1 Introduction

Crustal strain rates are fundementally important quantities for assessing seismic hazard. This is because the locations where strain is rapidly accumulating are the locations where we can expect strain energy to be released seismically. Maps of strain rate can be derived from geodetic measurements of ground displacements, and there are numerous methods for doing so. The classic and simplest way is to assume that the strain rate is constant in time and spatially uniform within subnetworks of the geodetic data. Linear least squares is then used to find the components of the strain rate tensor for each subnetwork (e.g Frank, 1966; Prescott, 1976; Savage et al., 1986; Feigl et al., 1993; Murray and Lisowski, 2000). Several algorithms have been developed to improve upon this procedure for calculating strain. Shen et al. (1996) and Shen et al. (2015) discuss an algorithm where, instead of using the immediately adjacent stations to calculate strain at a position, the strain is computed with a weighted average over the entire network where the weighting is smaller for more distant stations. Another strategy is to fit a set of interpolating basis functions to the velocity field and then compute the strain from the analytical derivative of the interpolant (e.g. Beavan and Haines, 2001; Tape et al., 2009). The aforementioned studies have all been concerned with estimating long term strain rates. Time dependent strain would be useful for studying geophysical processes which occur over timescales of days to years such as slow slip events, postseismic relaxation, or volcanic deformation. Ohtani et al. (2010) describes a Kalman filter based method for computing time dependent strain by fitting a set of basis functions to a time dependent displacement field and enforcing temporal smoothness in the basis function coefficients.

In essence, estimating strain rates is a matter of numerically calculating the spatial derivative of the geodetically observed velocity field. Calculating the spatial derivative is complicated by the fact that the velocity estimates tend to be noisy, and differentiation will only amplify the noise. Furthermore, the velocities are not observed on a regular grid, preventing the use of standard finite difference methods for computing derivatives. In this paper we demonstrate that both of these complications can be elegantly handled with the recently popularized Radial Basis Function-Finite Difference (RBF-FD) method (Wright and Fornberg, 2006). Our method is computationally efficient and stable (there is no inversion of an ill conditioned matrix). It also does not require tuning an intangible penalty parameter, and instead just requires the user to specify a minimum spatial wavelength for the underlying geophysical signal. Because of the simplicity of out method, it can easily be extended to time dependent estimates of strain by independently estimating strain for velocity fields for different time epochs. Additionally, our method can be easily extending to account for known discontinuities in the displacement field from, for example, a creeping fault.

and rather than tuning an there is no need to tune penalty parameters or hyper parameters. Due to the ease of Our procedure also can be For Due to the speed and reliability of recovering a final strain Our method is entirely For these it does not require tuning hyper tuning, and compuOur we do not claim

The aforementioned studies do not describe the Thi This procedure for calculating strain has been imrpoved upon by One algorithm which improves u underlying assumptions are with this procedure is that The underlying assumptions for this procedure are generally Several algorithms have been developed to mprove to find be constant and uniform over a region and then find the

The strain rate is merely the spatial derivatives of the velocity field; however, computing the spatial derivative is complicated by 1) the irregular spacing of geodetic observations and 2) the velocity estimates tend to be noisey and some degree of spatial smoothing or regularization is necessary. calculating spatial derivative has however, there are difficulties in doing so. Calculating strain is essentially,

energy, which may be released seismically, is building up an

One can derive a map of strain rates from geodetically observed ground deformation; however, one must devise a strat Past studies which used triangulation surveys estimated strain by finding the best strain field which is uniform in space and constant in time for each triangulation network (e.g. Prescott, 1976; Savage et al., 1991; Murray and Lisowski, 2000). The assumption of Modern studies using GPS data have used a This approach produces strain studies studies would estimate strain rates by finding the best method is to find the best fitting uniform strain rate tensor for smaller of the data (e.g Savage et al., 1991; Murray and Lisowski, 2000). of the eeis to assume a uniform strain field

These challenges can be overcome by assuming that Numerous procedures have been used to overcome these challenges and most of them have been used to calculate long-term strain rates from time averaged velocities (e.g. Frank, 1966; Savage et al., 1991; Shen et al., 1996, 2015). A more difficult task is to calculate time dependent strain rates from time dependent velocities, which are even noisier and require a more sophisticated smoothing approach, such as the Network Strain Filter introduced by (cite).Despite its difficulty, there is great motivation for calculating time dependent strain rates.

from Most often, the displacements

some smoothing and must must be smoothed out in order to is generally noisy and . data must be smoothed so as not to different and there have been numerous procedures have been proposed to used for t have been proposed to do so.

however, the process is not straight forward because There are two challenges with process of estimating strain is not entirely straight forward because there are two challenges that

however, the process for deriving strain rates is not necessarily straight he , Maps of strain Unfortunately, measuring strain accumulation on is

because it can be used to assess the energy being accumulated that could be released seismically.

on the Surface strain rates on the Crustal strain rates are a fundementally important quantity is a fundemental quantity which is fundemental

2 Method

2.1 Smoothing

We discuss a method to smooth observed data, u_{obs}, which can be spatial, temporal, or higher dimensional data, and the observations are generally irregularly spaced. Methods which handle this task often fall in at least one of two catagories. One approach is to fit a set of basis functions to the observations using least-squares or regularized least-squares (e.g. Fasshauer, 2007). Smoothing splines can be viewed in this context where the basis functions being fit to the observation are polyharmonic splines. The second catagory is Gaussian process regression, which is a Bayesian technique where a stochastic prior model is assumed for the underlying signal (e.g. Rasmussen and Williams, 2006). Kriging is among the better known examples of Gaussian process regression. The two approaches are not always distinct. For example, Kimeldorf and Wahba (1970) showed that smoothing splines can be cast as a Bayesian estimation problem with the appropriate stochastic prior model. Our method is an example of Gaussian process regression where our smoothed solution, u_{post}, incorporates u_{obs} and a stochastic prior model for the underlying signal which we are trying to recover. We constrain u_{post} with the observation equation

$$u_{\text{post}} = u_{\text{obs}} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{C}_{\text{obs}}), \tag{1}$$

and the prior model

$$u_{\text{prior}} \sim \mathcal{N}(0, \mathbf{C}_{\text{prior}}), \tag{2}$$

where ϵ and u_{prior} are considered to be Gaussian processes with zero mean and covariances $\mathbf{C}_{\text{obs}}$ and $\mathbf{C}_{\text{prior}}$ respectively. The solution for u_{post} minimizes the objective function

$$||u_{\text{post}} - u_{\text{obs}}||^2_{\mathbf{C}_{\text{obs}}} + ||u_{\text{post}}||^2_{\mathbf{C}_{\text{prior}}} \tag{3}$$

and is itself a Gaussian process with a distribution described by

$$u_{\text{post}} \sim \mathcal{N}(\bar{u}_{\text{post}}, \mathbf{C}_{\text{post}}). \tag{4}$$

We use $\bar{u}_{\text{post}}$ and $\mathbf{C}_{\text{post}}$ to denote the mean and covariance of u_{post} respectively. Using Bayesian linear regression (e.g. Tarantola, 2005) these values are found to be

$$\begin{aligned} \bar{u}_{\text{post}} &= (\mathbf{C}_{\text{obs}}^{-1} + \mathbf{C}_{\text{prior}}^{-1})^{-1}\mathbf{C}_{\text{obs}}^{-1}u_{\text{obs}} \\ \mathbf{C}_{\text{post}} &= (\mathbf{C}_{\text{obs}}^{-1} + \mathbf{C}_{\text{prior}}^{-1})^{-1}. \end{aligned} \tag{5}$$

$\mathbf{C}_{\text{obs}}$ is presumably well known, while $\mathbf{C}_{\text{prior}}$ needs to be chosen based on an understanding of the underlying signal which we are trying to estimate. With a judicious choice of $\mathbf{C}_{\text{prior}}$, eq. (5) can be made equivalent to several well established smoothing methods. In particular, we demonstrate how eq. (5) can be viewed as a low-pass filter with a well defined cutoff frequency. This is first demonstrated for filtering one-dimensional data, and the extension to higher dimensions follows naturally.

2.1.1 One-dimensional smoothing

For one-dimensional data we consider a prior which can be stated implicitly as

$$\mathbf{D}_n u_{\text{prior}} = q, \quad q \sim \mathcal{N}(0, \lambda^2), \tag{6}$$

where $\mathbf{D}_n$ is an n'th order differentiation matrix, and q is white noise with constant variance λ^2. If we momentarily ignore the fact that $\mathbf{D}_n$ is not invertible then we can explicitly write our prior covariance as

$$\mathbf{C}_{\text{prior}} = \lambda^2(\mathbf{D}_n^T\mathbf{D}_n)^{-1}, \tag{7}$$

and the filtered mean and covariance are described by

$$\begin{aligned}\bar{u}_{\text{post}} &= (\mathbf{C}_{\text{obs}}^{-1} + \frac{1}{\lambda^2}\mathbf{D}_n^T\mathbf{D}_n)^{-1}\mathbf{C}_{\text{obs}}^{-1}u_{\text{obs}} \\ \mathbf{C}_{\text{post}} &= (\mathbf{C}_{\text{obs}}^{-1} + \frac{1}{\lambda^2}\mathbf{D}_n^T\mathbf{D}_n)^{-1}.\end{aligned} \tag{8}$$

This filtered solution is closely tied to several well established methods of smoothing. For example, one can immediately recognize eq. (8) as an example of Tikhonov regularization. We also note a similarity between eq. (8) and smoothing splines. To see this similarity, we first recall that in one-dimension, a smoothing spline is defined as the function, $f(t)$, which minimizes,

$$\sum_{i=1}^{P}(u_{\text{obs}}^i - f(t_i))^2 + \alpha\int_{t_1}^{t_P} f^{(n)}(t)dt, \tag{9}$$

where u_{obs}^i is an observation at time t_i, P is the number of observations, α is a smoothing parameter, and $f^{(n)}$ denotes the n'th time derivative of f. If we ignore data uncertainties (i.e. $\mathbf{C}_{\text{obs}} = \mathbf{I}$), we can see from eq. (3) that $\bar{u}_{\text{post}}$ is the discrete function which minimizes

$$||u_{\text{obs}} - \bar{u}_{\text{post}}||_2^2 + \frac{1}{\lambda^2}||D_n\bar{u}_{\text{post}}||_2^2. \tag{10}$$

If u_{obs} is sampled at evenly spaced increments then, with the appropriate choice of smoothing parameters, eq. (10) can be recognized as a discretized form of (9) and we would thus expect $f(t)$ and $\bar{u}_{\text{post}}$ to be effectively the same. The similarities break down when dealing with non-uniformly spaced observations.

Whether we use a smoothing spline or eq. (8) to filter u_{obs}, we must choose the appropriate penalty parameter. One common method for choosing an appropriate penalty parameter is generalized cross-validation (Craven and Wahba, 1979), which yields a filtered solution with the maximum predictive power. There is merit to using an entirely objective approach such as cross-validation, and this would be approriate if there is no prior knowledge of the signal's characteristic wavelength. Otherwise, it may be better to chose a penalty parameter that damps out all the high frequency oscillations which are known to be noise. We elaborate on this point by demonstrating that eq. (8) can also be viewed as a low-pass filter with a cutoff frequency determined by λ.

We wish to transform $\bar{u}_{\text{post}}$ in eq. (8) to the frequency domain and in order to do so we assume that u_{obs} has a constant sampling rate. We also require that ϵ and u_{prior} are stationary stochastic processes (i.e. their statistical properties are invariant to time shifts). We require them to be stationary so that matrix multiplication by $\mathbf{C}_{\text{obs}}$ and $\mathbf{C}_{\text{prior}}$ can be viewed as convolution in the time domain and thus multiplication in the frequency domain. We then consider the simplifying case where ϵ is white noise with constant variance σ^2 (i.e. $\mathbf{C}_{\text{obs}} = \sigma^2\mathbf{I}$), and $\mathbf{D}_n$ is the periodic spectral differentiation matrix (e.g. Trefethen, 2000). Under a discrete Fourier transform, $\mathbf{D}_n$ has the properties

$$\mathcal{F}[\mathbf{D}_n g] = (2\pi i\omega)^n\hat{g} \tag{11}$$

and

$$\mathcal{F}[\mathbf{D}_n^T g] = (-2\pi i\omega)^n\hat{g}, \tag{12}$$

where ω is the frequency domain variable, g is an arbitrary vector and $\hat{g}$ is its discrete Fourier transform. The discrete Fourier transform of $\bar{u}_{\text{post}}$ is then

$$\hat{u}_{\text{post}}(\omega) = \frac{\frac{1}{\sigma^2}}{\frac{1}{\sigma^2} + \frac{(2\pi\omega)^{2n}}{\lambda^2}}\hat{u}_{\text{obs}}(\omega). \tag{13}$$

We make the change of variables

$$\lambda^2 = (2\pi\omega_c)^{2n}\sigma^2 \tag{14}$$

which changes the hyperparameter from λ to ω_c. The reason for this change of variables becomes apparent when we simplify eq. (13) to

$$\hat{u}_{\text{post}}(\omega) = \frac{1}{1 + \left(\frac{\omega}{\omega_c}\right)^{2n}}\hat{u}_{\text{obs}}(\omega). \tag{15}$$

We can recognize eq. (15) as an n'th order low-pass Butterworth filter with cut-off frequency ω_c. In the limit as $n \to \infty$ eq. (15) becomes an ideal low-pass filter which removes all frequencies above ω_c and leaves lower frequencies unaltered. Of course, an ideal low-pass filter is often undesirable because it will tend to produce

ringing artifacts in the filtered solutions. When modeling u_{prior} as Brownian motion or integrated Brownian motion, where $n = 1$ and $n = 2$ respectively, the transfer function is tapered across ω_c, which ameliorates ringing in the filtered solution.

By demonstrating that eq. (8) can be made equivalent to a low-pass filter, it may not be clear why we would ever use eq. (8) when it is far more efficient to filter in the frequency domain through the Fast Fourier Transform. In order to make use of the Fast Fourier Transform, the observations must be made at a constant sampling rate and the observation noise must be white with constant variance. In contrast, these conditions do not need to be met in order to evaluate eq. (8). The question is then whether eq. (8) still effectively acts as a low-pass filter when the idealized conditions are not met.

We address this question here with two numerical demonstrations. In this first demonstration we generate 100 samples of synthetic data over the time interval $0 < t < 1$ yr. The true signal in the syntethic data is a sine wave with a frequency of 1 yr^{-1} and 1 mm amplitude. We obscure the synthetic data with white noise that does not have a constant variance. We randomly pick a variance for each datum from a uniform random distribution ranging from 0.5 to 2 mm^2. For the period from 1 to 2 years, we assign a variance of 25 mm^2 to the data, effectively making the data uniformitive for the filtered solution. The synthetic data and its estimated power spectral density are plotted in Figure 1. Our goal is to use eq. (8) as a low-pass filter which damps out frequencies that are higher than our signal. Since our data does not have constant variance, we cannot choose a cutoff frequency from eq. (14). Instead we define a characteristic data variance as

$$\frac{1}{\bar{\sigma^2}} = \frac{1}{P}\text{tr}\left(\mathbf{C}_{\text{obs}}^{-1}\right), \tag{16}$$

where P is the number of observations, and we relate λ to ω_c by

$$\lambda^2 = (2\pi\omega_c)^{2n}\bar{\sigma}^2. \tag{17}$$

In this demonstration D_n is still the spectral differentiation matrix and we use $n = 2$. Our cutoff frequency is chosen to be $\omega_c = 2$ yr^{-1}. The filtered solution and its frequency content are plotted Figure 1. Based on panel B, eq. (8) is indeed acting as a low-pass filter with the specified cut-off frequency even though the variances are not constant. However, the frequency content of subsections of $\bar{u}_{\text{post}}$ appears to change depending on the variance of u_{obs}. For example, Over the interval $1 < t < 2$, where the variance of u_{obs} is higher, u_{post} lacks the higher frequency oscillations that can be seen in the remainder of the time series. This is desirable behavior because we do not want u_{post} contorting to fit dubious data. It is also worth pointing out that u_{post} asymptotically approaches a solution that is not significantly different from that shown in Figure 1 when the variance for u_{obs} increases to infinity over the interval $1 < t < 2$. This too is desirable behaviour because one could then surmise that an appropriate way to handle missing data with eq. (8) is to assign u_{obs} an infinite variance for the missing time period. With this insight, we can then extend the application of eq. (8) from data smoothing to data interpolation and extrapolation.

We present a similar demonstration to show that eq. (8) effectively acts as a low-pass filter even when the observations are not uniformly spaced. We take samples over the time interval $(0, 10)$ yr, where the sampling rate is 0.05 yr^{-1} for the first 5 years and 0.0125 yr^{-1} for the last 5 years. We use the same underlying signal from the previous demonstration and we add noise with a constant variance of 1 mm^2. We keep $n = 2$ but since we are dealing with non-uniformly spaced data, we can no longer use the periodic spectral differentiation matrix, and $\mathbf{D}_n$ is instead a first-order accurate finite difference matrix for arbitrarily spaced data (Fornberg, 1996). Once again, our cutoff frequency is $\omega_c = 2$ yr^{-1}. The synthetic data and filtered solution are shown in panel A of Figure 2. The power spectral density for the time intervals $(0, 5)$ and $(5, 10)$ yr are shown in panel B and C, respectively. In both panels B and C, the spectral density of the filtered solution diverges from the spectral density of the synthetic observations at the expected cutoff frequency of ω_c. For frequencies above $\sim$5 yr^{-1}, the rate that the spectral content decays is slower than what is predicted by the transfer function in eq. (15). This is most pronounced in panel C, and we attribute the discrepancy to our lower order differentiation matrix. Barring the high frequencies which have negligable power in $\bar{u}_{\text{post}}$, the power spectral density of $\bar{u}_{\text{post}}$ relative to u_{obs} is accurately described by the transfer function in eq. (15), despite our use of non-uniform spacing.

2.1.2 Smoothing in Higher Dimensions

We expand our discussion to smoothing data which is observed in d-dimensional space. The filtered solution is still given by eq. (5), and we consider the prior model

$$\mathbf{L}_n u_{\text{prior}} = q, \quad q \sim \mathcal{N}(0, \lambda^2) \tag{18}$$

where $\mathbf{L}_n$ is a differentiation matrix which approximates the operation

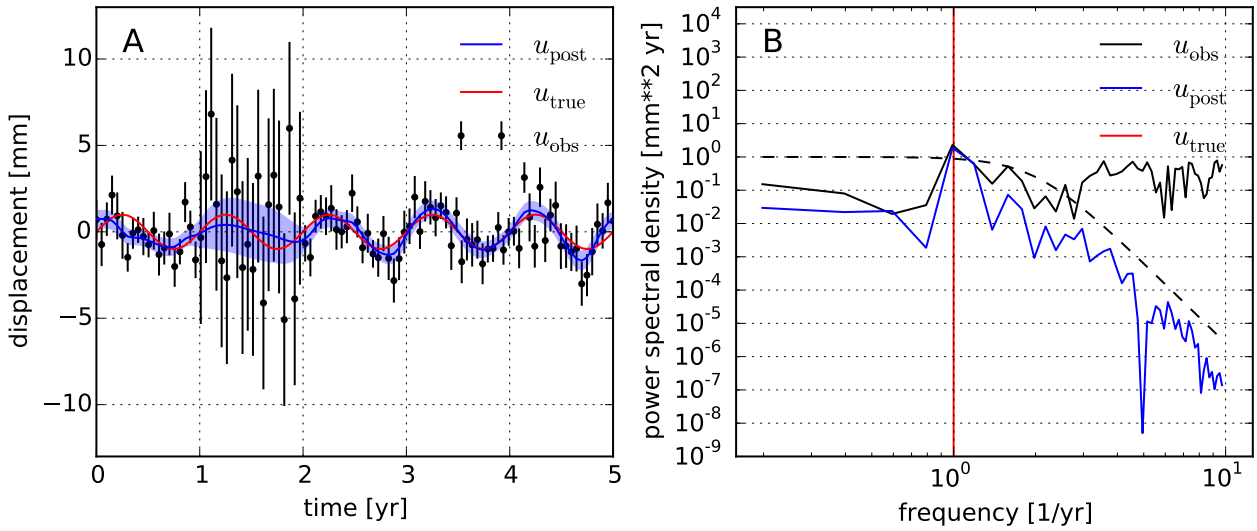

Figure 1: Panel A shows u_{obs} (black scatter points), u_{post} (blue line), and the true signal which we are trying to recover (red line). The lines on each scatter point and the light blue region show the one standard deviation uncertainty for the observations and filtered solution respectively. Panel B shows the estimated power spectral density for the observed, filtered, and true signal. The black dashed line is the squared transfer function in eq. (15) with $\omega_c = 2\ \text{yr}^{-1}$, which roughly indicates how eq. (8) scales the frequency content of u_{obs}. If the data variances were constant then the frequency content of u_{obs} and $\bar{u}_{\text{post}}$ would be exactly related by the black dashed line.

$$\sum_{i=1}^{d} \frac{\partial^n}{\partial x_i^n} \tag{19}$$

and n is an even integer. The corresponding covariance matrix is then

$$\mathbf{C}_{\text{prior}} = \lambda^2 \left(\mathbf{L}_n^T \mathbf{L}_n\right)^{-1}. \tag{20}$$

Using the change of variables from eq. (17), the solution in the time domain is

$$\begin{aligned} \bar{u}_{\text{post}} &= (\mathbf{C}_{\text{obs}}^{-1} + \frac{1}{(2\pi\omega_c)^{2n}\bar{\sigma}^2}\mathbf{L}_n^T\mathbf{L}_n)^{-1}\mathbf{C}_{\text{obs}}^{-1}u_{\text{obs}} \\ \mathbf{C}_{\text{post}} &= (\mathbf{C}_{\text{obs}}^{-1} + \frac{1}{(2\pi\omega_c)^{2n}\bar{\sigma}^2}\mathbf{L}_n^T\mathbf{L}_n)^{-1}. \end{aligned} \tag{21}$$

If we again assume that the observation have constant variance and L_n is the corresponding spectral differentiation matrix, then the d-dimensional discrete Fourier transform of $\bar{u}_{\text{post}}$ is

$$\hat{u}_{\text{post}}(\omega_1, \ldots, \omega_d) = \frac{1}{1 + \left(\sum_{i=1}^{d} \left(\frac{\omega_i}{\omega_c}\right)^n\right)^2}\hat{u}_{\text{obs}}. \tag{22}$$

The transfer function in eq. (22) can once again be recognized as a low-pass filter. Namely, in the limit as $n \to \infty$, the transfer function becomes a d-dimensional box which is zero for all the frequency tuples, $(\omega_1, \ldots, \omega_n)$, which have a component greater than ω_c.

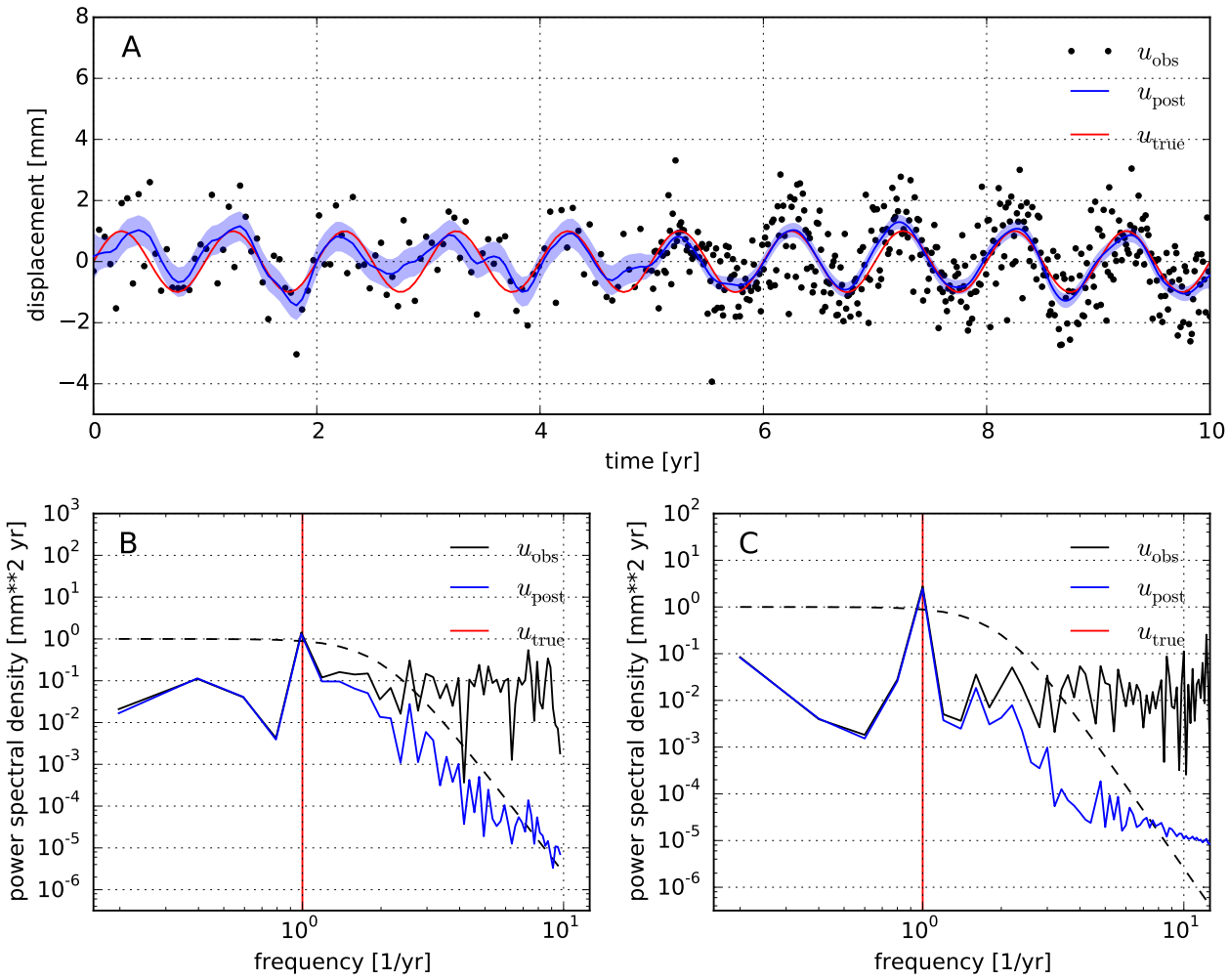

Figure 2: Panel A shows u_{obs} (black scatter points), u_{post} (blue line), and the true signal which we are trying to recover (red line). The light blue region show the one standard deviation uncertainty for u_{post}. The uncertainties for the observations are all 1 mm and are not shown for the sake of clarity. Panel B shows the estimated power spectral density for the observed, filtered, and true signal over the time interval (0,5) yr. Panel C is the same as panel B except for the time interval (5,10) yr. In both panel B and C the black dashed line is the squared transfer function in eq. (15) with $\omega_c = 2\ \mathrm{yr}^{-1}$.

2.2 Differentiating Scattered Data

3 Applications

3.1 Strain Rate in Southern California

3.2 Time Depenent Strain Rate in Cascadia

4 Discussion and Conclusion

References

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

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 403:377–403.

Fasshauer, G. E. (2007). *Meshfree Approximation Method with Matlab*. World Scientific Publishing Co., Singapore.

Feigl, K. L., Agnew, D. C., Bock, Y., and Dong, D. (1993). Space Geodetic Measurement of Crustal Deformation in Central and Southern California, 1984–1992. *Journal of Geophysical Research*, 98(B12):21 677–21 712.

Fornberg, B. (1996). *A practical guide to pseudospectral methods*. Cambridge University Press.

Frank, C. F. (1966). Deduction of earth strains from survey data. *Bulletin of the Seismological Society of America*, 56(1):35–42.

Kimeldorf, G. S. and Wahba, G. (1970). A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines. *The Annals of Mathematical Statistics*, 41(2):495–502.

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

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

Prescott, W. H. (1976). An extension of Frank's method for obtaining crustal shear strains from survey data. *Bulletin of the Seismological Society of America*, 66(6):1847–1853.

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

Savage, J. C., Lisowski, M., and Prescott, W. H. (1991). Strain Accumulation in Western United-States. *Journal of Geophysical Research*, 96(B9):14493–14507.

Savage, J. C., Prescott, W. H., and Gu, G. (1986). Strain accumulation in southern California, 1973–1984. *Journal of Geophysical Research*, 91(B7):7455–7473.

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

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

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

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.

Trefethen, L. N. (2000). *Spectral Methods in Matlab*. SIAM.

Wright, G. B. and Fornberg, B. (2006). Scattered node compact finite difference-type formulas generated from radial basis functions. *Journal of Computational Physics*, 212(1):99–123.