

Inversion using spatially variable mixed ℓ_p norms

Dominique Fournier^{ID} and Douglas W. Oldenburg

*Geophysical Inversion Facility, Department of Earth, Ocean, and Atmospheric Sciences, University of British Columbia, Canada.
E-mail: fournndo@gmail.com, doug@eoas.ubc.ca*

Accepted 2019 March 20. Received 2019 March 18; in original form 2018 December 13

SUMMARY

Non-uniqueness in the geophysical inverse problem is well recognized and so too is the ability to obtain solutions with different character by altering the form of the regularization function. Of particular note is the use of ℓ_p norms with $p \in [0, 2]$ which gives rise to sparse or smooth models. Most algorithms are designed to implement a single ℓ_p norm for the entire model domain. This is not adequate when the fundamental character of the model changes throughout the volume of interest. In such cases we require a generalized regularization function where each sub-volume of the model domain has penalties on smallness and roughness and its own suite of ℓ_p parameters.

Solving the inverse problem using mixed ℓ_p norms in the regularization (especially for $p < 1$) is computationally challenging. We use the Lawson formulation for the ℓ_p norm and solve the optimization problem with Iterative Reweighted Least Squares. The algorithm has two stages; we first solve the ℓ_2 -norm problem and then we switch to the desired suite of ℓ_p norms; there is one value of p for each term in the objective function. To handle the large changes in numerical values of the regularization function when p values are changed, and to ensure that each component of the regularization is contributing to the final solution, we successively rescale the gradients in our Gauss–Newton solution. An indicator function allows us to evaluate our success in finding a solution in which components of the objective function have been equally influential.

We use our algorithm to generate an ensemble of solutions with mixed ℓ_p norms. This illuminates some of the non-uniqueness in the inverse problem and helps prevent overinterpretation that can occur by having only one solution. In addition, we use this ensemble to estimate the suite of p values that can be used in a final inversion. First, the most common features of our ensemble are extracted using principal component analysis and edge detection procedures; this provides a reference model. A correlation of each member of the ensemble with the reference model, carried out in a windowed domain, then yields a set of p values for each model cell. The efficacy of our technique is illustrated on a synthetic 2-D cross-well example. We then apply our technique to the field example that motivated this research, the 3-D inversion of magnetic data at a kimberlite site in Canada. Since the final regularization terms have different sets of p values in different regions of model space we are able to recover compact regions associated with the kimberlite intrusions, continuous linear features with sharp edges that are associated with dykes and a background that is relatively smooth. The result has a geologic character that would not have been achievable without the use of spatially variable mixed norms.

Key words: North America; Magnetic anomalies: modelling and interpretation; Numerical modelling; Inverse theory; Tomography.

1 INTRODUCTION

In the general inverse problem we are given data $\mathbf{d}^{\text{obs}} = \{d_i^{\text{obs}}\}$, some estimate of their uncertainties $\{\sigma_i\}$ and an ability to carry out forward simulations $F_i[m] = d_i^{\text{pred}}$. We search for the model m that

gave rise to the data. A key issue related to the inverse problem is that there are an infinite number of possible models that can satisfy the observed data, that is, the inverse problem is ill-posed. Various assumptions must be made to reduce the solution to a few plausible candidates. A widespread and successful approach is to pose the

inversion as an optimization problem

$$\begin{aligned} \min_{\mathbf{m}} \phi(\mathbf{m}) &= \phi_d + \beta \phi_m \\ \text{s.t. } \phi_d &\leq \phi_d^*. \end{aligned} \quad (1)$$

In eq. (1), ϕ_d is the misfit function

$$\phi_d = \sum_{i=1}^N \left(\frac{d_i^{\text{pred}} - d_i^{\text{obs}}}{\sigma_i} \right)^2, \quad (2)$$

and ϕ_m is the regularization, or model objective function. It plays a central role in determining the character of the final model. It can take many forms but a generic representation, when we attempt to find 3-D functions that characterize an earth volume, is

$$\phi_m = \sum_{j=s,x,y,z} \alpha_j \int_V |f_j(m)|^{p_j} dV. \quad (3)$$

The functions f_j are user-defined but most often have the following form:

$$f_s = m, \quad f_x = \frac{dm}{dx}, \quad f_y = \frac{dm}{dy}, \quad f_z = \frac{dm}{dz}. \quad (4)$$

Thus $f_s(m)$ measures the size of m and $f_x(m), f_y(m)$ and $f_z(m)$ measure the roughness along orthogonal directions in 3-D. These functionals can also include a reference model m^{ref} . For ease of notation we shall omit explicit reference to m^{ref} , this is equivalent to setting $m^{\text{ref}} = 0$. The parameters p_j specify the values of the ℓ_p norm (generally $0 \leq p \leq 2$) that can be applied to each of the terms; their choice controls the character of the final model. Solutions with progressively fewer non-zero values of f_j are obtained by successively reducing p from $p = 2 \rightarrow 0$. The other important elements in the regularization are the α 's; they are user-defined scalars that adjust the relative influence of the various terms in the regularization function. Finally β is a scalar trade-off parameter (often called the Tikhonov parameter) that adjusts the relative weighting of ϕ_m and ϕ_d in the optimization problem. It is usually chosen so that the final misfit reaches a target value ϕ_d^* or is acceptably small.

The optimization problem is often solved by discretizing the system, linearizing the equations about a current model and using a gradient descent algorithm, such as a Gauss–Newton approach, where we attempt to find a solution that has zero gradients

$$\begin{aligned} \mathbf{g} &= \frac{\partial \phi(\mathbf{m})}{\partial \mathbf{m}} = \frac{\partial \phi_d}{\partial \mathbf{m}} \\ &+ \beta \left[\alpha_s \frac{\partial \phi_s}{\partial \mathbf{m}} + \alpha_x \frac{\partial \phi_x}{\partial \mathbf{m}} + \alpha_y \frac{\partial \phi_y}{\partial \mathbf{m}} + \alpha_z \frac{\partial \phi_z}{\partial \mathbf{m}} \right] = \mathbf{0}, \end{aligned} \quad (5)$$

where \mathbf{m} refers to a discrete model based on the choice of parametrization. A solution to eq. (5) can readily be calculated by gradient descent methods (Hestenes & Stiefel 1952; Nocedal & Wright 1999). Most often ℓ_2 -norms are used and this leads to a quadratic regularization function. For instance, the discretized form of eq. (3) is

$$\phi_m = \alpha_s \|\mathbf{V}_s \mathbf{m}\|_2^2 + \sum_{j=x,y,z} \alpha_j \|\mathbf{V}_j \mathbf{G}_j \mathbf{m}\|_2^2, \quad (6)$$

such that \mathbf{V}_s , \mathbf{V}_x , \mathbf{V}_y and \mathbf{V}_z are diagonal matrices relating to the ‘volumes’ of the cells used in discretizing the problem, and \mathbf{G}_x , \mathbf{G}_y and \mathbf{G}_z are discrete gradient operators. What is sometimes sought, at least as a first pass, is that if $\alpha_s \phi_s$ and $\alpha_x \phi_x$ have about the same numerical value then the two penalty terms are contributing equally. Dimensional analysis shows that for a uniform discretization h :

$$\frac{[\phi_x]}{[\phi_s]} = [h]^{-p}. \quad (7)$$

Thus for the ℓ_2 -norm regularization $\alpha_s/\alpha_x = h^{-2}$. If the forward problem is linear then the inverse problem (for a fixed β) is a linear system and easily solved.

Although convenient, the ℓ_2 -norm can be restrictive from the viewpoint of exploring model space. This has prompted the use of other ℓ_p -norms. The ℓ_1 -norm has received considerable attention. When applied to the smallness term it yields ‘spikey’ or impulsive models; when applied to the derivative term, it yields layered or blocky solutions (Li 1993; Farquharson & Oldenburg 1998; Daubechies *et al.* 2010). The benefits of ℓ_0 -norm have also been realized in generating minimum support models. When applied to the smallness component it gives a sparse model, when applied to the derivative term it gives compact objects (Last & Kubik 1983; Barbosa & Silva 1994; Portniaguine 1999; Ajo-Franklin *et al.* 2007). Other solutions on the full range for $0 \leq p \leq 2$ can help widen the search, but ℓ_0 , ℓ_1 and ℓ_2 often form the main applications.

In most applications the same norm has been applied uniformly to all components $f_j(m)$ making up the regularization, either globally or locally (Sun & Li 2014). A next step, which we are taking in this paper, is to use different norms for different components of eq. (3). The motivation is that the character of the model might change throughout the volume of interest; for instance it could be smooth and sparse in one region, and small and blocky in another. As an example, consider the magnetic data shown in Fig. 1. This is a total field magnetic map taken over the Tli Kwi Cho (TKC) kimberlite deposit, Northwest Territories. Different geologic structural features are evident: (a) linear structures corresponding to dykes, (b) compact bodies corresponding to kimberlite pipes and (c) a background host unit that is smoothly varying. An inversion algorithm that uses different ℓ_p -norms in different regions could capture the essence of this geology.

Our paper has two parts. In the first, we develop a stable two-stage algorithm for carrying out the optimization of spatially variable mixed norm (SVMN) objective functions. The essential part of this algorithm is to introduce an iterative re-scaling of the partial derivatives so that different components of the objective function can impact the solution throughout the inversion process. To illustrate our procedure we use the 2-D seismic tomography problem of Sun & Li (2014). We define an indicator function that allows us to evaluate our success in finding a solution in which components of the objective function have been equally influential. To implement the general ℓ_p -norm measure we use the Lawson approximation (Lawson 1961).

The development of a stable algorithm for solving multicomponent objective functions with variable ℓ_p -norm allows us to generate solutions that have different character and, partially, explore solution space. Our approach is to carry out a set of inversions that use different combinations of ℓ_p -norms. The suite of models generated can be subjected to further analysis. We use Principal Component Analysis (PCA) and edge detection algorithms to define dominant features and then use those images to determine ℓ_p -norms relevant to any particular sub-domain of the model. The efficacy of our approach is shown by carrying out SVMN inversions of the magnetic data shown in Fig. 1.

2 METHODOLOGY

In this study we focus on the general ℓ_p -norm measures of the form

$$\phi_s^p = \sum_i |m_i|^p \quad (8)$$

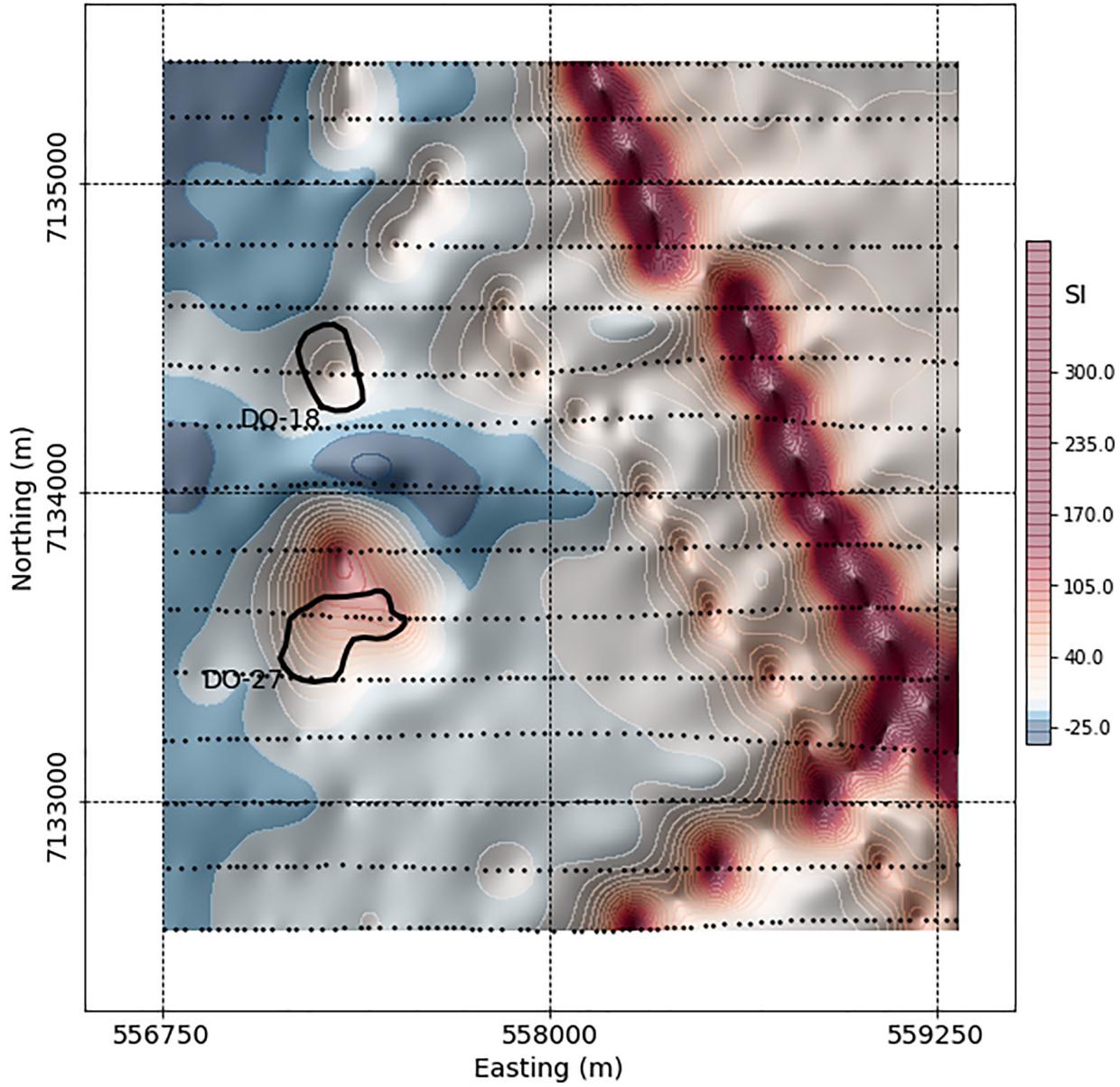


Figure 1. TMI data map from an airborne DIGHEM survey flown in 1992 over the DO-18/27 kimberlite pipes, formerly known as the Tli Kwi Cho (TKC) deposit, Northwest Territories. The study area comprises both linear and compact features of interest.

to constrain the inverse problem in eq. (1). Several approximations to the ℓ_p -norm have been implemented so that the function remains differentiable at $m_i = 0$. Approximations such as the Huber norm (Huber 1964)

$$\sum_i |m_i|^p \approx \sum_i \begin{cases} m_i^2, & |m_i| \leq \epsilon, \\ 2\epsilon|m_i| - \epsilon^2, & |m_i| > \epsilon, \end{cases}$$

and the Ekblom norm (Ekblom 1973):

$$\sum_i |m_i|^p \approx \sum_i (m_i^2 + \epsilon^2)^{p/2} \quad (9)$$

have been used to approximate the ℓ_1 -norm, where a small threshold value ϵ is added (Li 1993; Gorodnitsky & Rao 1997; Farquharson & Oldenburg 1998). Similarly, the ℓ_p -norm can be approximated by

Lawson's algorithm (Lawson 1961):

$$\sum_i |m_i|^p \approx \sum_i \frac{m_i^2}{(m_i^2 + \epsilon^2)^{1-p/2}}. \quad (10)$$

This formulation has received considerable attention in the literature and emerged as an important method in geophysical inversion. Under this approximation, the ℓ_p -norm penalty eq. (3) takes the form:

$$\int_V |f(m)|^{p_j} dV \approx \int_V \frac{f(m)^2}{(f(m)^2 + \epsilon_j^2)^{1-p_j/2}} dV, \quad (11)$$

where once again $f(m)$ from eq. (4) can either be a measure of the model or its spatial gradients, and m may, or may not be replaced by $m - m^{\text{ref}}$. The function can be discretized and evaluated through the Iterative Reweighted Least-Squares (IRLS) approach where the denominator is replaced by model parameters from the most recent

iteration. The smallest model component can be written as

$$\phi_s^{p_s} = \sum_i \frac{m_i^2}{((m_i^{(k-1)})^2 + \epsilon_s^2)^{1-p_s/2}} V_i, \quad (12)$$

where $m_i^{(k-1)}$ are model parameters obtained at a previous iteration and V_i are ‘volume’ terms connected with the discretization. In eq. (12) we have explicitly written the objective function as $\phi_s^{p_s}$ to indicate that we are evaluating a smallest model component with an ℓ_p -norm with $p = p_s$. The approximated norm can be expressed in linear form as

$$\phi_s^{p_s} = \|\mathbf{V}_s \mathbf{R}_s \mathbf{m}\|_2^2, \quad (13)$$

$$\mathbf{R}_s = \text{diag} \left[\left((\mathbf{m}^{(k-1)})^2 + \epsilon_s^2 \right)^{p_s/2-1} \right]^{1/2}. \quad (14)$$

Carrying out the same IRLS operation on the derivative components yields

$$\phi_j^{p_j} = \|\mathbf{V}_j \mathbf{R}_j \mathbf{D}_j \mathbf{m}\|_2^2, \quad (15)$$

where the IRLS weights are calculate by

$$\mathbf{R}_j = \text{diag} \left[\left((\mathbf{D}_j \mathbf{m}^{(k-1)})^2 + \epsilon_j^2 \right)^{p_j/2-1} \right]^{1/2}. \quad (16)$$

In this study, we replace the gradient terms \mathbf{G}_j with a finite difference operator

$$\mathbf{D}_j = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & -1 & 1 \end{bmatrix}. \quad (17)$$

This makes ϕ_s and ϕ_j dimensionally equivalent and the default scaling values can be set to $\alpha_s = \alpha_j = 1$. This simplifies greatly the implementation of the IRLS when dealing with discretization using variable cell sizes. The final regularization function is thus

$$\phi_m^p = \alpha_s \|\mathbf{V}_s \mathbf{R}_s \mathbf{m}\|_2^2 + \sum_{j=x,y,z} \alpha_j \|\mathbf{V}_j \mathbf{R}_j \mathbf{D}_j \mathbf{m}\|_2^2. \quad (18)$$

The core IRLS procedure described in Table 1 involves two main stages:

(1) Stage 1 solves the inverse problem using ℓ_2 -norms ($p_s = p_j = 2$). The assumption is made that the globally convex ℓ_2 -norm regularized inversion is a good approximation of the true solution and it is used to form the initial IRLS weights defined in eq. (14). The β parameter is controlled by a cooling schedule that starts with a high value and is successively decreased until $\phi_d \approx \phi_d^*$.

(2) Stage 2 starts from the solution obtained in Stage 1 and solves the inverse problem for specified p_s , p_x , p_y and p_z values. This is done iteratively using the regularization in eq. (18) and a standard Gauss–Newton procedure. A gradient descent direction $\delta\mathbf{m}$ is found by solving

$$\mathbf{H} \delta\mathbf{m} = \mathbf{g}, \quad (19)$$

where \mathbf{H} is the approximate Hessian and \mathbf{g} is the gradient of the objective function. We use the Conjugate Gradient method (Hestenes & Stiefel 1952) to solve this system.

The model update at the k th iteration is

$$\mathbf{m} = \mathbf{m}^{(k-1)} + \alpha \delta\mathbf{m}, \quad (20)$$

where the step length α is found by a line-search back-stepping method (Nocedal & Wright 1999). Gradient steps are only performed if the data misfit remains within the user-defined tolerance η_{ϕ_d} .

$$\frac{|\phi_d - \phi_d^*|}{\phi_d^*} \leq \eta_{\phi_d}. \quad (21)$$

If outside the tolerance, the algorithm repeats the Gauss–Newton calculation with the previous $\mathbf{m}^{(k-1)}$ and a different β -value, either lower or higher depending on the achieved ϕ_d . This β -search step is an important component in the workflow when the minimization switches between an ℓ_2 to an ℓ_p objective function because ϕ_m^p can vary markedly. This can force a change of β by a few orders of magnitude in some cases. Once an appropriate β has been found such that eq. (21) is respected, the model update $\mathbf{m}^{(k)}$ is accepted and used for the next iteration cycle. The IRLS process continues until the change in regularization falls below some pre-defined tolerance η_{ϕ_m}

$$\frac{|\phi_m^{(k-1)} - \phi_m^{(k)}|}{\phi_m^{(k)}} < \eta_{\phi_m}. \quad (22)$$

We set to $\eta_{\phi_m} = 10^{-5}$ (0.01 per cent change) in all our experiments.

The choice of threshold ϵ parameters remains a subject of disagreement among researchers. In the early work of Last & Kubik (1983), it was suggested that the threshold value should be small or near machine error ($\epsilon < 10^{-8}$) in order to best approximate the ℓ_p -norm. The same strategy was later adopted by others (Barbosa & Silva 1994; Stocco *et al.* 2009). Other researchers, such as in Ajo-Franklin *et al.* (2007), observed instabilities with small values and opted for a wider range ($10^{-4} < \epsilon < 10^{-7}$). More recently, Sun & Li (2014) proposed an ϵ -search phase to highlight regions reacting favourably to the ℓ_p -norm regularization, followed by a final inversion step with fixed threshold value ($\epsilon \ll 10^{-4}$). A similar strategy has also been proposed by Zhdanov & Tolstaya (2004) after selecting an optimal point on a trade-off curve.

In this study, we opt for a cooling strategy. Threshold values ϵ_j 's are initialized at a large value then monotonically reduced such that

$$\epsilon_j^{(k)} = \frac{\|f_j(\mathbf{m}^{(0)})\|_\infty}{\eta^k}, \quad (23)$$

where η is a user-defined cooling rate constant and $\|f_j(\mathbf{m}^{(0)})\|_\infty$ denotes the largest function value obtained at the end of Stage 1 of the algorithm for $j = s, x, y, z$. At the start of Stage 2, the Lawson approximation with large ϵ is effectively an ℓ_2 -norm. Thus there is only a small change in regularization between Stages 1 and 2 of the algorithm. This is desired since the ℓ_p -norm regularization is highly nonlinear. As the iteration process continues and $\epsilon \rightarrow 0$, the emphasis of the regularization function sweeps through the range of model values. Function values that are driven to the minimum of the regularization function are maintained at that value unless needed by one of the competing functions. This process continues until the algorithm reaches the convergence criteria presented in eqs (21) and (22). From a user standpoint, the cooling strategy is attractive as it eliminates the requirement to predetermine optimal ϵ threshold values and instead relies on a cooling rate η . We found experimentally that $\eta \approx 1.25$ generally yielded an optimal trade-off between computational time (number of iterations) and convergence to suitable solution.

Table 1. IRLS algorithm in pseudo-code made of two stages: **Stage 1** Initialization with convex least-squares inversion, **Stage 2** IRLS updates with inner β -search steps.

Stage 1: Initialization (ϕ_m^2)	Stage 2: IRLS (ϕ_m^p)	β -Search
$\min_m \phi_d + \beta \phi_m^2$	while $\frac{ \phi_m^{(k-1)} - \phi_m^{(k)} }{\phi_m^{(k)}} > \eta_{\phi_m}$	Solve $\mathbf{H} \delta \mathbf{m} = \mathbf{g}$
s.t. $\phi_d = \phi_d^*$	do β -Search	$\mathbf{m} = \mathbf{m}^{(k-1)} + \alpha \delta \mathbf{m}$
$\beta^{(0)}, \mathbf{m}^{(0)}, \mathbf{R}^{(0)}, \phi_m^{(0)}$	$k := k + 1$	if $\frac{ \phi_d - \phi_d^* }{\phi_d^*} > \eta_{\phi_d}$
—	$\beta^{(k)}, \mathbf{m}^{(k)}, \mathbf{R}^{(k)}$	adjust β , re-do
—	—	else
—	—	continue

2.1 Synthetic 2-D example

To demonstrate our methodology we use the 2-D traveltime tomography example presented in Sun & Li (2014). The synthetic velocity model shown in Fig. 2(a) is made up of a smooth velocity high and a blocky velocity low centred at a 1000 m along the x -axis; the background velocity is 2000 m s $^{-1}$. Contour lines corresponding to the 25th and 75th percentile values for both anomalies are also plotted. An array of 11 transmitters and 13 receivers is positioned on either side of the model which is discretized into 32×64 square 25 m cells. First arrival data are calculated for each transmitter–receiver pair by taking the line integral of slowness (reciprocal of velocity) along each ray path; this yields a total of 143 observations. Gaussian noise of 5 per cent is added to the simulated data shown in Fig. 2(b). We will first attempt to invert this data with mixed-norm penalties that are applied on the entire model.

2.1.1 Stage 1: ℓ_2 -norm solution

Starting with Stage 1, we first solve this inverse problem with the conventional ℓ_2 norm regularization ($p_s = p_x = p_y = 2$). After convergence of the algorithm ($\phi_d \approx \phi_d^*$), we recover the model presented in Fig. 3(a). We can distinguish two velocity anomalies at roughly the right depth with respect to the true model. We note however that the shape and horizontal extent of both targets are poorly defined. Anomalies are stretched along the straight-ray paths associated with the experiment.

Visually, the solution exhibits characteristics of remaining near the reference model of 2000 m s $^{-1}$ while also attempting to be smooth. Numerical evaluation of the components of the regularization function for this solution yields:

(

This might suggest that ϕ_s , ϕ_x and ϕ_y are roughly equal in importance. As we later attempt to improve this solution with variable ℓ_p -norm measures, we will need a general metric to evaluate our success in finding a solution in which components of the objective function have been equally influential. Rather than comparing the actual values of the norms, we propose to quantify the relative importance of each term based on their partial derivatives, or gradients. From eq. (5) we expect to find an optimal solution where the sum of the gradients vanishes, either because all components are equal to zero, or because multiple gradients have opposite signs. To quantify the size of the gradients we use the infinity norm

$$\|\mathbf{g}_j\|_\infty = \left\| \frac{\partial \phi_j}{\partial \mathbf{m}} \right\|_\infty. \quad (24)$$

The $\|\mathbf{g}_j\|_\infty$ metric is appealing for a few reasons: (i) it is directly linked to the minimization process because we use gradient descent methods, (ii) it does not depend on the dimension M of the parameter

space as do other measures that involve a sum of components of the vector, (iii) the theoretical maximum can be calculated analytically for any given ℓ_p -norm function. These properties will become useful in the following section when we attempt to balance different norm penalties applied on a cell-by-cell basis.

We quantify the relative importance of the regularizations terms with a proportionality ratio:

$$\lambda_\infty = \frac{\alpha_s \|\mathbf{g}_s\|_\infty}{\alpha_x \|\mathbf{g}_x\|_\infty + \alpha_y \|\mathbf{g}_y\|_\infty}. \quad (25)$$

For the ℓ_2 -norm solution presented in Fig. 3, we calculate a proportionality ratio near unity ($\lambda_\infty = 1.65$). This confirms that, with the current α -scaling strategy, smallness and smoothness penalties are contributing near equally to the solution. As we further generalize the regularization function for arbitrary ℓ_p -norm measures, we will attempt to maintain this proportionality ratio ($\lambda_\infty \approx 1$) between competing functions. From Fig. 3 we conclude that the ℓ_2 -norm regularization applied uniformly to the entire model domain does not provide an acceptable solution. We want to find more compact solutions but that display both smooth and abrupt boundaries. We therefore employ more general ℓ_p norms.

2.1.2 Stage 2: ℓ_p -norm penalties

The idea of combining different norms for the simultaneous recovery of smooth and blocky features has partially been explored by Sun & Li (2014). They demonstrated the benefits of dividing model space into regions with different ℓ_p -norm penalties. The choice of norms was limited to be either l_1 or l_2 norm. Little has been published however on the independent mixing of model and gradient norms on the range $p \in [0, 2]$, although this problem was initially addressed in (Fournier 2015).

Before attempting to vary the penalties locally (i.e. within the model domain) we first tackle the general mixed-norm problem. As an entry point we attempt to minimize the following objective function:

$$\min_{\mathbf{m}} \phi(\mathbf{m}) = \phi_d + \beta \left[\alpha_s \phi_s^0 + \sum_{j=x,y} \alpha_j \phi_j^2 \right] \quad (26)$$

$$\text{s.t. } \phi_d \leq \phi_d^*.$$

Based upon our previous work, we expect the solution to be sparse, in terms of non-zero model parameters ($p_s = 0$), while smooth with respect to the model gradients ($p_x, p_y = 2$). This should be more suitable for imaging smooth and compact velocity anomalies. Using the standard IRLS strategy ($\alpha_s = \alpha_x = \alpha_y = 1$), we obtain the solution presented in Fig. 4(a). Instead of the expected smooth model, we have recovered scattered anomalies concentrated near the velocity targets. The solution has been dominated by ϕ_s^0 . The influence from the smoothness penalties ϕ_x^2 and ϕ_y^2 has been marginal. Comparing

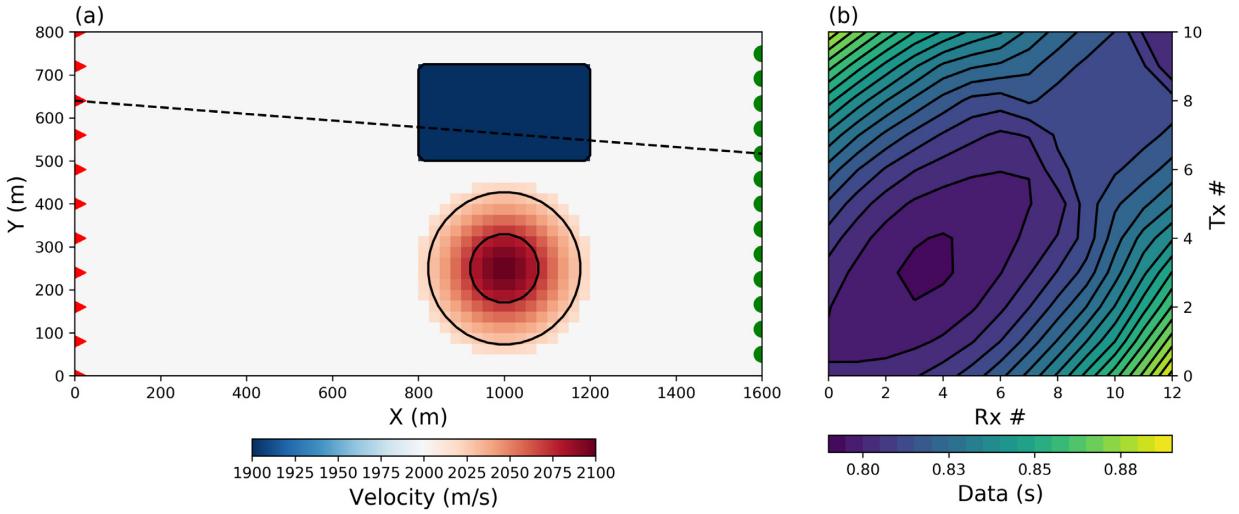


Figure 2. (a) Synthetic 2-D traveltome tomography problem made up of a rectangular velocity low and a smooth Gaussian velocity high. Contour lines (25th and 75th percentile) are shown in black for the true position and shape of the velocity anomalies. Transmitters (red) and receivers (green) are positioned on opposite sides of the model domain. (b) The data map for the 143 line integral data calculated for each transmitter-receiver pair.

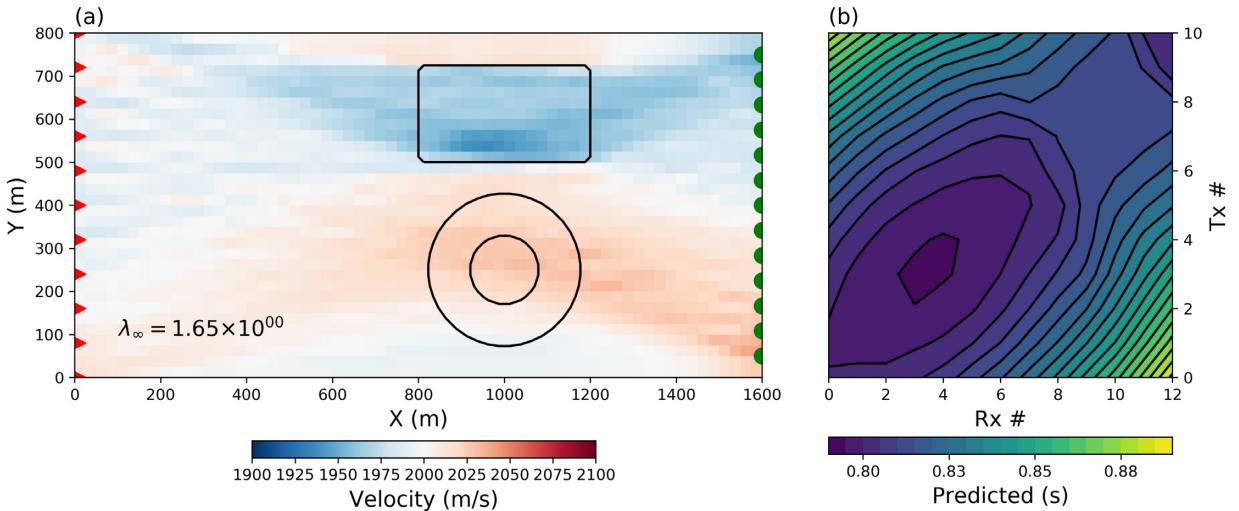


Figure 3. (a) Recovered smooth model and (b) predicted data. Contour lines are shown in (a) for the true position and shape of the velocity anomalies.

the partial derivatives of the objective function confirms this. After convergence, the calculated proportionality ratio is $\lambda_\infty = 7.31 \times 10^8$. This is a significant change from the end of Stage 1 where $\lambda_\infty \approx 1$. Clearly Stage 2 of the IRLS took the solution away from the proportionality condition, which translated into the poor recovery of continuous velocity anomalies.

2.1.3 Scaled ℓ_p -norm penalties

A more desirable solution can be obtained if proportionality among the components of the objective function is preserved throughout the IRLS process. Since the inverse problem is solved using gradients of the composite objective function $\phi(\mathbf{m})$, the relative magnitude of the individual gradients is a driving force in controlling the iteration step in eq. (19). Taking the partial derivatives of the linearized Lawson norm as prescribed in eq. (11) yields

$$g^p = \frac{\partial \phi^p}{\partial m} = \frac{f(m)}{(f(m)^2 + \epsilon^2)^{1-p/2}} V. \quad (27)$$

In eq. (27) the superscript p on the gradient, g^p , refers to the associated p -value and is not to be interpreted as an exponential. These are plotted in Fig. 5(a) as a function of $f(m)$ and for various values of p . We note that the magnitude of the derivative increases rapidly for small p and ϵ values as $f(m) \rightarrow 0$. This property of the IRLS approximation is important because, when attempting to combine different norm penalties within the same objective function, there will be a systematic bias towards small ℓ_p -norm penalties. To circumvent this bias we define the following gradient-based scaling

$$\gamma = \left[\frac{\|g^2\|_\infty}{\|g^p\|_\infty} \right]^{1/2}. \quad (28)$$

By using this scaling, we can equalize the size of the gradients associated with any ℓ_p norm. We can easily compute $\|g^p\|_\infty$ for any function $f(m)$ by taking the partial derivative of eq. (27) and setting it to zero. This maximum gradient of the Lawson approximation occurs at $f(m)^*$

$$f(m)^* = \begin{cases} \infty \text{ or } f(m)_{\max}, & p \geq 1 \\ \frac{\epsilon}{\sqrt{1-p}}, & p < 1, \end{cases} \quad (29)$$

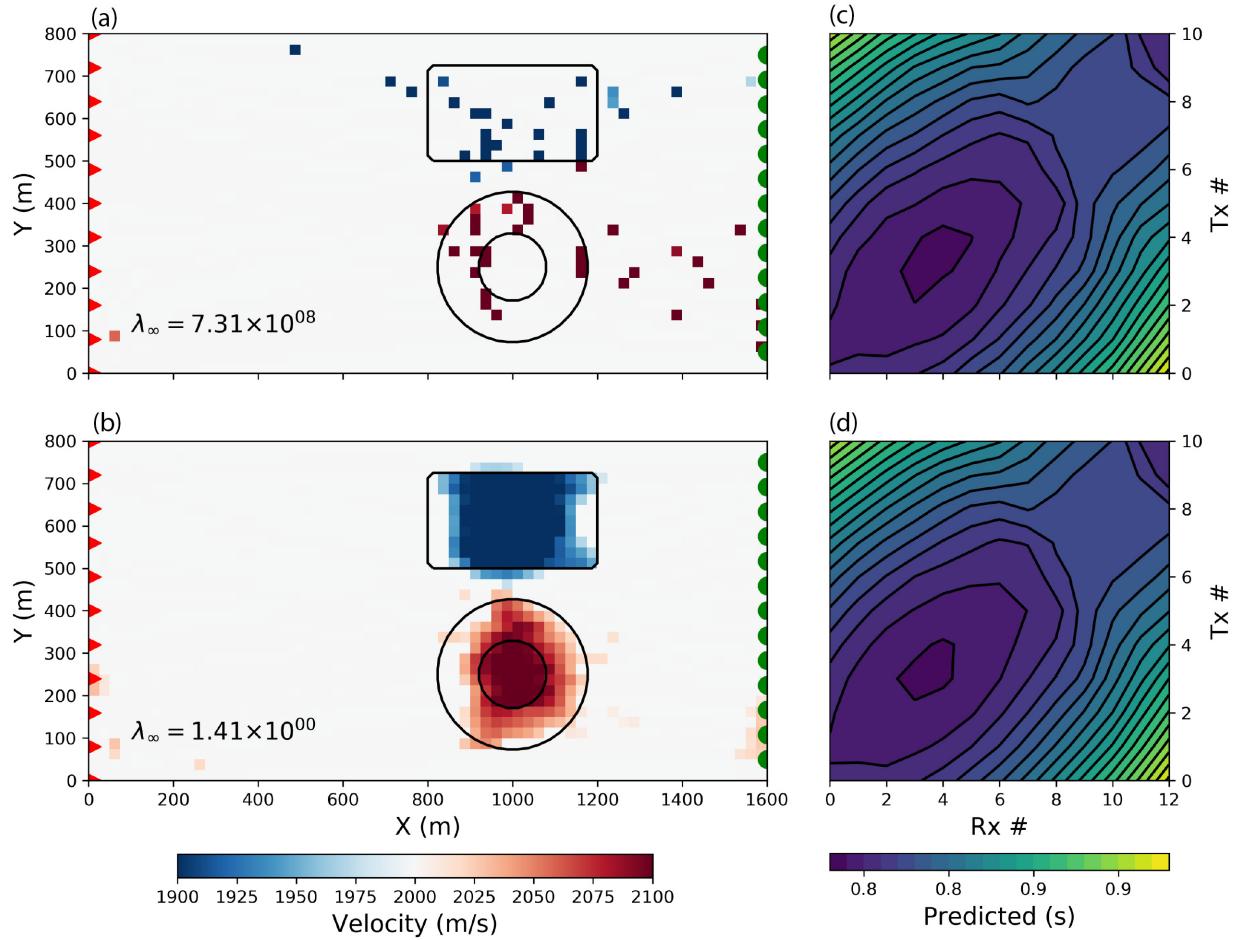


Figure 4. Recovered mixed-norm models using (a) the conventional IRLS and (b) the scaled-IRLS algorithm for $p_s = 0$ and $p_x = p_y = 2$. The contour lines are shown for the true position and shape of the velocity anomalies. (c), (d) Predicted data plot obtained with both approaches.

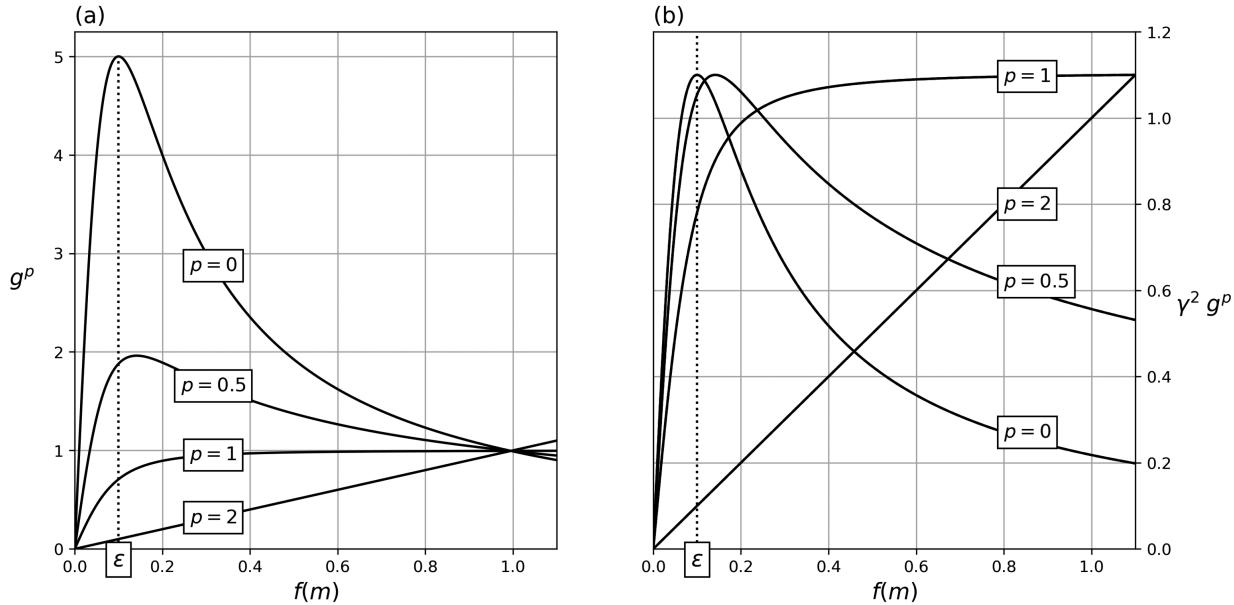


Figure 5. (a) Derivatives of the Lawson approximation over range of model function $f(m)$ and for a fix threshold parameter $\epsilon = 10^{-1}$ chosen for visual purpose only. For $p < 2$, gradients increase rapidly as $f(m) \rightarrow 0$ resulting in a regularization function dominated by sparse norms. (b) Gradients after applying the γ scaling, bringing all maximums to be equal.

from which we can calculate $\|\mathbf{g}^p\|_\infty$ by substituting $f(m)^*$ into eq. (27). Fig. 5(b) presents the scaled gradients for different approximated ℓ_p norms. Note that the largest derivative of any norm is at most as large as the ℓ_2 -norm penalty for $f(m) \in \mathbb{R}$. This equalization of the norm derivatives guarantees that multiple penalties can co-exist and impact the solution at every step of the IRLS, regardless of the chosen $\{p, \epsilon\}$ values.

By applying the scaling to the vector gradients of the model and derivatives term, we define a Scaled-IRLS (S-IRLS) regularization function of the form:

$$\hat{\phi}_m^p = \alpha_s \|\mathbf{V}_s \hat{\mathbf{R}}_s \mathbf{m}\|_2^2 + \sum_{j=x,y,z} \alpha_j \|\mathbf{V}_j \hat{\mathbf{R}}_j \mathbf{D}_j \mathbf{m}\|_2^2 \quad (30)$$

such that IRLS weights in eqs (14) and (16) are replaced by

$$\begin{aligned} \hat{\mathbf{R}}_s &= \gamma_s \mathbf{R}_s \\ \hat{\mathbf{R}}_j &= \gamma_j \mathbf{R}_j. \end{aligned} \quad (31)$$

The scaling parameters γ_s and γ_j are:

$$\gamma_s = \left[\frac{\|\mathbf{g}_s^2\|_\infty}{\|\mathbf{g}_s^{p_s}\|_\infty} \right]^{1/2}, \quad \gamma_j = \left[\frac{\|\mathbf{g}_j^2\|_\infty}{\|\mathbf{g}_j^{p_j}\|_\infty} \right]^{1/2}. \quad (32)$$

This re-scaling is done for two reasons. First, at the transition between Stages 1 and 2, it preserves the balance between the misfit and regularization terms and thus no large adjustment in the trade-off parameter β is needed. Second, this ensures that proportionality between ϕ_s^0 and ϕ_j^2 is preserved during the ℓ_p -norm inversion.

Two options are possible to compute the γ -scalings: (i) take the maximum absolute gradient directly from the gradient values in eq. (27), or (ii) calculate $\|\mathbf{g}_j^p\|_\infty$ analytically as prescribed in eq. (29). We have found that Option 2 is more stable since it is based upon a theoretical maximum of the gradient and not on a particular realization of that maximum that arises from the distribution of values in the current model $\mathbf{m}^{(k)}$.

The outcome of the re-scaling strategy is shown in Fig. 4(b). The solution seems to have our desired properties of being sparse in terms of model values and having smooth edges. It is interesting to note from Figs 4(c) and (d) that both solutions honour the observed data within the same tolerance, yet the S-IRLS solution is a better representation of the true model. The final calculated proportionality ratio $\lambda_\infty = 1.41$ confirms that the scaling strategy was successful in balancing the impact of the sparse and smooth regularization functions.

3 AUTOMATED MIXED-NORM MODELLING

In the preceding section we have developed a methodology to find a solution that minimizes a combination of model norms. Different combinations of norms allow us to generate a variety of models that adequately fit the data and sample model space in a systematic manner. This flexibility promotes two objectives. First, inverting for diverse models provides immediate insight about non-uniqueness and can prevent practitioners from overinterpreting a single inversion result. This is one of the most important aspects of our work. The second objective is more challenging. As outlined in the introduction, we want to use different combinations of norms on different parts of the model space so that we obtain solutions that are geologically informative. This selection process can rapidly become overwhelming for large problems in complex geological settings. A semi-automated approach that can capture the heterogeneity of

the Earth, with minimal intervention from an expert, is needed. Our strategy is:

- (1) Run a suite of inversions over a range of p parameters applied to the model norm and its spatial gradients. We will assume that the suit of models is a representative sampling of diverse solutions from the ℓ_p space.
- (2) Form an *average* model (\mathbf{m}_P) that captures most of the variability in the solution space
- (3) Extract p parameters for windowed portions of model space based on the correlation between the average \mathbf{m}_P and individual models.
- (4) Carry out a final SVMN inversion with local p parameters.

There are numerous ways to implement the above strategy. Here, we use a combination of PCA and edge detection algorithms. We illustrate our approach with a 2-D synthetic example and a 3-D field example.

3.1 Exploring the model space

Having dealt with scaling issues between ℓ_p -norm measures, we can now exploit the full flexibility of the regularization function in eq. (3). Our goal is to generate a suite of models under a broad range of assumptions. To demonstrate the flexibility of our algorithm, we carry out seven additional inversions, using a combination of norms on a range of $p_s, p_x, p_y \in [0, 1, 2]$ values ($p_x = p_y$ in all cases). The solutions, nine in total, are presented in Fig. 6. All models have a final misfit $\phi_d^* \approx 143$ and use the same ℓ_2 -norm solution to initiate the IRLS steps.

We make the following observations. The two velocity anomalies centred at 1000 m on the x -axis are dominant features. There is a progressive transition from a smooth model (upper left) to a blocky solution (lower right) as p_s, p_x and p_y decrease. The upper body (velocity low) appears to be most often represented as a blocky body with sharp edges. The lower anomaly (velocity high) tends to be more smooth. Away from the anomalous regions, the velocity is relatively smooth and close to the background reference model of 2000 m s^{-1} .

3.2 Average PCA model

We assume that the suite of models presented in Fig. 6 contains sufficient variability to be representative of our solution space and we wish to use these to extract the main features. There are numerous approaches to accomplish this and they vary in complexity from simple averaging to advanced machine learning algorithm. We use a PCA (Pearson 1901; Hotelling 1933). Considering each model in Fig. 6 as a *data* vector, the principal components of our solution space can be written as

$$[\mathbf{m}_1, \dots, \mathbf{m}_{nM}] \approx \mathbf{A} \mathbf{W} \quad (33)$$

such that PCA vectors along the columns of $\mathbf{A} \in \mathbb{R}^{nM \times nV}$ contains a subset of nV eigenvectors spanning the model space. These vectors encode the principal source of variation across the nine models recovered. The corresponding weights $\mathbf{W} \in \mathbb{R}^{nV \times nM}$, also known as loadings, are scalars relating how each principal component contributes to each model. We use the PCA algorithm from the open-source Python library Scikit-Learn.decomposition.PCA (Pedregosa *et al.* 2011). The number of principal components to be used in the analysis is determined by the experimenter. Fig. 7 presents the four largest principal components covering in

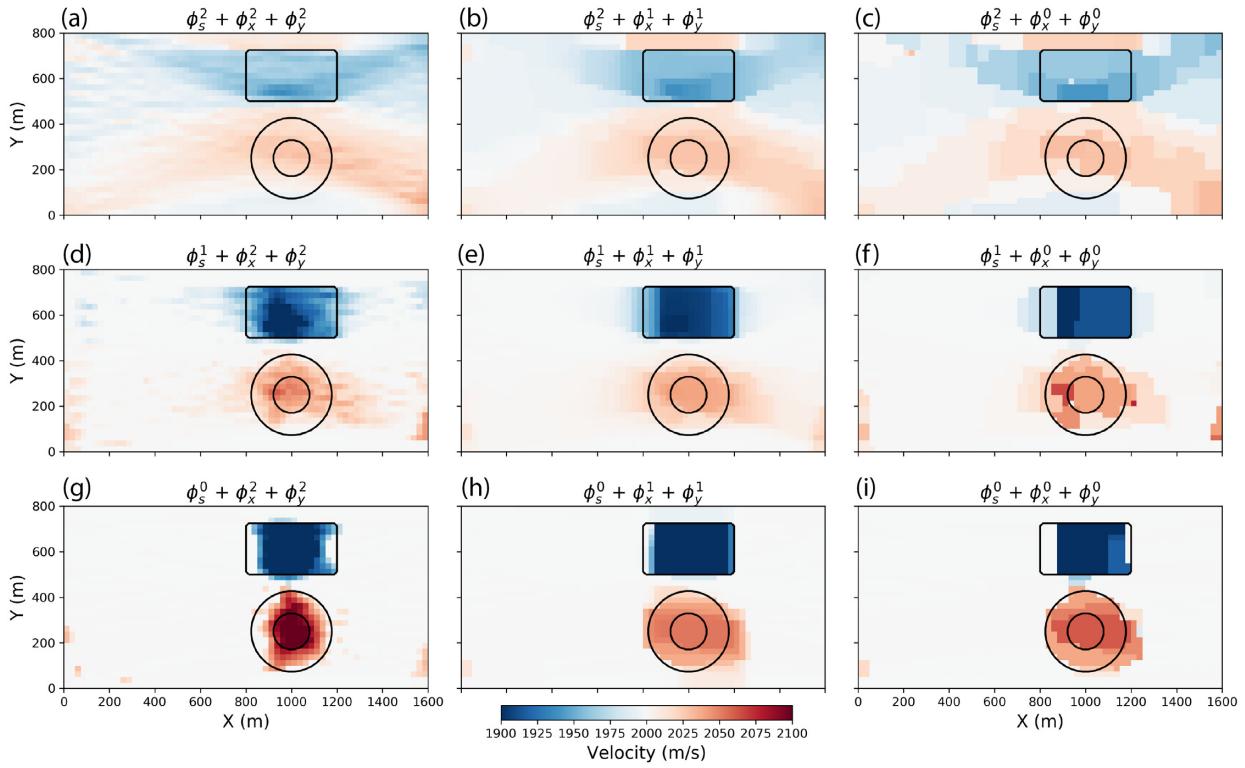


Figure 6. Suite of inverted models for various combination of norms $p_s, p_x = p_y \in [0, 1, 2]$ ($\alpha_s = \alpha_x = \alpha_y = 1$). The contour lines (25th and 75th percentile) are shown in black for the true position and shape of the velocity anomalies.

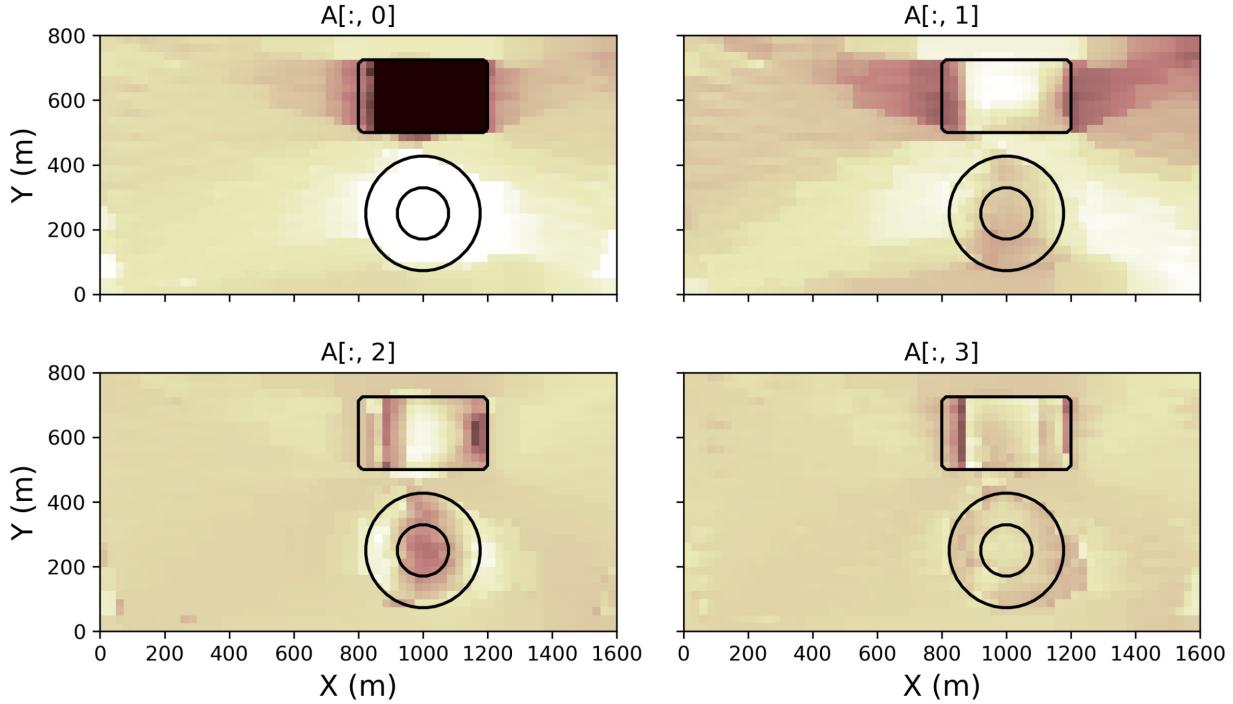


Figure 7. PCA vectors covering 75 per cent of the model variances.

this case over 75 per cent of the variance present in the model space.

Next, we generate a representative model by computing a weighted averaged model based on the positive PCA loadings such

that

$$\mathbf{m}_P = \frac{\sum_{i=1}^{nV} \sum_{j=1}^{nM} W_{ij} \mathbf{m}_i}{\sum_{i=1}^{nV} \sum_{j=1}^{nM} W_{ij}}. \quad (34)$$

The average model \mathbf{m}_P is presented in Fig. 8. We note the close

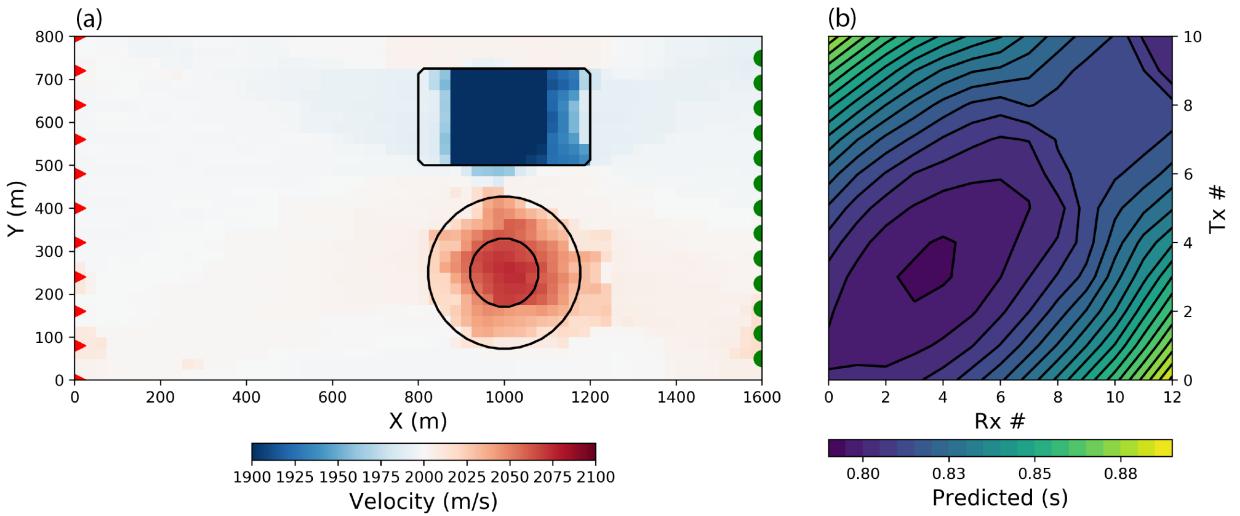


Figure 8. (a) Averaged PCA model \mathbf{m}_P (b) predicted data. The contour lines are shown in (a) for the true position and shape of the velocity anomalies.

resemblance with the true model. The background is fairly uniform near 2000 m s^{-1} ; the low-velocity anomaly appears to be a block with a velocity near 1900 m s^{-1} and the high-velocity anomaly appears in a smooth feature with a maximum velocity near 2100 m s^{-1} .

3.3 Parameters extraction

Next, we want to extract optimal inversion parameters on a cell-by-cell basis in order to best describe local features. To do so, we resort to a pattern recognition approach. In order to remove biases towards extreme model values, we transform our model space into a simpler parametric representation. We use a Canny edge detection algorithm from the open-source library `Scikit-Image.feature.Canny` (Pedregosa *et al.* 2011). Fig. 9 shows the parametric edges extracted from all nine inversions and the average PCA model \mathbf{m}_P .

From this simplified representation of each model, we perform a moving window correlation $r_{\mathbf{m}_i \mathbf{m}_P}$ between the average PCA model \mathbf{m}_P and each of the \mathbf{m}_i solutions:

$$r_{\mathbf{m}_i \mathbf{m}_P} = \frac{\sum_{j=1}^n (m_{ij} - \bar{m}_i)(m_{pj} - \bar{m}_P)}{\sqrt{\sum_{j=1}^n (m_{ij} - \bar{m}_i)^2 \sum_{j=1}^n (m_{pj} - \bar{m}_P)^2}}, \quad (35)$$

where \bar{m}_i and \bar{m}_P are the average model and PCA model values inside the window denoted by the subscript j . The parameters p_s, p_x and p_y associated with the highest correlation are used in a weighted average as defined in eq. (34). The process is repeated over the entire model space. For our example we use 20×20 pixels window. The recovered \mathbf{p}_s and $\mathbf{p}_x, \mathbf{p}_y$ values are presented in Fig. 10(a) and (b), respectively. We note that the norm on the model gradients is generally larger in the bottom region of the model corresponding to the location of the smooth positive anomaly.

We now define Scaled-IRLS weights on a cell-by-cell basis as:

$$\begin{aligned} \hat{\mathbf{R}}_s &= \text{diag} \left[\gamma_s \circ \left((\mathbf{m}^{(k-1)} - \mathbf{m}^{\text{mref}})^2 + \epsilon_s^2 \right)^{\circ \mathbf{p}_s/2-1} \right]^{1/2}, \\ \hat{\mathbf{R}}_x &= \text{diag} \left[\gamma_x \circ \left((\mathbf{D}_x (\mathbf{m}^{(k-1)} - \mathbf{m}^{\text{mref}}))^2 + \epsilon_x^2 \right)^{\circ \mathbf{p}_x/2-1} \right]^{1/2}, \\ \hat{\mathbf{R}}_y &= \text{diag} \left[\gamma_y \circ \left((\mathbf{D}_y (\mathbf{m}^{(k-1)} - \mathbf{m}^{\text{mref}}))^2 + \epsilon_y^2 \right)^{\circ \mathbf{p}_y/2-1} \right]^{1/2}, \end{aligned} \quad (36)$$

where $\circ \mathbf{p}_s, \circ \mathbf{p}_x$ and $\circ \mathbf{p}_y$ define the element-wise Hadamard power and $\gamma_s \circ, \gamma_x \circ$ and $\gamma_y \circ$ are the element-wise scaling multiplications

defined in eq. (31). In this fashion, the approximated mixed norm regularization can vary on a cell-by-cell basis.

Finally we proceed with our SVMN inversion. The data are now inverted using the extracted local parameters, applied on a cell-by-cell basis, and the result is shown in Fig. 11. There is a good correspondence with the true model: both the low-velocity blocky anomaly and the high-velocity smooth anomaly are imaged at the right location, with the right shape and near their respective seismic velocity. It is worth noting that we have achieved this result without direct input from the user, other than choosing a window size at the parameter extraction phase.

4 CASE STUDY: DO-18/27 AIRBORNE MAGNETICS

We now use our algorithm to invert airborne magnetic survey acquired over the DO-18/27 diamondiferous kimberlite pipes previously shown in Fig. 1. The deposit is located 30 km SE from the Lac de Gras region, Northwest Territories. The DO-18/27 kimberlites pipes are more conductive and susceptible than the host archean rocks and easily identifiable from airborne surveys (Pell 1997; Devriese *et al.* 2017; Fournier *et al.* 2017). They are also associated with a large network of magnetic dyke swarms. Our goal is to better recover the shape of compact kimberlite pipes as well as the narrow and elongated dykes.

At the time of acquisition, the inducing field strength, inclination and declination were $59\,500 \text{ nT}$, 83° and 19.5° , respectively. As a first step, we rotate the magnetic data 30° clockwise in order to align the strike of dykes along the Cartesian directions (not shown here). We assign 5 nT uncertainty on the data. We discretize the Earth into 25 m cubic cells and invert the data with the conventional smooth ℓ_2 -norm assumption with the open-source SimPEG package (Cockett *et al.* 2015). As shown in Fig. 12(a), the magnetic dykes appear to break between each survey line due to a large flight line separation. Considering the regional geology, the magnetic dykes are likely to be continuous throughout the region and they should be better represented by plate-like bodies.

As previously done on the synthetic 2-D model, we proceed with a series of inversions to sample solution space. Since we are now dealing with a 3-D problem, a large number of combinations of norms for $p_{s,x,y,z} = 2, 1, 0$ is possible. We are mostly interested in

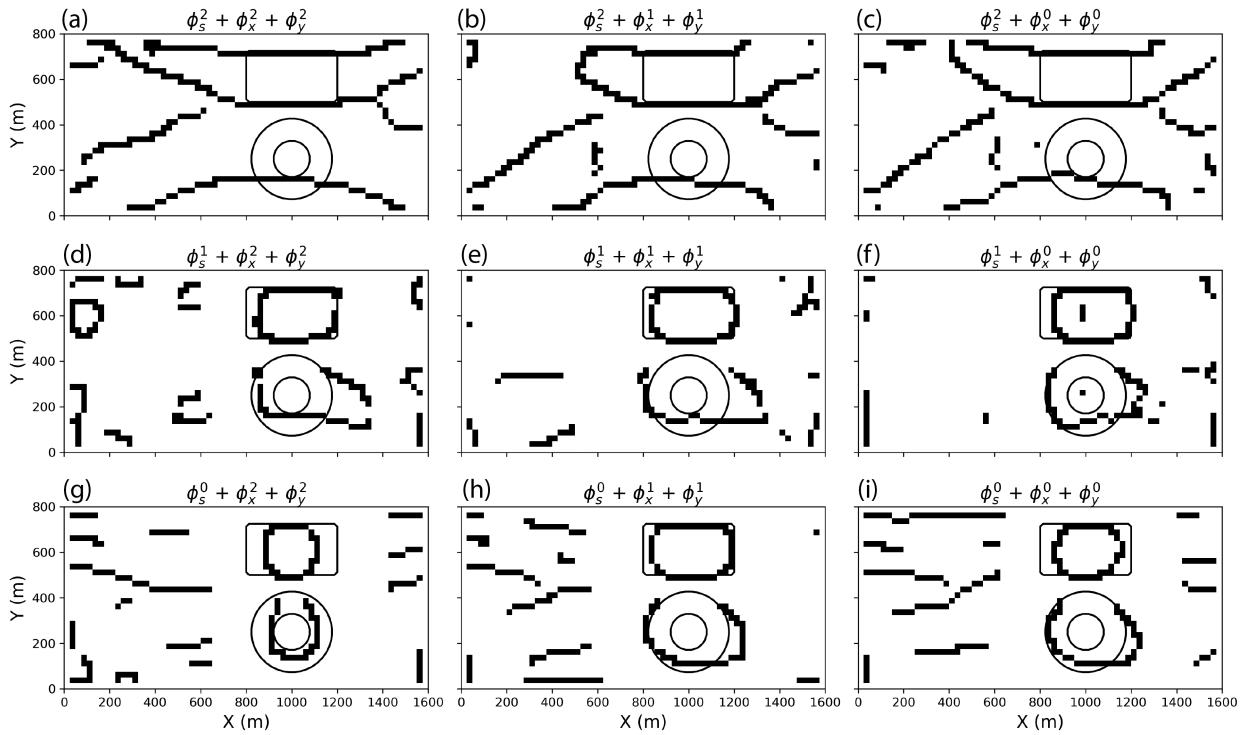


Figure 9. Parametric representation of the nine inverted models derived from the Canny edge detection algorithm.

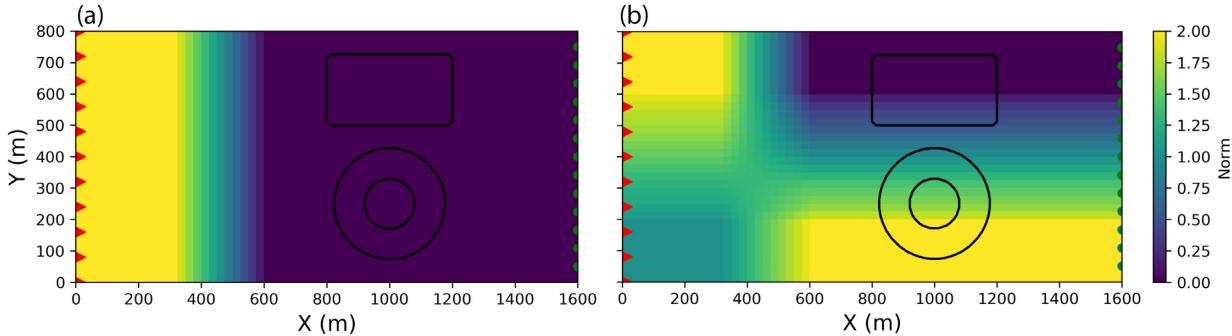


Figure 10. Local p values on (a) the model norm $\phi_s^{p_s}$ and (b) model gradients $\phi_x^{p_x}, \phi_y^{p_y}$ extracted from the solution space.

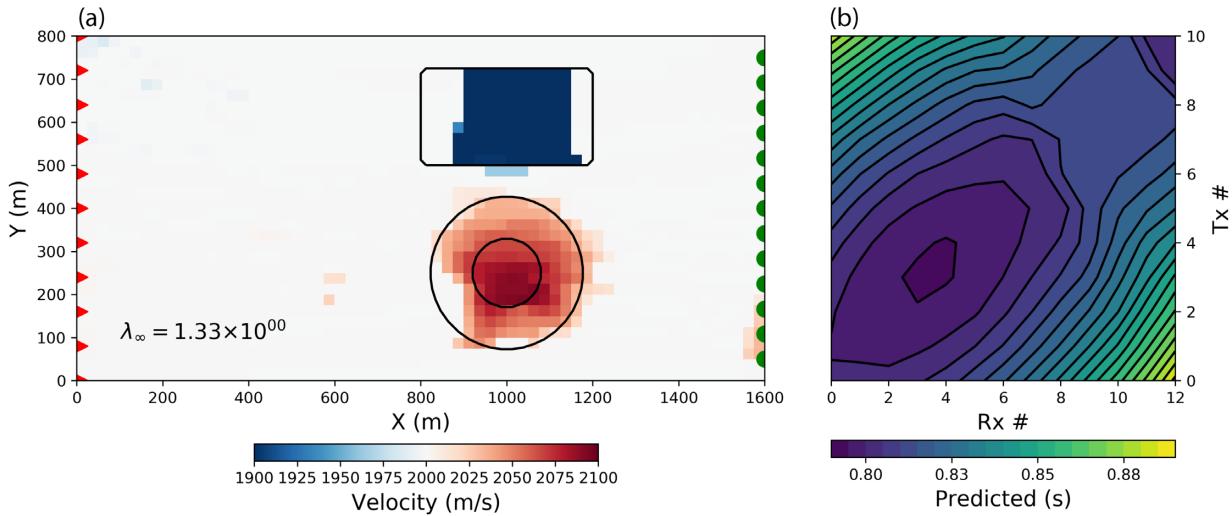


Figure 11. (a) The final SVMN inversion result (b) predicted data. Contour lines are shown in (a) for the true position and shape of the velocity anomalies.

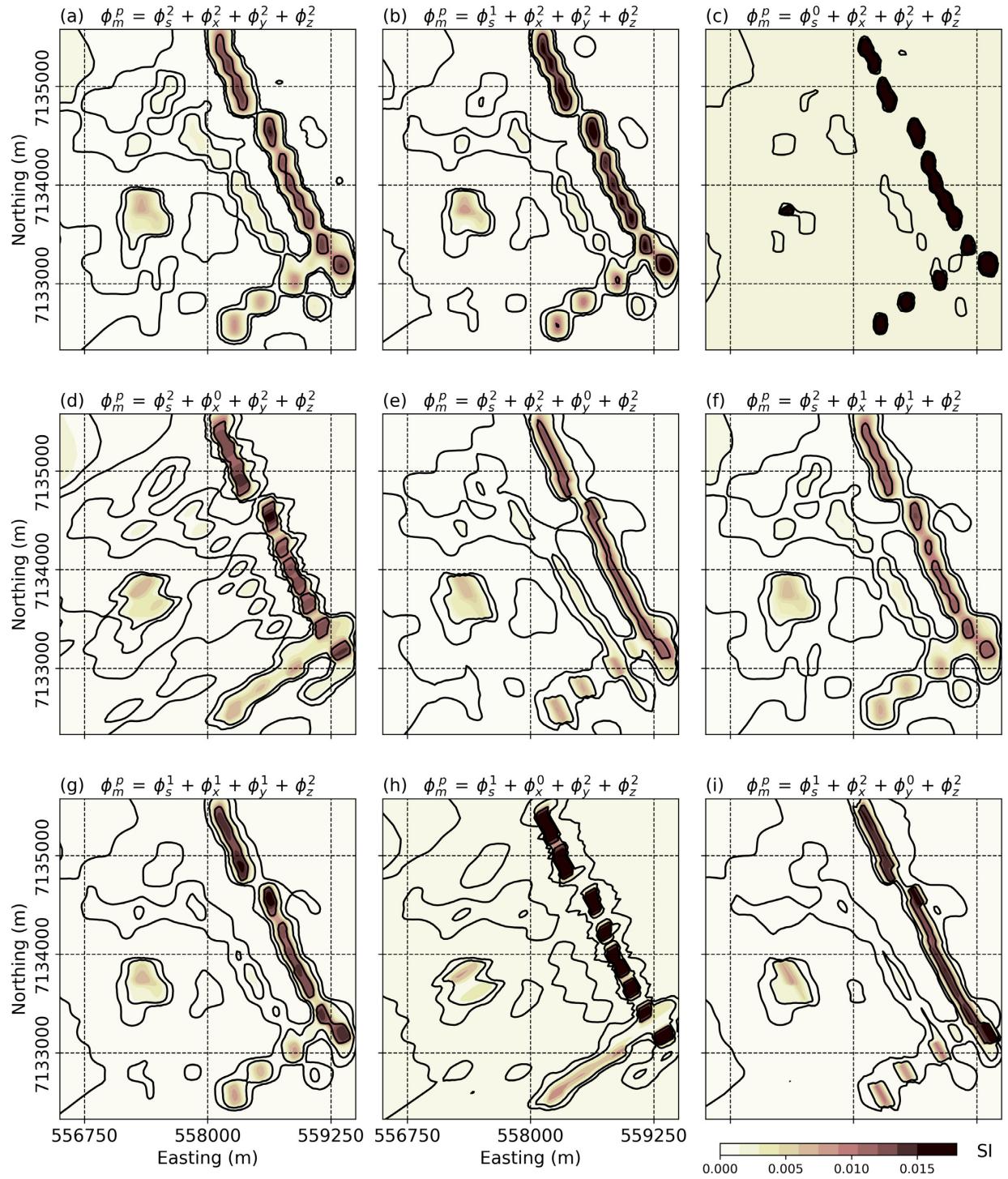


Figure 12. Horizontal sections through nine inversions of the airborne magnetic data over TKC after rotating back to UTM coordinates. Sparsity parameters are imposed along different directions to enforce lateral continuity.

accentuating trends along the dykes and better resolving the shape of the kimberlite pipes. Lowering the norm applied to specific model gradients promotes the recovery of piecewise continuous bodies along the same direction. So rather than applying an ℓ_p norm on all model gradients equally, we want to reduce the p value on each gradient independently. This would, in theory, require 81 inversions if we set $p_{s,x,y,z} = 2, 1, 0$. We narrow down our analysis to eight models based on the following criteria. (i) We expect the geology

to extend vertically, but the weak depth resolution of magnetic data does not allow us to distinguish clear trends for steeply dipping bodies. So we restrict our analysis to smooth vertical models ($p_z = 2$). (ii) We want to concentrate on continuity and character of the dykes in the two orthogonal directions. We therefore vary $p_{x,y} = 2, 1, 0$. Figs 12(b)–(i) presents horizontal sections at 75 m below the surface for eight subsequent inversion trials. We note that sparsity imposed on the model gradients along a specific Cartesian direction

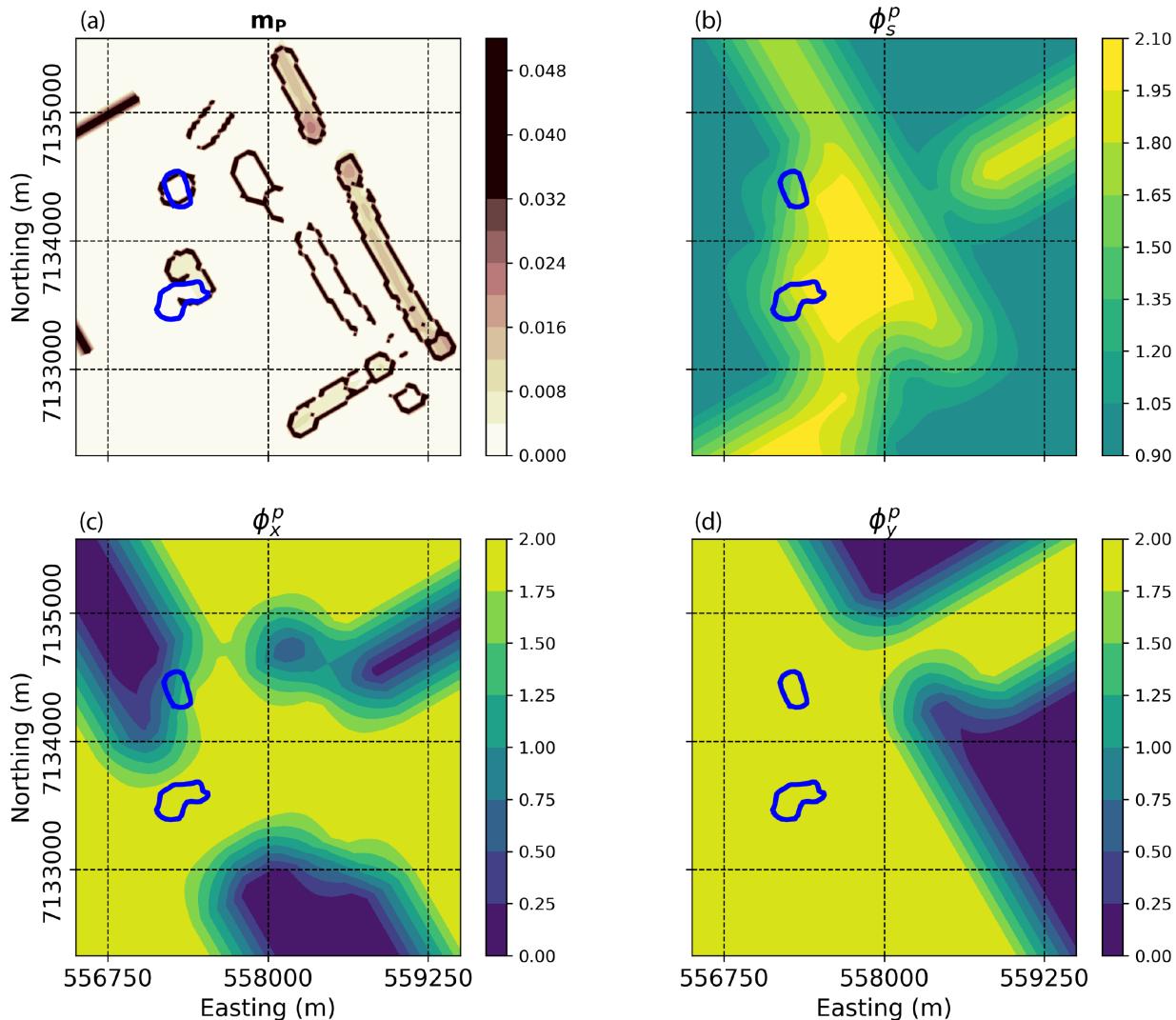


Figure 13. Estimated parameters from the correlation between all the inverted models. (a) The averaged PCA model \mathbf{m}_P and its Canny edge representation. (b–d) Average p values for ϕ_s^P , ϕ_x^P and ϕ_y^P , respectively. The outlines of the kimberlite pipes are shown for reference.

helps in recovering continuous features. While it is possible to manually divide the survey area into regions with variable norms (Fournier *et al.* 2016), we will attempt to extract these preferential orientations through the methodology prescribed in the previous section.

The following analysis steps are performed in 2-D along the same horizontal section at 75 m depth. We restrict our analysis to this horizontal depth section for three reasons: (i) we want to sample the cross-section of the DO-27 kimberlite pipe buried under a thick layer of glacial till, (ii) near surface features recovered in most models are dominated by the flight line separation, which we are trying to suppress with this process, (iii) we have stronger confidence in the geometry of the geology in 2-D.

We first perform a model averaging through a PCA followed by an edge detection analysis. The selection of an optimal window size and detection parameters used by the Canny algorithm is problem specific as it depends on the size and shape of features of interest. We used a 10×10 cells moving window (250×250 m). Fig. 13(a) presents the recovered \mathbf{m}_P overlaid by the edge solution. The different dyke orientations are clearly visible, as well as the circular anomalies associated with the kimberlites pipes. We then proceed

with a parameter extraction step based on the correlation between \mathbf{m}_P and the solution space. Sparse norms for the model and its gradients are shown in Figs 13(b)–(d).

In a final step, we proceed with the SVMN inversion. This yields the model presented in Fig. 14. The solution highlights the important features in the region:

- (1) Two continuous and narrow anomalies striking at 315°N . Both dykes show the same apparent break, which may correspond to left-lateral faulting event at Northing 71346500.
- (2) A third dyke striking at 60°N appears to overprint the signature of the weakest dykes.
- (3) The magnetic anomaly associated with DO-18 in the north coincides with the location of the known pipe.
- (4) The southern anomaly related to DO-27 shows higher susceptibility along the northern edge of the pipe. This anomaly was the initial target identified by the exploration team and later drilled. It is now understood that the main magnetic signal originating from DO-27 is associated with the hypabyssal kimberlite (HK) unit, a sheet-like intrusive mafic unit that has low diamondiferous content. Our inversion is thus consistent with the geology.

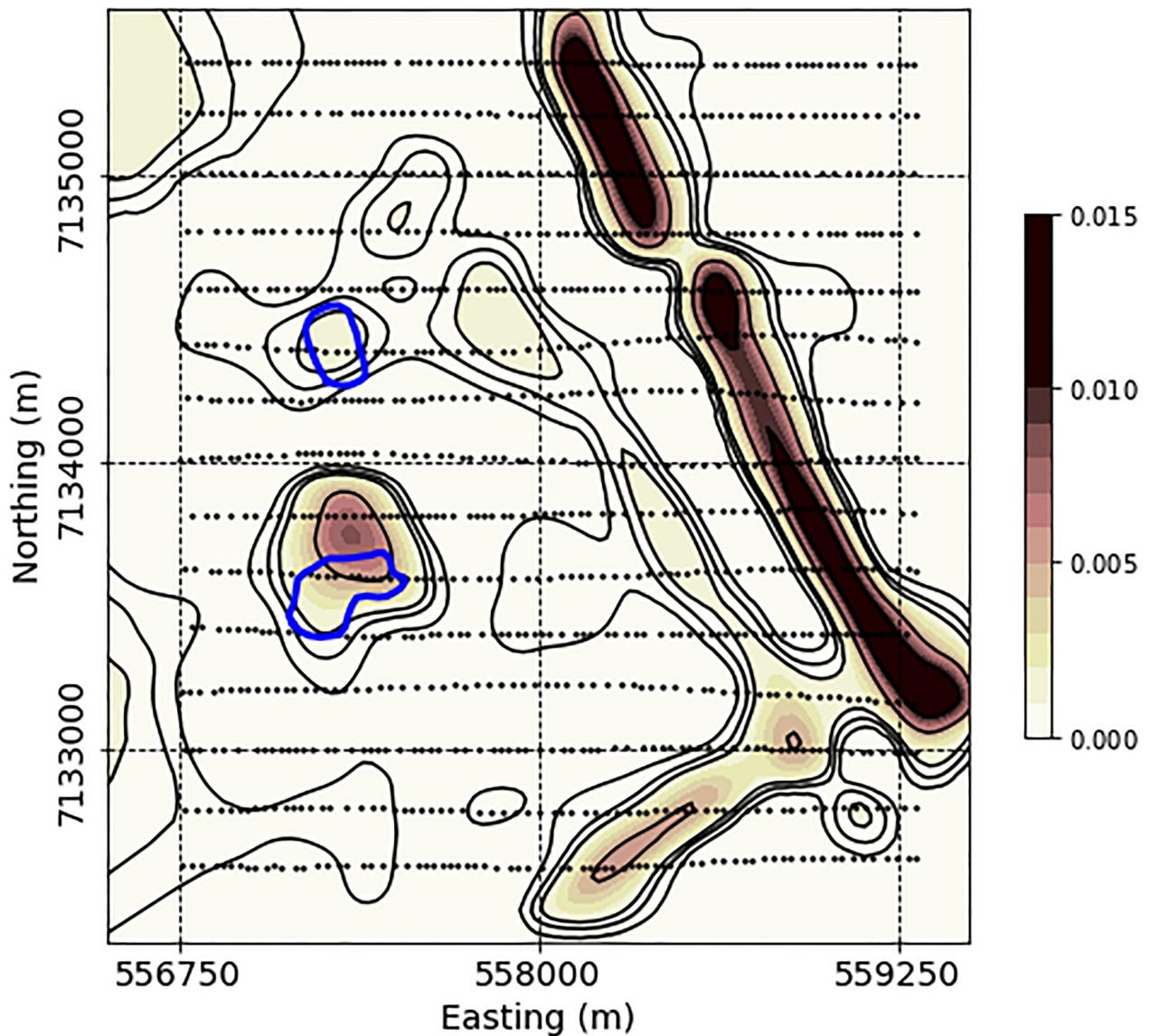


Figure 14. Final mixed-norm susceptibility model recovered over the DO-18/27 kimberlite pipes shown with blue outlines.

This final solution is an improvement over all nine previous inversions as it has managed to recover both elongate and continuous dykes striking in different directions, as well as compact isolated kimberlite pipes. A similar result has previously been shown in Fournier (2015), but that required tedious manual effort to select the norm models. We have accomplished the same result by using only the geophysical data and an entirely automated process for selecting parameters.

5 CONCLUSION

The geophysical inverse problem is inherently non-unique and the character of the solution is controlled by a regularization function that consists of multiple terms each of which is designed to affect the final outcome in a specific way. *Smallness* penalties force the solution to be close to a reference model; *roughness* penalties are focused on spatial variation. When composited into a final regularization, each term is pre-multiplied by a constant (α in our notation). Selecting appropriate values of α is challenging because the numerical value of the penalty depends upon the selected p norm as well

as the numerical discretization used to evaluate derivatives. As the number of penalty terms grows, and especially when each involves a different p -norm, pre-specifying the α values becomes challenging, time consuming and sometimes ineffective. In addition, solving an ℓ_p -norm problem when $p < 1$ poses additional challenges because the problem is no longer convex. To address these challenges, we take the following approach: (i) we replace the ℓ_p norms with the Lawson norm and, at each iteration, solve a local linear problem using IRLS; (ii) we use a two-stage process where, in the first stage, all penalties are ℓ_2 and in the second stage they are replaced by their final ℓ_p values; (iii) we replace the spatial derivatives with finite differences. This makes all penalty terms dimensionally equivalent and allows the α parameters to have a default value of unity; (iv) we rescale the gradients at each IRLS iteration so that all penalty terms can be equally influential in the final solution; (v) we introduce a proportionality ratio of the gradient vectors, λ_∞ , to measure success in achieving equal contribution from different terms in the regularization; (vi) we cool the parameter ϵ in the Lawson norm and use λ_∞ to indicate whether a slower cooling is needed. Although

we have portrayed our algorithm as one that produces a solution in which all terms in the regularization are equally effective in the final inversion, it is trivial to incorporate different weightings; λ_∞ then reflects the desired relative weightings.

Our inversion algorithm allows us explore two important items. The first is to sample the solution space associated with ℓ_p norms. The character of the model changes markedly with p value and inversions with different combinations of norms provides insight about which features might be required by the data and their geologic character. The importance of seeing multiple solutions to the same inverse problem cannot be overestimated.

The second important item is to construct a solution that encompasses different types of geologic features with a single domain. The geology at Tli Kwi Cho is a motivating example. At this level of complexity however, there are too many p parameters for a user to specify manually and some way to automate the procedure is required. We provide one strategy: (i) carry out a suite of ℓ_p inversions and build an ensemble of models; (ii) extract common features using a PCA analysis and build an average model \mathbf{m}_P ; (iii) in a windowed environment, perform correlations of ensemble member with \mathbf{m}_P ; for computing correlations we map the models to their Canny Edge representation; (iv) the suite of p values corresponding to the largest correlation coefficient is assigned to the cell at the centre of the moving window; (v) a final inversion, in which each cell has its own suite of p values, is carried out. This should be a candidate for a best model to interpret geologically.

Implementing the above process using SVMN has provided us with an image of magnetic susceptibility at TKC that is better related to geology than any other single inversion we have done. The methodology by which we generate and select the ensemble of possible solutions and the analysis by which we extract the SVMN parameters deserves further investigation. Finally, from a geometric perspective, we have so far used penalty terms on the model gradients measured along the three Cartesian directions. This is only appropriate if geological features are intersecting at right angles. Future work could increase the flexibility of the inversion by allowing gradient measures at variable angles.

ACKNOWLEDGEMENTS

The Authors would like to thank Peregrine Diamond Ltd. for giving us access to the airborne magnetic data. Special thanks goes to the Scikit-learn and SimPEG (Simulation and Parameter Estimation in Geophysics) development teams for their contribution to the open-source community.

REFERENCES

- Ajo-Franklin, J.B., Minsley, B.J. & Daley, T.M., 2007. Applying compact constraints to differential traveltime tomography, *Geophysics*, **72**(4), 67–75.
- Barbosa, V.C.F. & Silva, J.B.C., 1994. Generalized compact gravity inversion, *Geophysics*, **59**, 57–68.
- Cockett, R., Kang, S., Heagy, L.J., Pidlisecky, A. & Oldenburg, D.W., 2015. SimPEG: an open source framework for simulation and gradient based parameter estimation in geophysical applications, *Comput. Geosci.*, **85**(A), 142–154.
- Daubechies, I., DeVore, R., Fornasier, M. & Gunturk, C.S., 2010. Iteratively re-weighted least squares minimization for sparse recovery, *Commun. Pure Appl. Math.*, **63**(1), 1–35.
- Devriese, S.G.R., Davis, K. & Oldenburg, D.W., 2017. Inversion of airborne geophysics over the do-27/do-18 kimberlites part 1: potential fields, *Interpretation*, **5**(3), T299–T311.
- Ekblom, H., 1973. Calculation of linear best ℓ_p -approximation, *BIT*, **13**, 292–300.
- Farquharson, C.G. & Oldenburg, D.W., 1998. Non-linear inversion using general measures of data misfit and model structure, *Geophys. J. Int.*, **134**, 213–227.
- Fournier, D., 2015. *A Cooperative Magnetic Inversion Method with ℓ_p -Norm Regularization*, Msc. thesis, The University of British Columbia.
- Fournier, D., Davis, K. & Oldenburg, D.W., 2016. Robust and flexible mixed-norm inversion, in *SEG Technical Program Expanded Abstracts*, pp. 1542–1547, Soc. Explor. Geophys., Tulsa.
- Fournier, D., Kang, S., McMillan, M.S. & Oldenburg, D.W., 2017. Inversion of airborne geophysics over the do-27/do-18 kimberlites part 2: electromagnetics, *Interpretation*, **5**(3), T313–T325.
- Gorodnitsky, I.F. & Rao, B.D., 1997. Sparse singal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm, *IEEE Trans. Signal Process.*, **45**(3), 600–616.
- Hestenes, M.R. & Stiefel, E., 1952. Method of conjugate gradient for solving linear systems, *J. Res. Natl. Bur. Stand.*, **49**(6), 409–436.
- Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components, *J. Educ. Psychol.*, **24**(6 & 7), 417–441.
- Huber, P.J., 1964. Robust estimation of a location parameter, *Ann. Math. Stat.*, **35**, 73–101.
- Last, B.H. & Kubik, K., 1983. Compact gravity inversion, *Geophysics*, **48**, 713–721.
- Lawson, C.L., 1961. *Contribution to the Theory of Linear Least Maximum Approximation*, Ph.D. thesis, University of California.
- Li, Y., 1993. A globally convergent method for ℓ_p problems, *SIAM J. Sci. Stat. Comput.*, **3**, 609–629.
- Nocedal, J. & Wright, S.J., 1999. *Numerical Optimatization*, Springer Science.
- Pearson, K., 1901. On lines and plane of closest fit to systems of points in space, *Phil. Mag.*, **6**, 2(11), 559–572.
- Pedregosa, F. et al., 2011. Scikit-learn: machine learning in Python, *J. Mach. Learn. Res.*, **12**, 2825–2830.
- Pell, J.A., 1997. Kimberlites in the Slave Craton, Northwest Territories, Canada, *Geosci. Canada*, **24**(2), 77–90.
- Portniaguine, O.N., 1999. *Image Focussing and Data Compression in the Solution of Geophysical Inverse Problems*, PhD thesis, The University of Utah.
- Stocco, S., Godio, A. & Sambuelli, L., 2009. Modelling and compact inversion of magnetic data: a Matlab code, *Comput. Geosci.*, **35**, 2111–2118.
- Sun, J. & Li, Y., 2014. Adaptive ℓ_p inversion for simultaneous recovery of both blocky and smooth feature in geophysical model, *Geophys. J. Int.*, **197**, 882–899.
- Zhdanov, M. & Tolstaya, E., 2004. Minimum support nonlinear parametrization in the solution of a 3D magnetotelluric inverse problem, *Inverse Probl.*, **20**, 937–952.