We discuss a method to filter observed data, u_{obs}, in a Bayesian framework, where u_{obs} can be spatial and/or temporal data and there is also no requirement that the observations be made over a regular grid. Our filtered estimate, u_{post}, incorporates u_{obs} and prior knowledge of the statistical properties of the underlying signal. In this sense, this work is closely tied to Kriging, Kalman filtering, and, more generally, Gaussian process regression (e.g. Rasmussen et al 2006). 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 filtered solution is itself a Gaussian process which has a distribution described by

$$u_{\text{post}} \sim \mathcal{N}(\bar{u}_{\text{post}}, \mathbf{C}_{\text{post}}). \tag{3}$$

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

$$\begin{aligned} \bar{u}_{\text{post}} &= (\mathbf{C}_{\text{obs}}^{-1} + \mathbf{C}_{\text{prior}}^{-1})^{-1}\mathbf{C}_{\text{obs}}^{-1}u_{\text{obs}} \\ \mathbf{C}_{\text{post}} &= (\mathbf{C}_{\text{obs}}^{-1} + \mathbf{C}_{\text{prior}}^{-1})^{-1}. \end{aligned} \tag{4}$$

$\mathbf{C}_{\text{obs}}$ 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. Below we discuss how a judicious choice of $\mathbf{C}_{\text{prior}}$ can turn eq. (4) into a low-pass or high-pass filter. This is first demonstrated for filtering one-dimensional data. The extension to N-dimensions follows naturally.

One-dimensional smoothing

When the signal is only assumed to covary in time, u_{prior} is commonly treated as Brownian motion (eg. Me Segall, McQuire, Murray, etc.). We can treat u_{prior} as Brownian motion by assuming that its velocity is white noise with constant variance λ^2. That is to say

$$\mathbf{D}_1 u_{\text{prior}} = q, \quad q \sim \mathcal{N}(0, \lambda^2), \tag{5}$$

where $\mathbf{D}_N$ is a differentiation matrix which estimates an N'th order derivative. It is also common to treat u_{prior} as integrated Brownian motion. In that case, we would just replace $\mathbf{D}_1$ with $\mathbf{D}_2$. In either case, the appropriate choice of the $\mathbf{C}_{\text{prior}}$ is

$$\mathbf{C}_{\text{prior}} = \lambda^2(\mathbf{D}_N^T\mathbf{D}_N)^{-1}. \tag{6}$$

The filtered solution is then described by

$$\begin{aligned} \bar{u}_{\text{post}} &= (\mathbf{C}_{\text{obs}}^{-1} + \frac{1}{\lambda^2}\mathbf{D}_N^T\mathbf{D}_N)^{-1}\mathbf{C}_{\text{obs}}^{-1}u_{\text{obs}} \\ \mathbf{C}_{\text{post}} &= (\mathbf{C}_{\text{obs}}^{-1} + \frac{1}{\lambda^2}\mathbf{D}_N^T\mathbf{D}_N)^{-1}. \end{aligned} \tag{7}$$

To gain an intuitive understanding of how eq. (6) controls the filtered solution, we transform $\bar{u}_{\text{post}}$ in eq. (7) to the frequency domain. In order to transform to the frequency domain, u_{obs} must have a constant sampling rate and ϵ and u_{prior} must be stationary stochastic processes. 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 circulant spectral differentiation matrix (e.g. Trefethen). Under a discrete Fourier transform, $\mathbf{D}_N$ has the properties

$$\mathcal{F}[\mathbf{D}_N f] = (2\pi i\omega)^N \hat{f} \tag{8}$$

and

$$\mathcal{F}[\mathbf{D}_N^T f] = (-2\pi i\omega)^N \hat{f}, \tag{9}$$

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

$$\hat{u}_{\text{post}}(\omega) = \frac{\frac{1}{\sigma^2}}{\frac{1}{\sigma^2} + \frac{(2\pi\omega)^{2N}}{\lambda^2}} \hat{u}_{\text{obs}}(\omega). \tag{10}$$

In practice, λ is a hyperparameter that must be chosen by the user and there are numerous ways that this could be accomplished. For example, one could use maximum likelihood methods, a trade-off curve, or simply vary λ until the filtered signal looks appropriate when compared to the observations. We make the change of variables

$$\lambda^2 = (2\pi\omega_c)^{2N}\sigma^2 \tag{11}$$

which changes the hyperparameter from λ to ω_c. The reason for this change of variables becomes apparent when we simplify eq. (10) to

$$\hat{u}_{\text{post}}(\omega) = \frac{1}{1 + \left(\frac{\omega}{\omega_c}\right)^{2N}} \hat{u}_{\text{obs}}(\omega). \tag{12}$$

We can recognize eq. (12) as an N'th order low-pass Butterworth filter with cut-off frequency ω_c. Our free parameter ω_c now has a far more tangible meaning and can be chosen objectively based on a prior understanding of the characteristic wavelength of the underlying signal.

In the limit as $N \to \infty$ eq. (12) 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 frequency response is tapered across ω_c, which ameliorates ringing in the filtered solution.

We have demonstrated that under certain conditions eq. (7) can be made equivalent to a low-pass filter. Of course, one could also filter the observed data through the Fast Fourier Transform to obtain an equivalent result but in a fraction of the time. 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. (7). The question is then whether eq. (7) still effectively acts as a low-pass filter when the idealized conditions are not met.

We answer this question with two demonstration. In the first demonstration we generate XXX samples of $u_{\text{obs}}(t)$ with a constant sampling rate over the interval $0 < t < 1$. Our samples of $u_{\text{obs}}(t)$ are white noise where the variances for each observation are randomly selected from a uniform distribution ranging from 0.1 to 10. In our second demonstration we generate XXX samples of $u_{\text{obs}}(t)$ at values of t which have a uniform random distribution ranging from 0 to 1. the values for $u_{\text{obs}}(t)$ are again white noise but we use a constant variance of 1. In both demonstrations, u_{prior} is treated as integrated Brownian motion (i.e. $N = 2$). We compute $\mathbf{D}_2$ with a first-order accurate finite difference scheme. We then compute $\bar{u}_{\text{post}}$ and $\mathbf{C}_{\text{post}}$ from eq. (7) and select λ^2 to be consistent with a cut-off frequency of XXX. For the first demonstration, where the variance is not constant, we modify eq. 11 so that λ^2 is in terms of a characteristic variance, $\bar{\sigma}^2$, so that

$$\lambda^2 = (2\pi\omega_c)^{2N}\bar{\sigma}^2 \tag{13}$$

where

$$\frac{1}{\bar{\sigma}^2} = \frac{1}{P}\text{tr}\left(\mathbf{C}_{\text{obs}}^{-1}\right) \tag{14}$$

and P is the number of observations. We show the filtered time series and covariances at select positions in figures X. In order to assess whether the filtered solution is acting as a low-pass filter we make use of the Wiener-Khinchin theorem, which states that the power spectral density and autocovariance of a stationary random process are Fourier transform pairs. We would then expect each row of the covariance matrix to resemble

$$c(t) = \mathcal{F}^{-1}\left[\left|\frac{1}{1 + \left(\frac{\omega}{\omega_c}\right)^{2N}}\right|^2\right] \tag{15}$$

after an appropriate time shift.

then compare the inverse Fourier transform of the squared frequency response in eq. 12 to the covariances for our demonstrations. If the two covariance functions resemble eachother then we can conclude that eq.

We assess whether eq. 7

$u_{\text{obs}}(t) \sim \mathcal{N}(0, 1)$. XXX samples of observations with uniform spacing from 0 to 1. Each o create XXX samples of white noise , each observations made with a uniform random distribution over the time interval (0,1). The variance of the observations are also randomly chosen from a uniform distribution over the interval (0.1,10). We use integrated Brownian motion for our prior model (i.e. $N = 1$), and D_1 is a first-order accurate finite difference operator (e.g. Fornberg).

The value of eq. (7) is that it offers a way to perform a low-pass filter when the conditions required for a discrete Fourier transform are not met. In the preceding Fourier analysis, we needed to make assumptions which are generally not true for real world applications. Namely, we assume This raises the question of whether eq. 7 will still effectively act as a low-pass filter when the idealized conditions are not met. In this example, we use observations made with a uniform random distribution over the time interval (0,1). The variance of the observations are also randomly chosen from a uniform distribution over the interval (0.1,10). We use integrated Brownian motion for our prior model (i.e. $N = 1$), and D_1 is a first-order accurate finite difference operator (e.g. Fornberg).

For example, if u_{obs} has an irregular sample rate or if the observation noise does not have a constant variance then we can no longer work in the frequency domain but we can still evaluate eq. (7). We can also still use eq. (11) as an intuitive guide for choosing λ. If the observation noise does not have a constant variance then σ in eq. (11) needs to be replaced with a characteristic variance, $\bar{\sigma^2}$, which we choose to be

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

where P is the number of observations. We provide an example to demonstrate that eq. (7) effectively acts as a low-pass filter when the data is aperiodically sampled and the observation noise does not have a constant variance. We compare each row in the correlation matrix for this example, determined from eq. 7, to the autocorrelation predicted function

illustrate filtering irregularly sampled data time non-constant variance.

We find that an appropiaret which we choose to be $\bar{\sigma} = 1/(1/\sigma)_{rms}$.

It is worth pointing out that the solution for u_{smooth} in eq. **??** is identical to the solution that would be obtained through Kalman filtering followed by smoothing. Kalman filtering and smoothing are recursive algorithms that operate along the time dimension. It is thus difficult to envision how a Kalman filter can be extended to smoothing data in higher dimensions. In contrast, extending eq. **??** to higher dimensions comes simply and naturally as we describe in the next section.

Smoothing in higher dimensions

We expand our discussion from one-dimensional smoothing to smoothing in two-dimensional space. While the remainder of this paper discusses smoothing in two-dimensions, the extension to smoothing in higher dimensions will prove to be trivial. We now allow u_{prior} to have non-zero covariance along multiple dimensions. In other words, we now treat u_{prior} as a random field rather than a stochastic process. The smoothed solution that we seek is still given by eq. 4 and our task is once again to find an appropriate choice of $\mathbf{C}_{\text{prior}}$. Below we consider the case where the M'th order Laplacian of u_{prior} is white noise. That is to say

$$\boldsymbol{\Delta}_M u_{\text{prior}}(x_1, x_2) = q, \quad q \sim \mathcal{N}(0, \lambda^2) \tag{17}$$

where

$$\boldsymbol{\Delta}_M = \frac{\partial^{2M}}{\partial x_1^{2M}} + \frac{\partial^{2M}}{\partial x_2^{2M}}. \tag{18}$$

The corresponding covariance matrix is then

$$\mathbf{C}_{\text{prior}} = \lambda^2 \left(\boldsymbol{\Delta}_M^T \boldsymbol{\Delta}_M\right)^{-1}. \tag{19}$$

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

$$\lambda = (2\pi\omega_c)^{2M}\sigma \tag{20}$$

we obtain the solution

$$u_{\text{smooth}}(x_1, x_2) = \left(\mathbf{I} + \left(\frac{1}{2\pi\omega_c}\right)^{4M} \boldsymbol{\Delta}_M^T \boldsymbol{\Delta}_M\right)^{-1} u_{\text{obs}}(x_1, x_2), \tag{21}$$

which in the two-dimensional frequency domain is

$$\hat{u}_{\text{smooth}}(\omega_1, \omega_2) = \frac{1}{1 + \left(\frac{\omega_1^{2M} + \omega_2^{2M}}{\omega_c^{2M}}\right)^2} \hat{u}_{\text{obs}}(\omega_1, \omega_2). \tag{22}$$

The transfer function in eq. (22) 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. (18) 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 21) that eq. 4 can effectively act as a low pass filter with a judicious choice of $\mathbf{C}_{\text{prior}}$. One may then wonder why eq. (**??** and 21) 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 21) 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 21) 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.