doi:10.26443/seismica.v2i2.388

Bayesian eikonal tomography using Gaussian processes

Jack B. Muir * [1]

[1]Department of Earth Sciences, University of Oxford, now at Fleet Space Technologies

Author contributions: *Conceptualization*: Jack Muir. *Software*: Jack Muir. *Formal Analysis*: Jack Muir. *Visualization*: Jack Muir. *Writing — Original draft*: Jack Muir.

Abstract Eikonal tomography has become a popular methodology for deriving phase velocity maps from surface wave phase delay measurements. Its high efficiency makes it popular for handling datasets deriving from large-N arrays, in particular in the ambient-noise tomography setting. However, the results of eikonal tomography are crucially dependent on the way in which phase delay measurements are predicted from data, a point which has not been thoroughly investigated. In this work, I provide a rigorous formulation for eikonal tomography using Gaussian processes (GPs) to smooth observed phase delay measurements, including uncertainties. GPs allow the posterior phase delay gradient to be analytically derived. From the phase delay gradient, an excellent approximate solution for phase velocities can be obtained using the saddlepoint method. The result is a full Bayesian posterior distribution for phase velocities of surface waves, incorporating the nonlinear wavefront bending inherent in eikonal tomography, with no sampling required. On studying these posterior distributions, the outcomes of these analyses imply that the uncertainties reported for eikonal tomography are often underestimated.

Non-technical summary Eikonal tomography is an imaging method that uses slight variations between seismic waves trapped at the surface of the Earth to infer information about the properties beneath the surface. To be able to perform the best possible eikonal tomography, we need to be able to predict in between measurements of these variations at different seismic recording stations as best we can. Furthermore, end-users of seismic tomography require information about the uncertainty of the images. In this paper, I perform this prediction using Gaussian processes (GPs), a method with particularly nice mathematical properties. The GP prediction results in robust uncertainty measurements for our imaging problem without many of the computational difficulties associated with other uncertainty quantification methods.

Production Editor: Gareth Funning
Handling Editor: Suzan van der Lee
Copy & Layout Editor: Kirsty Bayliss

Received: February 20, 2023
Accepted: November 6, 2023
Published: December 12, 2023
Updated: January 24, 2026

1 Introduction

Surface wave tomography is a cornerstone imaging technique for the investigation of the crust and upper mantle. However, due to the significant non-planarity of scattered surface waves, interpretation of surface wave data is not straightforward (e.g., Wielandt, 1993). Despite this issue, the increasing proliferation of dense seismic arrays, combined with the advent of ambient-noise correlation methods, has motivated intense study into surface wave tomographic techniques. To ameliorate the great cost of nonlinear ray tracing for large inverse problems, a large part of this study has focused on methods that derive surface wave properties from only local information contained in the wavefield. Beginning with a wavefield perturbation approach (e.g., Friederich et al., 1994; Friederich and Wielandt, 1995; Pollitz, 2008), theoretical efforts in local surface wave inversion have since concentrated on direct measurement of wavefield derivatives (e.g Lin et al., 2009; Lin and Ritzwoller, 2011; de Ridder and Biondi, 2015; de Ridder and Maddison, 2018). Likely owing to its simplicity, the most popular extant method is eikonal tomography (Lin et al., 2009), which relies on the determination of the wavefield phase gradient across an entire local or regional array. For a single surface wave mode propagating with phase velocity C_p, frequency ω, phase delay T and amplitude A, the Helmholtz equation implies that (Tromp and Dahlen, 1993)

$$\frac{1}{C_p^2} = |\nabla T|^2 - \frac{\nabla^2 A}{\omega^2 A}. \tag{1}$$

*Corresponding author: jack.muir@fleet.space

Simplifying this relationship under the assumption that the frequency of the wave is large compared to perturbations in the wave amplitude gives us the eikonal equation:

$$C_p = \frac{1}{|\nabla T|}. \tag{2}$$

Eikonal tomography uses Equation 2 to directly infer local phase velocity from local phase gradient. A distinction compared to local gradiometry is that calculation of the phase gradient is performed simultaneously for all desired locations by fitting a delay curve across an array, rather than by local analysis of sub-arrays (Langston, 2007a, e.g.,). The assumption that the wavefront is smooth relative to frequency is strong, but the difficulty associated with measuring wavefront curvature accurately has ensured that eikonal tomography remains a central technique in array analysis. Application of eikonal tomography in practice has typically resulted in images comparable to other tomographic methods and Helmholtz tomography (which uses Equation 1 directly), especially when results are averaged azimuthally (Bodin and Maupin, 2008; Lin et al., 2009; Lehujeur and Chevrot, 2020). While the typical use case of eikonal tomography is surface-wave phase-velocity inversion (a 2D problem), other potential use cases of eikonal tomography could include 1D linear inversions along DAS arrays (e.g. Yang et al., 2022) or 3D inversions of first arrivals within mine arrays (e.g. Mandic et al., 2018), so interpolation schemes that work well in arbitrary dimensions are useful for eikonal tomography workflows.

In this work, I employ Gaussian process theory (Rasmussen and Williams, 2006) to derive semi-analytic closed-form approximations for the posterior distribution of eikonal-equation-based phase-velocity measurements using the saddlepoint method (Butler, 2007). In this case, semi-analytic means that the posterior approximations have a single parameter that must be solved using constrained minimization techniques — no Monte Carlo methods need be used. As a result, the approximate posterior can be calculated very quickly. As an intermediate result, I derive fully analytic posteriors for the gradient of phase delay. The delay gradient posteriors can be sampled using standard multivariate normal random number generators, which provides an efficient way to compute arbitrary statistics of the GP posterior when the semi-analytic approximations are difficult to obtain.

2 Eikonal tomography from derivatives of Gaussian processes

The least well-defined problem in eikonal tomography is how to go from point measurements of phase delay to the phase delay gradient map (Lin et al., 2009). It is in this process that the practitioner has the greatest control over the resulting phase velocity map; intuitively, we can immediately see that over-smoothing the map will result in a measurement of C_p that is too large; conversely, maps that are too rough will result in too small C_p. Past studies have typically employed splines (either in tension (e.g., Lin et al., 2009; Lin and Ritzwoller, 2011) or smoothing (Chevrot and Lehujeur, 2022)) to perform prediction. The spline framework is a robust general interpolation or smoothing method, however in its basic formulation it gives a single maximum-likelihood estimate of the prediction, with no associated uncertainty information. As I later show in the paper, phase velocities derived by eikonal tomography are biased due to the presence of uncertainty, so it is important to understand the scale of uncertainties when creating eikonal tomography maps.

This study aims to place the problem of estimating an optimal phase gradient map on a robust Bayesian footing, where all assumptions are explicit, adjustable, and optimizable in the face of the data. In this study, the problem of predicting phase delay measurements is posed as a Gaussian process (GP) regression (often referred to as Kriging in geostatistical literature) — we will see that this framework meets the desiderata for estimating phase gradients. GPs are a particular framework for defining distributions over function spaces (Rasmussen and Williams, 2006). GPs have the property that any finite collection of points sampled from them will have a multivariate Gaussian distribution. A GP is defined by a mean function $m(x)$ and covariance function $k(x, x')$, which generate the mean and covariance matrix of a finite collection of points drawn from the GP. Concretely, for any collection of points $(x_1, x_2, ...x_n)$ and associated function values $(d_1, d_2, ...d_n)$, the GP model assumes that

$$\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix} \sim N\left(\begin{bmatrix} m(x_1) \\ m(x_2) \\ \vdots \\ m(x_n) \end{bmatrix}, \begin{bmatrix} k(x_1,x_1) & k(x_1,x_2) & \dots & k(x_1,x_n) \\ k(x_2,x_1) & k(x_2,x_2) & \dots & k(x_2,x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n,x_1) & k(x_n,x_2) & \dots & k(x_n,x_n) \end{bmatrix}, \right) \tag{3}$$

where $N(\mu, \Sigma)$ is a multivariate Gaussian with mean vector μ and covariance matrix Σ. In the context of regression, this leads to a powerful result — if we assume a GP prior for an unknown function, and we then observe data with a Gaussian likelihood, the posterior distribution for the unknown function will also be a GP. Thus, GPs fully generalize finite linear regression and Gaussian inverse problems to the function space setting (Valentine and Sambridge,

2020a,b). As differentiation is a linear operation, derivatives of GPs are again also GPs. We will use these properties to derive closed-form posterior distributions for the derivatives of observed data under a GP prior. While the motivating example is eikonal tomography, these techniques are applicable to regression problems generally. Derivatives of GPs have long been used in the dynamical control community (e.g. Solak et al., 2002; Rasmussen, 2003). Closer in spirit to seismology, GP derivatives have also been applied to the identification of geodetic transients (Hines and Hetland, 2018). The presentation described here is generalized from McHutchon (2014).

In this manuscript, bold font refers to collections of observed data and capitals to matrices. Boldfont capitals are therefore collections of n data in m coordinates and will have dimensions $n \times m$. Coordinates (i.e., x) may be vector quantities but will not be boldfont. To begin, assume that there are measurements $(\boldsymbol{X}, \boldsymbol{d})$ of the observed phase delay $\boldsymbol{d}$ at points $\boldsymbol{X}$. Assume that the data $\boldsymbol{d}$ are noisy; for the purposes of exposition this is taken to be identically distributed Gaussian noise η with the distribution $N(0, \sigma)$, but arbitrary multivariate Gaussian noise distributions with data covariance C_D are also easily handled by GP theory. This implies that there is an unknown true phase delay field $T(x)$ with

$$\boldsymbol{d} = T(\boldsymbol{X}) + \eta. \tag{4}$$

The objective of eikonal tomography is to know the field $T(x)$ so that we can differentiate it and get C_p. I assume that

$$T(x) = T_0(x) + r(x) \tag{5}$$

where r is a zero-mean GP and $T_0(x)$ is a reference phase delay field, for example for a laterally homogeneous medium. Therefore, $T(x)$ is a GP with mean $T_0(x)$.

$$T(x) \sim GP(T_0(x), k(x, x')), \tag{6}$$

where $k(x, x')$ is the assumed covariance function. It should be noted that r is a distribution over functions; it incorporates uncertainty due to errors in the observed data and the effect of heterogeneities on the travel time function. r in effect models the residuals between the reference model T_0 and the observed data. For the examples in this work, I will use a squared-exponential kernel with independent length scales in each dimension for the covariance function:

$$k(x, x') = a^2 \exp\left(-\sum_{i=1}^{m} \frac{(x_i - x_i')^2}{2l_i^2}\right). \tag{7}$$

This covariance function promotes very smooth fields with characteristic amplitude a (it is infinitely differentiable), and provides a degree of flexibility that improves regression performance in the face of inhomogeneous data observation and complex travel time fields due to the independent length scales l_i. This covariance function allows the observed data points to "talk" to one another and build a smooth underlying interpolation field that captures variations in travel time due to structural heterogeneity. The hyperparameters $a, l_1, \ldots$ etc. are optimized by minimizing the negative log marginal likelihood of the GP model given the observed data — this is further discussed in Section 2.2.

I also assume that $T_0(x) = s_0|x|$ for a fixed reference slowness s_0. Let $K_{\boldsymbol{X}\boldsymbol{X}'}$ be the matrix of evaluating k with rows given by $\boldsymbol{X}$ and columns by $\boldsymbol{X}'$. The fundamental idea of GP regression is that, given this problem setup, then the observed data $\boldsymbol{d}$ and the predicted data $T(\boldsymbol{X}')$ has the joint multivariate Gaussian distribution

$$\begin{bmatrix} \boldsymbol{d} \\ T(\boldsymbol{X}') \end{bmatrix} \sim N\left(\begin{bmatrix} T_0(\boldsymbol{X}) \\ T_0(\boldsymbol{X}') \end{bmatrix}, \begin{bmatrix} K_{\boldsymbol{X}\boldsymbol{X}} + \sigma^2 I & K_{\boldsymbol{X}\boldsymbol{X}'} \\ K_{\boldsymbol{X}'\boldsymbol{X}} & K_{\boldsymbol{X}'\boldsymbol{X}'} \end{bmatrix}.\right) \tag{8}$$

By conditioning $T(\boldsymbol{X}')$ on the observed data $\boldsymbol{d}$ (i.e. finding the distribution of $T(\boldsymbol{X}')$ given fixed $\boldsymbol{d}$) we have (Rasmussen and Williams, 2006)

$$T(\boldsymbol{X}')|\boldsymbol{d} \sim N(T_0(\boldsymbol{X}') + K_{\boldsymbol{X}'\boldsymbol{X}}(K_{\boldsymbol{X}\boldsymbol{X}} + \sigma^2 I)^{-1}(\boldsymbol{d} - T_0(\boldsymbol{X})), K_{\boldsymbol{X}'\boldsymbol{X}'} - K_{\boldsymbol{X}'\boldsymbol{X}}(K_{\boldsymbol{X}\boldsymbol{X}} + \sigma^2 I)^{-1} K_{\boldsymbol{X}\boldsymbol{X}'}). \tag{9}$$

Note that data error models with Gaussian covariance just require replacing $\sigma^2 I$ with C_D.

Figure 1 shows an example application of GP regression for obtaining $T(x)|\boldsymbol{d}$, with comparison to the approach based on regression using splines (e.g., Lin et al., 2009; Lin and Ritzwoller, 2011) — in this case, using smoothing splines (e.g., Chevrot and Lehujeur, 2022). This example emulates a typical local surface wave application, using 100 data points uniformly distributed within the inversion region with 0.2 s added Gaussian noise. The squared slowness is obtained using the method of manufactured solutions from the eikonal equation to avoid any errors in the simulated data, and the synthetic phase delay field is strongly perturbed away from the reference model to highlight differences between the GP and spline based methods. The GP mean and standard deviation are given analytically, and show substantial differences with the smoothing spline fit — here, the spline smoothing parameter

is automatically set by the FitPack routine (Dierckx, 1993). In comparison to the GP, the spline performs less well, especially in areas of data gaps. Figure 2 compares the GP reconstruction with the true values of the phase delay map. The GP mean closely fits the true values, although the level of uncertainty becomes quite substantial near the edges of the domain.

I can now calculate expectation values (the mean of the probability distribution) for the derivatives; note that from now on I implicitly condition on $\boldsymbol{d}$ but will not write it out for ease of notation, unless it seems particularly germane to do so. Since differentiation is a linear operation, and linear operations acting on normal distributions result in normal distributions, the components of ∇T must also be normally distributed, and are completely specified by their mean and covariance. The collection of means for component i are immediately given by recognizing that as the expectation operator is also linear, it commutes with the derivative operator:

$$\begin{aligned}\mathbb{E}\left[\frac{\partial T(\boldsymbol{X}')}{\partial x_i'}\right] &= \frac{\partial \mathbb{E}\left[T(\boldsymbol{X}')\right]}{\partial x_i'} \\ &= \frac{\partial T_0(\boldsymbol{X}')}{\partial x_i'} + \frac{\partial K_{\boldsymbol{X}'\boldsymbol{X}}}{\partial x_i'}(K_{\boldsymbol{X}\boldsymbol{X}} + \sigma^2 I)^{-1}(\boldsymbol{d} - T_0(\boldsymbol{X})).\end{aligned} \tag{10}$$

Note that the mean value of the derivatives are calculated independently for each dimension; however as we will see they do have covariance between output points and between dimensions. For the covariance, consider $n \times n$ blocks of the covariance matrix of size $nd \times nd$ where d is the dimension and n is the number of output points. Note that I choose to order the hierarchy of the covariance matrix first by derivative coordinate, and second by data point index, as it makes the notation more convenient. As the covariance is bilinear,

$$\mathrm{Cov}\left(\frac{\partial T(\boldsymbol{X}')}{\partial x_i'}, \frac{\partial T(\boldsymbol{X}'')}{\partial x_j''}\right) = \frac{\partial^2 \mathrm{Cov}(T(\boldsymbol{X}'), T(\boldsymbol{X}''))}{\partial x_i' \partial x_j''} \tag{11}$$

where I introduce the dummy variable x'' to represent the second argument in the covariance ($\boldsymbol{X}' = \boldsymbol{X}''$, but we want to formally differentiate in respect to the second slot only when using x''). Continuing on,

$$\begin{aligned}\frac{\partial^2 \mathrm{Cov}(T(\boldsymbol{X}'), T(\boldsymbol{X}''))}{\partial x_i' \partial x_j''} &= \frac{\partial^2 \left(K_{\boldsymbol{X}'\boldsymbol{X}''} - K_{\boldsymbol{X}'\boldsymbol{X}}(K_{\boldsymbol{X}\boldsymbol{X}} + \sigma^2 I)^{-1} K_{\boldsymbol{X}\boldsymbol{X}''}\right)}{\partial x_i' \partial x_j''} \\ &= \frac{\partial^2 K_{\boldsymbol{X}'\boldsymbol{X}''}}{\partial x_i' \partial x_j''} - \frac{\partial K_{\boldsymbol{X}'\boldsymbol{X}}}{\partial x_i'}(K_{\boldsymbol{X}\boldsymbol{X}} + \sigma^2 I)^{-1}\frac{\partial K_{\boldsymbol{X}\boldsymbol{X}''}}{\partial x_j''}.\end{aligned} \tag{12}$$

So that I can compress the notation somewhat, let us define $\hat{K}_{\boldsymbol{X}\boldsymbol{X}} = K_{\boldsymbol{X}\boldsymbol{X}} + \sigma^2 I$ and $\Delta\boldsymbol{d} = \boldsymbol{d} - T_0(\boldsymbol{X})$. For the 2D case (noting that other dimensions immediately generalize), the conditional posterior is a multivariate Gaussian with mean given by Equation 10 and covariance given by Equation 12:

$$\begin{aligned}\nabla T(\boldsymbol{X}')|\boldsymbol{d} &= \begin{bmatrix}\frac{\partial T(\boldsymbol{X}')}{\partial x'} \\ \frac{\partial T(\boldsymbol{X}')}{\partial y'}\end{bmatrix}|\boldsymbol{d} \\ &\sim N\left(\begin{bmatrix}\frac{\partial T_0(\boldsymbol{X}')}{\partial x'} + \frac{\partial K_{\boldsymbol{X}'\boldsymbol{X}}}{\partial x'}\hat{K}_{\boldsymbol{X}\boldsymbol{X}}^{-1}\Delta\boldsymbol{d} \\ \frac{\partial T_0(\boldsymbol{X}')}{\partial y'} + \frac{\partial K_{\boldsymbol{X}'\boldsymbol{X}}}{\partial y'}\hat{K}_{\boldsymbol{X}\boldsymbol{X}}^{-1}\Delta\boldsymbol{d}\end{bmatrix}, \begin{bmatrix}\frac{\partial^2 K_{\boldsymbol{X}'\boldsymbol{X}''}}{\partial x'\partial x''} - \frac{\partial K_{\boldsymbol{X}'\boldsymbol{X}}}{\partial x'}\hat{K}_{\boldsymbol{X}\boldsymbol{X}}^{-1}\frac{\partial K_{\boldsymbol{X}\boldsymbol{X}''}}{\partial x''} & \frac{\partial^2 K_{\boldsymbol{X}'\boldsymbol{X}''}}{\partial x'\partial y''} - \frac{\partial K_{\boldsymbol{X}'\boldsymbol{X}}}{\partial x'}\hat{K}_{\boldsymbol{X}\boldsymbol{X}}^{-1}\frac{\partial K_{\boldsymbol{X}\boldsymbol{X}''}}{\partial y''} \\ \frac{\partial^2 K_{\boldsymbol{X}'\boldsymbol{X}''}}{\partial y'\partial x''} - \frac{\partial K_{\boldsymbol{X}'\boldsymbol{X}}}{\partial y'}\hat{K}_{\boldsymbol{X}\boldsymbol{X}}^{-1}\frac{\partial K_{\boldsymbol{X}\boldsymbol{X}''}}{\partial x''} & \frac{\partial^2 K_{\boldsymbol{X}'\boldsymbol{X}''}}{\partial y'\partial y''} - \frac{\partial K_{\boldsymbol{X}'\boldsymbol{X}}}{\partial y'}\hat{K}_{\boldsymbol{X}\boldsymbol{X}}^{-1}\frac{\partial K_{\boldsymbol{X}\boldsymbol{X}''}}{\partial y''}\end{bmatrix}\right),\end{aligned} \tag{13}$$

which is an exact distribution for the derivatives evaluated at $\boldsymbol{X}'$. Figure 3 shows the mean and covariance structure for the derivatives at two test points calculated using the above theory, compared to the true derivative of the phase delay, and finite-difference estimates computed using random draws of the GP estimate of the phase delay (i.e., Monte-Carlo finite-difference derivatives). Both the analytic and Monte-Carlo results closely agree with each other and with the true values for the derivatives. In Figure 4, I use the multivariate normal posterior for the derivatives to generate samples of the posterior for the squared slowness and compare it against the predictions from the smoothing spline. The GP posterior is in this case more accurate than the spline result, and also delivers uncertainty information.

Unfortunately, it turns out that this is as far as it is possible to go with exact distributions, as the velocity is a nonlinear function of the gradients in eikonal tomography. Thankfully, however, there is well-developed theory for approximating quadratic forms of normal random variables, and as $\frac{1}{C_p^2} = (\nabla T)^2$, which is a quadratic form of a normal random variable, it may be possible to try for a good approximation to the velocity. Before deriving one, however, there are two important issues to investigate — setting hyperparameters, and closed forms for the expectation value of velocity.

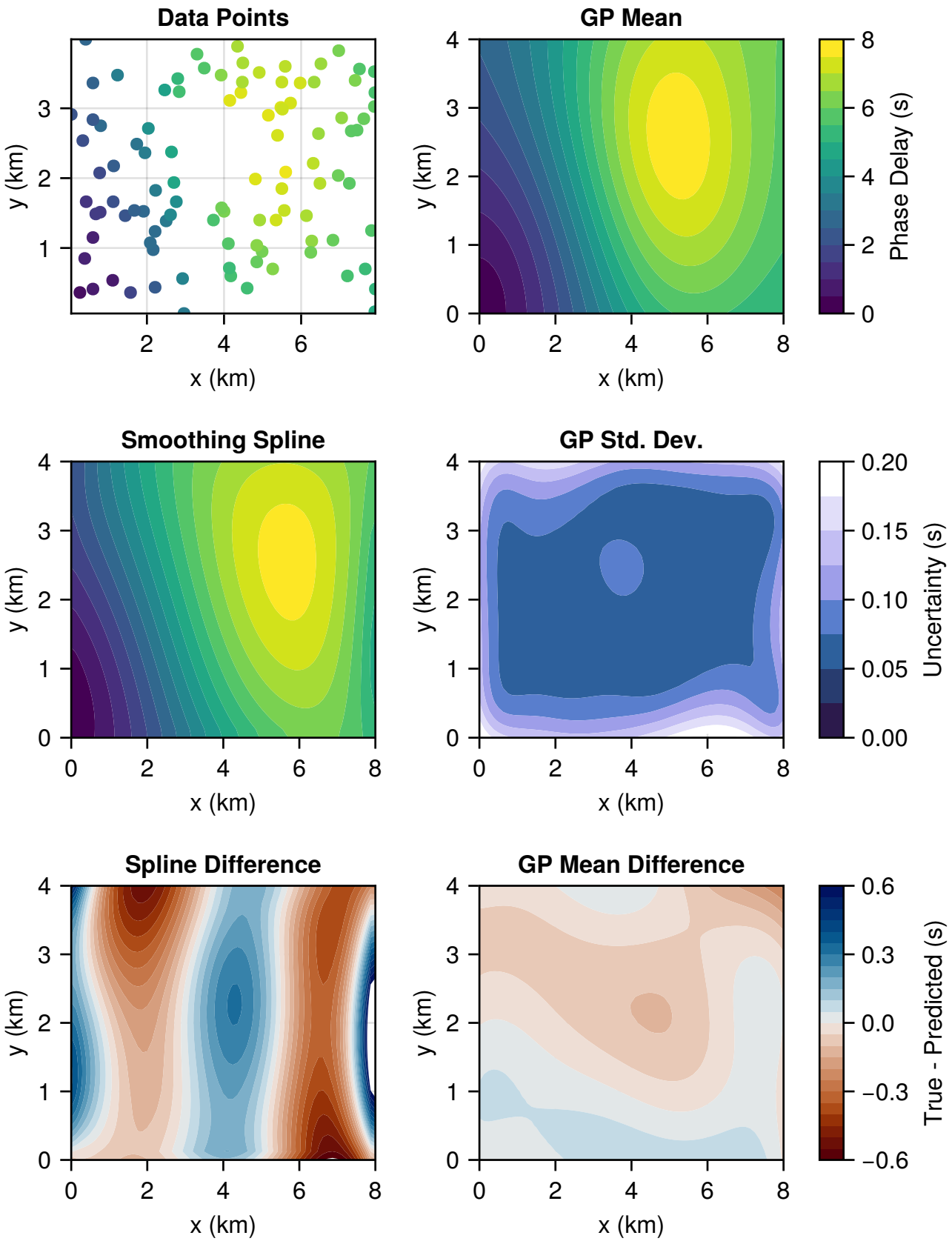

Figure 1 Comparison of the GP posterior (showing mean and point-wise standard deviation) of the estimated phase delay with a smoothing-spline based solution for an example phase delay data set with 100 randomly distributed points and 0.2 s Gaussian noise. The data is generated using the method of manufactured solutions, assuming a seismic source at $(0, 0)$. There are notable differences in the estimated phase delay, especially where there are gaps in the data coverage. The difference plots show the difference between the true phase delay field and the spline solution or the GP mean respectively. The colouring of the difference plots is arranged according to the usual seismic convention of blue being a fast and red being slow; in this case blue means that the predicted arrival is fast compared to the truth and vice versa.

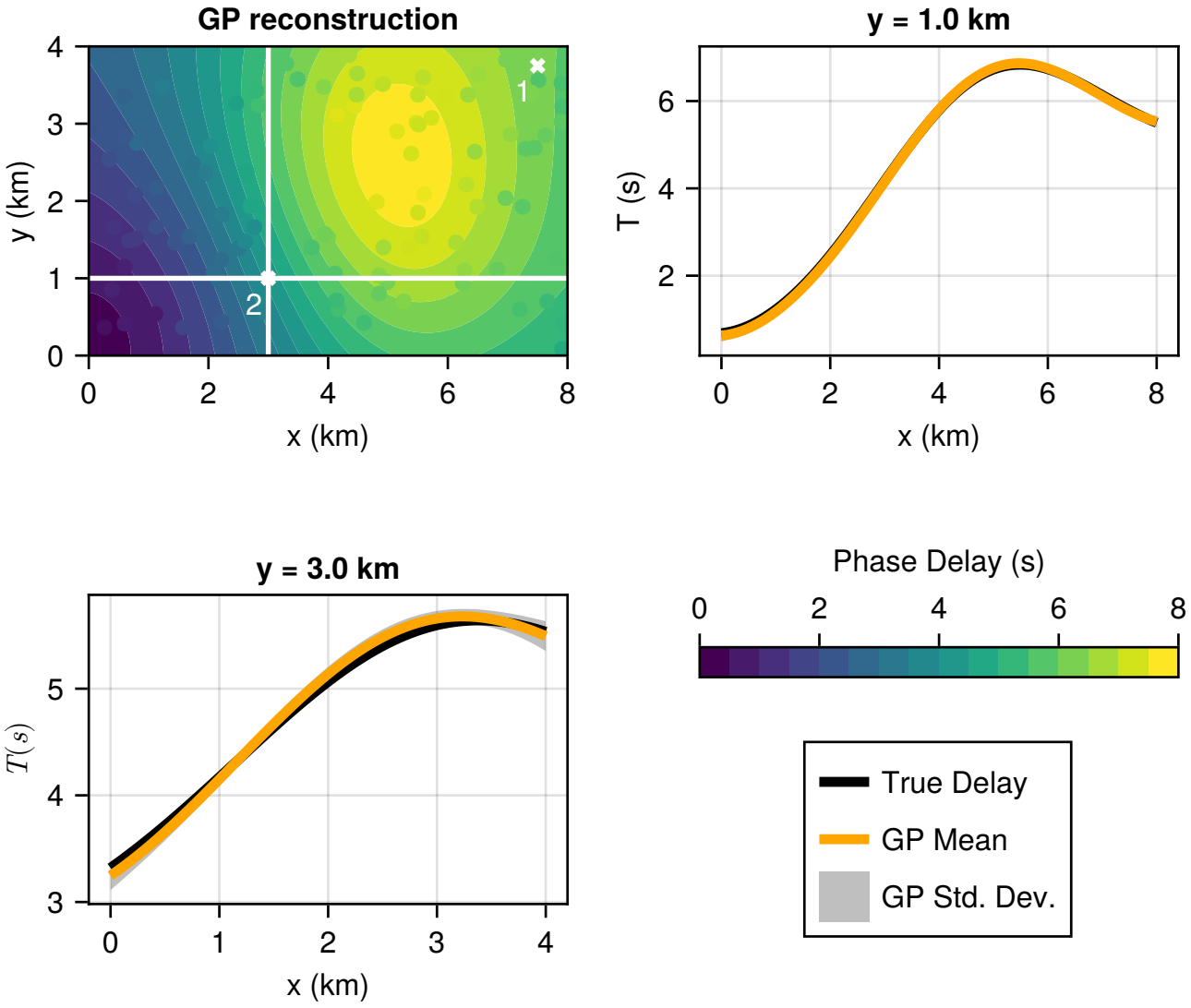

Figure 2 Cross-sections through the GP reconstruction of Figure 1, showing the true phase delay (black), GP mean (orange) and standard deviation (grey). The GP reconstruction is overlaid with the noisy observed delay values. The GP posterior closely follows the true phase delay curve, with substantially higher uncertainty near the edges of the domain, even before extrapolation. The test points used later in Figure 3 are shown by white crosses.

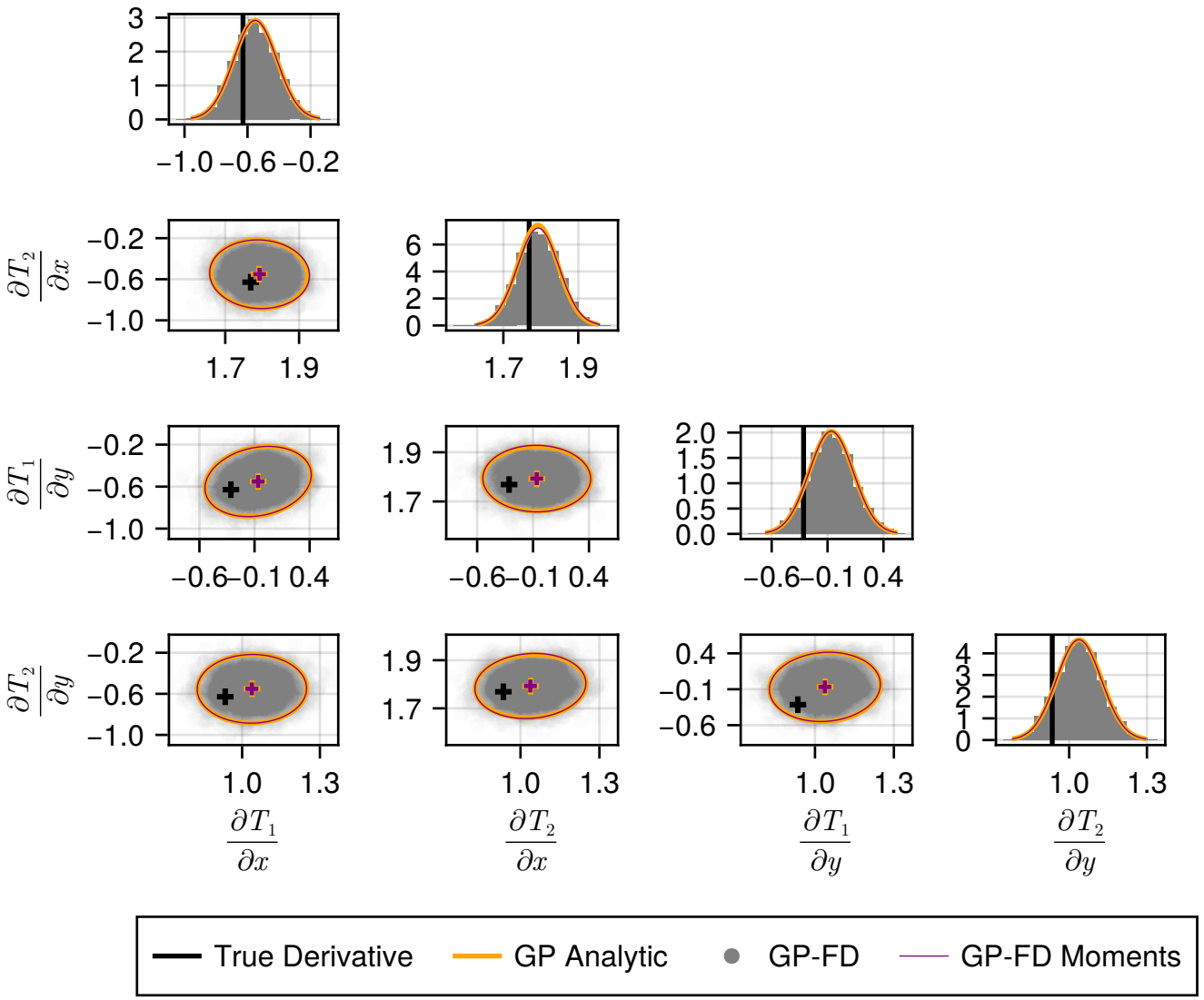

Figure 3 Corner plot showing the covariance of derivatives at two test points, and their individual histograms. The test points are T_1 at (7.5,3.75) (a near-edge point), and T_2 at (3.0, 1.0) (a more centered point). Black crosses and lines show the true value of the derivatives. Orange lines show the analytical GP based solutions derived in this paper, with ellipses drawn at the 95% credible level and crosses showing the mean. Grey circles and histograms show finite-difference (FD) based derivatives using Monte-Carlo samples of the GP posterior for phase delay, and red crosses and ellipses show the mean and estimated covariance at 95% confidence from the FD draws. For the 1D histograms, the y-axis represents the value of the PDF.

2.1 A realistic example — Rayleigh wave phase velocities near Ridgecrest, CA

Having investigated some of the features of the GP interpolation method for eikonal tomography using a synthetic with large amplitude perturbations, I now perform the same investigation for a realistic problem setup. I simulated phase velocity data, beginning with the V_P and V_S model of White et al. (2021) for the region immediately surrounding the fault traces of the Ridgecrest, CA July 2019 earthquake sequence. I converted White's model to UTM zone 11 coordinates within the area between 324–587 km easting, 3820–4094 km northing, and used the Nafe-Drake empirical relationship to obtain ρ from V_P (Brocher, 2005), and then the fundamental-mode Rayleigh-wave phase velocity at 30 s period was calculated for each point using surfdisp96 (Herrmann, 2013). The travel time field was calculated from the southwest corner of the domain (324 km easting, 3820 km northing; UTM zone 11) using the factored-eikonal fast-marching method (Treister and Haber, 2016). I interpolated the travel time field to the 154 station locations used in the generation of the White et al. (2021) model within the simulation domain box. Finally, I added Gaussian random noise with 0.1 s standard deviation to the simulated travel times to create the dataset.

The salient points of difference between this experiment and the previous ones are: firstly, the strength of the velocity perturbations is much smaller, resulting in smaller travel time effects; secondly, the distribution of stations is highly non-uniform resulting in variable spatial resolution; and thirdly, the travel time is calculated using a numerical

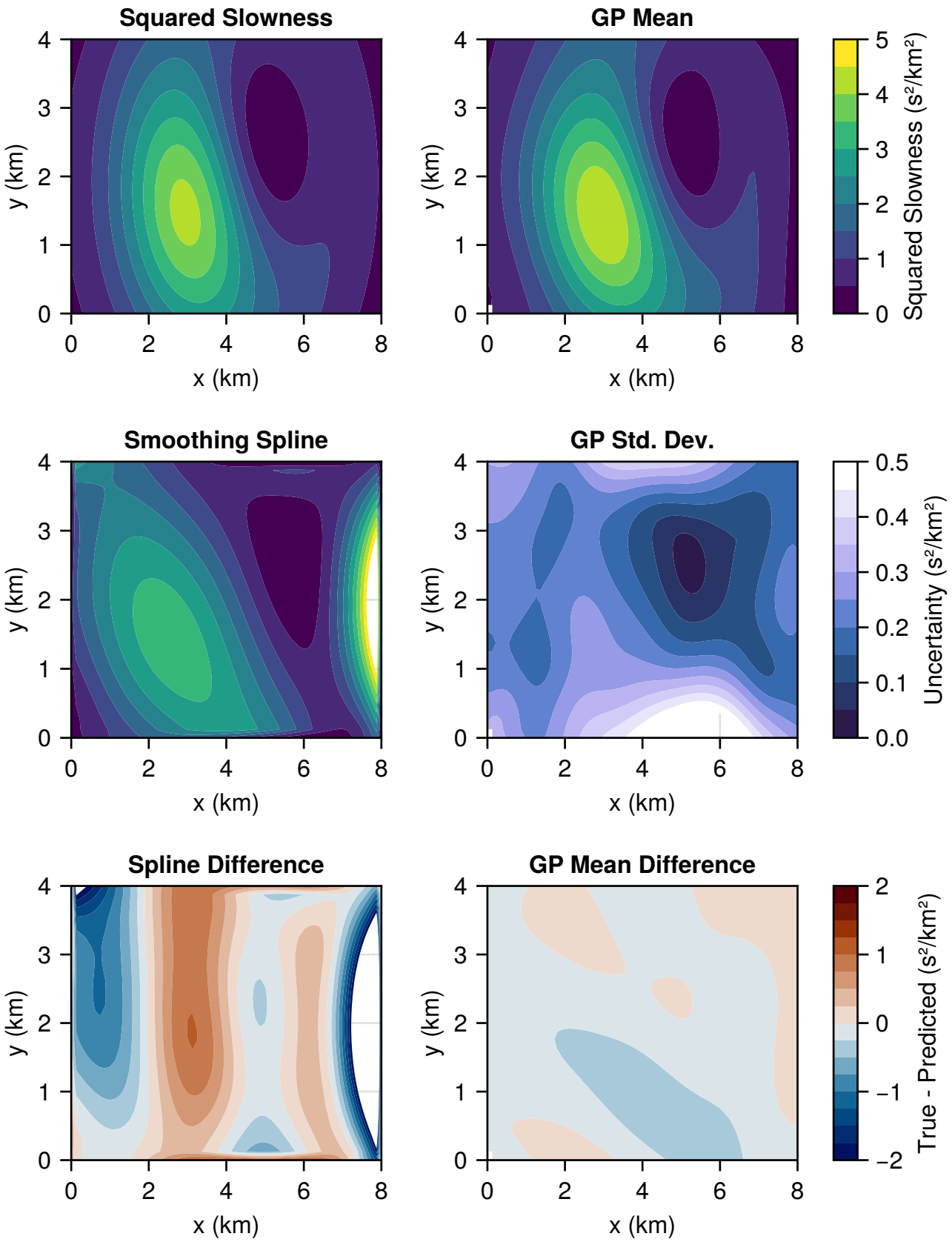

Figure 4 Comparison of the true squared slowness against results calculated using a squared-exponential Gaussian process with tuned hyperparameters. The GP mean and standard deviation are calculated by drawing 100,000 predicted travel time gradients. The spline squared slowness has been calculated using 5^{th} order centred finite differences. The GP result has a mean closer to the truth, and additionally adds uncertainty information, when compared to the smoothing spline. The colouring of the difference plots is arranged according to the usual seismic convention of blue being fast and red being slow; in this case blue means that the predicted slowness is smaller compared to the truth and vice versa; note that this induces a colour flip compared to Figure 1.

method, and so may contain minor errors (although these will be mitigated by using the high-accuracy algorithm of Treister and Haber (2016)). Figures 5, 6 and 7 are the equivalents of Figures 1, 2 and 4, respectively. Even in this substantially different setting, the GP based interpolation performs better than the spline. In particular, the spline based method appears to have trouble with the edges of the domain when the density of stations is highly non-uniform, resulting in artifacts near the southwest corner in the derived phase velocity field, whereas the GP based method does not suffer from these issues.

2.2 Finding good values for GP hyperparameters

The hyperparameters of the GP may be optimized by maximizing the log marginal likelihood of observations, where the marginalization is performed over the unknown function values $T(\boldsymbol{X})$ (Rasmussen and Williams, 2006). This gives the type-II maximum likelihood estimate; the hyperparameters have a point-estimate, whereas the function values have a full posterior distribution given that point-estimate. The log marginal likelihood for GP regression is given by

$$\log p(\boldsymbol{d}|\theta, \boldsymbol{X}) = -\frac{1}{2}\Delta\boldsymbol{d}^T \hat{K}_{\boldsymbol{XX}}^{-1}(\theta)\Delta\boldsymbol{d} - \frac{1}{2}\log|\hat{K}_{\boldsymbol{XX}}(\theta)| - \frac{n}{2}\log(2\pi), \tag{14}$$

where the covariance matrix $\hat{K}_{\boldsymbol{XX}}(\theta)$ is treated as a function of the hyperparameters θ, and n is the number of data. Intuitively, the log marginal likelihood parsimoniously balances data misfit (the first term) with the level of uncertainty (the second term). For a 2D squared-exponential kernel with independent length scales, independent Gaussian data noise, and a laterally homogeneous medium as a reference model, the hyperparameters are $\theta = (a, l_1, l_2, \sigma, s_0)$.

2.3 A special exact case for eikonal tomography: The expectation value of squared slowness given normally distributed derivatives

Consider without loss of generality a 2D case. The squared slowness is given by $1/C_p^2 = \left(\frac{\partial T}{\partial x}\right)^2 + \left(\frac{\partial T}{\partial y}\right)^2 = T_x^2 + T_y^2$. Assume the phase gradient is given by a multivariate Gaussian random variable

$$\begin{aligned} \nabla T &= \begin{bmatrix} T_x \\ T_y \end{bmatrix} \\ &\sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \nu_{xy} \\ \nu_{xy} & \sigma_y^2 \end{bmatrix}\right) \\ &= N(\mu, \Sigma) \end{aligned} \tag{15}$$

that describes the joint distribution of the two derivatives T_x, T_y, and let S^2 be the random variable describing the distribution of slowness squared. This is, for example, the distribution that arises for the derivatives of a single point conditioned on observations under GP regression as described above. Then $\mathbb{E}[S^2] = \mathbb{E}[\nabla T^T \nabla T]$. Note that $Cov[\nabla T, \nabla T] = \mathbb{E}[\nabla T \nabla T^T] - \mathbb{E}[\nabla T]\mathbb{E}[\nabla T]^T$. As the slowness squared is a scalar, I can take the trace to proceed as follows, following Kendrick (2002):

$$\begin{aligned} \mathbb{E}[S^2] &= \mathbb{E}[\nabla T^T \nabla T] \\ &= \mathbb{E}[tr(\nabla T^T \nabla T)] \\ &= tr(\mathbb{E}[\nabla T^T \nabla T]) \\ &= tr(\mathbb{E}[\nabla T]\mathbb{E}[\nabla T]^T + Cov[\nabla T, \nabla T]) \\ &= tr(\mu\mu^T + \Sigma) \\ &= \mu_x^2 + \mu_y^2 + \sigma_x^2 + \sigma_y^2 \\ &> \mu_x^2 + \mu_y^2 = (\mathbb{E}[T_x])^2 + (\mathbb{E}[T_y])^2 \end{aligned} \tag{16}$$

It is instructive to note that the expectation value of squared slowness is strictly greater than the sum-of-squares of the mean derivatives, so that velocities are "biased" lower after accounting for errors. Note that this is true for any calculation that assumes the derivatives have a Gaussian distribution, not just the Gaussian process framework analysed here. This assumption is implicit in any interpolation scheme that is linear in the observed data, if the data has Gaussian uncertainty.

3 Approximation of the posterior using the saddlepoint method

The analytic results obtained for the derivative ∇T have already given us a great deal. Any expectation value that depends on these derivatives (in particular, moments of the phase velocity) can be calculated using the Monte-Carlo

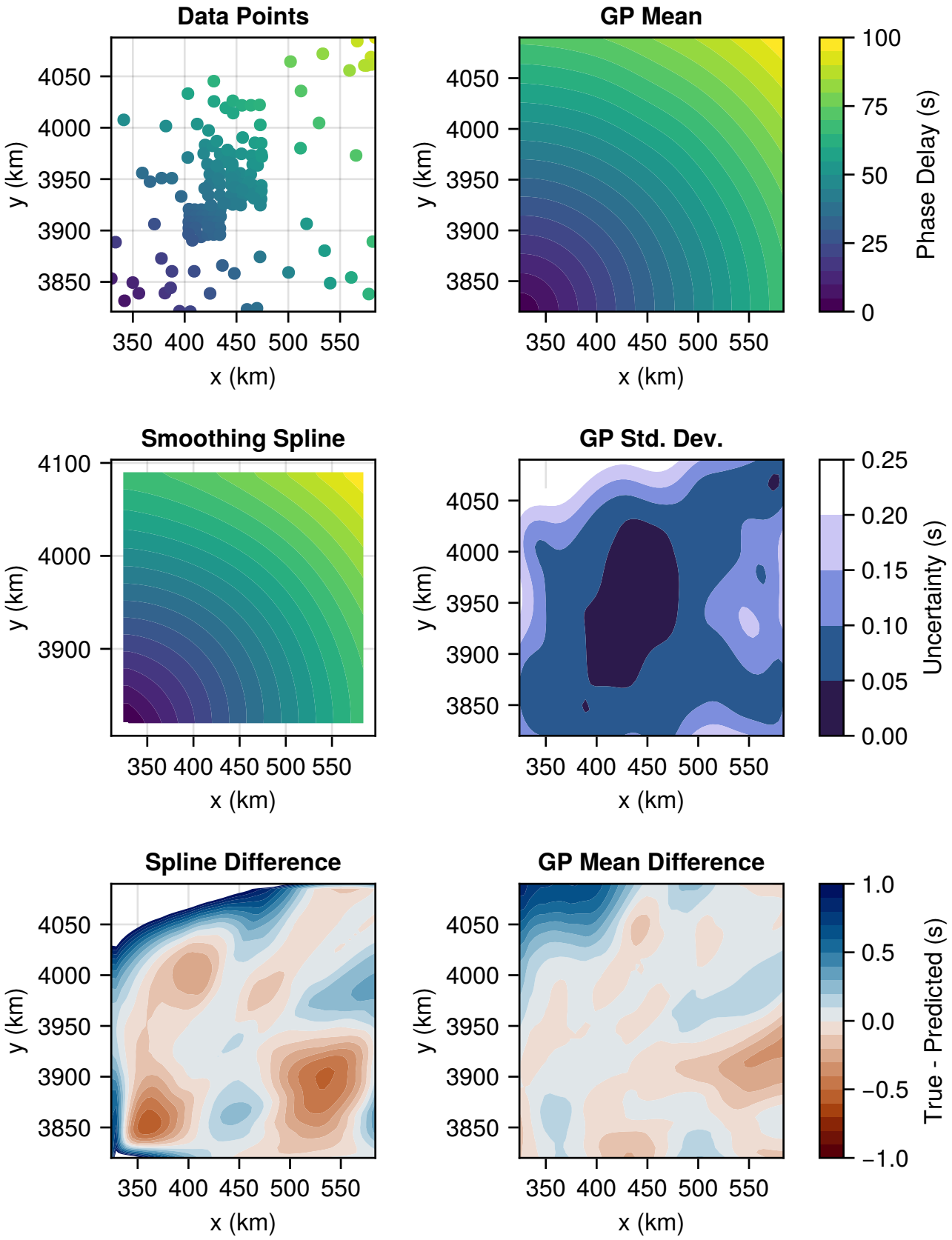

Figure 5 Comparison of the GP posterior (showing mean and point-wise standard deviation) of the phase delay with a smoothing-spline based solution for the example phase delay dataset derived from the model of White et al. (2021), with 154 station locations and 0.1 s added Gaussian noise. There are notable differences in the estimated phase delay, especially where there are gaps in the data coverage. The colouring of the difference plots is arranged according to the usual seismic convention of blue being a fast and red being slow; in this case blue means that the predicted arrival is fast compared to the truth and vice versa.

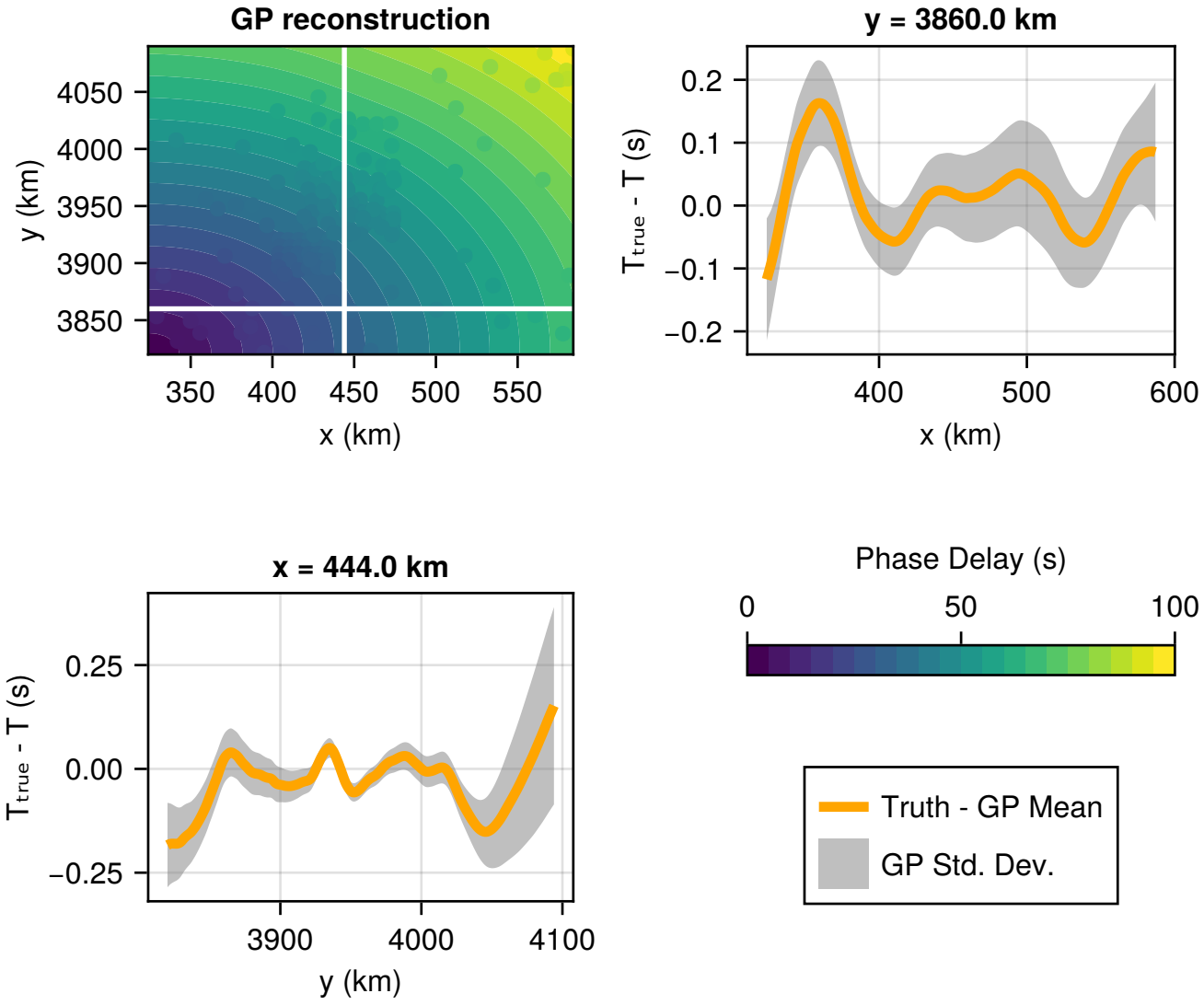

Figure 6 Cross-sections through the GP reconstruction of Figure 5, showing the residual between the true travel time field and the GP mean (orange) and standard deviation (grey). The GP reconstruction is overlaid with the noisy observed delay values.

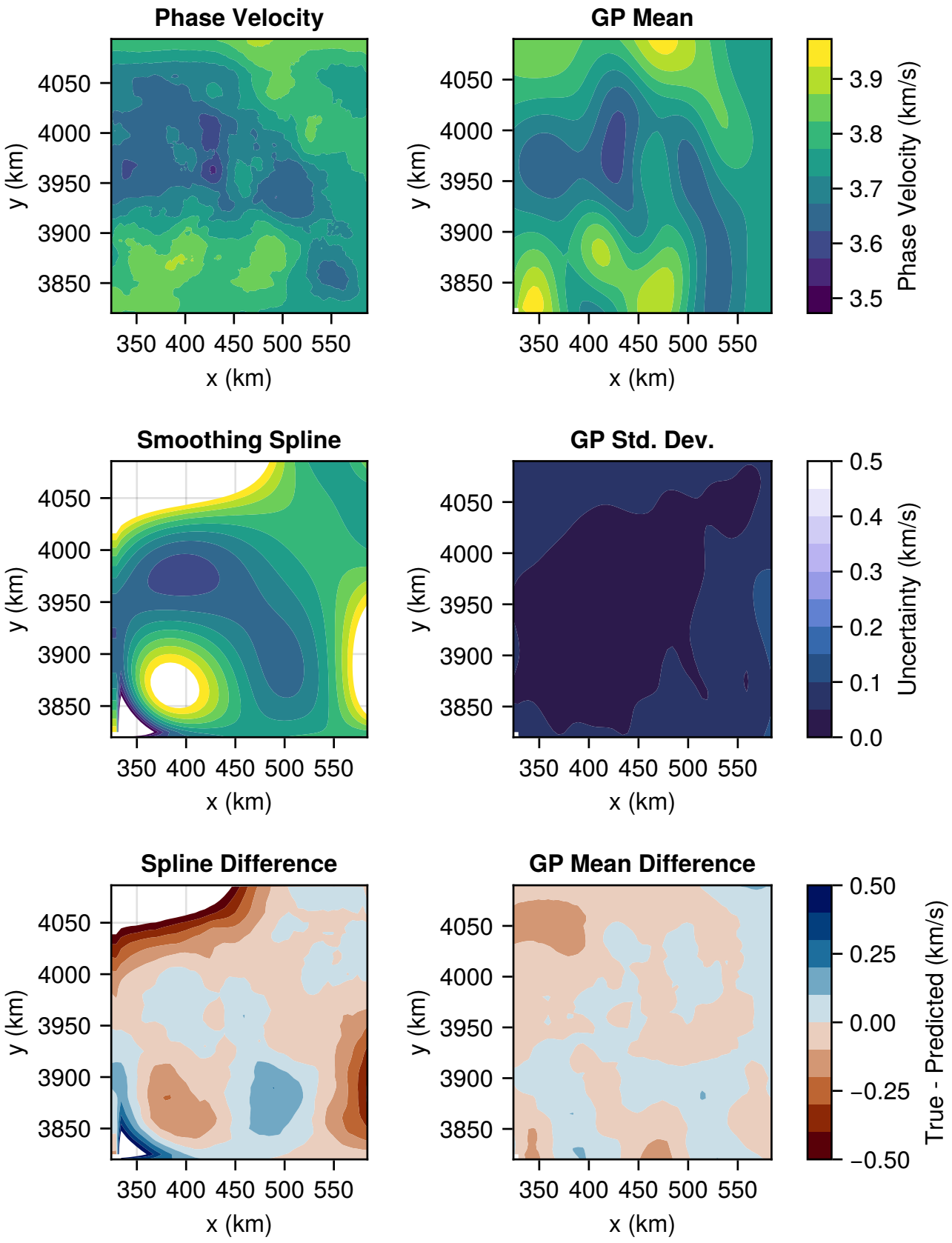

Figure 7 Comparison of the true phase velocity against results calculated using a squared-exponential Gaussian process with tuned hyperparameters for the Ridgecrest, CA velocity model of White et al. (2021). The GP mean and standard deviation are calculated by drawing 100,000 predicted travel time gradients. The spline phase velocity has been calculated using 5^{th} order centred finite differences. The GP result has a mean closer to the truth, and additionally adds uncertainty information, when compared to the smoothing spline.

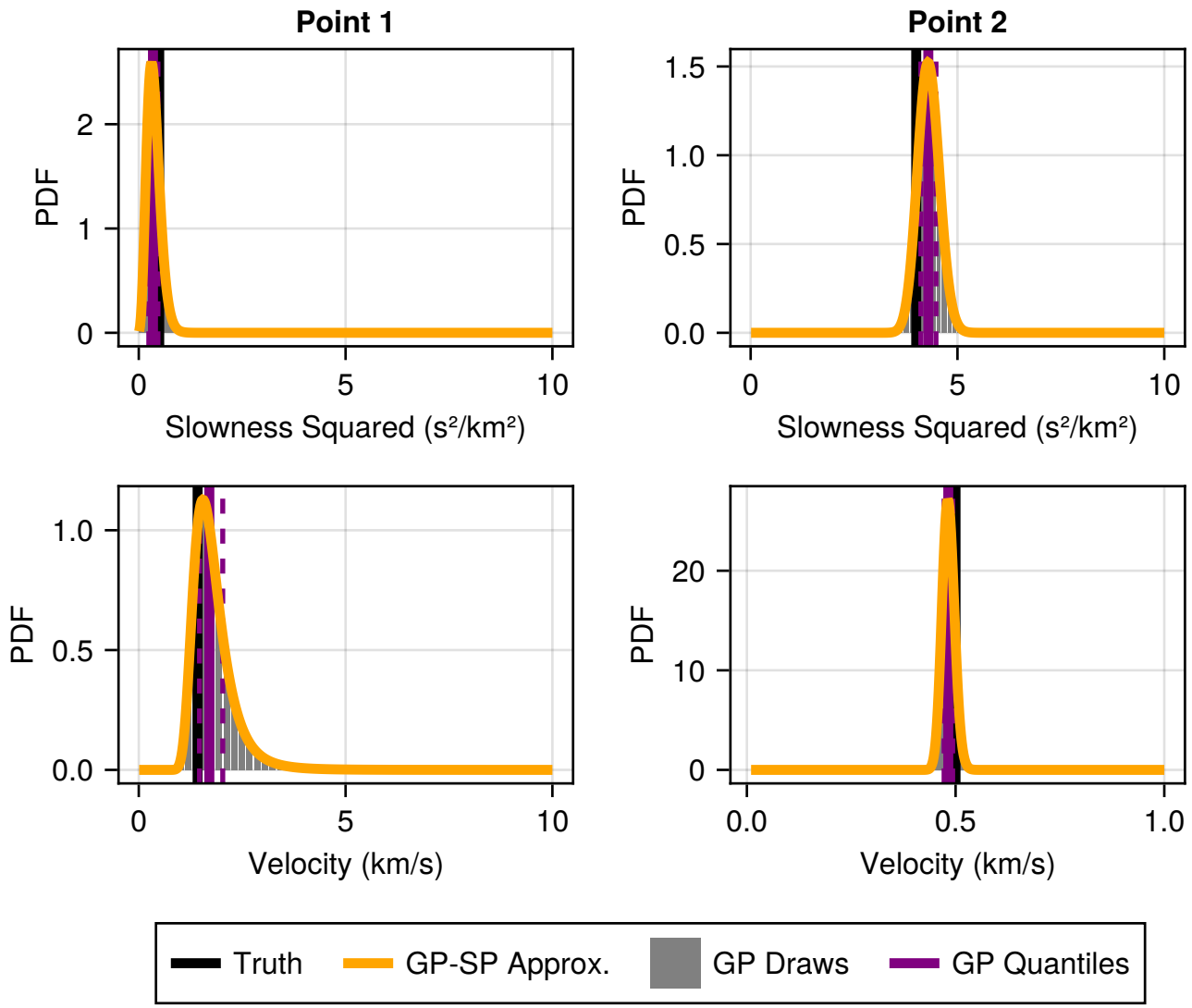

Figure 8 Comparison of the empirical CDF and PDF (grey) for the squared slowness and phase velocity for the near-edge point 1 (7.5,3.75) and more centered point 2 (3.0, 1.0) with the saddlepoint (SP) approximation (orange). For the PDF, the true value is also shown in black and the median, 25^{th} and 75^{th} percentiles of the empirical PDF are shown in purple. The empirical distributions are truncated between 0.01 and 10 for plotting purposes, other than for the velocity of Point 2 which is truncated between 0.01 and 1 due to minimal probability mass above 1.

method — i.e., by drawing many random samples of ∇T and then calculating the desired statistics on this random sample. Because it is possible to draw directly from the posterior of ∇T given Equation 13, every sample can be used and is independent (unlike in Markov-Chain Monte-Carlo). As such, these expectation values will usually converge quickly. However, there are cases where it is still useful to have approximations of the posterior that can be even more quickly calculated; for instance if the eikonal tomography derived phase velocities are being used in a joint inverse problem, or if accurate statistics for extreme values need to be calculated. A frequently used simple approximation would be to use Laplace's method directly on the posterior distribution for $||\nabla T||^2$ or C_p. The approximate posterior using this technique is the best fitting Gaussian distribution. However, looking at Figure 8, it is clear that neither distribution is close to Gaussian, and may not in fact have a clear mode to fit.

Instead of approximating the posterior directly, I instead use the saddlepoint approximation. The saddlepoint approximation for the distribution of random variables was originally proposed by Daniels (1954), with Butler (2007) giving a thorough account of the basic method. Very roughly, the idea is to examine the cumulant generating function (CGF) for a scalar random variable U

$$\begin{aligned} K(s) &= \log \mathbb{E}[\exp(sU)] \\ &= \log \int_{\mathcal{U}} e^{su} f(u) du, \end{aligned} \tag{17}$$

where $f(u)$ is the probability distribution of U and $\mathcal{U}$ is its domain of support. u could be, for example, the slowness squared or the phase velocity at a particular point, while s is a scalar auxiliary variable with units that are the inverse of those of u. Note that $K(s)$ is different from the various K covariance matrices (we have maintained use of K in both cases as they are by far the most common symbols used in the literature for both cases). The existence of the CGF requires that there is some interval $a < 0 < b$ such that the above integral converges. Daniels (1954) showed that the CGF could be used to derive a highly accurate approximation of the PDF (see Butler, 2007, for more information),

$$\hat{f}(u) = \sqrt{\frac{1}{2\pi K''(\hat{s})}} \exp(K(\hat{s}) - \hat{s}u), \tag{18}$$

where $\hat{s}$ is the solution of $K'(s) = u$. $\hat{s}$ is a saddlepoint of the integrand in Equation 17, hence the name "saddlepoint approximation". If the application requires it, $\hat{f}(u)$ then typically has to be normalized to integrate to unity so that it is a true probability distribution, giving us

$$\bar{f}(u) = \frac{\hat{f}(u)}{\int_{\mathcal{U}} \hat{f}(u)du}. \tag{19}$$

If the application only requires the PDF up to proportionality (as is often the case), then the above normalization is not required, and the saddlepoint approximation requires no integration whatsoever. Butler (2007) shows that this optimization problem is well posed and gives a unique real solution for $\hat{f}$, if s is constrained to be inside the interval that contains 0 for which $K(s)$ converges. Serendipitously, this low order method often provides extremely good approximations to the PDF, as the CGF K contains the full information about the distribution of X. For sums of random variables (such as $||\nabla T||^2$), it is almost always easier to construct the CGF K analytically rather than the PDF f, as $K_{U+V}(s) = K_U(s) + K_V(s)$, whereas $f_{U+V}(x) = f_U(x) * f_V(x)$ where U and V are arbitrary random variables and $*$ is the convolution operator. Therefore, when using the saddlepoint approximation to obtain the PDF, multiple potentially slowly converging convolution integrals are converted into a simple root-finding problem with a unique solution. Let us now apply this concept to deriving the PDFs of $||\nabla T||^2$ and C_p from our closed form posteriors for phase delay derivatives ∇T. To do this, my goal is to write the distribution of $||\nabla T||^2$ in a form for which I can determine the CGF $K_{||\nabla T||^2}$, and then use the saddlepoint approximation to obtain the posterior PDF $\hat{f}_{||\nabla T||^2}$, from which I can also obtain the posterior PDF $\hat{f}_{C_p}$ using a change-of-variables formula.

For simplicity, I approximate the posterior for a single point x' given data $(\mathbf{X}, \mathbf{d})$. I have shown that $\nabla T(x')|\boldsymbol{d} \sim N(\mu, \Sigma)$ for a d dimensional mean vector μ and a $d \times d$ covariance matrix Σ. Therefore we can write $\nabla T(x')|\boldsymbol{d}$ in non-centered form using a coordinate transform,

$$\nabla T(x')|\boldsymbol{d} = Q\Lambda^{1/2}h + \mu, \tag{20}$$

where $Q\Lambda Q^T = \Sigma$ is an eigenvalue decomposition of Σ and h is a d-dimensional standard normal variable $h \sim N(0, I)$. Q contains the normalized eigenvectors as its columns and Λ is a diagonal matrix of corresponding eigenvalues. Assuming that the phase delay measurements are taken in different locations, all of the terms in Λ are positive as then Σ, as a non-degenerate covariance matrix, is positive definite. I can then write

$$\begin{aligned} ||\nabla T(x')||^2 &= (Q\Lambda^{\frac{1}{2}}h + \mu)^T(Q\Lambda^{\frac{1}{2}}h + \mu) \\ &= (Qh + \bar{\mu})^T\Lambda(Qh + \bar{\mu}) \\ &= (h + Q^T\bar{\mu})^TQ^T\Lambda Q(h + Q^T\bar{\mu}) \end{aligned} \tag{21}$$

where $\bar{\mu} = \Lambda^{-\frac{1}{2}}\mu$. The eigenvalues collected in Λ are labelled λ_i, with corresponding components of $\bar{\mu}$ labelled $\bar{\mu}_i$. The quadratic form in Equation 21 can be written as a sum over non-central chi-squared distributions (Imhof, 1961; Butler and Paolella, 2008). The degree of freedom of each non-central chi-squared corresponds to the multiplicity of the eigenvalues of Σ, which will for our purposes always be distinct, giving

$$||\nabla T(x')||^2 = \sum_{i=1}^{m} \lambda_i \chi^2(1, \bar{\mu}_i^2). \tag{22}$$

Because of the summation property of the CGF, the CGF of $||\nabla T(x')||^2$ is then (Butler and Paolella, 2008)

$$K_{||\nabla T(x')||^2}(s) = \sum_{i=1}^{m} \left[-\frac{1}{2}\log(1 - 2s\lambda_i) + \frac{s\lambda_i\bar{\mu}_i^2}{1 - 2s\lambda_i} \right], \tag{23}$$

and the derivatives are given by

$$K^{(j)}_{||\nabla T(x')||^2}(s) = 2^{j-1}(j-1)! \sum_{i=1}^{m} \lambda_i^j (1 - 2s\lambda_i)^{-j} \left(1 + \frac{j\bar{\mu}_i^2}{1 - 2s\lambda_i} \right). \tag{24}$$

The domain of convergence in which the root of $K'(s) = u$ is sought is the largest open interval containing zero for which $K_{||\nabla T(x')||^2}(s)$ is defined, which from looking at Equation 23 is $s \in (-\infty, \frac{1}{2\lambda_{max}})$, where λ_{max} is the largest eigenvalue of Σ. Applying the saddlepoint approximation given the above K gives us the saddlepoint distribution $\hat{f}_{||\nabla T(x')||^2}(u)$ for the squared slowness, which can be normalized to give

$$\bar{f}_{||\nabla T(x')||^2}(u) = \frac{\hat{f}_{||\nabla T(x')||^2}(u)}{\int_0^\infty \hat{f}_{||\nabla T(x')||^2}(u)}. \tag{25}$$

The transformation between squared slowness $||\nabla T||^2$ and phase velocity C_p is given by $C_p = g(||\nabla T||^2)$ with $g(u) = \frac{1}{\sqrt{u}}$, which is a monotone decreasing function. The appropriate Jacobian transformation rule to obtain the approximate PDF of phase velocity is then (Kadane, 2011)

$$\begin{aligned} \bar{f}_{C_p(x')}(u) &= -\bar{f}_{||\nabla T(x')||^2}(g^{-1}(u))\frac{dg^{-1}}{du}(u) \\ &= \frac{2\bar{f}_{||\nabla T(x')||^2}\left(\frac{1}{u^2}\right)}{u^3}. \end{aligned} \tag{26}$$

The approximate distributions $\bar{f}_{||\nabla T(x')||^2}(u)$ and $\bar{f}_{C_p(x')}(u)$ are plotted against a histogram of 1,000,000 draws of the squared slowness and phase velocity using the analytic derivatives in Figure 8, showing that the saddlepoint approximations are a close fit. Higher order saddlepoint approximation terms and approximations for the cumulative distribution function (CDF) are collected in Butler (2007).

The saddlepoint method can be further applied to the joint distribution function for phase velocity or squared slowness at two different points to derive the approximate spatial covariance (Al-Naffouri et al., 2016). Because the underlying posterior distributions for the derivatives is given by a GP, the covariance completely describes the spatial behaviour of the velocity distribution, and so the ability to calculate the distribution for any two arbitrary points is sufficient to fully characterize the posterior. However, the resulting root-finding problem will be in two variables rather than one and is substantially more complicated than the forms derived here, so they are left for future work.

4 Discussion

4.1 Implications for sample statistics

Most eikonal tomography applications report per-station per-frequency error statistics by computing the standard error in the mean phase velocity over multiple sources; this averaging process is the second essential step in robust eikonal tomographic results (Lin et al., 2009). Studies typically appeal to the central limit theorem to justify the use of the sample standard error formula and sample mean for quantifying the data distribution. The reported standard errors are then used to weight data in further inversions — a typical use case is to perform 1D Bayesian inversion beneath each station using the mean values and the reported error. Previous methods do not optimally smooth the phase delay regression that underlies eikonal tomography, potentially producing biased results, and do not produce uncertainty estimates for each source. However, uncertainties reported in studies using these methods are often extremely low, amounting to a few percent of the estimated phase velocity.

In our GP framework, Monte Carlo sampling can be used to directly estimate the distribution for sample statistics such as the mean over multiple sources, and hence compute the full PDFs for these averaged quantities. As a motivation, observe that both the empirical distribution for phase velocity and its saddlepoint approximation is heavy tailed for point 1 in Figure 8. This is a point relatively close to the edge, which can result in a distribution that is far from Gaussian. Taking point 1, I draw 4^n samples of velocity from the GP posterior for $n = 0 \ldots 6$, calculate the sample mean and median for the batch of 4^n, and then repeat 100,000 times to find the distribution in the sample statistics. Figure 9 shows the results. The sample mean converges only slowly to a normal distribution, and is still broad even with 16 samples. In comparison, the sample median is well-behaved and converges quickly as the sample size increases. For both sample statistics, the distribution for small numbers of samples is unsurprisingly quite similar to the underlying velocity distribution, and is consequently heavy tailed — this should be taking under consideration for applications such as fitting azimuthal anisotropy profiles to eikonal tomography results, where many azimuth bins near the edges of arrays will often have few contributing sources. Assuming that the standard error formula describes the uncertainty in the measurements is likely an underestimate in that case.

4.2 Future work

In this study, I present the simplest possible implementation of a GP framework for eikonal tomography with analytic derivatives of phase delay. The flexibility of GP modelling offers several opportunities for future improvements that

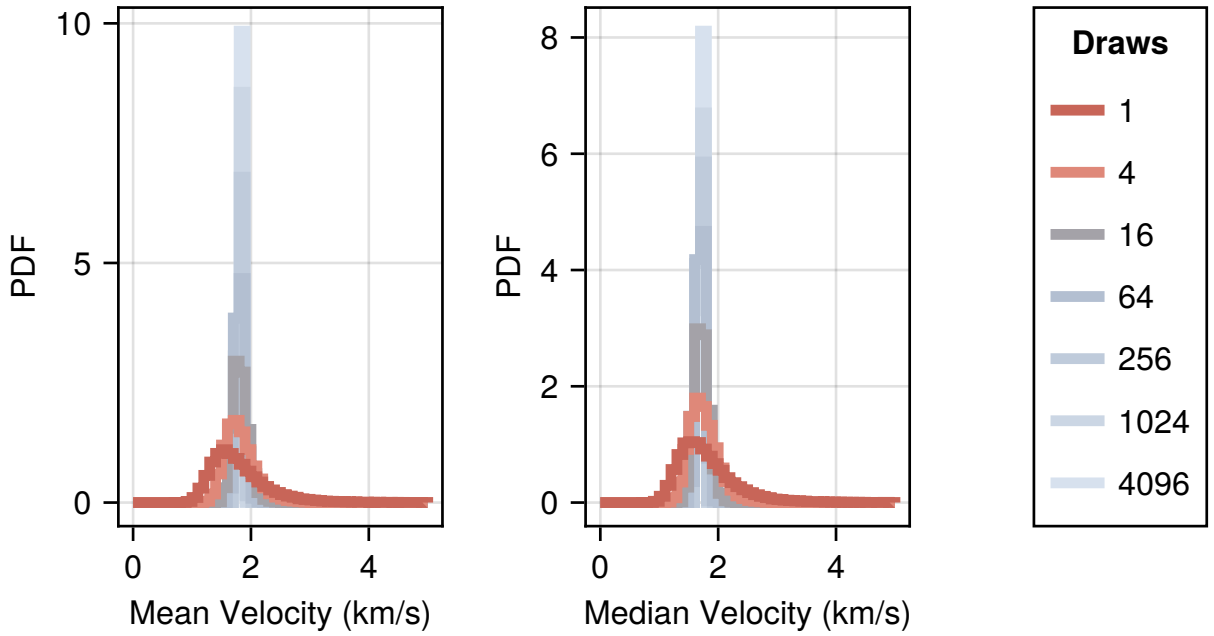

Figure 9 Comparison of the distribution of sample means and sample medians for the phase velocity at (7.5,3.75). The mean or median is calculated by drawing 4^n samples for $n = 0 \ldots 6$. This process is repeated 100,000 times to obtain the distributions of sample means and medians. Compared to the sample median, the sample mean converges to a normal distribution slowly.

should result in more robust inversions. The first of these is that multi-frequency eikonal inversion is naturally handled by GP modelling by assuming a space-frequency covariance function. The most simple model would use a separable function $k((x, f), (x', f')) = k_x(x, x')k_f(f, f')$. A smooth frequency covariance $k_f(f, f')$ would reduce the impact of missing data in particular frequency bins, which can be an issue due to spectral holes in surface wave trains.

Secondly, the squared-exponential kernel used in this study could be further improved to better represent the behaviour of true seismic wavefields; for instance, the problem could be recast in radial coordinates with a radial-azimuthal kernel as studied in Padonou and Roustant (2015). Due to the natural cylindrical symmetry of wave propagation, this may allow us to reduce the uncertainty in the eikonal tomography results. In particular, this kernel choice would be appropriate in use cases such as ambient-noise tomography where the seismic source is inside the array, resulting in highly non-planar wavefronts.

A third option would be to use the GP framework for smoothing the underlying full wavefield records before processing them for phase delay measurements or for other gradient based techniques such as wavefield gradiometry (e.g., Langston, 2007a,b; de Ridder and Biondi, 2015; de Ridder and Maddison, 2018) or full Helmholtz tomography (Lin and Ritzwoller, 2011). These applications would potentially require extending the GP derivative theory to higher order, but again noting that derivatives are linear, the resulting distributions for higher order spatial terms will also be GPs. The GP framework is especially well suited towards the inclusion of strain measurements in joint wavefield reconstruction (e.g., Muir and Zhan, 2021) as the appropriate covariance kernels can be calculated using the results in Equation 13 — an enticing prospect considering the proliferation of distributed acoustic sensing (DAS) strain sensors (Zhan, 2020). GP based techniques have also been used in geodesy to investigate transient strain rates (e.g., Hines and Hetland, 2018), and the saddlepoint approximation techniques investigated here could offer a way to more accurate quantification of strain invariants arising from geodetic analysis.

Finally, as the number of phase delay measurements increases across stations and frequency bins, the size of the data covariance matrix $\hat{K}$ increases. For n measurements, the cost of inverting this matrix scales like $O(n^3)$, so very large collections of measurements pose a challenge for GP based inversion. Due to the popularity of GPs in machine learning research, there are a wide range of sparse GP approximations that produce almost identical results and still result in analytic derivatives once the sparsity structure is determined (e.g., Titsias, 2009; Lindgren et al., 2011; Wilson and Nickisch, 2015). Employing these methods would allow efficient upscaling of the methodology presented

here to multi-frequency inversion of USArray-scale datasets.

4.3 Conclusions

This study derives an analytic posterior distribution for phase delay derivatives, and then derives approximate posteriors for phase velocity using the saddlepoint approximation applied to the eikonal equation. The result is a fully Bayesian eikonal tomography that requires no MCMC sampling to characterize the posterior. As such, computations are easily implemented and highly efficient. Using the GP framework as a basis, I investigated two important effects that impact the interpretation of eikonal tomography results, namely the effect of the inclusion of data uncertainty on the expectation value of velocity and the behaviour of sample statistics, both of which suggest that the uncertainty in eikonal tomography results is greater than previously assessed. The GP framework presents a fully interpretable way forward to improve eikonal tomography in the future, with many opportunities for future work due to the flexible and robust nature of GP modelling.

Data and code availability

The Pluto notebook and data (White et al., 2021) used to generate the results may be found on Zenodo (Muir, 2023). To run the Pluto notebook, users must first install Julia (Bezanson et al., 2017), then use the inbuilt package manager to install Pluto (enter the REPL, hit] to enter package management mode, then `install Pluto`, and `backspace` to re-enter REPL mode). Typing `using Pluto; Pluto.run()` will bring up the Pluto notebook environment from where the notebook can be opened by navigating to it through the filesystem. This will automatically run the notebook, including installing all version-controlled modules required.

Acknowledgements

I acknowledge the support of the European Commission via Marie Skłodowska-Curie Actions Individual Fellowship award number 101027079. I also acknowledge Professor Ron Butler of Southern Methodist University for forwarding the technical report supporting Butler and Paolella (2008), and Professor Victor Tsai of Brown University for providing a pre-submission internal review that was very helpful, as well as the helpful comments of Suzan van der Lee and two anonymous reviewers.

Erratum

An error was found in equation 14 of the original text, which was missing a logarithm around the determinant of the covariance matrix on the right hand side. This equation relates to setting the optimal parameters of the GP covariance matrix. This logarithm was also missing in the associated codes. We have corrected the code and rerun the analysis; we have found that the results have been uniformly substantially improved (i.e. smaller errors and variance for the GP model in both tests). The updated code notebook may be found at DOI https://doi.org/10.5281/zenodo.18240492. The author thanks Esteban Jimenez for identifying the error in the manuscript.

References

Al-Naffouri, T. Y., Moinuddin, M., Ajeeb, N., Hassibi, B., and Moustakas, A. L. On the Distribution of Indefinite Quadratic Forms in Gaussian Random Variables. *IEEE Transactions on Communications*, 64(1):153–165, Jan. 2016. doi: 10.1109/TCOMM.2015.2496592.

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, 59(1):65–98, Jan. 2017. doi: 10.1137/141000671.

Bodin, T. and Maupin, V. Resolution Potential of Surface Wave Phase Velocity Measurements at Small Arrays. *Geophysical Journal International*, 172(2):698–706, Feb. 2008. doi: 10.1111/j.1365-246X.2007.03668.x.

Brocher, T. A. Empirical Relations between Elastic Wavespeeds and Density in the Earth's Crust. *Bulletin of the Seismological Society of America*, 95(6):2081–2092, Dec. 2005. doi: 10.1785/0120050077.

Butler, R. W. *Saddlepoint Approximations with Applications*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2007. doi: 10.1017/CBO9780511619083.

Butler, R. W. and Paolella, M. S. Uniform Saddlepoint Approximations for Ratios of Quadratic Forms. *Bernoulli*, 14(1):140–154, Feb. 2008. doi: 10.3150/07-BEJ6169.

Chevrot, S. and Lehujeur, M. Eikonal Surface Wave Tomography with Smoothing Splines—Application to Southern California. *Geophysical Journal International*, 229(3):1927–1941, June 2022. doi: 10.1093/gji/ggac034.

Daniels, H. E. Saddlepoint Approximations in Statistics. *The Annals of Mathematical Statistics*, 25(4):631–650, 1954.

de Ridder, S. A. L. and Biondi, B. L. Near-surface Scholte Wave Velocities at Ekofisk from Short Noise Recordings by Seismic Noise Gradiometry. *Geophysical Research Letters*, 42(17):7031–7038, Sept. 2015. doi: 10.1002/2015GL065027.

de Ridder, S. A. L. and Maddison, J. R. Full Wave Field Inversion of Ambient Seismic Noise. *Geophysical Journal International*, Aug. 2018. doi: 10.1093/gji/ggy328.

Dierckx, P. *Curve and surface fitting with splines*. Oxford University Press, 1993. doi: 10.1093/oso/9780198534419.001.0001.

Friederich, W. and Wielandt, E. Interpretation of Seismic Surface Waves in Regional Networks: Joint Estimation of Wavefield Geometry and Local Phase Velocity. Method and Numerical Tests. *Geophysical Journal International*, 120(3):731–744, 1995. doi: 10.1111/j.1365-246X.1995.tb01849.x.

Friederich, W., Wielandt, E., and Stange, S. Non-Plane Geometries of Seismic Surface Wavefields and Their Implications For Regional Surface-Wave Tomography. *Geophysical Journal International*, 119(3):931–948, 1994. doi: 10.1111/j.1365-246X.1994.tb04026.x.

Herrmann, R. B. Computer Programs in Seismology: An Evolving Tool for Instruction and Research. *Seismological Research Letters*, 84(6): 1081–1088, Nov. 2013. doi: 10.1785/0220110096.

Hines, T. T. and Hetland, E. A. Revealing Transient Strain in Geodetic Data with Gaussian Process Regression. *Geophysical Journal International*, 212(3):2116–2130, Mar. 2018. doi: 10.1093/gji/ggx525.

Imhof, J. P. Computing the Distribution of Quadratic Forms in Normal Variables. *Biometrika*, 48(3/4):419–426, 1961. doi: 10.2307/2332763.

Kadane, J. B. *Principles of Uncertainty*. Chapman and Hall/CRC, 2011.

Kendrick, D. *Stochastic Control for Economic Models*. University of Texas, Austin, 2 edition, 2002.

Langston, C. A. Spatial Gradient Analysis for Linear Seismic Arrays. *Bulletin of the Seismological Society of America*, 97(1):265–280, Feb. 2007a. doi: 10.1785/0120060100.

Langston, C. A. Wave Gradiometry in Two Dimensions. *Bulletin of the Seismological Society of America*, 97(2):401–416, Apr. 2007b. doi: 10.1785/0120060138.

Lehujeur, M. and Chevrot, S. On the Validity of the Eikonal Equation for Surface-Wave Phase-Velocity Tomography. *Geophysical Journal International*, 223(2):908–914, Nov. 2020. doi: 10.1093/gji/ggaa352.

Lin, F.-C. and Ritzwoller, M. H. Helmholtz Surface Wave Tomography for Isotropic and Azimuthally Anisotropic Structure: Helmholtz Surface Wave Tomography. *Geophysical Journal International*, 186(3):1104–1120, Sept. 2011. doi: 10.1111/j.1365-246X.2011.05070.x.

Lin, F.-C., Ritzwoller, M. H., and Snieder, R. Eikonal Tomography: Surface Wave Tomography by Phase Front Tracking across a Regional Broad-Band Seismic Array. *Geophysical Journal International*, 177(3):1091–1110, June 2009. doi: 10.1111/j.1365-246X.2009.04105.x.

Lindgren, F., Rue, H., and Lindström, J. An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach: Link between Gaussian Fields and Gaussian Markov Random Fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, Sept. 2011. doi: 10.1111/j.1467-9868.2011.00777.x.

Mandic, V., Tsai, V. C., Pavlis, G. L., Prestegard, T., Bowden, D. C., Meyers, P., and Caton, R. A 3D Broadband Seismometer Array Experiment at the Homestake Mine. *Seismological Research Letters*, 89(6):2420–2429, Aug. 2018. doi: 10.1785/0220170228.

McHutchon, A. *Nonlinear Modelling and Control Using Gaussian Processes*. PhD thesis, Cambridge, 2014.

Muir, J. GP-Eikonal. 2023. doi: 10.5281/zenodo.10036494.

Muir, J. B. and Zhan, Z. Wavefield-Based Evaluation of DAS Instrument Response and Array Design. *Geophysical Journal International*, 229 (1):21–34, Dec. 2021. doi: 10.1093/gji/ggab439.

Padonou, E. and Roustant, O. Polar Gaussian Processes for Predicting on Circular Domains. page 24, 2015. doi: hal-01119942v1.

Pollitz, F. F. Observations and Interpretation of Fundamental Mode Rayleigh Wavefields Recorded by the Transportable Array (USArray). *Journal of Geophysical Research: Solid Earth*, 113(B10), 2008. doi: 10.1029/2007JB005556.

Rasmussen, C. E. Gaussian Processes to Speed up Hybrid Monte Carlo for Expensive Bayesian Integrals. *Bayesian Statistics*, 7:651–659, 2003.

Rasmussen, C. E. and Williams, C. K. I. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, Mass, 2006.

Solak, E., Murray-smith, R., Leithead, W., Leith, D., and Rasmussen, C. Derivative Observations in Gaussian Process Models of Dynamic Systems. In *Advances in Neural Information Processing Systems*, volume 15. MIT Press, 2002.

Titsias, M. Variational Learning of Inducing Variables in Sparse Gaussian Processes. In *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, pages 567–574. PMLR, Apr. 2009.

Treister, E. and Haber, E. A Fast Marching Algorithm for the Factored Eikonal Equation. *Journal of Computational Physics*, 324:210–225, 2016.

Tromp, J. and Dahlen, F. A. Variational Principles for Surface Wave Propagation on a Laterally Heterogeneous Earth - III. Potential Representation. *Geophysical Journal International*, 112:195–209, 1993.

Valentine, A. P. and Sambridge, M. Gaussian Process Models—II. Lessons for Discrete Inversion. *Geophysical Journal International*, 220(3): 1648–1656, Mar. 2020a. doi: 10.1093/gji/ggz521.

Valentine, A. P. and Sambridge, M. Gaussian Process Models—I. A Framework for Probabilistic Continuous Inverse Theory. *Geophysical*

Journal International, 220(3):1632–1647, Mar. 2020b. doi: 10.1093/gji/ggz520.

White, M. C. A., Fang, H., Catchings, R. D., Goldman, M. R., Steidl, J. H., and Ben-Zion, Y. Detailed Traveltime Tomography and Seismic Catalogue around the 2019 Mw 7.1 Ridgecrest, California, Earthquake Using Dense Rapid-Response Seismic Data. *Geophysical Journal International*, 227(1):204–227, June 2021. doi: 10.1093/gji/ggab224.

Wielandt, E. Propagation and Structural Interpretation of Non-Plane Waves. *Geophysical Journal International*, 113(1):45–53, Apr. 1993. doi: 10.1111/j.1365-246X.1993.tb02527.x.

Wilson, A. G. and Nickisch, H. Kernel Interpolation for Scalable Structured Gaussian Processes (KISS-GP). *Proceedings of the 32 nd International Conference on Machine Learning*, pages 1775–1784, 2015.

Yang, Y., Atterholt, J. W., Shen, Z., Muir, J. B., Williams, E. F., and Zhan, Z. Sub-Kilometer Correlation Between Near-Surface Structure and Ground Motion Measured With Distributed Acoustic Sensing. *Geophysical Research Letters*, 49:e2021GL096503, 2022. doi: 10.1029/2021GL096503.

Zhan, Z. Distributed Acoustic Sensing Turns Fiber-Optic Cables into Sensitive Seismic Antennas. *Seismological Research Letters*, 91(1):1–15, Jan. 2020. doi: 10.1785/0220190112.