum likelihood method (e.g., *Langbein and Johnson*, 1997) because the REML method accounts for the improper prior that we assigned to \mathbf{a} (*Hines and Hetland*, 2017). We independently estimate θ for each station, and so \mathbf{d}_* consists of displacements for an individual station. We are assuming $u(p) = 0$ when estimating the noise hyperparameters for this section. The distribution of inferred α and β are shown in Figure 5.2. The amplitude of FOGM noise, β , for the easting and northing components is notable low and are clustered around $0.5 \text{ mm/yr}^{0.5}$. The corresponding estimates of α tend to cluster around 0 yr^{-1} , suggesting that noise can be described with Brownian motion. We also estimate hyperparameters for the vertical component of displacements, under the hope that vertical deformation gradients could reveal some geophysical signal. The amplitude of FOGM noise for the vertical component is relatively large with a median value of $13.5 \text{ mm/yr}^{0.5}$. The inferred values for α are also higher for the vertical component with a median value of 8.21 yr^{-1} . In Figure 5.3, we use the median values of α and β to generate two random samples of FOGM noise for each component. The samples span five years and over these five years the easting and northing samples drift by about 1 mm. 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.*, 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 estimated transient displacements. However, estimated transient strain rates are insensitive to common mode error. We therefore do not 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.

5.5.2 Prior model

We next establish our 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 , we use the squared expo-

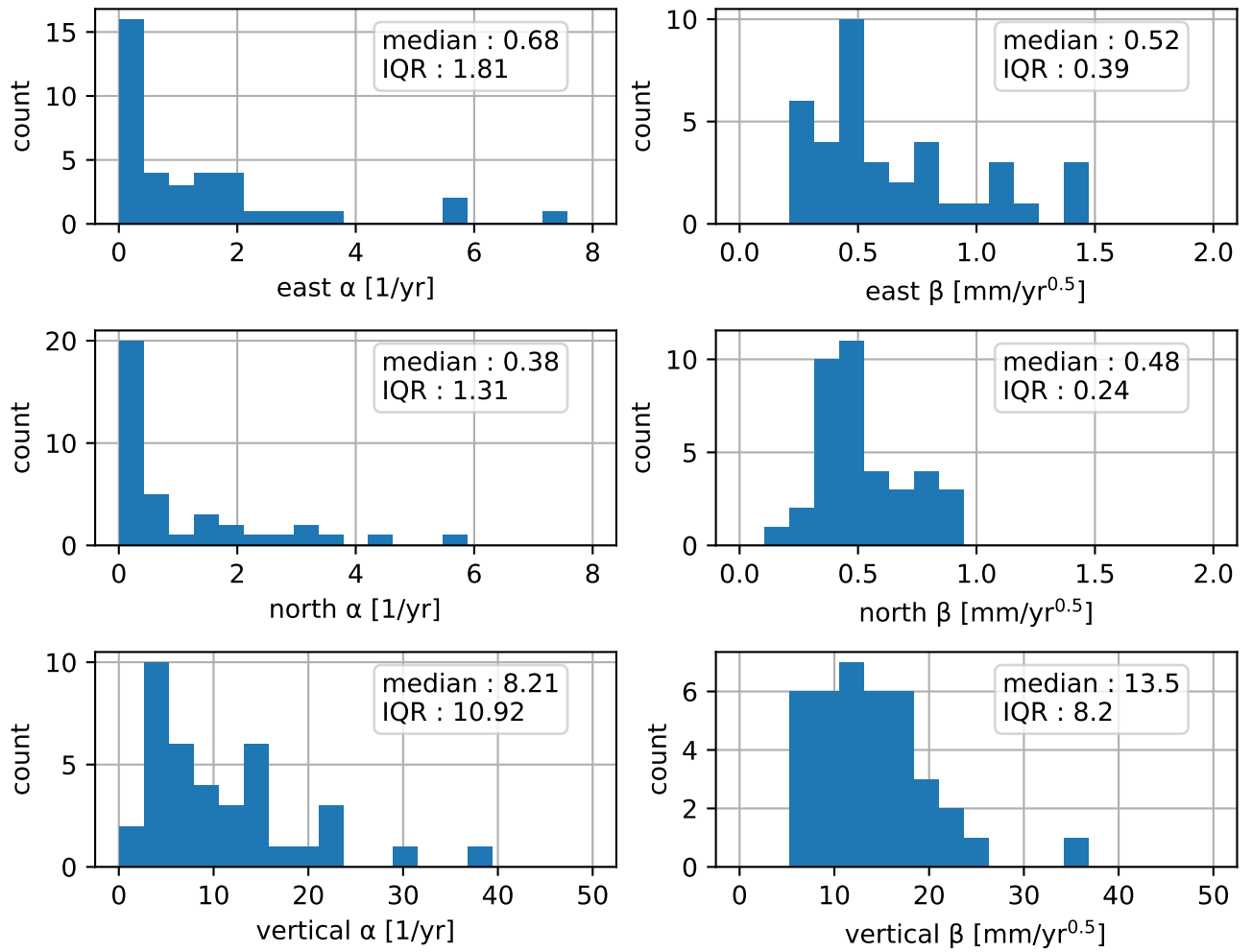


Figure 5.2: Distribution of estimated FOGM hyperparameters (eq. 5.20). Hyperparameters are estimated with the REML method for 38 stations in Cascadia that are east of 121°W . “IQR” is the interquartile range.

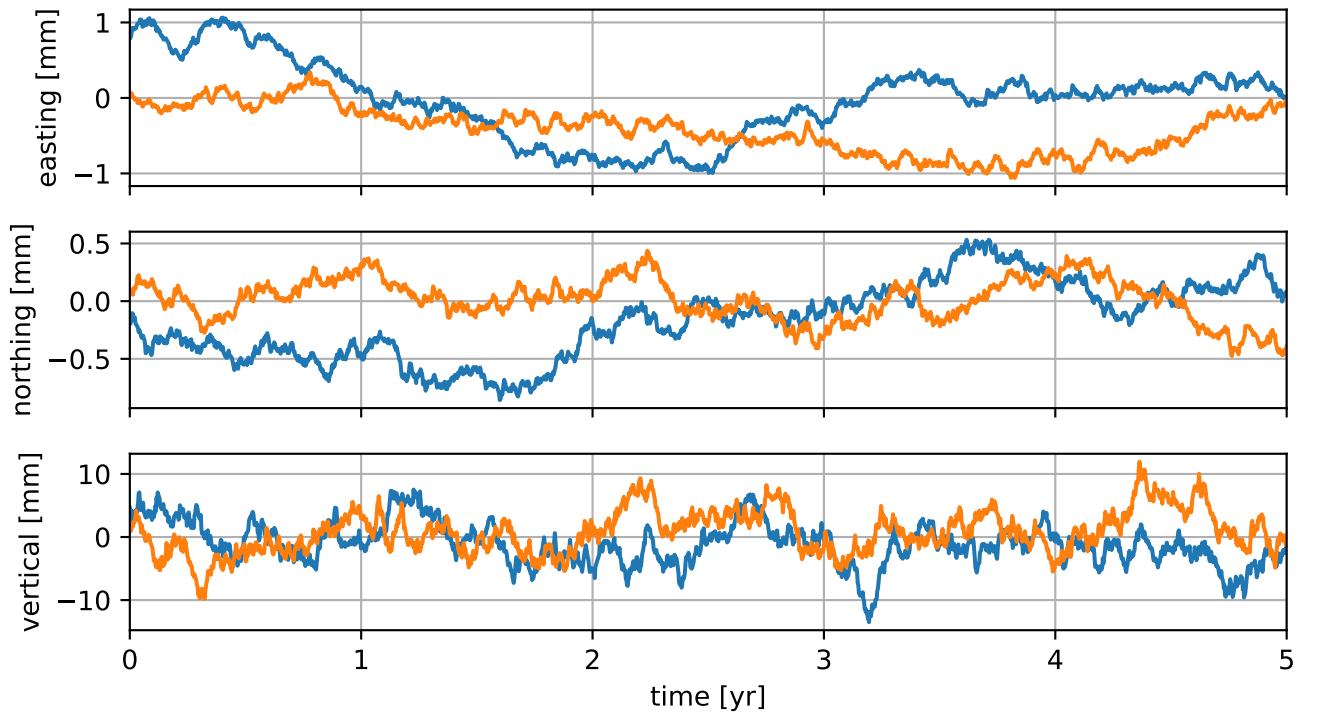


Figure 5.3: Two FOGM noise samples for each component. The FOGM hyperparameters have been set to the median values from Figure 5.2.

nential (SE) covariance function,

$$X(\vec{x}, \vec{x}') = \exp\left(\frac{-\|\vec{x} - \vec{x}'\|_2^2}{2\ell^2}\right). \quad (5.23)$$

The SE covariance function is commonly used in kriging (e.g., *Cressie*, 1993) and Gaussian process regression (e.g., *Rasmussen and 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 and Kato* (1998) demonstrated that the SE accurately describes the covariance of secular GNSS derived velocities in Japan.

We consider three potential models for the temporal covariance of u . First, we consider the one-dimensional SE covariance function,

$$T(t, t') = \phi^2 \exp\left(\frac{-|t - t'|^2}{2\tau^2}\right). \quad (5.24)$$

Note that T includes the hyperparameter ϕ , which serves to scale the covariance function C_u . Second, we consider integrated Brownian motion (IBM). IBM has zero mean and its covariance function can be found by integrating the covariance function for Brownian motion as

$$T(t, t') = \int_0^t \int_0^{t'} \phi^2 \min(s, s') ds' ds \quad (5.25)$$

$$= \frac{\phi^2}{2} \min(t, t')^2 \left(\max(t, t') - \frac{1}{3} \min(t, t') \right), \quad t, t' \geq 0. \quad (5.26)$$

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 and Mathews*, 1997; *McGuire and Segall*, 2003; *Ohtani et al.*, 2010; *Hines and Hetland*, 2016). It should be emphasized $t = 0$ is a reference time at which the Gaussian process is exactly zero. 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 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. In the following analysis, we make the quite arbitrary choice that t is zero on the first epoch of \mathbf{d}_* . 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 are 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

$$T(t, t') = \phi^2 \left(1 - \frac{|t - t'|}{\tau}\right)_+^5 \left(\frac{8|t - t'|^2}{\tau^2} + \frac{5|t - t'|}{\tau} + 1\right), \quad (5.27)$$

Table 5.1: Optimal hyperparameters for the prior on transient displacements determined with the REML method. The temporal covariance function is indicated by the “ T ” column. The SE, IBM, and Wendland covariance functions are defined in eqs. (5.24), (5.25), and (5.27), respectively. The spatial covariance function, X , is the squared exponential (eq. 5.23) in all cases. The hyperparameters are estimated for each of the seven SSEs considered in this study, and the tabulated values indicate the median and interquartile ranges of estimates. The “diff log(REML)” column compares the log REML likelihood to the log REML likelihood when using the SE covariance function for T . Negative values indicate that observations are more consistent with the SE covariance function.

T	direction	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

where

$$(t)_+ = \begin{cases} t, & t > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (5.28)$$

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 5.4. 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 5.4), which can be adversely affected by ice or snow during the winter (*Lisowski et al.*, 2008). After cleaning the dataset, we divide it into seven subsets which 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 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 \mathbf{d} , and they are omitted from \mathbf{d} in this REML analysis for Cascadia SSEs. The estimated hyperparameters for u are summarized in Table 1. Based on the interquartile ranges, the estimated hyperparameters 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 SSEs. For the IBM model, there are several anomalously large values for ℓ and ϕ , which results in large interquartile ranges.

Next we identify which covariance function for T best describes the SSEs. One approach

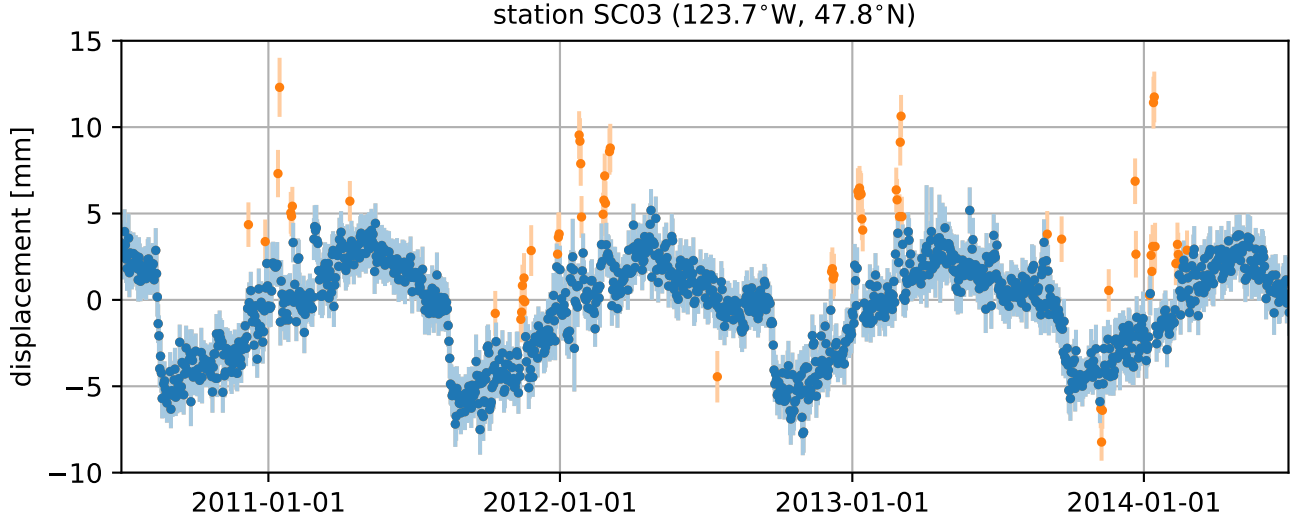


Figure 5.4: Detrended easting component of displacements at station SC03, which is located on Mount Olympus in Washington. The orange markers indicate outliers that were automatically detected using the algorithm from Section 5.4. The error bars show one standard deviation uncertainties. Note that outliers tend to be observed in the winter, suggesting that they were caused by snow or ice.

is to compare the REML likelihoods for each covariance function, similar to the analysis in *Langbein* (2004). In Table 1, we summarize how the log REML likelihoods for the Wendland and IBM covariance functions compare to the SE covariance function. Based on the differences in log 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 displacements for each covariance function. We consider the data prediction vector to be $\hat{\mathbf{d}} = (u(\mathbf{P}) + \mathbf{G}\mathbf{a}) | \mathbf{d}_*$. Following a similar procedure as in Section 5.3, it can be shown that $\hat{\mathbf{d}}$ is normally distributed with mean

$$\mu_{\hat{\mathbf{d}}} = \begin{bmatrix} C_u(\mathbf{P}, \mathbf{P}) & \mathbf{G} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{G} \\ \mathbf{G}^T & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{d}_* \\ \mathbf{0} \end{bmatrix} \quad (5.29)$$

and covariance

$$\mathbf{C}_{\hat{\mathbf{d}}} = C_u(\mathbf{P}, \mathbf{P}) - \begin{bmatrix} C_u(\mathbf{P}, \mathbf{P}) & \mathbf{G} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{G} \\ \mathbf{G}^T & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} C_u(\mathbf{P}, \mathbf{P}) \\ \mathbf{G}^T \end{bmatrix}. \quad (5.30)$$

We compute $\hat{\mathbf{d}}$ using SE, Wendland, and IBM covariance functions for T and the median hyperparameters from Table 1. Figure 5.5 compares the easting component of \mathbf{d}_* to $\hat{\mathbf{d}}$ for the winter 2015-2016 SSE at station P436, which is among the stations that record the strongest

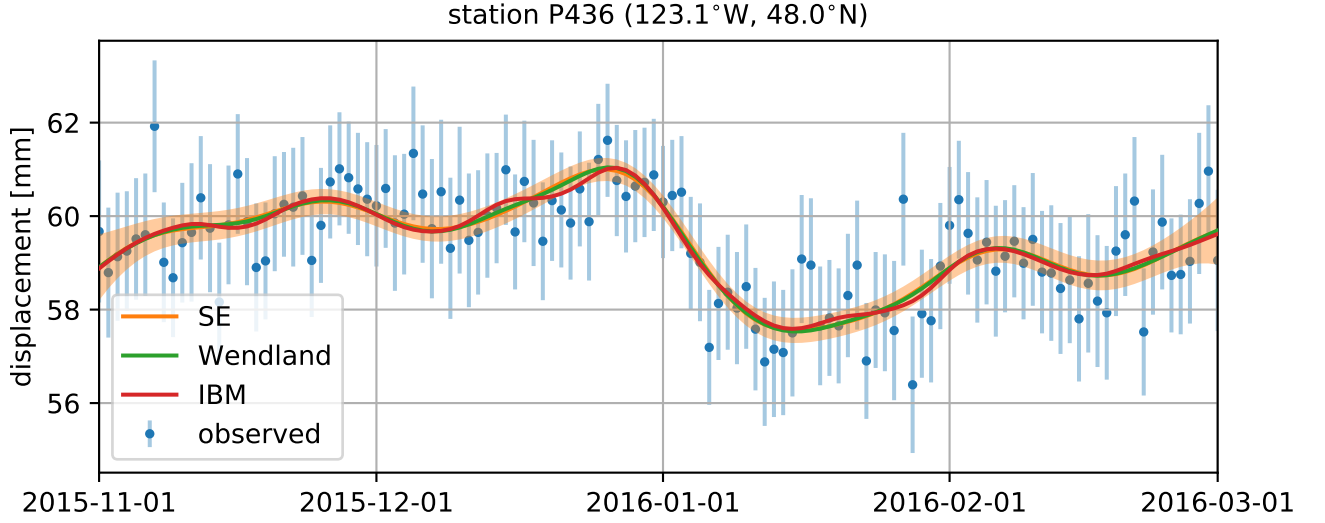


Figure 5.5: Observed easting component of displacements at station P436 and predicted displacements when using different covariance functions for T . The one standard deviation uncertainties are shown for the observations and the predicted displacements when using the SE covariance function. For clarity, uncertainties are not shown for the IBM and Wendland covariance functions, but they are nearly equivalent to the uncertainties for the SE covariance function.

signal. The data prediction vector reasonable fits the displacements throughout the SSE, regardless of the choice of T . The prediction for the IBM covariance function contains slightly more high frequency, and perhaps spurious, features. The predictions for the Wendland and SE covariance functions are nearly indistinguishable. Overall, the predicted displacements are not strongly sensitive to the choice of temporal covariance function. In our estimates of transient strain discussed in the next section, we ultimately settle on the Wendland covariance function for T and use the median values from Table 1 for the hyperparameters. We choose the Wendland covariance function over the SE covariance function because of its computational advantages.

5.5.3 Transient Strain Rates

Having established a noise model and a prior for transient displacements, we use the cleaned GNSS dataset to calculate transient strain rates in the Puget Sound region. We calculate transient strain rates 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 5.6 we show the transient strain rates on January 1, 2016, which coincides with the height of an SSE. We have included an animation showing the map view of strain rates through time as supplementary material. The strain rates shown in Figure 5.6 are generally similar to the strain rates during the other six SSEs considered in this study. The SSEs cause trench perpendicular compression in the Olympic Peninsula and extension east of Puget Sound. The strain transitions from compressional to extensional strain around the southern tip of Vancouver Island, which

coincides with the location of where thrust slip tends to be inferred for SSEs in the Puget Sound region (e.g., *Dragert et al.*, 2001; *Wech et al.*, 2009; *Schmidt and Gao*, 2010). Thus, this pattern of strain is to be expected. During the period in between SSEs, secular strain rates indicate trench perpendicular compression throughout this study region (*Murray and Lisowski*, 2000; *McCaffrey et al.*, 2007, 2013). When comparing inferred strain rates from SSEs to the secular strain rates, we see that SSEs are concentrating tectonically accumulated strain energy towards the trench, and presumably pushing the subduction zone closer to failure.

In Figure 5.7 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 compare the signal-to-noise ratio from eq. (5.15) to the frequency of seismic tremor (Figure 5.8). A signal-to-noise ratio greater than ~ 3 can be interpreted as a detected geophysical signal. For each detected event there is a corresponding peak in seismic tremor. We are also able to clearly identify transient strain associated with a more subtle SSE in August 2014. In between peaks in seismic tremor, the signal-to-noise ratio is consistently between 0 and 2, suggesting that all the transient strain detected at this location is associated with SSEs.

The results we have presented thus far indicate that we are identifying the strain that we should expect to see. There are, however, subtle features in our estimated transient strain rates which we were not expecting. For example, there is a brief period of east-west extension on the Olympic Peninsula several days prior to some of the SSEs. This feature can be seen before the summer 2012 and winter 2015-2016 SSEs in Figure 5.7 as well as in the supplementary animation. While this deformation is noteworthy, a discussion on the mechanisms causing it is outside the scope of this study.

5.6 Discussion

Our results demonstrate that GPR is an effective tool for estimating transient strain from GNSS data, which can then be used to detect geophysical processes. One may argue that geophysical signal can also be detected by merely inspecting the GNSS displacement time series. Indeed, the SSEs identified in Figure 5.8 do produce visible displacements in the GNSS data. However, the GNSS data also contains outliers and non-tectonic deformation that is localized to individual stations. In contrast, our estimates of transient strain only identify features that are sufficiently spatially and temporally coherent, based on our chosen prior model. Furthermore, our estimates of transient strain are insensitive to common mode noise, which is highly spatially correlated noise resulting from factors such as reference frame error. Common mode noise can obscure geophysical signal in the GNSS data, but it gets canceled out when computing the transient strain. Lastly, our estimates of transient strain rates are a spatial and temporal derivative of displacements, and thus any geophysical signal in the transient strain rates tends to be more pronounced than in the GNSS data. For these reasons, we argue that transient strain rates estimated with the method described in Section 5.3 can illuminate geophysical signal that may not be discernible from the noise in the GNSS displacement data.

In addition to detecting geophysical processes, the GNSS derived transient strain rates

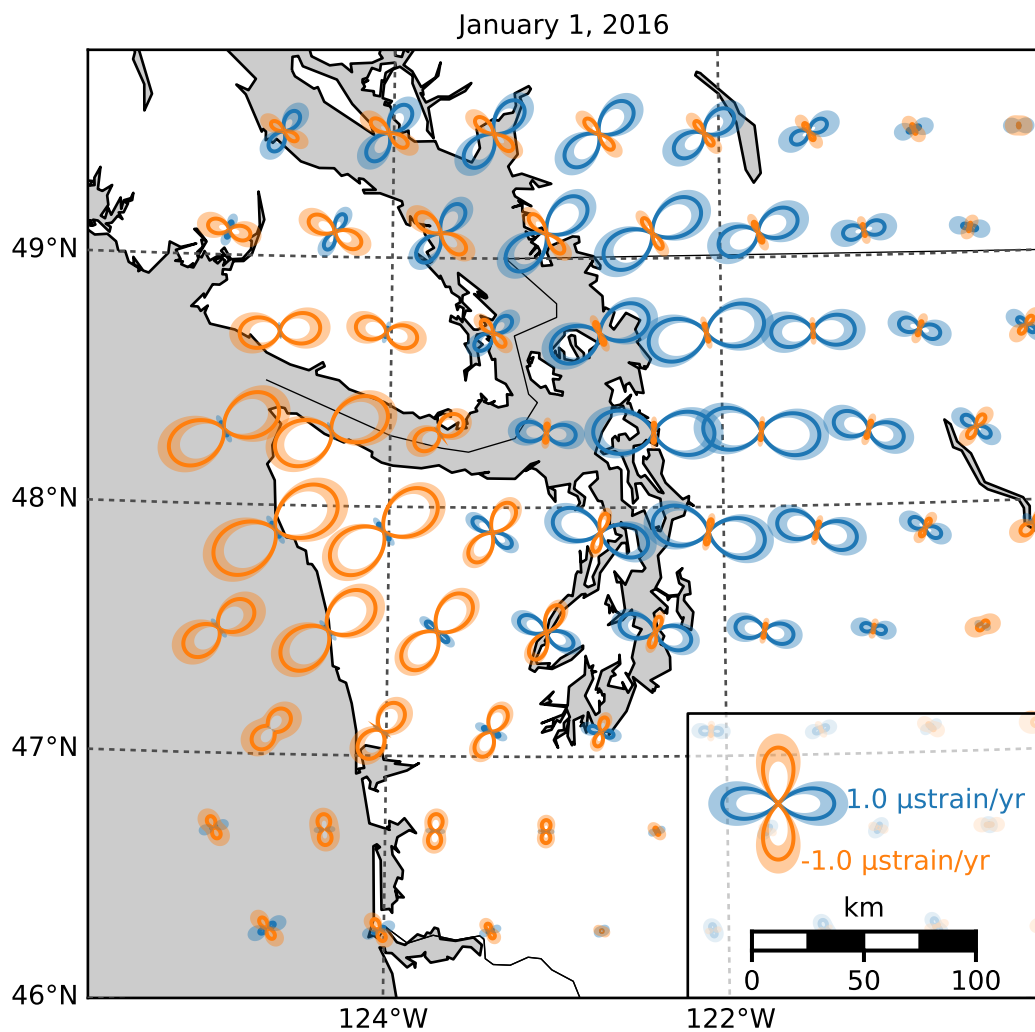


Figure 5.6: Estimated transient strain rates during the Winter 2015-2016 SSE. Strain glyphs show the normal strain rate along each azimuth, where orange indicates compression and blue indicates extension. The shaded regions indicate one standard deviation uncertainties in the normal strain rates.

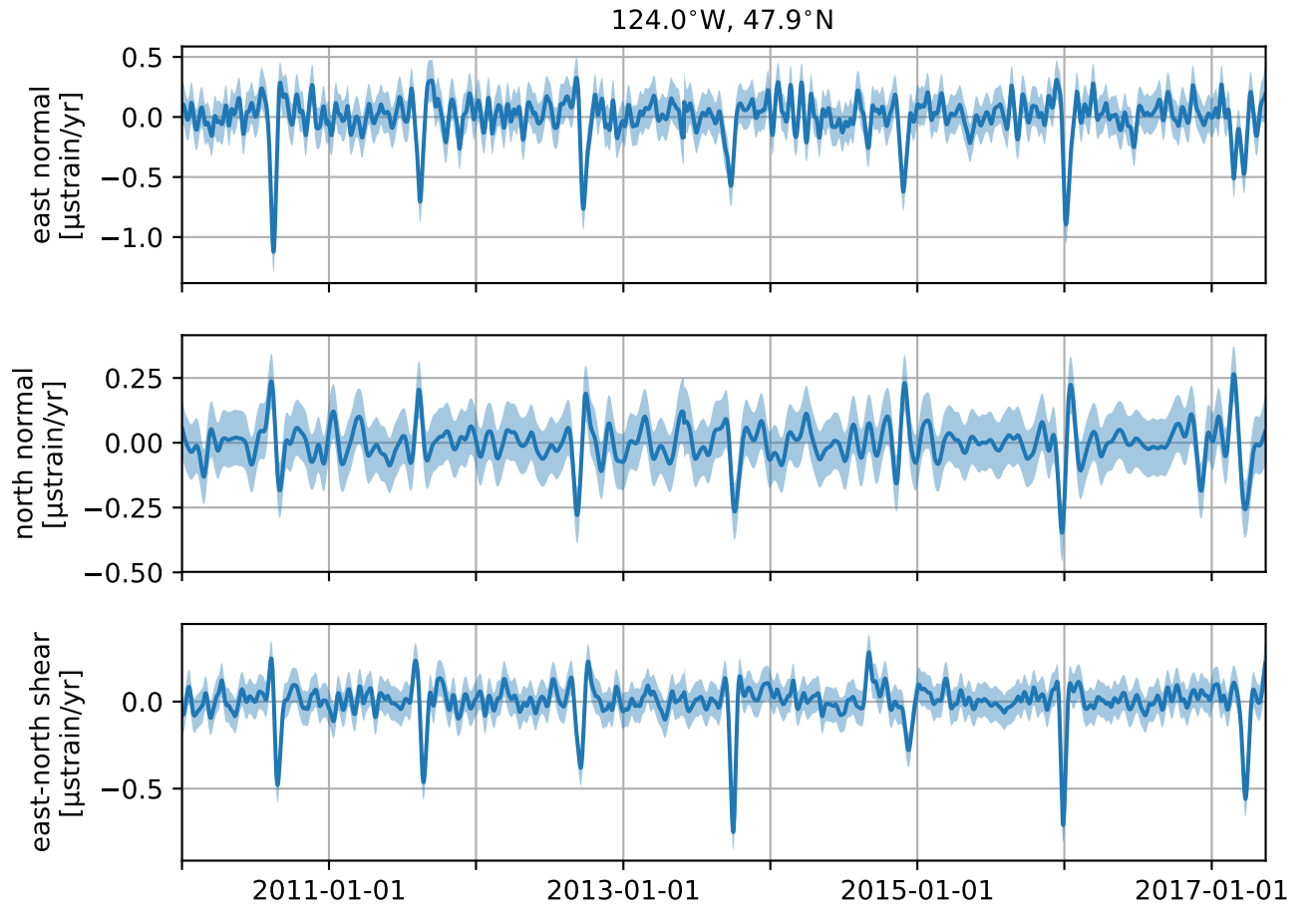


Figure 5.7: Three components of the transient horizontal strain rate tensor estimated at the position shown in Figure 5.1. The shaded regions indicate one standard deviation uncertainty.

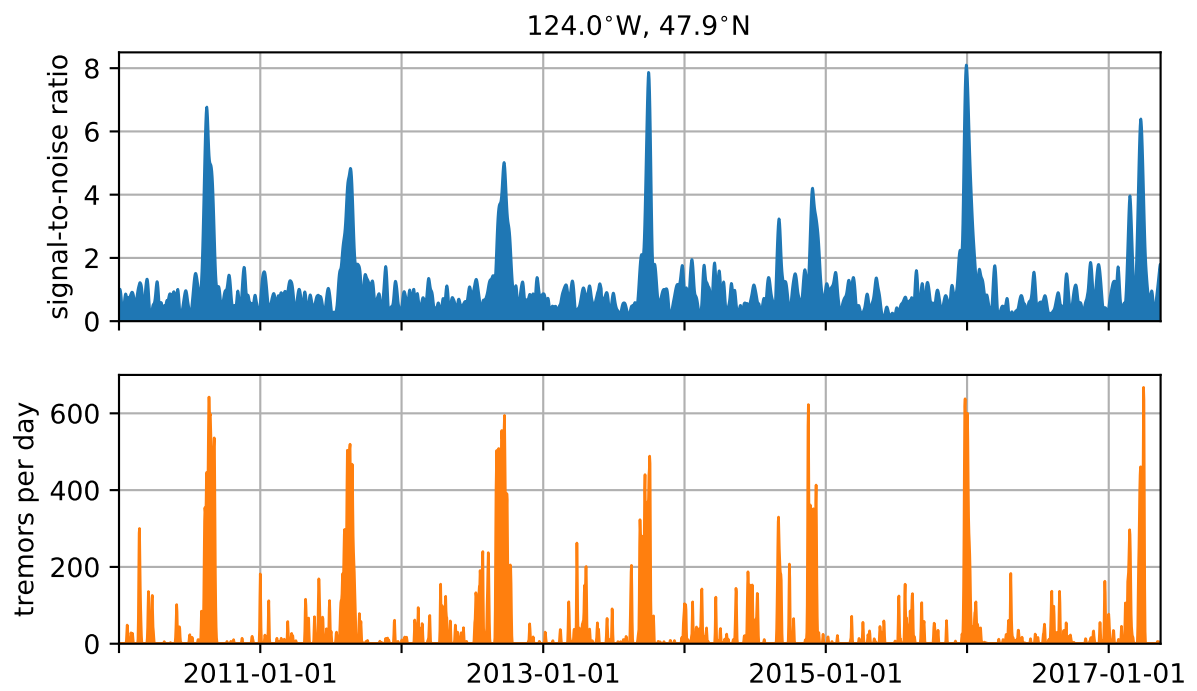


Figure 5.8: (top) Signal-to-noise ratio (eq. 5.15) at the position shown in Figure 5.1. (bottom) Frequency of tremors in the region shown in Figure 5.1.

can be used to better understand the data from borehole strain meters (BSMs). The Plate Boundary Observatory maintains 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 and 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 Strait 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.

There is potential for a more thorough analysis of the spatio-temporal noise in GNSS data, η , than what was performed in Section 5.5.1. We did not explore the spatial covariance of η , which would describe common mode noise. We are able to ignore common mode error in this study; however, 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 can also improve upon the seasonal 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 (i.e., the number of sinusoids needed to describe the observations). 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

$$T(t, t') = \phi^2 \exp \left(\frac{-\sin(\pi|t - t'|)^2}{2\tau^2} \right). \quad (5.31)$$

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 5.5.1.

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 eqs. (5.10) and (5.11) 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 a compact Wendland covariance function. Using a compact covariance function for our prior turns eqs. (5.10) and (5.11) into 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. (5.10) and (5.11) 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.

5.7 Conclusion

In this paper 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, 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. While this paper just focuses on estimating transient strain rates, we believe that GPR is a powerful tool that can be applied to a wide range of geophysical problems.

5.8 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.

5.9 References

- Beavan, J., and J. Haines (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., C. Kreemer, W. C. Hammond, and J. Gazeaux (2016), MIDAS robust trend estimator for accurate GPS station velocities without step detection, *Journal of Geophysical Research : Solid Earth*, 121, 2054–2068, doi:10.1002/2015JB012552. Received.
- Chen, Y., T. a. Davis, and W. W. Hager (2008), Algorithm 887 : CHOLMOD , Supernodal

- Sparse Cholesky Factorization and Update/downdate, *ACM Transactions on Mathematical Software*, 35(3), 1–12, doi:10.1145/1391989.1391995.
- Cressie, N. (1993), *Statistics for Spatial Data*, rev. ed., John Wiley & Sons, New York, doi:10.1111/j.1365-3121.1992.tb00605.x.
- Dong, D., P. Fang, Y. Bock, M. K. Cheng, and S. Miyazaki (2002), Anatomy of apparent seasonal variations from GPS-derived site position time series, *J. Geophys. Res.*, 107(B4), 2075, doi:10.1029/2001JB000573.
- Dong, D., P. Fang, Y. Bock, F. Webb, L. Prawirodirdjo, S. Kedar, and P. Jamason (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, doi:10.1029/2005JB003806.
- Dragert, G., K. Wang, and T. S. James (2001), A silent slip event on the deeper Cascadia subduction interface., *Science*, 292, 1525–1528, doi:10.1126/science.1060152.
- Dragert, H., and K. Wang (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, doi:10.1029/2011JB008609.
- El-Fiky, G. S., and T. Kato (1998), Continuous distribution of the horizontal strain in the Tohoku district, Japan, predicted by least-squares collocation, *Journal of Geodynamics*, 27(2), 213–236, doi:10.1016/S0264-3707(98)00006-4.
- Feigl, K. L., R. W. King, and T. H. Jordan (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, doi:10.1029/JB095iB03p02679.
- Field, E. H., et al. (2014), Uniform California Earthquake Rupture Forecast, version 3 (UCERF3) -The time-independent model, *Bulletin of the Seismological Society of America*, 104(3), 1122–1180, doi:10.1785/0120130164.
- Freed, a. M., and J. Lin (2001), Delayed triggering of the 1999 Hector Mine earthquake by viscoelastic stress transfer., *Nature*, doi:10.1038/35075548.
- Gibbs, B. P. (2011), *Advanced Kalman Filtering, Least-Squares and Modeling: A Practical Handbook*, John Wiley & Sons, Hoboken, NJ, doi:10.1002/9780470890042.
- Gladwin, M. T., R. L. Gwyther, R. Hart, M. Francis, and M. J. S. Johnston (1987), Borehole tensor strain measurements in California, *Journal of Geophysical Research: Solid Earth*, 92(B8), 7981–7988, doi:10.1029/JB092iB08p07981.
- Hart, R. H. G., M. T. Gladwin, R. L. Gwyther, D. C. Agnew, and F. K. Wyatt (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.
- Herring, T. A., et al. (2016), Plate Boundary Observatory and related networks: GPS data analysis methods and geodetic product, *Reviews of Geophysics*, pp. 1–50, doi: 10.1002/2016RG000529. Received.
- Hines, T. T., and E. A. Hetland (2016), Rheologic constraints on the upper mantle from five years of postseismic deformation following the El Mayor-Cuapah earthquake, *Journal of Geophysical Research: Solid Earth*, *121*, doi:10.1002/2016JB013114.
- Hines, T. T., and E. A. Hetland (2017), Unbiased characterization of noise in geodetic data, *submitted to Journal of Geodesy*.
- Hodgkinson, K., D. Agnew, and E. Roeloffs (2013), Working With Strainmeter Data, *Eos, Transactions American Geophysical Union*, *94*(9), 91–91, doi:10.1002/2013EO090011.
- Holt, W. E., and G. Shcherbenko (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, doi: 10.1785/0220130004.
- Kato, T., G. S. El-Fiky, E. N. Oware, and S. Miyazaki (1998), Crustal strains in the Japanese islands as deduced from dense GPS array, *Geophysical Research Letters*, *25*(18), 3445–3448, doi:10.1029/98GL02693.
- Langbein, J. (2004), Noise in two-color electronic distance meter measurements revisited, *Journal of Geophysical Research: Solid Earth*, *109*(4), 1–16, doi:10.1029/2003JB002819.
- Langbein, J. (2008), Noise in GPS displacement measurements from Southern California and Southern Nevada, *Journal of Geophysical Research: Solid Earth*, *113*(5), 1–12, doi: 10.1029/2007JB005247.
- Langbein, J., and H. Johnson (1997), Correlated errors in geodetic time series: Implications for time-dependent deformation, *Journal of Geophysical Research*, *102*(B1), 591–603, doi: 10.1029/96JB02945.
- Lisowski, M., D. Dzurisin, R. P. Denlinger, and E. Y. Iwatsubo (2008), Analysis of GPS-Measured Deformation Associated with the 2004–2006 Dome-Building Eruption of Mount St. Helens, Washington, *Tech. Rep. September 1984*.
- Lohman, R. B., and J. R. Murray (2013), The SCEC Geodetic Transient-Detection Validation Exercise, *Seismological Research Letters*, *84*(3), 419–425, doi:10.1785/0220130041.
- Mackay, D. J. C. (1998), Introduction to Gaussian processes, *Neural Networks and Machine Learning*, *168*(1996), 133–165, doi:10.1007/s10067-003-0742-1.
- Mao, A., G. A. Harrison, and H. Dixon (1999), Noise in GPS coordinate time series, *Journal of Geophysical Research*, *104*(B2), 2797–2816.

- McCaffrey, R., A. I. Qamar, R. W. King, R. Wells, G. Khazaradze, C. A. Williams, C. W. Stevens, J. J. Vollick, and P. C. Zwick (2007), Fault locking, block rotation and crustal deformation in the Pacific Northwest, *Geophysical Journal International*, *169*(3), 1315–1340, doi:10.1111/j.1365-246X.2007.03371.x.
- McCaffrey, R., R. W. King, S. J. Payne, and M. Lancaster (2013), Active tectonics of northwestern U.S. inferred from GPS-derived surface velocities, *Journal of Geophysical Research: Solid Earth*, *118*, 709–723, doi:10.1029/2012JB009473.
- McGuire, J. J., and P. Segall (2003), Imaging of aseismic fault slip transients recorded by dense geodetic networks, *Geophysical Journal International*, *155*, 778–788, doi:10.1111/j.1365-246X.2003.02022.x.
- Meade, B. J., and B. H. Hager (2005), Block models of crustal motion in southern California constrained by GPS measurements, *Journal of Geophysical Research: Solid Earth*, *110*, 1–19, doi:10.1029/2004JB003209.
- Miyazaki, S., J. J. McGuire, and P. Segall (2003), A transient subduction zone slip episode in southwest Japan observed by the nationwide GPS array, *Journal of Geophysical Research*, *108*(B2), 1–15, doi:10.1029/2001JB000456.
- Moritz, H. (1978), Least-Squares Collocation, *Reviews of Geophysics*, *16*(3), 421–430, doi:10.1029/RG016i003p00421.
- Murray, M. H., and M. Lisowski (2000), Strain accumulation along the Cascadia subduction zone in western Washington, *Geophysical Research Letters*, *27*(22), 3631–3634.
- Ohtani, R., J. J. McGuire, and P. Segall (2010), Network strain filter: A new tool for monitoring and detecting transient deformation signals in GPS arrays, *Journal of Geophysical Research: Solid Earth*, *115*(12), 1–17, doi:10.1029/2010JB007442.
- 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*, 3 ed., McGraw-Hill, New York.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (2007), *Numerical Recipes: The Art of Scientific Computing*, 3 ed., Cambridge University Press, Cambridge, doi:10.1137/1031025.
- Rasmussen, C. E., and C. K. I. Williams (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, doi:10.1029/2009JB006407.

- Roeloffs, E. A. (2006), Evidence for Aseismic Deformation Rate Changes Prior To Earthquakes, *Annual Review of Earth and Planetary Sciences*, *34*(1), 591–627, doi:10.1146/annurev.earth.34.031405.124947.
- Rogers, G., and H. Dragert (2003), Episodic tremor and slip on the Cascadia subduction zone: the chatter of silent slip., *Science*, *300*, 1942–1943, doi:10.1126/science.1084783.
- Sandwell, D. T., and P. Wessel (2016), Interpolation of 2-D vector data using constraints from elasticity, *Geophysical Research Letters*, pp. 1–7, doi:10.1002/2016GL070340.Received.
- Schmidt, D. A., and H. Gao (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, doi:10.1029/2008JB006045.
- Segall, P., and M. Mathews (1997), Time dependent inversion of geodetic data, *Journal of Geophysical Research*, *102*(B10), 22,391–22,409.
- Shen, Z., M. Wang, Y. Zeng, and F. Wang (2015), Optimal Interpolation of Spatially Discretized Geodetic Data, *Bulletin of the Seismological Society of America*, *105*(4), 2117–2127, doi:10.1785/0120140247.
- Shen, Z. K., D. D. Jackson, B. X. Ge, and X. G. Bob (1996), Crustal deformation across and beyond the Los Angeles basin from geodetic measurements, *Journal of Geophysical Research*, *101*(B12), 27,927–27,957, doi:10.1029/96JB02544.
- Tape, C., P. Musé, M. Simons, D. Dong, and F. Webb (2009), Multiscale estimation of GPS velocity fields, *Geophysical Journal International*, *179*(2), 945–971, doi:10.1111/j.1365-246X.2009.04337.x.
- Wdowinski, S., J. Zhang, P. Fang, and J. Genrich (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, doi:10.1785/gssrl.
- Wech, A. G., K. C. Creager, and T. I. Melbourne (2009), Seismic and geodetic constraints on Cascadia slow slip, *Journal of Geophysical Research: Solid Earth*, *114*(10), 1–9, doi:10.1029/2008JB006090.
- Wendland, H. (2005), *Scattered data approximation*, 348 pp., doi:10.2277/0521843359.
- Williams, S. D. P., Y. Bock, P. Fang, P. Jamason, R. M. Nikolaidis, L. Prawirodirdjo, M. Miller, and D. J. Johnson (2004), Error analysis of continuous GPS position time series, *Journal of Geophysical Research: Solid Earth*, *109*(B3), doi:10.1029/2003JB002741.
- 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., Y. Bock, H. Johnson, P. Fang, S. Williams, J. Genrich, S. Wdowinski, and J. Behr (1997), Southern California Permanent GPS Geodetic Array: Error analysis of daily position estimates and site velocities, *Journal of Geophysical Research*, *102*(B8), 18,035–18,055.