submitted to *Geophys. J. Int.*

Fast non-linear gravity inversion in spherical coordinates with application to the South American Moho

Leonardo Uieda[1,2], Valéria C. F. Barbosa[2]

[1]*Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil. e-mail: leouieda@gmail.com*

[2]*Observatório Nacional, Rio de Janeiro, Brazil.*

SUMMARY

Key words: Moho; gravity inversion; spherical coordinates; tesseroid; South America

1 INTRODUCTION

2 METHODOLOGY

In potential field methods, we must isolate the target anomalous density distribution prior to modeling and inversion. In our case, the target is the relief of the real Moho undulating around a reference Moho. We do this by removing all other effects from the gravity observations. The first correction is to remove the scalar gravity of an ellipsoidal reference Earth (the Normal Earth), hereafter denoted as γ. This effect is calculated on the same point P where the gravity observation was made (Fig 1a-b). $\gamma(P)$ is calculated using the closed-form solution presented by Li & Götze (2001). The difference between the observed gravity at point P ($g(P)$) and Normal gravity at the same point is known as the gravity disturbance,

$$\delta(P) = g(P) - \gamma(P). \tag{1}$$

The disturbance contains only the gravitational effects of density distributions that are anomalous with respect to the Normal Earth (see Fig. 1c). This includes all masses above the surface of the ellipsoid (the topography), the mass deficiency of the oceans, the mass deficiency of sedimentary basins, crustal sources (e.g., igneous intrusions, lateral density changes, etc), heterogeneities below

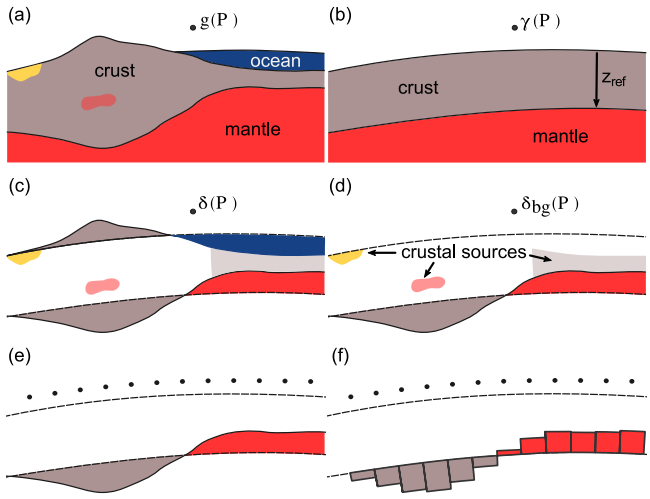

Figure 1. Sketch of the stages in gravity data correction and the discretization of the anomalous Moho relief using tesseroids. (a) The Earth and the measured gravity at point P ($g(P)$). (b) The Normal Earth and the calculated normal gravity at point P ($\gamma(P)$). z_{ref} is the depth of the Normal Earth Moho. (c) The gravity disturbance ($\delta(P)$) and the corresponding density anomalies after removal of the normal gravity: topography, oceans, crustal heterogeneities, and the anomalous Moho. (d) The Bouguer disturbance ($\delta_{bg}(P)$) after topographic correction and the remaining density anomalies. (e) All density anomalies save the anomalous Moho are assumed to have been removed before inversion. (f) The discretization of the anomalous Moho in tesseroids. Grey tesseroids will have a negative density contrast while red tesseroids will have a positive one.

the upper mantle, and the effect of the difference between the real Moho topography and the Moho of the Normal Earth.

In order to invert for the anomalous Moho relief, we must first isolate its gravitational attraction. Thus, all other effects must be either removed or assumed negligible. Here, we will remove the effect of the topography and oceans in order to obtain the full Bouguer disturbance (Fig 1d),

$$\delta_{bg}(P) = \delta(P) - g_{topo}(P). \tag{2}$$

We will remove the effect of sedimentary basins but assume that the effects of other crustal and mantle sources are negligible. Thus, the only effect left will be that of the anomalous Moho relief (Fig 1e). The gravitational attraction of the topography, oceans, and basins are calculated in a spherical Earth approximation by forward modeling using tesseroids (Fig. 2). The tesseroid effects are calculated numerically using Gauss-Legendre Quadrature (GLQ) integration (Asgharzadeh et al. 2007). The accuracy of the GLQ integration is improved by the adaptive discretization scheme of Uieda et al. (2016).

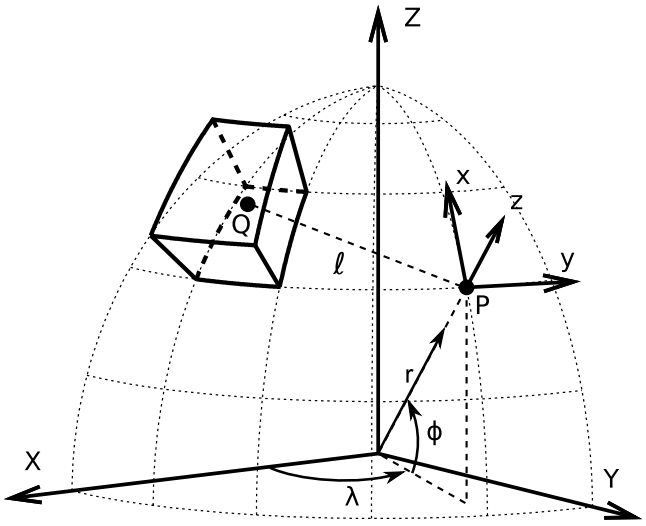

Figure 2. Sketch of a tesseroid (spherical prism) in a geocentric coordinate system (X, Y, Z). Observations are made at point P with respect to it's local North-oriented coordinate system (x, y, z). After Uieda (2015).

2.1 Parametrization

We parameterize the forward problem by discretizing the anomalous Moho into a grid of $M_{lon} \times M_{lat} = M$ juxtaposed tesseroids (Fig 1f). The true (real Earth) Moho varies in depth with respect to the Moho of the Normal Earth. Hereafter we will refer to the depth of the Normal Earth Moho as z_{ref} (see Fig. 1b). In cases where the true Moho is above z_{ref}, the top of the kth tesseroid is the Moho depth z_k, the bottom is z_{ref}, and the density-contrast ($\Delta\rho$) is positive (red tesseroids in Fig 1f). If the Moho is below z_{ref}, the top of the tesseroid is z_{ref}, the bottom is z_k, and $\Delta\rho$ is negative (grey tesseroids in Fig 1f).

Considering that the absolute value of the density-contrasts of the tesseroids is a fixed parameter, the predicted gravity anomaly of the Moho is a non-linear function of the parameters z_k, $k = 1, \ldots, M$,

$$d_i = f_i(\mathbf{p}), \tag{3}$$

in which d_i is the ith element of the N-dimensional predicted data vector $\mathbf{d}$, $\mathbf{p}$ is the M-dimensional parameter vector containing the M Moho depths (z_k), and f_i is the ith non-linear function that maps the parameters onto the data. The functions f_i are the radial component of the gravitational attraction of the tesseroid Moho model.

2.2 Inverse problem

We wish to estimate the parameter vector $\mathbf{p}$ from a set of observed gravity anomaly data $\mathbf{d}^o$. The least-squares estimate is the one that minimizes the data-misfit function

$$\phi(\mathbf{p}) = [\mathbf{d}^o - \mathbf{d}(\mathbf{p})]^T[\mathbf{d}^o - \mathbf{d}(\mathbf{p})]. \tag{4}$$

Function $\phi(\mathbf{p})$ is non-linear with respect to $\mathbf{p}$. Thus, we can determine its minimum using gradient-based iterative optimization methods like Gauss-Newton or Steepest Descent. Such methods start from an initial estimate $\mathbf{p}^0$ and iteratively update the estimate until a minimum is reached.

For the Gauss-Newton method, the update at the kth iteration, $\mathbf{\Delta p} = \mathbf{p}^{k+1} - \mathbf{p}^k$, is the solution of the linear system

$$\mathbf{H}^k\mathbf{\Delta p} = -\nabla\phi^k, \tag{5}$$

in which $\nabla\phi^k$ and $\mathbf{H}^k$ are, respectively, the gradient vector and the Hessian matrix of $\phi(\mathbf{p})$.

The Steepest Descent method uses only the gradient direction to update the initial estimate (Kelley 1987). The update at the kth iteration is achieved by equating the Hessian in Eq. 5 to the identity matrix,

$$\mathbf{\Delta p} = -\nabla\phi^k. \tag{6}$$

Thus, it does not require the computation and storage of the Hessian matrix nor the solution of linear systems. However, the Steepest Descent method has poor convergence when the current solution is close to the minimum of the goal function (Kelley 1987).

The gradient vector and the Gauss-Newton approximation of the Hessian matrix of $\phi(\mathbf{p})$ are, respectively,

$$\nabla\phi^k = -2\mathbf{A}^T[\mathbf{d}^o - \mathbf{d}(\mathbf{p}^k)], \tag{7}$$

and

$$\mathbf{H}^k \approx 2\mathbf{A}^T\mathbf{A}, \tag{8}$$

in which $\mathbf{A}$ is the Jacobian or sensitivity matrix,

$$A_{ij} = \frac{\partial f_i}{\partial p_j}(\mathbf{p^k}). \tag{9}$$

2.3 Regularization

Non-linear inversions for the relief of an interface (like the Moho) are ill-posed and require additional constraints in the form of regularization (Silva et al. 2001). A common approach is to use the first-order Tikhonov regularization to impose smoothness on the solution. The cost function for smoothness regularization is given by

$$\theta(\mathbf{p}) = \mathbf{p}^T \mathbf{R}^T \mathbf{R} \mathbf{p}, \tag{10}$$

where $\mathbf{R}$ is an $L \times M$ finite-difference matrix representing the L first-order differences between adjacent tesseroids.

The solution $\hat{\mathbf{p}}$ to the regularized inverse problem is the one that minimizes the goal function

$$\Gamma(\mathbf{p}) = \phi(\mathbf{p}) + \mu\theta(\mathbf{p}), \tag{11}$$

in which μ is the regularization parameter that controls the balance between fitting the observed data and obeying the smoothness constraint.

The goal function $\Gamma(\mathbf{p})$ is also non-linear with respect to $\mathbf{p}$ and can be minimized using the Gauss-Newton or Steepest Descent methods. The gradient vector and Hessian matrix of the goal function are, respectively,

$$\nabla\boldsymbol{\Gamma}^k = -2\mathbf{A}^T[\mathbf{d}^o - \mathbf{d}(\mathbf{p}^k)] + 2\mu\mathbf{R}^T\mathbf{R}\mathbf{p}^k, \tag{12}$$

and

$$\mathbf{H}^k = 2\mathbf{A}^T\mathbf{A} + 2\mu\mathbf{R}^T\mathbf{R}. \tag{13}$$

The parameter updates for the regularized Gauss-Newton and Steepest Descent methods, respectively, then become

$$\left[\mathbf{A}^T\mathbf{A} + \mu\mathbf{R}^T\mathbf{R}\right]\boldsymbol{\Delta}\mathbf{p} = \mathbf{A}^T[\mathbf{d}^o - \mathbf{d}(\mathbf{p}^k)] - \mu\mathbf{R}^T\mathbf{R}\mathbf{p}^k, \tag{14}$$

and

$$\boldsymbol{\Delta}\mathbf{p} = \mathbf{A}^T[\mathbf{d}^o - \mathbf{d}(\mathbf{p}^k)] - \mu\mathbf{R}^T\mathbf{R}\mathbf{p}^k, \tag{15}$$

Producing the regularized solution using the above equations is computationally costly because of two main factors: (1) the evaluation and storage of the dense $N \times M$ Jacobian matrix $\mathbf{A}$ and (2) the

solution of the resulting $M \times M$ equation system (not required for Steepest Descent). In practice, the derivatives in the Jacobian (Eq. 9) are often calculated through a first-order finite-difference approximation. Thus, evaluating $\mathbf{A}$ requires $2 \times M \times N$ forward modeling operations for each iteration of the gradient descent algorithm. These computations are performed for each iteration of the optimization.

2.4 Bott's method

Bott (1960) developed an efficient method to determined the basement relief of a sedimentary basin from gravity observations. The method requires data on a regular grid of $N_x \times N_y = N$ observations. The basement relief is then discretized into an equal grid of $M_x \times M_y = M$ elements with $M_x = N_x$ and $M_y = N_y$. Bott's iterative method starts with an initial estimate of the basement relief $\mathbf{p}^0$ equal to the null vector and updates the estimate using the formula

$$\boldsymbol{\Delta}\mathbf{p} = \frac{\mathbf{d}^o - \mathbf{d}(\mathbf{p}^k)}{2\pi G \Delta\rho}, \tag{16}$$

in which G is the gravitational constant and $\Delta\rho$ is the basin density contrast. The iterative process stops when the inversion residuals $\mathbf{r} = \mathbf{d}^o - \mathbf{d}(\mathbf{p}^k)$ fall below the assumed noise level of the data.

Silva et al. (2014) showed that Bott's method can be formulated as a special case of the Gauss-Newton method (Eq. 5) by setting the Jacobian matrix (Eq. 9) to

$$\mathbf{A} = 2\pi G \Delta\rho \mathbf{I}, \tag{17}$$

in which $\mathbf{I}$ is the identity matrix. In this framework, Bott's method uses a Bouguer plate approximation of the gravitational effect of the relief, $d_i = 2\pi G \Delta\rho z_i$. The derivative of d_i with respect to the parameter z_i is $2\pi G \Delta\rho$, thus linearizing the Jacobian matrix. However, the non-linearity of the predicted data $\mathbf{d}(\mathbf{p}^k)$ is preserved.

We propose that Bott's method can also be formulated as a special case of the Steepest Descent method (Eq. 6) by setting the Jacobian matrix to

$$\mathbf{A} = \frac{1}{4\pi G \Delta\rho}\mathbf{I}. \tag{18}$$

In practice, both formulations lead to Eq. 16. One of the advantages of Bott's method over the traditional Gauss-Newton or Steepest Descent is eliminating the computation and storage of the dense Jacobian matrix $\mathbf{A}$. Furthermore, Bott's method also does not require the solution of equation systems. However, a disadvantage of Bott's method is that it suffers from instability (Silva et al. 2014). A common approach to counter this issue is to apply a smoothing filter after the inversion to the unstable estimate, as in Silva et al. (2014).

2.5 Regularized Bott's method in spherical coordinates

We propose a regularized version of Bott's method to invert for the relief of the anomalous Moho in spherical coordinates. Our formulation maintains the regularized solutions for Gauss-Newton (Eq. 14) and Steepest Descent (Eq. 15) but replaces the full Jacobian matrix with the Bouguer plate approximations (respectively, Eq. 17 and 18). This linearizes the Jacobian matrix and reduces it to a sparse diagonal matrix, thus eliminating the cost of computing and storing $\mathbf{A}$. Matrix arithmetic operations can be performed efficiently by taking advantage of the sparse nature of matrices $\mathbf{A}$ and $\mathbf{R}$. The same is true for solving the equation system in the Gauss-Newton method (Eq. 14). However, the computational cost of forward modeling is still present. Particularly, forward modeling using tesseroids is more computationally intensive than using right-rectangular prisms because of the numerical integration and adaptive discretization. Benchmarks suggest that forward modeling accounts for approximately 99% of the computation time for a Gauss-Newton inversion (see supplementary material). Hence, this formulation allows us to retain the efficiency of Bott's method while stabilizing the solution through the well established formalism of Tikhonov regularization.

Check this number

2.6 Estimating the regularization parameter

Better reference

The regularization parameter μ controls how much smoothness is applied to the inversion result. An optimal value of μ will stabilize and smooth the solution while not compromising the fit to the observed data. Two widely used methods to estimate an optimal μ are the L-curve criterion and cross-validation (Hansen 1992). Here, we will adopt the hold-out method of cross-validation (Kim 2009). The hold-out method consists of splitting the observed data set into two independent parts: a training set $\mathbf{d}^o_{inv}$ and a testing set $\mathbf{d}^o_{test}$. The training set is used in the inversion while the testing set is kept back and used to judge the quality of the chosen value of μ. For a value of the regularization parameter μ_k, the training set is inverted using μ_k to obtain an estimate $\hat{\mathbf{p}}^k$. This estimate is used to calculate predicted data on the same points as the testing set via forward modeling ($\mathbf{d}^k_{test} = \mathbf{f}(\hat{\mathbf{p}}^k)$). The metric chosen to evaluate μ_k is the mean square error (MSE) of the misfit between the observed and predicted testing data sets,

Not the best reference

$$MSE_k = \frac{\|\mathbf{d}^o_{test} - \mathbf{d}^k_{test}\|^2}{N_{test}}, \tag{19}$$

in which N_{test} is the number of data in the testing set. The optimal value of μ will be the one that minimizes the MSE, i.e. the one that best predicts the testing data. We emphasize that the inversion is performed only on the training data set.

The algorithm for the hold-out cross-validation is summarized as follows:

(i) Divide the observed data into the training ($\mathbf{d}^o_{inv}$) and testing ($\mathbf{d}^o_{test}$) sets.

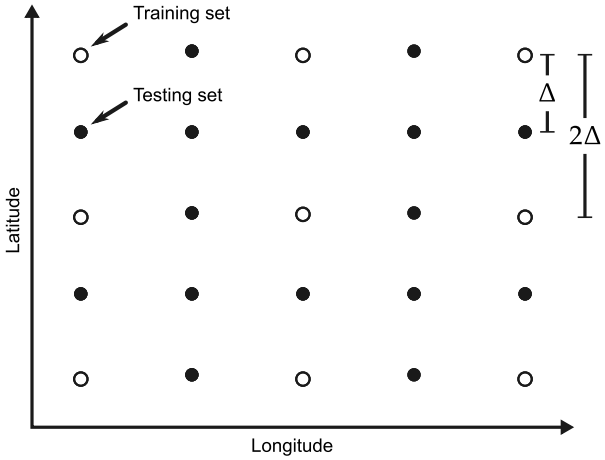

Figure 3. Sketch of a data grid separated into the training (white dots with black outlines) and testing (black dots) data sets. The training data set is still displayed on a regular grid but with twice the grid spacing of the original data grid.

(ii) For each $\mu_k \in [\mu_1, \mu_2, \ldots, \mu_{N_\mu}]$:

(a) Estimate $\hat{\mathbf{p}}^k$ by inverting the training set $\mathbf{d}^o_{inv}$.

(b) Use $\hat{\mathbf{p}}^k$ to calculate the predicted testing set $\mathbf{d}^k_{test}$.

(c) Calculate the mean square error MSE_k using Eq. 19.

(iii) The final solution is the $\hat{\mathbf{p}}^k$ corresponding to the smallest MSE_k.

The separation of the training and testing data sets is commonly done by taking random samples from the full data set. However, we cannot perform the separation in this way because Bott's method requires data on a regular grid as well as having model elements directly below each data point. Thus, we take as our training set the points from the observed data grid that fall on a similar grid but with twice the grid spacing (white dots with black outlines in Fig. 3). All other points from the original data grid make up the testing data set (black dots in Fig. 3). This separation will lead to a testing data set with more points than the training data set. A way to balance this loss of data in the inversion is to generate a data grid with half of the desired grid spacing, either through interpolation or from a spherical harmonic model.

2.7 Estimating z_{ref} and $\Delta\rho$

The depth of the Normal Earth Moho (z_{ref}) and the density-contrast of the anomalous Moho ($\Delta\rho$) are other hyper-parameters of the inversion. That is, their value influences the final solution but they are not estimated during the inversion. Both hyper-parameters cannot be determined from the gravity data alone. Estimating z_{ref} and $\Delta\rho$ requires information that is independent of the gravity data, such

as knowledge of the parameters at certain points. This information can be used in a manner similar to the cross-validation described in the previous section. In this study, we use point estimates of the Moho depth to determine the optimal values of z_{ref} and $\Delta\rho$. These points will generally come from seismologic studies, like receiver functions, surface wave dispersion, and deep refraction experiments.

Let $\mathbf{z}_s^o$ be a vector of N_s known Moho depths. We use the mean square error (MSE) as a measure of how well a given inversion output $\hat{\mathbf{p}}^k$ fits the know depths. The optimal values of z_{ref} and $\Delta\rho$ are the ones that best fit the independent known Moho depths (i.e., produce the smallest MSE). However, the points do not necessarily coincide with the model elements of the inversion. Before computing the MSE, we interpolate $\hat{\mathbf{p}}^k$ on the known points to obtain the predicted depths $\mathbf{z}_s^k$. The MSE is defined as

$$MSE = \frac{\|\mathbf{z}_s^o - \mathbf{z}_s^k\|^2}{N_s}. \tag{20}$$

The algorithm for estimating z_{ref} and $\Delta\rho$ is:

(i) For each combination of $z_{ref,l} \in [z_{ref,1}, z_{ref,2}, \ldots, z_{ref,N_z}]$ and $\Delta\rho_m \in [\Delta\rho_1, \Delta\rho_2, \ldots, \Delta\rho_{N_\rho}]$:

(a) Perform the inversion on the training data set $\mathbf{d}_{inv}^o$ using $z_{ref,l}$, $\Delta\rho_m$, and the previously estimated value of μ. The inversion output is the vector $\hat{\mathbf{p}}^{l,m}$.

(b) Interpolate $\hat{\mathbf{p}}^{l,m}$ on the known points to obtain the predicted depths $\mathbf{z}_s^{l,m}$.

(c) Calculate the MSE between $\mathbf{z}_s^o$ and $\mathbf{z}_s^{l,m}$ using Eq. 20.

(ii) The final solution is the $\hat{\mathbf{p}}^{l,m}$ corresponding to the smallest MSE.

A similar approach was used by Martins et al. (2010) to estimate the parameters defining the density-contrast variation with depth of a sedimentary basin. van der Meijde et al. (2013) also had a similar methodology for dealing with the hyper-parameters, though in a less formalized way.

2.8 Software implementation

The inversion method proposed here is implemented in the Python programming language. The software is freely available under the terms of the BSD 3-clause open-source software license. Our implementation relies on the open-source libraries scipy and numpy (Jones et al. 2001, `http://scipy.org`) for array-based computations, matplotlib (Hunter 2007, `http://matplotlib.org`) and seaborn (Waskom et al. 2015, `http://stanford.edu/~mwaskom/software/seaborn`) for plots and maps, and Fatiando a Terra (Uieda et al. 2013, `http://www.fatiando.org`) for geophysics specific tasks, particularly for forward modeling using tesseroids. We use the scipy.sparse package for sparse matrix arithmetic and linear algebra.

The computational experiments (e.g., data processing, synthetic tests, real data application) were

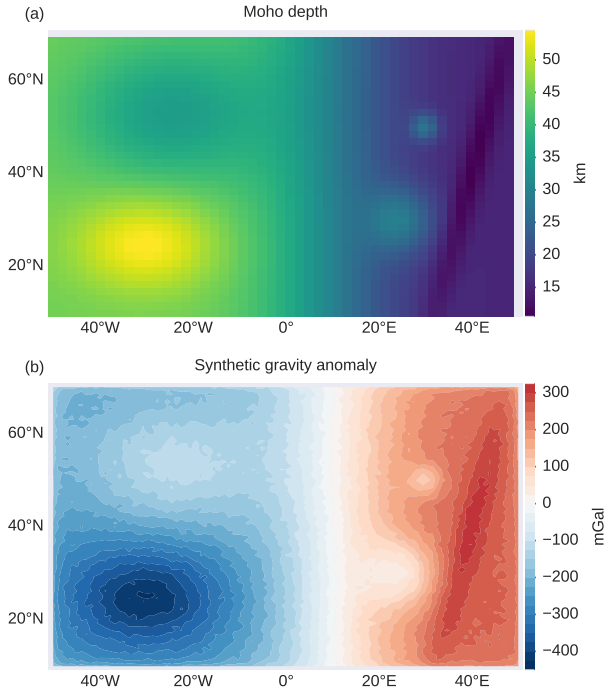

Figure 4. A simple Moho model made of tesseroids for synthetic data application. (a) The Moho depth of the model in kilometers. The model transitions from a deep Moho in the right to a shallow Moho in left, simulating the transition between a continental and an oceanic Moho. Each pixel in the pseudo-color image corresponds to a tesseroid of the model. (b) Noise-corrupted synthetic gravity data generated from the model shown in (a).

performed in Jupyter (formerly IPython) notebooks (Pérez & Granger 2007, http://jupyter.org/). The notebook files combine the source code used to run the experiments, the results and figures generated by the code, and rich text to explain and document the analysis.

All source code, data, and Jupyter notebooks used here can be found at the online repository https://github.com/pinga-lab/paper-moho-inversion-tesseroids. An archived version is also available at http://dx.doi.org/DOI (made available upon publication). The repository also contains instructions for installing the necessary software and reproducing all results presented here.

3 APPLICATION TO SYNTHETIC DATA FROM A SIMPLE MODEL

Bla bla bla.

4 APPLICATION TO SYNTHETIC DATA FROM THE CRUST1.0 MODEL

Bla bla bla.

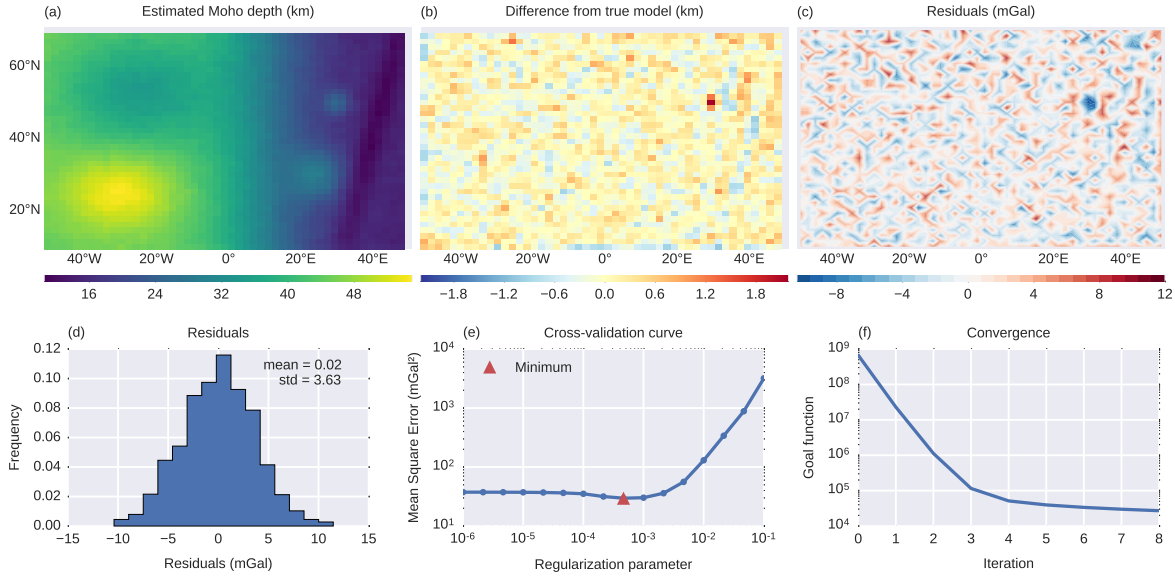

Figure 5. Results from the inversion of the simple synthetic data. (a) The estimated Moho depth. (b) The difference between the true model depths and the estimated depths. (c) The inversion residuals (observed data minus the data predicted by the estimate). (d) Histogram of the residuals. Also shown are the calculated mean and standard deviation (std) of the residuals. Note that the data were contaminated with normally distributed pseudo-random noise with zero mean and 5 mGal standard deviation. (e) Cross-validation curve used to determine the optimal regularization parameter (Eq. 11). The minimum Mean Square Error (Eq. 19) is found at $\mu = 0.00046$ (red triangle). (f) Goal function value (Eq. 11) per Gauss-Newton iteration showing the convergence of the gradient descent.

5 APPLICATION TO THE SOUTH AMERICAN MOHO

5.1 Gravity and seismic data

The gravitational effect of the topography is removed using the ETOPO1 digital terrain model (Amante & Eakins 2009, `http://dx.doi.org/10.7289/V5C8276M`). The effect of sedimentary basins is removed using the CRUST1.0 model (Laske et al. 2013, `http://igppweb.ucsd.edu/~gabi/rem.html`). Lateral variations in density along the Moho cannot be properly accounted for in regions where information coverage is sparse and readily accessible models are not available, like in the South American and African continents. For the purposes of this study, we will assume that all other crustal sources and lateral variations in density are negligible.

5.2 Inversion results

Bla bla bla.

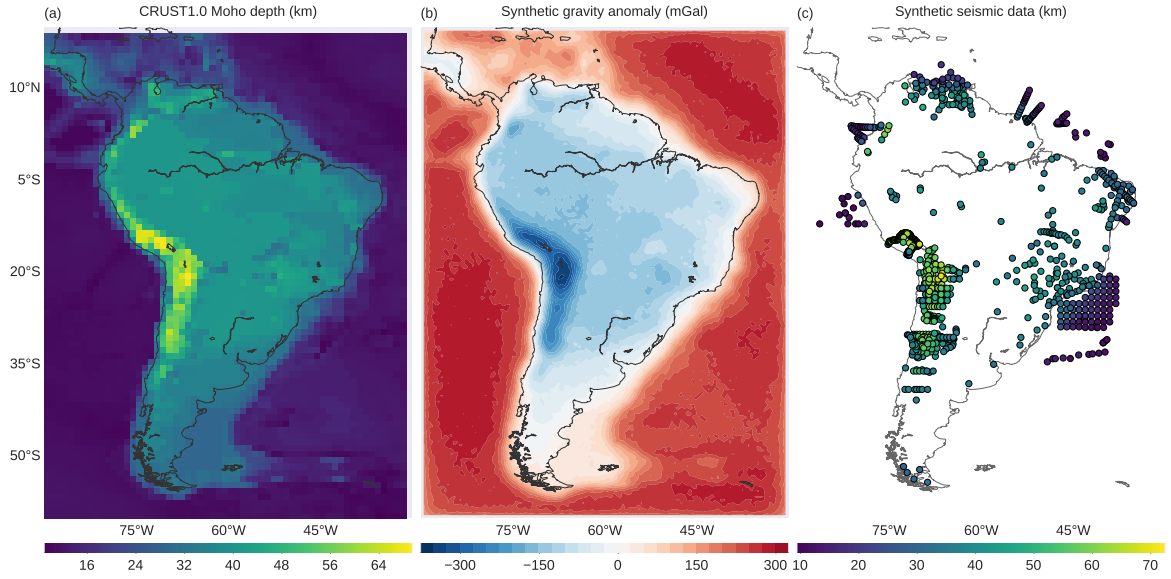

Figure 6. Synthetic data of a model derived from CRUST1.0. The model is made of tesseroids with an constant density-contrast of $\Delta\rho = 350\ kg/m^3$ and assuming a reference level of $z_{ref} = 30\ km$. (a) The Moho depth of the model in kilometers. Each pixel in the pseudo-color image corresponds to a tesseroid of the model. (b) Noise-corrupted synthetic gravity data generated from the model. (c) Synthetic seismic data simulating point estimates of Moho depth. The point estimates were obtained by interpolating the Moho depth in (a).

6 CONCLUSIONS

Meh.

7 ACKNOWLEDGMENTS

The authors are indebted to the developers and maintainers of the open-source software without which this work would not have been possible.

REFERENCES

Amante, C. & Eakins, B. W., 2009. ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis, *NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA.*.

Asgharzadeh, M. F., von Frese, R. R. B., Kim, H. R., Leftwich, T. E., & Kim, J. W., 2007. Spherical prism gravity effects by Gauss-Legendre quadrature integration, *Geophysical Journal International*, **169**(1), 1–11.

Assumpção, M., Feng, M., Tassara, A., & Julià, J., 2012. Models of crustal thickness for South America from seismic refraction, receiver functions and surface wave tomography, *Tectonophysics*.

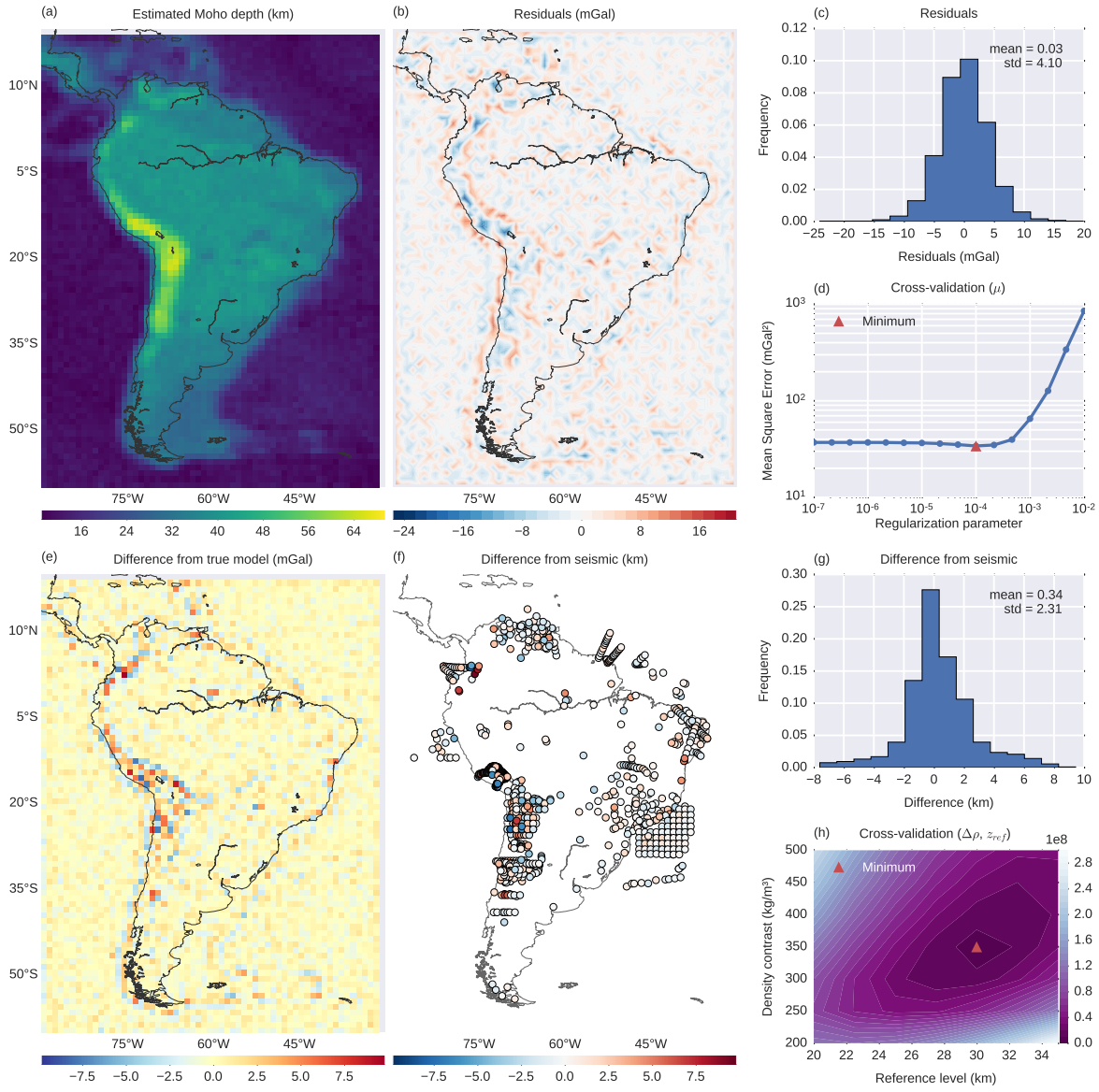

Figure 7. Inversion results from the CRUST1.0 synthetic data. (a) The estimated Moho depth. (b) The inversion residuals (observed minus predicted data). (c) Histogram of the residuals shown in (b). (d) Cross-validation curve used to determine the regularization parameter (Eq. 11). The minimum Mean Square Error (Eq. 19) is found at $\mu = 0.0001$ (red triangle). (e) difference between the CRUST1.0 model depths (Fig. 6a) and the estimated depths. (f) Difference between the synthetic seismic observations (Fig. 6c) and the estimated depths. (g) Histogram of the differences shown in (f). (h) Cross-validation results used to determine the reference level (z_{ref}) and the density-contrast ($\Delta\rho$). The colored contours represent the Mean Square Error (Eq. 20) in km^2. The minimum (red triangle) is found at $z_{ref} = 30\ km$ and $\Delta\rho = 350\ kg/m^3$.

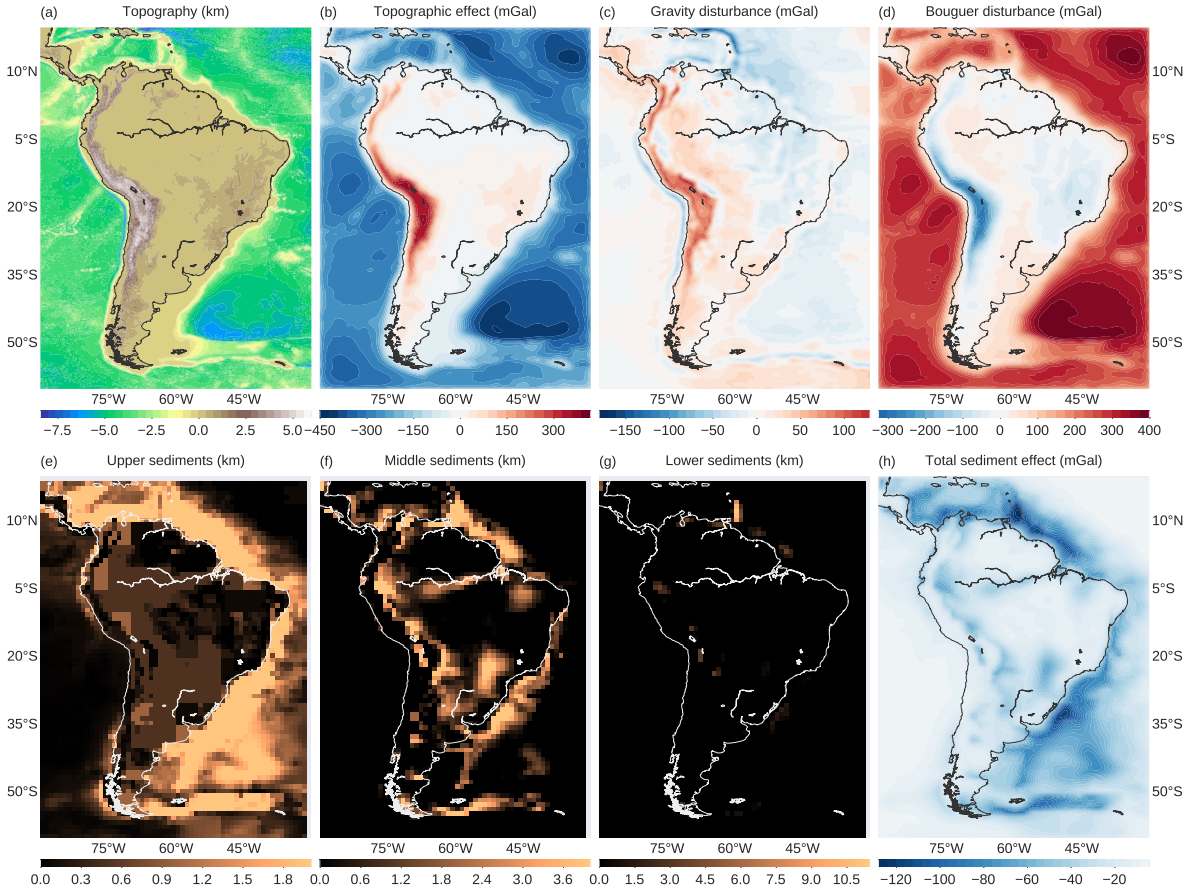

Figure 8. Gravity data for South America and the models used in the data corrections. (a) Topography from ETOPO1 sampled at 0.1° grid spacing. (b) Gravitational attraction of the topography calculated at the observation height using tesseroids. (c) The gravity disturbance (Eq. 1) calculated from the raw gravity data. (d) The Bouguer disturbance (Eq. 2) obtained by subtracting (b) from (c). The upper (e), middle (f), and lower (g) sediment layer thicknesses from the CRUST1.0 model. (h) The total gravitational attraction of the sediment layers shown in (e), (f), and (g), calculated using tesseroids.

Bott, M. H. P., 1960. The use of Rapid Digital Computing Methods for Direct Gravity Interpretation of Sedimentary Basins, *Geophysical Journal International*, **3**(1), 63–67.

Hansen, P., 1992. Analysis of Discrete Ill-Posed Problems by Means of the L-Curve, *SIAM Review*, **34**(4), 561–580.

Hunter, J. D., 2007. Matplotlib: A 2D graphics environment, *Computing in Science & Engineering*, **9**(3), 90–95.

Jones, E., Oliphant, T., Peterson, P., & others, 2001. SciPy: Open source scientific tools for Python, http://www.scipy.org/. Accessed 22-08-2015.

Kelley, C. T., 1987. *Iterative Methods for Optimization*, Society for Industrial and Applied Mathematics, Philadelphia, 1st edn.

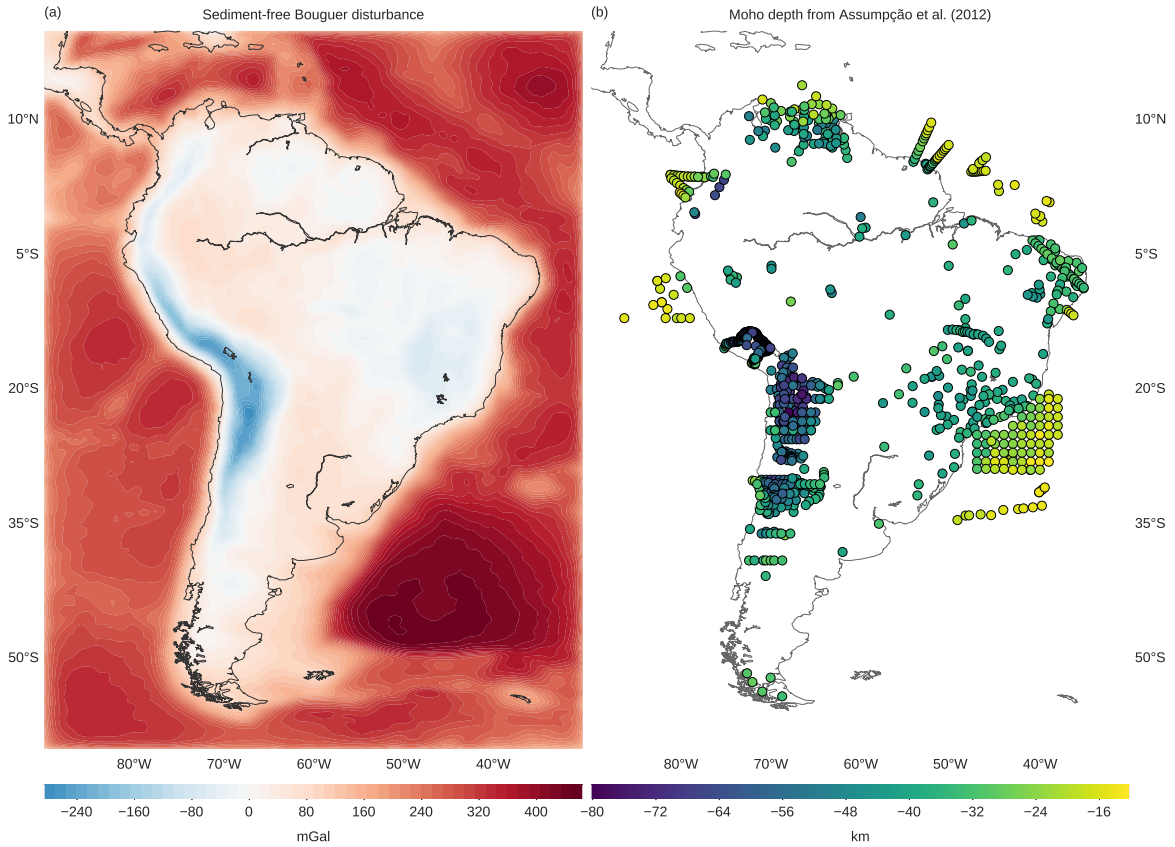

Figure 9. Input data for the South American Moho inversion. (a) Sediment-free Bouguer disturbance for South America. Obtained by subtracting the total sediment gravitational effect (Fig. 8h) from the Bouguer disturbance (Fig. 8d). (b) Seismological Moho depth estimates from Assumpção et al. (2012).

Kim, J.-H., 2009. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap, *Computational Statistics & Data Analysis*, **53**(11), 3735–3745.

Laske, G., Masters, G., Ma, Z., & Pasyanos, M., 2013. Update on CRUST1.0 - A 1-degree Global Model of Earth's Crust, in *Geophysical Research Abstracts*, vol. 15 of **EGU2013-2658**.

Li, X. & Götze, H., 2001. Ellipsoid, geoid, gravity, geodesy, and geophysics, *GEOPHYSICS*, **66**(6), 1660–1668.

Martins, C., Barbosa, V., & Silva, J., 2010. Simultaneous 3D depth-to-basement and density-contrast estimates using gravity data and depth control at few points, *GEOPHYSICS*, **75**(3), I21–I28.

Pérez, F. & Granger, B. E., 2007. IPython: A System for Interactive Scientific Computing, *Computing in Science & Engineering*, **9**(3), 21–29.

Silva, J., Medeiros, W., & Barbosa, V., 2001. Potential-field inversion: Choosing the appropriate technique to solve a geologic problem, *GEOPHYSICS*, **66**(2), 511–520.

Silva, J., Santos, D., & Gomes, K., 2014. Fast gravity inversion of basement relief, *GEOPHYSICS*, pp. G79–G91.

Uieda, L., 2015. A tesserioid (spherical prism) in a geocentric coordinate system with a local-North-oriented

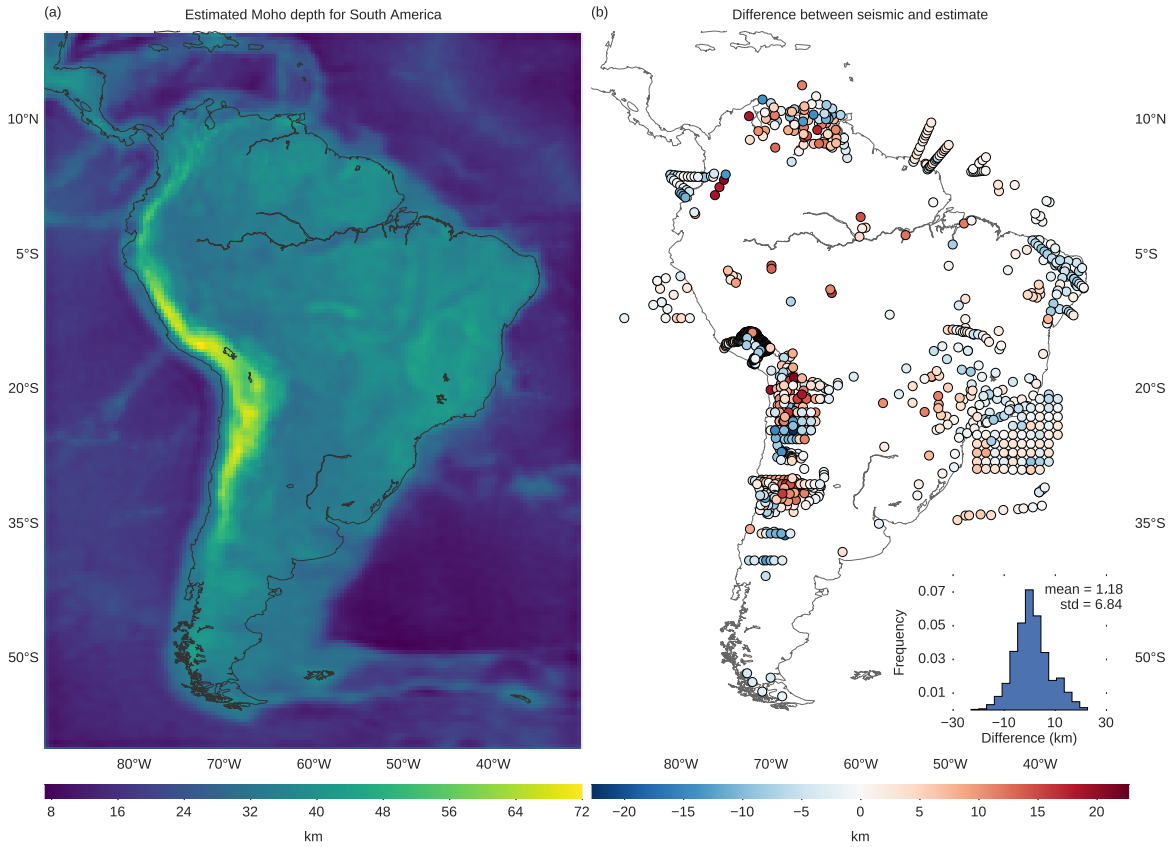

Figure 10. Inversion results for the South American Moho. (a) The estimated Moho depth of South America. (b) Differences between the seismological depths of Assumpção et al. (2012) and our gravity-derived estimate shown in (a). The inset in (b) shows a histogram of the differences along with their calculated mean and standard deviation (std).

coordinate system, *figshare*, http://dx.doi.org/10.6084/m9.figshare.1495525. Accessed 02-11-2015.

Uieda, L., Oliveira Jr, V. C., & Barbosa, V. C. F., 2013. Modeling the Earth with Fatiando a Terra, in *Proceedings of the 12th Python in Science Conference*, pp. 91 – 98.

Uieda, L., Barbosa, V. C. F., & Braitenberg, C., 2016. Tesseroids: forward modeling gravitational fields in spherical coordinates, *Geophysics*, In press (accepted).

van der Meijde, M., Julià, J., & Assumpção, M., 2013. Gravity derived Moho for South America, *Tectonophysics*.

Waskom, M., Botvinnik, O., Hobson, P., Warmenhoven, J., Cole, J. B., Halchenko, Y., Vanderplas, J., Hoyer, S., Villalba, S., Quintero, E., Miles, A., Augspurger, T., Yarkoni, T., Evans, C., Wehner, D., Rocher, L., Megies, T., Coelho, L. P., Ziegler, E., Hoppe, T., Seabold, S., Pascual, S., Cloud, P., Koskinen, M., Hausler, C., kjemmett, Milajevs, D., Qalieh, A., Allan, D., & Meyer, K., 2015. seaborn: v0.6.0 (June 2015), *Zenodo*, http://dx.doi.org/10.5281/zenodo.19108. Accessed 02-11-2015.

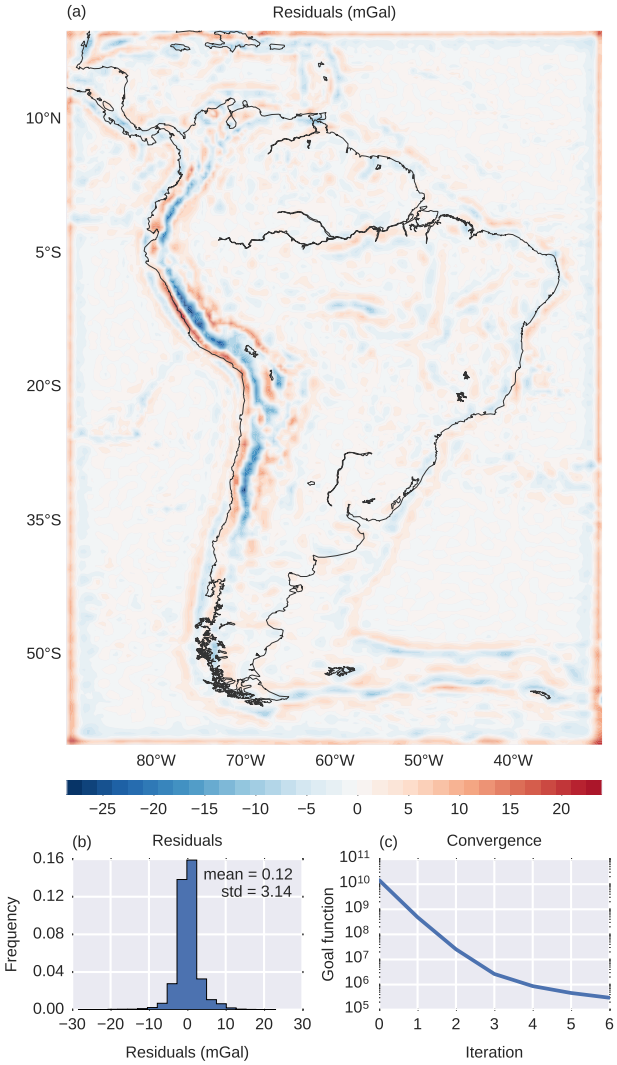

Figure 11. Residuals for the South American Moho inversion. The residuals are the observed data in Fig. 9a minus the data predicted by the estimate in Fig. 10a. Shown as (a) a map and (b) a histogram with the calculated mean and standard deviation. (c) The value of the goal function (Eq. 11) per Gauss-Newton iteration showing the convergence of the algorithm. Note that the y-axis is in logarithmic scale.

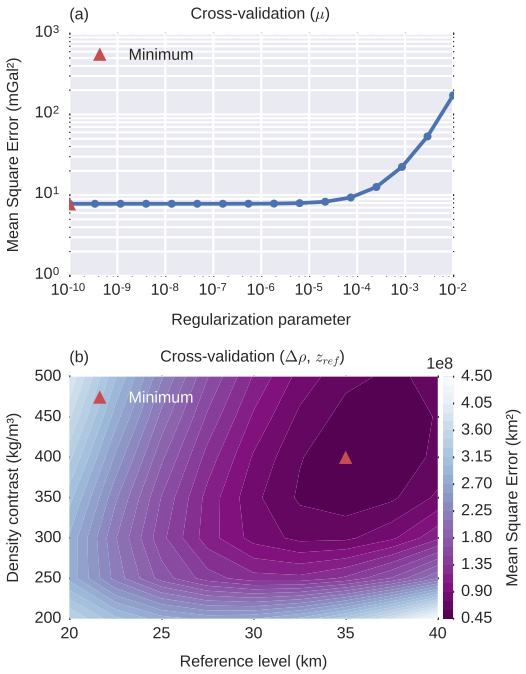

Figure 12. Cross-validation results for the South American Moho inversion. (a) Cross-validation to determine the regularization parameter μ (Eq. 11). The minimum Mean Square Error (Eq. 19), shown as a red triangle, corresponds to $\mu = 10^{-10}$. (b) Cross-validation to determine the reference level (z_{ref}) and the density-contrast ($\Delta\rho$). The colored contours represent the Mean Square Error (Eq. 20). The minimum (red triangle) is found at $z_{ref} = 35\ km$ and $\Delta\rho = 400\ kg/m^3$.