UNIVERSITY OF LIVERPOOL

Estimating the accuracy of Moho depth estimates from gravity inversion
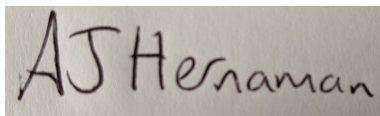
by

Aidan Hernaman

A dissertation submitted to the Department of Earth, Ocean, and Ecological Sciences, University of Liverpool, in partial fulfilment of the requirements for the degree of Master of Science in the subject of Geophysics and Geology.

May, 2021

Declaration

I, Aidan Hernaman, confirm that the work submitted in this dissertation is my own, and that appropriate credit has been given where reference is made to the work of others.

Signature:

Date: Thursday 13th May, 2021

Acknowledgements

First of all, I'm extremely grateful to Leonardo Uieda for being my supervisor for this Masters Project. I would also like to thank Eleanor Dacre and Lisa Palmer for their comments on the paper which helped improve the quality of this thesis.

Abstract

For a long time, models of the Moho discontinuity have been created from a variety of different methods including gravitational and seismological studies. However, very few of these models developed have uncertainty estimates, this is especially the case where there is limited seismic data available. In areas such as South America and Africa due to the economic and environmental challenges, over vast regions of the continent, there are little to no seismic point estimates. For these areas to have relevant Moho models either gravitational data needs to be used or the seismic data has to be interpolated; there is no way to tell how accurate these methods are to attaining a true Moho depth model. To determine a method's accuracy a method of cross-validation specifically repeated random sub-sample validation will be used to quantify errors on a gravitationally derived model of South America with the help of seismic point estimates. The results from this cross-validation will evaluate the accuracy of gravitational models in regions where there is no seismic data to compare it to. Additionally, for regions where the model significantly underestimates the Moho depth in comparison to the seismic data, there is likely an unmodelled mass present. The Paraná Basin, South America is thought to have large igneous intrusions resulting in a shallower Moho than expected. In this study, these intrusions will be modelled in an attempt to decrease the errors on the Moho model currently used. The cross-validation gives an indication that the specific model used has a reasonably small misfit from the point estimates but that the size of the error may vary geographically across South America as the error values attained are an average for the whole model.

Contents

List of Figures

Chapter 1

Introduction

The Mohorovičić discontinuity or Moho for short is the physical boundary signified by the change in many properties such as mineralogy, density, and temperature among other things, but it is mainly known as the change from the crust to the upper mantle. Ever since the discovery of this boundary observed through a significant change in seismic p-wave velocity on either side of the Moho by seismologist Andrija Mohorovičić in 1909 there have been many people using multiple methods to try and quantify the depths to this boundary. Some estimates of the Moho discontinuity include *Laske et al.* (2013), *Assumpção et al.* (2013), and *Reguzzoni and Sampietro* (2015), all of which have slightly different models depending on types of methods and data. Geophysical techniques such as seismology and gravimetry have been utilized to estimate the depth of the discontinuity and topography over a local to a global scale. This issue is known as a geophysical inverse problem and uses data in the form of gravity, seismic, or another method to determine the depth to the Moho over a certain area and produce a model. Why is this important, the determination of the Moho is critical for many reasons. One of the main problems that it helps solve is knowledge about how plates deform and move, it is not the answer to the whole issue, but it is a contributing factor. Especially in regions of continental crusts where their behaviour is somewhat less predictable and understood when compared to oceanic plates. In continental settings, the crustal thickness is indicative of the stress of the lithosphere and helps with the determination through models of inter-plate faulting that often lead to earthquake events. The Moho thickness can also provide an insight into geothermal heat flow helping with heat flux models of the Earth where often without these depths to the crust-mantle interface these values need to be estimated as are unknown. Finally, determination of the thickness of the crust improves understanding of where deeper Earth minerals may be found. Or where specific minerals have been found in the past correlating this to a specific Moho thickness may indicate as to the type of setting to find explore

in the future. In this paper, the moho model will be derived from gravitational data and removing the local gravity disturbances to get the regional field that is almost entirely based on the depth to the Crust-Mantle boundary with the help of other parameters that need to be constrained including the density, reference Moho, and regularisation parameter. The problem with using gravitational data to model the Moho is that gravity values calculated to be the regional field do not only include the effect of the discontinuity but also the effects of unmodelled masses in the Earth's crust that have not been removed when taking away all other effects that contribute to the strength of the raw gravity data. These unmodelled masses lead to the emergence of uncertainties in the model created, however with these uncertainties being of an unknown magnitude then it is difficult to quantify and correct for these unknown masses as the location, size, density, and number are impossible to determine. This is not just the case in the *Uieda and Barbosa* (2016), but almost all gravitationally derived Moho models suffer the same fate including *van der Meijde et al.* (2013), *Tugume et al.* (2013), and *Reguzzoni and Sampietro* (2015). All of these models produced are generally accurate representations of the Moho surface however none of these includes errors or uncertainty values to coincide with their models. Although these error estimates are often not seen in seismologically derived Moho models either, one of the only recognizable papers that try to calculate the uncertainty of their estimates is *Szwillus et al.* (2019) who obtain uncertainties by interpolating Moho depth even then these uncertainties calculated are larger than the errors coming from the P-wave velocity. In this paper, the aim is to find uncertainties in a gravitationally calculated Moho model by using cross-validation with seismic constraints. With the hopes of finding the difference between a model calculated from gravity data and seismic point estimates to see how good the gravitational estimates of the depth to the moho are where there are not any seismic points that the model can constrain to. Continents such as Europe and North America have extensive seismological surveys that span most of the area so it would be considered pointless to see how well this method of uncertainty estimation works as there are no substantial areas of land without seismic data. This is why South America has been chosen as most of the surveyed areas by either reflection or refraction are situated along the coast with few based towards the centre of the continent mainly due to the difficultly of surveying a result of the magnitude and density of vegetation in the Amazon rainforest. However, South America is also limited with seismic data due to lack of financial funding

as surveys over a large scale may not be economically viable for countries or companies that are interested in carrying out one. So with the vast area of the Amazon having little to no seismic points to constrain the gravitationally computed Moho discontinuity model it will be a good test to see how good the model is when there are no seismic point estimates to compare to. Therefore this paper will use the Moho model of South America created by *Uieda and Barbosa* (2016) and implement a cross-validation approach to calculate the uncertainties through the average differences between gravitational estimates and seismic point estimates of the crustal thickness or depth to Moho. The method builds upon the cross-validation originally used in the paper by randomly selecting a training and testing set and calculating not only the uncertainty but how many seismic point constraints are needed to accurately quantify the size of this error. In addition to using cross-validation, in the synthetic models created of South America, the effect of trying to model a previously unknown mass will be tackled by adding in underplating in the Parana basin addressed in *Mariani et al.* (2013). Through this the possibility of locating unmodelled masses could be determined, if the cross-validation of seismic point estimates provides a difference from the gravity model then it is likely that in the area there is a denser/less dense mass in the crust depending on if the difference is positive or negative.

Chapter 2

Data and Methodology

To accurately obtain a model of the Moho depth one must first remove all other effects that contribute to overall gravitational values recorded over an area (see Figure 2.1). This is achieved through removing the scalar gravity of an ellipsoidal reference Earth (the Normal Earth) and then the removal of all other effects. Initially, though the effect of the Normal Earth needs to be removed from the same point as to where the gravity observation was made and is calculated from the closed-form solution in *Li and Götze* (2001). The value obtained here is called the gravity disturbance and can be seen in equation 2.1 below.

$$\delta(\mathbf{P}) = g(\mathbf{P}) - \gamma(\mathbf{P}) \tag{2.1}$$

The gravity disturbance is still not a direct result of the change in density associated with the Moho discontinuity but also an amalgamation of topography with reference to the normal ellipsoid, variations of density in the crust (e.g. sedimentary basins and igneous intrusions), anomalies below the upper mantle, and mass deficiency due to the oceans. The effects of topography and sedimentary basins are removed through a topography correction in equation 2.2.

$$\delta_{\mathbf{bg}}(\mathbf{P}) = \delta(\mathbf{P}) - g_{topo}(\mathbf{P}) \tag{2.2}$$

This is the method used in *Uieda and Barbosa* (2016) and assumes that the effects of other crustal and mantle sources are negligible, and after this correction, the gravitational values attained are purely a result of the density variations on either side of the Moho discontinuity. Seeing as the data is obtained for a sufficiently large area (South America) this correction for topography amongst other things is calculated using tesseroids as part of a spherical Earth approximation. Tesseroids (Figure 2.2) are spherical prisms that are used in place of rectangular prisms as they account for the curvature of the Earth. Effects of the tesseroids are calculated using a GLQ integration presented in *Asgharzadeh et al.* (2007) and improved upon in *Uieda* (2015) through the adaptive discretization scheme.

4
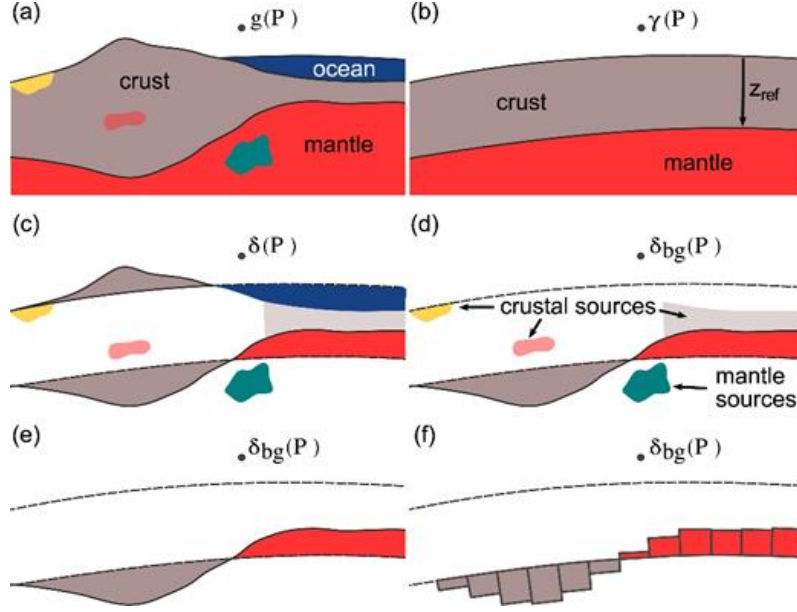
Figure 2.1:   Step by step stages of the removal of gravitational effects. (a) The measured gravity at point P($g(P)$) with reference to the Earth. (b) The Normal Earth and normal gravity at P($\gamma(P)$). (c) Removal of density anomalies e.g. oceans and topography to get ($\delta(P)$) the gravity disturbance. (d) The crust and mantle sources left after obtaining the Bouguer disturbance ($\delta_{bg}(P)$) by removing topography. (e) Assuming there are no unmodelled masses the remaining signal is the Moho and its corresponding depth. (f) Discretization of the anomalous Moho into tesseroids. Grey tesseroids have a negative density contrast while red ones have a positive contrast. After *Uieda and Barbosa* (2016).

2.1   Overview of the methodology from Uieda (2017)

As the cross-validation to estimate uncertainty in the model is implemented as part of the code used in the *Uieda and Barbosa* (2016) paper the method in calculating the model is largely similar. For context, an overview of this method will be given but for more detail see *Uieda and Barbosa* (2016). Upon the calculation of the Bouguer disturbance from the removal of topography, sediments etc. the forward model is parameterized by discretizing the anomalous Moho onto tesseroids. The forward model aims to calculate the difference between the Normal Earth Moho and the true Moho depth, and depending on which is shallower will result in either a positive (red) or negative (grey) density contrast displayed
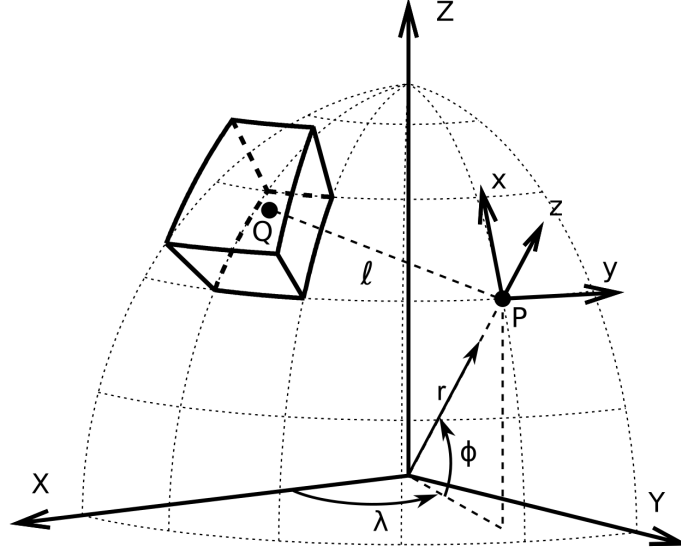
Figure 2.2: Sketch of a tesseroid system with geocentric coordinates (X Y Z). Gravitational observations are made at point P with respect to its local north-orientated coordinate system (x y z). From *Uieda* (2015).

by the colour of the tesseroids, Figure 2.1. The overall absolute value of the density contrast is a certain parameter, this produces a nonlinear problem with the equation 2.3.

$$\mathbf{d} = f(\mathbf{p}) \tag{2.3}$$

Where d is the data vector, p is the parameter vector containing Moho depths, and f is the non-linear function. Leading on to the inverse problem the parameter vector is estimated using least-squares that reduces the misfit to the data, equation 2.4.

$$\phi(\mathbf{p}) = [\mathbf{d}^o - \mathbf{d}(\mathbf{p})]^T [\mathbf{d}^o - \mathbf{d}(\mathbf{p})] \tag{2.4}$$

Where d^o is the observed gravity data, the equation means that this is a non-linear inverse problem, but we can calculate the parameters using optimization, where a perturbation vector Δp^0 is iterated until a minimum is reached which leads to the value of $\phi(p)$. The optimization of the least-squares estimate however is not enough for estimating the relief associated with the Moho and needs regularization in the form of a first-order Tikhonov regularization *Tikhonov and Arsenin* (1977) to ensure smoothness in the model and avoid unstable solutions in Moho depth, to provide a realistic model *Silva et al.* (2001), see

equation 2.5 below.

$$\theta(\mathbf{p}) = \mathbf{p}^T \mathbf{R}^T \mathbf{R} \mathbf{p} \tag{2.5}$$

R a matrix composed of first-order differences between Moho depths. This along with the least-squares estimate leads to an inverse problem that is solved by minimising the goal function, equation 2.6,

$$\mathbf{\Gamma}(\mathbf{p}) = \phi(\mathbf{p}) + \mu\theta(\mathbf{p}) \tag{2.6}$$

μ is the regularization parameter that helps control the fit to the observed data and the smoothness. After the rearrangement and substitution of equations, we arrive at a linear equation system that can calculate the update (Δp) with reference to the Normal Earth Moho,

$$[\mathbf{A^k}^T \mathbf{A}^k + \mu\mathbf{R}^T\mathbf{R}]\mathbf{\Delta p}^k = \mathbf{A^k}^T[\mathbf{d}^o - \mathbf{d}(\mathbf{p}^k)] - \mu\mathbf{R}^T\mathbf{R}\mathbf{p}^k \tag{2.7}$$

Where Ak is the Jacobian matrix, and Δp^k is the parameter perturbation vector. Bott's method *Bott* (1960) calculates the thickness of a sedimentary basin based on gravitational data, the method is iterative so recalculates a new vector of basement depths from the previous calculation until a value where the residuals (equation numerator) fall below the noise level, see equation 2.8.

$$\mathbf{\Delta p}^k = \mathbf{d}^o - \mathbf{d}(\mathbf{p}^k)/2\pi G\mathbf{\Delta}\rho \tag{2.8}$$

Where Δp is the density contrast between the sediment and the reference density, and G is the gravitational constant ($6.67\text{x}10\text{-}11$ $m^{-3}kg^{-1}s^{-2}$). However, *Silva et al.* (2014) showed that Bott's method can be written as,

$$\mathbf{A} = 2\pi G\mathbf{\Delta}\rho\mathbf{I} \tag{2.9}$$

and the main advantage to this is that this method does not need the solution of the equation system, but rather a constant diagonal matrix, A. This scales the model depths to fit the gravitational data. For calculating the depth to the Moho *Uieda and Barbosa* (2016) uses Bott's method in the inversion process and adapts it onto a spherical coordinate system using tesseroids. Stating in this paper that this method retains the efficiency of Bott's method while accounting for the stability problem previously present in the method. Following this step, the final part of the method involves calculating the hyperparameters (regularization parameter μ, Moho density-contrast Δp, and depth of

the Normal Earth Moho z_{ref}) which will be used in the inversion process. The calculation of the regularization parameter is through a method of holdout cross-validation from *Hansen* (1992) and from this optimal regularization value, the other two hyperparameters (Moho density contrast Δp, and depth of the Normal Earth Moho z_{ref}) can be calculated. The main way these parameters are calculated is by finding the smallest Mean Square Error (MSE) through a cross-validation method which compares known Moho depths from seismic point estimates to calculated ones from the hyperparameters and picks the values with the smallest associated MSE.

2.2 Implementation of cross-validation in error estimation

Cross-validation (CV) often used for large data sets in order to see how well the model produced from said data set performs independently (i.e. when there is no data to base the model on). The result of cross-validation is often an MSE or mean square error value which is the accuracy of the new predicted independent model. It is often the goal of the cross-validation in the first place, and how this can be minimised. For many cases of CV, the data set is split up into a training and testing set with the training set being used to find the best solution or model attained from the smallest cross-validation value while the testing (validating) set is kept separate. It is then compared to the model created from the training set to get an idea of the size of errors, and how well a model will perform for a completely independent set. The cross-validation procedure also helps prevent overfitting of the model or selection bias where some points tend to skew the overall model more than others. And in order to minimise these problems the most, the process is repeated multiple times with different training and testing sets along with the variation in the size of these subsets. Many types of CV are relevant for different case-specific things, although mostly the methods are split into two main types: exhaustive and non-exhaustive. Exhaustive CV is where all possible combinations of separating the full data into training and testing sets are used leading to a limited number of iterations that can be run. This method often works best for small volumes of data as with larger sets the computational time becomes uneconomical and an overall waste of time. Non-exhaustive CV does not use all possible combinations but rather a large enough number of iterations to be considered representative of the full data set. For this method in particular a procedure

of non-exhaustive repeated random sub-sampling validation also known as the Monte Carlo method. It works by as the name states repeatedly selecting a random selection of the data into a training and testing set, demonstrated in Figure 2.3, with the sizes of each
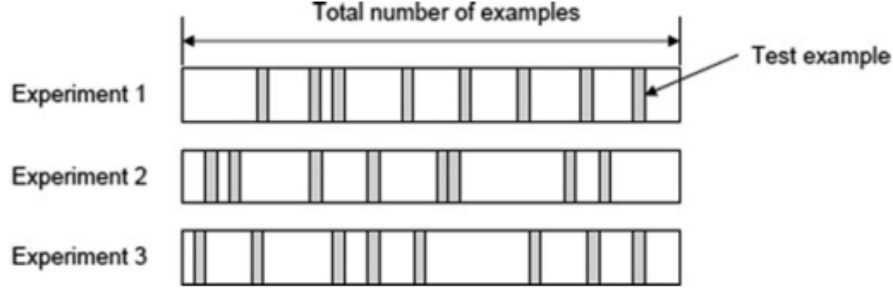

Figure 2.3: Procedure of repeated random sub-sample validation with the data split into different training (white) and testing (grey) sets for each iteration. After *Mejia et al.* (2016).

set being determined by the user. The training data is used to find the best model or solution with the associated lowest cross-validation score, and this then being compared to the testing or validating set to find the associated errors on the model. The procedure used here is similar but not to be confused with the exhaustive counterpart leave-p-out cross-validation which is the exact same process except it uses all combinations of the data, which hasn't been used here as random sampling is easier to implement. The data used here is seismic point data that is compared to a gravitationally derived moho model from selected hyperparameters. All the different models from different hyperparameter combinations are weighed up against a training set of seismic point estimates to find the model with the smallest variance or best match to these point estimates. It is then compared to the rest of the seismic data "held back" to find the Mean Square Error (MSE) and subsequently the Root Mean Square Error (RMSE), which is the average uncertainty of the model in kilometres. With 100 iterations per size and there being 3 sizes of training sets each being the closest integer value to fractions 2/3, 3/4, and 4/5 of the full data this would lead to a large enough proportion of all possible combinations to attain a representative insight to how well the model performs for an independent set. The full data consists of 937 seismic point estimates from *Assumpção et al.* (2013), of which the

locations can be seen in Figure 2.4, meaning that the different training subset sizes are 625,
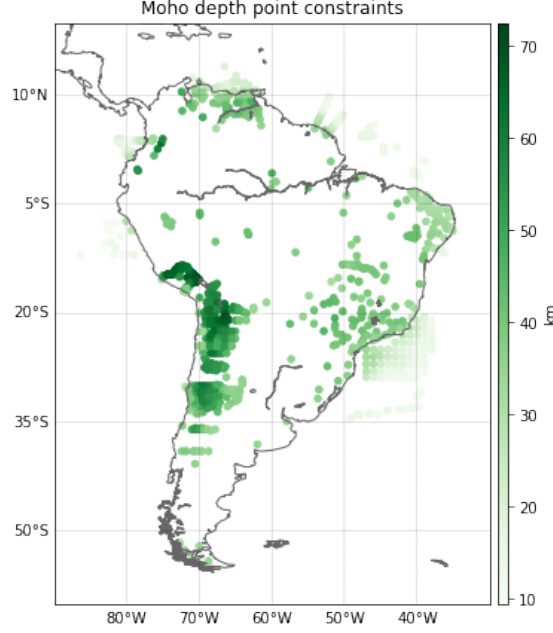

Figure 2.4: Location and depth of seismic point estimates from *Assumpção et al.* (2013). There are 937 total point estimates.

703 and 750, respectively. The training size always has to make up a larger proportion of the full data than the validating set as the model initially attained is representative of the overall data and hence the results of the MSE calculation is significant. This is supported by *Berrar* (2019) who stated that 70-90% of the full data should be part of the training set to be considered useful. A single iteration of this procedure works through randomly selecting a select number of elements in an array of 937 points, these element positions are the data points in the latitude, longitude and seismic estimates selected to be part of the training set and the leftover elements not used are placed in separate latitude, longitude and seismic estimates array and are held back. The training set is used to find the best model through a function that gets the cross-validation scores for all solutions from which the best solution is selected that has the smallest CV score. This solution is then compared to the testing set arrays returning the MSE value between the best solution and the point estimates. This value is then stored in an array with all the other iterations and is then plotted as a histogram to find the mean and standard deviation of the MSE to obtain an estimate of the uncertainty of the model in predicting the Moho depth where

there is no seismic data available. However, with all methods comes the disadvantages, repeated random sub-sample validation (RRSSV) suffers from some randomly generated selection bias, where some datum may not be selected for any iteration as a part of the validating or testing subset but on the other hand some datum may have been selected multiple times, possibly skewing the MSE result. Additionally albeit unlikely testing sets selected for separate iterations may be identical, but this should not be a problem given the sufficiently large data set so the chance of exact same subsets in different iterations is very small.

2.3 Using underplating to explain MSE values

The RMSE values reached will likely not be negligible in comparison to the model and the reasoning behind this is unmodelled or hidden masses. For instance, when calculating the Moho depth for a point if an unmodelled mass is present and has a positive density contrast, in relation to the surrounding subsurface, then the gravity model will underestimate the depth of the crust-mantle boundary. To overcome this problem these hidden masses can be modelled and included in the model calculation to produce a gravitationally derived moho that has a depth more similar to that of the seismic point estimates for a region. *Mariani et al.* (2013) tried to overcome the unusually thick crust in the Paraná basin, Brazil when compared to simple isostatic models. This was done in the form of adding underplating in the area and seeing how it changes the Moho estimates in the area. Given the success of the method, the dimensions, and properties of the underplating will be implemented into this study by combining it with the full data. The summation of these two data sets helps calculate the Moho synthetic gravity anomaly. Although the exact values needed for the intrusion are not stated in the paper they can be estimated from a specific figure, see Figure 2.5. The values ultimately used here map out a square intrusion with a density contrast of 200kg/m^3 and a depth of -30km to -45km, the lateral extent of this underplating is from -55 to -49 degrees for the west and east longitude points respectively and -27 to -21 degrees for the north and south latitude dimensions. Adding this in should increase the Moho depth in the area, however, the most likely outcome is that it will increase the MSE averages reached in the cross-validation procedure. Although, it is an interesting avenue in calculating and adding previously unmodelled masses into models hopefully increasing
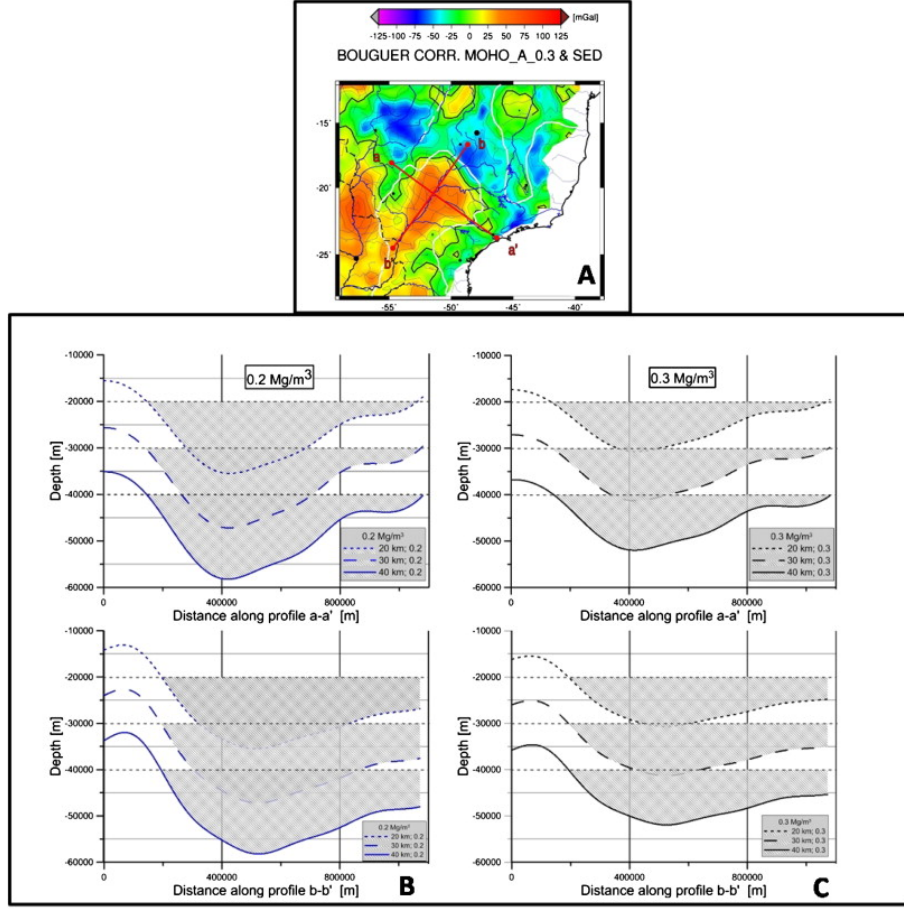
Figure 2.5: Underplating models with a $0.2Mg/m^3$ and $0.3Mg/m^3$ density contrasts for the Paraná Basin, South America. The models are calculated through inversion of gravity data. After *Mariani et al.* (2013).

the accuracy of gravitationally derived models.

2.4 Software Implementation

This inversion and error estimation method put forward in the methodology is executed in the Python programming language. Software is available under the BSD 3 clause open-source software license. The code in this project depend on open-source libraries scipy and numpy (*Harris et al.*, 2020) for number computations, matplotlib (*Hunter* (2007), http://matplotlib.org) and seaborn (*Waskom et al.* (2015), https://github.com/mwaskom/seaborn/tree/v0.6.0) for plots and maps, Fatiando a Terra

(*Uieda et al.* (2013), <http://www.fatiando.org>) for geophysics tasks. scipy.sparse package is implemented for use on sparse matrix arithmetic and linear algebra and solves the linear equation system equation. The use of Jupyter notebooks (*Pérez and Granger* (2007), <http://jupyter.org/>), which merge the source code, results, and figures of the project. All source code, Jupyter notebooks, data, and error estimate results are available through an online repository (<https://github.com/compgeolab/moho-uncertainty>).

Chapter 3

Results

To test the cross-validation (CV) approach to error estimation this code has been added into the *Uieda and Barbosa* (2016) synthetic-crust1 with Moho depth information extracted from the CRUST1.0 model (*Laske et al.*, 2013). As previously mentioned this procedure needs the availability of seismic point estimates, the data for this is from Assumpção et al. (2013). These Moho depth estimates along with their geographical location can be seen in Figure 2.4 and in total there are 937 points. The cross-validation approach used is repeated random sub-sample validation and as mentioned in the methodology randomly splits the full seismic data set into a training and testing (validating) set, with the training set compared to the solution to attain cross-validation values, the best solution is then selected from the smallest cross-validation value which is then scored against the testing set to attain the Mean Square Errors and subsequently the square root of these values which are the difference between the model and the point estimates. This error gives an indication to the average uncertainty in the overall model depth, and mainly to how good the model is where seismic data is not present which is largely the case for South America as most seismic point estimates are situated near the coast.

3.1   Cross-validation results from the synthetic model

In this run, the seismic point data will be split into a training and testing set, for a range of different proportions, these include the training size being 2/3, 3/4, and finally 4/5 of the full data. For each training size, the data that makes up this subset will be selected randomly from the full set for 100 iterations. The remaining data that was not selected for each iteration will be put into the validating set and held back for later scoring. As there are 937 separate seismic point estimates when the data is split these proportions need to be rounded to the nearest integer hence 2/3 is 625 data points, 3/4 is 703 points,

and finally 4/5 rounds to 750 points. Figure 3.1 shows the results of the cross-validation in the form of histograms showing the RMS values for all the iterations. All of these display somewhat of a normal distribution that should be more profound if more iterations were run. The mean values for all these histograms are very similar with all the values ranging between 2300-2350 metres with the highest value, 2344m, associated with the smallest training size of 625 and the smallest RMS value correlating to the largest training size. The standard deviation (std), which is a measure of the tightness of the spread to the mean value, increases with larger training sizes. This means that for larger training sizes the RMS values are more spread out with points for the largest training size of 750 having values that range from 1900-2700m with a standard deviation of 164.6. On the other hand, the other two sizes, 625 and 703, have std values of 103.0 and 124.8 respectively with RMS values not reaching below 2000m. *Szwillus et al.* (2019) uses a similar method of cross-validation to estimate Moho uncertainty, except the method used is seismic interpolation and is on a global scale rather than just South America Figure 3.2. The average Moho uncertainty calculated was 4.5km for South America however, the range of values was much larger with uncertainties in some places reaching 12km although these values were in places where no seismic data was present. This result is just over 2km higher than the mean uncertainty values seen in the histograms of around 2.3-2.4km and is likely due to the differing method. However, when compared to the model, these RMS values are quite small in comparison to the Moho depths from the model which on average is probably between 30-40km across the continent, where most of the seismic point estimates are located. The difference between the individual point estimates and the model in that location is shown in Figure 3.3. The point estimates generally tend to agree with the model, however, in few places like the Andes the model is underpredicting the Moho depth when compared to the point estimates, this could have given rise to the higher standard deviation for the larger training sets as a majority of the points held back for the validating set for some iterations may have been points from the Andes. This is especially likely seeing as a reasonable proportion of the full 937 points are situated in the mountain range.
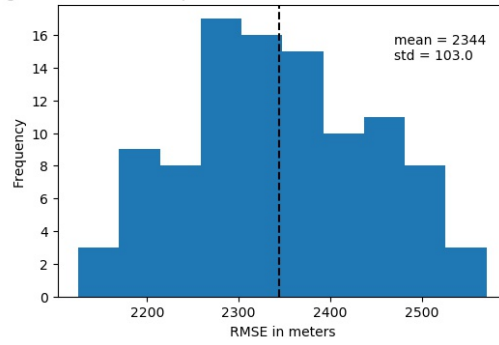
3.2 Cross-validation results after adding in underplating

This trial run is identical in every way to the synthetic-crust1 model except an intrusion has been added in the Paraná Basin. There are clusters of seismic point estimates situated in the same and surrounding area. Like the synthetic model the training sizes are 625, 703 and 750 with the full data set consisting of 937 points, each size was run with 100 iterations to create histograms of RMS values shown in Figure 3.4. These histograms like those without the intrusion display a fairly normal distribution, however, the mean values are higher. This result was expected as the inclusion of the underplating will increase the difference in that area between the model and seismic point estimates. The mean values of each training size are 2552m, 2531m and 2491m respectively with the value decreasing as the training size increases, these are around 200m higher than the equivalent training size without the intrusion. Like the results of the model without the intrusion too the standard deviation increases with larger training sizes. In comparison, these standard deviations are higher with the std value for size 625 being 131.1 which is around 28m higher than its counterpart. The ranges of the RMS values though do not exceed 1900-2700m the higher std values are explained by a larger proportion of the values attained through cross-validation being near the edges of the range. These results in comparison to the overall Moho depths are not that large as again the depths of the model are very similar to that of the model without the intrusion with the mean value being somewhere between 30-40km, so an error of about 2.5km or around 6-8%. This error is low and so indicates that the model fits the seismic data very well. These discrepancies may in part be due to the Andes problem stated above but also to the large difference between the model and the point estimates in the Paraná Basin meaning that if the majority of these points with large disparities (see Figure **??**) are selected as part of the testing set then the RMS value increases.
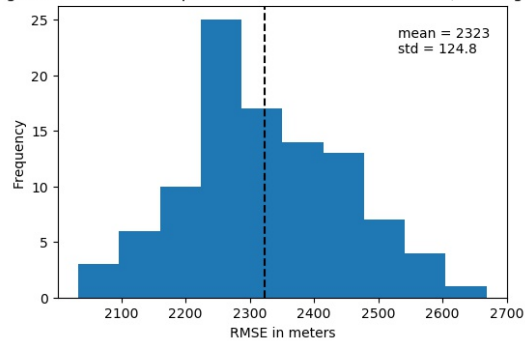
3.3 Software and run times

For the code with and without the intrusion added both took 1hr and 57 minutes with 3 different training sizes and 100 iterations per individual size, i.e. 300 iterations in total. This was performed on a laptop computer with an AMD Ryzen 5 3500U 2.1GHz processor.
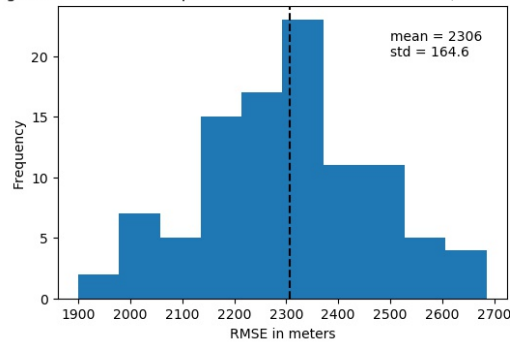
Figure 3.1: RMS values from cross validation. Training sizes top to bottom are: 625, 703, and 750 all have 100 iterations. The dotted black line indicates the mean RMS value which is stated in the top right corner along with the standard deviation (std).

Figure 3.2: Interpolated Moho depth and the corresponding uncertainty values for South America. After *Szwillus et al.* (2019).

Figure 3.3: Plot of estimated Moho depth without the added underplating with seismic point estimates superimposed. The colours of the point estimates (circles) represent the difference between the model and the seismic data.

Figure 3.4:   RMS values from cross validation with Paraná Basin underplating included in the model. Training sizes top to bottom are: 625, 703, and 750 all have 100 iterations. The dotted black line indicates the mean RMS value which is stated in the top right corner along with the standard deviation (std).
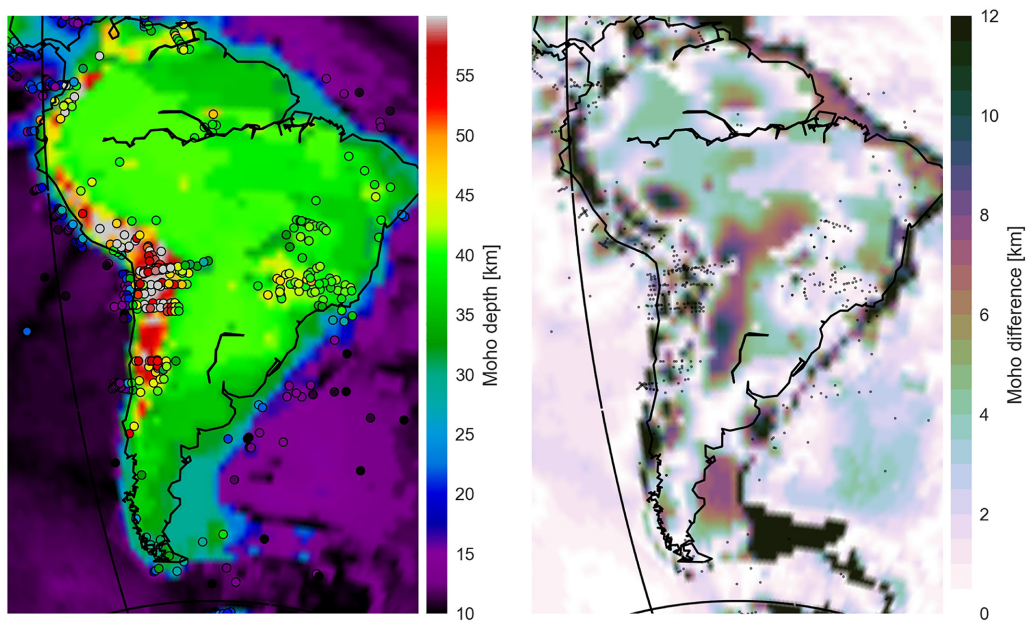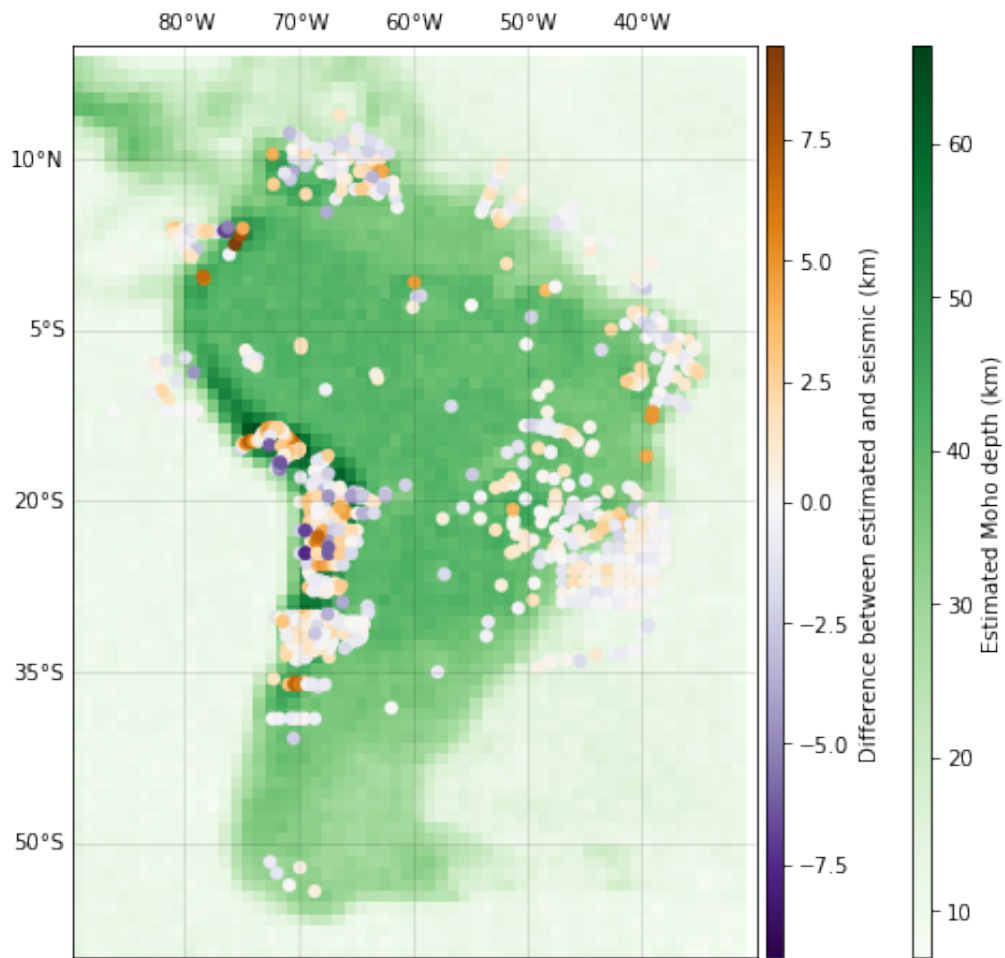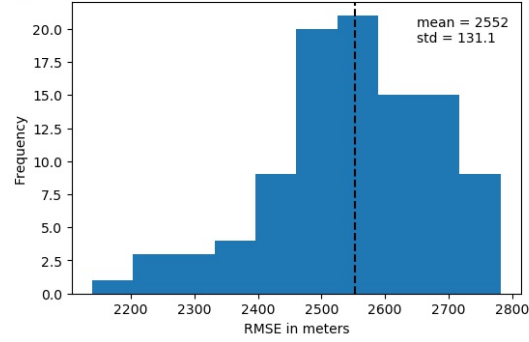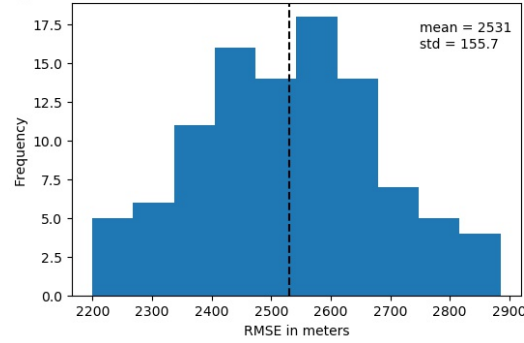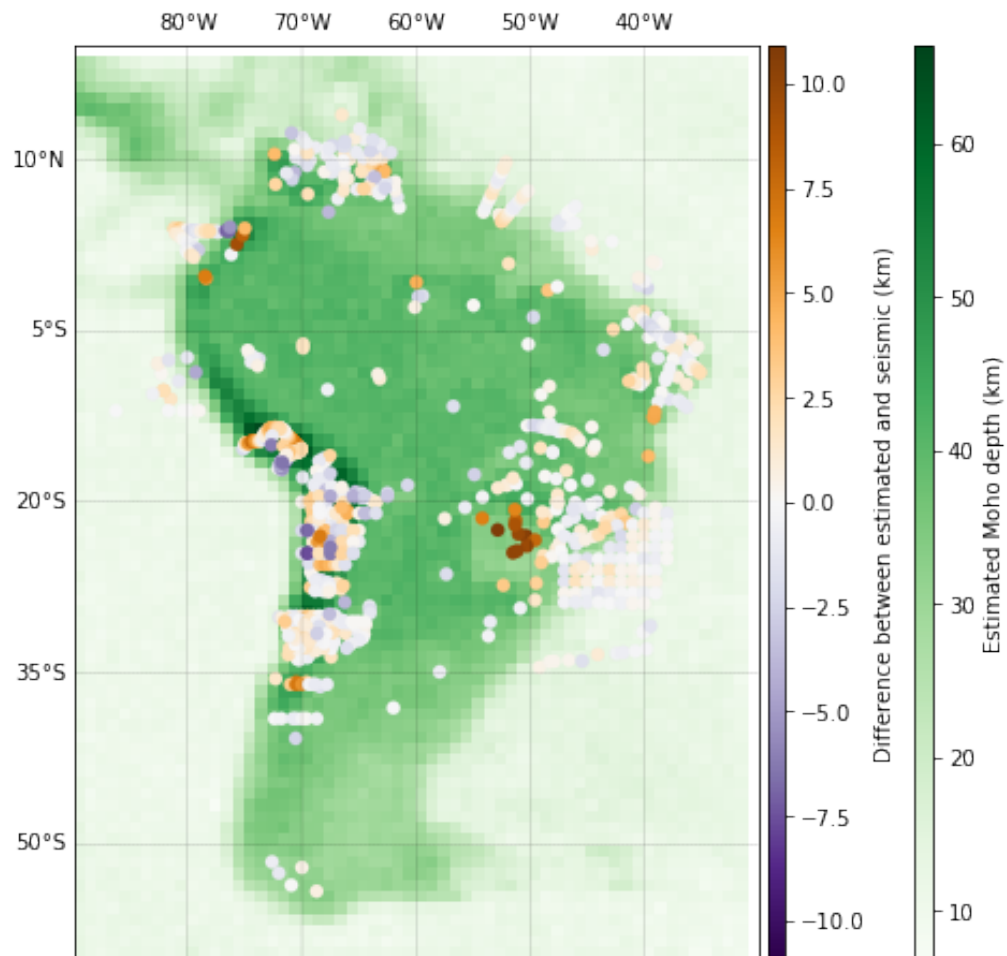
Figure 3.5: Plot of estimated Moho depth with the added underplating with seismic point estimates superimposed. The colours of the point estimates (circles) represent the difference between the model and the seismic data.

Chapter 4

Summary(Discussion, Conclusion, and Future Work)

The results obtained from the procedure of cross-validation indicate that there is an almost negligible difference between the mean of the errors when comparing all the different training sizes. This result is the same for both models with and without the added underplating of *Mariani et al.* (2013) with variations in both being around 0.1km. This means that for this model the size of the testing set does not matter because all sizes above 2/3 of the data will arrive at the same conclusion. It is worth noting though that the model with the intrusion implemented has on average higher RMS values than that without the intrusion, this was initially expected. The residual values between the model and the point estimates show that the error on the whole model was around 2.3-2.6km but this likely varies from region to region. However, this is a viable method to estimate Moho model errors using cross-validation that can be extended past just South America and to different parts of the Earth given that there are seismic point estimates to carry out the method of repeated random sub-sample validation.

4.1 Modelling a subducting slab to combat uncertainty

Even though the method was successful in estimating the error on the model this RMS error value only gives one singular uncertainty on the model, this can be skewed by few large disparities between the model and the point estimates. This is very much so the case, in the Andes due to the active plate tectonics in the area leading to a sharp increase in Moho depth when compared to the surrounding depths, something which cannot be modelled properly due to the regularization parameter which keeps the model smooth and without sharp vertical variations over short distances. Given that the regularization

parameter was selected through hold-out cross-validation, it is not plausible to change this value as it can cause instability in other parts of the model, where Moho depths are estimated well when compared to previous results. One solution that *Uieda and Barbosa* (2016) suggested was to implement a separate smoothness regularization for areas in South America such as the Andes or not to have one for these regions at all. Another solution to this problem that may be easier to implement into the code would be to model the subducting Nazca plate using a Slab2 model from rockhound (*Uieda et al.*, 2020). The only issue with this is that in the model the crust and lithosphere thickness and density would have to be assumed and then these can be mapped onto tesseroids. This procedure will address the shallow Moho under the Andes and the surrounding area and should in theory decrease the MSE values obtained from the repeated random sub-sample validation. On top of this, it would be straightforward to see how much this method would improve the error estimates on the model by comparing the difference between identically derived models with the only change being one notebook including the Slab2 model and the other not. With the deep Moho values (upwards of 40km) being underestimated in South America for many models (*van der Meijde et al.*, 2013) (*Reguzzoni and Sampietro*, 2015). One way to obtain better MSE values would be to change the location of the model to an area that lacks significant tectonic activity, for example, Africa. Although, no region is free from tectonic movement and the East African Rift Zone is the most prominent tectonic regime here. Despite this, the Moho depths in this region do not surpass 50km, and so this is a viable choice. However, if using the same procedure of cross-validation here, Africa would present the same problems as South America with the locations and clustering of seismic point data around the coast of the continent. This issue like that of South America would be due to the environment in central Africa along with financial hardship as seismic surveys are expensive to carry out. Although overall the RMS values or errors on the model should be lower as the model can be fit better to the seismic point estimates through regularization keeping the Moho surface smooth. Even if the focus was turned back to South America the CRUST1.0 model used (*Laske et al.*, 2013) doesn't consider and model the subducting slab either, though taking this into account the final Moho depth model produced using the code from *Uieda and Barbosa* (2016) has discrepancies with the CRUST1.0 model. The main reason for this is that CRUST1.0 uses seismic point estimates and extrapolates them for regions where no data is present and

South America this most notably is the Amazon Rainforest. On the other hand, the Uieda model is computed using gravitational data, rather than seismic. One area where these models somewhat disagree is in the Andes and to the east of the vast mountain range. The CRUST model predicts a deeper Moho through extrapolation due to the isostatic balancing from the Andes whereas this part is modelled to be much shallower in *Uieda and Barbosa* (2016) as there is gravity data for the foreland basins.

4.2 Blocked testing sets

The procedure used in this study as mentioned used repeated random sub-sample validation to estimate the errors on the model as a result the singular error value attained is the average error on the whole model across the continent of South America. When this is taken into account as a large-scale representation, there is a lack of information given as the areas of larger and smaller disparity are not known. Rather than using "random" sampling of the data to separate it into the training and testing set instead the data can be split up using a method of "Blocking". By selecting points for the testing set by geographical area rather than randomly an uncertainty estimate can be calculated for just the region that the seismic point estimates span. By using this method the likely larger errors seen in the Andes can be separated from the areas of the continent where the model and the point estimates agree very well with each other leading to smaller RMS values in blocked areas with small disparities. However, "Blocking" with limited seismic estimates in some regions and large quantities of these estimates grouped in tight clusters may lead to different validating sizes, which as seen in the results is not much of a problem as the 3 different sizes used achieved very similar mean RMS values. For this method to work the blocked regions must be of equal or similar size as a testing size of 100 points cannot be compared to a section with only 10 seismic point estimates. So when implementing "Blocking" for South America areas where there are a large proportion of seismic points such as the Andes may have to be split into two or more regions. However, rather than gaining an error estimate for the model for the Andes or another high cluster region as a whole there would be multiple error values attained and would account as a separate RMS value per region. This uneven cluster would lead to errors for regions of possibly vastly different sizes depending on the testing size collected, with the smaller the subset

the more areas there are and likely the more computationally expensive the process will be.

4.3 Degrees of freedom in density estimations

The code implemented by *Uieda and Barbosa* (2016) uses cross-validation to estimate the best model hyperparameters before producing the final model. Overall, the misfit between the point estimates and the model is small, but in some places, the model has large disparities with the seismic point estimates. This is partially due to the singular density contrast used across the whole continent for the model. South America is a tectonically active region and will have geological units vastly ranging in densities from the less dense sedimentary rocks and basins to the very dense igneous intrusions. So one way of accounting for this is adding in more degrees of freedom in density estimations. This method has been implemented in *Haas et al.* (2020) using seismic regionalization that splits up the continent of South America into 6 different density contrasts (see Figure 4.1). These
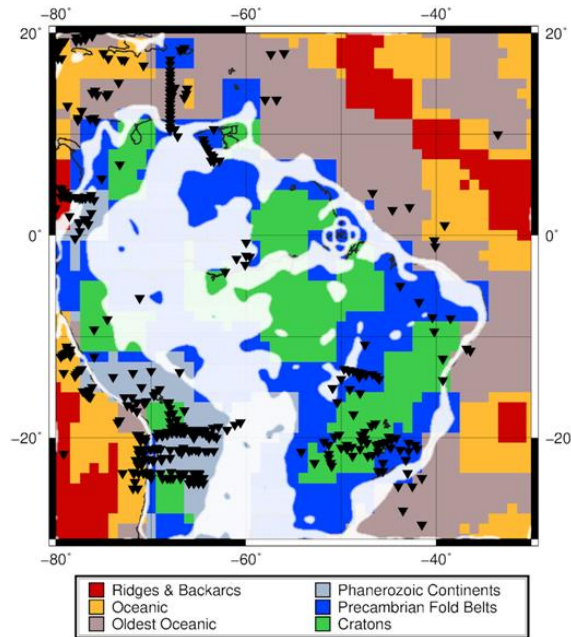

Figure 4.1:   Geologic units from seismic regionalization superimposed with estimated Moho depth of between 28km to 32km. The black triangles represent seismic stations. After *Haas et al.* (2020).

density values are attained by using surface-wave tomographic models seen in *Schaeffer and Lebedev* (2015). If this method were implemented into this code then it is likely that the mean errors between the model and the seismic constraints would decrease as the model created through inversion will better fit the seismic points whilst still having a smooth solution through a regularization coefficient. However, the main issue that arises is that the model requires the user to manually choose how many different regions there will be. For every extra region added to the model there will be an exponential increase in the computational time, and after a certain number of different density contrasts adding more will not improve the model. Hence a trade-off is needed between the number of regions and the computational time. In *Haas et al.* (2020) as mentioned 6 regions with different density contrasts were used and comparing this model to *Uieda and Barbosa* (2016) the frequency of residuals is clustered more towards smaller values whereas the latter model has a slightly higher representation when it comes to larger residual Moho depths (see Figure 4.2). A possible future avenue could be to put the computational time aside and
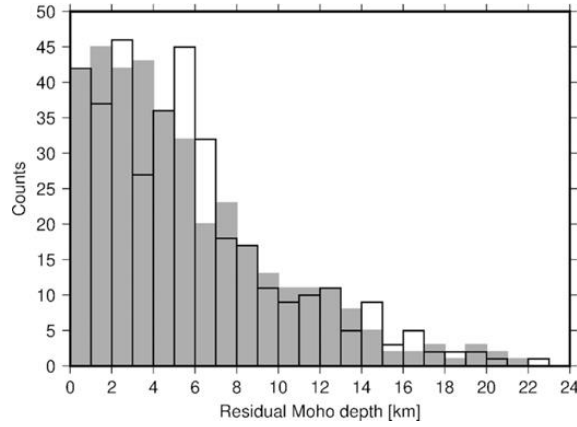

Figure 4.2: Histogram comparing *Haas et al.* (2020) (grey bars) residual Moho depths to *Uieda and Barbosa* (2016) model (black bars). The bin size for each model is 1km. From *Haas et al.* (2020).

see how many different density regions can be implemented into the code until adding more in does not decrease the mean size of the Moho residuals between the model and point estimates.

References

Asgharzadeh, M. F., R. R. B. von Frese, H. R. Kim, T. E. Leftwich, and J. W. Kim, Spherical prism gravity effects by Gauss-Legendre quadrature integration, *Geophysical Journal International*, *169*(1), 1–11, doi:10.1111/j.1365-246X.2007.03214.x, 2007.

Assumpção, M., M. Feng, A. Tassara, and J. Julià, Models of crustal thickness for south america from seismic refraction, receiver functions and surface wave tomography, *Tectonophysics*, *609*, 82–96, doi:10.1016/j.tecto.2012.11.014, 2013.

Berrar, D., Cross-validation, in *Encyclopedia of Bioinformatics and Computational Biology*, pp. 542–545, Elsevier, doi:10.1016/b978-0-12-809633-8.20349-x, 2019.

Bott, M. H. P., The use of Rapid Digital Computing Methods for Direct Gravity Interpretation of Sedimentary Basins, *Geophysical Journal International*, *3*(1), 63–67, doi:10.1111/j.1365-246X.1960.tb00065.x, 1960.

Haas, P., J. Ebbing, and W. Szwillus, Sensitivity analysis of gravity gradient inversion of the moho depth—a case example for the amazonian craton, *Geophysical Journal International*, *221*(3), 1896–1912, doi:10.1093/gji/ggaa122, 2020.

Hansen, P., Analysis of Discrete Ill-Posed Problems by Means of the L-Curve, *SIAM Review*, *34*(4), 561–580, doi:10.1137/1034115, 1992.

Harris, C. R., et al., Array programming with NumPy, *Nature*, *585*(7825), 357–362, doi:10.1038/s41586-020-2649-2, 2020.

Hunter, J. D., Matplotlib: A 2D graphics environment, *Computing in Science & Engineering*, *9*(3), 90–95, doi:10.1109/MCSE.2007.55, 2007.

Laske, G., G. Masters, Z. Ma, and M. Pasyanos, Update on CRUST1.0 - A 1-degree Global Model of Earth's Crust, in *EGU General Assembly Conference Abstracts*, vol. 15, pp. EGU2013–2658, 2013.

Li, X., and H. Götze, Ellipsoid, geoid, gravity, geodesy, and geophysics, *GEOPHYSICS*, *66*(6), 1660–1668, doi:10.1190/1.1487109, 2001.

Mariani, P., C. Braitenberg, and N. Ussami, Explaining the thick crust in paraná basin, brazil, with satellite GOCE gravity observations, *Journal of South American Earth Sciences*, *45*, 209–223, doi:10.1016/j.jsames.2013.03.008, 2013.

Mejia, G., A. Gómez Montoya, and O. Quintero, Reconocimiento de emociones utilizando la transformada wavelet estacionaria en señales eeg multicanal, doi:10.13140/RG.2.2.18377.24167, 2016.

Pérez, F., and B. E. Granger, IPython: A System for Interactive Scientific Computing, *Computing in Science & Engineering*, *9*(3), 21–29, doi:10.1109/MCSE.2007.53, 2007.

Reguzzoni, M., and D. Sampietro, GEMMA: An earth crustal model based on GOCE satellite data, *International Journal of Applied Earth Observation and Geoinformation*, *35*, 31–43, doi:10.1016/j.jag.2014.04.002, 2015.

Schaeffer, A., and S. Lebedev, Global heterogeneity of the lithosphere and underlying mantle: A seismological appraisal based on multimode surface-wave dispersion analysis, shear-velocity tomography, and tectonic regionalization, in *The Earth's heterogeneous mantle*, pp. 3–46, Springer, 2015.

Silva, J., W. Medeiros, and V. Barbosa, Potential-field inversion: Choosing the appropriate technique to solve a geologic problem, *GEOPHYSICS*, *66*(2), 511–520, doi:10.1190/1.1444941, 2001.

Silva, J., D. Santos, and K. Gomes, Fast gravity inversion of basement relief, *Geophysics*, *79*(5), G79–G91, doi:10.1190/geo2014-0024.1, 2014.

Szwillus, W., J. C. Afonso, J. Ebbing, and W. D. Mooney, Global crustal thickness and velocity structure from geostatistical analysis of seismic data, *Journal of Geophysical Research: Solid Earth*, *124*(2), 1626–1652, doi:10.1029/2018jb016593, 2019.

Tikhonov, A. N., and V. Y. Arsenin, *Solutions of ill-posed problems*, Scripta series in mathematics, Winston, Washington : New York, 1977.

Tugume, F., A. Nyblade, J. Julià, and M. van der Meijde, Precambrian crustal structure in africa and arabia: Evidence lacking for secular variation, *Tectonophysics*, *609*, 250–266, doi:10.1016/j.tecto.2013.04.027, 2013.

Uieda, L., A tesserioid (spherical prism) in a geocentric coordinate system with a local-north-oriented coordinate system, doi:10.6084/M9.FIGSHARE.1495525.V1, 2015.

Uieda, L., and V. C. Barbosa, Fast nonlinear gravity inversion in spherical coordinates with application to the south american moho, *Geophysical Journal International*, *208*(1), 162–176, doi:10.1093/gji/ggw390, 2016.

Uieda, L., V. C. Oliveira Jr, and V. C. F. Barbosa, Modeling the Earth with Fatiando a Terra, in *Proceedings of the 12th Python in Science Conference*, edited by S. van der Walt, J. Millman, and K. Huff, pp. 91 – 98, 2013.

Uieda, L., S. R. Soler, and A. Pesce, Rockhound: Download geophysical models/datasets and load them in Python, doi:10.5281/zenodo.3627166, 2020.

van der Meijde, M., J. Julià, and M. Assumpção, Gravity derived moho for south america, *Tectonophysics*, *609*, 456–467, doi:10.1016/j.tecto.2013.03.023, 2013.

Waskom, M., et al., seaborn: v0.6.0, 2015.