

# Computational aspects of the equivalent-layer technique: review

Vanderlei C. Oliveira Jr.<sup>1</sup>, Diego Takahashi<sup>1</sup>, André L. A. Reis<sup>2</sup> and Valéria C. F. Barbosa<sup>1,\*</sup>

<sup>1</sup>*Observatório Nacional, Department of Geophysics, Rio de Janeiro, Brasil*

<sup>2</sup>*Universidade do Estado do Rio de Janeiro, Department of Applied Geology, Rio de Janeiro, Brasil*

Correspondence\*: Valéria C.F. Barbosa  
valcris@on.br

## 2 ABSTRACT

3 Equivalent-layer technique is a powerful tool for processing potential-field data in the space  
4 domain. However, the greatest hindrance for using the equivalent-layer technique is its high  
5 computational cost for processing massive data sets. The large amount of computer memory  
6 usage to store the full sensitivity matrix combined with the computational time required for  
7 matrix-vector multiplications and to solve the resulting linear system, are the main drawbacks  
8 that made unfeasible the use of the equivalent-layer technique for a long time. More recently, the  
9 advances in computational power propelled the development of methods to overcome the heavy  
10 computational cost associated with the equivalent-layer technique. We present a comprehensive  
11 review of the computation aspects concerning the equivalent-layer technique addressing how  
12 previous works have been dealt with the computational cost of this technique. Historically, the  
13 high computational cost of the equivalent-layer technique has been overcome by using a variety  
14 of strategies such as: moving data-window scheme, column- and row-action updates of the  
15 sensitivity matrix, reparametrization, sparsity induction of the sensitivity matrix, iterative methods  
16 using the full sensitivity matrix, iterative deconvolution by using the concept of block-Toeplitz  
17 Toeplitz-block (BTTB) matrices and direct deconvolution. We compute the number of floating-point  
18 operations of some of these strategies adopted in the equivalent-layer technique to show their  
19 effectiveness in reducing the computational demand. Numerically, we also address the stability of  
20 some of these strategies used in the equivalent-layer technique by comparing with the stability  
21 via the classic equivalent-layer technique with the zeroth-order Tikhonov regularization. We  
22 show that even for the most computationally efficient methods, which can save up to  $10^9$  flops,  
23 the stability of the linear system is maintained. The two most efficient strategies, iterative and  
24 direct deconvolutions, can process large datasets quickly and yield good results. However, direct  
25 deconvolution has some drawbacks. Real data from Carajás Mineral Province, Brazil, is also  
26 used to validate the results showing a potential field transformation.

27 **Keywords:** equivalent layer, gravimetry, fast algorithms, computational cost, stability analysis

## 1 INTRODUCTION

The equivalent-layer technique has been used by exploration geophysicists for processing potential-field data since the late 1960s (Dampney, 1969). This technique is based on a widely accepted principle, which states that a discrete set of observed potential-field data due to 3D sources can be approximated by that due to a discrete set of virtual sources (such as point masses, dipoles, prisms, doublets). From a theoretical point of view, the equivalent-layer technique is grounded on potential theory (Kellogg, 1967) and consists in considering that the potential field data can be approximated by a linear combination of harmonic functions describing the potential field due to the virtual sources. These sources, commonly called equivalent sources, are arranged on a layer with finite horizontal dimensions and located below the observations. In the classical approach, a linear inverse problem is solved to estimate the physical property of each equivalent source subject to fit the observations. Then, the estimated physical-property distribution on the equivalent layer is used to accomplish the desired potential-field transformation (e.g., interpolation, upward/downward continuation, reduction to the pole). The later step is done by multiplying the estimated physical-property distribution by the matrix of Green's functions associated with the desired potential-field transformation.

Because the linear inverse problem to be solved in the equivalent-layer technique is set up with a full sensitivity matrix, its computational cost strongly depends on the number of potential-field observations and can be very inefficient for dealing with massive data sets. To overcome this problem, computationally efficient methods based on equivalent-layer technique have arose in the late 1980s. This comprehensive review discusses specific strategies aimed at reducing the computational cost of the equivalent-layer technique. These strategies are addressed in the following articles: Leão and Silva (1989); Cordell (1992); Xia et al. (1993); Mendonça and Silva (1994); Guspí and Novara (2009); Li and Oldenburg (2010); Oliveira Jr. et al. (2013); Siqueira et al. (2017); Jirigalatu and Ebbing (2019); Takahashi et al. (2020, 2022); Mendonça (2020); and Soler and Uieda (2021);

To our knowledge, the first method towards improving the efficiency was proposed by Leão and Silva (1989), who used an overlapping moving-window scheme spanning the data set. The strategy adopted in Leão and Silva (1989) involves solving several smaller, regularized linear inverse problems instead of one large problem. This strategy uses a small data window and distributes equivalent sources on a small regular grid at a constant depth below the data surface, with the sources' window extending beyond the boundaries of the data window. Because of the spatial layouts of observed data and equivalent sources in Leão and Silva (1989), the small sensitivity submatrix containing the coordinates of the data and equivalent sources within a window remains constant for all data windows. This holds true regardless of the specific locations of the data and equivalent sources within each window. For each position of the data window, this scheme consists in computing the processed field at the center of the data window only, and the next estimates of the processed field are obtained by shifting the data window across the entire dataset. More recently, Soler and Uieda (2021) extended the method introduced by Leão and Silva (1989) to accommodate irregularly spaced data collected on a non-flat surface. Unlike Leão and Silva (1989), in the generalization proposed by Soler and Uieda (2021), the sensitivity submatrix that includes the coordinates of the data and equivalent sources needs to be computed for each window. Soler and Uieda (2021) developed a computational approach to further enhance the efficiency of the equivalent-layer technique by combining two strategies. The first one — the block-averaging source locations — reduces the number of model parameters and the second strategy — the gradient-boosted algorithm — reduces the size of the linear system to be solved by iteratively fitting the

70 equivalent source model along overlapping windows. It is worth noting that the equivalent-layer strategy of  
71 using a moving-window scheme either in Leão and Silva (1989) or in Soler and Uieda (2021) is similar to  
72 discrete convolution.

73 As another strategy to reduce the computational workload of the equivalent-layer technique, some authors  
74 have employed column- and row-action updates, which are commonly applied to image reconstruction  
75 methods (e.g., Elfving et al., 2017). These methods involve iterative calculations of a single column  
76 and a single row of the sensitivity matrix, respectively. Following the strategy column-action update,  
77 Cordell (1992) proposed a computational method in which a single equivalent source positioned below a  
78 measurement station is iteratively used to compute both the predicted data and residual data for all stations.  
79 In Cordell's method, a single column of the sensitivity matrix is calculated per iteration, meaning that  
80 a single equivalent source contributes to data fitting in each iteration. Guspí and Novara (2009) further  
81 extended Cordell's method by applying it to scattered magnetic observations. Following the strategy of  
82 column-action update, Mendonça and Silva (1994) developed an iterative procedure where one data point  
83 is incorporated at a time, and a single row of the sensitivity matrix is calculated per iteration. This strategy  
84 adopted by Mendonça and Silva (1994) is known as *equivalent data concept*. This concept is based on  
85 the principle that certain data points within a dataset are redundant and, as a result, do not contribute to  
86 the final solution. On the other hand, there is a subset of observations known as equivalent data, which  
87 effectively contributes to the final solution and fits the remaining redundant data. In their work, Mendonça  
88 and Silva (1994) adopted an iterative approach to select a substantially smaller subset of equivalent data  
89 from the original dataset.

90 The next strategy involves reparametrizing the equivalent layer with the objective of solving a smaller  
91 linear inverse problem by reducing the dimension of the model space. Oliveira Jr. et al. (2013) reduced  
92 the model parameters by approximating the equivalent-source layer by a piecewise-polynomial function  
93 defined on a set of user-defined small equivalent-source windows. The estimated parameters are the  
94 polynomial coefficients for each window and they are much smaller than the original number of equivalent  
95 sources. By using the subspace method, Mendonça (2020) reparametrizes the equivalent layer, which  
96 involves reducing the dimension of the linear system from the original parameter-model space to a lower-  
97 dimensional subspace. The subspace bases span the parameter-model space and they are constructed by  
98 applying the singular value decomposition to the matrix containing the gridded data.

99 Following the strategy of sparsity induction, Li and Oldenburg (2010) transformed the full sensitivity  
100 matrix into a sparse one using orthonormal compactly supported wavelets. Barnes and Lumley (2011)  
101 proposed an alternative approach to introduce sparsity based on the use of quadtree discretization to group  
102 equivalent sources far from the computation points. Those authors explore the induced sparsity by using  
103 specific iterative methods to solve the linear system.

104 The strategy named iterative methods estimates iteratively the parameter vector that represents a  
105 distribution over an equivalent layer. Xia and Sprowl (1991) and Xia et al. (1993) have developed efficient  
106 iterative algorithms for updating the distribution of physical properties within the equivalent layer in the  
107 wavenumber and space domains, respectively. Specifically, in Xia and Sprowl's (1991) method the physical-  
108 property distribution is updated by using the ratio between the squared depth to the equivalent source and  
109 the gravitational constant multiplied by the residual between the observed and predicted observation at the  
110 measurement station. Siqueira et al. (2017) developed an iterative solution where the sensitivity matrix  
111 is transformed into a diagonal matrix with constant terms through the use of the *excess mass criterion*  
112 and of the positive correlation between the observed gravity data and the masses on the equivalent layer.  
113 The fundamentals of the Siqueira et al.'s method is based on the Gauss' theorem (e.g., Kellogg, 1967,

114 p. 43) and the total excess of mass (e.g., Blakely, 1996, p. 60). All these iterative methods use the full  
115 and dense sensitivity matrix to calculate the predicted data and residual data in the whole survey data per  
116 iteration. Hence, the iterative methods proposed by Xia and Sprowl (1991), Xia et al. (1993) and Siqueira  
117 et al. (2017) neither compress nor reparametrize the sensitivity matrix. Jirigalatu and Ebbing (2019) also  
118 proposed an iterative equivalent layer that uses the full and dense sensitivity matrix. However, in their  
119 approach, Jirigalatu and Ebbing (2019) efficiently compute the predicted data and residual data for the  
120 entire survey per iteration in the wavenumber domain.

121 Following the strategy of the iterative deconvolution, Takahashi et al. (2020, 2022) developed fast and  
122 effective equivalent-layer techniques for processing, respectively, gravity and magnetic data by modifying  
123 the forward modeling to estimate the physical-property distribution over the equivalent layer through a  
124 2D discrete fast convolution. These methods took advantage of the Block-Toeplitz Toeplitz-block (BTTB)  
125 structure of the sensitivity matrices, allowing them to be calculated by using only their first column. In  
126 practice, the forward modeling uses a single equivalent source, which significantly reduces the required  
127 RAM memory.

128 The method introduced by Takahashi et al. (2020, 2022) can be reformulated to eliminate the need for  
129 conjugate gradient iterations. This reformulation involves employing a *direct deconvolution* approach (e.g.,  
130 Aster et al., 2019, p. 220) with *Wiener filter* (e.g., Gonzalez and Woods, 2002, p. 263).

131 Here, we present a comprehensive review of diverse strategies to solve the linear system of the equivalent  
132 layer alongside an analysis of the computational cost and stability of these strategies. To do this analysis,  
133 we are using the floating-point operations count to evaluate the performance of a selected set of methods  
134 (e.g., Leão and Silva (1989); Cordell (1992); Oliveira Jr. et al. (2013); Siqueira et al. (2017); Mendonça  
135 (2020); Takahashi et al. (2020); Soler and Uieda (2021); and direct deconvolution). To test the stability, we  
136 are using the linear system sensitivity to noise as a comparison parameter for the fastest of these methods  
137 alongside the classical normal equations. A potential-field transformation will also be used to evaluate the  
138 quality of the equivalent sources estimation results using both synthetic and real data from Carajás Mineral  
139 Province, Brazil.

140 In the following sections, we will address the theoretical bases of the equivalent-layer technique, including  
141 aspects such as the sensitivity matrix, layer depth and spatial distribution and the total number of equivalent  
142 sources. Then, we will explore the general formulation and solution of the linear inverse problem for  
143 the equivalent-layer technique, including discussions on linear system solvers. Additionally, we will  
144 quantify the required arithmetic operations for a given equivalent-layer method, assessing the number of  
145 floating-point operations involved. Next, we will evaluate the stability of the estimated solutions obtained  
146 from applying specific equivalent-layer methods. Finally, we will delve into the computational strategies  
147 adopted in the equivalent-layer technique for reducing computational costs. These strategies encompass  
148 various approaches, such as the moving data-window scheme, column- and row-action updates of the  
149 sensitivity matrix, reparametrization, sparsity induction of the sensitivity matrix, iterative methods using the  
150 full sensitivity matrix, iterative deconvolution using the concept of block-Toeplitz Toeplitz-block (BTTB)  
151 matrices, and direct deconvolution.

## 2 FUNDAMENTALS

152 Let  $\mathbf{d}$  be a  $D \times 1$  vector, whose  $i$ -th element  $d_i$  is the observed potential field at the position  $(x_i, y_i, z_i)$ ,  
 153  $i \in \{1 : D\}$ , of a topocentric Cartesian system with  $x$ ,  $y$  and  $z$  axes pointing to north, east and down,  
 154 respectively. Consider that  $d_i$  can be satisfactorily approximated by a harmonic function

$$f_i = \sum_{j=1}^P g_{ij} p_j, \quad i \in \{1 : D\}, \quad (1)$$

155 where,  $p_j$  represents the scalar physical property of a virtual source (i.e., monopole, dipole, prism) located  
 156 at  $(x_j, y_j, z_j)$ ,  $j \in \{1 : P\}$  and

$$g_{ij} \equiv g(x_i - x_j, y_i - y_j, z_i - z_j), \quad z_i < \min\{z_j\}, \quad \forall i \in \{1 : D\}, \quad (2)$$

157 is a harmonic function, where  $\min\{z_j\}$  denotes the minimum  $z_j$ , or the vertical coordinate of the shallowest  
 158 virtual source. These virtual sources are called *equivalent sources* and they form an *equivalent layer*. In  
 159 matrix notation, the potential field produced by all equivalent sources at all points  $(x_i, y_i, z_i)$ ,  $i \in \{1 : D\}$ ,  
 160 is given by:

$$\mathbf{f} = \mathbf{G}\mathbf{p}, \quad (3)$$

161 where  $\mathbf{p}$  is a  $P \times 1$  vector with  $j$ -th element  $p_j$  representing the scalar physical property of the  $j$ -th  
 162 equivalent source and  $\mathbf{G}$  is a  $D \times P$  matrix with element  $g_{ij}$  given by equation 2.

163 The equivalent-layer technique consists in solving a linear inverse problem to determine a parameter  
 164 vector  $\mathbf{p}$  leading to a predicted data vector  $\mathbf{f}$  (equation 3) *sufficiently close to* the observed data vector  $\mathbf{d}$ ,  
 165 whose  $i$ -th element  $d_i$  is the observed potential field at  $(x_i, y_i, z_i)$ . The notion of *closeness* is intrinsically  
 166 related to the concept of *vector norm* (e.g., Golub and Van Loan, 2013, p. 68) or *measure of length* (e.g.,  
 167 Menke, 2018, p. 41). Because of that, almost all methods for determining  $\mathbf{p}$  actually estimate a parameter  
 168 vector  $\tilde{\mathbf{p}}$  minimizing a length measure of the difference between  $\mathbf{f}$  and  $\mathbf{d}$  (see subsection 3.1). Given an  
 169 estimate  $\tilde{\mathbf{p}}$ , it is then possible to compute a potential field transformation

$$\mathbf{t} = \mathbf{A}\tilde{\mathbf{p}}, \quad (4)$$

170 where  $\mathbf{t}$  is a  $T \times 1$  vector with  $k$ -th element  $t_k$  representing the transformed potential field at the position  
 171  $(x_k, y_k, z_k)$ ,  $k \in \{1 : T\}$ , and

$$a_{kj} \equiv a(x_k - x_j, y_k - y_j, z_k - z_j), \quad z_k < \min\{z_j\}, \quad \forall k \in \{1 : T\}, \quad (5)$$

172 is a harmonic function representing the  $kj$ -th element of the  $T \times P$  matrix  $\mathbf{A}$ .

### 173 2.1 Spatial distribution and total number of equivalent sources

174 There is no well-established criteria to define the optimum number  $P$  or the spatial distribution of the  
 175 equivalent sources. We know that setting an equivalent layer with more (less) sources than potential-field  
 176 data usually leads to an underdetermined (overdetermined) inverse problem (e.g., Menke, 2018, p. 52–53).  
 177 Concerning the spatial distribution of the equivalent sources, the only condition is that they must rely on a  
 178 surface that is located below and does not cross that containing the potential field data. Soler and Uieda  
 179 (2021) present a practical discussion about this topic.

180 From a theoretical point of view, the equivalent layer reproducing a given potential field data set cannot  
 181 cross the true gravity or magnetic sources. This condition is a consequence of recognizing that the equivalent  
 182 layer is essentially an indirect solution of a boundary value problem of potential theory (e.g., Roy, 1962;  
 183 Zidarov, 1965; Dampney, 1969; Li et al., 2014; Reis et al., 2020). In practical applications, however, there  
 184 is no guarantee that this condition is satisfied. Actually, it is widely known from practical experience (e.g.,  
 185 Gonzalez et al., 2022) that the equivalent-layer technique works even for the case in which the layer cross  
 186 the true sources.

187 Regarding the depth of the equivalent layer, Dampney (1969) proposed a criterion based on horizontal data  
 188 sampling, suggesting that the equivalent-layer depth should be between two and six times the horizontal grid  
 189 spacing, considering evenly spaced data. However, when dealing with a survey pattern that has unevenly  
 190 spaced data, Reis et al. (2020) adopted an alternative empirical criterion. According to their proposal,  
 191 the depth of the equivalent layer should range from two to three times the spacing between adjacent  
 192 flight lines. The criteria of Dampney (1969) and Reis et al. (2020) are valid for planar equivalent layers.  
 193 Cordell (1992) have proposed and an alternative criterion for scattered data that leads to an undulating  
 194 equivalent layer. This criterion have been slightly modified by Guspí et al. (2004), Guspí and Novara  
 195 (2009) and Soler and Uieda (2021), for example, and consists in setting one equivalent source below  
 196 each datum at a depth proportional to the horizontal distance to the nearest neighboring data points. Soler  
 197 and Uieda (2021) have compared different strategies for defining the equivalent sources depth for the  
 198 specific problem of interpolating gravity data, but they have not found significant differences between them.  
 199 Regarding the horizontal layout, Soler and Uieda (2021) proposed the block-averaged sources locations  
 200 in which the survey area is divided into horizontal blocks and one single equivalent source is assigned  
 201 to each block. The horizontal coordinates of the single source in a given block is defined by the average  
 202 horizontal coordinates of the observation points at the block. According to Soler and Uieda (2021), this  
 203 block-averaged layout may prevent aliasing of the interpolated values, specially when the observations  
 204 are unevenly sampled. This strategy also reduces the number of equivalent sources without affecting the  
 205 accuracy of the potential-field interpolation. Besides, it reduces the computational load for estimating the  
 206 physical property on the equivalent layer.

## 207 2.2 Matrix G

208 Generally, the harmonic function  $g_{ij}$  (equation 2) is defined in terms of the inverse distance between the  
 209 observation point  $(x_i, y_i, z_i)$  and the  $j$ -th equivalent source at  $(x_j, y_j, z_j)$ ,

$$\frac{1}{r_{ij}} \equiv \frac{1}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}}, \quad (6)$$

210 or by its partial derivatives of first and second orders, respectively given by

$$\partial_\alpha \frac{1}{r_{ij}} \equiv \frac{-(\alpha_i - \alpha_j)}{r_{ij}^3}, \quad \alpha \in \{x, y, z\}, \quad (7)$$

211 and

$$\partial_{\alpha\beta} \frac{1}{r_{ij}} \equiv \begin{cases} \frac{3(\alpha_i - \alpha_j)^2}{r_{ij}^5}, & \alpha = \beta, \\ \frac{3(\alpha_i - \alpha_j)(\beta_i - \beta_j)}{r_{ij}^5} - \frac{1}{r_{ij}^3}, & \alpha \neq \beta, \end{cases} \quad \alpha, \beta \in \{x, y, z\}. \quad (8)$$

212 In this case, the equivalent layer is formed by punctual sources representing monopoles or dipoles (e.g.,  
 213 Dampney, 1969; Emilia, 1973; Leão and Silva, 1989; Cordell, 1992; Oliveira Jr. et al., 2013; Siqueira et al.,  
 214 2017; Reis et al., 2020; Takahashi et al., 2020; Soler and Uieda, 2021; Takahashi et al., 2022). Another  
 215 common approach consists in not defining  $g_{ij}$  by using equations 6–8, but other harmonic functions  
 216 obtained by integrating them over the volume of regular prisms (e.g., Li and Oldenburg, 2010; Barnes and  
 217 Lumley, 2011; Li et al., 2014; Jirgalatu and Ebbing, 2019). There are also some less common approaches  
 218 defining the harmonic function  $g_{ij}$  (equation 2) as the potential field due to plane faces with constant  
 219 physical property (Hansen and Miyazaki, 1984), doublets (Silva, 1986) or by computing the double  
 220 integration of the inverse distance function with respect to  $z$  (Guspí and Novara, 2009).

221 A common assumption for most of the equivalent-layer methods is that the harmonic function  $g_{ij}$   
 222 (equation 2) is independent on the actual physical relationship between the observed potential field and  
 223 their true sources (e.g., Cordell, 1992; Guspí and Novara, 2009; Li et al., 2014). Hence,  $g_{ij}$  can be  
 224 defined according to the problem. The only condition imposed to this function is that it decays to zero  
 225 as the observation point  $(x_i, y_i, z_i)$  goes away from the position  $(x_j, y_j, z_j)$  of the  $j$ -th equivalent source.  
 226 However, several methods use a function  $g_{ij}$  that preserves the physical relationship between the observed  
 227 potential field and their true sources. For the case in which the observed potential field is gravity data,  $g_{ij}$   
 228 is commonly defined as a component of the gravitational field produced at  $(x_i, y_i, z_i)$  by a point mass or  
 229 prism located at  $(x_j, y_j, z_j)$ , with unit density. On the other hand,  $g_{ij}$  is commonly defined as a component  
 230 of the magnetic induction field produced at  $(x_i, y_i, z_i)$  by a dipole or prism located at  $(x_j, y_j, z_j)$ , with unit  
 231 magnetization intensity, when the observed potential field is magnetic data.

232 The main challenge in the equivalent-layer technique is the computational complexity associated with  
 233 handling large datasets. This complexity arises because the sensitivity matrix  $\mathbf{G}$  (equation 3) is dense  
 234 regardless of the harmonic function  $g_{ij}$  (equation 2) employed. In the case of scattered potential-field  
 235 data, the structure of  $\mathbf{G}$  is not well-defined, regardless of the spatial distribution of the equivalent sources.  
 236 However, in a specific scenario where (i) each potential-field datum is directly associated with a single  
 237 equivalent source located directly below it, and (ii) both the data and sources are based on planar and  
 238 regularly spaced grids, Takahashi et al. (2020, 2022) demonstrate that  $\mathbf{G}$  exhibits a block-Toeplitz Toeplitz-  
 239 block (BTTB) structure. In such cases, the product of  $\mathbf{G}$  and an arbitrary vector can be efficiently computed  
 240 using a 2D fast Fourier transform as a discrete convolution.

### 3 LINEAR INVERSE PROBLEM OF EQUIVALENT-LAYER TECHNIQUE

#### 241 3.1 General formulation

242 A general formulation for almost all equivalent-layer methods can be achieved by first considering that  
 243 the  $P \times 1$  parameter vector  $\mathbf{p}$  (equation 3) can be reparameterized into a  $Q \times 1$  vector  $\mathbf{q}$  according to:

$$\mathbf{p} = \mathbf{H} \mathbf{q}, \quad (9)$$

244 where  $\mathbf{H}$  is a  $P \times Q$  matrix. The predicted data vector  $\mathbf{f}$  (equation 3) can then be rewritten as follows:

$$\mathbf{f} = \mathbf{G} \mathbf{H} \mathbf{q}. \quad (10)$$

245 Note that the original parameter vector  $\mathbf{p}$  is defined in a  $P$ -dimensional space whereas the reparameterized  
 246 parameter vector  $\mathbf{q}$  (equation 9) lies in a  $Q$ -dimensional space. For convenience, we use the terms  $P$ -space  
 247 and  $Q$ -space to designate them.

248 In this case, the problem of estimating a parameter vector  $\tilde{\mathbf{p}}$  minimizing a length measure of the difference  
 249 between  $\mathbf{f}$  (equation 3) and  $\mathbf{d}$  is replaced by that of estimating an auxiliary vector  $\tilde{\mathbf{q}}$  minimizing the goal  
 250 function

$$\Gamma(\mathbf{q}) = \Phi(\mathbf{q}) + \mu \Theta(\mathbf{q}), \quad (11)$$

251 which is a combination of particular measures of length given by

$$\Phi(\mathbf{q}) = (\mathbf{d} - \mathbf{f})^\top \mathbf{W}_d (\mathbf{d} - \mathbf{f}), \quad (12)$$

252 and

$$\Theta(\mathbf{q}) = (\mathbf{q} - \bar{\mathbf{q}})^\top \mathbf{W}_q (\mathbf{q} - \bar{\mathbf{q}}), \quad (13)$$

253 where the regularization parameter  $\mu$  is a positive scalar controlling the trade-off between the data-misfit  
 254 function  $\Phi(\mathbf{q})$  and the regularization function  $\Theta(\mathbf{q})$ ;  $\mathbf{W}_d$  is a  $D \times D$  symmetric matrix defining the relative  
 255 importance of each observed datum  $d_i$ ;  $\mathbf{W}_q$  is a  $Q \times Q$  symmetric matrix imposing prior information on  $\mathbf{q}$ ;  
 256 and  $\bar{\mathbf{q}}$  is a  $Q \times 1$  vector of reference values for  $\mathbf{q}$  that satisfies

$$\bar{\mathbf{p}} = \mathbf{H} \bar{\mathbf{q}}, \quad (14)$$

257 where  $\bar{\mathbf{p}}$  is a  $P \times 1$  vector containing reference values for the original parameter vector  $\mathbf{p}$ .

258 After obtaining an estimate  $\tilde{\mathbf{q}}$  for the reparameterized parameter vector  $\mathbf{q}$  (equation 9), the estimate  $\tilde{\mathbf{p}}$  for  
 259 the original parameter vector (equation 3) is computed by

$$\tilde{\mathbf{p}} = \mathbf{H} \tilde{\mathbf{q}}. \quad (15)$$

260 The reparameterized vector  $\tilde{\mathbf{q}}$  is obtained by first computing the gradient of  $\Gamma(\mathbf{q})$ ,

$$\nabla \Gamma(\mathbf{q}) = -2 \mathbf{H}^\top \mathbf{G}^\top \mathbf{W}_d (\mathbf{d} - \mathbf{f}) + 2\mu \mathbf{W}_q (\mathbf{q} - \bar{\mathbf{q}}). \quad (16)$$

261 Then, by considering that  $\nabla \Gamma(\tilde{\mathbf{q}}) = \mathbf{0}$  (equation 16), where  $\mathbf{0}$  is a vector of zeros, as well as adding and  
 262 subtracting the term  $(\mathbf{H}^\top \mathbf{G}^\top \mathbf{W}_d \mathbf{G} \mathbf{H}) \bar{\mathbf{q}}$ , we obtain

$$\tilde{\boldsymbol{\delta}}_q = \mathbf{B} \boldsymbol{\delta}_d, \quad (17)$$

263 where

$$\tilde{\mathbf{q}} = \tilde{\boldsymbol{\delta}}_q + \bar{\mathbf{q}}, \quad (18)$$

$$\boldsymbol{\delta}_d = \mathbf{d} - \mathbf{G} \mathbf{H} \bar{\mathbf{q}}, \quad (19)$$

$$\mathbf{B} = (\mathbf{H}^\top \mathbf{G}^\top \mathbf{W}_d \mathbf{G} \mathbf{H} + \mu \mathbf{W}_q)^{-1} \mathbf{H}^\top \mathbf{G}^\top \mathbf{W}_d, \quad (20)$$

266 or, equivalently (Menke, 2018, p. 62),

$$\mathbf{B} = \mathbf{W}_q^{-1} \mathbf{H}^\top \mathbf{G}^\top (\mathbf{G} \mathbf{H} \mathbf{W}_q^{-1} \mathbf{H}^\top \mathbf{G}^\top + \mu \mathbf{W}_d^{-1})^{-1}. \quad (21)$$

267 Evidently, we have considered that all inverses exist in equations 20 and 21.

268 The  $Q \times D$  matrix  $\mathbf{B}$  defined by equation 20 is commonly used for the case in which  $D > Q$ , i.e., when  
 269 there are more data than parameters (overdetermined problems). In this case, we consider that the estimate

270  $\tilde{\mathbf{q}}$  is obtained by solving the following linear system for  $\tilde{\boldsymbol{\delta}}_q$  (equation 18):

$$\left( \mathbf{H}^\top \mathbf{G}^\top \mathbf{W}_d \mathbf{G} \mathbf{H} + \mu \mathbf{W}_q \right) \tilde{\boldsymbol{\delta}}_q = \mathbf{H}^\top \mathbf{G}^\top \mathbf{W}_d \boldsymbol{\delta}_d . \quad (22)$$

271 On the other hand, for the cases in which  $D < Q$  (underdetermined problems), matrix  $\mathbf{B}$  is usually defined  
272 according to equation 21. In this case, the general approach involves estimating  $\tilde{\mathbf{q}}$  in two steps. The first  
273 consists in solving a linear system for a dummy vector, which is subsequently used to compute  $\tilde{\mathbf{q}}$  by a  
274 matrix-vector product as follows:

$$\begin{aligned} \left( \mathbf{G} \mathbf{H} \mathbf{W}_q^{-1} \mathbf{H}^\top \mathbf{G}^\top + \mu \mathbf{W}_d^{-1} \right) \mathbf{u} &= \boldsymbol{\delta}_d , \\ \tilde{\boldsymbol{\delta}}_q &= \mathbf{W}_q^{-1} \mathbf{H}^\top \mathbf{G}^\top \mathbf{u} \end{aligned} \quad (23)$$

275 where  $\mathbf{u}$  is a dummy vector. After obtaining  $\tilde{\boldsymbol{\delta}}_q$  (equations 22 and 23), the estimate  $\tilde{\mathbf{q}}$  is computed with  
276 equation 18.

## 277 3.2 Formulation without reparameterization

278 Note that, for the particular case in which  $\mathbf{H} = \mathbf{I}_P$  (equation 9), where  $\mathbf{I}_P$  is the identity of order  $P$ ,  
279  $P = Q$ ,  $\mathbf{p} = \mathbf{q}$ ,  $\bar{\mathbf{p}} = \bar{\mathbf{q}}$  (equation 14) and  $\tilde{\mathbf{p}} = \tilde{\mathbf{q}}$  (equation 15). In this case, the linear system (equations 22  
280 and 23) is directly solved for

$$\tilde{\boldsymbol{\delta}}_p = \tilde{\mathbf{p}} - \bar{\mathbf{p}} , \quad (24)$$

281 instead of  $\tilde{\boldsymbol{\delta}}_q$  (equation 18).

## 282 3.3 Linear system solvers

283 According to their properties, the linear systems associated with over and underdetermined problems  
284 (equations 22 and 23) can be solved by using *direct methods* such as LU, Cholesky or QR factorization, for  
285 example (Golub and Van Loan, 2013, sections 3.2, 4.2 and 5.2). These methods involve factorizing the  
286 linear system matrix in a product of “simple” matrices (i.e., triangular, diagonal or orthogonal). Here, we  
287 consider the *Cholesky factorization*, (Golub and Van Loan, 2013, p. 163).

288 Let us consider a real linear system  $\mathbf{M} \mathbf{x} = \mathbf{y}$ , where  $\mathbf{M}$  is a symmetric and positive definite matrix  
289 (Golub and Van Loan, 2013, p. 159). In this case, the Cholesky factorization consists in computing

$$\mathbf{M} = \mathcal{G} \mathcal{G}^\top , \quad (25)$$

290 where  $\mathcal{G}$  is a lower triangular matrix called *Cholesky factor* and having positive diagonal entries. Given  $\mathcal{G}$ ,  
291 the original linear system is replaced by two triangular systems, as follows:

$$\begin{aligned} \mathcal{G} \mathbf{s} &= \mathbf{y} \\ \mathcal{G}^\top \mathbf{x} &= \mathbf{s} \end{aligned} \quad (26)$$

292 where  $\mathbf{s}$  is a dummy vector. For the overdetermined problem (equation 22),  $\mathbf{M} =$   
293  $(\mathbf{H}^\top \mathbf{G}^\top \mathbf{W}_d \mathbf{G} \mathbf{H} + \mu \mathbf{W}_q)$ ,  $\mathbf{x} = \tilde{\boldsymbol{\delta}}_q$  and  $\mathbf{y} = (\mathbf{H}^\top \mathbf{G}^\top \mathbf{W}_d \boldsymbol{\delta}_d)$ . For the underdetermined problem  
294 (equation 23),  $\mathbf{M} = (\mathbf{G} \mathbf{H} \mathbf{W}_q^{-1} \mathbf{H}^\top \mathbf{G}^\top + \mu \mathbf{W}_d^{-1})$ ,  $\mathbf{x} = \mathbf{u}$  and  $\mathbf{y} = \boldsymbol{\delta}_d$ .

295 The use of direct methods for solving large linear systems may be problematic due to computer (i) storage  
 296 of large matrices and (ii) time to perform matrix operations. This problem may be specially complicated in  
 297 equivalent-layer technique for the cases in which the sensitivity matrix  $\mathbf{G}$  does not have a well-defined  
 298 structure (sec. 2.2)

299 These problems can be overcome by solving the linear system using an iterative method. These methods  
 300 produce a sequence of vectors that typically converge to the solution at a reasonable rate. The main  
 301 computational cost associated with these methods is usually some matrix-vector products per iteration. The  
 302 *conjugate gradient* (CG) is a very popular iterative method for solving linear systems in equivalent-layer  
 303 methods. This method was originally developed to solve systems having a square and positive definite  
 304 matrix. There are two adapted versions of the CG method. The first is called *conjugate gradient normal*  
 305 *equation residual* (CNR) Golub and Van Loan (2013, sec. 11.3) or *conjugate gradient least squares*  
 306 (CGLS) (Aster et al., 2019, p. 165) and is used to solve overdetermined problems (equation 22). The second  
 307 is called *conjugate gradient normal equation error* (CGNE) method Golub and Van Loan (2013, sec. 11.3)  
 308 and is used to solve the underdetermined problems (equation 23). Algorithm 1 outlines the CGLS method  
 309 applied to the overdetermined problem (equation 22).

## 4 FLOATING-POINT OPERATIONS

310 Two important factors affecting the efficiency of a given matrix algorithm are the storage and amount of  
 311 required arithmetic. Here, we quantify this last factor associated with different computational strategies to  
 312 solve the linear system of the equivalent-layer technique (section 7). To do it, we opted by counting *flops*,  
 313 which are floating point additions, subtractions, multiplications or divisions (Golub and Van Loan, 2013,  
 314 p. 12–14). This is a non-hardware dependent approach that allows us to do direct comparison between  
 315 different equivalent-layer methods. Most of the flops count used here can be found in Golub and Van Loan  
 316 (2013, p. 12, 106, 107 and 164).

317 Let us consider the case in which the overdetermined problem (equation 22) is solved by Cholesky  
 318 factorization (equations 25 and 26) directly for the parameter vector  $\tilde{\mathbf{p}}$  by considering the particular case in  
 319 which  $\mathbf{H} = \mathbf{I}_P$  (equation 9 and subsection 3.2),  $\mu = 0$  (equation 11),  $\mathbf{W}_d = \mathbf{I}_D$  (equation 12) and  $\bar{\mathbf{p}} = \mathbf{0}$   
 320 (equation 14), where  $\mathbf{I}_P$  and  $\mathbf{I}_D$  are the identities of order  $P$  and  $D$ , respectively. Based on the information  
 321 provided in table 1, the total number of flops can be determined by aggregating the flops required for  
 322 various computations. These computations include the matrix-matrix and matrix-vector products  $\mathbf{G}^\top \mathbf{G}$   
 323 and  $\mathbf{G}^\top \mathbf{d}$ , the Cholesky factor  $\mathcal{G}$ , and the solution of triangular systems. Thus, we can express the total  
 324 number of flops as follows:

$$f_{\text{Cholesky}} = 1/3D^3 + 2D^2 + 2(P^2 + P)D. \quad (27)$$

325 The same particular overdetermined problem can be solved by using the CGLS method (Algorithm 1).  
 326 In this case, we use table 1 again to combine the total number of flops associated with the matrix-vector  
 327 and inner products defined in line 3, before starting the iteration, and the 3 saxpys, 2 inner products and 2  
 328 matrix-vector products per iteration (lines 7 – 12). By considering a maximum number of iterations ITMAX,  
 329 we obtain

$$f_{\text{CGLS}} = 2PD + \text{ITMAX}(4PD + 4D). \quad (28)$$

330 The same approach used to deduce equations 27 and 28 is applied to compute the total number of flops for  
 331 the selected equivalent-layer methods discussed in section 7.

To simplify our analysis, we do not consider the number of flops required to compute the sensitivity matrix  $\mathbf{G}$  (equation 3) or the matrix  $\mathbf{A}$  associated with a given potential-field transformation (equation 4) because they depend on the specific harmonic functions  $g_{ij}$  and  $a_{ij}$  (equations 2 and 5). We also neglect the required flops to compute  $\mathbf{H}$ ,  $\mathbf{W}_d$ ,  $\mathbf{W}_q$  (equations 9, 12 and 13),  $\bar{p}$  (equation 14), retrieve  $\tilde{\mathbf{q}}$  from  $\tilde{\delta}_q$  (equation 18) and computing  $\delta_d$  (equation 19).

## 5 NUMERICAL STABILITY

All equivalent-layer methods aim at obtaining an estimate  $\tilde{\mathbf{p}}$  for the parameter vector  $\mathbf{p}$  (equation 3), which contains the physical property of the equivalent sources. Some methods do it by first obtaining an estimate  $\tilde{\mathbf{q}}$  for the reparameterized parameter vector  $\mathbf{q}$  (equation 9) and then using it to obtain  $\tilde{\mathbf{p}}$  (equation 15). The stability of a solution  $\tilde{\mathbf{p}}$  against noise in the observed data is rarely addressed. Here, we follow the numerical stability analysis presented in Siqueira et al. (2017).

For a given equivalent-layer method (section 7), we obtain an estimate  $\tilde{\mathbf{p}}$  assuming noise-free potential-field data  $\mathbf{d}$ . Then, we create  $L$  different noise-corrupted data  $\mathbf{d}^\ell$ ,  $\ell \in \{1 : L\}$ , by adding  $L$  different sequences of pseudorandom Gaussian noise to  $\mathbf{d}$ , all of them having zero mean. From each  $\mathbf{d}^\ell$ , we obtain an estimate  $\tilde{\mathbf{p}}^\ell$ . Regardless of the particular equivalent-layer method used, the following inequality (Aster et al., 2019, p. 66) holds true:

$$\Delta p^\ell \leq \kappa \Delta d^\ell, \quad \ell \in \{1 : L\}, \quad (29)$$

where  $\kappa$  is the constant of proportionality between the model perturbation

$$\Delta p^\ell = \frac{\|\tilde{\mathbf{p}}^\ell - \tilde{\mathbf{p}}\|}{\|\tilde{\mathbf{p}}\|}, \quad \ell \in \{1 : L\}, \quad (30)$$

and the data perturbation

$$\Delta d^\ell = \frac{\|\mathbf{d}^\ell - \mathbf{d}\|}{\|\mathbf{d}\|}, \quad \ell \in \{1 : L\}, \quad (31)$$

with  $\|\cdot\|$  representing the Euclidean norm. The constant  $\kappa$  acts as the condition number associated with the pseudo-inverse in a given linear inversion. The larger (smaller) the value of  $\kappa$ , the more unstable (stable) is the estimated solution. Because of that, we designate  $\kappa$  as *stability parameter*. Equation 29 shows a linear relationship between the model perturbation  $\Delta p^\ell$  and the data perturbation  $\Delta d^\ell$  (equations 30 and 31). We estimate the  $\kappa$  (equation 29) associated with a given equivalent-layer method as the slope of the straight line fitted to the *numerical stability curve* formed by the  $L$  points  $(\Delta p^\ell, \Delta d^\ell)$ .

## 6 NOTATION FOR SUBVECTORS AND SUBMATRICES

Here, we use a notation inspired on that presented by Van Loan (1992, p. 4) to represent subvectors and submatrices. Subvectors of  $\mathbf{d}$ , for example, are specified by  $\mathbf{d}[\mathbf{i}]$ , where  $\mathbf{i}$  is a list of integer numbers that “pick out” the elements of  $\mathbf{d}$  forming the subvector  $\mathbf{d}[\mathbf{i}]$ . For example,  $\mathbf{i} = (1, 6, 4, 6)$  gives the subvector  $\mathbf{d}[\mathbf{i}] = [d_1 \ d_6 \ d_4 \ d_6]^\top$ . Note that the list  $\mathbf{i}$  of indices may be sorted or not and it may also have repeated indices. For the particular case in which the list has a single element  $\mathbf{i} = (i)$ , then it can be used to extract the  $i$ -th element  $d_i \equiv \mathbf{d}[i]$  of  $\mathbf{d}$ . Sequential lists can be represented by using the colon notation. We consider two types of sequential lists. The first has starting index is smaller than the final index and increment of 1.

The second has starting index is greater than the final index and increment of  $-1$ . For example,

$$\begin{aligned}\mathbf{i} = (3 : 8) &\Leftrightarrow \mathbf{d}[\mathbf{i}] = [d_3 \ d_4 \ \dots \ d_8]^\top \\ \mathbf{i} = (8 : 3) &\Leftrightarrow \mathbf{d}[\mathbf{i}] = [d_8 \ d_7 \ \dots \ d_3]^\top \\ \mathbf{i} = (: 8) &\Leftrightarrow \mathbf{d}[\mathbf{i}] = [d_1 \ d_2 \ \dots \ d_8]^\top \\ \mathbf{i} = (3 :) &\Leftrightarrow \mathbf{d}[\mathbf{i}] = [d_3 \ d_4 \ \dots \ d_D]^\top\end{aligned}$$

355 where  $D$  is the number of elements forming  $\mathbf{d}$ .

The notation above can also be used to define submatrices of a  $D \times P$  matrix  $\mathbf{G}$ . For example,  $\mathbf{i} = (2, 7, 4, 6)$  and  $\mathbf{j} = (1, 3, 8)$  lead to the submatrix

$$\mathbf{G}[\mathbf{i}, \mathbf{j}] = \begin{bmatrix} g_{21} & g_{23} & g_{28} \\ g_{71} & g_{73} & g_{78} \\ g_{41} & g_{43} & g_{48} \\ g_{61} & g_{63} & g_{68} \end{bmatrix}.$$

Note that, in this case, the lists  $\mathbf{i}$  and  $\mathbf{j}$  “pick out”, respectively, the rows and columns of  $\mathbf{G}$  that form the submatrix  $\mathbf{G}[\mathbf{i}, \mathbf{j}]$ . The  $i$ -th row of  $\mathbf{G}$  is given by the  $1 \times P$  vector  $\mathbf{G}[i, :]$ . Similarly, the  $D \times 1$  vector  $\mathbf{G}[:, j]$  represents the  $j$ -th column. Finally, we may use the colon notation to define the following submatrix:

$$\mathbf{i} = (2 : 5), \mathbf{j} = (3 : 7) \Leftrightarrow \mathbf{G}[\mathbf{i}, \mathbf{j}] = \begin{bmatrix} g_{23} & g_{24} & g_{25} & g_{26} & g_{27} \\ g_{33} & g_{34} & g_{35} & g_{36} & g_{37} \\ g_{43} & g_{44} & g_{45} & g_{46} & g_{47} \\ g_{53} & g_{54} & g_{55} & g_{56} & g_{57} \end{bmatrix},$$

356 which contains the contiguous elements of  $\mathbf{G}$  from rows 2 to 5 and from columns 3 to 7.

## 7 COMPUTATIONAL STRATEGIES

357 The linear inverse problem of the equivalent-layer technique (section 3) for the case in which there are  
358 large volumes of potential-field data requires dealing with:

- 359 (i) the large computer memory to store large and full matrices;
- 360 (ii) the long computation time to multiply a matrix by a vector; and
- 361 (iii) the long computation time to solve a large linear system of equations.

362 Here, we review some strategies aiming at reducing the computational cost of the equivalent-layer technique.  
363 We quantify the computational cost by using flops (section 4) and compare the results with those obtained  
364 for Cholesky factorization and CGLS (equations 27 and 28). We focus on the overall strategies used by the  
365 selected methods.

### 366 7.1 Moving window

367 The initial approach to enhance the computational efficiency of the equivalent-layer technique is  
368 commonly denoted *moving window* and involves first splitting the observed data  $d_i, i \in \{1 : D\}$ , into  
369  $M$  overlapping subsets (or data windows) formed by  $D^m$  data each,  $m \in \{1 : M\}$ . The data inside the

370  $m$ -th window are usually adjacent to each other and have indices defined by an integer list  $\mathbf{i}^m$  having  
 371  $D^m$  elements. The number of data  $D^m$  forming the data windows are not necessarily equal to each other.  
 372 Each data window has a  $D^m \times 1$  observed data vector  $\mathbf{d}^m \equiv \mathbf{d}[\mathbf{i}^m]$ . The second step consists in defining  
 373 a set of  $P$  equivalent sources with scalar physical property  $p_j$ ,  $j \in \{1 : P\}$ , and also split them into  $M$   
 374 overlapping subsets (or source windows) formed by  $P^m$  data each,  $m \in \{1 : M\}$ . The sources inside the  
 375  $m$ -th window have indices defined by an integer list  $\mathbf{j}^m$  having  $P^m$  elements. Each source window has a  
 376  $P^m \times 1$  parameter vector  $\mathbf{p}^m$  and is located right below the corresponding  $m$ -th data window. Then, each  
 377  $\mathbf{d}^m \equiv \mathbf{d}[\mathbf{i}^m]$  is approximated by

$$\mathbf{f}^m = \mathbf{G}^m \mathbf{p}^m, \quad (32)$$

378 where  $\mathbf{G}^m \equiv \mathbf{G}[\mathbf{i}^m, \mathbf{j}^m]$  is a submatrix of  $\mathbf{G}$  (equation 3) formed by the elements computed with equation  
 379 2 using only the data and equivalent sources located inside the window  $m$ -th. The main idea of the moving-  
 380 window approach is using the  $\tilde{\mathbf{p}}^m$  estimated for each window to obtain (i) an estimate  $\tilde{\mathbf{p}}$  of the parameter  
 381 vector for the entire equivalent layer or (ii) a given potential-field transformation  $\mathbf{t}$  (equation 4). The main  
 382 advantages of this approach is that (i) the estimated parameter vector  $\tilde{\mathbf{p}}$  or transformed potential field are  
 383 not obtained by solving the full, but smaller linear systems and (ii) the full matrix  $\mathbf{G}$  (equation 3) is never  
 384 stored.

385 Leão and Silva (1989) presented a pioneer work using the moving-window approach. Their method  
 386 requires a regularly-spaced grid of observed data on a horizontal plane  $z_0$ . The data windows are defined by  
 387 square local grids of  $\sqrt{D'} \times \sqrt{D'}$  adjacent points, all of them having the same number of points  $D'$ . The  
 388 equivalent sources in the  $m$ -th data window are located below the observation plane, at a constant vertical  
 389 distance  $\Delta z_0$ . They are arranged on a regular grid of  $\sqrt{P'} \times \sqrt{P'}$  adjacent points following the same  
 390 grid pattern of the observed data. The local grid of sources for all data windows have the same number  
 391 of elements  $P'$ . Besides, they are vertically aligned, but expands the limits of their corresponding data  
 392 windows, so that  $D' < P'$ . Because of this spatial configuration of observed data and equivalent sources,  
 393 we have that  $\mathbf{G}^m = \mathbf{G}'$  (equation 32) for all data windows (i.e.,  $\forall m \in \{1 : M\}$ ), where  $\mathbf{G}'$  is a  $D' \times P'$   
 394 constant matrix.

395 By omitting the normalization strategy used by Leão and Silva (1989), their method consists in directly  
 396 computing the transformed potential field  $t_c^m$  at the central point  $(x_c^m, y_c^m, z_0 + \Delta z_0)$  of each data window  
 397 as follows:

$$t_c^m = (\mathbf{a}')^\top \mathbf{B}' \mathbf{d}^m, \quad m \in \{1 : M\}, \quad (33)$$

398 where  $\mathbf{a}'$  is a  $P' \times 1$  vector with elements computed by equation 5 by using all equivalent sources in the  
 399  $m$ -th window and only the coordinate of the central point in the  $m$ -th data window and

$$\mathbf{B}' = (\mathbf{G}')^\top \left[ \mathbf{G}' (\mathbf{G}')^\top + \mu \mathbf{I}_{D'} \right]^{-1} \quad (34)$$

400 is a particular case of matrix  $\mathbf{B}$  associated with underdetermined problems (equation 21) for the particular  
 401 case in which  $\mathbf{H} = \mathbf{W}_q = \mathbf{I}_{P'}$  (equations 9 and 13),  $\mathbf{W}_d = \mathbf{I}_{D'}$  (equation 12),  $\bar{\mathbf{p}} = \mathbf{0}$  (equation 14), where  
 402  $\mathbf{I}_{P'}$  and  $\mathbf{I}_{D'}$  are identity matrices of order  $P'$  and  $D'$ , respectively, and  $\mathbf{0}$  is a vector of zeros. Due to the  
 403 presumed spatial configuration of the observed data and equivalent sources,  $\mathbf{a}'$  and  $\mathbf{G}'$  are the same for all  
 404 data windows. Hence, only the data vector  $\mathbf{d}^m$  is modified according to the position of the data window.  
 405 Note that equation 33 combines the potential-field transformation (equation 4) with the solution of the  
 406 undetermined problem (equation 23).

407 The method proposed by Leão and Silva (1989) can be outlined by the Algorithm 2. Note that Leão and  
 408 Silva (1989) directly compute the transformed potential  $t_c^m$  at the central point of each data window without  
 409 explicitly computing and storing an estimated for  $\mathbf{p}^m$  (equation 32). It means that their method allows  
 410 computing a single potential-field transformation. A different transformation or the same one evaluated at  
 411 different points require running their moving-data window method again.

412 The total number of flops in Algorithm 2 depends on computing the  $P' \times D'$  matrix  $\mathbf{B}'$  (equation 34) in  
 413 line 6 and use it to define the  $1 \times P'$  vector  $(\mathbf{a}')^\top \mathbf{B}'$  (line 7) before starting the iterations and computing  
 414 an inner product (equation 33) per iteration. We consider that the total number of flops associated with  $\mathbf{B}'$   
 415 is obtained by the matrix-matrix product  $\mathbf{G}' (\mathbf{G}')^\top$ , its inverse and then the premultiplication by  $(\mathbf{G}')^\top$ . By  
 416 using table 1 and considering that inverse is computed via Cholesky factorization, we obtain that the total  
 417 number of flops for lines 6 and 7 is  $2(D')^2 P' + 7(D')^3/6 + 2(D')^2 P'$ . Then, the total number of flops for  
 418 Algorithm 2 is

$$f_{\text{LS89}} = 7/6(D')^3 + 4P'(D')^2 + M 2P'. \quad (35)$$

419 Soler and Uieda (2021) generalized the method proposed by Leão and Silva (1989) for irregularly spaced  
 420 data on an undulating surface. A direct consequence of this generalization is that a different submatrix  
 421  $\mathbf{G}^m \equiv \mathbf{G}[\mathbf{i}^m, \mathbf{j}^m]$  (equation 32) must be computed for each window. Differently from Leão and Silva  
 422 (1989), Soler and Uieda (2021) store the computed  $\tilde{\mathbf{p}}^m$  for all windows and subsequently use them to obtain  
 423 a desired potential-field transformation (equation 4) as the superposed effect of all windows. The estimated  
 424  $\tilde{\mathbf{p}}^m$  for all windows are combined to form a single  $P \times 1$  vector  $\tilde{\mathbf{p}}$ , which is an estimate for original  
 425 parameter vector  $\mathbf{p}$  (equation 3). For each data window, Soler and Uieda (2021) solve an overdetermined  
 426 problem (equation 22) for  $\tilde{\mathbf{p}}^m$  by using  $\mathbf{H} = \mathbf{W}_q = \mathbf{I}_{P^m}$  (equations 9 and 13),  $\mathbf{W}_d^m$  (equation 12) equal to  
 427 a diagonal matrix of weights for the data inside the  $m$ -th window and  $\bar{\mathbf{p}} = \mathbf{0}$  (equation 14), so that

$$[(\mathbf{G}^m)^\top \mathbf{W}_d^m \mathbf{G}^m + \mu \mathbf{I}_{P'}] \tilde{\mathbf{p}}^m = (\mathbf{G}^m)^\top \mathbf{W}_d^m \mathbf{d}^m. \quad (36)$$

428 Unlike Leão and Silva (1989), Soler and Uieda (2021) do not adopt a sequential order of the data windows;  
 429 rather, they adopt a randomized order of windows in their iterations. The overall steps of the method  
 430 proposed by Soler and Uieda (2021) are defined by the Algorithm 3. For convenience, we have omitted the  
 431 details about the randomized window order, normalization strategy employed and block-averaged sources  
 432 layout proposed by those authors (see subsection 2.1). Note that this algorithm starts with a residuals vector  
 433  $\mathbf{r}$  that is iteratively updated. The iterative algorithm in Soler and Uieda (2021) estimates a solution ( $\tilde{\mathbf{p}}^m$  in  
 434 equation 36) using the data and the equivalent sources that fall within a moving-data window; however, it  
 435 calculates the predicted data and the residual data in the whole survey data. Next, the residual data that fall  
 436 within a new position of the data window is used as input data to estimate a new solution within the data  
 437 window which, in turn, is used to calculate a new predicted data and a new residual data in the whole  
 438 survey data.

439 The computational cost of Algorithm 3 can be defined in terms of the linear system (equation 36) to be  
 440 solved for each window (line 10) and the subsequent updates in lines 11 and 12. We consider that the linear  
 441 system cost can be quantified by the matrix-matrix and matrix-vector products  $(\mathbf{G}^m)^\top \mathbf{G}^m$  and  $(\mathbf{G}^m)^\top \mathbf{d}^m$ ,  
 442 respectively, and solution of the linear system (line 10) via Cholesky factorization (equations 25 and 26).  
 443 The following updates represent a saxpy without scalar-vector product (line 11) and a matrix-vector product  
 444 (line 12). In this case, according to table 1, the total number of flops associated with Algorithm 3 is given

445 by:

$$f_{\text{SU21}} = M \left[ \frac{1}{3}(P')^3 + 2(D')(P')^2 + (4D')P' \right], \quad (37)$$

446 where  $P'$  and  $D'$  represent, respectively, the average number of equivalent sources and data at each window.

## 447 7.2 Column-action update

448 Cordell (1992) proposed a *column-action update* strategy similar to those applied to image reconstruction  
 449 methods (e.g., Elfving et al., 2017). His approach, that was later used by Guspí and Novara (2009), relies  
 450 on first defining one equivalent source located right below each observed data  $d_i$ ,  $i \in \{1 : D\}$ , at a vertical  
 451 coordinate  $z_i + \Delta z_i$ , where  $\Delta z_i$  is proportional to the distance from the  $i$ -th observation point  $(x_i, y_i, z_i)$  to  
 452 its closest neighbor. The second step consists in updating the physical property  $p_j$  of a single equivalent  
 453 source,  $j \in \{1 : D\}$  and remove its predicted potential field from the observed data vector  $\mathbf{d}$ , producing a  
 454 residuals vector  $\mathbf{r}$ . At each iteration, the single equivalent source is the one located vertically beneath the  
 455 observation station of the maximum data residual. Next, the predicted data produced by this single source  
 456 is calculated over all of the observation points and a new data residual  $\mathbf{r}$  and the  $D \times 1$  parameter vector  $\mathbf{p}$   
 457 containing the physical property of all equivalent sources are updated iteratively. During each subsequent  
 458 iteration, Cordell's method either incorporates a single equivalent source or adjusts an existing equivalent  
 459 source to match the maximum amplitude of the current residual field. The convergence occurs when all of  
 460 the residuals are bounded by an envelope of prespecified expected error. At the end, the algorithm produces  
 461 an estimate  $\tilde{\mathbf{p}}$  for the parameter vector yielding a predicted potential field  $\mathbf{f}$  (equation 3) satisfactorily  
 462 fitting the observed data  $\mathbf{d}$  according to a given criterion. Note that the method proposed by Cordell (1992)  
 463 iteratively solves the linear  $\mathbf{G}\tilde{\mathbf{p}} \approx \mathbf{d}$  with a  $D \times D$  matrix  $\mathbf{G}$ . At each iteration, only a single column of  $\mathbf{G}$   
 464 (equation 3) is used. An advantage of this *column-action update approach* is that the full matrix  $\mathbf{G}$  is never  
 465 stored.

466 Algorithm 4 delineates the Cordell's method. We have introduced a scale factor  $\sigma$  to improve convergence.  
 467 Note that a single column  $\mathbf{G}[:, i_{\max}]$  of the  $D \times D$  matrix  $\mathbf{G}$  (equation 3) is used per iteration, where  $i_{\max}$   
 468 is the index of the maximum absolute value in  $\mathbf{r}$ . As pointed out by Cordell (1992), the method does not  
 469 necessarily decrease monotonically along the iterations. Besides, the method may not converge depending  
 470 on how the vertical distances  $\Delta z_i$ ,  $i \in \{1 : D\}$ , controlling the depths of the equivalent sources are set.  
 471 According to Cordell (1992), the maximum absolute value  $r_{\max}$  in  $\mathbf{r}$  decreases robustly at the beginning  
 472 and oscillates within a narrowing envelope for the subsequent iterations.

473 Guspí and Novara (2009) generalized Cordell's method to perform reduction to the pole and other  
 474 transformations on scattered magnetic observations by using two steps. The first step involves computing  
 475 the vertical component of the observed field using equivalent sources while preserving the magnetization  
 476 direction. In the second step, the vertical observation direction is maintained, but the magnetization  
 477 direction is shifted to the vertical. The main idea employed by both Cordell (1992) and Guspí and Novara  
 478 (2009) is an iterative scheme that uses a single equivalent source positioned below a measurement station  
 479 to compute both the predicted data and residual data for all stations. This approach entails a computational  
 480 strategy where a single column of the sensitivity matrix  $\mathbf{G}$  (equation 3) is calculated per iteration.

481 The total number of flops in Algorithm 4 consists in computing the scale factor  $\sigma$  (line 5), computing an  
 482 initial approximation for the parameter vector and the residuals (lines 6 and 7) and finding the maximum  
 483 absolute value in vector  $\mathbf{r}$  (line 8) before the while loop. Per iteration, there is a saxpy (line 13) and another  
 484 search for the maximum absolute value in vector  $\mathbf{r}$  (line 14). By considering that selecting the maximum  
 485 absolute value in a  $D \times 1$  vector is a  $D \log_2(D)$  operation (e.g., Press et al., 2007, p. 420), we get from

486 table 1 that the total number of flops in Algorithm 38 is given by:

$$f_{C92} = 4D^2 + 6D + D \log_2(D) + \text{ITMAX} [2D + D \log_2(D)] . \quad (38)$$

487 **7.3 Row-action update**

488 To reduce the total processing time and memory usage of equivalent-layer technique, Mendonça and  
 489 Silva (1994) proposed a strategy called *equivalent data concept*. The equivalent data concept is grounded  
 490 on the principle that there is a subset of redundant data that does not contribute to the final solution and  
 491 thus can be dispensed. Conversely, there is a subset of observations, called equivalent data, that contributes  
 492 effectively to the final solution and fits the remaining observations (redundant data). Iteratively, Mendonça  
 493 and Silva (1994) selected the subset of equivalent data that is substantially smaller than the original dataset.  
 494 This selection is carried out by incorporating one data point at a time.

495 The method presented by Mendonça and Silva (1994) is a type of algebraic reconstruction technique  
 496 (ART) (e.g., van der Sluis and van der Vorst, 1987, p. 58) or *row-action update* (e.g., Elfving et al., 2017)  
 497 to estimate a parameter vector  $\tilde{\mathbf{p}}$  for a regular grid of  $P$  equivalent sources on a horizontal plane  $z_0$ . Such  
 498 methods iterate on the linear system rows to estimate corrections for the parameter vector, which may  
 499 substantially save computer time and memory required to compute and store the full linear system matrix  
 500 along the iterations. The convergence of such methods strongly depends on the linear system condition. The  
 501 main advantage of such methods is not computing and storing the full linear system matrix, but iteratively  
 502 using its rows. In contrast to common row-action algorithms, the rows in Mendonça and Silva (1994) are  
 503 not processed sequentially. Instead, in Mendonça and Silva (1994), the rows are introduced according to  
 504 their residual magnitudes (maximum absolute value in  $\mathbf{r}$ ), which are computed based on the estimate over  
 505 the equivalent layer from the previous iteration. The particular row-action method proposed by Mendonça  
 506 and Silva (1994) considers that

$$\mathbf{d} = \begin{bmatrix} \mathbf{d}_e \\ \mathbf{d}_r \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{G}_e \\ \mathbf{R}_r \end{bmatrix}, \quad (39)$$

507 where  $\mathbf{d}_e$  and  $\mathbf{d}_r$  are  $D_e \times 1$  and  $D_r \times 1$  vectors and  $\mathbf{G}_e$  and  $\mathbf{G}_r$  are  $D_e \times P$  and  $D_r \times P$  matrices,  
 508 respectively. Mendonça and Silva (1994) designate  $\mathbf{d}_e$  and  $\mathbf{d}_r$  as, respectively, *equivalent* and *redundant*  
 509 data. With the exception of a normalization strategy, Mendonça and Silva (1994) calculate a  $P \times 1$  estimated  
 510 parameter vector  $\tilde{\mathbf{p}}$  by solving an underdetermined problem (equation 23) involving only the equivalent  
 511 data  $\mathbf{d}_e$  (equation 39) for the particular case in which  $\mathbf{H} = \mathbf{W}_p = \mathbf{I}_P$  (equations 9 and 13),  $\mathbf{W}_d = \mathbf{I}_{D_e}$   
 512 (equation 12) and  $\bar{\mathbf{p}} = \mathbf{0}$  (equation 14), which results in

$$(\mathbf{F} + \mu \mathbf{I}_{D_e}) \mathbf{u} = \mathbf{d}_e \quad , \quad \tilde{\mathbf{p}} = \mathbf{G}_e^\top \mathbf{u} \quad , \quad (40)$$

513 where  $\mathbf{F}$  is a computationally-efficient  $D_e \times D_e$  matrix that approximates  $\mathbf{G}_e \mathbf{G}_e^\top$ . Mendonça and Silva  
 514 (1994) presume that the estimated parameter vector  $\tilde{\mathbf{p}}$  obtained from equation 40 leads to a  $D_r \times 1$  residuals  
 515 vector

$$\mathbf{r} = \mathbf{d}_r - \mathbf{G}_r \tilde{\mathbf{p}} \quad (41)$$

516 having a maximum absolute value  $r_{\max} \leq \epsilon$ , where  $\epsilon$  is a predefined tolerance.

517 The overall method of Mendonça and Silva (1994) is defined by Algorithm 5. It is important noting  
 518 that the number  $D_e$  of equivalent data in  $\mathbf{d}_e$  increases by one per iteration, which means that the order

519 of the linear system in equation 40 also increases by one at each iteration. Those authors also propose  
 520 a computational strategy based on Cholesky factorization (e.g., Golub and Van Loan, 2013, p. 163) for  
 521 efficiently updating  $(\mathbf{F} + \mu \mathbf{I}_{D_e})$  at a given iteration (line 16 in Algorithm 5) by computing only its new  
 522 elements with respect to those computed in the previous iteration.

## 523 7.4 Reparameterization

524 Another approach for improving the computational performance of equivalent-layer technique consists  
 525 in setting a  $P \times Q$  reparameterization matrix  $\mathbf{H}$  (equation 9) with  $Q \ll P$ . This strategy has been used  
 526 in applied geophysics for decades (e.g., Skilling and Bryan, 1984; Kennett et al., 1988; Oldenburg et al.,  
 527 1993; Barbosa et al., 1997) and is known as *subspace method*. The main idea relies in reducing the linear  
 528 system dimension from the original  $P$ -space to a lower-dimensional subspace (the  $Q$ -space). An estimate  
 529  $\tilde{\mathbf{q}}$  for the reparameterized parameter vector  $\mathbf{q}$  is obtained in the  $Q$ -space and subsequently used to obtain  
 530 an estimate  $\tilde{\mathbf{p}}$  for the parameter vector  $\mathbf{p}$  (equation 3) in the  $P$ -space by using equation 9. Hence, the key  
 531 aspect of this *reparameterization approach* is solving an appreciably smaller linear inverse problem for  $\tilde{\mathbf{q}}$   
 532 than that for the original parameter vector  $\tilde{\mathbf{p}}$  (equation 3).

533 Oliveira Jr. et al. (2013) have used this approach to describe the physical property distribution on the  
 534 equivalent layer in terms of piecewise bivariate polynomials. Specifically, their method consists in splitting  
 535 a regular grid of equivalent sources into source windows inside which the physical-property distribution  
 536 is described by bivariate polynomial functions. The key aspect of their method relies on the fact that the  
 537 total number of coefficients required to define the bivariate polynomials is considerably smaller than the  
 538 original number of equivalent sources. Hence, they formulate a linear inverse problem for estimating the  
 539 polynomial coefficients and use them later to compute the physical property distribution on the equivalent  
 540 layer.

541 The method proposed by Oliveira Jr. et al. (2013) consists in solving an overdetermined problem (equation  
 542 22) for estimating the polynomial coefficients  $\tilde{\mathbf{q}}$  with  $\mathbf{W}_d = \mathbf{I}_D$  (equation 12) and  $\bar{\mathbf{q}} = \mathbf{0}$  (equation 14), so  
 543 that

$$( \mathbf{H}^\top \mathbf{G}^\top \mathbf{G} \mathbf{H} + \mu \mathbf{W}_q ) \tilde{\mathbf{q}} = \mathbf{H}^\top \mathbf{G}^\top \mathbf{d}, \quad (42)$$

544 where  $\mathbf{W}_q = \mathbf{H}^\top \mathbf{W}_p \mathbf{H}$  is defined by a matrix  $\mathbf{W}_p$  representing the zeroth- and first-order Tikhonov  
 545 regularization (e.g., Aster et al., 2019, p. 103). Note that, in this case, the prior information is defined in the  
 546  $P$ -space for the original parameter vector  $\mathbf{p}$  and then transformed to the  $Q$ -space. Another characteristic of  
 547 their method is that it is valid for processing irregularly-spaced data on an undulating surface.

548 Mendonça (2020) also proposed a reparameterization approach for the equivalent-layer technique. Their  
 549 approach, however, consists in setting  $\mathbf{H}$  as a truncated singular value decomposition (SVD) (e.g., Aster  
 550 et al., 2019, p. 55) of the observed potential field. Differently from Oliveira Jr. et al. (2013), however, the  
 551 method of Mendonça (2020) requires a regular grid of potential-field data on horizontal plane. Another  
 552 difference is that these authors uses  $\mathbf{W}_q = \mathbf{I}_Q$  (equation 13), which means that the regularization is defined  
 553 directly in the  $Q$ -space.

554 We consider an algorithm (not shown) that solves the overdetermined problem (equation 22) by combining  
 555 the reparameterization with CGLS method (Algorithm 1). It starts with a reparameterization step defined  
 556 by defining a matrix  $\mathbf{C} = \mathbf{G} \mathbf{H}$  (equation 10). Then, the CGLS (Algorithm 1) is applied by replacing  $\mathbf{G}$   
 557 with  $\mathbf{C}$ . In this case, the linear system is solved by the reparameterized parameter vector  $\tilde{\mathbf{q}}$  instead of  $\tilde{\mathbf{p}}$ .  
 558 At the end, the estimated  $\tilde{\mathbf{q}}$  is transformed into  $\tilde{\mathbf{p}}$  (equation 15). Compared to the original CGLS shown  
 559 in Algorithm 1, the algorithm discussed here has the additional flops associated with the matrix-matrix

product to compute  $\mathbf{C}$  and the matrix-vector product of equation 15 outside the while loop. Then, according to table 1, the total number of flops given by:

$$f_{\text{reparam.}} = 2Q(DP + D) + 2PQ + \text{ITMAX} (4QD + 4D) . \quad (43)$$

The important aspect of this approach is that, for the case in which  $Q \ll P$  (equation 9), the number of flops per iteration can be substantially decreased with respect to those associated with Algorithm 1. In this case, the flops decrease per iteration compensates the additional flops required to compute  $\mathbf{C}$  and obtain  $\tilde{\mathbf{p}}$  from  $\tilde{\mathbf{q}}$  (equation 15).

## 7.5 Sparsity induction

Li and Oldenburg (2010) proposed a method that applies the discrete wavelet transform to introduce sparsity into the original dense matrix  $\mathbf{G}$  (equation 3). Those authors approximate a planar grid of potential-field data by a regularly-spaced grid of equivalent sources, so that the number of data  $D$  and sources  $P$  is the same, i.e.,  $D = P$ . Specifically, Li and Oldenburg (2010) proposed a method that applies the wavelet transform to the original dense matrix  $\mathbf{G}$  and sets to zero the small coefficients that are below a given threshold, which results in an approximating sparse representation of  $\mathbf{G}$  in the wavelet domain. They first consider the following approximation

$$\mathbf{d}_w \approx \mathbf{G}_s \mathbf{p}_w , \quad (44)$$

where

$$\mathbf{d}_w = \mathcal{W} \mathbf{d} , \quad \mathbf{p}_w = \mathcal{W} \mathbf{p} , \quad (45)$$

are the observed data and parameter vector in the wavelet domain;  $\mathcal{W}$  is a  $D \times D$  orthogonal matrix defining a discrete wavelet transform; and  $\mathbf{G}_s$  is a sparse matrix obtained by setting to zero the elements of

$$\mathbf{G}_w = \mathcal{W} \mathbf{G} \mathcal{W}^\top \quad (46)$$

with absolute value smaller than a given threshold.

Li and Oldenburg (2010) solve a normalized inverse problem in the wavelet domain. Specifically, they first define a matrix

$$\mathbf{G}_L = \mathbf{G}_s \mathbf{L}^{-1} \quad (47)$$

and a normalized parameter vector

$$\mathbf{p}_L = \mathbf{L} \mathbf{p}_w , \quad (48)$$

where  $\mathbf{L}$  is a diagonal and invertible matrix representing an approximation of the first-order Tikhonov regularization in the wavelet domain. Then they solve an overdetermined problem (equation 22) to obtain an estimate  $\tilde{\mathbf{p}}_L$  for  $\mathbf{p}_L$  (equation 48), with  $\mathbf{G}_L$  (equation 47),  $\mathbf{H} = \mathbf{I}_P$  (equations 9),  $\mu = 0$  (equation 11),  $\mathbf{W}_d = \mathbf{I}_D$  (equation 12) and  $\bar{p} = 0$  (equation 14) via conjugate-gradient method (e.g., Golub and Van Loan, 2013, sec. 11.3). Finally, Li and Oldenburg (2010) compute an estimate  $\tilde{\mathbf{p}}$  for the original parameter vector given by

$$\tilde{\mathbf{p}} = \mathcal{W}^\top (\mathbf{L}^{-1} \tilde{\mathbf{p}}_L) , \quad (49)$$

where the term within parenthesis is an estimate  $\tilde{\mathbf{p}}_w$  of the parameter vector  $\mathbf{p}_w$  (equation 45) in the wavelet domain and matrix  $\mathcal{W}^\top$  represents an inverse wavelet transform.

589 Barnes and Lumley (2011) also proposed a computationally efficient method for equivalent-layer  
 590 technique by inducing sparsity into the original sensitivity matrix  $\mathbf{G}$  (equation 3). Their approach consists  
 591 in setting a  $P \times Q$  reparameterization matrix  $\mathbf{H}$  (equation 9) with  $Q \approx 1.7 P$ . Note that, differently from  
 592 Oliveira Jr. et al. (2013) and Mendonça (2020), Barnes and Lumley (2011) do not use the reparameterization  
 593 with the purpose of reducing the number of the parameters. Instead, they use a reparameterization scheme  
 594 that groups distant equivalent sources into blocks by using a bisection process. This scheme leads to  
 595 a quadtree representation of the physical-property distribution on the equivalent layer, so that matrix  
 596  $\mathbf{GH}$  (equation 10) is notably sparse. Barnes and Lumley (2011) explore this sparsity in solving the  
 597 overdetermined problem for  $\tilde{\mathbf{q}}$  (equation 42) via conjugate-gradient method (e.g., Golub and Van Loan,  
 598 2013, sec. 11.3).

599 It is difficult to predict the exact sparsity obtained from the methods proposed by Li and Oldenburg (2010)  
 600 and Barnes and Lumley (2011) because it depends on several factors, including the observed potential-field  
 601 data. According to Li and Oldenburg (2010), their wavelet approach results in a sparse matrix having  $\approx 2\%$   
 602 of the elements in  $\mathbf{G}_w$  (equation 46). The reparameterization proposed by Barnes and Lumley (2011) leads  
 603 to a sparse matrix  $\mathbf{GH}$  (equation 10) with only  $\approx 1\%$  of non-zero elements. These sparsity patterns can be  
 604 efficiently explored, for example, in computing the required matrix-vector products along the iterations of  
 605 the CGLS method (Algorithm 1).

## 606 7.6 Iterative methods using the full matrix $\mathbf{G}$

607 Xia and Sprowl (1991) introduced an iterative method for estimating the parameter vector  $\tilde{\mathbf{p}}$  (equation 3),  
 608 which was subsequently adapted to the Fourier domain by Xia et al. (1993). Their method uses the full  
 609 and dense sensitivity matrix  $\mathbf{G}$  (equation 3) (without applying any compression or reparameterization, for  
 610 example) to compute the predicted data at all observation points per iteration. More than two decades later,  
 611 Siqueira et al. (2017) have proposed an iterative method similar to that presented by Xia and Sprowl (1991).  
 612 The difference is that Siqueira et al.'s algorithm was deduced from the *Gauss' theorem* (e.g., Kellogg, 1967,  
 613 p. 43) and the *total excess of mass* (e.g., Blakely, 1996, p. 60). Besides, Siqueira et al. (2017) have included  
 614 a numerical analysis showing that their method produces very stable solutions, even for noise-corrupted  
 615 potential-field data.

616 The iterative method proposed by Siqueira et al. (2017) is outlined in Algorithm 6, presumes an equivalent  
 617 layer formed by monopoles (point masses) and can be applied to irregularly-spaced data on an undulating  
 618 surface. Instead of using the element of area originally proposed by Siqueira et al. (2017), we introduce the  
 619 scale factor  $\sigma$ , which can be automatically computed from the observed potential-field data. Note that the  
 620 residuals  $\mathbf{r}$  are used to compute a correction  $\Delta\mathbf{p}$  for the parameter vector at each iteration (line 11), which  
 621 requires a matrix-vector product involving the full matrix  $\mathbf{G}$ . Interestingly, this approach for estimating  
 622 the physical property distribution on an equivalent layer is the same originally proposed by Bott (1960)  
 623 for estimating the basement relief under sedimentary basins. The methods of Xia and Sprowl (1991) and  
 624 Siqueira et al. (2017) were originally proposed for processing gravity data, but can be potentially applied  
 625 to any harmonic function because they actually represent iterative solutions of the classical *Dirichlet's problem*  
 626 or the *first boundary value problem of potential theory* (Kellogg, 1967, p. 236) on a plane.

627 Recently, Jirigalatu and Ebbing (2019) presented another iterative method for estimating a parameter  
 628 vector  $\tilde{\mathbf{p}}$  (equation 3). With the purpose of combining different potential-field data, their method basically  
 629 modifies that shown in Algorithm 6 by changing the initial approximation and the iterative correction for  
 630 the parameter vector. Specifically, Jirigalatu and Ebbing (2019) replace line 5 by  $\tilde{\mathbf{p}} = \mathbf{0}$ , where  $\mathbf{0}$  is a vector  
 631 of zeros, and line 11 by  $\Delta\mathbf{p} = \omega \mathbf{G}^\top \mathbf{r}$ , where  $\omega$  is a positive scalar defined by trial and error. Note that

632 this modified approach requires two matrix-vector products involving the full matrix  $\mathbf{G}$  per iteration. To  
 633 overcome the high computational cost of these two products, Jirigalatu and Ebbing (2019) set an equivalent  
 634 layer formed by prisms and compute their predicted potential field in the wavenumber domain by using the  
 635 Gauss-FFT technique Zhao et al. (2018).

636 The iterative method proposed by Siqueira et al. (2017) (Algorithm 6) requires computing the scale factor  
 637  $\sigma$  (line 5), computing an initial approximation for the parameter vector and the residuals (lines 6 and 7)  
 638 before the main loop. Inside the main loop, there is a half saxpy (lines 11 and 12) to update the parameter  
 639 vector, a matrix-vector product (line 13) and the residuals update (line 14). Then, we get from table 1 that  
 640 the total number of flops is given by:

$$f_{\text{SOB17}} = 4D^2 + 6D + \text{ITMAX} (2D^2 + 3D) . \quad (50)$$

641 Note that the number of flops per iteration in  $f_{\text{SOB17}}$  (equation 50) has the same order of magnitude, but is  
 642 smaller than that in  $f_{\text{CGLS}}$  (equation 28).

## 643 7.7 Iterative deconvolution

644 Recently, Takahashi et al. (2020, 2022) proposed the *convolutional equivalent-layer method*, which  
 645 explores the structure of the sensitivity matrix  $\mathbf{G}$  (equation 3) for the particular case in which (i) there  
 646 is a single equivalent source right below each potential-field datum and (ii) both data and sources rely  
 647 on planar and regularly spaced grids. Specifically, they consider a regular grid of  $D$  potential-field data  
 648 at points  $(x_i, y_i, z_0)$ ,  $i \in \{1 : D\}$ , on a horizontal plane  $z_0$ . The data indices  $i$  may be ordered along the  
 649  $x$ - or  $y$ -direction, which results in an  $x$ - or  $y$ -oriented grid, respectively. They also consider a single  
 650 equivalent source located right below each datum, at a constant vertical coordinate  $z_0 + \Delta z$ ,  $\Delta z > 0$ . In  
 651 this case, the number of data and equivalent sources are equal to each other (i.e.,  $D = P$ ) and  $\mathbf{G}$  (equation  
 652 3) assumes a *doubly block Toeplitz* (Jain, 1989, p. 28) or *block-Toeplitz-Toeplitz-block* (BTTB) (Chan and  
 653 Jin, 2007, p. 67) structure formed by  $N_B \times N_B$  blocks, where each block has  $N_b \times N_b$  elements, with  
 654  $D = N_B N_b$ . This particular structure allows formulating the product of  $\mathbf{G}$  and an arbitrary vector as a *fast*  
 655 *discrete convolution* via *Fast Fourier Transform* (FFT) (Van Loan, 1992, section 4.2).

656 Consider, for example, the particular case in which  $N_B = 4$ ,  $N_b = 3$  and  $D = 12$ . In this case,  $\mathbf{G}$   
 657 (equation 3) is a  $12 \times 12$  block matrix given by

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}^0 & \mathbf{G}^1 & \mathbf{G}^2 & \mathbf{G}^3 \\ \mathbf{G}^{-1} & \mathbf{G}^0 & \mathbf{G}^1 & \mathbf{G}^2 \\ \mathbf{G}^{-2} & \mathbf{G}^{-1} & \mathbf{G}^0 & \mathbf{G}^1 \\ \mathbf{G}^{-3} & \mathbf{G}^{-2} & \mathbf{G}^{-1} & \mathbf{G}^0 \end{bmatrix}_{D \times D}, \quad (51)$$

658 where each block  $\mathbf{G}^n$ ,  $n \in \{(1 - N_B) : (N_B - 1)\}$ , is a  $3 \times 3$  Toeplitz matrix. Takahashi et al. (2020,  
 659 2022) have deduced the specific relationship between blocks  $\mathbf{G}^n$  and  $\mathbf{G}^{-n}$  and also between a given block  
 660  $\mathbf{G}^n$  and its transposed  $(\mathbf{G}^n)^\top$  according to the harmonic function  $g_{ij}$  (equation 2) defining the element  $ij$   
 661 of the sensitivity matrix  $\mathbf{G}$  (equation 3) and the orientation of the data grid.

662 Consider the matrix-vector products

$$\mathbf{G} \mathbf{v} = \mathbf{w} \quad (52)$$

663 and

$$\mathbf{G}^\top \mathbf{v} = \mathbf{w} , \quad (53)$$

664 involving a  $D \times D$  sensitivity matrix  $\mathbf{G}$  (equation 3) defined in terms of a given harmonic function  $g_{ij}$   
 665 (equation 2), where

$$\mathbf{v} = \begin{bmatrix} \mathbf{v}^0 \\ \vdots \\ \mathbf{v}^{N_B-1} \end{bmatrix}_{D \times 1}, \quad \mathbf{w} = \begin{bmatrix} \mathbf{w}^0 \\ \vdots \\ \mathbf{w}^{N_B-1} \end{bmatrix}_{D \times 1}, \quad (54)$$

666 are arbitrary partitioned vectors formed by  $N_B$  sub-vectors  $\mathbf{v}^n$  and  $\mathbf{w}^n$ ,  $n \in \{0 : (N_B - 1)\}$ , all of them  
 667 having  $N_b$  elements. Equations 52 and 53 can be computed in terms of an auxiliary matrix-vector product

$$\mathbf{G}_c \mathbf{v}_c = \mathbf{w}_c, \quad (55)$$

668 where

$$\mathbf{v}_c = \begin{bmatrix} \mathbf{v}_c^0 \\ \vdots \\ \mathbf{v}_c^{N_B-1} \\ \mathbf{0} \end{bmatrix}_{4D \times 1}, \quad \mathbf{w}_c = \begin{bmatrix} \mathbf{w}_c^0 \\ \vdots \\ \mathbf{w}_c^{N_B-1} \\ \mathbf{0} \end{bmatrix}_{4D \times 1}, \quad (56)$$

669 are partitioned vectors formed by  $2N_b \times 1$  sub-vectors

$$\mathbf{v}_c^n = \begin{bmatrix} \mathbf{v}_c^n \\ \mathbf{0} \end{bmatrix}_{2N_b \times 1}, \quad \mathbf{w}_c^n = \begin{bmatrix} \mathbf{w}_c^n \\ \mathbf{0} \end{bmatrix}_{2N_b \times 1}, \quad (57)$$

670 and  $\mathbf{G}_c$  is a  $4D \times 4D$  *doubly block circulant* (Jain, 1989, p. 28) or *block-circulant circulant-block* (BCCB)  
 671 (Chan and Jin, 2007, p. 76) matrix. What follows aims at explaining how the original matrix-vector products  
 672 defined by equations 52 and 53, involving a  $D \times D$  BTTB matrix  $\mathbf{G}$  exemplified by equation 51, can be  
 673 efficiently computed in terms of the auxiliary matrix-vector product given by equation 55, which has a  
 674  $4D \times 4D$  BCCB matrix  $\mathbf{G}_c$ .

675 Matrix  $\mathbf{G}_c$  (equation 55) is formed by  $2N_B \times 2N_B$  blocks, where each block  $\mathbf{G}_c^n$ ,  $n \in \{(1 - N_B) : (N_B - 1)\}$  is a  $2N_b \times 2N_b$  circulant matrix. For the case in which the original matrix-vector product is that  
 676 defined by equation 52, the first column of blocks forming the BCCB matrix  $\mathbf{G}_c$  is given by  
 677

$$\mathbf{G}_c[:, :2N_b] = \begin{bmatrix} \mathbf{G}_c^0 \\ \mathbf{G}_c^{-1} \\ \vdots \\ \mathbf{G}_c^{1-N_B} \\ \mathbf{0} \\ \mathbf{G}_c^{N_B-1} \\ \vdots \\ \mathbf{G}_c^1 \end{bmatrix}_{4D \times 2N_b}, \quad (58)$$

678 with blocks  $\mathbf{G}_c^n$  having the first column given by

$$\mathbf{G}_c^n[:, 1] = \begin{bmatrix} \mathbf{G}^n[:, 1] \\ 0 \\ (\mathbf{G}^n[1, N_b : 2])^\top \end{bmatrix}_{2N_b \times 2N_b}, \quad n \in \{(1 - N_B) : (N_B - 1)\}, \quad (59)$$

679 where  $\mathbf{G}^n$  are the blocks forming the BTTB matrix  $\mathbf{G}$  (equation 51). For the case in which the original  
 680 matrix-vector product is that defined by equation 53, the first column of blocks forming the BCCB matrix  
 681  $\mathbf{G}_c$  is given by

$$\mathbf{G}_c[:, : 2N_b] = \begin{bmatrix} \mathbf{G}_c^0 \\ \mathbf{G}_c^1 \\ \vdots \\ \mathbf{G}_c^{N_B-1} \\ \mathbf{0} \\ \mathbf{G}_c^{1-N_B} \\ \vdots \\ \mathbf{G}_c^{-1} \end{bmatrix}_{4D \times 2N_b}, \quad (60)$$

682 with blocks  $\mathbf{G}_c^n$  having the first column given by

$$\mathbf{G}_c^n[:, 1] = \begin{bmatrix} (\mathbf{G}^n[1, :])^\top \\ 0 \\ \mathbf{G}^n[N_b : 2, 1] \end{bmatrix}_{2N_b \times 2N_b}, \quad n \in \{(1 - N_B) : (N_B - 1)\}. \quad (61)$$

683 The complete matrix  $\mathbf{G}_c$  (equation 55) is obtained by properly downshifting the block columns  $\mathbf{G}_c[:, : 2N_b]$   
 684 defined by equation 58 or 60. Similarly, the  $n$ -th block  $\mathbf{G}_c^n$  of  $\mathbf{G}_c$  is obtained by properly downshifting  
 685 the first columns  $\mathbf{G}_c^\ell[:, 1]$  defined by equation 59 or 61.

686 Note that  $\mathbf{G}_c$  (equation 55) is a  $4D \times 4D$  matrix and  $\mathbf{G}$  (equation 51) is a  $D \times D$  matrix. It seems weird  
 687 to say that computing  $\mathbf{G}_c \mathbf{v}_c$  is more efficient than directly computing  $\mathbf{G} \mathbf{v}$ . To understand this, we need first  
 688 to use the fact that BCCB matrices are diagonalized by the 2D unitary discrete Fourier transform (DFT)  
 689 (e.g., Davis, 1979, p. 31). Because of that,  $\mathbf{G}_c$  can be written as

$$\mathbf{G}_c = (\mathcal{F}_{2N_B} \otimes \mathcal{F}_{2N_b})^* \Lambda (\mathcal{F}_{2N_B} \otimes \mathcal{F}_{2N_b}), \quad (62)$$

690 where the symbol “ $\otimes$ ” denotes the Kronecker product (e.g., Horn and Johnson, 1991, p. 243),  $\mathcal{F}_{2N_B}$  and  
 691  $\mathcal{F}_{2N_b}$  are the  $2N_B \times 2N_B$  and  $2N_b \times 2N_b$  unitary DFT matrices (e.g., Davis, 1979, p. 31), respectively,  
 692 the superscript “ $*$ ” denotes the complex conjugate and  $\Lambda$  is a  $4D \times 4D$  diagonal matrix containing the  
 693 eigenvalues of  $\mathbf{G}_c$ . Due to the diagonalization of the matrix  $\mathbf{G}_c$ , equation 55 can be rewritten by using  
 694 equation 62 and premultiplying both sides of the result by  $(\mathcal{F}_{2N_B} \otimes \mathcal{F}_{2N_b})$ , i.e.,

$$\Lambda (\mathcal{F}_{2N_B} \otimes \mathcal{F}_{2N_b}) \mathbf{v}_c = (\mathcal{F}_{2N_B} \otimes \mathcal{F}_{2N_b}) \mathbf{w}_c. \quad (63)$$

695 By following Takahashi et al. (2020), we rearrange equation 63 as follows

$$\mathcal{L} \circ (\mathcal{F}_{2N_B} \mathbf{v}_c \mathcal{F}_{2N_b}) = \mathcal{F}_{2N_B} \mathbf{W}_c \mathcal{F}_{2N_b} \quad (64)$$

696 where “ $\circ$ ” denotes the Hadamard product (e.g., Horn and Johnson, 1991, p. 298) and  $\mathcal{L}$ ,  $\mathbf{V}_c$  and  $\mathbf{W}_c$  are  
 697  $2N_B \times 2N_b$  matrices obtained by rearranging, along their rows, the elements forming the diagonal of  $\Lambda$   
 698 (equation 62), vector  $\mathbf{v}_c$  and vector  $\mathbf{w}_c$  (equation 56), respectively. Then, by premultiplying both sides of  
 699 equation 64 by  $\mathcal{F}_{2N_B}^*$  and then postmultiplying both sides by  $\mathcal{F}_{2N_b}^*$ , we obtain

$$\mathcal{F}_{2N_B}^* [\mathcal{L} \circ (\mathcal{F}_{2N_B} \mathbf{v}_c \mathcal{F}_{2N_b})] \mathcal{F}_{2N_b}^* = \mathbf{W}_c. \quad (65)$$

700 Finally, we get from equation 62 that matrix  $\mathcal{L}$  can be computed by using only the first column  $\mathbf{G}_c[:, 1]$  of  
 701 the BCCB matrix  $\mathbf{G}_c$  (equation 55) according to (Takahashi et al., 2020)

$$\mathcal{L} = \sqrt{4D} \mathcal{F}_{2N_B} \mathcal{C} \mathcal{F}_{2N_b}, \quad (66)$$

702 where  $\mathcal{C}$  is a  $2N_B \times 2N_b$  matrix obtained by rearranging, along its rows, the elements of  $\mathbf{G}_c[:, 1]$  (equation  
 703 55). It is important noting that the matrices  $\mathcal{C}$  and  $\mathcal{L}$  (equation 66) associated with the BTTB matrix  $\mathbf{G}$   
 704 (equation 51) are different from those associated with  $\mathbf{G}^\top$ .

705 The whole procedure to compute the original matrix-vector products  $\mathbf{G}\mathbf{v}$  (equation 52) and  $\mathbf{G}^\top\mathbf{v}$   
 706 (equation 53) consists in (i) rearranging the elements of the vector  $\mathbf{v}$  and the first column  $\mathbf{G}[:, 1]$  of matrix  
 707  $\mathbf{G}$  into the matrices  $\mathcal{V}_c$  and  $\mathcal{C}$  (equations 65 and 66), respectively; (ii) computing terms  $\mathcal{F}_{2N_B} \mathcal{A} \mathcal{F}_{2N_b}$  and  
 708  $\mathcal{F}_{2N_B}^* \mathcal{A} \mathcal{F}_{2N_b}^*$ , where  $\mathcal{A}$  is a given matrix, and a Hadamard product to obtain  $\mathcal{W}_c$  (equation 65); and (iii)  
 709 retrieve the elements of vector  $\mathbf{w}$  (equation 52) from  $\mathcal{W}_c$  (equation 65). It is important noting that the steps  
 710 (i) and (iii) do not have any computational cost because they involve only reorganizing elements of vectors  
 711 and matrices. Besides, the terms  $\mathcal{F}_{2N_B} \mathcal{A} \mathcal{F}_{2N_b}$  and  $\mathcal{F}_{2N_B}^* \mathcal{A} \mathcal{F}_{2N_b}^*$  in step (ii) represent, respectively, the  
 712 2D Discrete Fourier Transform (2D-DFT) and the 2D Inverse Discrete Fourier Transform (2D-IDFT) of  $\mathcal{A}$ .  
 713 These transforms can be efficiently computed by using the 2D Fast Fourier Transform (2D-FFT). Hence,  
 714 the original matrix-vector products  $\mathbf{G}\mathbf{v}$  (equation 52) and  $\mathbf{G}^\top\mathbf{v}$  (equation 53) can be efficiently computed  
 715 by using the 2D-FFT.

716 Algorithms 7 and 8 show pseudo-codes for the convolutional equivalent-layer method proposed by  
 717 Takahashi et al. (2020, 2022). Note that those authors formulate the overdetermined problem (equation  
 718 22) of obtaining an estimate  $\tilde{\mathbf{p}}$  for the parameter vector  $\mathbf{p}$  (equation 3) as an *iterative deconvolution* via  
 719 *conjugate gradient normal equation residual* (CGNR) Golub and Van Loan (2013, sec. 11.3) or *conjugate*  
 720 *gradient least squares* (CGLS) (Aster et al., 2019, p. 165) method. They consider  $\mathbf{H} = \mathbf{I}_P$  (equation 9),  
 721  $\mu = 0$  (equation 11),  $\mathbf{W}_d = \mathbf{W}_q = \mathbf{I}_P$  (equations 12 and 13) and  $\bar{\mathbf{p}} = \mathbf{0}$  (equation 14). As shown by  
 722 Takahashi et al. (2020, 2022), the CGLS produces stable estimates  $\tilde{\mathbf{p}}$  for the parameter vector  $\mathbf{p}$  (equation  
 723 3) in the presence of noisy potential-field data  $\mathbf{d}$ . This is a well-known property of the CGLS method (e.g.,  
 724 Aster et al., 2019, p. 166).

725 The key aspect of Algorithm 7 is replacing the matrix-vector products of CGLS (Algorithm 1) by fast  
 726 convolutions (Algorithm 8). A fast convolution requires one 2D-DFT, one 2D-IDFT and an entrywise  
 727 product of matrices. We consider that the 2D-DFT/IDFT are computed with 2D-FFT and requires  
 728  $\lambda(4D) \log_2(4D)$  flops, where  $\lambda = 5$  is compatible with a radix-2 FFT (Van Loan, 1992, p. 16), and  
 729 the entrywise product  $24D$  flops because it involves two complex matrices having  $4D$  elements (Golub  
 730 and Van Loan, 2013, p. 36). Hence, Algorithm 8 requires  $\lambda(16D) \log_2(4D) + 26D$  flops, whereas a  
 731 conventional matrix-vector multiplication involving a  $D \times D$  matrix requires  $2D^2$  (table 1). Finally,  
 732 Algorithm 7 requires two 2D-FFTs (lines 4 and 5), one fast convolution and an inner product (line 8)  
 733 previously to the while loop. Per iteration, there are three saxpys (lines 12, 15 and 16), two inner products  
 734 (lines 14 and 17) and two fast convolutions (lines 13 and 17), so that:

$$f_{\text{T0B20}} = \lambda(16D) \log_2(4D) + 26D + \text{ITMAX} [\lambda(16D) \log_2(4D) + 58D]. \quad (67)$$

## 735 7.8 Direct deconvolution

736 The method proposed by Takahashi et al. (2020, 2022) can be reformulated to avoid the iterations of the  
 737 conjugate gradient method. This alternative formulation consists in considering that  $\mathbf{v} = \mathbf{p}$  and  $\mathbf{w} = \mathbf{d}$  in

738 equation 52, where  $\mathbf{p}$  is the parameter vector (equation 3) and  $\mathbf{d}$  the observed data vector. In this case, the  
 739 equality “=” in equation 52 becomes an approximation “ $\approx$ ”. Then, equation 64 is manipulated to obtain

$$\mathcal{V}_c \approx \mathcal{F}_{2N_B}^* \left[ (\mathcal{F}_{2N_B} \mathcal{W}_c \mathcal{F}_{2N_b}) \circ \check{\mathcal{L}} \right] \mathcal{F}_{2N_b}^*, \quad (68)$$

740 where

$$\check{\mathcal{L}} = \mathcal{L}^* \oslash (\mathcal{L} \circ \mathcal{L}^* + \zeta \mathbf{1}), \quad (69)$$

741  $\mathbf{1}$  is a  $4D \times 4D$  matrix of ones, “ $\oslash$ ” denotes entrywise division and  $\zeta$  is a positive scalar. Note that  $\zeta = 0$   
 742 leads to  $\mathbf{1} \oslash \mathcal{L}$ . In this case, the entrywise division may be problematic due to the elements of  $\mathcal{L}$  having  
 743 absolute value equal or close to zero. So, a small  $\zeta$  is set to avoid this problem in equation 69. Next, we use  
 744  $\check{\mathcal{L}}$  to obtain a matrix  $\mathcal{V}_c$  from equation 68. Finally, the elements of the estimated parameter vector  $\tilde{\mathbf{p}}$  are  
 745 retrieved from the first quadrant of  $\mathcal{V}_c$ . This procedure represents a *direct deconvolution* (e.g., Aster et al.,  
 746 2019, p. 220) using a *Wiener filter* (e.g., Gonzalez and Woods, 2002, p. 263).

747 The required total number of flops associated with the direct deconvolution aggregates one 2D-FFT  
 748 to compute matrix  $\mathcal{L}$  (equation 66), one entrywise product  $\mathcal{L} \circ \mathcal{L}^*$  involving complex matrices and one  
 749 entrywise division to compute  $\check{\mathcal{L}}$  (equation 69) and a fast convolution (Algorithm 8) to evaluate equation  
 750 68, which results in:

$$f_{\text{deconv.}} = \lambda (12D) \log_2(4D) + 72D. \quad (70)$$

751 Differently from the convolutional equivalent-layer method proposed by Takahashi et al. (2020, 2022), the  
 752 alternative direct deconvolution presented here produces an estimated parameter vector  $\tilde{\mathbf{p}}$  directly from  
 753 the observed data  $\mathbf{d}$ , in a single step, avoiding the conjugate gradient iterations. On the other hand, the  
 754 alternative method presented here requires estimating a set of tentative parameter vectors  $\tilde{\mathbf{p}}$  for different  
 755 predefined  $\zeta$ . Besides, there must be criterion to chose the best  $\tilde{\mathbf{p}}$  from this tentative set. This can be  
 756 made, for example, by using the well-known *L-curve* (Hansen, 1992). From a computational point of view,  
 757 the number of CGLS iterations in the method proposed by Takahashi et al. (2020, 2022) is equivalent to  
 758 the number of tentative estimated parameter vectors required to form the L-curve in the proposed direct  
 759 deconvolution.

## 8 NUMERICAL SIMULATIONS

### 760 8.1 Flops count

761   Figure 1 shows the total number of flops for solving the overdetermined problem (equation 22) with  
 762 different equivalent-layer methods (equations 27, 28, 35, 37, 38, 43, 50, 67, and 70), by considering  
 763 the particular case in which  $\mathbf{H} = \mathbf{I}_P$  (equation 9 and subsection 3.2),  $\mu = 0$  (equation 11),  $\mathbf{W}_d = \mathbf{I}_D$   
 764 (equation 12) and  $\bar{\mathbf{p}} = \mathbf{0}$  (equation 14), where  $\mathbf{I}_P$  and  $\mathbf{I}_D$  are the identities of order  $P$  and  $D$ , respectively.  
 765 The flops are computed for different number of potential-field data ranging from 10,000 to 1,000,000.  
 766 Figure 1 shows that the moving data-window strategy by using Leão and Silva's 1989 method and direct  
 767 deconvolution are the fastest methods.

768   The control parameters to run the equivalent-layer methods shown in Figure 1 are the following: i) in  
 769 CGLS, reparameterization approaches (e.g., Oliveira Jr. et al., 2013; Mendonça, 2020), Siqueira et al.  
 770 (2017), and Takahashi et al. (2020) (equations 28 38), 43, 50 and 67) we set ITMAX = 50; ii) Cordell  
 771 (1992) we set ITMAX = 6  $D$ ; iii) in Leão and Silva (1989) (equation 35) we set  $D' = 49$  ( $7 \times 7$ ) and  
 772  $P' = 225$  ( $15 \times 15$ ); and iv) in Soler and Uieda (2021) (equation 37) we set  $D' = P' = 900$  ( $30 \times 30$ ).

### 773 8.2 Synthetic potential-field data

774   We create a model composed of several rectangular prisms that can be split into three groups. The first  
 775 is composed of 300 small cubes (not shown) with top at 0 m and side lengths defined according to a  
 776 pseudo-random variable having uniform distribution from 100 to 200 m. Their density contrasts are defined  
 777 by a pseudo-random variable uniformly distributed from 1000 to 2000 kg/m<sup>3</sup>. These prisms produce the  
 778 short-wavelength component of the simulated gravity data. The 4 prisms forming the second group of our  
 779 model (indicated by A-D in Figure 2) have tops varying from 10 to 100 m and bottom from 1010 to 1500 m.  
 780 They have density contrasts of 1500, -1800, -3000 and 1200 kg/m<sup>3</sup> and side lengths varying from 1000  
 781 to 4000 m. These prisms produce the mid-wavelength component of the simulated gravity data. There is  
 782 also a single prism (indicated by E in Figure 2) with top at 1000 m, bottom at 1500 m and side lengths of  
 783 4000 and 6000 m. This prism has density contrast is -900 kg/m<sup>3</sup> and produces the long-wavelength of our  
 784 synthetic gravity data.

785   We have computed noise-free gravity disturbance and gravity-gradient tensor components produced by  
 786 our model on a regularly spaced grid of  $50 \times 50$  points at  $z = -100$  m (Figure 3). We have also simulated  
 787 additional  $L = 20$  gravity disturbance data sets  $\mathbf{d}^\ell$ ,  $\ell \in \{1 : L\}$ , by adding pseudo-random Gaussian noise  
 788 with zero mean and crescent standard deviations to the noise-free data (not shown). The standard deviations  
 789 vary from 0.5% to 10% of the maximum absolute value in the noise-free data, which corresponds to 0.21  
 790 and 4.16 mGal, respectively.

### 791 8.3 Stability analysis and gravity-gradient components

792   We set a planar equivalent layer of point masses having one source below each datum at a constant  
 793 vertical coordinate  $z \approx 512.24$  m. This depth was set by following the Dampney's (1969) criterion (see  
 794 Subsection 2.1), so that the vertical distance  $\Delta z$  between equivalent sources and the simulated data is equal  
 795 to  $3 \times$  the grid spacing ( $\Delta x = \Delta y \approx 204.08$  m). Note that, in this case, the layer has a number of sources  
 796  $P$  equal to the number of data  $D$ .

797   We have applied the Cholesky factorization (equations 25 and 26), CGLS (Algorithm 1), the iterative  
 798 method of Siqueira et al. (2017) (Algorithm 6), the iterative deconvolution (Algorithms 7 and 8) proposed

799 by Takahashi et al. (2020) (Algorithm 7) and the direct deconvolution (equations 68 and 69) with four  
 800 different values for the parameter  $\zeta$  to the 21 gravity data sets.

801 For each method, we have obtained one estimate  $\tilde{\mathbf{p}}$  from the noise-free gravity data  $\mathbf{d}$  and  $L = 20$   
 802 estimates  $\tilde{\mathbf{p}}^\ell$  from the noise-corrupted gravity data  $\mathbf{d}^\ell$ ,  $\ell \in \{1 : L\}$ , for the planar equivalent layer of point  
 803 masses, totaling 21 estimated parameter vectors and 20 pairs  $(\Delta p^\ell, \Delta d^\ell)$  of model and data perturbations  
 804 (equations 30 and 31). Figure 4 shows the numerical stability curves computed with each method for the  
 805 synthetic gravity data.

806 All these 21 estimated parameters vectors were obtained by solving the overdetermined problem (equation  
 807 22) with the same method for the particular case in which  $\mathbf{H} = \mathbf{I}$  (equation 9 and subsection 3.2),  
 808  $\mathbf{W}_d = \mathbf{W}_q = \mathbf{I}$  (equations 12 and 13) and  $\bar{\mathbf{p}} = \mathbf{0}$  (equation 14), where  $\mathbf{I}$  is the identity of order  $D$ .

809 Figure 4 shows how the numerical stability curves vary as the level of the noise is increased. We can see  
 810 that for all methods, a linear tendency is observed as it is expected. The inclination of the straight line  
 811 indicates the stability of each method. As shown in Figure 4, the direct deconvolution with  $\zeta = 0$  exhibits a  
 812 high slope, which indicates high instability and emphasizes the necessity of using the Wiener filter ( $\zeta > 0$   
 813 in equation 69).

814 The estimated stability parameters  $\kappa$  (equation 29) obtained for the Cholesky factorization, CGLS and  
 815 iterative deconvolution are close to each other (Figures 4). They are slightly smaller than that obtained for  
 816 the iterative method of Siqueira et al. (2017) (S0B17). Note that by varying the parameter  $\zeta$  (equation 69) it  
 817 is possible to obtain different stability parameters  $\kappa$  for the direct deconvolution. There is no apparent rule  
 818 to set  $\zeta$ . A practical criterion can be the maximum  $\zeta$  producing a satisfactory data fit. Overshoot values  
 819 tend to exaggeratedly smooth the predicted data.

820 We inverted the noise-corrupted gravity disturbance with the highest noise level (not shown) to estimate  
 821 an equivalent layer (not shown) via iterative deconvolution (Algorithm 7). Figure 5(G) shows the residuals  
 822 (in mGal) between the predicted and noise-corrupted gravity disturbances. As we can see, the residuals  
 823 are uniformly distributed on simulated area and suggest that the equivalent layer produces a good data fit.  
 824 This can be verified by inspecting the histogram of the residuals between the predicted and noise-corrupted  
 825 gravity disturbances shown in panel (G) of Figure 6.

826 Using the estimated layer, we have computed the gravity-gradient data (not shown) at the observations  
 827 points. Figures 5 (A)–(F) show the residuals (in Eötvös) between the predicted (not shown) and noise-free  
 828 gravity-gradient data (Figure 3). These figures show that the iterative deconvolution (Algorithm 7) could  
 829 predict the six components of the gravity-gradient tensor with a good precision, which can also be verified  
 830 in the corresponding histograms shown in Figure 6.

831 In the supplementary material, we show the residuals between the gravity data predicted by the equivalent  
 832 layer estimated by using the following methods: i) the CGLS method (Algorithm 1); ii) the Cholesky  
 833 factorization (equations 25 and 26); iii) the iterative method proposed by Siqueira et al. (2017) (Algorithm  
 834 6); iv) the direct deconvolution with optimal value of  $\zeta = 10^{-22}$  (equation 69); and v) the iterative method  
 835 proposed by Cordell (1992) (Algorithm 4).

## 9 APPLICATIONS TO FIELD DATA

836 In this section, we show the results obtained by applying the iterative deconvolution (Algorithm 7) to a  
 837 field data set over the Carajás Mineral Province (CMP) in the Amazon craton (Moroni et al., 2001; Villas  
 838 and Santos, 2001). This area (Figure 7) is known for its intensive mineral exploration such as iron, copper,  
 839 gold, manganese, and, recently, bauxite.

### 840 9.1 Geological setting

841 The Amazon Craton is one of the largest and least-known Archean-Proterozoic areas in the world,  
 842 comprehending a region with a thousand square kilometers. It is one of the main tectonic units in South  
 843 America, which is covered by five Phanerozoic basins: Maranhão (Northeast), Amazon (Central), Xingu-  
 844 Alto Tapajós (South), Parecis (Southwest), and Solimões (West). The Craton is limited by the Andean  
 845 Orogenic Belt to the west and the by Araguaia Fold Belt to the east and southeast. The Amazon craton has  
 846 been subdivided into provinces according to two models, one geochronological and the other geophysical-  
 847 structural (Amaral, 1974; Teixeira et al., 1989; Tassinari and Macambira, 1999). Thus, seven geological  
 848 provinces with distinctive ages, evolution, and structural patterns can be observed, namely: (i) Carajás with  
 849 two domains - the Mesoarchean Rio Maria and Neoarchean Carajás; (ii) Archean-Paleoproterozoic Central  
 850 Amazon, with Iriri-Xingu and Curuá-Mapuera domains; (ii) Trans-Amazonian (Ryacian), with the Amapá  
 851 and Bacajá domains; (iv) the Orosinian Tapajós-Parima, with Peixoto de Azevedo, Tapajós, Uaimiri, and  
 852 Parima domains; (v) Rondônia-Juruena (Statherian), with Jamari, Juruena, and Jauru domains; (vi) The  
 853 Statherian Rio Negro, with Rio Negro and Imeri domains; and (vii) Sunsás (Meso-Neoproterozoic), with  
 854 Santa Helena and Nova Brasilândia domains (Santos et al., 2000). Nevertheless, we focus this work only  
 855 on the Carajás Province.

856 The Carajás Mineral Province (CMP) is located in the east-southeast region of the craton (Figure 7),  
 857 within an old tectonically stable nucleus in the South American Plate that became tectonically stable at the  
 858 beginning of Neoproterozoic (Salomao et al., 2019). This area has been the target of intensive exploration  
 859 at least since the final of the '60s, after the discovery of large iron ore deposits. There are several greenstone  
 860 belts in the region, among them are the Andorinhas, Inajá, Cumaru, Carajás, Serra Leste, Serra Pelada, and  
 861 Sapucaia (Santos et al., 2000). The mineralogic and petrologic studies in granite stocks show a variety of  
 862 minerals found in the province, such as amphibole, plagioclase, biotite, ilmenite, and magnetite (Cunha  
 863 et al., 2016).

### 864 9.2 Potential-field data

865 The field data used here were obtained from an airborne survey conducted by Lasa Prospecções S/A.  
 866 and Microsurvey Aerogeofísica Consultoria Científica Ltda. between April/2013 and October/2014. The  
 867 survey area covers  $\approx 58000 \text{ km}^2$  between latitudes  $-8^\circ / -5^\circ$  and longitudes  $-53^\circ / -49.5^\circ$  referred to the  
 868 WGS-84 datum. We obtained the horizontal coordinates  $x$  and  $y$  already in the UTM zone 22S. The flight  
 869 and tie lines are spaced at 3 km and 12 km, with orientation along directions  $N - S$  and  $E - W$ , respectively.  
 870 The data are placed at an approximately constant distance of 900 m above the ground. Figure 8 shows the  
 871  $D = 500,000$  aerogravimetric data on a grid of  $1000 \times 500$  observation points with  $\Delta x = 358.12 \text{ m}$  and  
 872  $\Delta y = 787.62 \text{ m}$ .

### 873 9.3 Potential-field transformation

874 We applied the equivalent-layer technique to the observed data (Figure 8) with the purpose of illustrating  
 875 how to estimate the gravity-gradient tensor over the study area. We used an equivalent layer layout with

876 one source located below each datum (so that  $P = D$ ) on a horizontal plane having a vertical distance  
877  $\Delta z \approx 2362.86$  m from the observation plane. This setup is defined by setting  $\Delta z \approx 3 dy$ , which follows the  
878 same strategy of Reis et al. (2020). We solve the linear inverse problem for estimating the physical-property  
879 distribution on the layer by using the iterative deconvolution (Algorithm 7) with a maximum number of 50  
880 iterations. Actually, the algorithm have converged with only 18 iterations.

881 Figure 9(G) shows the histogram of the residuals between the predicted (not shown) and observed data  
882 (Figure 8). As we can see, the iterative deconvolution produced an excellent data fit. By using the estimated  
883 layer, we have computed the gravity-gradient tensor components at the observation points. The results are  
884 shown in Figures 9(A)–(F).

885 Considering the processing time, the iterative deconvolution took  $\approx 1.98$  s to execute the 18  
886 iterations for estimating the physical-property distribution on the layer by inverting the  $D = 500,000$   
887 observed data. The code was run in a modest computer with 16,0 GiB of memory and processor  
888 12th Gen Intel Core i9 – 12900H  $\times 20$ . Given the estimated equivalent layer, the gravity-gradient  
889 components shown in Figure 9 were computed in  $\approx 3.41$  s. These results demonstrate the efficiency  
890 of the iterative deconvolution method in processing large datasets.

## 10 DISCUSSION

891 The review discusses strategies utilized to reduce the computational cost of the equivalent-layer technique  
892 for processing potential-field data. These strategies are often combined in developed methods to efficiently  
893 handle large-scale data sets. Next, the computational strategies are addressed.

894 The first one is the moving data-window scheme spanning the data set. This strategy solves several much  
895 smaller, regularized linear inverse problems instead of a single large one. Each linear inversion is solved  
896 using the potential-field observations and equivalent sources within a given moving window and can be  
897 applied to both regularly or irregularly spaced data sets. If the data and the sources are distributed on planar  
898 and regularly spaced grids, this strategy offers a significant advantage because the sensitivity submatrix of  
899 a given moving window remains the same for all windows. Otherwise, the computational efficiency of the  
900 equivalent-layer technique using the moving-window strategy decreases because the sensitivity submatrix  
901 for each window must be computed.

902 The second and third strategies, referred to as the column-action and row-action updates, involve  
903 iteratively calculating a single column and a single row of the sensitivity matrix, respectively. By following  
904 the column-action update strategy, a single column of the sensitivity matrix is calculated during each  
905 iteration. This implies that a single equivalent source contributes to the fitting of data in each iteration.  
906 Conversely, in the row-action update strategy, a single row of the sensitivity matrix is calculated per  
907 iteration, which means that one potential-field observation is incorporated in each iteration, forming a new  
908 subset of equivalent data much smaller than the original data. Both strategies (column- and row-action  
909 updates) have a great advantage because a single column or a single row of the sensitivity matrix is  
910 calculated iteratively. However, to our knowledge, the strategy of the column-action update presents some  
911 issues related to convergence, and the strategy of the row-action update can also have issues if the number  
912 of equivalent data is not significantly smaller than the original number of data points.

913 The fourth strategy is the sparsity induction of the sensitivity matrix using wavelet compression, which  
914 involves transforming a full sensitivity matrix into a sparse one with only a few nonzero elements. The  
915 developed equivalent-layer methods using this strategy achieve sparsity by setting matrix elements to  
916 zero if their values are smaller than a predefined threshold. We highlight two methods that employ the  
917 sparsity induction strategy. The first method, known as wavelet-compression equivalent layer, compresses  
918 the coefficients of the original sensitivity matrix using discrete wavelet transform, achieves sparsity in the  
919 sensitivity matrix, and solves the inverse problem in the wavelet domain without an explicit regularization  
920 parameter. The regularized solution in the wavelet domain is estimated using a conjugate gradient (CG)  
921 least squares algorithm, where the number of iterations serves as a regularization factor. The second  
922 equivalent-layer method that uses the sparsity induction strategy applies quadtree discretization of the  
923 parameters over the equivalent layer, achieves sparsity in the sensitivity matrix, and solves the inverse  
924 problem using CG algorithm. In quadtree discretization, equivalent sources located far from the observation  
925 point are grouped together to form larger equivalent sources, reducing the number of parameters to be  
926 estimated. Computationally, the significant advantage of the equivalent-layer methods employing wavelet  
927 compression and quadtree discretization is the sparsity induction in the sensitivity matrix, which allows  
928 for fast iteration of the CG algorithm. However, we acknowledge that this strategy requires computing  
929 the full and dense sensitivity matrix, which can be considered a drawback when processing large-scale  
930 potential-field data.

931 The fifth strategy is the reparametrization of the original parameters to be estimated in the equivalent-layer  
932 technique. In this strategy, the developed equivalent-layer methods reduce the dimension of the linear

933 system of equations to be solved by estimating a lower-dimensional parameter vector. We highlight two  
934 methods that used the reparametrization strategy: i) the polynomial equivalent layer (PEL) and; ii) the  
935 lower-dimensional subspace of the equivalent layer. In the PEL, there is an explicit reparametrization  
936 of the equivalent layer by representing the unknown distribution over the equivalent layer as a set of  
937 piecewise-polynomial functions defined on a set of equivalent-source windows. The PEL method estimates  
938 the polynomial coefficients of all equivalent-source windows. Hence, PEL reduces the dimension of the  
939 linear system of equations to be solved because the polynomial coefficients within all equivalent-source  
940 windows are much smaller than both the number of equivalent sources and the number of data points.  
941 In the lower-dimensional subspace of the equivalent layer, there is an implicit reparametrization of the  
942 equivalent layer by reducing the linear system dimension from the original and large-model space to a  
943 lower-dimensional subspace. The lower-dimensional subspace is grounded on eigenvectors of the matrix  
944 composed by the gridded data set. The main advantage of the reparametrization of the equivalent layer is to  
945 deal with lower-dimensional linear system of equations. However, we acknowledge that this strategy may  
946 impose an undesirable smoothing effect on both the estimated parameters over the equivalent layer and the  
947 predicted data.

948 The sixth strategy involves an iterative scheme in which the estimated distribution over the equivalent  
949 layer is updated iteratively. Following this strategy, the developed equivalent-layer methods differ either in  
950 terms of the expression used for the estimated parameter correction or the domain utilized (wavenumber or  
951 space domains). The iterative estimated correction may have a physical meaning, such as the excess mass  
952 constraint. All the iterative methods are efficient as they can handle irregularly spaced data on an undulating  
953 surface, and the updated corrections for the parameter vector at each iteration are straightforward, involving  
954 the addition of a quantity proportional to the data residual. However, they have a disadvantage because the  
955 iterative strategy requires computing the full and dense sensitivity matrix to compute the predicted and  
956 residual data in all observation stations per iteration.

957 The seventh strategy is called iterative deconvolutional of the equivalent layer. This strategy deals with  
958 regularly spaced grids of data stations and equivalent sources which are located at a constant height and  
959 depth, respectively. Specifically, one source is placed directly below each observation station, which results  
960 in sensitivity matrices with a BTTB (Block-Toeplitz Toeplitz-Block) structure. It is possible to embed the  
961 BTTB matrix into a matrix of Block-Circulant Circulant-Block (BCCB) structure, which requires only  
962 one equivalent source. This allows for fast matrix-vector product using a 2D fast Fourier transform (2D  
963 FFT). As a result, the potential-field forward modeling can be calculated using a 2D FFT with only one  
964 equivalent source required. The main advantages of this strategy are that the entire sensitivity matrices  
965 do not need to be formed or stored; only their first columns are required. Additionally, it allows for a  
966 highly efficient iteration of the CG algorithm. However, the iterative deconvolutional of the equivalent  
967 layer requires observations and equivalent sources aligned on a horizontal and regularly-spaced grid.

968 The eighth strategy is a direct deconvolution method, which is a mathematical process very common in  
969 geophysics. However, to our knowledge, direct deconvolution has never been used to solve the inverse  
970 problem associated with the equivalent-layer technique. From the mathematical expressions in the iterative  
971 deconvolutional equivalent layer with BTTB matrices, direct deconvolution arises naturally since it is an  
972 operation inverse to convolution. The main advantage of applying the direct deconvolution strategy in  
973 the equivalent layer is that it avoids, for example, the iterations of the CG algorithm. However, the direct  
974 deconvolution is known to be an unstable operation. To mitigate this instability, the Wiener deconvolution  
975 method can be adopted.

976 Table 2 presents a list of computational strategies used in the equivalent-layer technique to reduce  
977 the computational demand. The table aims to emphasize the important characteristics, advantages, and  
978 disadvantages of each computational strategy. Additionally, it highlights the available methods that use  
979 each strategy.

## 11 CONCLUSION

We have presented a comprehensive review of the strategies used to tackle the intensive computational cost associated with processing potential-field data using the equivalent-layer technique. Each of these strategies is rarely used individually; rather, some developed equivalent-layer methods combine more than one strategy to achieve computational efficiency when dealing with large-scale data sets. We focus on the following specific strategies: (1) the moving data-window scheme; (2) the column-action and row-action updates; (3) the sparsity induction of the sensitivity matrix; (4) the reparametrization of the original parameters; (5) the iterative scheme using the full sensitivity matrix; (6) the iterative deconvolution; and (7) the direct deconvolution. Taking into account the mathematical bases used in the above-mentioned strategies, we have identified five groups: i) the reduction of the dimensionality of the linear system of equations to be solved; ii) the generation of a sparse linear system of equations to be solved; iii) the explicit iterative method; iv) the improvement in forward modeling; and v) the deconvolution using the concept of block-Toeplitz Toeplitz-block (BTTB) matrices.

We show in this review that the computational cost of the equivalent layer can vary from up to  $10^9$  flops depending on the method without compromising the linear system stability. The moving data-window scheme and direct deconvolution are the fastest methods; however, they both have drawbacks. To be computationally efficient, the moving data-window scheme and the direct deconvolution require data and equivalent sources that are distributed on planar and regularly spaced grids. Moreover, they both require choosing an optimum parameter of stabilization. We stress that the direct deconvolution has an additional disadvantage in terms of a higher data residual and border effects over the equivalent layer after processing. These effects can be seen from the upward continuation of the real data from Carajás.

We draw the readers' attention to the possibility of combining more than one aforementioned strategies for reducing the computational cost of the equivalent-layer technique.

## CONFLICT OF INTEREST STATEMENT

1002 The authors declare that the research was conducted in the absence of any commercial or financial  
1003 relationships that could be construed as a potential conflict of interest.

## AUTHOR CONTRIBUTIONS

1004 VCOJr and VCFB: Study conception and mathematical deductions. VCOJr: Algorithms. VCOJr and DT:  
1005 Python codes and synthetic data applications. VCOJr and ALAR: Real data applications. VCOJr and  
1006 VCFB: Result analysis and draft manuscript preparation.

## FUNDING

1007 VCFB was supported by fellowships from CNPq (grant 309624/2021-5) and FAPERJ (grant  
1008 26/202.582/2019). VCOJr was supported by fellowships from CNPq (grant 315768/2020-7) and FAPERJ  
1009 (grant E-26/202.729/2018). DT was supported by a Post-doctoral scholarship from CNPq (grant  
1010 300809/2022-0).

## ACKNOWLEDGMENTS

1011 We thank the Brazilian federal agencies CAPES, CNPq, state agency FAPERJ and Observatório Nacional  
1012 research institute and Universidade do Estado do Rio de Janeiro.

## DATA AVAILABILITY STATEMENT

1013 The datasets generated for this study can be found in the frontiers-paper Github repository link:  
1014 <https://github.com/XXXXXX/eqlayer-review-computational>.

## REFERENCES

- 1015 Amaral, G. (1974). *Geologia Pré-Cambriana da Região Amazônica*. Ph.D. thesis, Universidade de São  
1016 Paulo
- 1017 Aster, R. C., Borchers, B., and Thurber, C. H. (2019). *Parameter Estimation and Inverse Problems*  
1018 (Elsevier), 3 edn.
- 1019 Barbosa, V. C. F., Silva, J. B., and Medeiros, W. E. (1997). Gravity inversion of basement relief using  
1020 approximate equality constraints on depths. *Geophysics* 62, 1745–1757
- 1021 Barnes, G. and Lumley, J. (2011). Processing gravity gradient data. *GEOPHYSICS* 76, I33–I47. doi:10.  
1022 1190/1.3548548
- 1023 Blakely, R. J. (1996). *Potential Theory in Gravity and Magnetic Applications* (Cambridge University  
1024 press)
- 1025 Bott, M. H. P. (1960). The use of Rapid Digital Computing Methods for Direct Gravity Interpretation  
1026 of Sedimentary Basins. *Geophysical Journal International* 3, 63–67. doi:10.1111/j.1365-246X.1960.  
1027 tb00065.x
- 1028 Chan, R. H.-F. and Jin, X.-Q. (2007). *An introduction to iterative Toeplitz solvers*, vol. 5 (SIAM)
- 1029 Cordell, L. (1992). A scattered equivalent-source method for interpolation and gridding of potential-field  
1030 data in three dimensions. *Geophysics* 57, 629–636

- 1031 Cunha, I. R., Dall'Agnol, R., and Feio, G. R. L. (2016). Mineral chemistry and magnetic petrology of the  
1032 archean Planalto Suite, Carajás Province – Amazonian Craton: Implications for the evolution of ferroan  
1033 archean granites. *Journal of South American Earth Sciences* 67, 100–121
- 1034 Dampney, C. N. G. (1969). The equivalent source technique. *GEOPHYSICS* 34, 39–53. doi:10.1190/1.  
1035 1439996
- 1036 Davis, P. J. (1979). *Circulant matrices* (John Wiley & Sons, Inc.)
- 1037 Elfving, T., Hansen, P. C., and Nikazad, T. (2017). Convergence analysis for column-action methods in  
1038 image reconstruction. *Numerical Algorithms* 74, 905–924. doi:10.1007/s11075-016-0176-x
- 1039 Emilia, D. A. (1973). Equivalent sources used as an analytic base for processing total magnetic field  
1040 profiles. *GEOPHYSICS* 38, 339–348. doi:10.1190/1.1440344
- 1041 Golub, G. H. and Van Loan, C. F. (2013). *Matrix Computations*. Johns Hopkins Studies in the Mathematical  
1042 Sciences (Johns Hopkins University Press), 4 edn.
- 1043 Gonzalez, R. C. and Woods, R. E. (2002). *Digital Image Processing* (Prentice Hall), 2 edn.
- 1044 Gonzalez, S. P., Barbosa, V. C. F., and Oliveira Jr., V. C. (2022). Analyzing the ambiguity of the remanent-  
1045 magnetization direction separated into induced and remanent magnetic sources. *Journal of Geophysical  
1046 Research: Solid Earth* 127, 1–24. doi:10.1029/2022JB024151
- 1047 Guspí, F., Introcaso, A., and Introcaso, B. (2004). Gravity-enhanced representation of measured geoid  
1048 undulations using equivalent sources. *Geophysical Journal International* 159, 1–8. doi:10.1111/j.  
1049 1365-246X.2004.02364.x
- 1050 Guspí, F. and Novara, I. (2009). Reduction to the pole and transformations of scattered magnetic data using  
1051 Newtonian equivalent sources. *GEOPHYSICS* 74, L67–L73. doi:10.1190/1.3170690
- 1052 Hansen, P. C. (1992). Analysis of discrete ill-posed problems by means of the l-curve. *SIAM Review* 34,  
1053 561–580. doi:10.1137/1034115
- 1054 Hansen, R. O. and Miyazaki, Y. (1984). Continuation of potential fields between arbitrary surfaces.  
1055 *GEOPHYSICS* 49, 787–795. doi:10.1190/1.1441707
- 1056 Horn, R. A. and Johnson, C. R. (1991). *Topics in Matrix Analysis* (Cambridge University Press), 1 edn.
- 1057 Jain, A. K. (1989). *Fundamentals of Digital Image Processing* (Pearson), 1 edn.
- 1058 Jirigalatu, J. and Ebbing (2019). A fast equivalent source method for airborne gravity gradient data.  
1059 *Geophysics* 84, G75–G82. doi:10.1190/GEO2018-0366.1
- 1060 Kellogg, O. D. (1967). *Foundations of Potential Theory* (Springer-Verlag), reprint from the first edition of  
1061 1929 edn.
- 1062 Kennett, B., Sambridge, M., and Williamson, P. (1988). Subspace methods for large inverse problems with  
1063 multiple parameter classes. *Geophysical Journal International* 94, 237–247
- 1064 Leão, J. W. D. and Silva, J. B. C. (1989). Discrete linear transformations of potential field data. *Geophysics*  
1065 54, 497–507. doi:10.1190/1.1442676
- 1066 Li, Y., Nabighian, M., and Oldenburg, D. W. (2014). Using an equivalent source with positivity for low-  
1067 latitude reduction to the pole without striation. *GEOPHYSICS* 79, J81–J90. doi:10.1190/geo2014-0134.  
1068 1
- 1069 Li, Y. and Oldenburg, D. W. (2010). Rapid construction of equivalent sources using wavelets.  
1070 *GEOPHYSICS* 75, L51–L59. doi:10.1190/1.3378764
- 1071 Mendonça, C. A. (2020). Subspace method for solving large-scale equivalent layer and density mapping  
1072 problems. *GEOPHYSICS* 85, G57–G68. doi:10.1190/geo2019-0302.1
- 1073 Mendonça, C. A. and Silva, J. B. C. (1994). The equivalent data concept applied to the interpolation of  
1074 potential field data. *Geophysics* 59, 722–732. doi:10.1190/1.1443630
- 1075 Menke, W. (2018). *Geophysical data analysis: Discrete inverse theory* (Elsevier), 4 edn.

- 1076 Moroni, M., Girardi, V., and Ferrario, A. (2001). The Serra Pelada Au-PGE deposit, Serra dos Carajás (Pará  
1077 state, Brazil): geological and geochemical indications for a composite mineralising process. *Mineralium  
1078 Deposita* 36, 768–785
- 1079 Oldenburg, D., McGillivray, P., and Ellis, R. (1993). Generalized subspace methods for large-scale inverse  
1080 problems. *Geophysical Journal International* 114, 12–20
- 1081 Oliveira Jr., V. C., Barbosa, V. C. F., and Uieda, L. (2013). Polynomial equivalent layer. *GEOPHYSICS* 78,  
1082 G1–G13. doi:10.1190/geo2012-0196.1
- 1083 Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical recipes: the art of  
1084 scientific computing* (Cambridge University Press), 3 edn.
- 1085 Reis, A. L. A., Oliveira Jr., V. C., and Barbosa, V. C. F. (2020). Generalized positivity constraint on  
1086 magnetic equivalent layers. *Geophysics* 85, 1–45. doi:10.1190/geo2019-0706.1
- 1087 Roy, A. (1962). Ambiguity in geophysical interpretation. *GEOPHYSICS* 27, 90–99. doi:10.1190/1.  
1088 1438985
- 1089 Salomao, G. N., Dall’Agnol, R., Angelica, R. S., Figueiredo, M. A., Sahoo, P. K., Medeiros-Filho, C. A.,  
1090 et al. (2019). Geochemical mapping and estimation of background concentrations in soils of Carajás  
1091 mineral province, eastern Amazonian Craton, Brazil. *Geochemistry: Exploration, Environment, Analysis*  
1092 19, 431–447
- 1093 Santos, J. O. S., Hartmann, L. A., Gaudette, H. E., Groves, D. I., Mcnaughton, M. J., and Fletcher, I. R.  
1094 (2000). A new understanding of the provinces of the Amazon Craton based on integration of field  
1095 mapping and U-Pb and Sm-Nd geochronology. *Gondwana Research* 3, 453–488
- 1096 Silva, J. B. C. (1986). Reduction to the pole as an inverse problem and its application to low-latitude  
1097 anomalies. *GEOPHYSICS* 51, 369–382. doi:10.1190/1.1442096
- 1098 Siqueira, F., Oliveira Jr., V. C., and Barbosa, V. C. F. (2017). Fast iterative equivalent-layer technique for  
1099 gravity data processing: A method grounded on excess mass constraint. *GEOPHYSICS* 82, G57–G69.  
1100 doi:10.1190/GEO2016-0332.1
- 1101 Skilling, J. and Bryan, R. (1984). Maximum entropy image reconstruction-general algorithm. *Monthly  
1102 Notices of the Royal Astronomical Society*, Vol. 211, NO. 1, P. 111, 1984 211, 111
- 1103 Soler, S. R. and Uieda, L. (2021). Gradient-boosted equivalent sources. *Geophysical Journal International*  
1104 227, 1768–1783. doi:10.1093/gji/ggab297
- 1105 Takahashi, D., Oliveira Jr., V. C., and Barbosa, V. C. (2022). Convolutional equivalent layer for magnetic  
1106 data processing. *Geophysics* 87, 1–59
- 1107 Takahashi, D., Oliveira Jr., V. C., and Barbosa, V. C. F. (2020). Convolutional equivalent layer for gravity  
1108 data processing. *GEOPHYSICS* 85, G129–G141. doi:10.1190/geo2019-0826.1
- 1109 Tassinari, C. C. and Macambira, M. J. (1999). Geochronological provinces of the Amazonian Craton.  
1110 *Episodes* 22, 174–182
- 1111 Teixeira, W., Tassinari, C., Cordani, U. G., and Kawashita, K. (1989). A review of the geochronology of  
1112 the Amazonian Craton: Tectonic implications. *Precambrian Research* 42, 213–227
- 1113 van der Sluis, A. and van der Vorst, H. A. (1987). Numerical solution of large, sparse linear algebraic  
1114 systems arising from tomographic problems. In *Seismic tomography with applications in global  
1115 seismology and exploration geophysics*, ed. G. Nolet (D. Reidel Publishing Company), chap. 3. 49–83
- 1116 Van Loan, C. F. (1992). *Computational Frameworks for the fast Fourier transform*. Frontiers in Applied  
1117 Mathematics (SIAM)
- 1118 Villas, R. N. and Santos, M. (2001). Gold deposits of the Carajás mineral province: deposit types and  
1119 metallogenesis. *Mineralium Deposita* 36, 300–331

- 1120 Xia, J. and Sprowl, D. R. (1991). Correction of topographic distortion in gravity data. *Geophysics* 56,  
1121 537–541
- 1122 Xia, J., Sprowl, D. R., and Adkins-Helgeson, D. (1993). Correction of topographic distortions in potential-  
1123 field data; a fast and accurate approach. *Geophysics* 58, 515–523. doi:10.1190/1.1443434
- 1124 Zhao, G., Chen, B., Chen, L., Liu, J., and Ren, Z. (2018). High-accuracy 3D Fourier forward modeling  
1125 of gravity field based on the Gauss-FFT technique. *Journal of Applied Geophysics* 150, 294–303.  
1126 doi:10.1016/j.jappgeo.2018.01.002
- 1127 Zidarov, D. (1965). Solution of some inverse problems of applied geophysics. *Geophysical Prospecting*  
1128 13, 240–246. doi:10.1111/j.1365-2478.1965.tb01932.x

## 12 ALGORITHMS

---

**Algorithm 1:** Generic pseudo-code for the CGLS applied to the overdetermined problem (equation 22) for the particular case in which  $\mathbf{H} = \mathbf{I}_P$  (equation 9 and subsection 3.2),  $\mu = 0$  (equation 11),  $\mathbf{W}_d = \mathbf{I}_D$  (equation 12) and  $\tilde{\mathbf{p}} = \mathbf{0}$  (equation 14), where  $\mathbf{I}_P$  and  $\mathbf{I}_D$  are the identities of order  $P$  and  $D$ , respectively.

---

**Initialization :**

```

1 Compute  $\mathbf{G}$ ;
2 Set  $\mathbf{r} = \mathbf{d}$  and compute  $\delta = \|\mathbf{r}\|/D$  ;
3 Compute  $\vartheta = \mathbf{G}^\top \mathbf{r}$  and  $\rho_0 = \vartheta^\top \vartheta$  ;
4 Set  $\tilde{\mathbf{p}} = \mathbf{0}$ ,  $\tau = 0$  and  $\eta = \mathbf{0}$  ;
5  $m = 1$  ;
6 while ( $\delta > \epsilon$ ) and ( $m < \text{ITMAX}$ ) do
7   Update  $\eta \leftarrow \vartheta + \tau \eta$  ;
8   Compute  $\nu = \mathbf{G} \eta$  ;
9   Compute  $v = \rho_0 / (\nu^\top \nu)$  ;
10  Update  $\tilde{\mathbf{p}} \leftarrow \tilde{\mathbf{p}} + v \eta$  ;
11  Update  $\mathbf{r} \leftarrow \mathbf{r} - v \nu$  and  $\delta \leftarrow \|v \nu\|/D$  ;
12  Compute  $\vartheta = \mathbf{G}^\top \mathbf{r}$  and  $\rho = \vartheta^\top \vartheta$  ;
13  Compute  $\tau = \rho / \rho_0$  ;
14  Update  $\rho_0 \leftarrow \rho$  ;
15   $m \leftarrow m + 1$  ;
16 end

```

---



---

**Algorithm 2:** Generic pseudo-code for the method proposed by Leão and Silva (1989).

---

**Initialization :**

```

1 Set the indices  $\mathbf{i}^m$  for each data window,  $m \in \{1 : M\}$  ;
2 Set the indices  $\mathbf{j}^m$  for each source window,  $m \in \{1 : M\}$  ;
3 Set the constant depth  $z_0 + \Delta z_0$  for all equivalent sources ;
4 Compute the vector  $\mathbf{a}'$  associated with the desired potential-field transformation ;
5 Compute the matrix  $\mathbf{G}'$  ;
6 Compute the matrix  $\mathbf{B}'$  (equation 34) ;
7 Compute the vector  $(\mathbf{a}')^\top \mathbf{B}'$  ;
8  $m = 1$  ;
9 while  $m < M$  do
10   | Compute  $t_c^m$  (equation 33) ;
11   |  $m \leftarrow m + 1$  ;
12 end

```

---

---

**Algorithm 3:** Generic pseudo-code for the method proposed by Soler and Uieda (2021).**Initialization :**

```

1 Set the indices  $\mathbf{i}^m$  for each data window,  $m \in \{1 : M\}$  ;
2 Set the indices  $\mathbf{j}^m$  for each source window,  $m \in \{1 : M\}$  ;
3 Set the depth of all equivalent sources ;
4 Set a  $D \times 1$  residuals vector  $\mathbf{r} = \mathbf{d}$  ;
5 Set a  $P \times 1$  vector  $\tilde{\mathbf{p}} = \mathbf{0}$  ;
6  $m = 1$  ;
7 while  $m < M$  do
8   | Set the matrix  $\mathbf{W}_d^m$  ;
9   | Compute the matrix  $\mathbf{G}^m$  ;
10  | Compute  $\tilde{\mathbf{p}}^m$  (equation 36) ;
11  |  $\tilde{\mathbf{p}}[\mathbf{j}^m] \leftarrow \tilde{\mathbf{p}}[\mathbf{j}^m] + \tilde{\mathbf{p}}^m$  ;
12  |  $\mathbf{r} \leftarrow \mathbf{r} - \mathbf{G}[:, \mathbf{j}^m] \tilde{\mathbf{p}}^m$  ;
13  |  $m \leftarrow m + 1$  ;
14 end

```

---



---

**Algorithm 4:** Generic pseudo-code for the method proposed by Cordell (1992).**Initialization :**

```

1 Compute a  $D \times 1$  vector  $\Delta \mathbf{z}$  whose  $i$ -th element  $\Delta z_i$  is a vertical distance controlling the depth of
  the  $i$ -th equivalent source,  $i \in \{1 : D\}$  ;
2 Set a tolerance  $\epsilon$  ;
3 Set a maximum number of iterations ITMAX ;
4 Compute  $\mathbf{G}$  (equation 3) ;
5 Compute the scale factor  $\sigma = \mathbf{d}^\top (\mathbf{G} \mathbf{d}) / \mathbf{d}^\top \mathbf{d}$  ;
6 Set a  $D \times 1$  vector  $\tilde{\mathbf{p}} = \sigma \mathbf{d}$  ;
7 Set a  $D \times 1$  residuals vector  $\mathbf{r} = \mathbf{d} - \mathbf{G} \tilde{\mathbf{p}}$  ;
8 Define the maximum absolute value  $r_{\max}$  in  $\mathbf{r}$  ;
9  $m = 1$  ;
10 while ( $r_{\max} > \epsilon$ ) and ( $m < \text{ITMAX}$ ) do
11   | Define the coordinates  $(x_{\max}, y_{\max}, z_{\max})$  and index  $i_{\max}$  of the observation point associated with
      $r_{\max}$  ;
12   |  $\tilde{\mathbf{p}}[i_{\max}] \leftarrow \tilde{\mathbf{p}}[i_{\max}] + (\sigma r_{\max})$  ;
13   |  $\mathbf{r} \leftarrow \mathbf{r} - (\sigma r_{\max}) \mathbf{G}[:, i_{\max}]$  ;
14   | Define the new  $r_{\max}$  in  $\mathbf{r}$  ;
15   |  $m \leftarrow m + 1$  ;
16 end

```

---

---

**Algorithm 5:** Generic pseudo-code for the method proposed by Mendonça and Silva (1994).**Initialization :**

```

1 Set a regular grid of  $P$  equivalent sources at a horizontal plane  $z_0$  ;
2 Set a tolerance  $\epsilon$  ;
3 Set a  $D \times 1$  residuals vector  $\mathbf{r} = \mathbf{d}$  ;
4 Define the maximum absolute value  $r_{\max}$  in  $\mathbf{r}$  ;
5 Define the index  $i_{\max}$  of  $r_{\max}$  ;
6 Define the list of indices  $\mathbf{i}_r$  of the remaining data in  $\mathbf{r}$  ;
7 Define  $\mathbf{d}_e = \mathbf{d}[i_{\max}]$  ;
8 Compute  $(\mathbf{F} + \mu \mathbf{I}_{D_e})$  and  $\mathbf{G}_e$  ;
9 Compute  $\tilde{\mathbf{p}}$  (equation 40) ;
10 Compute  $\mathbf{r} = \mathbf{d}[\mathbf{i}_r] - \mathbf{G}[\mathbf{i}_r, :] \tilde{\mathbf{p}}$  ;
11 Define the maximum absolute value  $r_{\max}$  in  $\mathbf{r}$  ;
12 while ( $r_{\max} > \epsilon$ ) do
13   Define the index  $i_{\max}$  of  $r_{\max}$  ;
14   Define the list of indices  $\mathbf{i}_r$  of the remaining elements in  $\mathbf{r}$  ;
15    $\mathbf{d}_e \leftarrow \begin{bmatrix} \mathbf{d}_e \\ \mathbf{d}[i_{\max}] \end{bmatrix}$  ;
16   Update  $(\mathbf{F} + \mu \mathbf{I}_{D_e})$  and  $\mathbf{G}_e$  ;
17   Update  $\tilde{\mathbf{p}}$  (equation 40) ;
18   Update  $\mathbf{r} = \mathbf{d}[\mathbf{i}_r] - \mathbf{G}[\mathbf{i}_r, :] \tilde{\mathbf{p}}$  ;
19   Define the maximum absolute value  $r_{\max}$  in  $\mathbf{r}$  ;
20 end

```

---



---

**Algorithm 6:** Generic pseudo-code for the iterative method proposed by Siqueira et al. (2017).**Initialization :**

```

1 Set  $P$  equivalent sources on a horizontal plane  $z_0$  ;
2 Set a tolerance  $\epsilon$  ;
3 Set a maximum number of iterations ITMAX ;
4 Compute  $\mathbf{G}$  (equation 3) ;
5 Compute the scale factor  $\sigma = \mathbf{d}^\top (\mathbf{G}\mathbf{d}) / \mathbf{d}^\top \mathbf{d}$  ;
6 Set a  $D \times 1$  vector  $\tilde{\mathbf{p}} = \sigma \mathbf{d}$  ;
7 Compute the  $D \times 1$  residuals vector  $\mathbf{r} = \mathbf{d} - \mathbf{G}\tilde{\mathbf{p}}$  ;
8 Compute  $\delta = \|\mathbf{r}\|/D$  ;
9  $m = 1$  ;
10 while ( $\delta > \epsilon$ ) and ( $m < \text{ITMAX}$ ) do
11   Compute  $\Delta\mathbf{p} = \sigma \mathbf{r}$  ;
12   Update  $\tilde{\mathbf{p}} \leftarrow \tilde{\mathbf{p}} + \Delta\mathbf{p}$  ;
13   Compute  $\nu = \mathbf{G} \Delta\mathbf{p}$  ;
14   Update  $\mathbf{r} \leftarrow \mathbf{r} - \nu$  ;
15   Compute  $\delta = \|\nu\|/D$  ;
16    $m \leftarrow m + 1$  ;
17 end

```

---

---

**Algorithm 7:** Generic pseudo-code for the convolutional equivalent-layer method proposed by Takahashi et al. (2020, 2022).

---

**Initialization :**

- 1 Set the regular grid of  $P$  equivalent sources on a horizontal plane  $z_0$  ;
  - 2 Set a tolerance  $\epsilon$  and a maximum number of iterations ITMAX ;
  - 3 Compute the first column  $\mathbf{G}[:, 1]$  and row  $\mathbf{G}[1, :]$  of the sensitivity matrix  $\mathbf{G}$  (equation 3) for the particular case in which it has a BTTB structure (equation 51);
  - 4 Rearrange the elements of  $\mathbf{G}[:, 1]$  into matrix  $\mathcal{C}$ , compute its 2D-DFT via 2D-FFT and multiply by  $\sqrt{4D}$  to obtain a matrix  $\mathcal{L}'$  (equation 66);
  - 5 Rearrange the elements of  $\mathbf{G}[1, :]$  into matrix  $\mathcal{C}$ , compute its 2D-DFT via 2D-FFT and multiply by  $\sqrt{4D}$  to obtain a matrix  $\mathcal{L}''$  (equation 66);
  - 6 Set  $\tilde{\mathbf{p}} = \mathbf{0}$  ;
  - 7 Set  $\mathbf{r} = \mathbf{d}$  and compute  $\delta = \|\mathbf{r}\|/D$  ;
  - 8 Compute  $\vartheta = \mathbf{G}^\top \mathbf{r}$  (Algorithm 8) and  $\rho_0 = \vartheta^\top \vartheta$  ;
  - 9 Set  $\tau = 0$  and  $\eta = \mathbf{0}$  ;
  - 10  $m = 1$  ;
  - 11 **while** ( $\delta > \epsilon$ ) **and** ( $m < \text{ITMAX}$ ) **do**
  - 12     Update  $\eta \leftarrow \vartheta + \tau \eta$  ;
  - 13     Compute  $\nu = \mathbf{G} \eta$  (Algorithm 8);
  - 14     Compute  $v = \rho_0 / (\nu^\top \nu)$  ;
  - 15     Update  $\tilde{\mathbf{p}} \leftarrow \tilde{\mathbf{p}} + v \eta$  ;
  - 16     Update  $\mathbf{r} \leftarrow \mathbf{r} - v \nu$  and  $\delta \leftarrow \|v \nu\|/D$  ;
  - 17     Compute  $\vartheta = \mathbf{G}^\top \mathbf{r}$  (Algorithm 8) and  $\rho = \vartheta^\top \vartheta$  ;
  - 18     Compute  $\tau = \rho / \rho_0$  ;
  - 19     Update  $\rho_0 \leftarrow \rho$  ;
  - 20      $m \leftarrow m + 1$  ;
  - 21 **end**
- 

---

**Algorithm 8:** Pseudo-code for computing the generic matrix-vector products given by equations 52 and 53 via fast 2D discrete convolution for a given vector  $\mathbf{v}$  (equation 54) and matrix  $\mathcal{L}$  (equation 66).

---

- 1 Rearrange the elements of  $\mathbf{v}$  (equations 52 and 54) into the matrix  $\mathcal{V}_c$  (equation 65);
  - 2 Compute  $\mathcal{F}_{2N_B} \mathcal{V}_c \mathcal{F}_{2N_b}$  via 2D-FFT;
  - 3 Compute the Hadamard product with matrix  $\mathcal{L}$  (equation 66);
  - 4 Compute 2D-IDFT via 2D-FFT to obtain matrix  $\mathcal{W}_c$  (65);
  - 5 Retrieve  $\mathbf{w}$  (equations 52 and 54) from  $\mathbf{w}_c$  (equations 55–57);
-

## 13 TABLES

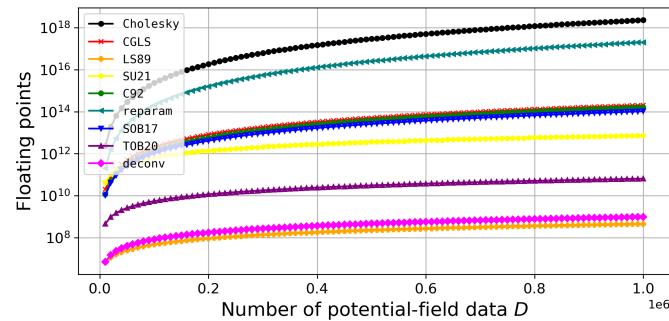
Reference	Term	flops
eq. 10	$\mathbf{G} \mathbf{H}$	$2DQP$
eq. 15	$\mathbf{H} \tilde{\mathbf{q}}$	$2PQ$
eq. 22	$(\mathbf{G} \mathbf{H})^\top (\mathbf{G} \mathbf{H})$	$2Q^2D$
eq. 22	$(\mathbf{G} \mathbf{H})^\top \boldsymbol{\delta}_d$	$2QD$
eq. 23	$(\mathbf{G} \mathbf{H}) (\mathbf{G} \mathbf{H})^\top$	$2D^2Q$
eq. 23	$(\mathbf{G} \mathbf{H})^\top \mathbf{u}$	$2QD$
eq. 25	lower triangle of $\mathcal{G}$	$D^3/3$ or $Q^3/3$
eq. 26	solve triangular systems	$2D^2$ or $2Q^2$
Alg. 1	$\boldsymbol{\eta} \leftarrow \boldsymbol{\vartheta} + \tau \boldsymbol{\eta}$	$2Q$
Alg. 1	$\boldsymbol{\vartheta}^\top \boldsymbol{\vartheta}$	$2Q$
Alg. 4	scale factor $\sigma$	$2DP + 4D$

**Table 1.** Total number of flops associated with some useful terms according to Golub and Van Loan (2013, p. 12). The number of flops associated with equations 25 and 26 depends if the problem is over or underdetermined. Note that  $P = Q$  for the case in which  $\mathbf{H} = \mathbf{I}_P$  (subsection 3.2). The term  $\boldsymbol{\eta} \leftarrow \boldsymbol{\vartheta} + \tau \boldsymbol{\eta}$  is a vector update called *saxpy* (Golub and Van Loan, 2013, p. 4). The terms defined here are references to compute the total number of flops throughout the manuscript.

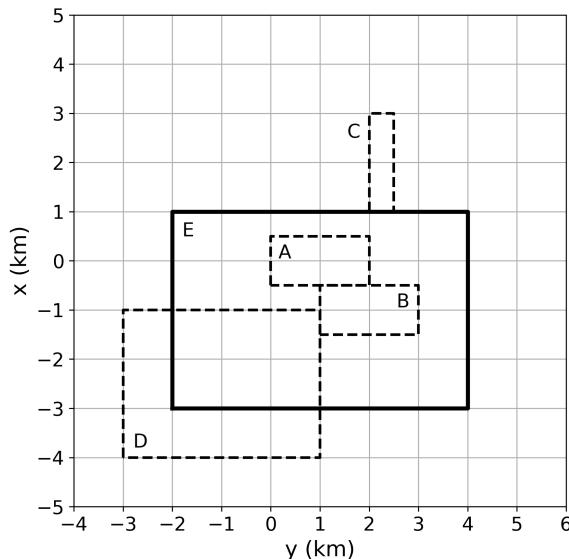
Computational strategies	Characteristics	Advantages	Disadvantages	articles
Moving data-window scheme	A single and small sensitivity submatrix for all moving windows	One of the fastest strategies	Regularly spaced grids of sources and data	Leão and Silva (1989)
Moving data-window scheme	Multiple and small sensitivity submatrices, one for each moving	Irregularly spaced grids of sources and data	Computational speed is reduced	Soler and Uieda (2021)
Column-action updates	A single equivalent source is used, iteratively	A single column of the sensitivity matrix is calculated	Issues related to convergence	Cordell (1992) Guspí and Novara (2009)
Row-action updates	Equivalent data concept	A subset of rows of the sensitivity matrix is calculated	Increasing the order of the linear system of equations, iteratively	Mendonça and Silva (1994)
Reparametrization of the original parameters	Reduction the dimension of the linear system of equations	Lower-dimensional linear system of equations	Undesirable smoothing effect	Oliveira Jr. et al. (2013) Mendonça (2020)
Sparsity induction of the sensitivity matrix	Sparse representation of the original dense sensitivity matrix	Fast iteration of the CG algorithm	Requires computing the full and dense sensitivity matrix	Li and Oldenburg (2010) Barnes and Lumley (2011)
Iterative methods using the full sensitivity matrix	The equivalent layer is updated, iteratively	Fast iterations	Requires computing the full and dense sensitivity matrix	Xia and Sprowl (1991) Xia et al. (1993) Siqueira et al. (2017) Jirigalatu and Ebbing (2019)
Iterative deconvolution	Block-Toeplitz Toeplitz-block (BTTB) matrices concept	One of the fastest strategies	Regularly spaced grids of sources and data	Takahashi et al. (2020) Takahashi et al. (2022)
Direct deconvolution	BTTB matrices concept	One of the fastest strategies	Solution instability	

**Table 2.** Computational strategies to overcome the intensive computational cost of the equivalent-layer technique for processing potential-field data and the corresponding articles.

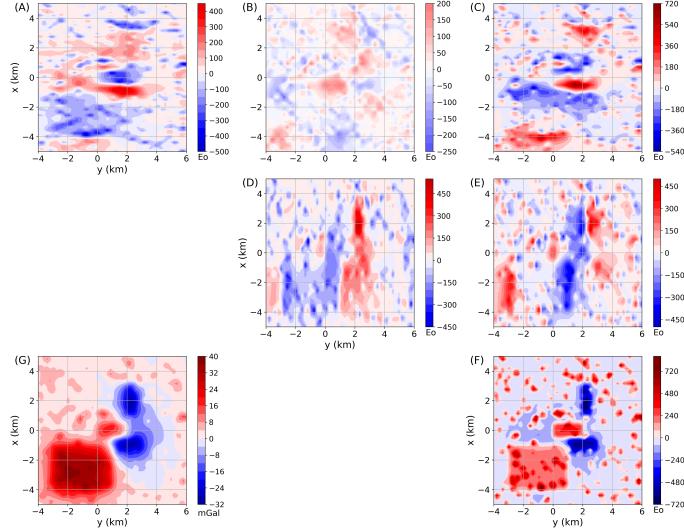
## 14 FIGURES



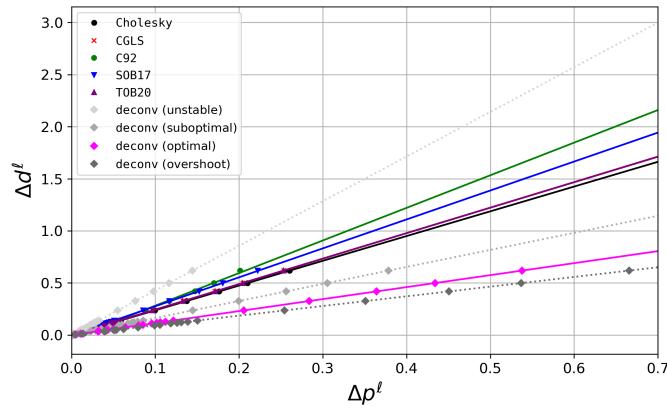
**Figure 1.** Total number of flops for different equivalent-layer methods (equations 27, 28, 35, 37, 38, 43, 50, 67, and 70). The number of potential-field data  $D$  varies from 10,000 to 1,000,000.



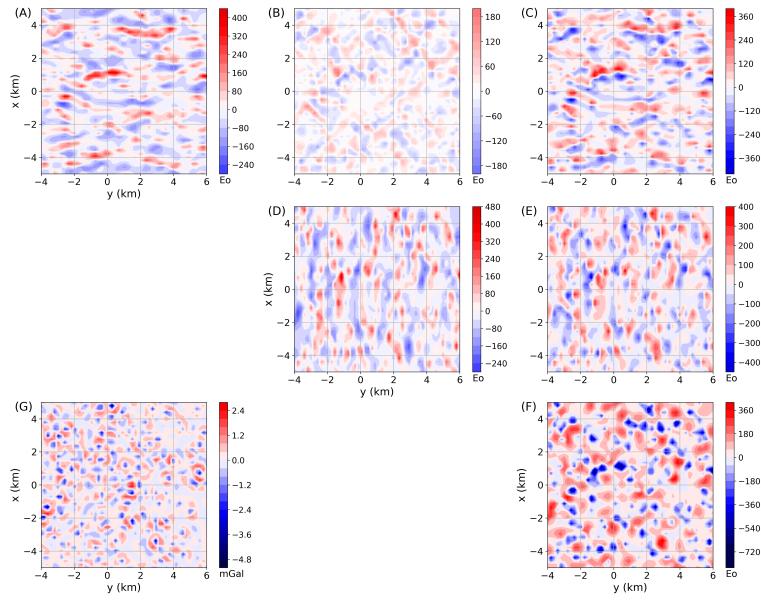
**Figure 2.** Synthetic prisms used in numerical simulations. The prisms A to D with horizontal projections represented by dashed lines have density contrasts of, respectively, 1500,  $-1800$ ,  $-3000$  and  $1200 \text{ kg/m}^3$ , tops varying from 10 to 100 m, bottom from 1010 to 1500 m and side lengths varying from 1000 to 4000 m. The prism E with horizontal projection represented by solid lines has a density contrast  $-900 \text{ kg/m}^3$ , top at 1000 m, bottom at 1500 m and side lengths of 4000 and 6000 m. Our model also have 300 additional small cubes (not shown), with top at 0 m and side lengths defined according to a pseudo-random variable having uniform distribution from 100 to 200 m. Their density contrasts vary randomly from 1000 to  $2000 \text{ kg/m}^3$ .



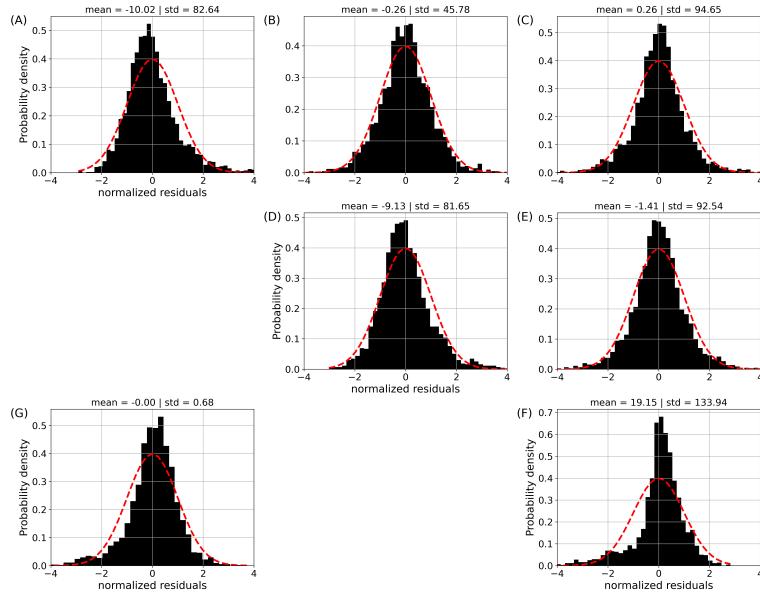
**Figure 3.** Noise-free gravity data produced by an ensemble of rectangular prisms (Figure 2). The data are located on a regular grid of  $50 \times 50$  points. Panels (A)–(F) show, respectively, the  $xx$ ,  $xy$ ,  $xz$ ,  $yy$ ,  $yz$  and  $zz$  component of the gravity-gradient tensor in Eötvös (E). Panel (G) shows the gravity disturbance in milligals (mGal).



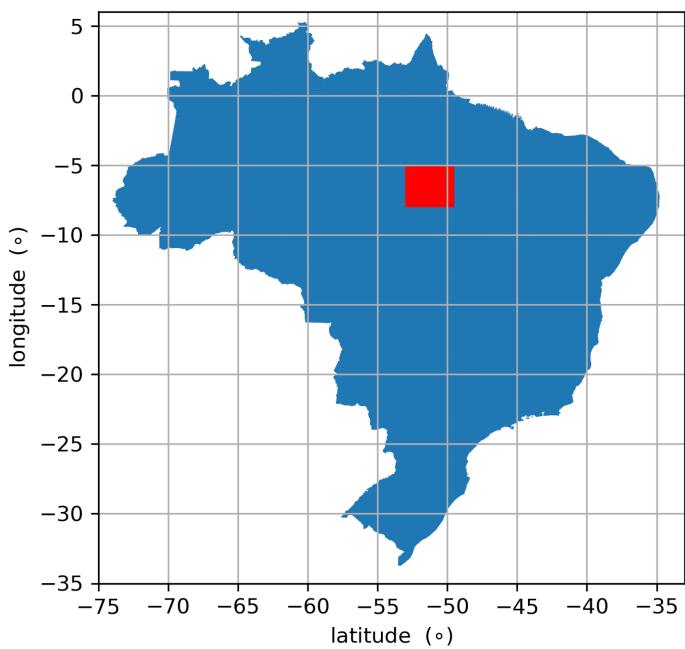
**Figure 4.** Numerical stability curves obtained for the 21 synthetic gravity data sets by using the Cholesky factorization with  $\mu \approx 2 \times 10^{-2}$ ; column-action update (C92) with 25,000 iterations ( $10 \times$  the number  $D$  of potential-field data); CGLS, iterative method (SOB17) and iterative deconvolution (T0B20) with 50 iterations each (Algorithms 1, 6 and 7); and the direct deconvolution (deconv.) computed with four different values for  $\zeta$  (equation 69): 0,  $10^{-18}$  (overshoot),  $10^{-22}$  (optimal) and  $10^{-28}$  (suboptimal). The stability parameter  $\kappa$  (equation 29) obtained for the eight curves described above are 2.37 (Cholesky); 2.45 (CGLS); 3.13 (C92); 2.78 (SOB17); 2.44 (T0B20); 4.28, 1.63, 1.15 and 0.93 (deconv. with null, suboptimal, optimal and overshoot  $\zeta$ ).



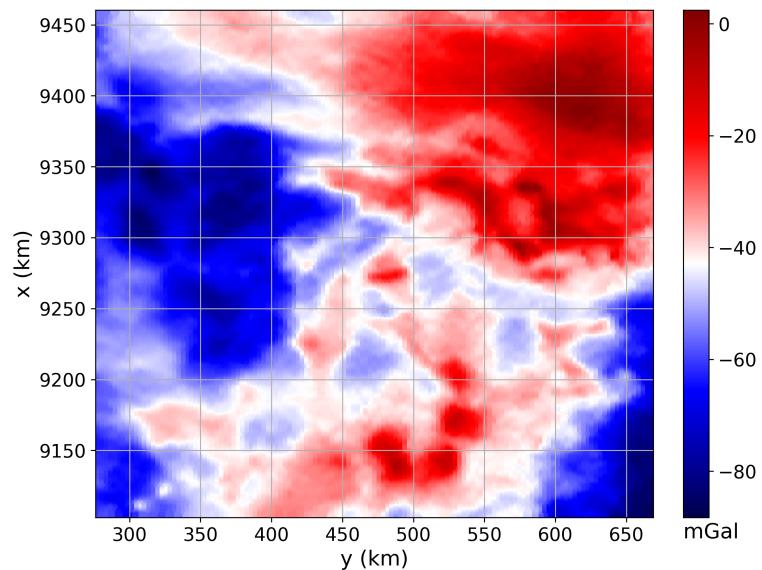
**Figure 5.** Residuals between the gravity data predicted by the equivalent layer estimated with the iterative deconvolution (TOB20) (Algorithm 7). The inverse problems was solved by using the noise-corrupted gravity disturbance having the maximum noise level (not shown). Panels (A)–(F) show the residuals between the predicted and noise-free gravity gradient data (Figure 3) associated with the  $xx$ ,  $xy$ ,  $xz$ ,  $yy$ ,  $yz$  and  $zz$  components of the gravity-gradient tensor, respectively. The values are in Eötvös. (G) Shows the residuals between the predicted and noise-corrupted gravity disturbances. The values are in milligals (mGal).



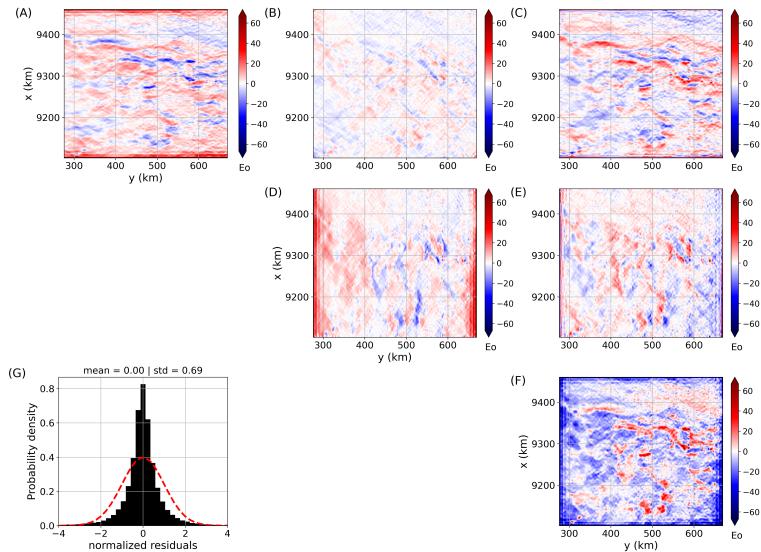
**Figure 6.** Histograms of the residuals shown in Figure 5. The residuals were normalized by removing the mean and dividing the difference by the standard deviation. Panels (A)–(F) show the histograms associated with the  $xx$ ,  $xy$ ,  $xz$ ,  $yy$ ,  $yz$  and  $zz$  components of the gravity-gradient tensor, respectively. (G) Shows the histogram of the residuals between the predicted and noise-corrupted gravity disturbances.



**Figure 7.** Location of the Carajás Mineral Province (CMP), Brazil. The coordinates are referred to the WGS-84 datum. The study area (shown in red) is located at the UTM zone 22S.



**Figure 8.** Field aerogravimetric data over Carajás, Brazil. There are  $D = 500,000$  observations located on regular grid of  $1,000 \times 500$  points.



**Figure 9.** Estimated gravity-gradient tensor components over Carajás, Brazil. Panels (A)–(F) show, respectively, the  $xx$ ,  $xy$ ,  $xz$ ,  $yy$ ,  $yz$  and  $zz$  components of the gravity-gradient tensor in Eötvös. Panel (G) shows the histogram of the residuals between predicted data (not shown) and field data (Figure 8). The residuals were normalized by removing the mean and dividing the difference by the standard deviation. The results were generated by applying the iterative deconvolution (TOB20) (Algorithm 7) with 50 iterations.