Data science methods to discern crustal and lithospheric configurations using gravity and Geoid anomalies

Dr. Terence Vockerodt

May 2022

Contents

Part I

Introduction

1.1 What are gravity anomalies?

Gravity anomalies are the difference between measured values of gravitational acceleration at a point on the Earth, and the value predicted by a theoretical model [1,2]. A simple model of the Earth assumes that it is an oblate spheroid or ellipsoid (a sphere with a wider equatorial radius than polar radius), with a constant density of material beneath the surface [2]. The Earth does have an ellipsoidal shape due to its rotation [2], but clearly this simple description is very different from our Earth, which does not have a smooth surface and has varying density composition. This causes changes in the gravitational field (relative to the oblate spheroid model) that manifest themselves in the measurements of gravitational acceleration.

Gravity anomalies are useful measurements, since they can be used to determine the subsurface structure of the Earth without intrusion. One can use gravity anomalies in order to scout for many things in the subsurface, such as ore/material deposits [1], ground-water reservoirs [1], faults [1], and geothermal sites that can generate renewable energy [2].

Typically the observed/measured gravity is corrected for effects from celestial bodies such as the Sun and Moon [2]. It is also corrected for the Earth's rotation, latitude, tidal effects and gravity meter fluctuations [1]. Gravity meter fluctuations are typically caused by the deformation of the spring in the spring-mass system used [2]. In general, gravity meter measurements are not made on the surface of the ellipsoid. The free air correction to the observed gravity essentially maps the gravity reading to the surface of the ellipsoid [1,2]. This is then further corrected for mass that exists between the observation point and the ellipsoid surface/sea level [1,2], which is known as the Bouguer anomaly. The isostatic anomaly is calculated by subtracting the effect of subsurface mountain roots from the Bouguer anomaly [1].

An example of a synthetic dataset that shows the variation of gravity anomalies with terrain is shown in Fig. 1.1. Here, a reference configuration is modified with thickening and thinning of the lithospheric layers - the crust, the upper mantle and the asthenosphere. The thickness difference causes patterns in the Bouguer, Geoid and Airy anomalies, which can be used as inputs into data science models to predict the configuration. The Airy anomaly is a particular example of an isostatic anomaly [3], and the Geoid anomaly is the anomaly in the measurement of gravity at the ocean surface [4].

1.2 Project specification

The key idea for this project is to investigate how data science can be used to discern different crustal and lithospheric configurations, relative to some reference configuration, and to develop a procedure to automate allocation of configurations. This relies on knowing the configuration directly below each set of anomalies in Fig.

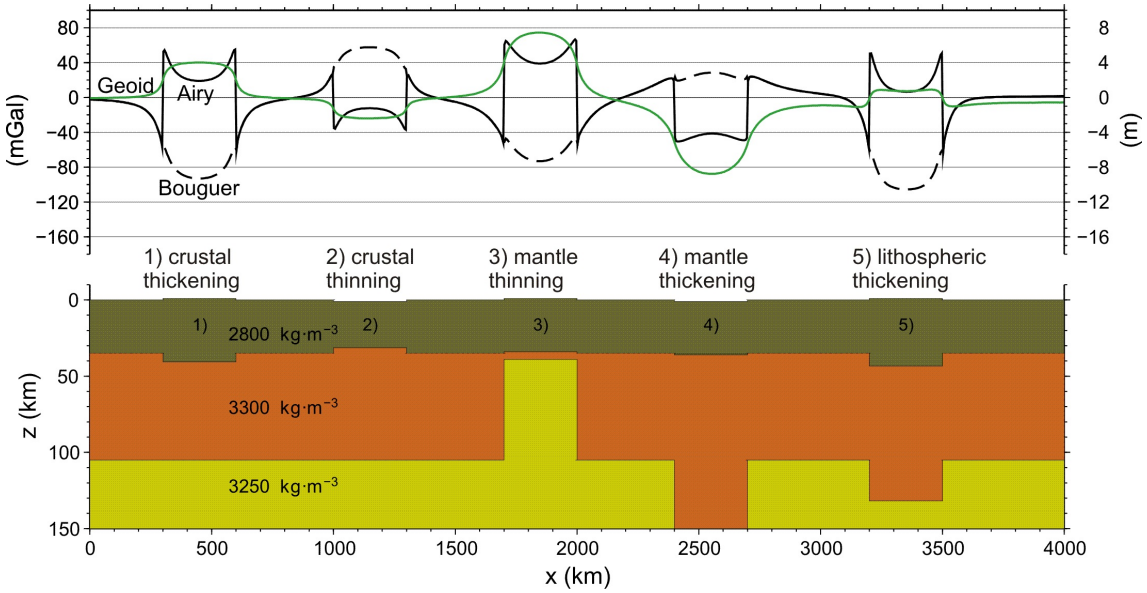

Figure 1.1: Synthetic dataset of Bouguer, Geoid and Airy gravity anomalies for six different crustal and lithospheric configurations, for a cross section of land under local isostasy. On the bottom graph: top layer is the crust, middle layer is the upper mantle, and bottom layer is the asthenosphere. Data generated using open source tAo code [5]. Figure made by user Gaianauta - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=20398751, and is an adaptation of Fig. 6-24 in reference [6].

1.1, making this a supervised procedure. However, this does not necessarily mean that the project has to use supervised learning techniques exclusively.

In geophysics, the process of determining a subsurface configuration from a known gravity reading is called inversion [2]. Typically with inversion, a computer model starts at an initial configuration and generates the resulting gravity anomalies. The model then iterates the configuration using some best fit metric to determine a closer fit to the measured real world anomalies. Initially, one may have a group of points being fitted using this procedure, and newer points can be assimilated into the model to better determine a configuration. Depending on the number of measured points available, it may take quite some time before an accurate configuration is known. The fewer measurements available, the more difficult it will be to determine an accurate configuration. This makes it difficult to classify less physically accessible regions due to lack of data points.

This project aims to use only the anomalies at a specific coordinate to predict the configuration of what is directly underneath. This relies on generating a set of configurations like in Fig. 1.1, which only takes one iteration of the scientific code used in typical inversion processes per configuration. This basis set or catalogue of configurations is then used to classify measured anomalies. This catalogue can be made to be simple and coarse like in Fig. 1.1, or detailed and complicated. It can even sample configurations from converged inversion codes for more realistic

configurations.

Once the data science methods have been developed, they can then be used to automatically predict sets of anomalies in a scalable way. At least, one must provide one of each anomaly shown in Fig. 1.1, but entire batches of anomalies can be predicted at a time due to how the data science methods are coded, which allows for scalable automation of the classification process.

1.3 Ethical concerns

I have mentioned before that the gravity anomalies can be used to scout the lithospheric configuration of the Earth's surface. This information can be used to determine whether or not precious ores are likely to be beneath the surface, for instance. Reckless mining operations that have no concern for the local ecology and cultural significance of the land should not be tolerated, in my opinion, as this naïve and inhumane approach is detrimental to the sustainable progress of humanity. Even if these tools are used for scouting potential renewable energy sites, proper respect to the ecology and culture of the land should be given. With an established understanding and rapport with the potential sites of study, one can then feel comfortable using these tools to further the analysis.

Part II

Methods

2.1 Data extraction

Since I do not have access to the data files directly, I used software called Engage Digitizer to extract the data points from Fig. 1.1. This allows one to apply a grid to the image of the data, and then data points can be obtained as coordinates on the grid. Whilst this extraction is not perfectly accurate, for this stage of the project it is irrelevant, as this synthetic dataset is just used to demonstrate the training process as a proof of concept. In reality, the training dataset used ought to have many more crustal and lithospheric configurations considered, such as ones for different rock densities.

When sampling points from the three curves, it was common to not have the x values of sampled points align correctly due to computer mouse precision. Luckily, the software has inbuilt interpolation, such that the sampled points of one or more curves can be matched with another curve using interpolation. Due to the apparent discontinuities in the Airy gravity anomaly, I opted to interpolate the Geoid and Bouguer anomaly sample points to the sampled Airy x values. This avoids using interpolation on the Airy anomaly points, which is highly inaccurate due to the discontinuity in the curve. After this interpolation, one has the three anomalies per x value, and we can move onto training a model.

2.2 K-means clustering

K-means clustering is an unsupervised learning algorithm that clusters data based on its proximity to a cluster's mean vector. The mean vector is then varied based on the data points that were assigned to that cluster, and then the cluster assignment process starts again. The mean vector of each cluster continues to vary until the clusters are stable. In our case, a mean vector would be a specific value for each Airy, Bouguer and Geoid anomaly.

This algorithm runs very quickly and will cluster data points with similar values to one another together. But there is one crucial detail I have overlooked so far - how can an unsupervised algorithm be used in a supervised procedure? The answer is related to cluster assignment. Suppose you have a set of anomalies, where you know with certainty what crustal and lithospheric configuration that they are associated with. Let's say that when you assign it to a cluster, it assigns to cluster number 1. Now, any other data points which also assign to cluster 1, assuming there is a decent number of clusters, are likely to also feature a similar configuration. If one has this kind of example anomaly set for every cluster, then one can associate unknown points regardless of which cluster they are in. These example data points can be generated synthetically, or they can use real world data - just as long as the configuration for the anomalies is known with certainty.

Unfortunately, since I have synthetic data only, I cannot assess the accuracy of such

a method for this purpose. I will investigate its ability to cluster the synthetic data, in order to see what the clusters look like qualitatively. In principle at least, this is the fastest method of those considered in this document.

2.3 Decision trees and random forests

Decision trees are a supervised learning algorithm, that are relatively intuitive to understand. Decision trees operate with a flowchart like structure, where bifurcations (splitting the data into two sets) in the data are made based on if their data values are greater/equal or less than some learned value. An example of a decision tree is shown later on in the results in Fig. 3.5. For each bifurcation only one aspect of the data is considered at a time, which in our case corresponds to one specific anomaly. Eventually, the bifurcations lead to separation of two different configurations.

Decision trees have an advantage in that they do not use redundant features, which can be used to inform later decisions of feature engineering. The results from decision trees are also invariable under scaling of the data, which makes it very intuitive to generate interpretable results since no rescaling procedures need to be employed on the decision boundaries. Decision trees are easier to understand than neural networks, but are not always accurate and tend to overfit the training data if they are deep enough. Random forests are generally an improvement to decision trees in that they take the average of many deep decision trees, which attempts to smooth out the variance from the overfitted deep trees.

Since this is a supervised learning algorithm, the anomalies must be associated with a unique label corresponding to their configuration. It is likely that the training set will not include many real world examples, since the anomalies are usually taken without then excavating to see what is underneath the point where the reading is (this would subsequently change the reading at that point, too). This promotes the use and development of synthetic datasets from scientific computer models, since any number of crustal and lithospheric configurations that can be imagined can be modelled without excavation. Eventually, the model ought to be used to classify real world gravity and Geoid anomalies, but for training the labelled synthetic dataset is essential.

The training takes longer on average than using an algorithm like K-means, which does not require an explicit 'training' phase. Since the random forest is an aggregation of many trees, this of course takes longer to train than a single decision tree, with the hope of providing more accuracy.

2.4 Artificial neural networks (ANNs)

Artificial neural networks (ANNs) are an attempt at modelling a biological brain within a computer, and they are the bread and butter of the machine learning methods. They are not intuitive to understand which makes their intermediate results difficult to work with, but they provide some of the best supervised learning accuracy out of all of the methods. Since data interpretability is lost using these methods, one often opts to scale the data points so that the training algorithm parameters are bounded more appropriately.

Much like the decision tree and random forest methods, the ANN needs the data to be labelled, which as mentioned before would require a sufficiently accurate scientific computer model to generate synthetic data. I expect this method to be the most accurate in classifying the synthetic dataset extracted from Fig. 1.1.

Part III

Results

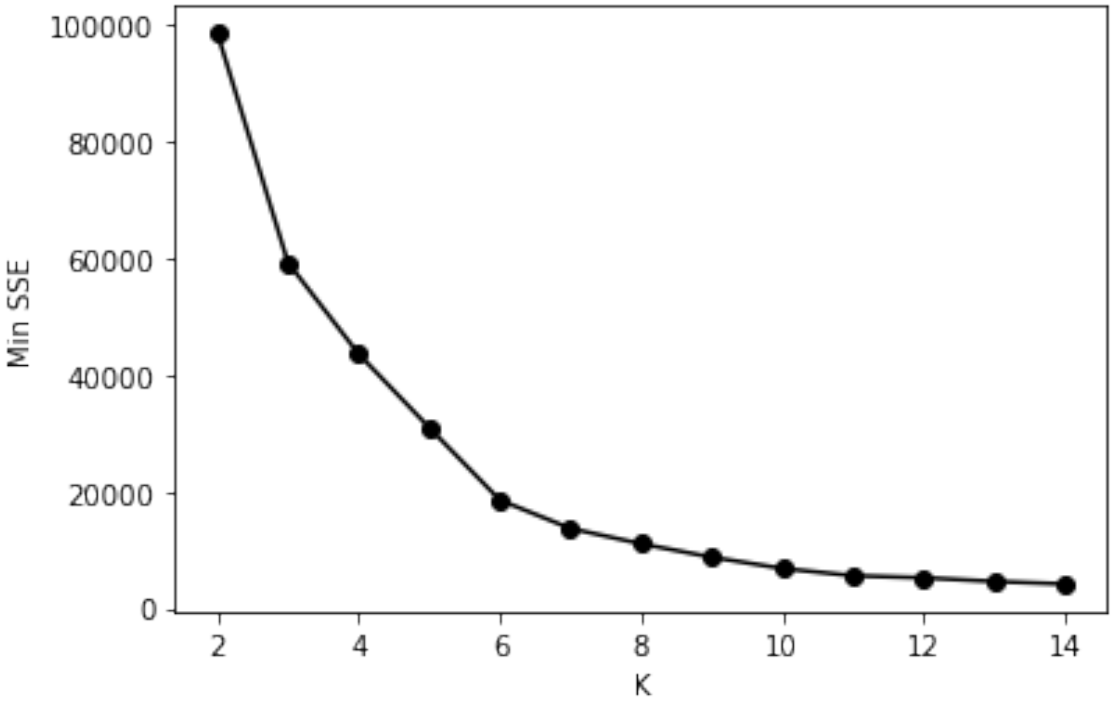

Figure 3.1: Elbow method plot for the K-means clustering method applied to Airy, Bouguer and Geoid anomalies from Fig. 1.1. Elbow located at $K = 6$, which is the same number of unique configurations that we consider.

3.1 Objectives

The types of results we are after are twofold. Firstly, we want to see qualitatively how the K-means method performs to better understand how it will cluster the data. The elbow method will be used as a metric to find the optimal K value. Once the K-means method is investigated, the supervised learning procedures will be investigated for their efficacy. This will involve standard classification report that monitors the accuracy of the model with respect to the six different configurations.

3.2 K-means clustering

For the K-means method, I set up the models in the Jupyter notebook entitled '1) K-means'. I start by using the elbow method to find the optimal number of clusters K to use. This involves creating several instances of the model with an iterating value for K. Using the three anomalies, the elbow method results are shown in Fig. 3.1. The elbow method analysis shows that the optimum number for K is 6, which coincidentally corresponds to the number of configurations that we consider in Fig. 1.1.

The model was then ran using $K = 6$, and the resulting clusters are shown in Fig.

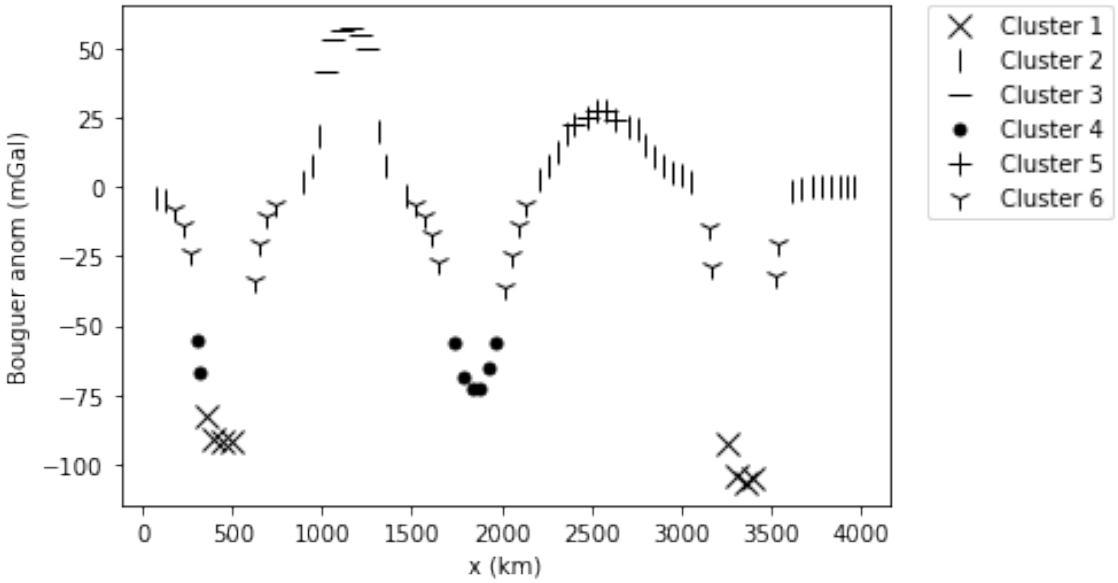

Figure 3.2: Example classification for $K = 6$ using the K-means clustering method applied to the Airy, Bouguer and Geoid anomalies from Fig. 1.1. The clustering effectively bands the data, without capturing the underlying patterns that distinguish the configurations.

3.2. There is a mismatch between what the clusters are doing and what we want the clusters to do. The clustering procedure essentially bands the data between two boundaries, similar to what a decision tree would do. However, this does not properly classify lots of the data points, and the classification should not band the data. Increasing K further just introduces more and more bands, which is not the kind of classification that we would like.

Part of the reason that the K-means method bands the data is because it only looks at the three anomalies in isolation when assigning to clusters. The anomalies are related to one another in that they all represent corrected versions of the observed gravity, and it is clear in Fig. 1.1 that there are some similarities between some of the anomalies. For instance, the Bouguer and Airy anomalies follow each other very closely for the default configuration, which is sensible because the default configuration does not contain subsurface mountain roots. When the configurations change from the default configuration, the Airy anomaly experiences near-asymptotic shifts in value. In fact, these two anomalies in tandem should provide a powerful indicator of configurations different to the default due to this behaviour. Therefore, I opted to include extra features that link two different anomalies together.

Consider two different anomalies Δg_N and Δg_M. The family of features that I decided to use are

$$f_{12}(\Delta g_N, \Delta g_M) = \Delta g_N^2 - \Delta g_M^2,$$
$$f_{22}(\Delta g_N, \Delta g_M) = \Delta g_N \Delta g_M,$$

where the first subscript index indicates the index of the function in the family and the second indicates the number of anomalies considered on the right hand side. This was to ensure that if additional features are needed that involved up to the three anomalies that these features could be labelled systematically. Essentially, it boils down to taking the difference of the two squares of the two anomalies, and calculating the cross term. These functions were applied for all possible dual combinations of the three anomalies in Fig. 1.1.

Using these new features, I employed the elbow method again shown in Fig. 3.3. This time, the elbow was calculated at $K = 5$, but this is close enough to $K = 6$ that I opt to use 6 instead. This is because we have 6 configurations to consider from Fig. 1.1. Fig. 3.4 shows the clustering for $K = 6$, including these new features. This time, the clustering is remarkably closer to the original data classification. Note that the clusters in Fig. 3.4 do not correspond to the clusters of Fig. 3.2, since it is random which set of points gets assigned to any particular numbered cluster. Whilst some points were misclassified as crustal thinning instead of crustal thickening, these new quadratic features enable a clustering that is significantly closer to the configurations.

Now, if one assigns real world data to these clusters, they can see which of the synthetic data also enters those clusters, which can help properly classify the real world data point configurations. From these results, the K-means method can be particularly strong at classifying the stationary points of the anomalies, although it is difficult to determine where those points are in practice without multiple gravity readings.

3.3 Decision trees and random forests

The code for the decision trees and random forests is located in the Jupyter notebook entitled '2) Decision trees and random forests'. Since this is a supervised learning algorithm, I must split the data into a test and train set at least. I chose the test train split to be 33% test data, 67% train data. After 15 different runs solely using the three anomalies as features, the average decision tree and random forest test set accuracy was 96% and 97% respectively, making these methods reasonable choices for the classification. Fig. 3.5 shows an example of a decision tree, with the relevant bifurcated groups shown above and below each decision boundary. The decision tree in Fig. 3.5 misclassifies two points as lithospheric thickening instead of crustal thickening. This is not unreasonable to expect for some models, because these two configurations both involve thickening of the subsurface to some degree and thus are

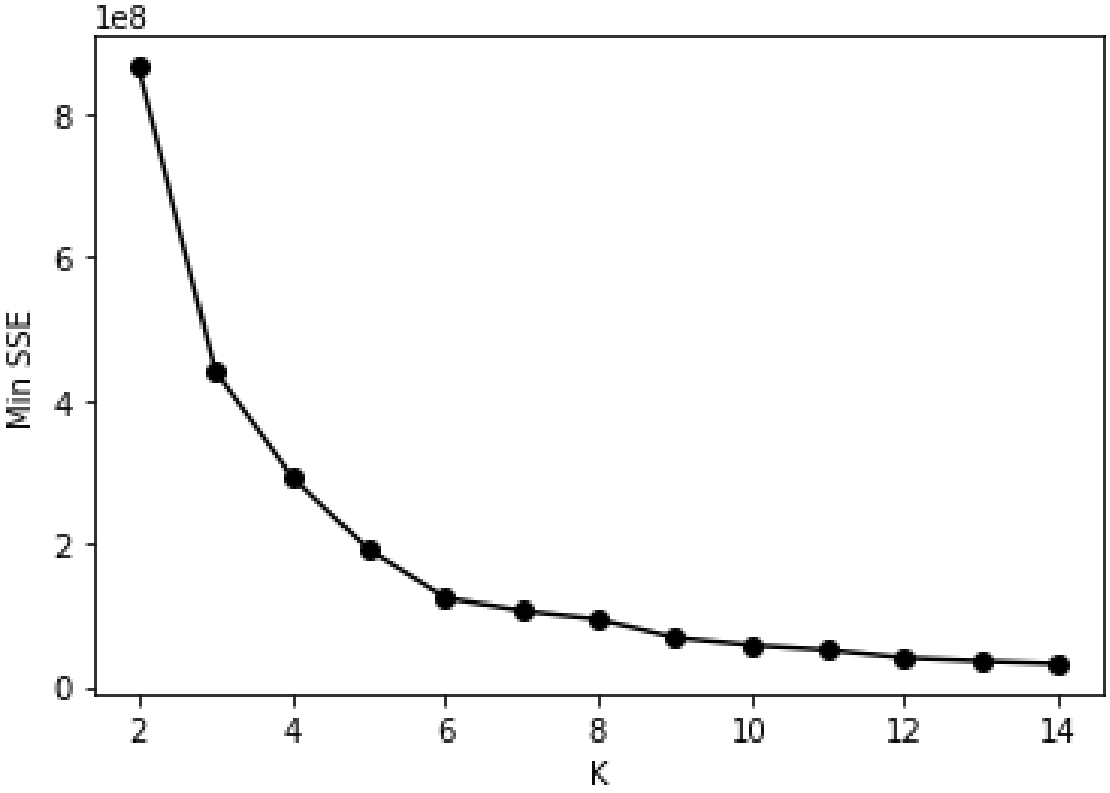

Figure 3.3: Elbow method plot for the K-means clustering method applied to Airy, Bouguer and Geoid anomalies from Fig. 1.1 and their quadratic combinations as explained in the text. Elbow located at $K = 5$, but subsequent results use $K = 6$ in order to distinguish between all the configurations we are considering from Fig. 1.1.

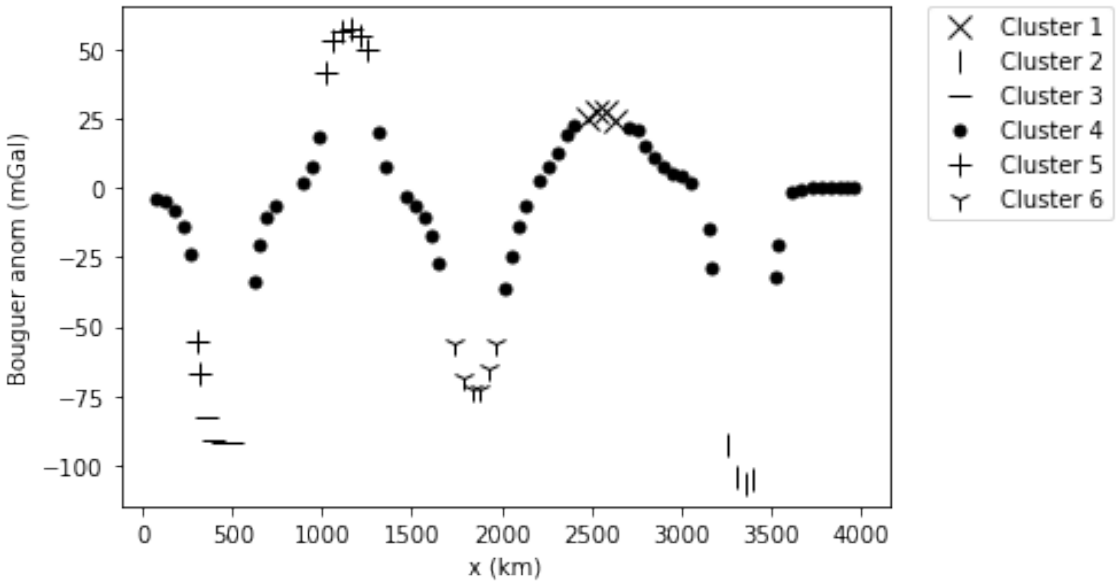

Figure 3.4: Example classification for $K = 6$ using the K-means clustering method applied to Airy, Bouguer and Geoid anomalies from Fig. 1.1 and their quadratic combinations as explained in the text. Some points are misclassified as crustal thinning instead of crustal thickening, which holds little to no physical basis since the configurations are starkly different to one another. Overall, the accuracy of the clustering as compared to the reference data is significantly better than Fig. 3.2 (note that the clusters are labelled differently between the two diagrams).

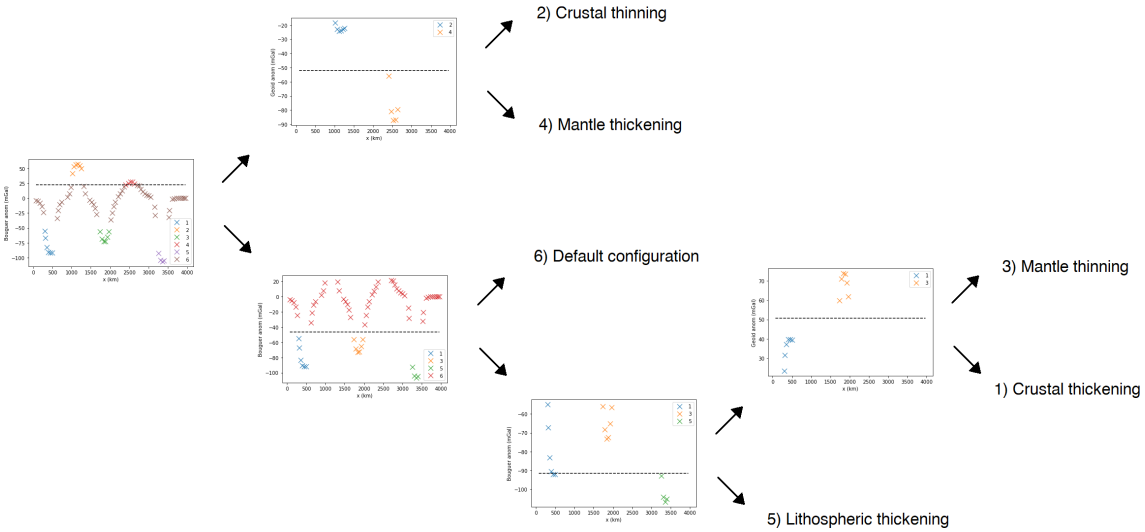

Figure 3.5: Example decision tree that uses the Airy, Bouguer and Geoid anomalies from Fig. 1.1, with 100% training set accuracy and 92% test set accuracy. Size of test set is 25, so this corresponds to 2 misclassified test set points, which are classified as lithospheric thickening instead of crustal thickening. Decision values shown as dashed lines on each graph. Reading from left to right - points above on or the decision line go up and right, and points below go down and right.

very physically similar. Fig. 3.6 shows an example of the quality of classification by the decision tree model, which again shows a thickening misclassification. However, despite being more accurate in general, some runs of the random forest (see Fig. 3.7) misclassified points as mantle thinning instead of crustal thickening, which holds little to no basis in reality since they are not even remotely similar processes physically speaking.

Since using quadratic features helped the K-means method produce more accurate results, I ran a random forest with higher order features. After 15 runs, the average accuracy for both test and train sets was found to be 100%. Upon checking the individual decision trees, none of the quadratic features were redundant.

3.4 ANN

The code for the ANN is located in the Jupyter notebook entitled '3) ANN'. For this task, I built the ANN with two hidden layers of size 50, each with a 25% dropout rate to prevent overfitting. Each hidden layer activates using relu, except for the output layer which activates using the softmax function since I use one hot encoding for the outputs. The features/inputs were scaled using the StandardScaler() method from sklearn.preprocessing, in order to constrain the weights to more computationally manageable values.

Using the same train test split as with the decision trees and random forests (33%

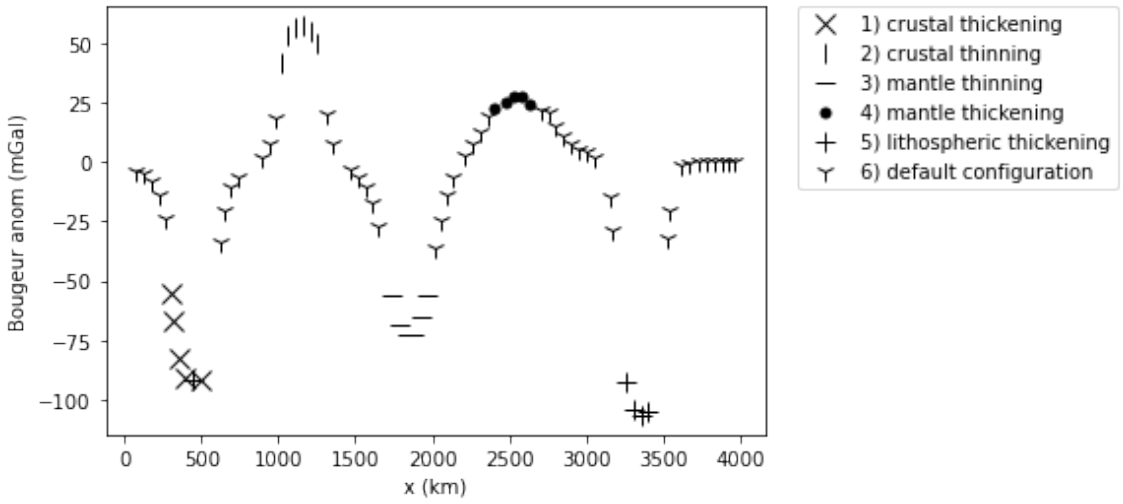

Figure 3.6: Example classification of points using a decision tree that uses the Airy, Bouguer and Geoid anomalies from Fig. 1.1. In this example, one point was misclassified as lithospheric thickening instead of crustal thickening.

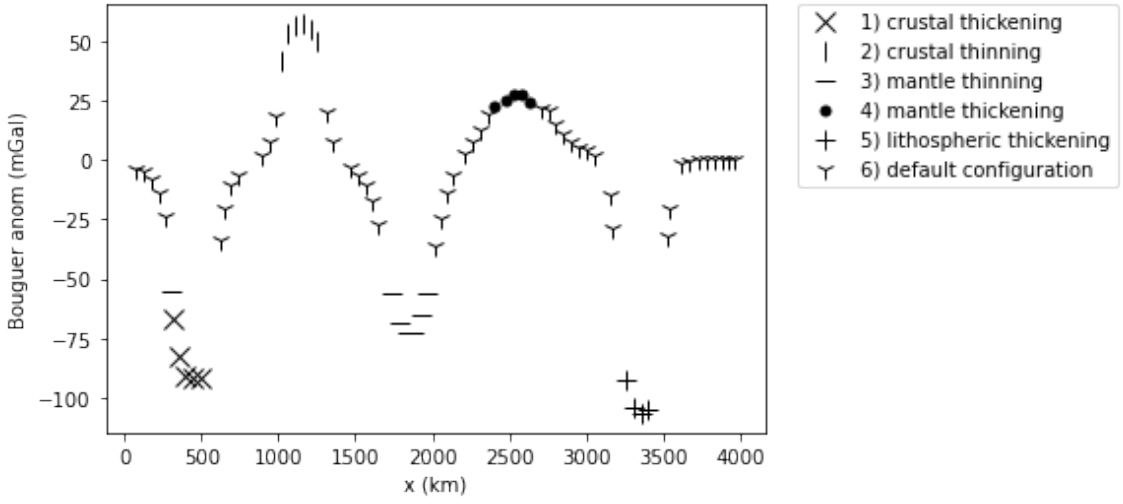

Figure 3.7: Example classification of points using a random forest that uses the Airy, Bouguer and Geoid anomalies from Fig. 1.1. In this example, one point was misclassified as mantle thinning instead of crustal thickening.

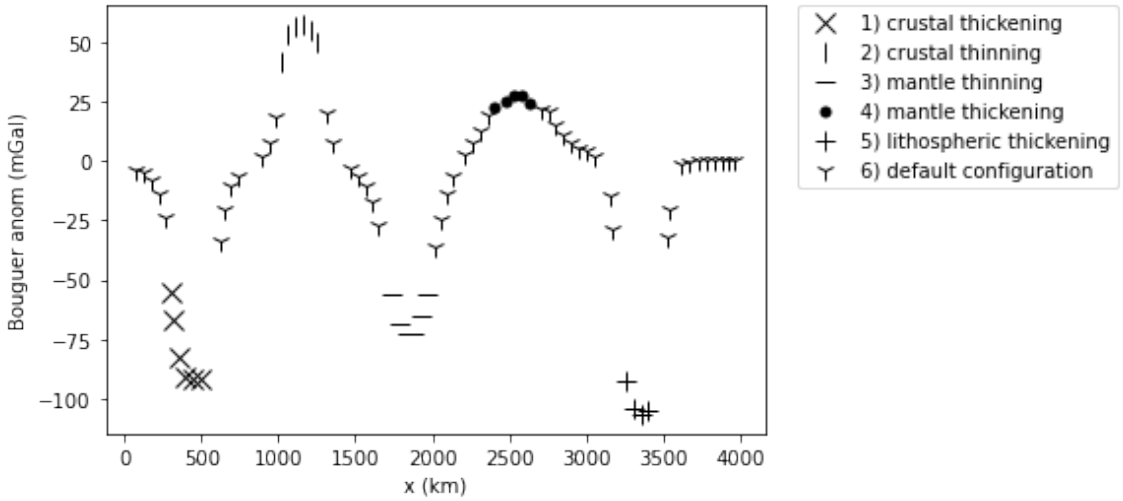

Figure 3.8: Example classification using an ANN that uses the Airy, Bouguer and Geoid anomalies from Fig. 1.1. The ANN consistently gets to 100% test and train accuracy.

test, 67% train), I trained the ANN. Within or at 200 epochs, the ANN converged to perfect accuracy, even without the use of quadratic features. In principle, an ANN can mimic these quadratic features in the first layer by learning the entire input training set as weights, which will essentially calculate convolutions of any input across the entire training set of data. Whilst this seems like a sensible way to interpret the ANN, it does not necessarily converge like this. One could initialise the weights of a particular layer as the training data if one wished to, but this will involve a larger network and thus may not benefit computation time.

Fig. 3.8 shows the classification using the ANN, which as mentioned before has 100% accuracy. Whilst there might seem to be no benefits to using quadratic features, including them reduced the average convergence time (no. epochs needed for perfect test and train accuracy) from 134 epochs to 65. The average convergence times were taken from 15 runs of the code. The introduction of the quadratic features reduces convergence times in this case by approximately a factor of 2, which is useful when considering larger and larger configuration basis sets.

Part IV

Conclusions and future work

4.1 Conclusions

To recap - this project set out to use data science methods in order to discern the crustal and lithospheric configurations, for a limited basis set of configurations generated from the tAo code [5]. This code produces 1D cross sectional values for various gravity anomalies. This project used some pre-generated gravity anomaly data and tested if an automated procedure could be developed using data science techniques.

I chose to use three classes of method to try and discern the different configurations - K-means, decision trees/random forests and an ANN. All of these methods performed with high accuracy with the introduction of quadratic features, which helped the algorithms categorise better by considering the values of two different anomalies simultaneously. However, without these extra features, the model accuracy was less, except in the case of the ANN which only converged faster with these extra features. Luckily, these features are very easy to generate and provide more information for the models to bring almost all of them to 100% accuracy.

This leads to several conclusions. Firstly, I would not have had the idea to investigate these quadratic features if not for using some of the 'less accurate but easier to interpret' models. In a vacuum, one may benefit from looking at non ANN models in order to investigate feature engineering principles, but for most people this is not feasible since one predictive model ought to be selected for most projects.

Secondly, the ANN performed the most consistently straight out of the box, but took by far the most time to train. There may be a significant number of subsurface configurations to consider if one wants an accurate classification of the subsurface. Other methods such as K-means performed reasonably well, but have an advantage in that they take less time to fit to the data. It will be interesting to see if K-means can maintain its accuracy if this configuration basis size increases.

Overall, all methods performed reasonably well, and I learnt a lot about their implementation as well as some feature engineering techniques.

4.2 Future work

The tAo code generates 1D cross sectional data [5], but naturally the Earth exists in 3D. An equivalent scientific code that can generate the configurations in a 3D space would be fantastic for this kind of project. Having this alone would be enough to move onto attempting to classify real world data.

This kind of project also can work in principle if one has missing data or only some of the anomalies calculated, it would just require training more ANNs for instance. Training on just two of the anomalies or even one of the anomalies may be feasible

to generate models for any kind of data input. Of course, one expects using as many anomalies as possible to be the best way of discerning the configurations, since more information is used by the algorithm to make the distinction.

Also, the configurations here are assumed to be static. There may be some way to involve timestamped gravity anomalies in order to determine dynamic configurations, too. This does not necessary have to involve time series analysis methods such as recurrent neural networks (RNNs) or SARIMA models, since typically RNNs and SARIMA models are used to predict the next however many timestamped data points in the future. Instead, what one would want is to take some sample of time (long enough to cover most dynamical subsurface events), and identify what kind of transition the subsurface went through to produce that change in the gravity anomaly. This idea may not have much practical real world application however, since one is not typically providing gravitometer readings during, say, an earthquake, but for slower subsurface movements there may be ample data to detect slow changes in the crust and lithosphere. Of course, one can always look at the static configurations before and after the subsurface event, provided that it does indeed settle into a static configuration!

This project aimed to generate an automated method of the inversion process. However, sometimes the anomalies produced by different configurations are degenerate [2], which means that gravity readings should be used alongside other methods for proper inversion. This project can be feasibly extended in order to include other geophysical measurements, so this presents as a non-issue but rather a consideration for the most accurate classification of subsurface configurations.

In this work, I made a simplification in that I only considered what configuration was underneath each x position as the labels for the training sets. This was convenient to do because of how the configurations were presented in Fig. 1.1, which had discontinuous, hard borders between each configuration. Obviously, proximity to other configurations causes changes to the anomalies at a single x position, even within the data from Fig. 1.1. The borders between real world configurations may not be as discontinuous.

Models such as ANNs are probably able to discern the subsurface configuration directly underneath, even with the removal of discontinuous borders between configurations. However, for more realistic configurations some clever thinking in how to decide the borders between different configurations must be decided. This can be achieved using decision boundaries. For example, if the difference in mantle thickness is above or below some decision value, then it can be labelled as being one configuration or another. In this scheme, any configuration has some distribution of thicknesses for each layer. This can be adjusted to be as coarse or as refined as one wishes. Even using a configuration schema like in Fig. 1.1 can hopefully provide a reasonable approximation to real world subsurface configurations, or at least provide better initial conditions for a more formal inversion process.

On the other hand, if configurations can be translated into a vector of outputs such

as crustal, mantle and asthenosphere thickness/density, this would be more useful classifying real world data. Both classes of outputs could be used to corroborate model results and remove ambiguity in the classification.

In general, this project can be extended using the following steps:

1. Use scientific software to generate synthetic gravity and geoid anomalies and other geophysical parameters, using a 3D model and an ample basis size. These can either be static configurations or have some dynamic behaviour.
2. Train the data on ANN(s), or equivalent. The number of predictive models needed corresponds to how much data is known for each longitude and latitude point simultaneously.
3. Use the predictive model from the step above to classify and plot real world data.

Bibliography

[1] "Introduction to potential fields: Gravity," *U.S. Geological Survey*, 1997. Online: https://doi.org/10.3133/fs23995 (accessed 13/05/2022).

[2] P. Altwegg, *Gravimetry for geothermal exploration methodology, computer programs and two case studies in the Swiss Molasse basin.* PhD thesis, 2015.

[3] J. F. Kirby, "On the pitfalls of Airy isostasy and the isostatic gravity anomaly in general," *Geophysical Journal International*, vol. 216, 2018.

[4] "Geoid - Wikipedia." Online: https://en.wikipedia.org/wiki/Geoid (accessed 22/07/2022).

[5] D. García-Castellanos, "tAo." Online: https://github.com/danigeos/tao-geo (accessed 13/05/2022).

[6] D. García-Castellanos, *Desarrollo de modelos numéricos de flexión litosférica: aplicación a fosas oceánicas y cuencas de antepaís.* PhD thesis, University of Barcelona, 1998.