Tectonic Environments of Cobalt in South America revealed through Support Vector Machines and Random Forests

Hugo Fahey (470427693)
Anna Freudenstein (490251254)
Tammy Satterthwaite (450158531)

1. Abstract

Cobalt is considered a critical mineral, which is mined mostly in one, politically unstable area. It is therefore important to explore potential alternate sources of cobalt. Cobalt is known to be associated with copper and nickel deposits occurring on subduction zones. We aim to determine which tectonic subduction zone parameters are most likely to result in the formation of cobalt deposits, as well as examining whether subduction zones can create rich cobalt deposits (>0.1% cobalt) that could potentially be exploited. In order to do this, we reconstruct plate subduction along the Atacama trench, connecting the subduction of the Nazca plate over time to a dataset consisting of the location, formation time and cobalt concentration of cobalt deposits in South America.

We then used machine learning methods to analyse different tectonic factors and use those most important to create a prediction model for the formation of cobalt deposits. We trained two supervised learning classifiers, support vector machines and random forests, on the dataset. We found that the speed of plate convergence (km/million years) is the most important factor in the formation of cobalt deposits, with the subduction obliquity, the distance to trench, and the age of the subducted seafloor also contributing. However, this subduction zone does not appear to produce rich cobalt deposits, making it unfavourable as a site of further cobalt exploration.

2. Introduction

Cobalt minerals occur in several diverse settings all of which display different examples of mineralisation. Cobalt deposits are diverse in terms of their geological setting, age, morphology, mineralogy, origin, and grade relations. Though pure cobalt is not found in nature, cobalt-bearing minerals and compounds are abundant and widespread (Slack et. al 2017). In general, most common rock-forming minerals do not contain significant amounts of cobalt where instead, cobalt-bearing minerals are included in or act as a substitute for other elements. Hence, cobalt is almost always a by- or co-product for other transition metals and is especially found in the place of iron, copper and nickel as they share similar chemical properties (Cobalt Institute).

Cobalt occurs in a wide variety of deposit types. Over 60% of the worlds cobalt mine production are found in sediment hosted ore deposits primarily found as a by-product for copper and iron (Figure 1). Being the most common cobalt deposit type, these major deposits are characterised by ore minerals contained within stratiform organic-rich pyritic shales and sandstones. Although sedimentary rock-hosted stratiform copper deposits account for the majority of cobalt production in the world, all this output occurs spatially distributed just in the Central African Copper-belt while other sedimentary rock-hosted stratiform copper deposits throughout the world contain only minor amounts of cobalt (Hitzman et al. 2017).

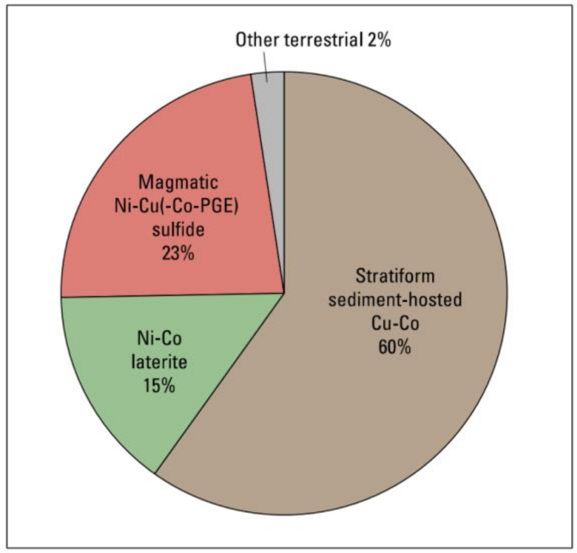

Figure 1. Pie chart showing the percentage of world cobalt mine production by deposit type. Other terrestrial deposits are grouped together and include iron oxide copper gold (IOCG) deposits, polymetallic (Ag-Ni-Co-As-Bi) and other cobalt rich vein deposits, and volcanogenic massive sulphide deposits. Sourced from Slack et al. (2017)

Around 15% of cobalt deposits are found in Ni-Co laterites where weathering of cobalt silicates and sulphides in ultramafic rocks remobilises and enriches cobalt as a by-product in residual weathered rocks (Hitzman et al. 2017). Grades of cobalt content in laterites varies from 0.1-1.5 % (*Cobalt Institute*).

Nearly 25% of cobalt deposits occur in magmatic Ni-Cu sulphide deposits where cobalt is a by-product of sulphides in Ni-Cu mafic-to-ultramafic intrusions (Hitzman et al. 2017). Found as trace amounts with a grade content of just 1.0 %, many of these deposits are very old occurring in rocks of Proterozoic and Archean age.

The remainder of terrestrial cobalt deposits occur in hydrothermal and volcanogenic deposit types. These can include iron oxide copper gold (IOCG) ore deposits, epigenetic Co-Cu-Au deposits, five element (Ni-Co-Ag-As-Bi) vein deposits, and a spectrum of synsedimentary and (or) diagenetic deposits. Though these attribute to a range of deposit styles and mineralisation processes, the key process is precipitation from hydrothermal activity passing through the host rock due to volcanogenic activity (Cobalt Institute). These cobalt deposits can be found along fault planes, fissures, veins, and cracks restricted to areas of hydrothermal activity associated with ultramafic rocks in areas of volcanic activity (Hitzman et al. 2017). These deposit types, however, occur in areas where little exploration has been done due to poorly defined exploration models and due to the difficulties in accessibility.

The largest known quantities of cobalt occur on the seafloor, contained within manganese nodules and cobalt-rich crusts. They are however not economically viable with current technology and economic conditions.

This study focuses on cobalt deposits within South America though a global assessment is also reviewed. Though cobalt deposits are widespread throughout the continent as will be assessed within the report, there are only two locations where cobalt is mined in South America. These both occur in nickel ore mines in Brazil where cobalt is mined and refined as concentrations in these Ni-Co laterite ore deposits (Shedd 2015).

South America has an extended and diverse geological history as seen in Figure 2. Once a part of the supercontinent Pangaea, South America split from Africa on its western edge around 130 Mya and separated from Antarctica from its south approximately 50 Mya (Alden 2018). South America currently has two major landforms. The western half is comprised of the Andes Mountains formed from the subduction of the Nazca Plate underneath the entire western edge of the South American Plate. Subduction has been constantly occurring over the last 200 million years (Rosenbaum et al. 2005). This active margin is prone to volcanic activities. The eastern half of the South American continent is comprised of several underlain cratons billions of years in age. In-between these cratons and the Andes are sedimentary basins (Alden 2018). The eastern half of the continent in juxtaposition is stable and has been unaffected by mountain building events for hundreds of millions of years attributing to millions of years of erosion.

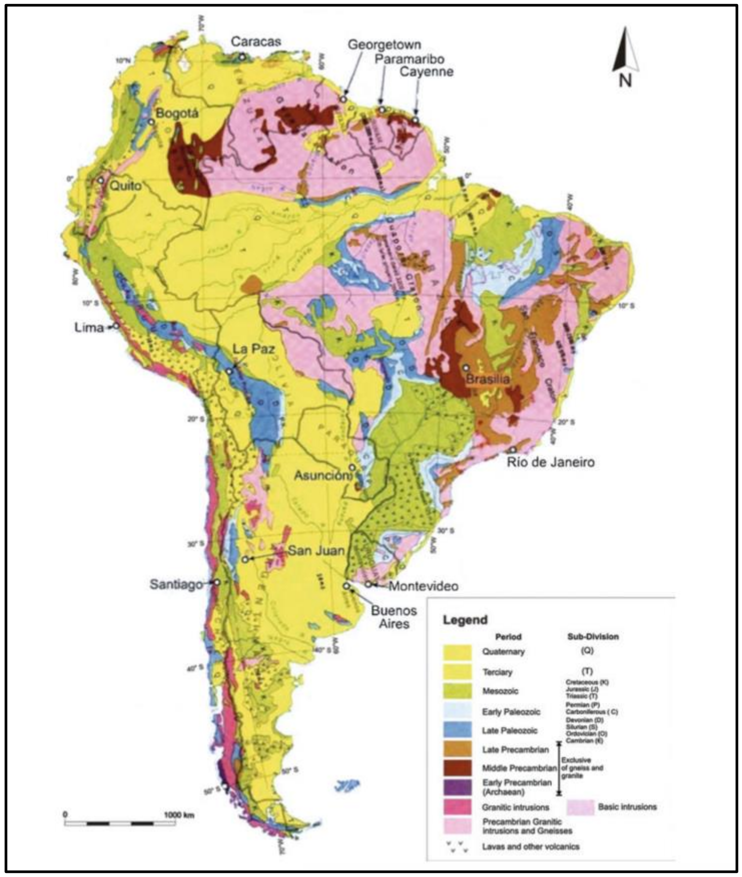

Figure 2. Geological map of South America
Source: Acevedo, et al. (2015)

The importance of cobalt is significant as it is used in numerous diverse commercial, industrial and military applications (Hitzman et al. 2017). Cobalt is fundamental to industry and essential for enabling technological development corresponding with a low carbon future (Cobalt Institute). The leading use of cobalt is for rechargeable battery electrodes as well as for superalloys which are used to make gas turbine engines. In addition to being strategic and critical for industrial uses, cobalt is also a central component for vitamin B12 which is vital for growth and vitality (Cobalt Institute and Hitzman et al. 2017).

Machine learning allows the exploration of ore deposits to be presented in a global context. Technology and open-source web applications have allowed the global scientific community to share and build on data while programming allows data to be manipulated and analysed. Programming is important for speeding up the input and output of data and to create innovative solutions and interpretations for global queries.

3. Data Set

Our dataset of 185,856 global cobalt deposits was sourced from the *EarthChem* database. This was after data cleaning , as before the correction of latitude and longitude contained 185,871 points. The data set documents the latitude, longitude, age and concentration of cobalt (parts per million) in each deposit. Of these, we restricted our analysis to the South American data points with a longitude between -90° and -25°, and a latitude between 15° and -60°, leaving us with a dataset with 4,919 points. These deposits formed between 1 and 4,940 million years ago. We then analysed the distribution of economically viable rich deposits (> 1000 ppm) in South America, restricting the dataset further leaving two deposits.

The majority of the global dataset came from the *United States Geological Survey* (USGS) sample database, explaining why the dataset has such a high concentration of points in North America as compared to other locations. Due to the higher input of sampling from the United States, the *EarthChem* cobalt data set is not representative of the worldwide distribution of cobalt deposits. It is bias in higher inclusion of deposits from within the United States. Specifically in our restricted South American data, most of the points came from the *Geochemistry of Rocks of the Oceans and Continents* (GEOROC) database, based in Germany. This data is collected from published papers and then compiled on GEOROC.

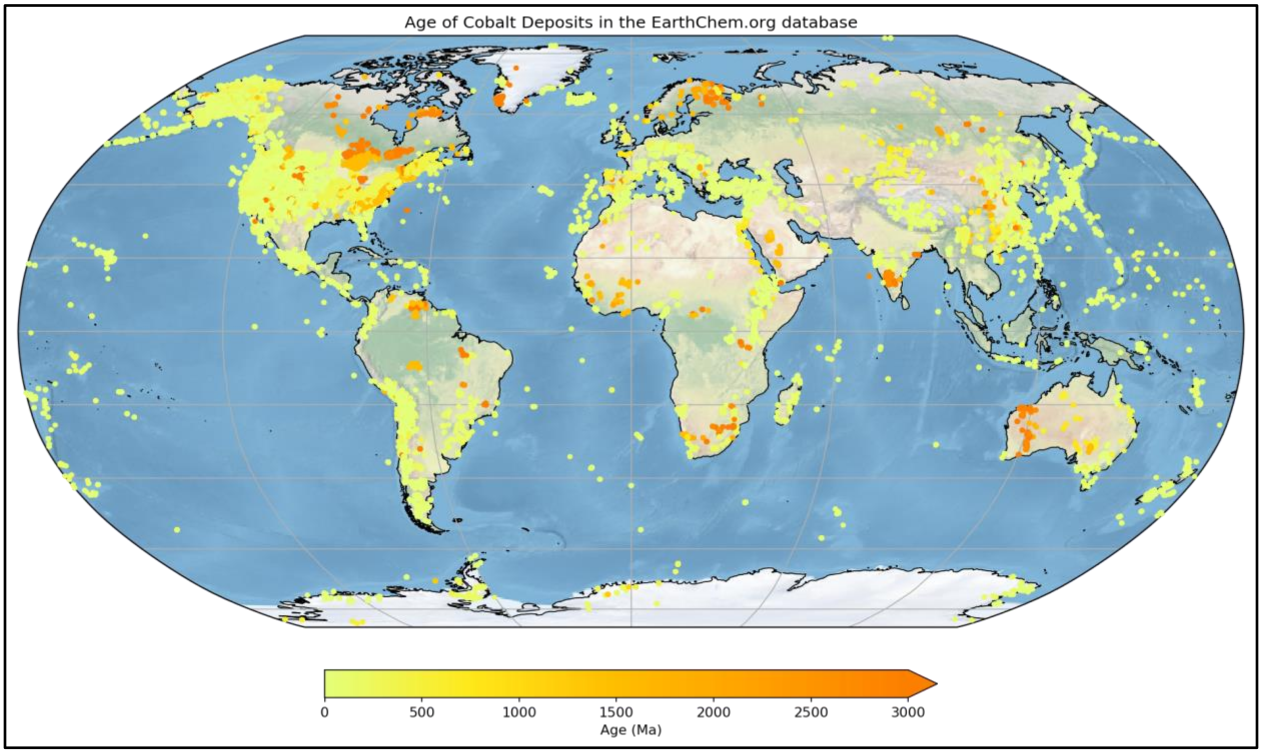

Figure 3. Map of cobalt deposit data from *EarthChem.org* database, coloured by age of the deposit. Created with python, utilising *matplotlib* and *cartopy* python open source libraries.

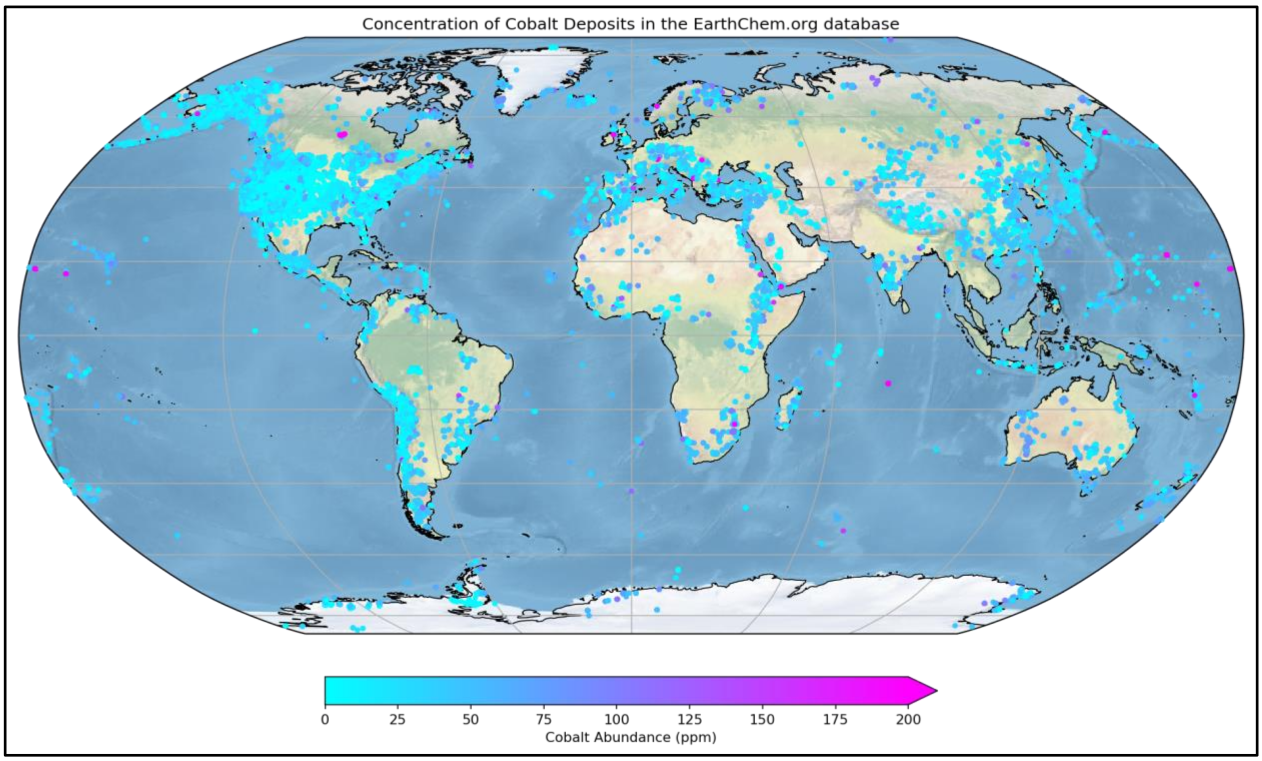

Figure 4. Map of cobalt deposit data from *EarthChem.org* database, coloured by abundance of cobalt in the deposit. Created with python, utilising *matplotlib* and *cartopy* python open source libraries.

There are three major groupings of ages of cobalt formation; approximately three billion years ago, just prior to two billion years ago, and within the past 100 million years, over which about 60% of the deposits formed (Figure 5-6a). The vast majority of cobalt occurs in small concentrations. The world wide data set has a median deposit concentration of 17.0 ppm (Figure 5b). The data restricted to South America follows the same general age and concentration trends as the world wide data.

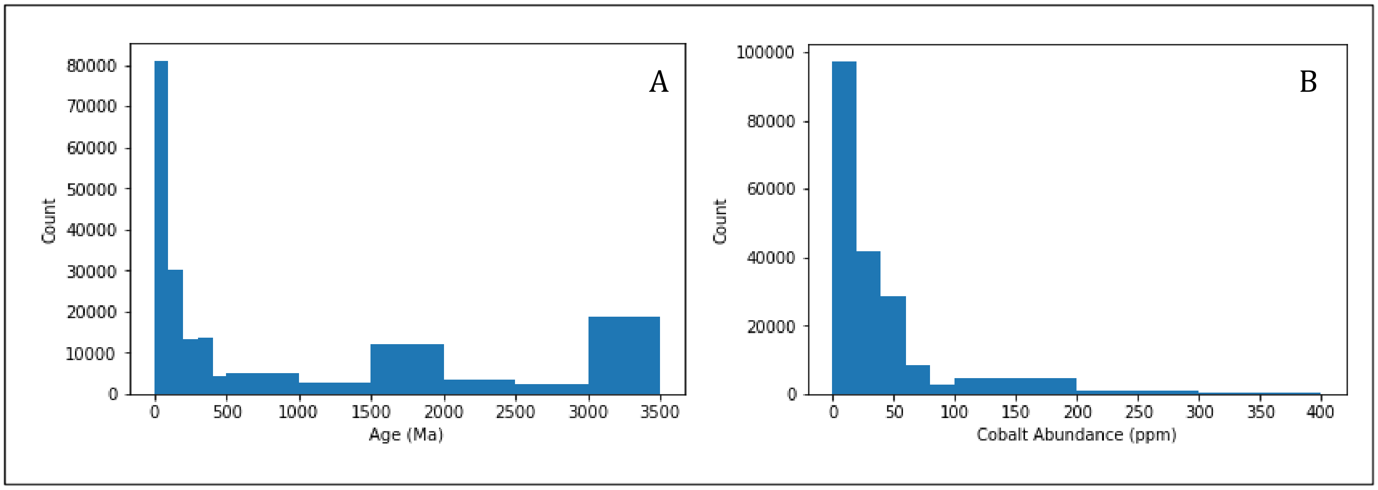

Figure 5. Histograms of world-wide cobalt deposit data from *EarthChem* database
A: Age of cobalt deposits
B: Concentration of cobalt within the deposits

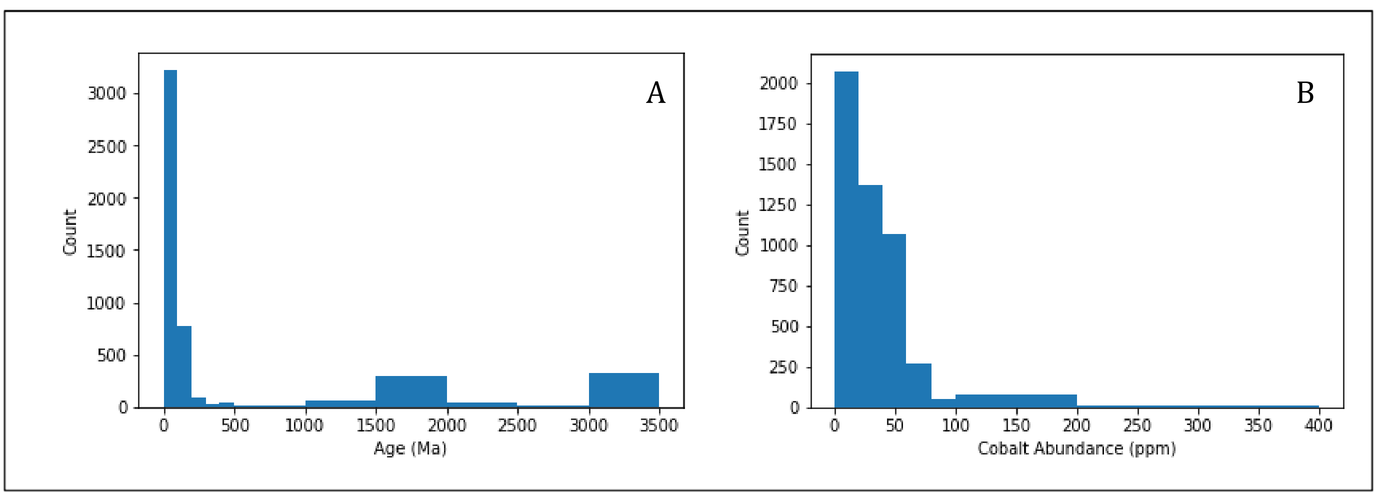

Figure 6. Histograms of cobalt deposit data restricted to South America, from *EarthChem* database.
A: Age of cobalt deposits
B: Concentration of cobalt within the deposits

4. Methods

4.1 Plate Reconstruction

We reconstructed plate movement using the python library of the *GPlates* (www.gplates.org) functionality. This let us correlate the ages of cobalt deposits from our *EarthChem* dataset with the plate dynamics of the South American subduction margin over 230 million years. The relative plate motion data is taken from plate circuit calculations from parameters in (Seton et al. 2012). This paper also serves as the source for our palaeo-age grids for the subduction. The angle of subduction was taken from the angle between the lateral direction of the trench and the direction of convergence (Butterworth et al. 2016).

The plate model data was discretized into sections every 1 million years and 0.5° along the trench. Each deposit was then linked to the closest spatiotemporal plate section and its associated parameters (i.e. convergence rate and subduction obliquity. For full parameter list see Data Analysis). This then allowed us to use machine learning methods to see if any relationship exists between parameters associated with plate movement and the formation of cobalt deposits (see Machine Learning Results).

4.2 Data Analysis

Various open source python libraries were utilised allowing for processing and depiction of the data. These included; *NumPy* - for mathematical data manipulation, *pandas* - for manipulating numerical tables and time series, *matplotlib* - for plotting figures, *cartopy* - for plotting latitude and longitude on a global map, *SciPy*, *Pickle*, and the aforementioned py*GPlates*.

We took two separate approaches to classifying the data for analysis. Initially we considered each separate data point as a distinct event in itself, which we then used to determine the tectonic parameters for each point individually.
Our second approach involved partitioning the entire region east of the subduction margin into 500 km segments and 10 million year time intervals. We have introduced this method in order to compensate for the likely presence of as-yet undiscovered cobalt deposits, as well as to reduce overinterpretation of clustered data points that are in fact part of the same formation event.

We focused our analysis on seafloor age, distance to trench edge, convergence rate and subduction obliquity, as these are the parameters that our machine learning analysis found to be most significant for the formation of cobalt deposits (see Figure 7 and Machine Learning). Although subducting plate normal velocity (km/million years) has a higher rating in this analysis than the subduction obliquity (°), we found this concept to be similar enough to the convergence rate that we decided to disregard the subducting plate normal velocity in further analysis, instead focusing on convergence rate (which our machine learning analysis rated as more important than subducting plate normal velocity) and subduction obliquity (the next highest rating after subducting plate normal velocity) for the purposes of our analysis.

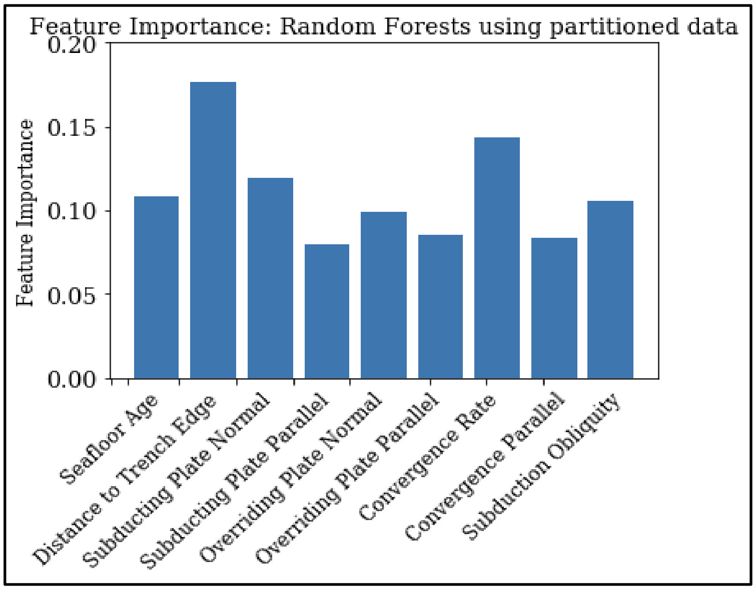

Figure 7. Graph displaying the most important features of the subduction zone to cobalt deposit formation, according to Random Forests analysis

We created a series of semi-randomised non-deposit data points which are examples of tectonic circumstances that did not result in cobalt being deposited. These serve as a comparison point for our deposit dataset, and were also used in training the machine learning algorithms (4.3 Machine Learning methods). For point of formation analysis, non-deposit points were set at the same latitude and longitude as deposits, but with a different formation time (0 - 230 million year age range). For partitioned data analysis, all grid squares containing one or more deposits are labelled a deposit partition, while all grid squares containing no deposits are labelled a non-deposit partition.

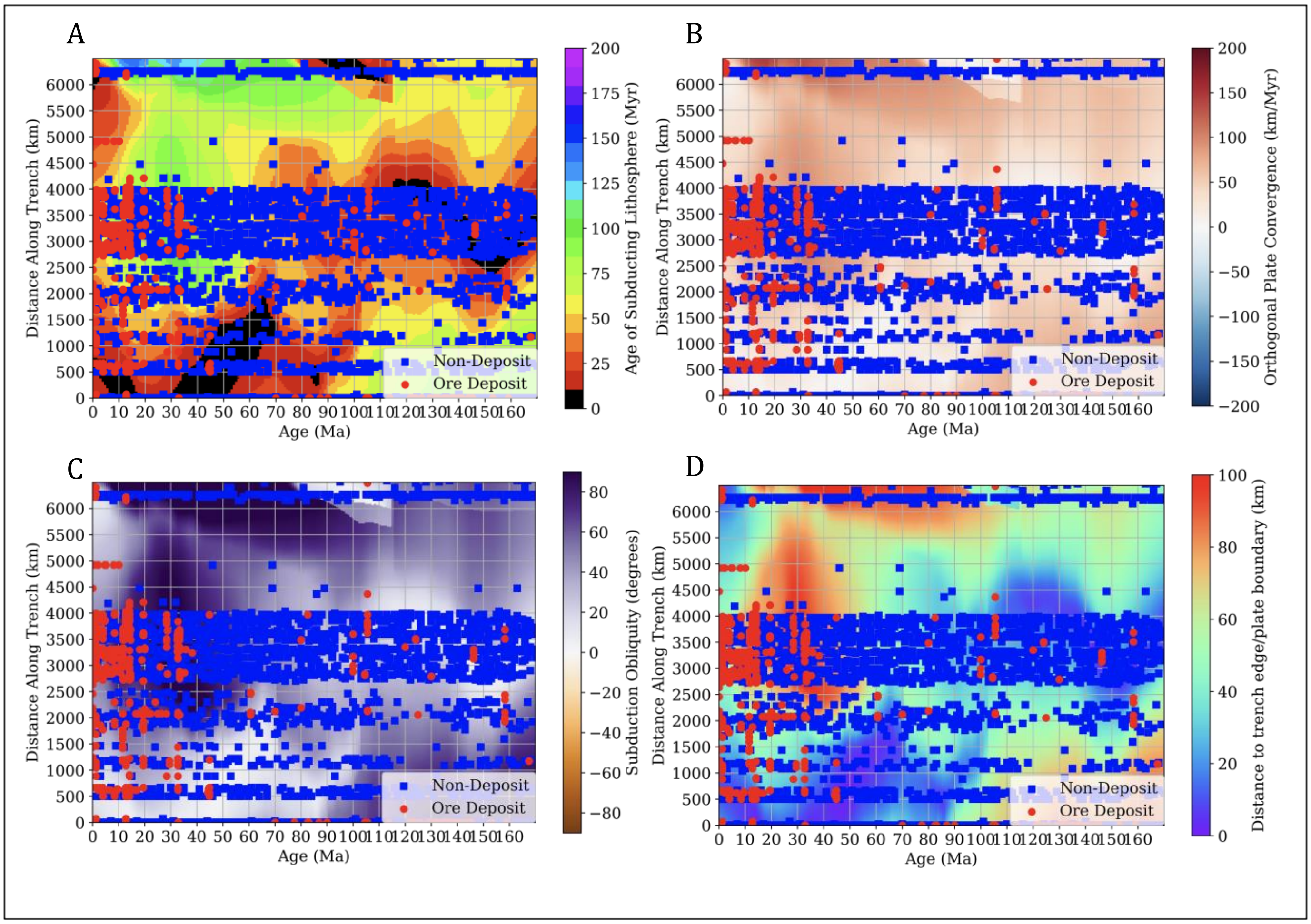

Figure 8: Graphs showing the four most important tectonic parameters: (A) Age of subducted lithosphere, (B) Orthogonal plate convergence rate, (C) Subduction obliquity, (D) Distance to plate boundary, and how they have changed over time on a spatiotemporal plot of the cobalt dataset. The blue and red dots show the non-deposit and deposit point of formation data, respectively. The grid on the graph shows the squares of the partitioned data. Grid squares containing a red deposit dot are taken as deposit regions, while those without are taken as non-deposit regions.

As many of our cobalt data points were for deposits with economically non-viable amounts of cobalt, as the overall average concentration (median) of the South American deposits was 17 ppm, or 0.0017%, we attempted to create a dataset consisting only of points in our dataset with a grade of > 0.1% (or 1000 ppm, the average cut-off grade for economic cobalt deposits). However, upon analysing this dataset we realised that only two cobalt deposits in the entirety of South America contained enough cobalt to be economically viable, and these deposits were not associated with subduction zones (Figure 9). Although we consider this insight valuable and worthy of discussion (see Discussion), we decided not to carry out further analysis, such as machine learning analysis, on this dataset as it was too small to produce statistically significant results.

Figure 9. Map of locations of cobalt deposits with a grade of >0.1% in South America

4.3 Machine Learning

Machine learning analysis allows us to confirm that the observations we made of relationships between subduction parameters and cobalt deposit formation in our data analysis are causal.
We considered two methods for our analysis, random forests and support vector machines (SVM). These methods are both supervised learning 'classifiers', which we trained on a sample subset of our data, and then tested on another subset. They then produced predictions which we could compare against the existing data.

SVM is a nonlinear binary classifier which can be used to make probabilistic classifications using the Platt scaling method. The kernel trick allows SVM to project the tectonic parameters nonlinearly into high dimensional space by transforming each dot product into a kernel function, allowing nonlinear data, such as our deposit-nondeposit dataset, to be separated along a "linear" (projected) boundary into binary classes. Although the kernel trick can increase the error of the SVM when examining the test data, our dataset was large enough that the algorithm still produced acceptable results (see Table 1). We used a nonlinear radial basis kernel as the foundation for our SVM, as used in Butterworth (2016) and Akbani et al. (2004).

Random forests is a machine learning method based on combining multiple decision trees, which assess the tectonic parameters, the likelihood of a particular parameter resulting in a deposit, and then produce a result which essentially shows the percentage of trees which 'voted' for a particular class. This gives us a result for the relative importance of each individual parameter, which cannot be assessed as easily using SVM (see Figure 12). Our program runs using the scikit-learn Python package Pedregosa et al. (2011).

In order to decide which of the two machine learning methods to use, we executed a fivefold cross validation process, where the data is split into five sets of 80% for the training subset and 20% for the testing subset. The average accuracy of the predictions of the respective machine learning methods are shown in Table 1. Using this method, we determined that both SVM and random forests produced predictions better than 0.5, so we elected to continue with both of these classification methods.

	Cross validation score	
Machine Learning Method	**Point of Formation Data**	**Partitioned Data**
Random Forests	0.79∓0.03	0.68∓0.12
SVM	0.77∓0.04	0.75∓0.08

Table 1. Table showing the results of the five-fold cross validation process, a method to estimate the accuracy of machine learning predictions. The closer the score is to 1, the better the machine learning method was able to predict the testing subset using the information from the training subset. A score of 0.5 would indicate that the machine learning did not predict better than random selection would.

5. Results

5.1 Data Analysis

As seen in Figure 10, younger deposits typically aged 50 Ma to present are located in a condensed cluster on the western margin of South America down to the south of the region. Within this heavy cluster of cobalt deposits, there are some older deposits ageing up to 200 Mya located around the main cobalt range. In comparison, older cobalt deposits aged between 75 Ma to 150 Ma are located on the eastern margin of South America in a more dispersed manner. Deposits of this age are also seen to be located in the north-west region of the continent and older cobalt deposits aged around 200 Ma are found towards the north.

Analysis of the cobalt dataset using histograms evaluating the influence of the four main tectonic parameters (Figure 11A) shows that cobalt deposition at this subduction zone peaks when the subducting slab is comparatively young (<30 million years old), with another peak around 50-60 Mya. Figure 11B shows there is a strong correlation between cobalt deposit formation and distance from the subduction zone itself. Cobalt deposits cluster within 100 km according to the histogram below. A peak in deposits between 2500 and 3000 km from the trench is likely an outlier, caused by older deposits on the east coast of South America that are not associated with the subduction zone (see Figure 10). The speed of subduction also has influence (Figure 11C), with a slower subduction speed (0 - 50 km/million years) resulting in the greatest number of deposits forming. Finally, a flat angle of subduction (0 - 40°) seems to be the most advantageous for deposit formation, although the correlation is not as clearly delineated as for the other parameters (Figure 11D).

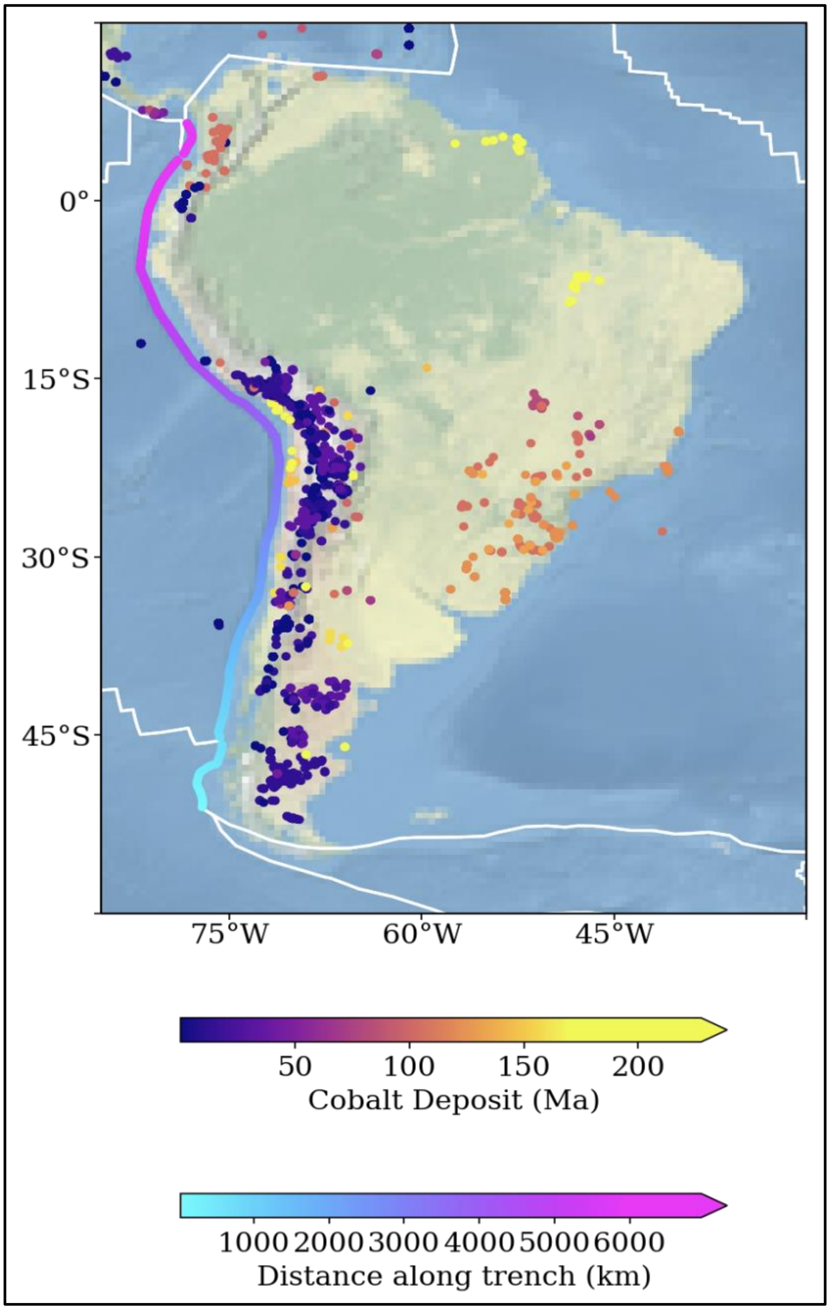

Figure 10. Map of cobalt deposits in South America, data from *EarthChem*. Displaying cobalt deposit age by colour

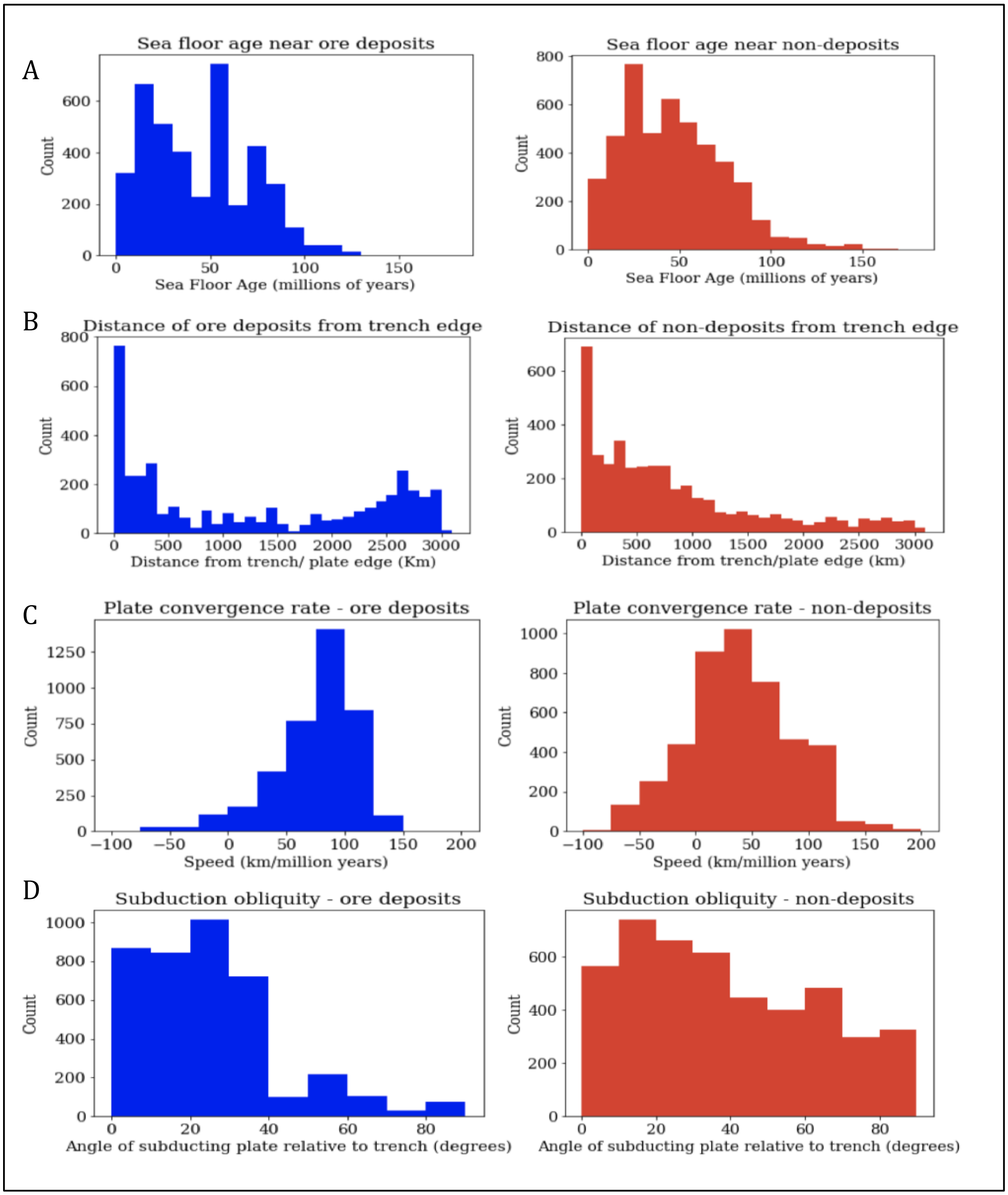

Figure 11. Histograms of each parameter, used in the machine learning, for the point of formation data for both the ore deposits and the random non-ore deposits. These histograms show the difference between the non-deposit and deposit data as it pertains to the four subduction parameters that we are examining in depth. Non-deposit plots (red) show the general trend of the subducting slab, while peaks in the deposit plots (blue) show where this trend caused a spike in deposit formation.

A) Histogram showing seafloor age (millions of years) against deposit count. We see a distinct peak in deposit formation when the subducting slab is 20-30 million years old, and another peak when the slab is between 50 and 60 million years old. B) Histogram showing distance of ore deposits from the plate edge (km) against deposit count. We see a distinct peak in deposit formation within 100 km of the plate edge. C) Histogram showing plate convergence rate (km/million years) against deposit count. Here we can see that a speed of subduction between 75 and 100 km/million years is most advantageous for deposit formation. D) Histogram showing subduction obliquity (°) against deposit count. This histogram shows that a flatter angle of subduction appears to correlate to more deposits forming (0-40°).

5.2 Machine Learning

The results for random forest analysis, showing the relative importance of four parameters, produce similar results for the different data classification methods. The convergence rate is rated as the most important factor in determining the spatiotemporal emergence of cobalt deposits, followed by the subduction obliquity, the distance to trench and the age of the subducted seafloor, which all have similar, and lesser, importance.

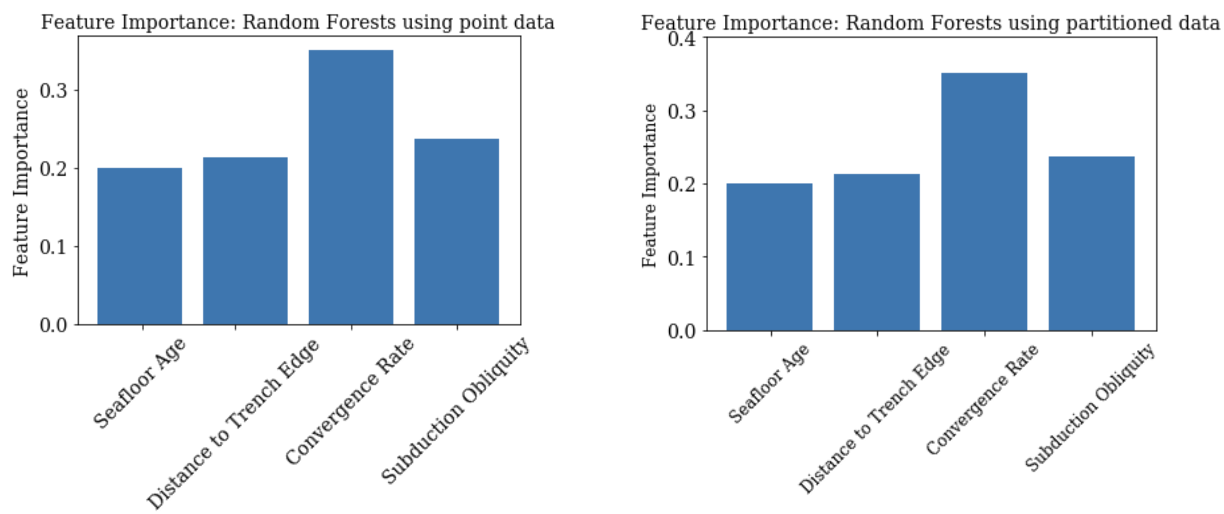

Figure 12. Graphs showing the relative importance of four tectonic parameters to the formation of cobalt deposits, using the different data classification methods.

The probability of a deposit forming at a particular spatiotemporal point according to the factors depicted in Figure 12 is shown in Figure 13. This also presents a comparison with the existing data.

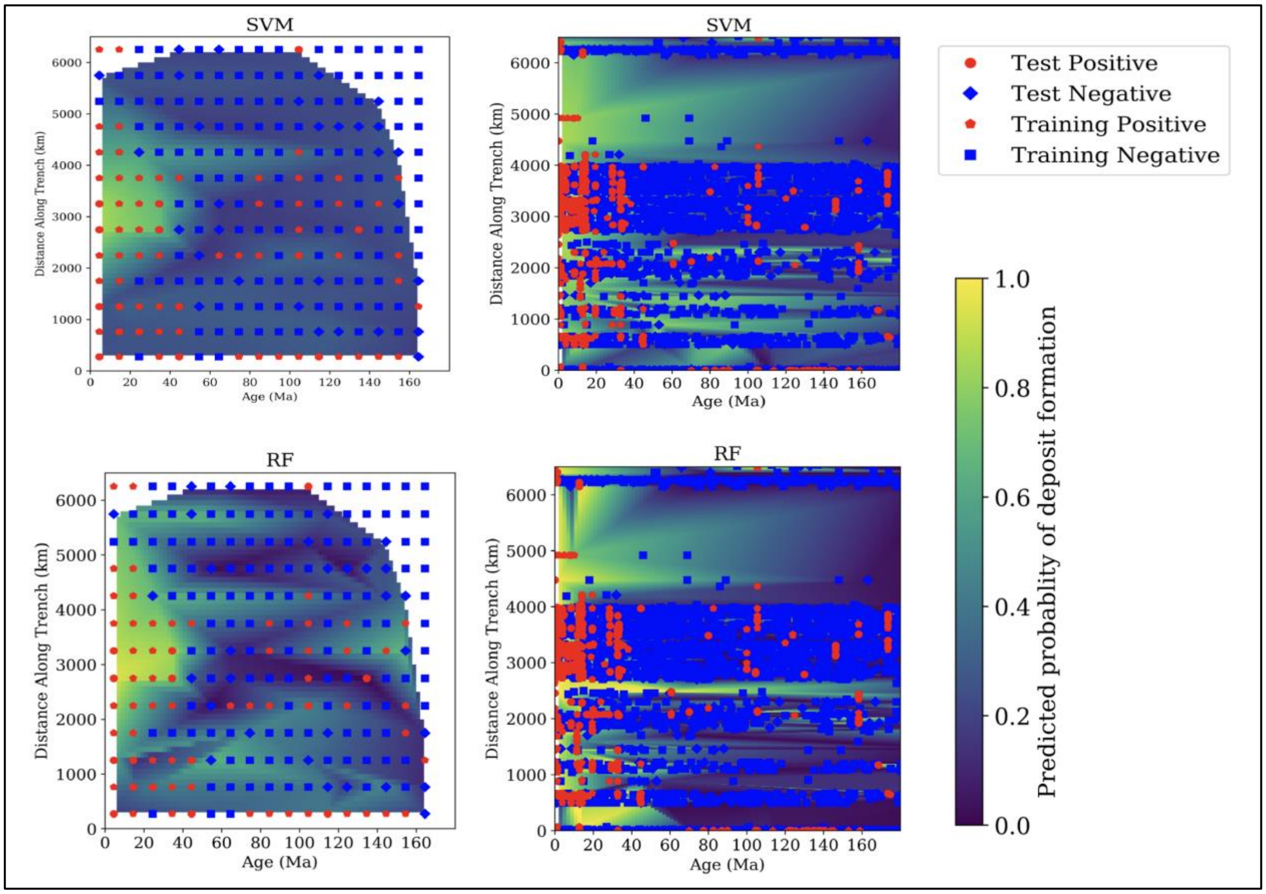

Figure 13. Diagrams showing the spatiotemporal probability of cobalt deposit formation, according to Support Vector Machines (SVM) and Random Forests (RF) analysis. The left hand column shows the results for partitioned data classification, while the right hand column shows the results for point of formation data classification. The training and testing sets (see Methods, section TBD) are shown as red circles and blue triangles for the testing set, and red pentagons and blue squares for the training set.

6. Discussion

In order to determine the relationship between subduction and cobalt deposits we investigated cobalt deposit data across South America; including latitude, longitude, age and cobalt concentrations. This was related to tectonic parameters of the subduction setting of the Nazca plate under the western margin of the South American plate. We looked at sea floor age, segment length, slab length, distance of the deposit to the trench edge, subducting plate velocity, overriding plate velocity, convergence rate, subduction obliquity and distance of the deposit along the margin. Initial machine learning testing using the random forest method (Figure 7) found sea floor age, distance from the plate boundary, plate convergence rate and subduction obliquity to be the dominant tectonic parameters in the prediction of cobalt ore deposit location. As such, these four parameters were incorporated into data analysis and machine learning methods.

As seen in Figure 11, younger deposits of cobalt are located on the western margin of South America correlating with the tectonic activity occurring in the same region. The forces that have been pushing the Andes up have been at work for 140 million years with subduction related volcanic processes still occurring today. As reviewed, subduction is a favourable process for forming ore deposits. Hydrothermal activity during tectonic convergence along the eastern continental margin has created porphyry copper and orogenic gold ore deposits containing trace amounts of cobalt by-products. Found in fault planes, fissures, veins and cracks associated with volcanic activity, the traces of cobalt by-products are formed from precipitation processes of hydrothermal activity passing through the host rock (Hitzman et al. 2017). The active continental margin on the western margin has produced the majority of known cobalt deposits in South America. As seen in Figures 10 and 11B, this heavy cluster of cobalt deposits on the western border contain low abundance levels of cobalt concentration. Though geographically abundant, the concentrations themselves are not rich.

After running an analysis on a restricted dataset looking at cobalt-rich ore deposits in South America, it was revealed that there are only a total of two known deposits with a cobalt grade content of 0.1% or higher in the region out of 4,919 known cobalt deposits. This correlates with the fact that there are only two locations of mined cobalt in South America, located in nickel ore deposit mines in Brazil. This supports our finding that subduction zones are not environments where enriched cobalt deposits are formed.

Sea floor age data for the cobalt deposits in South America displays two peaks of cobalt deposit occurrence; derived from sea floor 11-20 and 51-60 million years old. Older seafloor is thicker and denser and can become more enriched in metals as it ages due to the increased material building up on the seafloor via deposition from the above ocean. In this way, older sea floor can be associated with more abundant metal ore deposits. However, as the majority of cobalt deposits occur in small concentrations there is no need for older seafloor to be the source of the melt forming the ore deposits. The younger sea floor 11-20 million years old could result in lower concentration cobalt deposits compared to the sea floor over 50 million years old.

Majority of the cobalt deposits occur within 500 km of the plate boundary, with an additional cluster occuring 3000 km from the plate boundary. The ore deposits furthest away from the active subduction boundary are not associated with magmatism driving formation of ore deposits. The clusters of deposits located in a band close to the trench arise from hydrothermal activity driven by partial melting of the subduction wedge.

A study by Bertrand, Guillou-Frottier and Loiselet (2014) found association between relatively fast convergence rate of plates in a subduction zone followed by rapid reduction in convergence and increased partial melt bearing metals resulting in formation of porphyry type ore deposits. Our data analysis (Figure 11C) finds increased occurrence of cobalt deposits in conjunction with plate convergence rates of higher than 50 km/ million years. This corresponds to Bertrand, Guillou-Frottier and Loiselet (2014), with increased subduction rate drives increased melt production, driving formation of associated ore deposits containing some amount of cobalt.

Our subduction obliquity parameter associated angles of the subducting plate below 40° with occurence of more cobalt deposits. Low angle subduction allows for increased concentration of melt in the shallow crust promoting formation of ore deposits close to the surface (Rosenbaum et al. 2005).

In the studied South American subduction setting, higher abundance of cobalt deposits occur with increased partial melt due to relatively fast plate convergence and shallow angle subduction placing that melt closer to the surface, as opposed to a higher angle subduction setting where the melt would be much deeper.

With data and understanding of these subduction setting parameters and how they change both spatially and temporally, predictions can be made as to the occurrence of cobalt deposits linked with subduction driven magmatism.

Our implementation of machine learning produced locations across South America where subduction driven cobalt deposits are likely to occur given our primary subduction parameters of; sea floor age, distance from the plate boundary, plate convergence rate and subduction obliquity. Figure 13 displays reconstructed positions in space and time of where cobalt deposits likely have formed in South America. We utilised both random forest and SVM machine learning techniques to categorize the data and predict ore deposits given our primary subduction parameters. The accuracy of these two machine learning methods in predicting ore deposits and distinguishing from non-deposits are both much higher than would be expected from a random method (Table 1). Similar methods of applied machine learning based off physical parameters of a geological tectonic setting could be utilised in the exploration of new ore deposits.

Further study should examine other subduction settings and incorporate broader data sets of world wide cobalt deposits in addition to other metal ores. Additionally, parameters driving magmatic activity at subduction zones, such as temperature gradient through the crust, could be included for a more well rounded approach to the prediction of ore deposit formation due to subduction driven processes.

7. Conclusion

Though many cobalt deposits have formed in close proximity to the Atacama Trench, and although the subduction zone is responsible for many factors that appear to be associated with the formation of deposits, these deposits have extremely low concentrations of cobalt. This is likely due to cobalts' association with Ni-Cu deposits, that often form as the result of intrusions at subduction zones. Cobalt is therefore generally only present in trace amounts in this tectonic setting. Our findings conclude that subduction zones are not environments where enriched cobalt deposits are formed, making it unfavourable as a site of further cobalt exploration. Further study of cobalt deposits in different tectonic settings could give us more insight into cobalt deposition processes, as well as open new avenues for cobalt exploration.

<u>References</u>

Acevedo, R. D., Rocca, M. C. L., Ponce, J. F. and Stinco, S. G. (2015). Introduction. In: Impact craters in South America, *SpringerBriefs in Earth System Sciences: South America and the Southern Hemisphere*, pp. 1-5, https://doi-org.ezproxy1.library.usyd.edu.au/10.1007/978-3-319-13093-4_1

Alden, A. (2018). 'A look at South American geology', *Thoughtco*, Published 14 August 2018, https://www.thoughtco.com/south-american-geology-1441058

Bertrand, G., Guillou-Frottier, L. and Loiselet, C. (2014). Distribution of porphyry copper deposits along the western Tethyan and Andean subduction zones: Insights from a paleotectonic approach. *Ore Geology Reviews*, 60, pp.174-190,
https://doi.org/10.1016/j.oregeorev.2013.12.015

Butterworth, N., Quevedo, L., Morra, G., and Müller, R. (2012), Influence of overriding plate geometry and rheology on subduction, *Geochem. Geophys. Geosyst.*, 13, Q06W15, https://doi.org/10.1029/2011GC003968.

Cobalt Institute. (2017). Production and Supply, online source: https://www.cobaltinstitute.org/

Rosenbaum, G., Giles, D., Saxon, M., Betts, P., Weinberg, R. and Duboz, C. (2005). Subduction of the Nazca Ridge and the Inca Plateau: Insights into the formation of ore deposits in Peru. *Earth and Planetary Science Letters*, 239(1-2), pp.18-32,
https://doi.org/10.1016/j.epsl.2005.08.003

Hitzman, M. W., Bookstrom, A. A., Slack, J. F. and Zientek, M. L. (2017), Cobalt – styles of deposits and the search for primary deposits, *U.S. Geological Survey Open-File Report 2017-1155*, pp. 1-47, https://doi.org/10.3133/ofr20171155

Pedregosa, F., et al. (2011), Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.*, 12, 2825– 2830.

Seton, M., et al. (2012), Global continental and ocean basin reconstructions since 200 Ma, *Earth Sci. Rev.*, 113, 212– 270,
https://doi.org/10.1016/j.earscirev.2012.03.002

Shedd, K. B. (2015). Cobalt (advance release), *U.S. Geological Survey Minerals Yearbook – 2015*, pp. 19.1-19.19, online source: https://s3-us-west-2.amazonaws.com/prd-wret/assets/palladium/production/mineral-pubs/cobalt/myb1-2015-cobal.pdf

Slack, J. F., Kimball, B. E. Shedd, K. B. (2017). Cobalt, Chapter F of: *Critical Mineral Resources of the United States – Economic and Environmental Geology and Prospects for Future Supply: U.S. Geological Survey Professional Paper 1802*, pp. F1-F40, https://doi.org/10.3133/pp1802F

Acknowledgements

This work is based off the code of Butterworth et al. (2012). We thank Dr. Nathaniel Butterworth, Dr. Rohitash Chandra and Prof. Dietmar Muller for their contribution and guidance.

Appendix

For further insight, attached is our Github notebook containing complete code scripts and associated files; https://github.com/afreudenstein/3888cobalt