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 generally fall within 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). Of course the solution will be largely dependent on the subjective choice of basis functions. Common choices of basis functions include polynomials and radial basis functions (RBF's). The second catagory of smoothing methods is Gaussian process regression where a stochastic prior model is assumed for the underlying signal (e.g. Rasmussen and Williams, 2006). Kriging is perhaps the best known example of Gaussian process regression. This distinction is 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.

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 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

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

and

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

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
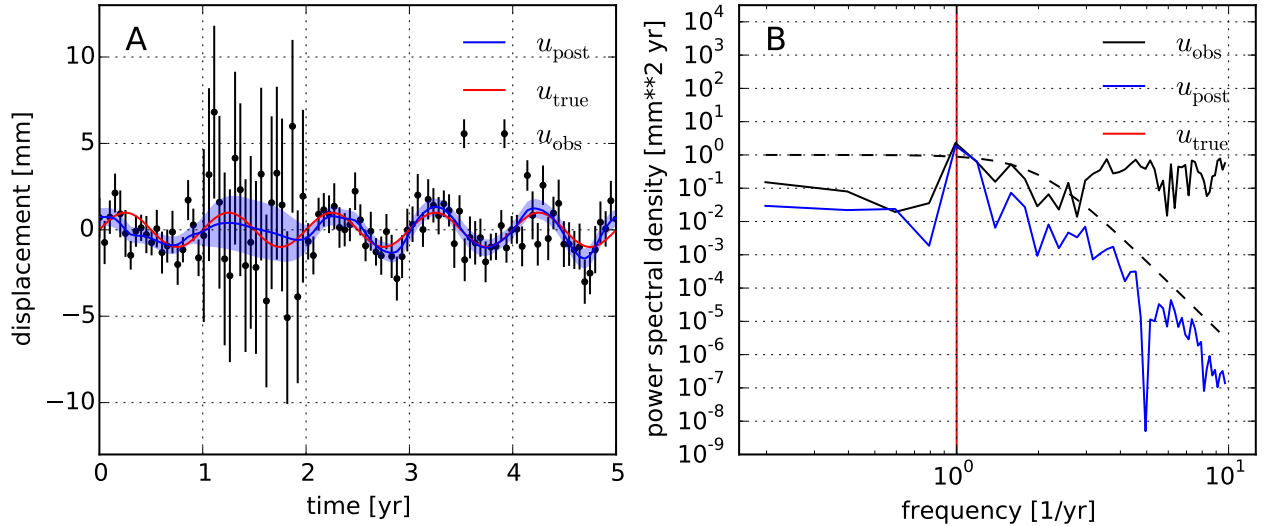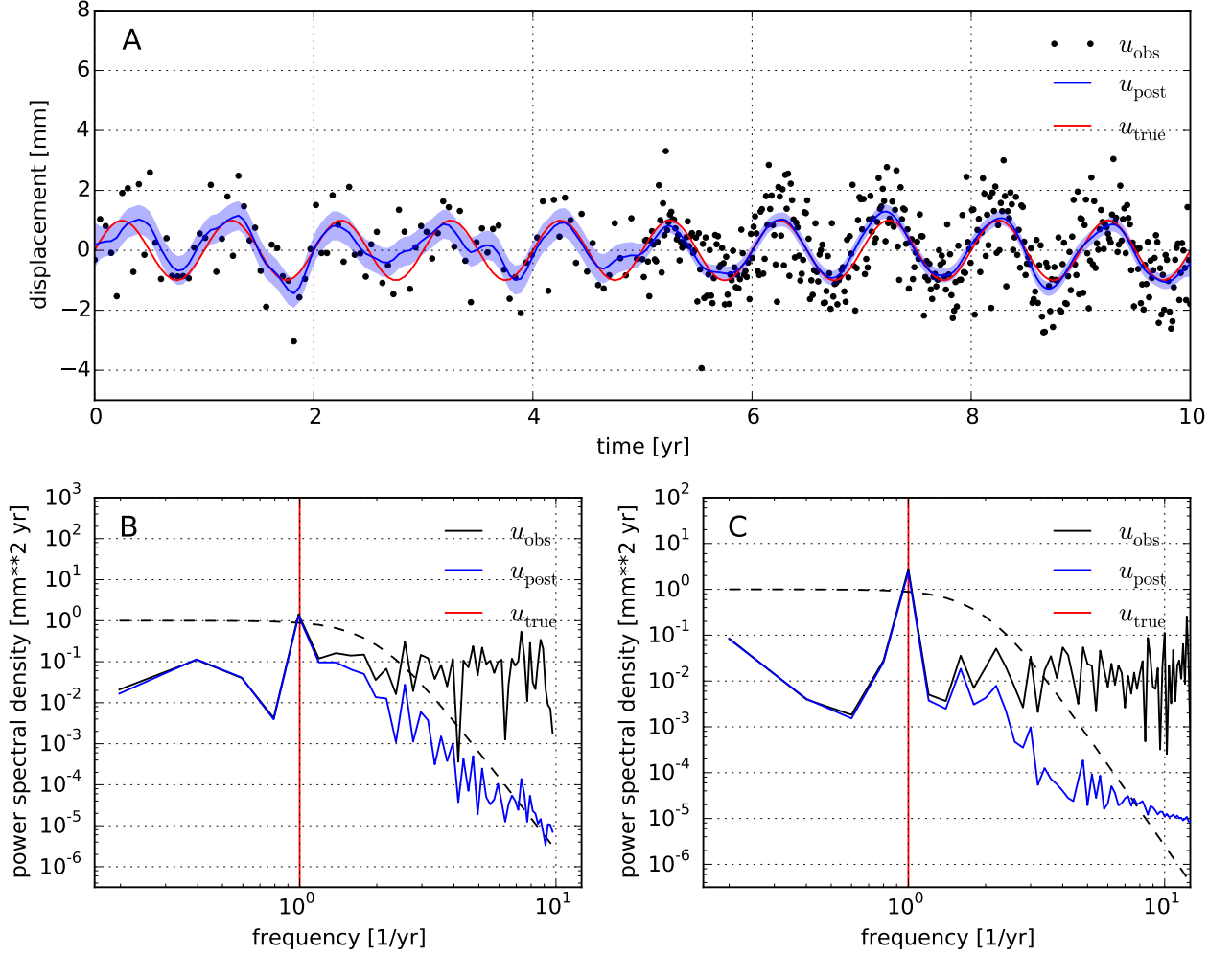
Figure 1: Panel A shows u_{obs} (black scatter points), u_{post} (blue line), and the true signal which we are trying to recover (red line). The lines on each scatter point and the light blue region show the one standard deviation uncertainty for the observations and filtered solution respectively. Panel 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 $\omega_c = 2$ 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 $\bar{u}_{post}$ would be exactly related by the black dashed line.

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

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

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 $\bar{u}_{post}$ 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 ~ 5 yr^{-1}, the rate that the spectral content decays is slower than what is predicted by the transfer function in eq.

3

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 $\omega_c = 2 \text{ yr}^{-1}$.

(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.

Smoothing in higher dimensions

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

$$\mathbf{L}_m u_{\text{prior}}(x_1, x_2) = q, \quad q \sim \mathcal{N}(0, \lambda^2) \tag{18}$$

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

$$\frac{\partial^{2m}}{\partial x_1^{2m}} + \frac{\partial^{2m}}{\partial x_2^{2m}}. \tag{19}$$

The corresponding covariance matrix is then

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

4

We again assume that the observation noise is uncorrelated with constant variance σ^2. Using the change of variables

$$\lambda = (2\pi\omega_c)^{2M}\sigma \tag{21}$$

we obtain the solution

$$u_{\text{smooth}}(x_1, x_2) = \left(\mathbf{I} + \left(\frac{1}{2\pi\omega_c}\right)^{4M}\mathbf{\Delta}_M^T\mathbf{\Delta}_M\right)^{-1}u_{\text{obs}}(x_1, x_2), \tag{22}$$

which in the two-dimensional frequency domain is

$$\hat{u}_{\text{smooth}}(\omega_1, \omega_2) = \frac{1}{1 + \left(\frac{\omega_1^{2M} + \omega_2^{2M}}{\omega_c^{2M}}\right)^2}\hat{u}_{\text{obs}}(\omega_1, \omega_2). \tag{23}$$

The transfer function in eq. (23) can once again be recognized as a low-pass filter. Namely, in the limit as $M \to \infty$ all the frequency pairs, (ω_1, ω_2), where $|\omega_1| > \omega_c$ or $|\omega_2| > \omega_c$ are removed. That is, the transfer function becomes a two-dimensional box centered at $(0, 0)$ with width $2\omega_c$. It is apparent from the transfer function why our prior model must be defined in terms of even order derivatives. If eq. (19) contained odd order derivatives, then all frequency pairs where $\omega_1 = -\omega_2$ would be unattenuated, leaving undesirable artifacts in the smoothed solution.

We have demonstrated with eq. (?? and 22) that eq. 5 can effectively act as a low pass filter with a judicious choice of $\mathbf{C}_{\text{prior}}$. One may then wonder why eq. (?? and 22) would ever be used to smooth data when it is far more efficient to perform the equivalent operation through the Fast Fourier Transform domain. The power in eq. (?? and 22) resides in the fact that they makes no assumption about when or where the observations have been made. In order to filter data through the Fast Fourier Transform, the observations must be evenly spaced on a regular grid, which is often not the case for geophysical data. Indeed, the focus of this paper is to smooth data which has been observed at scattered locations. We can use eq. (?? and 22) regardless of where the observations were made provided that we are able to create the differentiation matrix. We use the recently developed Radial Basis Function Finite-Difference method to form the differentiation matrices and we describe the procedure in the follow section.

References

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.

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

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

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

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