



3-D gravity-inversion using graph theory to delineate the framework of homogeneous sources

Rodrigo Bijani¹, Cosme F. Ponte-Neto¹, Dionisio U. Carlos², ¹Observatório Nacional (ON/MCTI), ²VALE/SA, Brasil

Copyright 2014, SBGf - Sociedade Brasileira de Geofísica.

Este texto foi preparado para a apresentação no VI Simpósio Brasileiro de Geofísica, Porto Alegre, 14 a 16 de outubro de 2014. Seu conteúdo foi revisado pelo Comitê Técnico do VI SimBGf, mas não necessariamente representa a opinião da SBGf ou de seus associados. É proibida a reprodução total ou parcial deste material para propósitos comerciais sem prévia autorização da SBGf.

Abstract

We present a new 3-D gravity-inversion to estimate the total-mass, the horizontal and vertical positions, in Cartesian coordinates, of an ensemble of point masses. To stabilize the solutions, we invoke the Graph theory and then solve the Minimum spanning tree (MST) problem. With this, we create a severe spatial restriction to the 3-D distribution of point masses in subsurface. The idea behind this methodology is aid other 3-D inverse methods with promising starting models by using a Heuristic method. Our method consists of a Genetic Algorithm with Elitism (EGA) to find the best solution in front of a set of possible solutions. This Heuristic method is a searching method that works well fronting multi-modal and non-linear functions. The inversion results applied to noise-corrupted synthetic gravity data show that, regardless the starting model, the estimated distribution of point masses offers valuable information about the skeletal of 3-D homogeneous sources in subsurface. Several tests with different values of regularizing parameter were made in order to bespeak the new heuristic regularizing function.

Introduction

Specifically, 3-D gravity-inversions infer the geometry or the density-contrast distribution of geological sources in the subsurface (Barbosa and Silva, 2011). Jointly with the noise in the data, the inherent non-uniqueness that exists in most of 3-D gravity-inversions promote severe difficulties in those estimation-problems (Jackson, 1972). As a result, the obtained models are ambiguous and/or unstable. Commonly, these inverse problems are solved by deterministic methods. These techniques present good convergence only when a promising starting model is achieved (Goldberg and Holland, 1988). Confronting this drawback are the controlled random search optimization methods (Montana, 1994; McCormack et al., 1999). These random process, also known as Heuristic methods, simulate natural phenomena algorithmically, like Simulated Annealing (SA), Ant Colony Optimization (ACO) and Genetic Algorithm (GA). The latter method starts from a random population and progressively modifies the set of solutions by simulating the evolutionary behaviour of biological systems, until an acceptable result is achieved (Parker, 1999). These algorithms have the capability of finding global minima in functions with multiple local minima

regardless starting models (Smith et al., 1992).

In this paper, we focus our efforts on a method that aids other 3-D gravity-inversions with good starting models. To deal with instability presented in most of inverse methods, we apply a new stabilizing function to a 3-D gravity-inversion to retrieve the skeleton of geologic sources. The stabilizing function, called equidistant function, is based on the Minimum Spanning Tree (MST) problem (Held and Karp, 1971; Deo, 1974; Graham and Hell, 1985; Gross and Yellen, 2005) to equalize the distance among neighbouring point masses. We couple the equidistant function to our genetic algorithm with elitism (EGA) to solve the gravity-inversion for two noise-corrupted synthetic data sets. For both synthetic tests, we run the inversion three times with different values of regularizing parameter (i.e., one without regularization, one with the right regularization and one over-regularized). With these examples, we show both a satisfactory description of the ambiguity inherent to the 3-D gravity-inversion and the influence of the equidistant function into the solutions.

The gravity inverse problem

Let \mathbf{g}^{obs} be a N-dimensional vector containing the vertical component of the gravitational field measured on N equally-spaced observation-points. This gravitational response is produced by a 3-D buried source confined beneath the earth's surface. Let \mathbf{p} be a 3-M vector with the 3-D Cartesian coordinates of M point masses and m is one value of mass for all point masses.

$$\mathbf{p} = (x_1, y_1, z_1, \dots, x_M, y_M, z_M) \quad (1)$$

Bunching m and \mathbf{p} , we create a 3-M + 1 parameter-vector \mathbf{q} illustrated as follows:

$$\mathbf{q} = (m, \mathbf{p}) \quad (2)$$

The formulation of the inverse problem consists of finding a 3-D spatial distribution of point masses (i.e., a vector \mathbf{q}) that fits the observed Gravity anomaly in a least squares sense. Mathematically, we minimize the following objective function:

$$\Gamma(\mathbf{q}) = \phi(\mathbf{q}) + \lambda \theta(\mathbf{p}) \quad (3)$$

where the real-scalar function $\phi(\mathbf{q})$ is the data-misfit function, given by

$$\phi(\mathbf{q}) = \|\mathbf{g}^{obs} - \mathbf{g}(\mathbf{q})\|^2 \quad (4)$$

where $\|\cdot\|^2$ is the Euclidean Norm. The function $\phi(\mathbf{q})$ computes the difference between measured and predicted

data at the same measurement points. The function $\theta(\mathbf{p})$, also known as stabilizing function, imposes physical and/or geologic attributes on the solution. Instead of using traditional Tikhonov regularizations (Backus and Gilbert, 1967; Levine, 1979), we develop a new one called the equidistant function. This geometric function (i.e., depends only on vector \mathbf{p}) imposes equidistant among the nearest point masses. A more detailed view about the equidistant function will be discussed next section. The real positive number λ , also known as the regularizing parameter, balances the relative importance between the data-misfit function and the equidistant function.

The equidistant function

A particular strategy, and the one which we propose here, is use the graph theory concepts. A graph is an ensemble of nodes, or vertices, that can be connected by weighted edges (Deo, 1974; Gross and Yellen, 2005; Hoory et al., 2006). With this in mind, we associate the set of M point masses with the vertices of an undirected graph. The edges of this graph are weighted by the distances among the point masses. This formulation allows us to solve a traditional problem in graph theory: The Minimum Spanning Tree (MST) problem (Held and Karp, 1971; Graham and Hell, 1985; Singh, 2009). The MST problem is to find a least cost spanning tree connecting all nodes in an edge weighted graph (Zhou and Gen, 1999). There are several algorithms to deal with the MST problem. In our methodology, we implement the Kruskal's algorithm (Kruskal, 1956; Cormen, 2001). Kruskal's algorithm finds a minimum spanning tree for a connected weighted graph. It finds a subset of the edges that forms a tree that includes every vertex, where the total weight of all the edges in the tree is minimized. The computational time for Kruskal's algorithm depends logarithmically on the number of edges of the graph. The configuration of point masses should respect the MST concept. This is not enough to produce stable solutions. So, we apply a penalty for the edges of the MST. This penalty imposes that the edges of a MST should have the same length. In summary, this additional attribute stabilizes the solutions by producing a homogeneous spatial distribution of point masses. Mathematically, the equidistant function can be defined as follows:

$$\theta(\mathbf{p}) = \sum_{i=1}^{M-1} [d_{ij}(\mathbf{p}) - d^*(\mathbf{p})]^2, \quad j = 1, \dots, M \quad (5)$$

and

$$d_{ij}(\mathbf{p}) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \quad (6)$$

where $d_{ij}(\mathbf{p})$ is the edge connecting the i -th and j -th adjacent point masses, M is the number of point masses, $d^*(\mathbf{p})$ is the average distance among the edges of a MST and \mathbf{p} is the parameter-vector containing the coordinates of the M point masses. This regularizing method needs no matrix operations and offers a restricted degree-of-freedom to the set of point masses. It is easy to implement and it is useful to every discrete Potential-field inverse problem.

Genetic Algorithm with Elitism (EGA)

Genetic algorithms consist of a random search algorithm based on the mechanics of natural selection and natural genetics (Goldberg and Holland, 1988; Montana, 1994). Genetic algorithms are widely used in optimization

problems due to its capacity of finding the global minima of multi-modal functions (Holland, 1992). Differently from deterministic methods, the genetic algorithms require not only a starting model, but a set of starting models (i.e., an initial population). The initial population is randomly selected inside a minimum and a maximum mass values. The same is made for the Cartesian coordinates of the point masses. These ranges are referred to as search limits. The search limits are defined by the interpreter. The former plays a crucial role in the convergence of Genetic algorithms. At the end of the inversion, the estimated mass should be as close as possible to the middle of the mass ranges. Otherwise, we should reset the mass ranges, based on how near the estimated mass is from the top or bottom of the mass ranges, and then run the genetic algorithm again. This is important to ensure that the algorithm will guide us to a promising minimum that fits the Gravity data.

The relevant stages in the implementation of the genetic algorithm, such as crossover, mutation and selection of parents are widely discussed in Goldberg and Holland (1988) and Parker (1999). We consider an extra stage in our implementation called elitism. At this stage, a sub-population of n_e individuals, decreasingly ordered in terms of the objective function $\phi(\mathbf{q})$, is replicated to the next generation. The elitism can be considered as a convergence accelerator, because it allows the appearance of a "super-man" at the last generations. Genetic algorithms with this strategy are referred to as genetic algorithms with elitism or EGA (Chakraborty and Chaudhuri, 2003). We implement the EGA to minimize the objective function (Equation 2) for a value of λ .

Synthetic examples

To validate our methodology and better expose the application of the equidistant function, we present two tests with synthetic data simulating gravitational anomalies caused by simple homogeneous sources. In both tests, the observed data was calculated on a regular grid of 41×33 observation points in the x - and y - directions, totaling 1353 equally-spaced data. The gravity data was corrupted with zero-mean Gaussian pseudo-random noise with a standard deviation of 1.0 mGal. A residual histogram is presented in order to complement the data-fit analysis of the best solution obtained by the EGA. The results are good if the residual histogram present a Gaussian behaviour. For both tests, we perform three inversions each. One with no regularization (i.e., $\lambda = 0$), one with optimal value of the regularizing parameter (obtained by the L-Curve method) and an over-regularized test. With this, we can clearly present the powerful of our regularizing method.

Vertical dyke

In this first test, we buried a vertical prism, with total mass of $m = 6800$ kg, elongated at z -direction, with top and base at 0.6 and 4.0 km, respectively. Both x - and y -directions vary from -1.0 km to +1.0 km. With this, we are simulating, in a simplified manner, a gravitational signal produced by a vertical dyke. To perform the inversion, we use $M = 20$ point masses, 100 individuals in the population, 10 individuals replicated in the elitism and 250 generations. We define the ranges for the horizontal coordinates by analysing the amplitude and position of the Gravity anomaly. For this

test, we set, in km, $x \in [-4.0, +4.0]$, $y \in [-6.0, +6.0]$, $z \in [+0.1, +7.0]$. The range for the estimated mass is defined, by guess, an order of magnitude difference between the maximum and the minimum value of mass.

Figure 1 shows the results without stabilization (i.e., $\lambda = 0$). Figure 1A is the perspective view of the vertical prism jointly with the point masses.

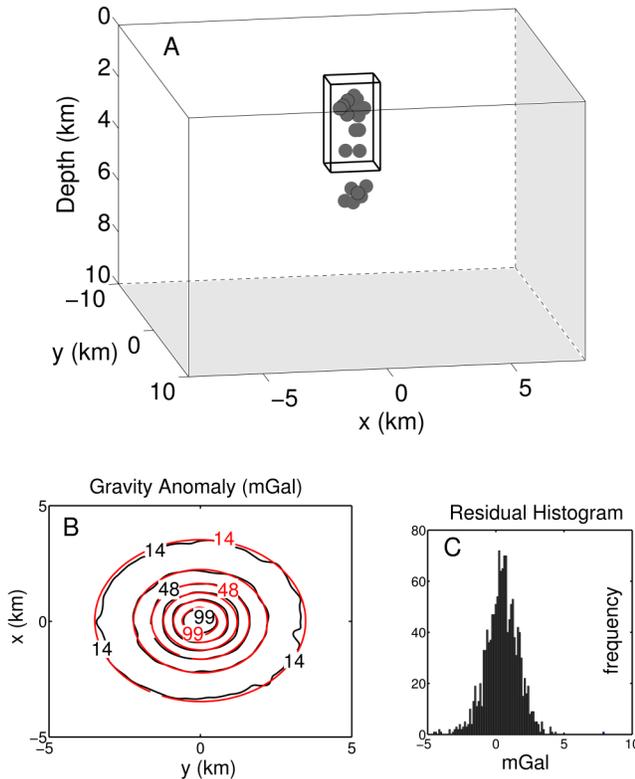


Figure 1: Solution without regularization for the vertical dyke example. In (A), we present the 3-D view of the true source (vertical prism) jointly with the point masses (gray points). (B) The fit between the observed data (black lines) and the predicted data (red lines). The black and red numbers represent the contour intervals. (C) The residual histogram with mean of 0.84 mGal and standard deviation of ± 1.22 mGal.

Clearly, there is a subgroup of point masses disconnected from the uppermost set. The estimated total-mass is 78000 kg, higher than the true 68000 kg. We observe that the deeper point masses do not contribute significantly to the data-fit. Despite a good fit between observed and predicted data (Figures 1B and 1C), we reject this unstable solution. Another reason to discard this solution is that the set of point masses are not homogeneously distributed along the interpretive model.

Figures 2 present the inversion results with the optimal value of the regularizing parameter (i.e., $\lambda = 1.0$). We can see in Figure 2A that the deeper point masses of the former example are now building the skeleton of the prism properly. It's possible to observe that the point masses are surrounding the symmetry axis of the vertical prism. The estimated total-mass is 78892 kg, where the total-mass of the true source is 68000 kg. In Figure 2B, we observe that

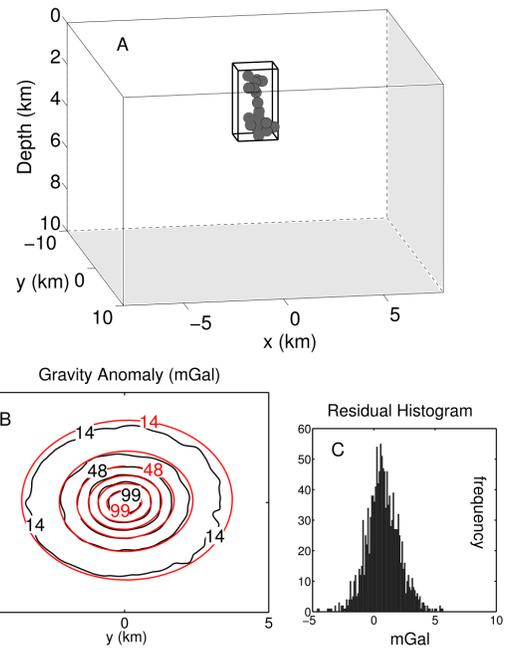


Figure 2: Solution with adequate regularization. (A) The 3-D view of the true source (vertical prism) together with the set of point masses. (B) The fit between observed (black lines) and predicted (red lines) data. The black and red numbers are the contour intervals. (C) The residual histogram with mean of 0.74 mGal and standard deviation of ± 1.34 mGal.

the fit between the observed and the predicted data is also good, corroborated by the residual histogram (Figure 2C).

To better illustrate the powerful of the equidistant function, we perform an inversion with an exaggerated regularizing parameter ($\lambda = 10^5$). Figure 3A shows the over-regularized solution, the point masses are extremely compact inside the prism, which means that the EGA prioritized the minimization of the equidistant function in equation 3. The over-regularized solution creates a compact spatial distribution of point masses that does not fit the true Gravity data. This test presents a large misfit between observed and predicted data (Figures 3B and C). The estimated total-mass is 94420 kg, again higher than the true value of 68000 kg.

Dipping dyke

The second test is a more complex model, where 4 prisms, with total mass of $m = 162000$ kg, are placed in order to form a staircase. With this, we are simulating a gravitational signal produced by a dipping dyke. For this test, we use $M = 40$ point masses, 100 individuals in the population, 10 individuals replicated in the elitism and 350 generations. For this test, we set, in km, $x \in [-5.0, +5.0]$, $y \in [-6.0, +6.0]$, $z \in [+0.1, +6.0]$. We define the ranges for the horizontal coordinates and the mass range by the same criteria of the former example.

Again, Figure 4 present unstable results (i.e., $\lambda = 0$). Figure 4A shows both the true source and the point masses. We observe that the deeper point masses are scattered. This result shows that the solution is

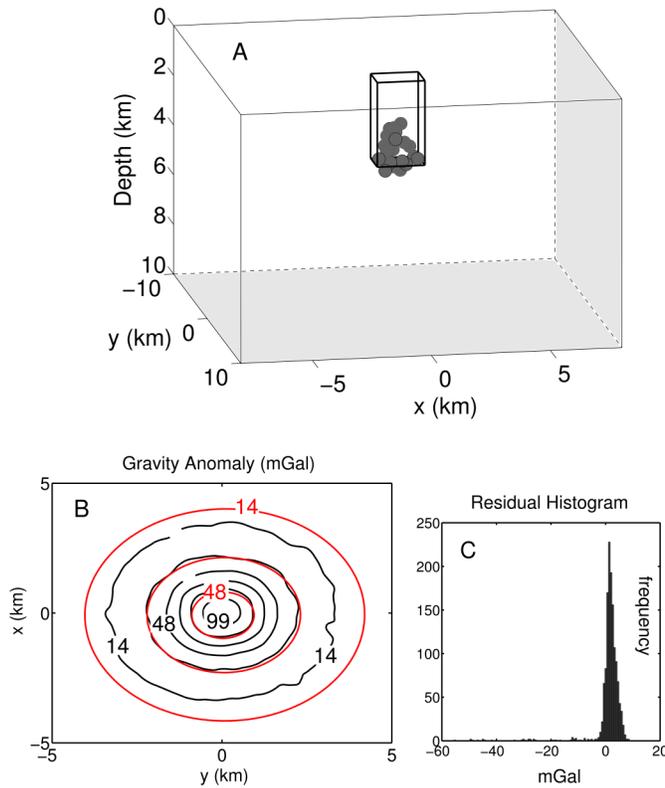


Figure 3: Over-regularized solution. The 3-D view of the true source (vertical prism) and the set of point masses (gray points). (B) The fit between observed (black lines) and predicted (red lines) data. The black and red numbers are the contour intervals. (C) The residual histogram with mean of 1.29 mGal and standard deviation of ± 5.53 mGal.

neither homogeneously distributed along the true source nor stabilized.

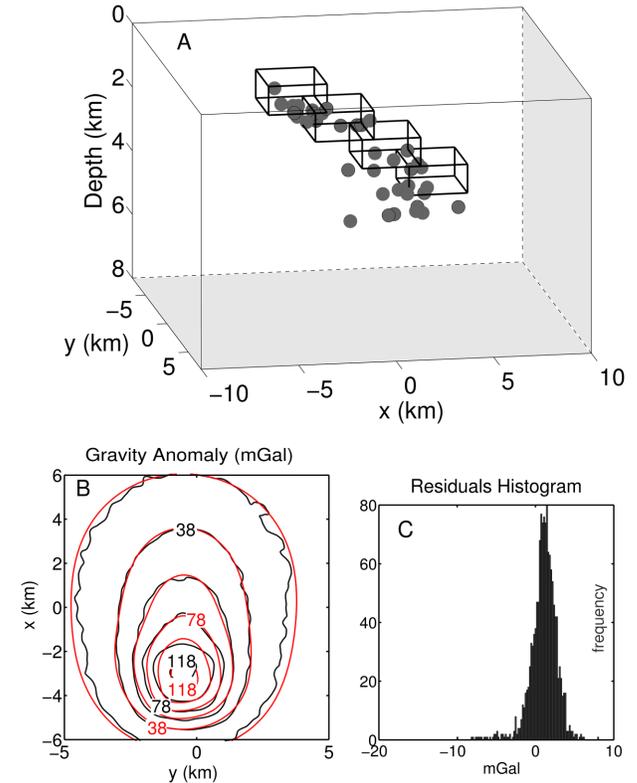


Figure 4: Solution without regularization. The 3-D view of the true source (vertical prism) and the set of point masses (black point masses). (B) The fit between observed (black lines) and predicted (red lines) data. The black and red numbers are the contour intervals. (C) The residual histogram with mean of 1.04 mGal and standard deviation of ± 1.73 mGal.

The estimated mass is 187870 kg, upper than the true mass of 162000 kg. There is a good fit between observed and predicted data (Figures 4B and C). We observe that the point masses are non-homogeneously distributed along the true source (Figure 4A). But as we know, this solution should be stabilized. For so, we re-run EGA with the equidistant function by setting λ as 4.4×10^{-2} .

Figure 5A presents the true source and the solution of the best individual of the 250th generation of the EGA. We now see the point masses more homogeneously distributed along the correct dip of the source, verifying the good performance of our equidistant function. In Figure 5B and C we present an acceptable fit between observed and predicted data. The value of the equidistant function for this test is $\theta(\mathbf{p}) = 2.70$.

Now we perform the inversion with an exaggerated $\lambda = 1.0$. The result (Figures 6A, B and C) show a compact solution without fitting the data. The point masses neither rescue the correct dip of the true source nor delineate properly the framework of the source. The residual histogram (Figure 6) presents a standard deviation of 8.88 mGal, higher than the correct value (1.0 mGal).

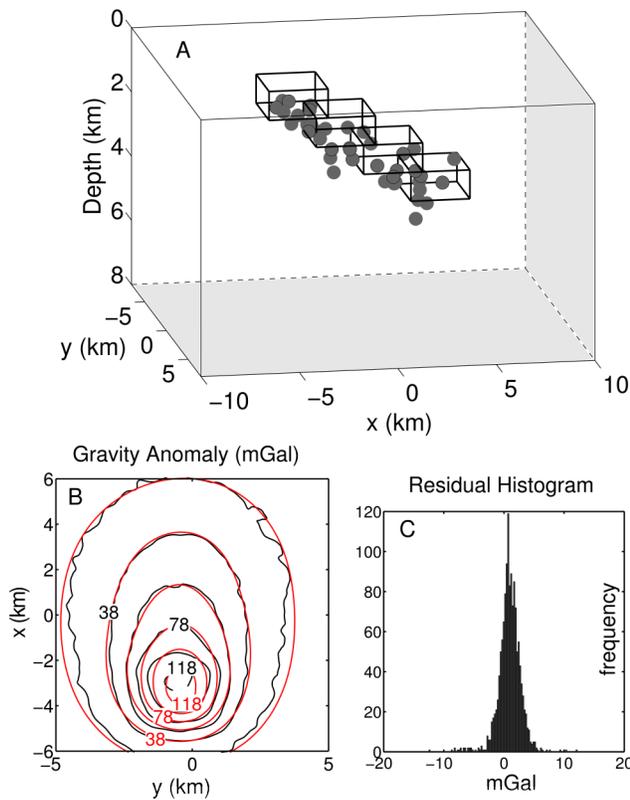


Figure 5: Solution with $\lambda = 0.44 \times 10^{-2}$. The 3-D view of the true source (vertical prism) and the set of point masses (black point masses). (B) The fit between observed (black lines) and predicted (red lines) data. The black and red numbers are the contour intervals. (C) The residual histogram with mean of 0.92 mGal and standard deviation of ± 1.87 mGal.

Conclusions

We have introduced a new 3-D gravity-inversion to delineate the skeleton of homogeneous sources through a set of point masses. This aid other more-complex 3-D inversions with their starting models. To stabilize the solutions, a new regularizing function, called equidistant function, has been added to the inversion. We parametrized the earth's subsurface as a set of simple point masses, all with the same mass. The ensemble of point masses was associated with a graph to solve the minimum spanning tree (MST) problem. This problem restricts the spatial distribution of point masses in order to adjust both predicted and observed Gravity data. We then minimize the equidistant function to equalize the distances among nearest point masses. We can say that the equidistant function worked as a "modelling clay". We consider the equidistant function ideal to work with Heuristic methods. Because of this, a Genetic Algorithm with Elitism (EGA) has been implemented to randomly create a set of initial solutions (i.e., an initial population) and, by following Darwin's evolutionary theory, find a stable and a homogeneous 3-D distribution of point masses. Our method also provided an over-estimation of the total-mass of the geologic source in subsurface. Tests on synthetic data showed that the equidistant function was able to

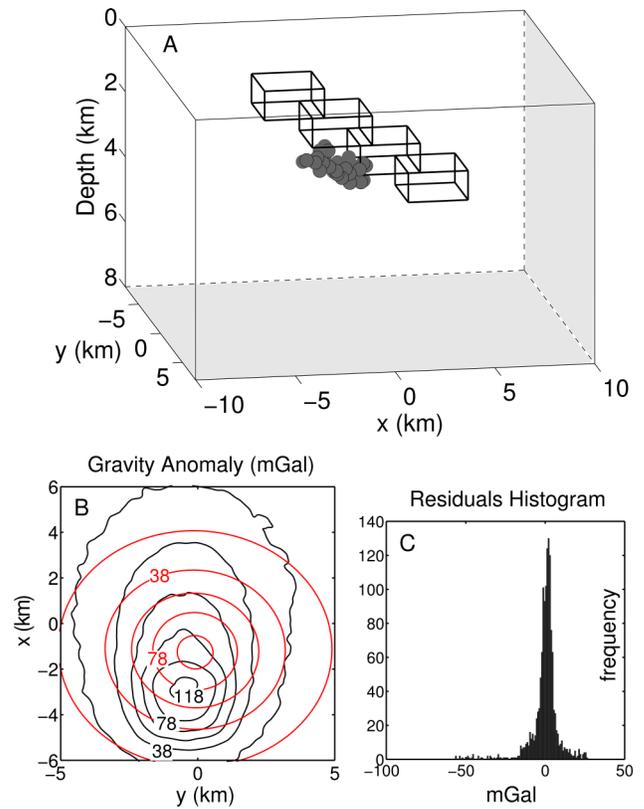


Figure 6: Over-regularized solution with $\lambda = 1.0$. The 3-D view of the true source (vertical prism) and the set of point masses (gray spheres). (B) The fit between observed (black lines) and predicted (red lines) data. The black and red numbers are the contour intervals. (C) The residual histogram with mean of 0.012 mGal and standard deviation of ± 8.88 mGal.

reconstruct elongated sources with correct dip information.

As a future work, we can devote some attention on the Multi-objective optimization approach, where the regularizing parameter will be unnecessary to stabilize the solutions. Other possibility lies in the Hybrid inversion, where we mix both Heuristic and deterministic methods. Our first intention is apply the same methodology to a magnetic inversion, in order to estimate the correct magnetization direction of 3-D homogeneous magnetic sources.

Acknowledgments

We appreciate Constantino de Melo Motta, Carlos Andres Bonilla Quintero, and Leonardo Uieda for valuable contributions during the write of the manuscript. Special thanks for Saulo Siqueira Martins for valuable contributions. First author also thanks CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for the scholarship.

References

- Backus, G. E., and J. F. Gilbert, 1967, Numerical applications of a formalism for geophysical inverse problems: *Geophysical Journal International*, **13**, 247–276.

- Barbosa, V. C. F., and J. B. C. Silva, 2011, Reconstruction of geologic bodies in depth associated with a sedimentary basin using gravity and magnetic data: *Geophysical Prospecting*, **59**, 1021–1034.
- Chakraborty, B., and P. Chaudhuri, 2003, On the use of genetic algorithm with elitism in robust and nonparametric multivariate analysis: **32**, 13–27.
- Cormen, T. H., 2001, *Introduction to algorithms*: MIT Press.
- Deo, N., 1974, *Graph theory with applications to engineering and computer science*: PHI Learning Pvt. Ltd.
- Goldberg, D. E., and J. H. Holland, 1988, Genetic algorithms and: *Machine Learning*, **3**, 95–99.
- Graham, R. L., and P. Hell, 1985, On the history of the minimum spanning tree problem: *IEEE Ann. Hist. Comput.*, **7**, 43–57.
- Gross, J. L., and J. Yellen, 2005, *Graph theory and its applications*, second edition: CRC Press.
- Held, M., and R. M. Karp, 1971, The traveling-salesman problem and minimum spanning trees: Part II: *Mathematical Programming*, **1**, 6–25.
- Holland, J. H., 1992, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*: MIT Press.
- Hoory, S., N. Linial, and A. Wigderson, 2006, Expander graphs and their applications: *Bulletin of the American Mathematical Society*, **43**, 439–561.
- Jackson, D. D., 1972, Interpretation of inaccurate, insufficient and inconsistent data: *Geophysical Journal of the Royal Astronomical Society*, **28**, 97–109.
- Kruskal, Joseph B., J., 1956, On the shortest spanning subtree of a graph and the traveling salesman problem: *Proceedings of the American Mathematical Society*, **7**, 48–50. (ArticleType: research-article / Full publication date: Feb., 1956 / Copyright © 1956 American Mathematical Society).
- Levine, H. A., 1979, Review: A. n. tikhonov and v. y. arsenin, solutions of ill posed problems: *Bulletin (New Series) of the American Mathematical Society*, **1**, 521–524.
- McCormack, M. D., R. F. Stoitsits, D. J. MacAllister, and K. D. Crawford, 1999, Applications of genetic algorithms in exploration and production: *The Leading Edge*, **18**, 716–718.
- Montana, D. J., 1994, Strongly typed genetic programming: *Evolutionary Computation*, **3**, 199–230.
- Parker, P. B., 1999, *Genetic algorithms and their use in geophysical problems*: Ph.D. Thesis, 140.
- Singh, A., 2009, An artificial bee colony algorithm for the leaf-constrained minimum spanning tree problem: *Applied Soft Computing*, **9**, 625–631.
- Smith, M. L., J. A. Scales, and T. L. Fischer, 1992, Global search and genetic algorithms: *The Leading Edge*, **11**, 22–26.
- Zhou, G., and M. Gen, 1999, Genetic algorithm approach on multi-criteria minimum spanning tree problem: *European Journal of Operational Research*, **114**, 141–152.