Trans-dimensional surface reconstruction with different classes of parameterization

Rhys Hawkins[1], Thomas Bodin[1], Malcolm Sambridge[2],Gaël Choblet[3], Laurent Husson[4]

[1]Laboratoire de Géologie de Lyon Terre, Planètes, Environnement, Ecole Normale Supérieure de Lyon, Lyon, France
[2]Research School of Earth Sciences, Australian National University, Canberra, Australia
[3]Université de Nantes, CNRS, Laboratoire de Planétologie et Géodynamique, Nantes, France.
[4]Université Grenoble Alpes, CNRS, ISTerre, Grenoble, France

Key Points:

- We present a new software for trans-dimensional surface reconstruction incorporating Hierarchical Error estimation, Hamiltonian Monte Carlo, and Parallel Tempering.
- We propose two alternative parameterizations to the ubiquitous Voronoi cells: Delaunay triangulation and Clough-Tocher interpolation.
- These alternate parameterizations may open up the application of trans-dimensional surface reconstruction to a wider variety of geophysical problems.

Abstract

The use of Bayesian trans-dimensional sampling in 2D and 3D imaging problems has recently become widespread in geophysical inversion. Its benefits include its spatial adaptability to the level of information present in the data and the ability to produce uncertainty estimates. The most used parameterization in Bayesian trans-dimensional inversions is Voronoi cells. Here we introduce a general software, TransTessellate2D, that allows 2D trans-dimensional inference with Voronoi cells and two alternative underlying parameterizations, Delaunay triangulation with linear interpolation and Clough-Tocher interpolation, which utilize the same algorithm but result in either C^0 or C^1 continuity. We demonstrate that these alternatives are better suited to the recovery of smooth models, and show that the posterior probability solution is less susceptible to multi-modalities which can complicate the interpretation of model parameter uncertainties.

1 Introduction

Geophysical inverse problems regularly involve observations with spatially varying sensitivity to the Earth's properties of interest. Examples include seismic tomography where the location of earthquakes are concentrated at tectonic plate boundaries or fault zones (Rawlinson, Fichtner, Sambridge, & Young, 2014), or assessing regional coastal inundation rates where tide gauge observations are sparsely located (Church & White, 2011), or climate reconstructions from bore hole temperature records (Hopcroft, Gallagher, & Pain, 2009), or estimates of global heat flow (Davies, 2013). A major problem is that the irregular spatial distribution the observations can cause instabilities in the inverse problem when regular grids are used. In a general inverse problem we formulate the problem as

$$\mathbf{G}\mathbf{m} = \mathbf{d} + \epsilon, \tag{1}$$

where $\mathbf{d}$ is a vector of our observations, $\mathbf{m}$ the vector of unknown Earth model parameters, $\mathbf{G}$ the forward model operator and ϵ representing errors. An irregular distribution of observations, where parts of grid are not constrained by the observations can result in a matrix $\mathbf{G}$ that is not full rank, in which case $\mathbf{G}$ is not invertable. Alternatively, or $\mathbf{G}$ may have one or more rows with near linear dependence resulting in poor conditioning of the inverse.

A standard approach to this problem is to regularize the problem by either damping the solution towards a reference model or by penalizing large spatial gradients through maximization of smoothness measures. Such damping or smoothing regularization are commonly performed uniformly across and while there exist criteria for the selection of these weights, they are not without limitations (Hanke, 1996; Hansen, 1999).

This problem has been well recognized within the community and various adaptive parameterization schemes have been implemented. These methods typically use a heuristic metric, such as the density of data coverage, in order to determine if a region should be inverted at a finer resolution (Chiao & Kuo, 2001; Inoue, Fukao, Tanabe, & Ogata, 1990; Kárason & van der Hilst, 2001; Sambridge & Faletič, 2003).

Where observations can be related to a continuous Earth model through sensitivity kernels in linear or near-linear problems, another method is the Backus-Gilbert or Optimal Local Averages (Backus, 1970a, 1970b, 1970c). In this approach the Earth model is continuously parameterized and the problem is one of solving for a local average at a point by using resolution constraints based on the sensitivity kernels. In the original formulation, this required a large computational effort for each point of the domain which limited application of this style of inversion. An alternate formulation was developed in the helio-seismology community which improves the efficiency of Backus-Gilbert inversions (Pjipers & Thompson, 1992) and has recently been applied to large scale seismic tomography problems (Zaroli, 2016).

Recently more general approaches have been proposed that use priors generated from training data (Lochbühler, Vrugt, Sadegh, & Linde, 2015) or structural information (de Pasquale & Linde, 2017) as a way to impose spatially varying model correlation. Alternatively, prior constraints can be controlled by hyper-parameters in a hierarchical Bayesian framework (Malinverno & Briggs, 2004; Valentine & Sambridge, 2018). Bayesian techniques use probabilistic prior information in conjunction with observations to obtain a posterior probability distribution of model parameters, commonly using Markov chain Monte Carlo (McMC) methods (Mosegaard & Tarantola, 1995; Sambridge & Mosegaard, 2002). In McMC methods, rather than searching for a single optimal model, an ensemble of plausible models are computed from which in addition to optimal models, estimates of uncertainty can be obtained.

An extension to traditional fixed model dimension McMC inversion, where the number of model parameters remains fixed, is trans-dimensional or reversible jump McMC (rjMcMC) (Denison, Holmes, Mallick, & Smith, 2002; Green, 1995). In this method the dimension of the model, that is the number of unknown parameters, is inverted for as part of the process. The often repeated claim of trans-dimensional inversion is that it results in a parsimonious solution, that is, the resulting Markov chain ensemble will converge toward models with an efficient number of parameters required to predict the observations within noise levels (neither under-parameterized nor over-parameterized models). This general approach has been utilized in a number of geophysical inverse problems across various disciplines (Bodin, Salmon, Kennett, & Sambridge, 2012; Bodin & Sambridge, 2009; Burdick & Lekić, 2017; Dettmer, Dosso, & Holland, 2011; Dettmer et al., 2016; Dettmer, Molnar, Steininger, Dosso, & Cassidy, 2012; Galetti, Curtis, Baptie, Jenkins, & Nicolson, 2016; Hawkins, Brodie, & Sambridge, 2017; Malinverno, 2002; Olugboji, Lekic, & McDonough, 2017; Piana Agostinetti, Giacomuzzi, & Malinverno, 2015; Piana Agostinetti & Malinverno, 2010; Saygin et al., 2016). Its general advantage over other approaches is that it produces parsimonious inference that results in better estimates of uncertainties as shown in comparisons with more traditional fixed dimensional inversions (Dettmer et al., 2016; Olugboji et al., 2017).

The most common parameterization used in these trans-dimensional inversions is the Voronoi cell (Bodin & Sambridge, 2009; Burdick & Lekić, 2017; Galetti et al., 2016; Saygin et al., 2016). When using Voronoi cells, a 2D or 3D region is parameterized as a collection of cell centers with associated Earth model parameters. The number and location of these nodes vary during the McMC inversion. Predictions of the Earth model values at a particular point in the domain correspond to the parameters of the nearest node, hence Voronoi cells represent nearest neighbor polygons or polyhedra under an L_2 norm. A disadvantage of Voronoi cells is that they are not optimal for representing smoothly varying functions. A second disadvantage is that the spatial gradient of the field is zero everywhere except at the boundaries where spatial gradients are discontinuous. This prevents their use in applications where the forward model requires spatial gradients or where posterior inferences on spatial gradients are useful. Iterative approaches, whereby the mean of a set of recent models in the Markov chain are used to generate smooth models from which approximations of spatial gradients can be obtained are possible, but add further approximations. Conversely, an advantage of Voronoi cells is that they excel at the discovery of spatial discontinuities.

Here we extend the Voronoi cell parameterization to Delaunay triangles with linear and cubic interpolants giving C^0 and C^1 continuous fields for 2D problems. These extensions complement other extensions to Voronoi cell parameterizations such as the Johnson—Mehl tessellation (Belhadj, Romary, Gesret, Noble, & Figliuzzu, 2018) and have analogs in 1D trans-dimensional parameterizations used in geophysical problems where "change points" are modelled with step functions (Ingham, Heslop, Roberts, Hawkins, & Sambridge, 2014) or changes in gradient are modelled with piece wise linear functions (Hopcroft, Gallagher, & Pain, 2007). We show that compared to the two alternatives, the Voronoi cell parameterization poorly recovers features in the inversion of smooth models and introduces multi-modal posteriors that complicate the interpretation of uncertainties. Conversely, the continuous parameterizations are able to better recover continuous fields but perform poorly when attempting to fit observations based on underlying discontinuous 2D fields.

Overall, we show that in cases where the estimated surface is likely to include discontinuities such as inference of tectonics from local GPS observations, a Voronoi cell parameterization is likely to be preferable. For intrinsically smooth 2D fields such as temperature, density, or gravity potentials, one of the new Delaunay parameterizations may be more appropriate. The framework provided by this software allows this parameterization choice which is important for optimal results, as espoused in recent 1D trans-dimensional studies discussing parameterization trade-offs (Gao & Lekić, 2018; Roy & Romanowicz, 2017).

In a last section, we show a synthetic joint inversion of 3 different data sets to constrain relative sea level rise: tide gauge measurements, satellite altimetry and GPS vertical land motion estimates. We jointly invert for two surfaces: absolute land motion, and absolute see level rise. This test further illustrates the fact that the choice of parameterization affects both the recovered structure and its estimated uncertainties as reported by Hawkins and Sambridge (2015).

When used for 2D regression problems, our Bayesian trans-dimensional software can be seen as an alternative to simple interpolation or kriging methods that generally assume a constant spatial correlation length (Oliver & Webster, 1990). Our method is more general and can include more complex forward modelling, likelihoods and error models while adapting solutions to have finer resolution where observations are sufficiently informative.

The regression examples shown here are simple by design, yet the software allows arbitrary forward models and likelihood functions to be used and is therefore more widely applicable to geophysical problems and beyond. Some potential examples include the reconstruction of gravity anomalies from satellite measurements (Sandwell & Smith, 1997), reconstruction of the Moho discontinuity from geophysical data (Bodin, Salmon, et al., 2012), interpolation of aeromagnetic data (Billings, Beatson, & Newsam, 2002), and regional historic climate reconstructions (Hopcroft et al., 2009).

2 Overview of the Algorithm

2.1 Parameterization

We consider the inversion of geophysical data constrained to a 2D field, for example, a region on the Earth's surface defined by latitude and longitude. We first introduce the three parameterizations considered, Voronoi cells, Delaunay triangulation with a linear interpolant, and Delaunay triangulation with a Clough-Tocher interpolant. For each of the parameterizations, a model consists of a set of mobile 2D points (or nodes) with one or more Earth model parameters associated with each point.

The vector of unknown model parameters is thus defined as the set of geographical locations and values associated to each node. Our three different parameterizations propose different ways to interpolate between nodes, and thus can be thought of as three different forward operators that generate a predicted surface from the vector of model parameters. Given a variable dimension model $\mathbf{m}$, for the forward model operator g can be written as $g = f \circ h$ where f is the user defined forward model and h is the parameterization operator that maps the model vector into predictions in Cartesian space. The parameterization operator can be seamlessly be replaced by h_{Voronoi}, h_{Delaunay} or $h_{\mathrm{Clough-Tocher}}$ representing the 3 alternate parameterizations. Note that since the vector of model parameters $\mathbf{m}$ contains the position of nodes, which makes the operator g non-linear even in the cases presented here where the user forward model f is a linear regression operator.

2.1.1 Voronoi cells

In the original introduction of the reversible jump approach of Green (1995), the last example presented was an application of image segmentation using Voronoi cells. This general algorithm has been extended to different geophysical problems such as resistivity tomography (Malinverno, 2002), seismic surface wave tomography (Bodin & Sambridge, 2009), body wave tomography (Burdick & Lekić, 2017; Piana Agostinetti et al., 2015), CSEM tomography (Ray & Key, 2012), finite fault inversion (Dettmer, Benavente, Cummins, & Sambridge, 2014), estimates of coastal inundation (Choblet, Husson, & Bodin, 2014), and reconstructing surfaces of geodetic uplift rates (Husson, Bodin, Spada, Choblet, & Comé, 2018).

In the Voronoi cell parameterization, the model is defined using a number of nodes representing the Voronoi cell centers. Each Voronoi cell is given a set of one or more Earth model parameters. The reconstructed surface parameter at a given point corresponds to the value of the nearest Voronoi cell node. This parameterization produces surfaces with constant values in each Voronoi cell and discontinuities at Voronoi cell edges.

The Voronoi cell approach would be seemingly implausible for the inversion of geophysical problems where heterogeneity is expected to be smooth. This seeming incongruity hasn't prevented the successful application of trans-dimensional Voronoi cells to geophysical inverse problems such as surface wave tomography inversion as the average of a large ensemble of such models will generate a smoothly varying posterior mean (Bodin & Sambridge, 2009; Galetti et al., 2016; Saygin et al., 2016).

2.1.2 Delaunay triangulation with linear interpolation

As an alternative, we propose a relatively simple modification to the Voronoi cell approach where the dual of the Voronoi cell, the Delaunay triangulation is used. In this parameterization, rather than the model nodes representing the center of Voronoi cells, they represent vertices of a triangulation of the domain.

In this case, rather than the values at a given spatial point being determined by a nearest node, the model nodes defining the triangle can be linearly interpolated to any point within the triangle by computing Barycentric coordinates (Sambridge, Braun, &

McQueen, 1995). This then provides a model that describes a continuous field over the domain but with discontinuities in the gradient at triangle edges.

2.1.3 Delaunay triangulation with Clough-Tocher interpolation

A further extension to the Delaunay triangulation replaces the linear interpolant based on the barycentric coordinates of a point with a cubic interpolant, a modified Clough-Tocher interpolant (Clough & Tocher, 1965; Mann, 1998).

In this parameterization, gradients are estimated at nodes from the values at neighboring nodes, analogously to 1D Cubic Hermite interpolation. The estimated node gradients are subsequently used to constrain the normal gradients at triangle edges so that within each triangle, a cubic interpolant is available that also maintains continuous gradients across each triangle edge. There is an extra computation burden in this method as a small two by two system has to be solved for each node of the model before a point can be interpolated. Details of the exact formulation used here appear in supplementary material.

2.1.4 Summary

To give an example of each of the three parameterizations available in the software, in Figure 1 we show maps of a 2D field where the same model vector $\mathbf{m}$ is used, defined with 5 nodes: one central node at (0,0) with a value of 1, and four corner nodes at ($\pm$1,$\pm$1) with values of 0, essentially a 2D delta function.

2.2 Bayesian formulation

In a Bayesian approach to inference, the solution we obtain is the *a posteriori* probability distribution or posterior (Mosegaard & Tarantola, 1995; Sambridge & Mosegaard, 2002). This is the probability density of the model space given the observed data, or written mathematically, $p(\mathbf{m}|\mathbf{d})$, where $\mathbf{m}$ is our vector of model parameters and $\mathbf{d}$ our vector of observations. The posterior distribution is defined through Bayes' theorem (Bayes, 1763)

$$p(\mathbf{m}|\mathbf{d}, \mathcal{I}) = \frac{p(\mathbf{m}|\mathcal{I})p(\mathbf{d}|\mathbf{m}, \mathcal{I})}{p(\mathbf{d}|\mathcal{I})}, \tag{2}$$

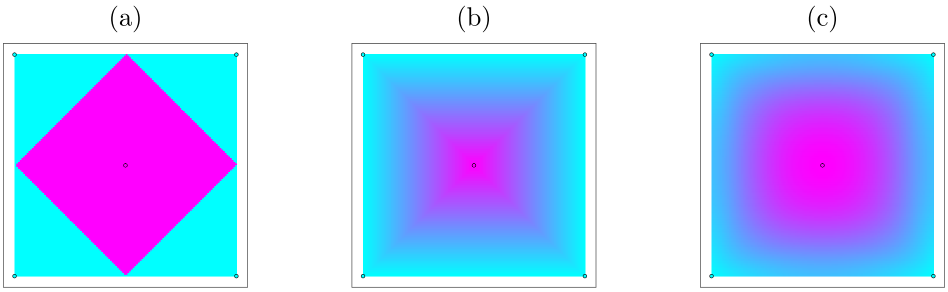

Figure 1. An example of the differences in each parameterization used in this study. Each of the parameterizations are defined with 5 points with a value of one at the center (0,0) and values of zero at the corners (±1,±1). In (a) is the Voronoi cell parameterization commonly used in trans-dimensional inversion, in (b) we show the linear Delaunay parameterization, and (c) the cubic Clough-Tocher parameterization.

where $p(\mathbf{m}|\mathcal{I})$ is the prior, $p(\mathbf{d}|\mathbf{m}, \mathcal{I})$ is the likelihood analogous to the measure of fit to the observations, and $p(\mathbf{d}|\mathcal{I})$ is normalization term often called the "evidence". The dependence $\mathcal{I}$ represents additional prior information within the formulation of problem and the chosen parameterization forms part of this dependence (Malinverno, 2002). In many non-linear geophysical inverse problems, this probability density function is approximated numerically using McMC techniques. As we will see in some synthetic examples, the posterior is highly dependent on choices in the formulation of the problem with the focus herein on the selected parameterization.

In simple problems, the posterior can be evaluated analytically, but in many cases numerical methods are required. Markov chain Monte Carlo (McMC) sampling approach can be applied to the numerator of the right-hand side of (2) to obtain an estimate of the posterior probability distribution up to the normalizing constant of the evidence, which is often difficult to compute explicitly (Sambridge, Gallagher, Jackson, & Rickwood, 2006), although numerical techniques are available (Brunetti, Linde, & Vrugt, 2017; Schöniger, Wöhling, Samaniego, & Nowak, 2014).

2.3 Markov chain Monte Carlo (McMC)

In an McMC inversion, an ensemble of plausible models is constructed, some of these models may not fit the observations optimally but nonetheless are representative of the tails or intermediate regions of multi-modal of posterior distributions. Models are included in this ensemble based on a criteria for the acceptance of proposed perturbations to model parameters. A commonly used acceptance criterion is the Metropolis-Hastings (Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953) where for a proposed transition from model $\mathbf{m}$ to $\mathbf{m}'$, the acceptance is given by

$$\alpha = \min\left\{1, \frac{p(\mathbf{m}'|\mathcal{I})}{p(\mathbf{m}|\mathcal{I})}\frac{p(\mathbf{d}|\mathbf{m}',\mathcal{I})}{p(\mathbf{d}|\mathbf{m},\mathcal{I})}\frac{Q(\mathbf{m}'\to\mathbf{m})}{Q(\mathbf{m}\to\mathbf{m}')}\right\}, \tag{3}$$

where the term $\frac{p(\mathbf{m}'|\mathcal{I})}{p(\mathbf{m}|\mathcal{I})}$ is the prior ratio, $\frac{p(\mathbf{d}|\mathbf{m}',\mathcal{I})}{p(\mathbf{d}|\mathbf{m},\mathcal{I})}$ the likelihood ratio, and $\frac{Q(\mathbf{m}'\to\mathbf{m})}{Q(\mathbf{m}\to\mathbf{m}')}$ the proposal ratio.

The proposal ratio represents the probability distribution that perturbs an initial model $\mathbf{m}$ to obtain $\mathbf{m}'$. At each iteration of an McMC inversion, a new model $\mathbf{m}'$ is created, its likelihood computed and the acceptance rate calculated. The new model is accepted with the probability α. If the new model is accepted, then $\mathbf{m}$ is set to $\mathbf{m}'$, otherwise $\mathbf{m}'$ is rejected and the current model is unchanged. After repeating this process for a suitably large number of iterations we obtain a set of models. It is customary to remove some number of initial models that are considered pre-converged or "burnin" models, after which is left a chain or ensemble of models that approximate the posterior distribution (Brooks, Gelman, Jones, & Meng, 2011).

2.4 Reversible Jumps

An extension to standard McMC is Birth/Death McMC (Geyer & Møller, 1994) and the more general Reversible Jump McMC (Denison et al., 2002; Green, 1995; Malinverno, 2002) where additional proposals are available that change the model dimension, that is, the number of Voronoi nodes or Delaunay vertices in our case. The acceptance criteria for reversible jump or trans-dimensional steps is

$$\alpha = \min\left\{1, \frac{p(\mathbf{m}'|\mathcal{I})}{p(\mathbf{m}|\mathcal{I})}\frac{p(\mathbf{d}|\mathbf{m}',\mathcal{I})}{p(\mathbf{d}|\mathbf{m},\mathcal{I})}\frac{Q(\mathbf{m}'\to\mathbf{m})}{Q(\mathbf{m}\to\mathbf{m}')}|\mathcal{J}|\right\}, \tag{4}$$

where the additional term $|\mathcal{J}|$ is the determinant of the Jacobian of the model transformation from one dimension or parameterization to another. This is required to preserve volume between the two dimensions.

In the context of the parameterizations discussed here, trans-dimensional steps change the number of nodes used to fit the observations. This is facilitated with proposals that add and remove a single nodal point and its associated parameter values, called Birth and Death proposals (Geyer & Møller, 1994). The number and distribution of nodes self-adapts to the resolving power of the data, which is in stark contrast to traditional methods that impose a globally fixed resolution in the formulation of the problem, that is, through a fixed grid.

2.5 Hamiltonian Monte Carlo

Through the careful tuning of proposal distributions, the acceptance rates of McMC inversion can approach optimal values. One inherent problem with McMC is that proposals are generally designed to be small perturbations from the current model, for example, perturbations sampled from a Gaussian distribution. The size of the perturbations need to be tuned to be small to achieve reasonable acceptance rates and as such can result in a high degree of correlation between neighboring models in a chain. This reduces the effectiveness of the chain to estimate the posterior by reducing the Effective Sample Size (ESS) (Brooks et al., 2011).

An advance over McMC is Hamiltonian Monte Carlo (Duane, Kennedy, Pendelton, & Roweth, 1987; Neal, 1994, 2011) (originally called Hybrid Monte Carlo), where an additional calculation of the gradient of the likelihood function is used to propose models that are less correlated, that is, further away from the current model while retaining a high likelihood and higher probability of acceptance. Hamiltonian Monte Carlo increases convergence rates and increases the effective sample size of a chain resulting in fewer iterations required for sampling a posterior. This comes at the expense of requiring calculation of gradients. Regardless of this extra cost, Hamiltonian Monte Carlo has recently been used in some non-linear geophysical inverse problems (Fichtner & Simutė, 2018; Fichtner, Zunini, & Gebraad, 2019; Sen & Biswas, 2017).

Both the more common Metropolis-Hastings McMC and HMC build an ensemble of plausible models through a Markov chain and are hence Markov chain methods. The primary difference that in McMC we sample and apply a perturbation to the current model from a probability distribution. In HMC, a random initial momentum is sampled and the model trajectory is simulated with Hamiltonian dynamics using the log of the posterior as the potential in the Hamiltonian equation.

In the problem considered in this paper, the model parameters consist of a set of node positions and their associated values. Since the existence of the gradient of the likelihood with respect to node positions is forward model dependent for the Voronoi cell parameterization, we use Metropolis-Hastings proposals for perturbing the location of nodes, Hamiltonian proposals perturbing the values associated with nodes and reversible jump proposals for changes of dimension although hybrid HMC/reversible jump proposals are possible (Sen & Biswas, 2017).

2.6 Likelihood and Hierarchical error estimation

The form of a likelihood function in an inverse problem in a Bayesian framework is primarily dictated by the expected distribution of errors in the formulation of the inverse problem with contributions from the data errors and forward modelling errors. The common assumption is that

$$\mathbf{d}_{\text{observed}} = g(\mathbf{m}_{\text{true}}) + \epsilon_{\text{data}} + \epsilon_{\text{theory}} + \epsilon_{\ldots}, \tag{5}$$

that is, our observations are a sum of the observations predicted from the true model plus some combination of known and unknown sources of errors. Here we have indicated two common sources of noise, ϵ_{data} represents measurement or observational errors, and ϵ_{theory} represents general theoretical errors that include simplifying approximations and numerical imprecision in forward modelling represented by the operator g, but also errors due to the inability of the parameterization to represent the true 2D field.

The likelihood for a particular set of predictions from a model, $g(\mathbf{m})$, becomes

$$p(\epsilon) = p(g(\mathbf{m}) - \mathbf{d}), \tag{6}$$

where ϵ without the subscript represents the combined noise.

A common choice of likelihood function is a Gaussian distribution. The rationale for this is that since we assume that the noise ϵ is a combination of various sources of error, the resulting distribution will be asymptotically Gaussian due to the central limit theorem (assuming the errors have finite variance).

A Gaussian likelihood is of the form

$$p(\mathbf{m}|\mathbf{d}, \mathcal{I}) = \frac{1}{\sqrt{2\pi}|C_d|} \exp\left\{-\frac{1}{2}(g(\mathbf{m}) - \mathbf{d})^T C_d^{-1} (g(\mathbf{m}) - \mathbf{d})\right\}, \tag{7}$$

where C_d is the covariance matrix of errors. The software allows writing of custom likelihood functions, however for simplicity we are using diagonal covariance matrices, that is, the errors are independent for each observation. In more complex and real data problems, this assumption would be overly simplistic and covariance or auto regressive errors would be more appropriate (Bodin, Sambridge, et al., 2012; Dettmer et al., 2012; Dosso & Wilmut, 2006; Kolb & Lekić, 2014). In the diagonal covariance matrix case, the Gaussian likelihood reduces to

$$p(\mathbf{m}|\mathbf{d}, \mathcal{I}) = \frac{1}{\prod_i \sqrt{2\pi}\sigma_i} \exp\left\{-\sum_i \frac{(g(\mathbf{m})_i - \mathbf{d}_i)^2}{2\sigma_i^2}\right\}. \tag{8}$$

The observational uncertainty is often estimated crudely in real world applications and will not account for other sources of error such as theoretical errors. This suggests that given

$$\epsilon = \epsilon_{\text{data}} + \epsilon_{\text{theory}} + \ldots, \tag{9}$$

that the σ value in (8) should in fact be set to

$$\sigma_i = \sqrt{\sigma_{i,\text{observation}}^2 + \sigma_{\text{theory}}^2} \tag{10}$$

where σ_{theory} is the unknown standard deviation of the theoretical noise. This unknown theoretical noise can be included as a hyper-parameter to be inverted for during the inversion using a hierarchical Bayes approach (Malinverno & Briggs, 2004).

Formulating a hierarchical error model is a complex procedure, and while the software supports an arbitrary number of hierarchical parameters, we have elected to assume that the theoretical errors are small relative to the data errors and use a single scaling term, that is

$$\sigma_i = \lambda \sigma_{i,\mathrm{observation}}, \tag{11}$$

where λ is the unknown scaling term. The benefit of this approach, in addition to its simplicity, is that it preserves the relative weighting of the inversion due to individual observational errors. In real data problems, such a simple hierarchical error model may not be appropriate. Again we stress that this is implemented in the user defined likelihood function and so the operation of hierarchical parameters can be modified to suit problems where the above assumptions are not appropriate.

2.7 Parallel Tempering

One of the common issues with trans-dimensional sampling is the often low acceptance rates for trans-dimensional proposals resulting in poor sampling of the posterior, which is particularly important for statistical inference on the number of parameters. While various approaches have been successfully applied to improve the acceptance rates of trans-dimensional proposals (Al-Awadhi, Hurn, & Jennison, 2004; Sen & Biswas, 2017), we incorporate Parallel Tempering (Dosso, Holland, & Sambridge, 2012; Earl & Deem, 2005; Sambridge, 2014) in the inversion to improve mixing between models with different dimensions.

Parallel Tempering uses multiple parallel chains at different temperatures T, with statistical inference performed only with the chains at $T = 1$. The effect of the temperature is in the acceptance criteria where it is applied to the likelihood ratio

$$\alpha = \min\left\{1, \frac{p(\mathbf{m}'|\mathcal{I})}{p(\mathbf{m}|\mathcal{I})}\left[\frac{p(\mathbf{d}|\mathbf{m}',\mathcal{I})}{p(\mathbf{d}|\mathbf{m},\mathcal{I})}\right]^{1/T}\frac{Q(\mathbf{m}'\to\mathbf{m})}{Q(\mathbf{m}\to\mathbf{m}')}|\mathcal{J}|\right\}, \tag{12}$$

where T is the temperature. At higher temperature, the effect of the likelihood ratio is diminished and the trans-dimensional proposals acceptance rates will tend to increase. Periodically, model exchanges are proposed between chains at different temperatures enabling better exploration of the posterior and mixing between models of different dimension.

2.8 Convergence

In an McMC/HMC simulation, a large number of candidate models are available from which statistical inference can be made. It is common practice to remove some number of models from the start of the chain, called "burn in" samples where the chain may contain unconverged models. In addition, chain thinning is often performed where only every nth model is retained from the chain to reduce the effect of correlation between neighboring models in the chain. While we utilize HMC to reduce this correlation potentially obviating the need for thinning, we retain McMC proposals for moves of type birth, death and hierarchical proposals, i.e. perturbations to λ in (11).

To ensure convergence within a trans-dimensional inversion, standard approaches such as the Gelman-Rubin statistic (Gelman & Rubin, 1992) are difficult to apply as the variance of an individual model parameters cannot be reliably calculated in a chain where the model dimension changes. The Gelman-Rubin statistic can be computed for hyper-parameters of the inversion such as the hierarchical error scale which does give some measure of the convergence between chains (Hawkins et al., 2017).

2.9 Summary

In the algorithm presented here, we simulate multiple Markov chains initialized with from random models using independently seeded random number generators. The chain is simulated for a fixed number of iterations and at each iteration, one of the following proposals is chosen at random

Value The value(s) of the nodes are perturbed using a Hamiltonian Monte Carlo proposal.

Move The position of a randomly chosen node is perturbed using an McMC proposal.

Birth A new node and its associated value(s) are added using an rjMcMC proposal.

Death A node is selected at random to be removed using an rjMcMC proposal.

Hierarchical The hierarchical scaling parameter is perturbed using an McMC proposal.

During the inversion and at a predefined rate, the independent chains perform a parallel tempering exchange swap. Only the chains with temperatures of one have their ensembles processed for statistical inferences. Details of each class of the proposal appear in appendices.

3 Synthetic Regression

As a first demonstration of the software and the effects of parameterization, we present an application to a 2D regression problem. We show that the posterior solution (and hence estimated surface uncertainties) strongly depends on the parameterization. The main point here is to demonstrate the effects of a poor parameterization choice. The first surface to reconstruct is shown in Figure 2(a), it is smooth and consists of the sum of four Gaussians. The second, shown in Figure 2(b) is a tessellated image with regions approximately corresponding to the first, but with straight edged discontinuities.

The synthetic observations were created by randomly generating 100 points within the region, illustrated with crosses in the figures, and sampling the true model at those points. Independent Gaussian noise was added to each observation with a standard deviation of 0.05, which is approximately a five percent error level given the range of values is approximately $0 \ldots 1$.

In total six inversions are computed using the three different parameterizations and the two synthetic data sets. The same settings were used for each inversion, that is, we use 28 parallel chains with 4 temperatures logarithmically spaced between 1 and 5. The initial model is randomly generated from the priors. The Hamiltonian step size and McMC proposal widths are tuned to obtained reasonable acceptance rates (approximately 0.80 for HMC and 0.24 for McMC). The prior on the values are set to uniform between -0.5 and 2, thus encompassing the range of the unknown Earth model parameter in this synthetic example. The prior on the hierarchical scaling is also uniform between 0.5 and 5. Each inversion was simulated for one million iterations.

In Figure 3 we show the results for the inversion of the synthetic regression data set created from the smooth model. In this case, for summary purposes, we have chosen to show the mean and standard deviation of the ensemble, however other choices are possible such as median and credible interval widths (we show images of absolute errors from the true model and maximum a posteriori probability (MAP) images in the supplementary material). In each case, the true model is recovered relatively well given the level of noise. In the mean models, the progressively smoother results are evident as higher order interpolants are used, that is in (a) Voronoi cells are effectively 0th order, followed in (c) by Delaunay with a linear interpolant and lastly in (e) Delaunay with a cubic interpolant.

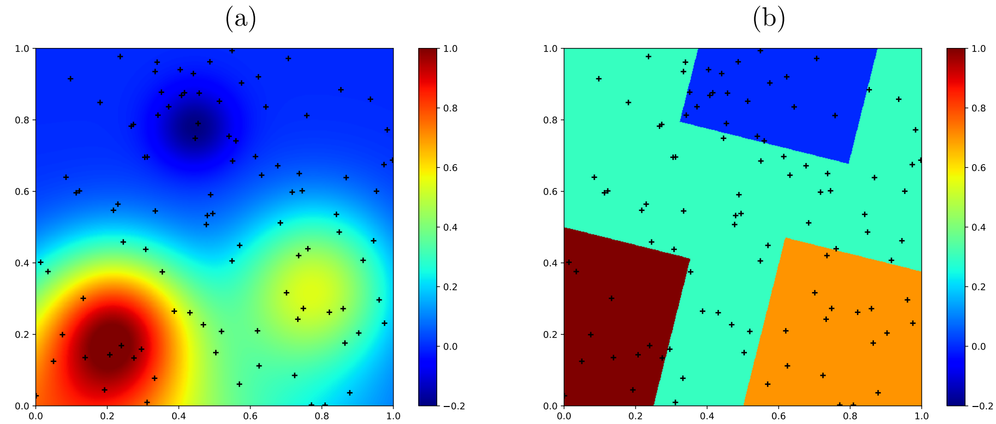

Figure 2. The true models used in the synthetic regression examples. In (a) the true model is smooth and consists of a sum of four Gaussians, wheres (b) is a tessellated approximation of the same model with straight edges and discontinuities. The randomly located observation points are indicated with crosses.

The standard deviation maps of the Voronoi cell parameterization in Figure 3(b) contain ring like structures. These features are caused by the combination of discontinuities in the Voronoi cell parameterization and their mobility. It has been claimed that this feature is only evident in non-linear forward models such as non-linear tomography (Galetti, Curtis, Meles, & Baptie, 2015), however we see they appear here in a linear regression forward model.

Trans-dimensional inversion with Voronoi cells introduces non-linearity to the problem through the dynamic number and location of Voronoi nodes. Strictly speaking, these filaments of large standard deviation occur due to the mobility of the Voronoi cells introducing multi-modalities in the posterior near Voronoi cell edges, which in turn leads to large standard deviations. We can see that the posterior standard deviation for the two Delaunay parameterizations by comparison are generally smaller.

In trans-dimensional inversion, and particularly for Voronoi cell parameterizations, simply plotting the standard deviation as a measure of uncertainty will not necessarily give an accurate appraisal of the posterior. If we instead take a transect through the ensemble along a particular horizontal line and look at the marginal probability distributions, we can visibly see how the distribution varies spatially. In Figure 4 we show the distribution along a horizontal transect for each of the parameterizations in (b), (c) and (d) with the location of the transect indicated with a dashed line in (a). What is clear

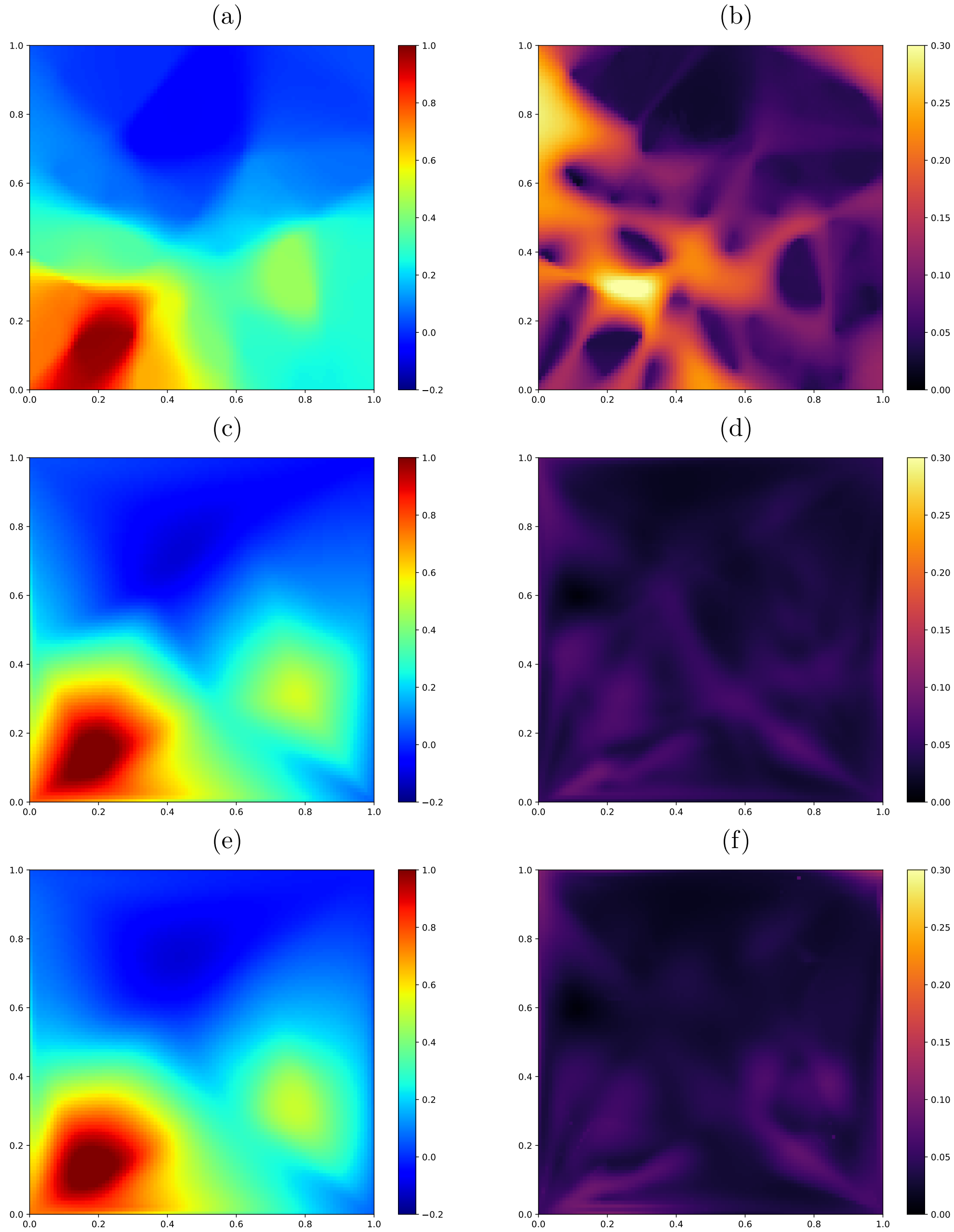

Figure 3. The summary plots for the inversion of the smooth model. In (a) and (b) are the mean and standard deviation results for the Voronoi cell parameterization. In (c) and (d) are the mean and standard deviation results for the Delaunay triangulation with linear interpolant parameterization. In (e) and (f) are the mean and standard deviation results for the Delaunay triangulation with Clough-Tocher interpolant.

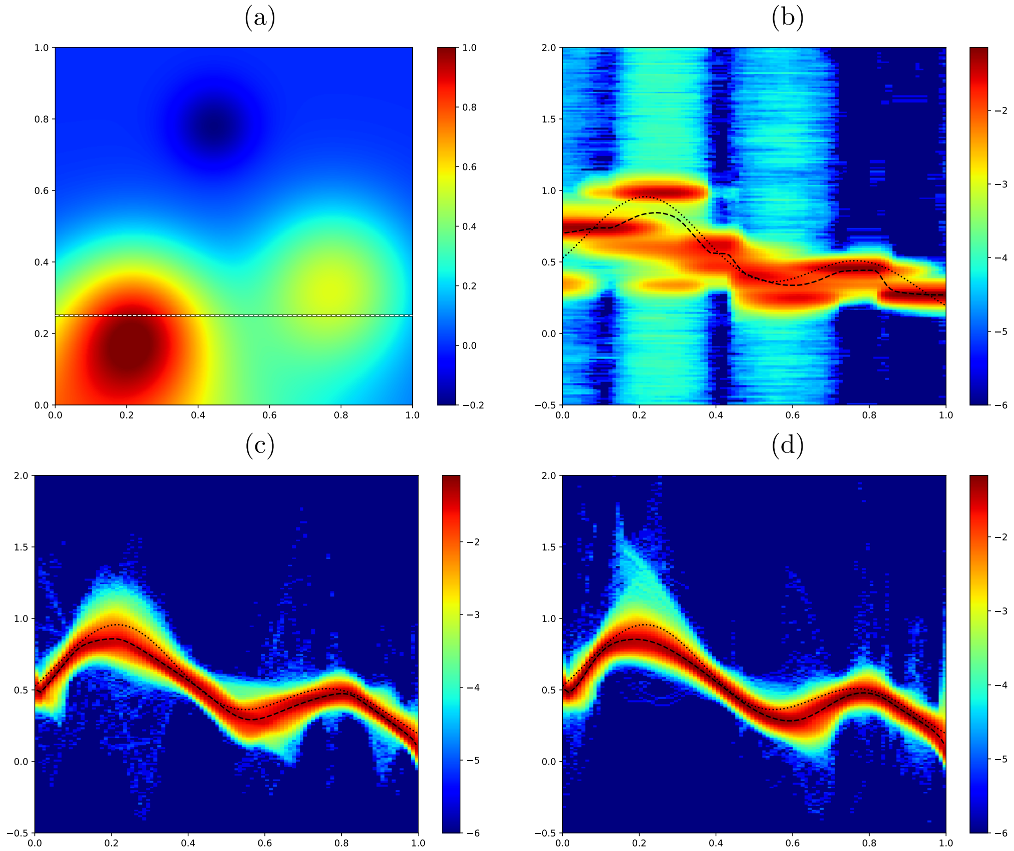

Figure 4. In (a) we show the location of a transect taken through the ensemble to show distribution of models. In (b) we show the distribution for the Voronoi cell parameterization, in (c) the Delaunay parameterization with linear interpolation, and (d) the Delaunay parameterization with Clough-Tocher interpolation. In each of the distribution plots, we show the $\log_{10}$ of the probability, the ensemble mean is plotted with a dashed line, and the true model with a dotted line.

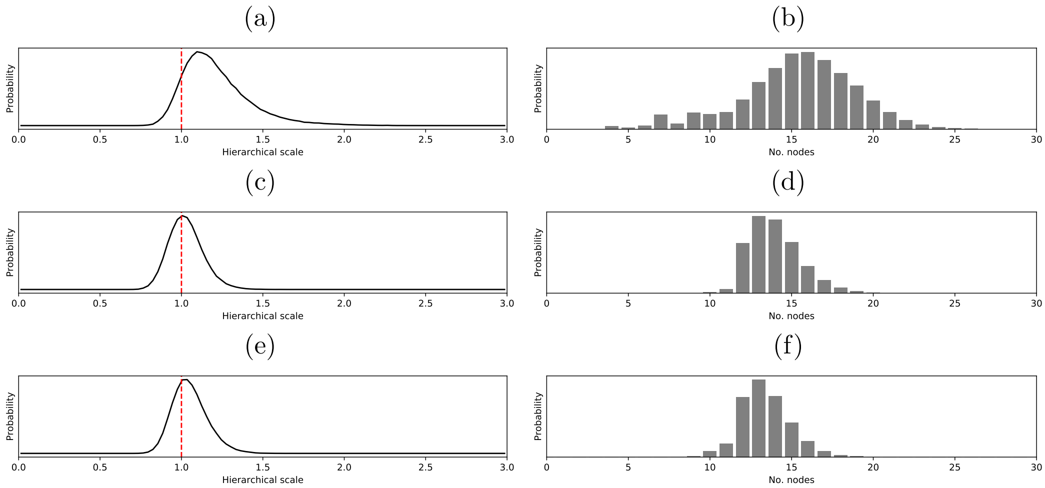

Figure 5. Posterior histograms from the inversion of the smooth data set for the hierarchical scaling factor are shown in (a) for the Voronoi parameterization, (c) for the Linear Delaunay parameterization and (e) for the Clough-Tocher Delaunay parameterization. Histograms for the number of model nodes are similarly shown in (b), (d) and (f) for the three parameterizations. The two Delaunay parameterizations in this case have hierarchical scaling factors close to one, and fewer number of model parameters.

is that the Voronoi cell parameterization has several regions where the distribution is multi-modal. For example, taking a vertical line at approximately $y = 0.25$ in Figure 4(b) would produce four peaks. Computing the standard deviation of such a multi-modal distribution would produce large values, and this is the underlying cause of the large magnitude standard deviations seen in Figure 3(b). Even though the Voronoi cell parameterization is a zeroth order discontinuous parameterization, from Figure 4(b) we can see that the ensemble mean (black dashed line) of the Voronoi model is smooth and reasonably approximates the true model (black dotted line).

Given the Voronoi parameterization is poor at representing a smooth Gaussian field, we should expect higher levels estimated for the hierarchical error scale in the Voronoi parameterization than for the Delaunay parameterizations. Recall that the level of estimated error given by the hierarchical parameter can be seen as the level of data fit achieved by the model and its parameterization. In Figure 5 we show the histograms of the number of nodes for all chains combined with the hierarchical scaling parameters for each of the inversions in (a), (c), and (e). Since this inversion is for a synthetic experiment where we know the true noise level, the hierarchical error scale should converge to approximately one when the parameterization is able to predict the observations to within noise level.

From the plots we can see that both the Delaunay parameterizations have histograms with modes of approximately one, whereas the Voronoi parameterization has a slightly higher mode and a longer tail. Also in Figure 5 we show the posterior histogram on the number of model nodes in (b), (d) and (f). We observe that the mean and the variance of the distribution is larger for Voronoi cells (b) than for Delaunay parameterization (d and f). Overall, the Voronoi parameterization uses more cells and produce a worse data fit than Delaunay parameterizations. Since the theoretical errors introduced by the poor parameterization choice of Voronoi cells in this case is non-zero, the hierarchical estimate of the error scaling term is greater than one.

In Figure 6 we show results for the inversion of of observation obtained from the discontinuous 2D field shown in Figure 2(b). Here the Voronoi cell parameterization has better recovered the true field than the two Delaunay parameterizations which only produce smooth approximations of the truth.

In this case, the standard deviation for the Voronoi cell parameterization has large values coincident with the the discontinuities in the 2D field. This is not surprising as the edges are not precisely constrained by the observations leading to uncertainty in their location which in turn will lead to a multi-modal posterior distribution proximate to true edges. As stated for the previous inversions, computing the standard deviation of a multi-modal distribution will naturally lead to a large uncertainties as shown in these results. Some authors (Burdick & Lekić, 2017; Cho, Gibson, & Zhu, 2018; Olugboji et al., 2017) have suggested that areas of large uncertainties can be used as a proxy of the location of discontinuities with models. In this synthetic example, it would appear that this in indeed a reliable proxy for the location of discontinuities, however compare this to the results for the smooth model in Figure 3(b) where we have similar large standard deviations in the inversion of a continuous model. Discontinuities in an underlying 2D field will lead to large standard deviations in the posterior, but large standard deviations do not necessarily imply discontinuities. It is a characteristic of Voronoi cell trans-dimensional inversion with mobile cells that they produce regions of multi-modal posteriors leading to ring like structures of large magnitude in maps of posterior standard deviation.

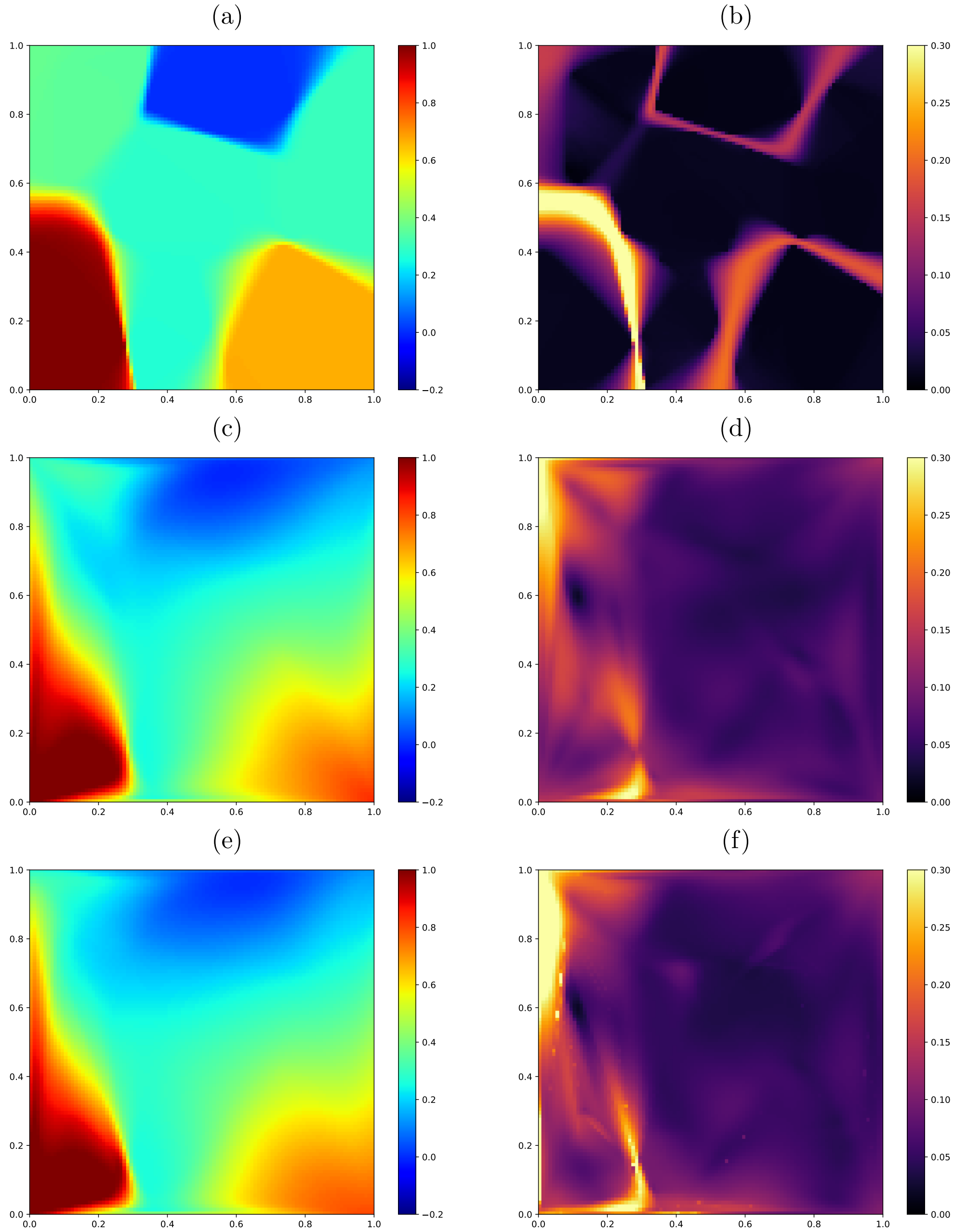

Figure 6. The summary plots for the inversion of the tessellated model. In (a) and (b) are the mean and standard deviation results for the Voronoi cell parameterization. In (c) and (d) are the mean and standard deviation results for the Delaunay triangulation with linear interpolant parameterization. In (e) and (f) are the mean and standard deviation results for the Delaunay triangulation with Clough-Tocher interpolant.

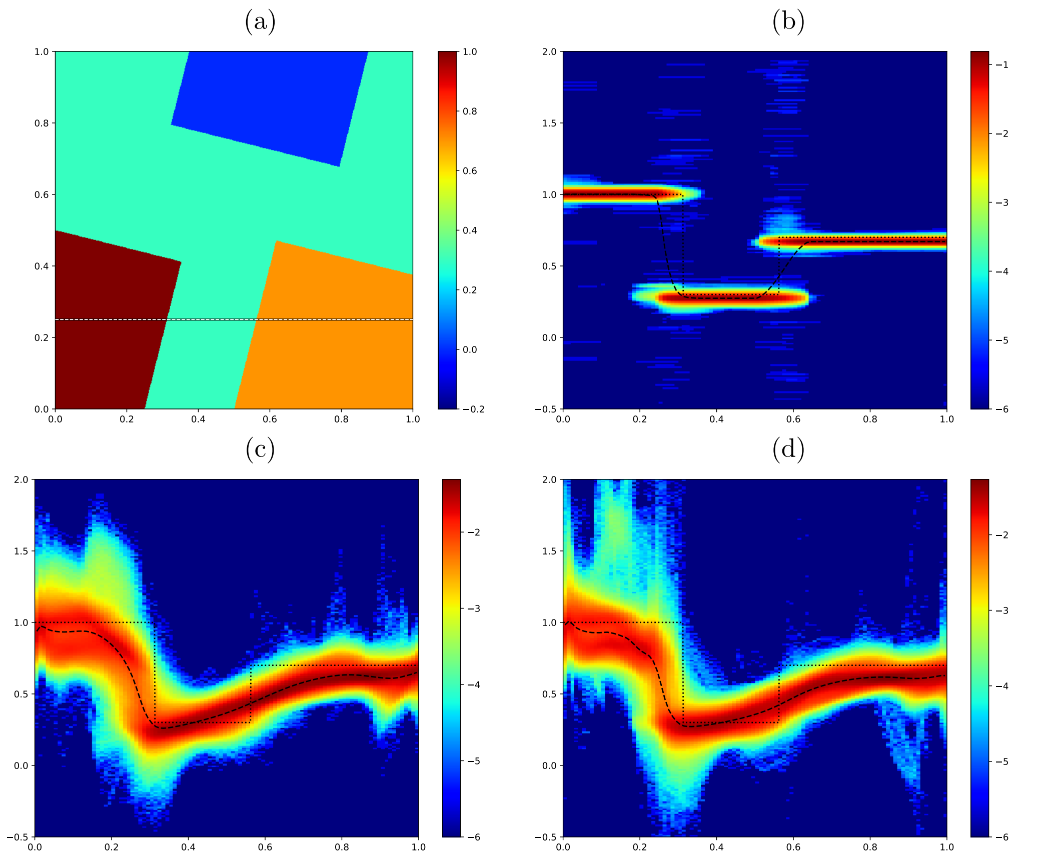

Figure 7. In (a) we show the location of a transect taken throw the ensemble to show distribution of models. In (b) we show the distribution for the Voronoi cell parameterization, in (c) the Delaunay parameterization with linear interpolation, and (d) the Delaunay parameterization with Clough-Tocher interpolation. In each of the distribution plots, we show the $\log_{10}$ of the probability, the ensemble mean is plotted with a dashed line, and the true model with a dotted line.

Once again in Figure 7 we show the posterior estimates of the reconstructed surface. In this inversion we can clearly see that the posterior for the Voronoi cell parameterization has very narrow posterior widths. The cause of the large standard deviations near discontinuities in the 2D field in Figure 6(b) can be clearly seen in this Figure 7(b) where the posterior is strongly bi-modal near discontinuities due to uncertainty in the location of discontinuities.

Both the Delaunay triangulation parameterizations have approximated the discontinuous model with a smooth function. Due to the poor ability of the parameterization to represent the true discontinuous surface, the uncertainties are much broader. It should be noted here that even though the mean model with the two Delaunay parameterizations may be a poor representation of the true discontinuous surface, the true model remains well within the higher probability region of the posterior. This is an important result: even in the case of a poorly chosen parameterization, the algorithm is able to adjust both the model complexity (number of nodes) and data uncertainty (through the scaling parameter λ) and to provide accurate surface uncertainties. If the parameterization choice is poor, there will be a corresponding increase in the estimated level of data errors (due to increased theory errors), which will be reflected in higher uncertainties in posterior estimates of the 2D field.

In Figure 8, for each of the inversions, we show posterior distribution for the hierarchical error scale parameter in (a), (c), and (e) and for the number of nodes in (b), (d) and (f). We can see that the Voronoi parameterization has fit the observations well to the level of added noise with the mode of the hierarchical scale posterior approximately one and relatively tightly constrained. In contrast, the hierarchical scale posteriors for the two Delaunay parameterizations have much larger modes and are more weakly constrained. A similar trend is observed in (b), (d) and (f) where we show the posterior histogram on the number of model nodes. The recovery of the noise level in the Voronoi cell parameterization here can generally only be achieved in synthetic tests where we know the noise model. In a real world problem where the true noise and forward modelling is more complex, such a tightly constrained result as in the Voronoi cell parameterization may not be possible due to approximations in the hierarchical error model.

Given this new general software framework for constraining 2D fields with a configurable parameterization, an obvious question arises as to which parameterization should

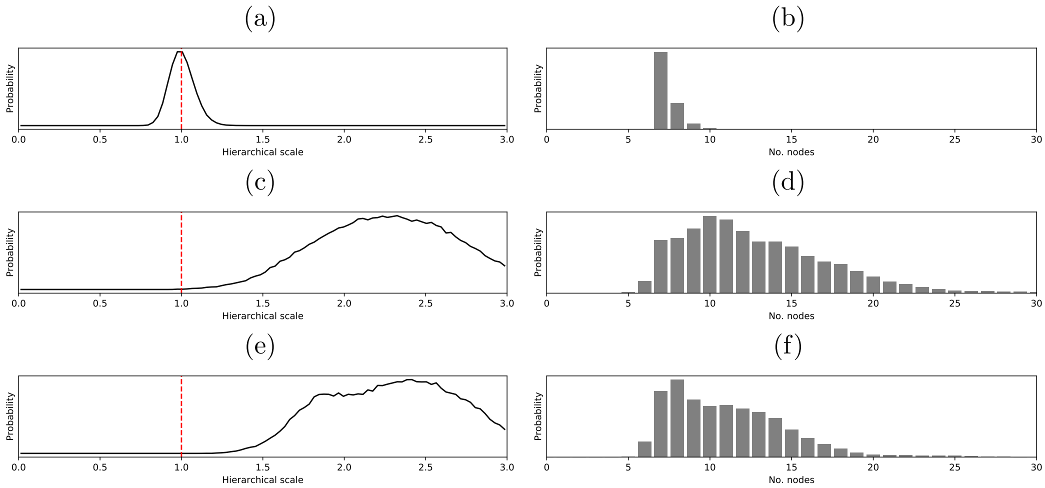

Figure 8. Posterior histograms from the inversion of the tessellated data set for the hierarchical scaling factor are shown in (a) for the Voronoi parameterization, (c) for the Linear Delaunay parameterization and (e) for the Clough-Tocher Delaunay parameterization. Histograms for the number of model nodes are similarly shown in (b), (d) and (f) for the three parameterizations. In this case, the Voronoi parameterization recovers the true level of noise and uses fewer parameters than the two Delaunay parameterizations.

be used. Many approximate criteria exist for model choice problems, however most assume a fixed number of parameters (Akaike, 1974; Schwarz, 1978). These criteria provide an approximation of Bayes factors or evidence ratios that can be used to select which model best fits our observations. The Deviance Information Criteria (DIC) has the advantage that it can be applied in trans-dimensional inversion (Hawkins & Sambridge, 2015; Steininger, Dosso, Holland, & Dettmer, 2014). The DIC variant we use for trans-dimensional inversions is given by

$$\mathrm{DIC} = \overline{D(\mathbf{m})} + \frac{1}{2}\mathrm{var}(D(\mathbf{m})), \tag{13}$$

where $D(\mathbf{m})$ is called the deviance and given by

$$D(\mathbf{m}) = -2\log p(\mathbf{m}|\mathbf{d}, \mathcal{I}) + \text{constant}, \tag{14}$$

where the constant is a function of the data and cancels for model comparison purposes. Here the mean and variance refer to the posterior expectations of the deviance which can be approximated from the Markov chain ensemble. The mean of the deviance gives a measure of the fit to the observations, where as its variance penalizes over parameterization as an over parameterized model leads to higher degrees of freedom and hence larger variance in the posterior deviance.

The attraction of this criteria is its simplicity to compute as we only need to calculate the mean and standard deviation of the ensemble negative log likelihoods. Many other criteria such as the AIC and BIC require the calculation of the maximum likelihood and the number of model parameters. In a trans-dimensional inversion, the number of model parameters is dynamic, with the maximum likelihood model likely belonging to an over-parameterization model within the ensemble.

In Table 1 we show the DICs computed for each of the parameterizations for the two inversions. We can see that in the inversion of the true smooth model, the two Delaunay parameterizations are significantly preferred to the Voronoi cell model. It is somewhat surprising that the Clough-Tocher parameterization is not preferred, however the difference between the Linear and Clough-Tocher Delaunay parameterization is small. For the tessellated true model, the preferences are reversed as expected. Before closing here, we again stress that the DIC is an approximate model comparison and is not without its limitations and criticisms. From a Bayesian perspective, the best approach for determining the support of one parameterization over the other is through computing Bayes factors (Kass & Raftery, 1995) which requires computation of the evidence which may be a future extension of this software.

4 Synthetic case study: relative sea level, absolute sea level and vertical land motion

As a further synthetic example, we now illustrate the potential of the software on a geophysical inverse problem involving three sets of disparate observations. The goal here is to estimate the relative sea level rise from a combination of tide gauges, satellite altimetry and GPS vertical land motion estimates. This problem involves reconstructing different surfaces that are either continuous and smoothly varying (absolute sea level), or have discrete transitions or sharp spatial gradients (vertical land motion). We therefore use this joint inversion as a canonical example, but many alternatives exists both within geophysics and other fields.

(a)

Parameterization	$\overline{D(\mathbf{m})}$	$\mathrm{var}(D(\mathbf{m}))$	DIC
Voronoi	-481.130	820.787	-70.736
Delaunay	-512.471	211.326	-406.808
Clough Tocher	-507.714	205.422	-405.003

(b)

Parameterization	$\overline{D(\mathbf{m})}$	$\mathrm{var}(D(\mathbf{m}))$	DIC
Voronoi	-516.945	23.317	-505.286
Delaunay	-355.045	1411.684	350.797
Clough Tocher	-352.460	1175.300	235.190

Table 1. The DIC values computed for each of the parameterizations for the inversion of the smooth and tessellated synthetic data. The lowest DIC value is the preferred model. In (a) for the smooth synthetic data, the Linear Delaunay parameterization is preferred whereas in (b) for the tessellated synthetic data the Voronoi parameterization is preferred.

Understanding sea level rise due to anthropogenic global warming has important ramifications for coastal communities which contain a large proportion of the world's population. The rates at which sea level currently changes along the coastline is determined by the local vertical land motion (mostly due to post-glacial isostatic rebound) and global sea level rise, due predominantly to melting glaciers and thermal expansion of the oceans (Cazenave & Cozannet, 2013; Church & White, 2011).

Tide gauges observing the sea level over long time periods are used for the direct measurement of relative sea level rates. However, tide gauges are subject to bias caused by man-made and natural local changes to coastlines, and by instrumental measurement errors. Their time series often have large uncertainties and associated record lengths strongly vary among stations. Deriving a comprehensive view of relative sea level change solely from tide gauges is therefore challenging.

While tide gauges measure directly the relative sea level rate, that is, the difference between absolute rates of sea level rise and vertical land motion, rates of absolute sea level have been accurately measured globally using satellite based microwave radar altimetry since 1992 (from the launch of Topex/Poseidon, followed by Jason 1 and Jason 2). However, while absolute sea level measurement from satellites are accurate in the open ocean, they are problematic near shorelines due to spurious signals from land reflections (Gommenginger et al., 2011). On land, the deployment of GNSS stations around the globe for measuring rates of vertical land motion (in addition to lateral movement) provide good constraints on recent rates of vertical land motion with a generally denser coverage than tide gauges (Blewitt, Kreemer, Hammond, & Gazeaux, 2015).

Previously, trans-dimensional Voronoi cells have been used to create maps of relative sea level rise directly from tide gauge observations (Choblet et al., 2014) and for estimating Glacial Isostatic Adjustment (GIA) from vertical land motion as inferred from GNSS (Global Navigation Satellite System) stations (Husson et al., 2018).

Here we propose a synthetic inversion to jointly construct maps of absolute sea level rise and vertical land motion from which relative sea level rise and therefore coastal inundation can be inferred, similar to previous regional studies that instead evaluated time series locally (Pfeffer & Allemand, 2016; Pfeffer, Spada, Mémin, Boy, & Allemand, 2017).

A distinction between this sea-level example and the previous theoretical example is that here we parameterize two independent 2D fields, one to represent the rate of absolute sea level rise, and the other the rate of vertical land motion, that is our model becomes

$$\mathbf{m} = \begin{bmatrix} \mathbf{m}_{\text{sea}} \\ \mathbf{m}_{\text{land}} \end{bmatrix}, \tag{15}$$

where both $\mathbf{m}_{\text{sea}}$ and $\mathbf{m}_{\text{land}}$ are trans-dimensional models and for each we can choose to use either a Voronoi or Delaunay parameterization independently. Given observations of absolute sea level change, vertical land motion and tide gauges, the likelihood can be composed

$$p(\mathbf{d}|\mathbf{m}, \mathcal{I}) = p(\mathbf{d}_{\text{sea}}|\mathbf{m}_{\text{sea}}, \mathcal{I})p(\mathbf{d}_{\text{land}}|\mathbf{m}_{\text{land}}, \mathcal{I})p(\mathbf{d}_{\text{tide}}|\mathbf{m}_{\text{sea}}, \mathbf{m}_{\text{land}}, \mathcal{I}). \tag{16}$$

Note that the relative sea level surface is not directly parameterized and is derived from the two parameterized 2D fields. This still allows posterior inference on relative sea level rise as we can compute its 2D field for each model pair in the ensemble and collect statistics as if it were parameterized independently.

For the choice of parameterization for the sea model, at long scale lengths, the rate of sea level change at annual time scales is spatially smooth and predominantly correlated with latitude. This strongly suggests that either of the two Delaunay parameterizations should be used to represent absolute sea level rise.

In contrast, the choice of parameterization for the land model is less evident. The uplift of land is a combination of generally smooth variation cause by deformational processes, but with strong lateral variations or discontinuities near active faults. If we assume that the gradual variation is small compared to the magnitude of the discontinuities near faults, then a Voronoi parameterization would be appropriate in tectonically active regions. Either of the two Delaunay parameterizations would be more suited in tectonically quiet regions.

In order to test the effect of different parameterization options, we set out to create a synthetic data set for sea level and vertical land motion rates using the region of Tasmania. In the following, “sea model” indicates absolute sea level rate (as observed by satellite altimetry), “land model” indicates vertical land motion (as measured by GNSS stations) and “tide gauge” indicates relative sea level rate (as measured by tide gauge stations). For the sea model, the rate of absolute sea level rise is set to a smooth function of latitude between 0 and 4 mm/year. In order to sub sample the continuous field into an irregular set of observations, sea level rate observations were created by generating random points in the ocean more than 10 km away from the coast line, sampling the true (yet synthetic) sea level rate and adding independent Gaussian noise with a standard deviation of 1 mm/year.

For the synthetic land model, we created a fictitious fault running diagonally down the center of Tasmania with a small negative uplift rate (-2.0 mm/year) on the western side and a larger positive uplift rate (5.0 mm/year) to the east. Observations are generated using random points on land to which we add independent Gaussian noise with a standard deviation of 1 mm/year. Lastly, we simply subtract the true sea model from the true land model to obtain the tide gauge model and create observations randomly located on the coast of Tasmania and add the same level of Gaussian noise. The syn-

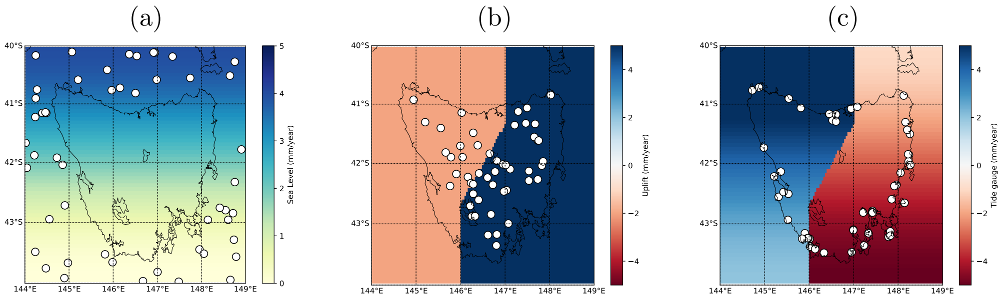

Figure 9. In (a) and (b) we show the true synthetic sea level rise and vertical land motion in mm/year. The derived tide gauge image is shown in (c). On each of the plots are shown the location of the randomly generated observations with small circles, that is the location of sea level observations are shown in (a), GPS observations in (b) and tide gauges in (c).

thetic models are shown in Figure 9 in addition to the observations indicated with small circles of which there are 50 of each type for a total of 150 in this data set.

For this inversion, similar to the previous example, we use 28 Markov chains and 4 temperatures logarithmically spaced between 1 and 5. Initial models are randomly generated and chains are simulated for 1 million iterations with 500,000 removed as burn in. A primary difference here is that we use three independent hierarchical error scaling terms, one for each class of observation, namely sea level altimetry, land based GPS, and tide gauge.

For a first test, we invert the observations using a Voronoi cell parameterization for both the sea and land models with the results shown in Figure 10. As is to be expected, the choice of the Voronoi cell parameterization for the sea level rate is a poor one that introduces large uncertainty in the sea level reconstruction in (b), which is then propagated to the derived tide gauge uncertainty in (f). For the recovery of the land model, the Voronoi cell parameterization is effective with high regions of uncertainty in the map restricted to the sea where there is no data, and along the fault where uncertainty in the faults location leads to multi-modality and therefore high standard deviation.

If we instead parameterize the sea with the Delaunay triangulation with linear interpolation, the results improve as shown in Figure 11. In (b) the standard deviation map of the sea level is lower, more homogeneous, and free of large magnitude standard deviations caused by mobile Voronoi cells. This lower uncertainty propagates to the tide

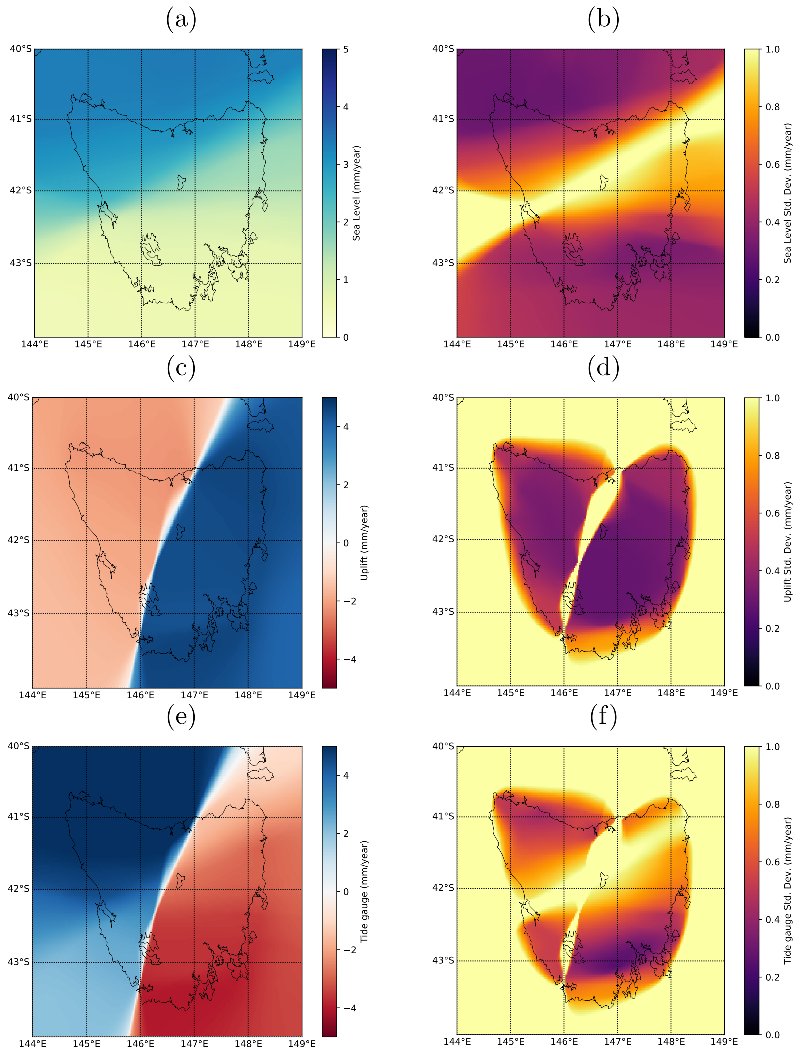

Figure 10. In (a), (c) and (e) we show the ensemble means of the rates of absolute sea level change (sea), vertical land motion (land) and relative sea level change (tide gauge) in mm/year, when using a Voronoi cell parameterization for both the sea and land model. The corresponding standard deviation maps are shown in (b), (d), and (f). Large uncertainties in the tide gauge standard deviation are caused by using the Voronoi cell parameterization to represent the sea level.

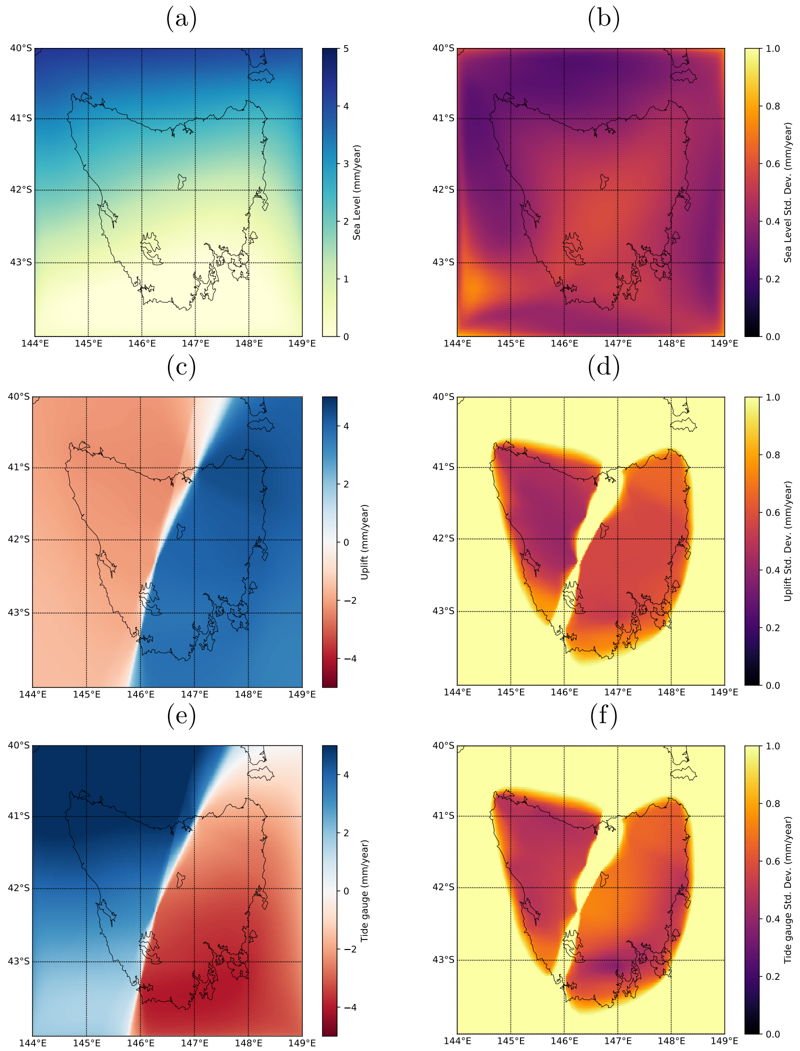

Figure 11. In (a), (c) and (e) we show the ensemble means of the rates of absolute sea level change (sea), vertical land motion (land) and relative sea level change (tide gauge) in mm/year, when using the Delaunay parameterization with linear interpolation for the sea model and Voronoi parameterization for the land. The maps of standard deviation are shown in (b), (d), and (f).

gauge uncertainty shown in (f) where the remaining regions of high uncertainty are due to high uncertainty from the land model in the sea and along the fault.

To explore the posterior of the relative sea level, we can generate a series of virtual tide gauges evenly spaced along the coast line and show the posterior along the coast as done by Choblet et al. (2014). Note again that this is not a 2D field parameterized in this inversion but one that can be inferred directly from the parameterized 2D absolute sea level and vertical land motion surfaces. In Figure 12(a), we show a generated a set of points running counter clockwise around the coastline of Tasmania starting just south of Hobart (the point marked as 0).

In Figure 12(b), the posterior histogram is shown for the inversion where the Voronoi parameterization was used for both the sea and land model. Compared to the previous example in Section 3, this posterior displays less multi-modality. Between Hobart and Bicheno we can see a strong constant signal in the tide gauge due to the Voronoi parameterization fitting the gradual change in sea level along the East coast with a constant function. In this case, the true model shown with a black dotted line is outside the regions of highest posterior probability, although still within the more broader region of probable models and therefore within uncertainties.

In Figure 12(c), the corresponding results for the inversion using the Delaunay triangulation parameterization with linear interpolant are shown. In contrast to the previous results, the median and uncertainties more faithfully track the true model. This is primarily due to the Delaunay parameterization being able to better model the gradual variation in sea level rates along the East and West coasts. Inversions were also performed with the Clough-Tocher interpolant with indistinguishable results relative to the linear interpolant for this problem and these are not shown for brevity. This is due to the relatively simple structure of the sea level which is equally well represented by linear and cubic interpolants given the level of noise.

5 Discussion

We have a presented a new inversion software to constrain 2D fields using Bayesian trans-dimensional sampling, and incorporating hierarchical error estimation, Hamiltonian Monte Carlo and Parallel Tempering. A novel aspect of this software is the choice of alternate parameterizations rather than the commonly used Voronoi cell parameterization. These alternate parameterizations are Delaunay triangulation with linear interpolation, and Delaunay triangulation with Clough-Tocher interpolation. In contrast to Voronoi cells, which produce a discontinuous 2D field, these alternative parameterizations produce C^0 and C^1 continuous 2D fields respectively, and will allow applications to problems where spatial gradients are required in either forward modelling or posterior inferences.

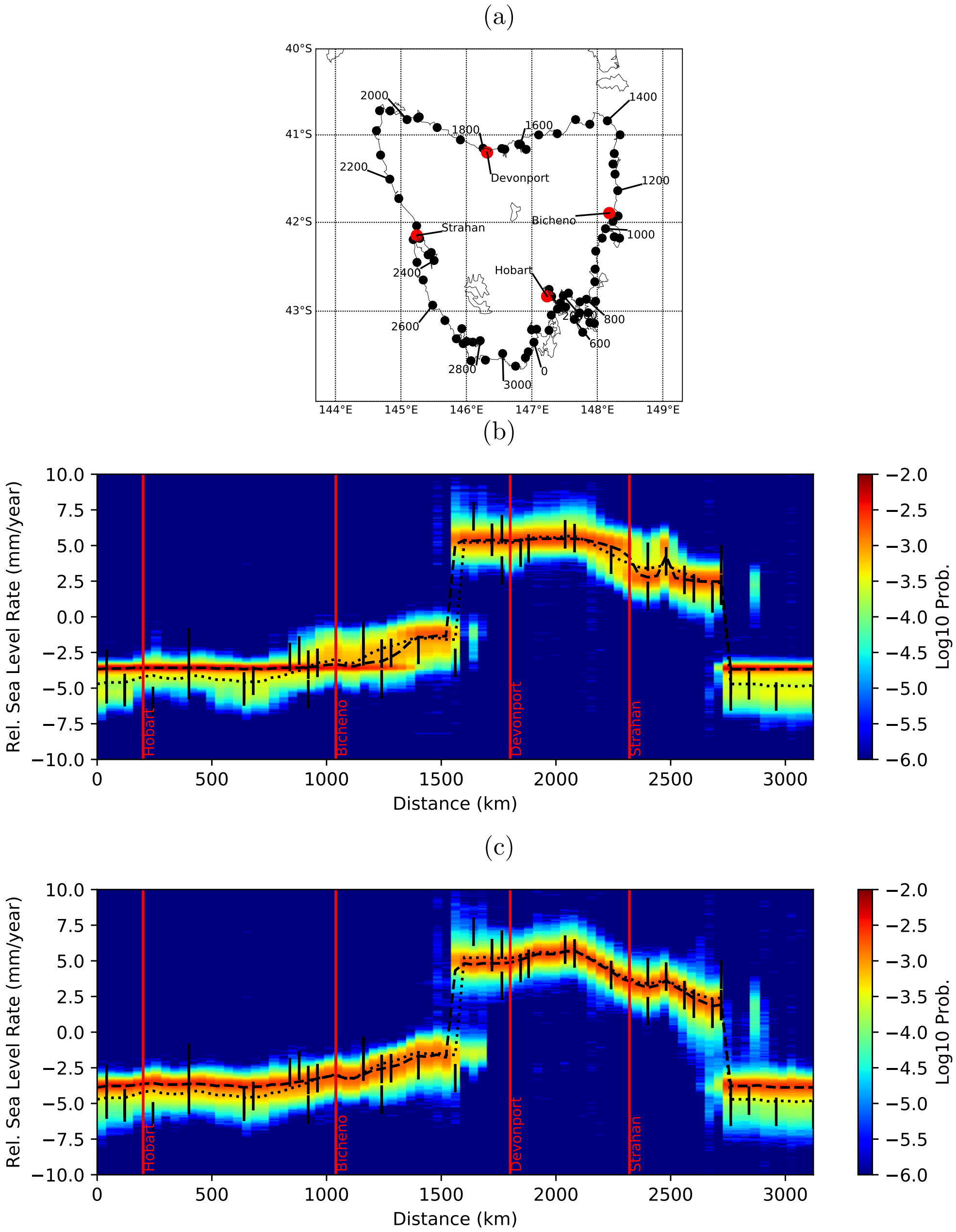

Figure 12. In (a) we show the set of evenly spaced virtual tide gauge points starting at point 0 south of Hobart and traversing the island in a counter clockwise sense. Selected points are marked with their approximate coastal distance in kilometers. In (b) we show the posterior distribution of the virtual tide gauges for the inversion with the Voronoi cell parameterization for the sea model and in (c) with the Delaunay triangulation parameterization with linear interpolant for the sea model. Both results use the Voronoi cell parameterization for the land model. In (b) and (c), the black dotted line is the true model, and the black dashed line is the median of the ensemble.

We showed in synthetic regression tests that the choice of parameterization is important as it both strongly affects the form of the posterior solution and the ability to recover true models. We also stress that even in cases where the parameterization choice is poor, the combination of trans-dimensional sampling and hierarchical error scaling ensures that while the posterior may contain poorer fitting models, the uncertainty estimates will be higher and correctly estimated. For the Voronoi cell parameterization, we showed that maps of standard deviation depict ring like structures with large magnitudes that are caused by multi-modal posteriors, and which can be difficult to interpret. Some have suggested these anomalies can be used as proxies for the location of discontinuities (Burdick & Lekić, 2017; Cho et al., 2018; Olugboji et al., 2017), however they similarly appear in synthetic tests of well converged posteriors when inverting known purely smooth models, regardless of the number of chains (see Appendix B). Hence we urge caution this interpretation of standard deviations: high standard deviations/multi-modalities in a Voronoi cell based inversions is a necessary but not sufficient condition for the existence of a discontinuity.

In contrast, inversions using the two alternate Delaunay triangulation parameterizations exhibit far less propensity for multi-modal posteriors, even when these parameterizations are ill suited such as for the discontinuous 2D field regression example. This leads to a more easily interpretable posterior. In some geophysical problems, the 1st or 2nd spatial derivative of the model may be important, in which case the Voronoi cell parameterization is inappropriate. The availability of this new software with these alternate parameterizations will open up trans-dimensional sampling to a wider variety of geophysical inverse problems and also to fields beyond geosciences.

In a trans-dimensional inversion, it may seem surprising that the parameterization can change the result of an inversion. For example, in the smooth model inversion, if the model dimension can change, why does a Voronoi cell parameterization not simply construct models with a large number of cells to approximate a smooth model? The key here is that there is not enough information in the observations to constrain the required number of Voronoi cells to accurately reflect the smooth model. This contributes to the theoretical errors of the formulation as a "parameterization error" which is to some degree approximated for in the hierarchical error estimation. If the inversions were simulated by fixing the noise to the true value, the results would provide better fits to the models than shown. However, this type of inversion assumes perfect knowledge of the observational and theory errors which is generally not the case in geophysical inverse problems.

In many geophysical problems, continuous models may be more appropriate than Voronoi cells. The poor representation of continuous fields of the Voronoi cell parameterization was a primary motivation for the development of Trans-dimensional trees (Hawkins & Sambridge, 2015). Similar to this software, the trans-dimensional tree approach with a wavelet parameterization requires a choice of which wavelet basis to choose and results in similar large uncertainties when poor choices are made. At this stage, running separate inversions and using approximate model choice criteria such as the DIC (Spiegelhalter, Best, Carlin, & van der Linde, 2002) to select which parameterization is better supported by the data seems a pragmatic albeit imperfect solution. Approximate criteria are not without their limitations and ultimately this could be resolved by accurate calculation of the evidence to compute Bayes factors, or by introducing some trans-dimensional method to propose local model parameterization changes.

A Proposal Details

A.1 Move proposals

Move proposals are standard McMC proposals using a Metropolis-Hastings (Hastings, 1970; Metropolis et al., 1953) rule. The proposal moves one node/vertex at a time and uses a Gaussian perturbation of the point so that the proposal density is

$$Q(\mathbf{m} \rightarrow \mathbf{m}') = \frac{1}{k} N(0, \sigma_x) N(0, \sigma_y) \tag{A.1}$$

where k is the number of nodes/vertices, and σ_x, σ_y are the standard deviations of the perturbations of the x and y coordinates of the node. Since the normal distribution is symmetric, the proposal ratio in the acceptance criteria will cancel leaving the prior ratio and likelihood ratio. Furthermore, we use a uniform prior for the positions of the nodes and therefore this also cancels leaving the acceptance criteria for move proposals as simply the likelihood ratio, that is

$$\alpha_{\text{move}}(\mathbf{m} \rightarrow \mathbf{m}') = \min \left\{ 1, \frac{p(\mathbf{d}|\mathbf{m}', \mathcal{I})}{p(\mathbf{d}|\mathbf{m}, \mathcal{I})} \right\}. \tag{A.2}$$

A.2 Hamiltonian Steps

Hamiltonian Monte Carlo proposals uses an auxiliary variable technique and calculation of the gradient of the posterior to generate new model proposals far away from the current model (Duane et al., 1987; Fichtner & Simutė, 2018; Neal, 2011; Sen & Biswas, 2017).

The auxiliary variable, $\mathbf{p}$, is analogous to momentum in a Hamiltonian dynamical system

$$\mathcal{H}(\mathbf{m}, \mathbf{p}) = U(\mathbf{m}) + K(\mathbf{p}), \tag{A.3}$$

where U is the potential function of the current model $\mathbf{m}$, and K the kinetic energy function of the momentum $\mathbf{p}$. The potential function is given by

$$U(\mathbf{m}) = -\log p(\mathbf{d}|\mathbf{m}, \mathcal{I})p(\mathbf{m}|\mathcal{I}), \tag{A.4}$$

that is, the negative log of the posterior and the kinetic energy function by

$$K(\mathbf{m}) = \frac{\mathbf{p}^T\mathbf{M}\mathbf{p}}{2}, \tag{A.5}$$

where $\mathbf{M}$ is the mass matrix. Recent advances in HMC (Fichtner et al., 2019) have shown that this mass matrix can be optimized to provide better sampling in fixed dimension inversions. In trans-dimensional sampling, the number of model parameters and hence the size of this mass matrix changes during the inversion so we have elected to use an identity matrix here. The approach implemented here could be improved further by further research into adapting optimal mass matrices trans-dimensional sampling.

A Hamiltonian Monte Carlo proposal samples an initial momentum vector $\mathbf{p}$ from a multi-dimensional normal distribution with zero mean and unit standard deviation. The Hamiltonian dynamical system is simulated for a configured number of steps with a tunable step size to obtain a proposed model $\mathbf{m}'$ and momentum $\mathbf{p}'$. This simulation requires the gradient of the potential function, $\frac{\partial U}{\partial \mathbf{m}}$, and to ensure that the proposal is reversible, the leap frog method is generally used (Neal, 2011).

In the cases of the three parameterizations used, Hamiltonian proposals only perturb the values at each of the Voronoi cell nodes/Delaunay triangulation vertices and not their locations. This is due to the fact that the gradient of the likelihood with respect to the location of the cells is undefined for Voronoi cells.

Once a proposed model is obtained, it is accepted or rejected according to the criteria

$$\alpha_{\mathrm{hmc}}(\mathbf{m} \rightarrow \mathbf{m}') = \min \left\{1, \exp\left(-\mathcal{H}(\mathbf{m}, \mathbf{p}) + \mathcal{H}(\mathbf{m}', \mathbf{p}')\right)\right\}. \tag{A.6}$$

The requirement of needing the gradient of the posterior with respect to the model parameter values may be prohibitive to compute or not available in some cases. The software frame also supports standard McMC proposal for change of values and the acceptance criteria for value proposals in this case is similar to the move proposal above.

A.3 Trans-dimensional Steps

In this software, trans-dimensional steps involve adding and remove nodes and their values. For simplicity, we have chosen to use the "birth from the prior" approach. In this approach, for a Birth proposal where a new node is generated, the new location and value are sampled from the prior. This means that the proposal density for a Birth proposal is

$$Q(\mathbf{m} \rightarrow \mathbf{m}') = p(x')p(y')p(z') \tag{A.7}$$

where x' and y' are the new node coordinates and z' the new node value. This strategy simplifies the Birth and Death acceptance criteria in two ways: first the proposal ratio cancels with the prior ratio and secondly the Jacobian is unity. This leaves an acceptance criteria for birth/death proposals as simply the likelihood ratio, that is

$$\alpha_{\mathrm{birth/death}} = \min \left\{1, \frac{p(\mathbf{d}|\mathbf{m}', \mathcal{I})}{p(\mathbf{d}|\mathbf{m}, \mathcal{I})}\right\}. \tag{A.8}$$

A.4 Hierarchical proposals

For hierarchical error scaling proposals, we use a standard McMC proposal with a normal distribution perturbation of λ and the acceptance criteria therefore is

$$\alpha_\lambda(\lambda \to \lambda') = \min\left\{1, \frac{p(\lambda')}{p(\lambda)} \frac{p(\mathbf{d}|\mathbf{m}', \mathcal{I})}{p(\mathbf{d}|\mathbf{m}, \mathcal{I})}\right\}, \tag{A.9}$$

where $p(\lambda)$ is the prior on the hierarchical scaling factor.

A.5 Parallel tempering

During a Parallel Tempering exchange proposal, two parallel chains are chosen at random to exchange their models. The acceptance criteria is

$$\alpha_{\text{exchange}}(\mathbf{m}_i \leftrightarrow \mathbf{m}_j) = \min\left\{1, \left[\frac{p(\mathbf{d}|\mathbf{m}_j, \mathcal{I})}{p(\mathbf{d}|\mathbf{m}_i, \mathcal{I})}\right]^{\frac{1}{T_i}} \left[\frac{p(\mathbf{d}|\mathbf{m}_i, \mathcal{I})}{p(\mathbf{d}|\mathbf{m}_j, \mathcal{I})}\right]^{\frac{1}{T_j}}\right\} \tag{A.10}$$

where i and j subscripts indicate the two different chains.

B Stability of variance estimates

In this short appendix we show the variance estimate for the regression problem of the smooth model with Voronoi cell parameterization under different configurations of the inversion. In Figure B.1(a) the we show the result from a single chain, (b) the same number of chains as the main result (28) but we use 10 million iterations instead of 1 million, (c) 56 chains and (d) 112 chains. In each case, the same pattern of variance as presented in Figure 3(b) is recovered indicating that our presented results are well converged and provide robust estimate of the posterior variance. Hence the posterior multi-modalities as discussed are robust features and not due to poor convergence.

Acknowledgments

Portions of this software development were funded by AuScope through the National Collaborative Research Infrastructure Strategy (NCRIS), an Australian Federal Government programme.

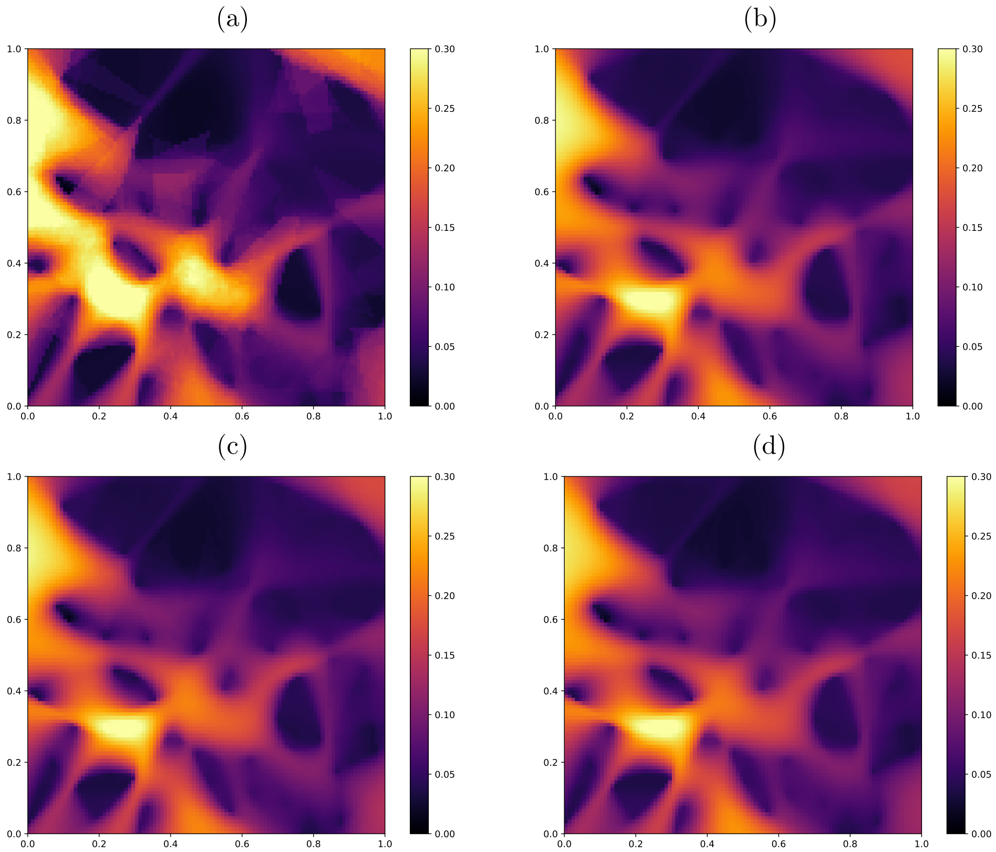

Figure B.1. A comparison of the the standard deviation obtained from differently configure inversions. In (a) we invert a single chain, (b) we use 28 independent chains as in the main body but simulate 10 million steps, in (c) we use 56 chains and (d) 112 chains. In each case the estimated standard deviation is in agreement with results presented in Figure 3(b)

RH and TB are funded by the European Unions Horizon 2020 research and innovation programme under Grant Agreement No. 716542.

The authors kindly thank Nicklas Linde, Andreas Fichtner and an anonymous reviewer for the constructive comments that helped to improve this manuscript.

The TransTesselate2D software is available from http://www.iearth.org.au

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *AC-19*(6), 716-723.

Al-Awadhi, F., Hurn, M., & Jennison, C. (2004). Improving the acceptance rate of reversible jump MCMC proposals. *Statistics and Probability Letters*, *69*, 189 - 198.

Backus, G. (1970a). Inference from inadequate and inaccurate data I. *Proceedings of the National Academy of Sciences*, *65*(1), 1-7.

Backus, G. (1970b). Inference from inadequate and inaccurate data II. *Proceedings of the National Academy of Sciences*, *65*(2), 281-287.

Backus, G. (1970c). Inference from inadequate and inaccurate data III. *Proceedings of the National Academy of Sciences*, *67*(1), 282-289.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, *53*, 370-418.

Belhadj, J., Romary, T., Gesret, A., Noble, M., & Figliuzzu, B. (2018). New parameterizations for bayesian seismic tomography. *Inverse Problems*, *34*(6).

Billings, S., Beatson, R., & Newsam, G. (2002). Interpolation of geophysical data using continuous global surfaces. *Geophysics*, *67*(6), 1810 - 1822.

Blewitt, G., Kreemer, C., Hammond, W. C., & Gazeaux, J. (2015). MIDAS robust trend estimator for accurate gps station velocities without step detection. *Journal of Geophysical Research: Solid Earth*, *121*(3), 2054-2068.

Bodin, T., Salmon, M., Kennett, B. L. N., & Sambridge, M. (2012). Probabilistic surface reconstruction from multiple data sets: An example for the Australian Moho. *Journal of Geophysical Research*, *117*, B10307. doi: 1029/2012JB009547

Bodin, T., & Sambridge, M. (2009). Seismic tomography with the reversible jump algorithm. *Geophysical Journal International*, *178*, 1411 - 1436.

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, L., & Rawlinson, N. (2012). Trans-dimensional inversion of receiver functions and surface wave dispersion. *Journal of Geophysical Research*, *117*, B02301. doi: 10.1029/2011JB008560

Brooks, S., Gelman, A., Jones, G. L., & Meng, X. (Eds.). (2011). *Handbook of markov chain monte carlo*. Chapman and Hall/CRC.

Brunetti, C., Linde, N., & Vrugt, J. A. (2017). Bayesian model selection in hydrogeophysics: Application to conceptual subsurface models of the South Oyster Bacterial Transport Site, Virginia, USA. *Advances in Water Resources*, *102*, 127 - 141.

Burdick, S., & Lekić, V. (2017). Velocity variations and uncertainty from transdimensional P-wave tomography of North America. *Geophysical Journal International*, *209*(2), 1337 - 1351.

Cazenave, A., & Cozannet, G. L. (2013). Sea level rise and its coastal impacts. *Earth's Future*, *2*, 15 - 34.

Chiao, L., & Kuo, B. (2001). Multiscale seismic tomography. *Geophysical Journal International*, *145*, 517 - 527.

Cho, Y., Gibson, R. L., & Zhu, D. (2018). Quasi 3D transdimensional Markov-chain Monte Carlo for seismic impedance inversion and uncertainty analysis. *Interpretation*, *6*(3), T613 - T624. doi: 10.1190/INT-2017-0136.1

Choblet, G., Husson, L., & Bodin, T. (2014). Probabilistic surface reconstruction of coastal sea level rise during the twentieth century. *Journal of Geophysical Research: Solid Earth*, *119*(12), 9206 - 9236. doi: 10.1002/2014JB011639

Church, J. A., & White, N. J. (2011). Sea-level rise from the late 19th to the early 21st century. *Surv. Geophys.*, *32*, 585 - 602.

Clough, R. W., & Tocher, J. L. (1965). Finite element stiffness matrics for analysis of plate bending. *Proceedings of Conference on Matrix Methods in Structural Analysis*.

Davies, J. H. (2013). Global map of solid Earth surface heat flow. *Geochemistry Geophysics Geosystems*, *14*(10), 4608 - 4622. doi: 10.1002/ggge.20271

Denison, D. G. T., Holmes, C. C., Mallick, B. K., & Smith, A. F. M. (2002). *Bayesian methods for non-linear classification and regression.* John Wiley and Sons.

de Pasquale, G., & Linde, N. (2017). On structure-based priors in Bayesian geophysical inversion. *Geophysical Journal International*, *208*(3), 1342 - 1358.

Dettmer, J., Benavente, R., Cummins, P. R., & Sambridge, M. (2014). Trans-dimensional finite-fault inversion. *Geophysical Journal International*, *199*, 735-751.

Dettmer, J., Dosso, S. E., & Holland, C. W. (2011). Sequential trans-dimensional Monte Carlo for range-dependent geoacoustic inversion. *J. Acoust. Soc. Am.*, *129*(4), 1794-1806.

Dettmer, J., Hawkins, R., Cummins, P. R., Hossen, J., Sambridge, M., Hino, R., & Inazu, D. (2016). Tsunami source uncertainty estimation: The 2011 Japan tsunami. *Journal of Geophysical Research: Solid Earth*, *121*, 4483 - 4505. doi: 10.1002/2015JB012764

Dettmer, J., Molnar, S., Steininger, G., Dosso, S. E., & Cassidy, J. F. (2012). Trans-dimensional inversion of microtremor array dispersion data with hierarchical autoregressive error models. *Geophysical Journal International*, *188*, 719 - 734.

Dosso, S. E., Holland, C. W., & Sambridge, M. (2012). Parallel tempering in strongly nonlinear geoacoustic inversion. *Journal of the Acoustic Society of America*, *132*(5), 3030-3040.

Dosso, S. E., & Wilmut, M. J. (2006). Data uncertainty estimation in matched-field geoacoustic inversion. *IEEE Journal of Oceanic Engineering*, *31*(2), 470 - 479.

Duane, S., Kennedy, A. D., Pendelton, B. J., & Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, *195*(2), 216 - 222.

Earl, D. J., & Deem, M. W. (2005). Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, *7*(23), 3910-3916.

Fichtner, A., & Simutė, S. (2018). Hamiltonian monte carlo inversion of seismic sources in complex media. *Journal of Geophysical Research: Solid Earth*, *123*(4), 2984 - 2999. doi: 10.1002/2017JB015249

Fichtner, A., Zunini, A., & Gebraad, L. (2019). Hamiltonian Monte Carlo solution of tomographic inverse problems. *Geophysical Journal International*, *216*(2), 1344 - 1363. doi: 10.1093/gji/ggy496

Galetti, E., Curtis, A., Baptie, B., Jenkins, D., & Nicolson, H. (2016). Transdimensional Love-wave tomography of the British Isles and shear-velocity structure of the East Irish Sea Basin from ambient-noise interferometry. *Geophysical Journal International*. doi: 10.1093/gji/ggw286

Galetti, E., Curtis, A., Meles, G. A., & Baptie, B. (2015). Uncertainty loops in travel-time tomography from nonlinear wave physics. *Physical Review Letters*, *114*, 148501. doi: 10.1103/PhysRevLett.114.148501

Gao, C., & Lekić, V. (2018). Consequences of parametrization choices in surface wave inversion: insights from transdimensional Bayesian methods. *Geophysical Journal International*, *215*(2), 1037 - 1063.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457 - 472.

Geyer, C. J., & Møller, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, *21*, 359-373.

Gommenginger, C., Thibaut, P., Fenoglio-Marc, L., Quartly, G., Deng, X., Gómez-Enri, J., … Gao, Y. (2011). Coastal altimetry. In S. Vignudelli, A. Kostianoy, & P. Cipollini (Eds.), (chap. Retracking Altimeter Waveforms Near the Coasts). Springer.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*(4), 711 - 732.

Hanke, M. (1996). Limitations of the L-curve method in ill-posed problems. *BIT Numerical Mathematics*, *36*(2), 287-301.

Hansen, P. C. (1999). *The l-curve and its use in the numerical treatment of inverse problems.* Department of Mathematical Modelling, Technical University of Denmark.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97 - 109.

Hawkins, R., Brodie, R., & Sambridge, M. (2017). Bayesian trans-dimensional inversion of Airborne Electromagnetic 2D conductivity profiles. *Exploration Geophysics*. doi: 10.1071/EG16139

Hawkins, R., & Sambridge, M. (2015). Geophysical imaging using trans-dimensional trees. *Geophysical Journal International*, *203*(2), 972-1000. doi: 10.1093/gji/ggv326

Hopcroft, P. O., Gallagher, K., & Pain, C. C. (2007). Inference of past climate from borehole temperature data using Bayesian reversible jump Markov chain Monte Carlo. *Geophysical Journal International*, *171*(3), 1430-1439. doi: 10.1111/j.1365-246X.2007.03596.x

Hopcroft, P. O., Gallagher, K., & Pain, C. C. (2009). A bayesian partition modelling approach to resolve spatial variability in climate records from borehole temperature inversion. *Geophysical Journal International*, *178*(2), 651-666. doi: 10.1111/j.1365-246X.2009.04192.x

Husson, L., Bodin, T., Spada, G., Choblet, G., & Comé, K. (2018). Bayesian surface reconstruction of geodetic uplift rates: mapping the global fingerprint of gia. *Journal of Geodynamics*, *122*, 25 - 40.

Ingham, E. M., Heslop, D., Roberts, A. P., Hawkins, R., & Sambridge, M. (2014). Is there a link between geomagnetic reversal frequency and paleointensity? a bayesian approach. *Journal of Geophysical Research: Solid Earth*, *119*(7), 5290 - 5304. doi: 10.1002/2014JB010947

Inoue, H., Fukao, Y., Tanabe, K., & Ogata, Y. (1990). Whole mantle P-wave travel time tomography. *Phys. Earth Planet. Inter.*, *59*, 294 - 328.

Kárason, H., & van der Hilst, R. D. (2001). Tomographic imaging of the lowermost mantle with differential times of refracted and diffracted core phases (PKP,P_{diff}). *Journal of Geophysical Research*, *106*(B4), 6569-6587.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773-795.

Kolb, J. M., & Lekić, V. (2014). Receiver function deconvolution using transdimensional hierarchical Bayesian inference. *Geophysical Journal International*, *197*(3), 1719 - 1735.

Lochbühler, T., Vrugt, J. A., Sadegh, M., & Linde, N. (2015). Summary statistics from training images as prior information in probabilistic inversion. *Geophysical Journal International*, *201*(1), 157-171.

Malinverno, A. (2002). Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. *Geophysical Journal International*, *151*, 675-688.

Malinverno, A., & Briggs, V. A. (2004). Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. *Geophysics*, *69*(4), 1005 - 1016. doi: 10.1190/1.1778243

Mann, S. (1998). *Cubic precision Clough-Tocher interpolation* (Tech. Rep. No. CS-98-15). Computer Science Department, University of Waterloo.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1986-1992.

Mosegaard, K., & Tarantola, A. (1995). Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research*, *100*(B7), 12431-12447.

Neal, R. M. (1994). An improved acceptance procedure for the hybrid Monte Carlo algorithm. *Journal of Computational Physics*, *111*, 194 - 203.

Neal, R. M. (2011). Handbook of Markov chain Monte Carlo. In (chap. McMC using Hamiltonian Dynamics). Chapman and Hall/CRC.

Oliver, M. A., & Webster, R. (1990). Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information Systems*, *4*(3), 313-332.

Olugboji, T. M., Lekic, V., & McDonough, W. (2017). A statistical assessment of seismic models of the U.S. continental crust using Bayesian inversion of ambient noise surface wave dispersion data. *Tectonics*, *36*.

Pfeffer, J., & Allemand, P. (2016). The key role of vertical land motions in coastal sea level variations: A global synthesis of multisatellite altimetry, tide gauge data and GPS measurements. *Earth and Planetary Science Letters*, *439*, 39 - 47.

Pfeffer, J., Spada, G., Mémin, A., Boy, J.-P., & Allemand, P. (2017). Decoding the origins of vertical land motions observed today at coasts. *Geophysical Journal International*, *210*, 148 - 165.

Piana Agostinetti, N., Giacomuzzi, G., & Malinverno, A. (2015). Local 3D earthquake tomography by trans-dimensional Monte Carlo sampling. *Geophysical Journal International*, *201*(3), 1598-1617.

Piana Agostinetti, N., & Malinverno, A. (2010). Receiver function inversion by trans-dimensional Monte Carlo sampling. *Geophysical Journal International*, *181*(2), 858-872.

Pjipers, F. P., & Thompson, M. J. (1992). Faster formulations of the optimally localized averages method for helioseismic inversions. *Astronomy and Astrophysics*, *262*, L33 - L36.

Rawlinson, N., Fichtner, A., Sambridge, M., & Young, M. (2014). Advances in geophysics. In (Vol. 55, chap. Seismic tomography and the assessment of uncertainty). Academic Press.

Ray, A., & Key, K. (2012). Bayesian inversion of marine CSEM data with a trans-dimensional self parametrizing algorithm. *Geophysical Journal International*, *191*, 1135 - 1151. doi: 10.1111/j.1365-246X.2012.05677.x

Roy, C., & Romanowicz, B. A. (2017). On the implications of a priori constraints in transdimensional Bayesian inversion for continental lithospheric layering. *Journal of Geophysical Research: Solid Earth*, *122*, 10118 - 10131.

Sambridge, M. (2014). A parallel tempering algorithm for probabilistic sampling and multimodal optimization. *Geophysical Journal International*, *192*, 357-374.

Sambridge, M., Braun, J., & McQueen, H. (1995). Geophysical parameterization and interpolation of irregular data using natural neighbours. *Geophysical Journal International*, *122*, 837-857.

Sambridge, M., & Faletič, R. (2003). Adaptive whole earth tomography. *Geochemistry Geophysics Geosystems*, *4*(3).

Sambridge, M., Gallagher, K., Jackson, A., & Rickwood, P. (2006). Trans-dimensional inverse problems, model comparison and the evidence. *Geophysical Journal International*, *167*, 528-542.

Sambridge, M., & Mosegaard, K. (2002). Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics*, *40*(3), 1-29.

Sandwell, D. T., & Smith, W. H. F. (1997). Marine gravity anomaly from Geosat and ERS 1 satellite altimetry. *J. Geophys. Res.*, *102*, 10,039 - 10,054.

Saygin, E., Cummins, P., Cipta, A., Hawkins, R., Pandhu, R., Murjaya, J., ... Kennett, B. L. N. (2016). Imaging architecture of the Jakarta basin, Indonesia with trans-dimensional Bayesian seismic noise tomography. *Geophysical Journal International*, *204*(2), 918 - 931. doi: 10.1093/gji/ggv466

Schöniger, A., Wöhling, T., Samaniego, L., & Nowak, W. (2014). Model selection on solid ground: Rigorous comparison of nine ways to evaluate Bayesian model evidence. *Water Resour. Res.*, *50*, 9484 - 9513.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461-464.

Sen, M. K., & Biswas, R. (2017). Transdimensional seismic inversion using the reversible jump Hamiltonian Monte Carlo algorithm. *Geophysics*, *82*(3), R119 - R134. doi: 10.1190/GEO2016-0010.1

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, *64*, 583-639.

Steininger, G., Dosso, S. E., Holland, C. W., & Dettmer, J. (2014). Estimating seabed scattering mechanisms via Bayesian model selection. *J. Acoust. Soc. Am.*, *136*(4), 1552-1562.

Valentine, A. P., & Sambridge, M. (2018). Optimal regularisation for a class of linear inverse problem. *Geophysical Journal International*, ggy303. doi: 10.1093/gji/ggy303

Zaroli, C. (2016). Global seismic tomography using Backus-Gilbert inversion. *Geophysical Journal International*, *207*, 876-888.