

## RESEARCH ARTICLE

10.1002/2016TC004289

## Key Points:

- Time-dependent plate reconstruction kinematics applied to South American porphyry copper deposit data
- Machine learning classification used to identify mechanisms for porphyry copper magmatism
- Subducting slab age, convergence rates and directions, and trench extent control porphyry genesis

## Supporting Information:

- Supporting Information S1
- Data Set S1
- Data Set S2
- Data Set S3
- Data Set S4
- Software S1
- Software S2
- Movie S1
- Figure S1
- Text S1

## Correspondence to:

N. Butterworth,  
natbutter@gmail.com

## Citation:

Butterworth, N., D. Steinberg, R. D. Müller, S. Williams, A. S. Merdith, and S. Hardy (2016), Tectonic environments of South American porphyry copper magmatism through time revealed by spatiotemporal data mining, *Tectonics*, 35, 2847–2862, doi:10.1002/2016TC004289.

Received 2 JUL 2016

Accepted 31 OCT 2016

Accepted article online 3 NOV 2016

Published online 9 DEC 2016

## Tectonic environments of South American porphyry copper magmatism through time revealed by spatiotemporal data mining

N. Butterworth<sup>1</sup>, D. Steinberg<sup>2</sup>, R. D. Müller<sup>1</sup>, S. Williams<sup>1</sup>, A. S. Merdith<sup>1,2</sup>, and S. Hardy<sup>2</sup>

<sup>1</sup>EarthByte Group, School of Geosciences, University of Sydney, Sydney, New South Wales, Australia, <sup>2</sup>Data61, CSIRO, Sydney, New South Wales, Australia

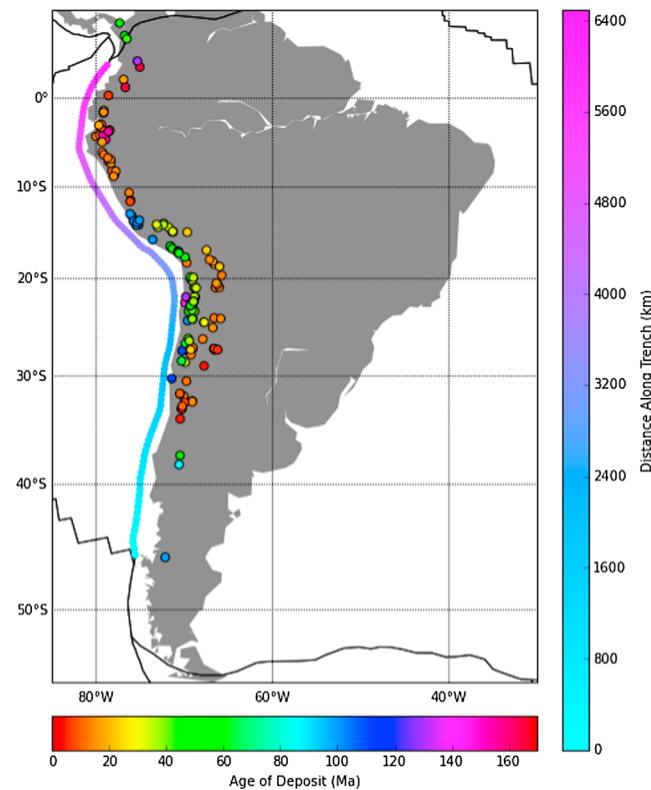
**Abstract** Porphyry ore deposits are known to be associated with arc magmatism on the overriding plate at subduction zones. While general mechanisms for driving magmatism are well established, specific subduction-related parameters linking episodes of ore deposit formation to specific tectonic environments have only been qualitatively inferred and have not been formally tested. We develop a four-dimensional approach to reconstruct age-dated ore deposits, with the aim of isolating the tectonomagmatic parameters leading to the formation of copper deposits during subduction. We use a plate tectonic model with continuously closing plate boundaries, combined with reconstructions of the spatiotemporal distribution of the ocean floor, including subducted portions of the Nazca/Farallon plates. The models compute convergence rates and directions, as well as the age of the downgoing plate through time. To identify and quantify tectonic parameters that are robust predictors of Andean porphyry copper magmatism and ore deposit formation, we test two alternative supervised machine learning methods; the “random forest” (RF) ensemble and “support vector machines” (SVM). We find that a combination of rapid convergence rates (~100 km/Myr), subduction obliquity of ~15°, a subducting plate age between ~25–70 Myr old, and a location far from the subducting trench boundary (>2000 km) represents favorable conditions for porphyry magmatism and related ore deposits to occur. These parameters are linked to the availability of oceanic sediments, the changing small-scale convection around the subduction zone, and the availability of the partial melt in the mantle wedge. When coupled, these parameters could influence the genesis and exhumation of porphyry copper deposits.

### 1. Introduction

The South American Cordillera is rich in porphyry deposits [Singer *et al.*, 2008]. The Andes is the longest continental arc, formed by long-lived subduction since the Jurassic along the Pacific margin [Maloney *et al.*, 2013]. During its evolution, the Andean margin was host to a series of magmatic events leading to the formation of porphyry copper deposits interspersed in distinct spatiotemporal clusters (Figure 1 and Movie S1 in the supporting information).

The close relationship between porphyry deposits and subduction is well established; however, there is no consensus on which subduction parameters primarily control the genesis of porphyry deposits [Cooke *et al.*, 2005; Richards, 2013]. Rather, the interplay of specific geodynamic parameters, resulting in a particular coupling between the downgoing and overriding plate, are likely crucial in determining and controlling types of magmatism and subsequent ore formation. Porphyry deposits are most often associated with calc-alkaline and adakitic magmatism in subduction zones and refertilization of the subcontinental lithospheric mantle [Thiéblemont *et al.*, 1997; Kay and Mpodozis, 2001; Griffin *et al.*, 2013], with two distinct melting phases required for porphyry genesis. An initial melting in the metasomatized mantle wedge generates relatively oxidized and sulfur-rich mafic magmas with incompatible chalcophile and siderophile elements [Bertrand *et al.*, 2014]. This is followed by an injection of dykes and sills into the melting, assimilation, storage, and homogenization zone of the lower crust resulting in a melt of a crust-mantle-derived hybrid magma [Sillitoe, 1972; Richards, 2003; Bertrand *et al.*, 2014]. The second melting phase has a high content of volatile and metalliferous elements and a relatively low density to allow its upward migration through the crust [Richards, 2003].

Numerous tectonomagmatic conditions have been implicated as controlling the spatiotemporal distribution of magmatism and ore deposits in subduction arcs [Sillitoe, 1972, 2010]. These include the following: a link between deposits in the periphery of stagnating subducted slabs [Khomich *et al.*, 2014], the rate and



**Figure 1.** Porphyry copper deposit locations in South America as compiled by Bertrand *et al.* [2014]. Each deposit is colored by its formation age and shown at its present day location. The present-day subduction trench, colored by the distance along the margin, is reconstructed back through time (supporting information Movie S1) to capture the subduction parameters of the slab associated with each porphyry deposit (referred to in Figure 2).

interaction between multiple conditions controls tectonic regime, magmatism, and subsequent porphyry copper genesis; however, the relative influence of various conditions remains to be established.

Understanding porphyry copper deposit formation mechanisms will prove useful for exploration targeting in the future. New geophysical and geochemical surface detection techniques prompted significant growth in ore deposit discoveries from the 1970s to the 1990s. Subsequently, a majority of near-surface deposits have been exploited, and discovery rates have declined [Schodde, 2011]. Either new detection techniques or improved methods of predictive modeling are required for sustained economic viability of ore exploitation, including porphyry copper deposits. Analytical techniques utilizing spatiotemporal data mining [Cracknell and Reading, 2014; Landgrebe *et al.*, 2013; Merdith *et al.*, 2013] provide a novel four-dimensional analytical approach to understand the interplay of geodynamic parameters associated with magmatism and porphyry deposits [Richards, 2013]. We seek to improve the efficiency of future exploration and advance the understanding of porphyry copper ore deposit formation processes by quantifying the interplay of tectonic parameters enabling porphyry magmatism. To do this, we employ plate tectonic models to simulate tectonic regimes acting along the South American trench since 200 Ma. We compare these tectonic regimes with data sets of known porphyry deposits associated with the trench. By using statistical analyses employed in modern machine learning methods we hope to establish the crucial tectonic conditions and configurations linked to ore deposit formation and where and when episodes of porphyry copper formation have occurred.

## 2. Porphyry Magmatism Data Set

One hundred and fifty-five South American porphyry copper deposits were compiled by Bertrand *et al.* [2014] from “Porphyry copper deposits of the World” [Singer *et al.*, 2008] and from the French Geological Survey

evolution of convergence between subducting and overriding plates [Bertrand *et al.*, 2014], the absolute velocity of the plate overriding a subducting slab and the development of extensional features [Ramos, 2010; Khomich *et al.*, 2014], the age of the subducting oceanic lithosphere [Capitanio *et al.*, 2011], the dip of the subducting slab and flat slab subduction [Jarrard, 1986; Kay and Mpodozis, 2001], trench migration and the obliquity of subduction [Macpherson and Hall, 2002; Jingwen *et al.*, 2013], and the lateral length of the subducting slab [Schellart, 2008]. Subduction of active mid-ocean ridges as well as aseismic ridges (extinct spreading ridges and oceanic plateaus) has also been associated with the development of porphyry-rich ore deposits on the overriding plate [Mcgeary *et al.*, 1985; Cooke *et al.*, 2005; Richards, 2013]. Subducting crustal features from the Nazca and Antarctic plates, including the Nazca-Antarctic Ridge, have been linked to the development of ore deposits in the Andes [Rosenbaum *et al.*, 2005]. Nevertheless, currently no single parameter has been satisfactorily identified as a primary determinant of overriding plate magmatism [Maloney *et al.*, 2013]. This implies that dynamic

database [e.g., *Billa et al.*, 2004]. The class (Cu potential or grade) of the ore deposits varies, and for all our analyses we do not distinguish between them. The age of mineralization for each data point is given either by an absolute age or by a median age of the stratigraphic sequence it belongs to. Deposits are emplaced from 45°S to 9°N within 1300 km of the Nazca-South American present day subduction trench. The ages range from 4.7 to 291.5 Ma, with 118 of the 155 deposits having formed in the Cenozoic. We remove some deposits from our analyses based on their ages and spatial location, explained in section 4. The age-dated modified data set is shown in Figure 1 with deposits in their present day locations. All data are available in the supporting Information.

### 3. Plate Reconstruction Methods

We carry out spatiotemporal data mining by combining geological data with past plate tectonic motions using pygplates, a python module that expands on the plate reconstruction functionality of GPlates ([www.gplates.org](http://www.gplates.org)). The plate reconstructions provide information on the subduction kinematics along the South American margin through time. For this study we use the plate model and associated rotation file of *Seton et al.* [2012] for all reconstructions. The Python scripts used for this are distributed in the supporting information. We use this plate model to reconstruct and track the set of age-dated magmatic events compiled by *Bertrand et al.* [2014] as far back as 200 Ma (the extent of the plate model) with a focus around the time of porphyry formation (Figure 1 and Movie S1 in the supporting information).

The plate kinematic parameters we analyze encompass both relative plate motions and absolute plate motions. Relative motion between South America and the subducting plates is derived from the plate circuit calculations using parameters given in *Seton et al.* [2012]. Absolute motions of these plates are computed by combining relative motions with an absolute reference frame linking relative plate motions to the deep mantle [*O'Neill et al.*, 2005]. Absolute motion of the subducting and overriding plates can thus be calculated as the stage rotation between the subducting plate and mantle. Convergence is calculated as the relative stage rotations of the overriding and subducting plate. We use the same sign conventions for velocities as *Bertrand et al.* [2014]; plate movement toward the overriding plate is defined as positive. Obliquity of subduction is determined by finding the angle between the direction of convergence and the lateral direction of the trench at each of the sampling points. This is defined as 0° when the subducting plate is perpendicular to the trench, positive when the angle of the subducting plate vector is more north and negative when the angle of the subducting plate vector is further south. We use the oceanic paleo-age grids of *Seton et al.* [2012] as an additional spatiotemporal data set related to the coupling of the downgoing plate with the overriding plate.

We transform the data into a matrix consisting of the location of the porphyry deposits through time, along with the tectonic parameters associated with a given deposit including age of subducting slab, convergence rate history, overriding and subducting plate velocity, obliquity of subduction, and distance from slab edge. A nearest neighbor search is used to find the tectonomagmatic and kinematic parameters around the subduction zone associated with each ore deposit at their time of formation. The key parameters of plate convergence rates, subducting seafloor ages, distance to trench edge, and subduction obliquity are shown in section 4 data analysis with the full results for other parameters shown in the supporting information.

To determine which specific set of tectonomagmatic conditions may be associated with groups of ore deposits, we discretize the South American subduction zone every 0.5° along the trench in 1 Myr time steps back to 200 Ma. We then link each ore deposit to its closest discretization and identify the various tectonomagmatic parameters associated with each discretized area.

### 4. Data Analysis

Two different data analysis approaches are used to test different assumptions of porphyry copper ore deposit formation and preservation. The first analysis (detailed in section 4.1) is performed on deposits individually by linking each point of ore formation with the subduction parameters of the closest segment of the subducting trench at its time of formation. This method is then applied to the same data set, but we restrict the time window to 70 Ma to account for older deposits being eroded. The second analysis is performed on the deposits partitioned into space-time clusters that account for entire regions of deposit formation rather than singular points (detailed in section 4.3). This captures porphyry deposits of the South American margin in distinct spatiotemporal clusters derived from the point data of *Bertrand et al.* [2014]. We then use these same

space-time clusters restricted to the 70 Ma and also remove the southernmost 1000 km of the South American trench as there are no deposits in this space-time region.

For statistical validation we produce pseudo-random sets of “nondeposits” that are sets of data points for which we have reason to believe that the tectonic parameters do not result in an ore deposit. This is a necessary step in the machine learning we employ, as the algorithms seeking to find how certain parameters lead to a correct categorization must also know what parameters are associated with an incorrect categorization. There are several methods for doing this [Carranza and Laborte, 2015; Zuo and Carranza, 2011; Cortes and Vapnik, 1995], but the methods are often suitable only for specific cases. We apply two different methodologies for developing nondeposits, one for each of our data analysis approaches.

For the first analysis, the nondeposit data set is built by taking the present day latitude-longitude location of the known deposits and assigning random formation ages between 0 and 200 Ma to each point. The nondeposit locations are then reconstructed back to their random “formation” age, and the tectonomagmatic properties associated with the formation of the nondeposit are recorded. This captures the spatial location for where we know it is geologically feasible for a deposit to form, but the geodynamic conditions at the random time likely were not appropriate for porphyry copper magmatism. This method biases the number of nondeposits to regions where there are already several known deposits while leaving some areas of the South American margin underrepresented. However, this method provides a well-balanced data set with an approximately equal number of deposits and nondeposits, which is a generally important characteristic for machine learning [Akbari *et al.*, 2004].

To produce nondeposits for the second data analysis approach, we begin by reconstructing the known deposits back to their formation time. The resulting time-space map (Figure 2) is then partitioned into 500 km space intervals and 10 Myr time intervals. When one or more known deposits fall into a partition, the entire square is considered a “deposit.” When no known deposits occur in a partition, the square is considered a nondeposit. This method minimizes biasing related to undersampling (as likely not all porphyry copper deposits in South America have been explored and found) and overinterpreting nearby clusters of deposits that may be associated with a single tectonomagmatic episode.

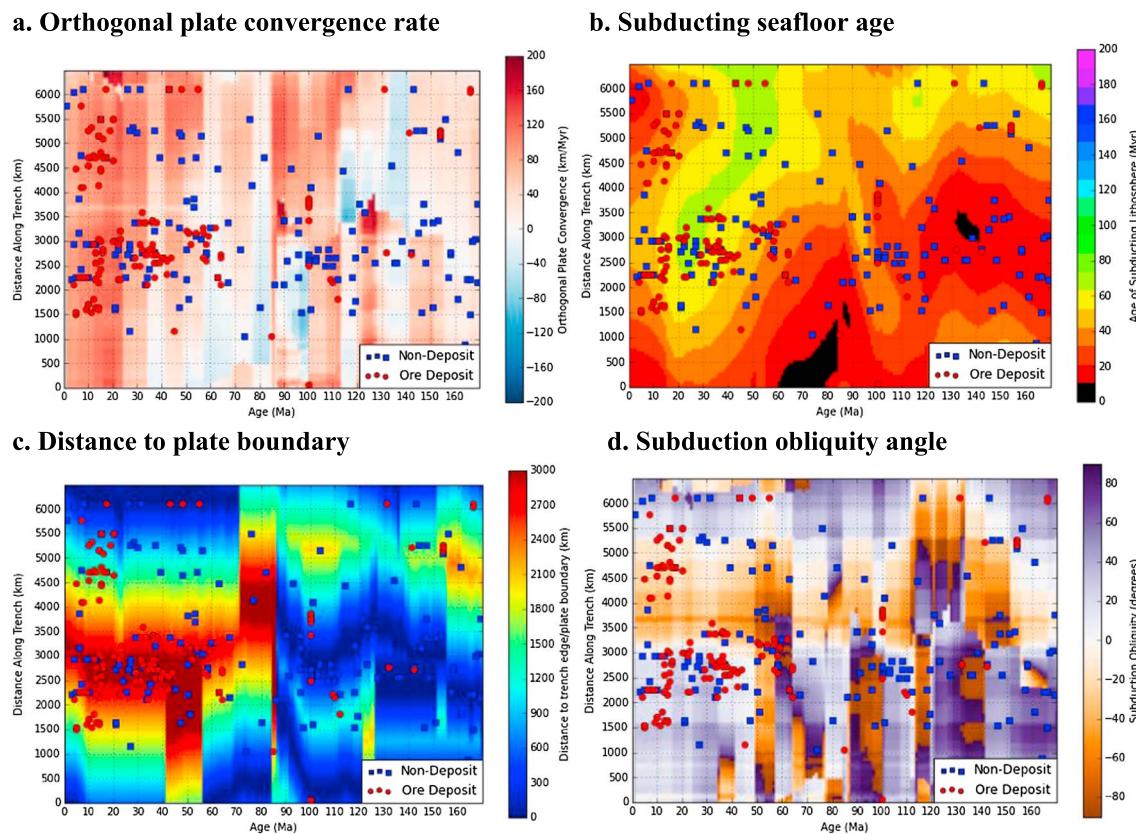
Here we focus on what we have found to be the four primary parameters: plate convergence, subducting slab age, subduction obliquity, and distance of ore deposit from the closest plate boundary along the subduction trench (Figures 2–4). For a full comparison of all parameters investigated see Figures S1–S4 in the supporting information.

#### 4.1. Point of Formation

We first consider each porphyry magmatism data point as distinct events. This allows us to determine the tectonomagmatic parameters associated with every ore deposit location individually. The typical size of a porphyry deposit is on the order of ~2 km thick and around 5 km wide [Singer *et al.*, 2008], and the minimum distance between each of the porphyries in the analyzed data set is ~10 km. The nondeposit data set is built by taking the present day location from the deposits and assigning random formation ages (between 0 and 200 Ma) to each point. This results in a nondeposit data set that captures locations where we know porphyry copper deposits may be formed. Of the 155 deposits in the original data set [Bertrand *et al.*, 2014], we excluded one deposit that is more than 10° from the closest subduction zone at its time of formation. A further eight deposits are older than 200 Ma (the extent of our plate model), and thus, we cannot obtain kinematic data for these deposits. As a consequence, there are 147 deposits and 154 nondeposits in our data partitioning.

Convergence at the South American trench displays a broadly bimodal speed distribution (Figure 3a). Convergence rates for the random sampling of points within South America (nondeposits) peak at 24–34 km/Myr, 65–76 km/Myr, and 86–97 km/Myr. Deposit locations indicate formation approximately coincides with convergence at 24–34 km/Myr, 76–86 km/Myr, and 96–117 km/Myr.

In Figure 3b there are few nondeposit locations correlated with seafloor younger than 11 Myr and few older than 71 Myr. This indicates that seafloor subducting beneath South America since 200 Ma is generally no older than ~80 Ma (Figure 3c) and that no ore deposits are associated with active ridge subduction at the Chile Triple Junction. Concurrently, there are few events of porphyry magmatism correlated with seafloor younger than 28 Myr and few older than 71 Myr. There is a slight peak in the amount of magmatic porphyries



**Figure 2.** Spatiotemporal porphyry copper deposit formation locations and their corresponding subduction parameters of (a) orthogonal convergence rate, (b) age of the subducting oceanic lithosphere, (c) the distance to the closest plate boundary, and (d) the angle between the direction of subduction and the trench along the Nazca-South American trench. The time component is the reconstruction of the present-day trench location. Distance along the trench can be compared geographically with Figure 1. Tectonomagmatic parameters associated with both data analysis approaches can be discerned from this figure. For the point of formation approach, porphyry copper deposits (positive cases) in their space-time formation location are the red circles, and nondeposits (negative cases) in their pseudo-random space-time formation location are blue squares. For the area of formation approach, the grid represents the spatiotemporal partitioning of the data along the trench in 500 km bins and through time in 10 Myr intervals. For partitioning the data, when a deposit or multiple deposits fall in a grid square it is considered a deposit. When no porphyry magmatism occurs in a square, it is considered a nondeposit. The average of each tectonic parameter is taken for the square. In Figure 2c blue hues indicate being near a plate boundary that is not the subduction zone itself (e.g., a trench-ridge intersection), and red hues indicate being far from a boundary.

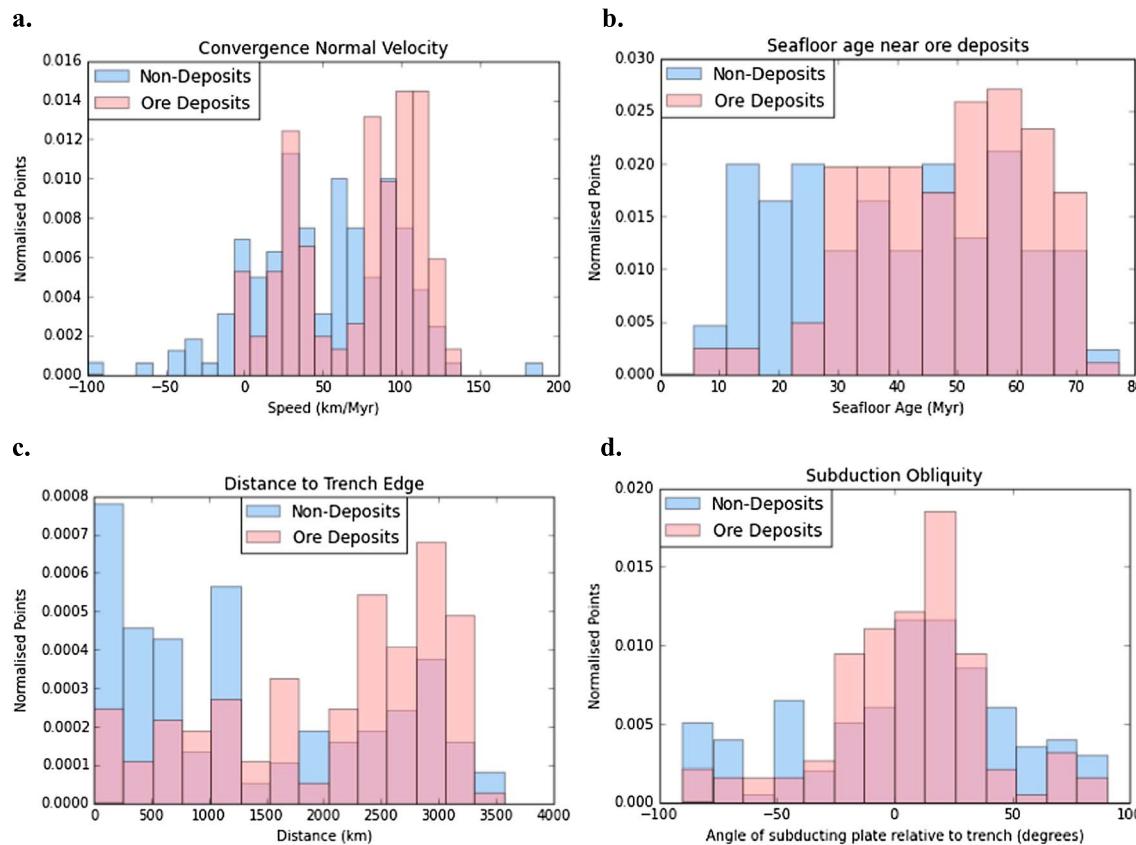
correlated with a specific seafloor age in the 55–60 Myr bin range. This indicates that seafloor older than 28 Myr and at least up to 71 Myr old (data are not tested beyond this range) may be associated with porphyry magmatism.

Figure 3c shows the location of deposit formation along the South American trench; that is, how far the deposits are from both northern and southern plate boundaries. While the nondeposit data cluster near the trench edge, deposits preferentially form when the distance to the trench edge is over 2000 km.

The angle that the subducting plate (Nazca, Farallon, Phoenix, Catequil, or Chazca) makes with the overriding South American plate is inferred from the nondeposit data and displays a normal distribution centered around trench-perpendicular subduction ( $0^\circ$ ). Concurrently, deposits tend to preferentially form when subduction obliquity is close to being trench-perpendicular, as indicated by the clustering of deposits between  $-12$  and  $38^\circ$ . Additionally there is a significant increase in ore formation when obliquity is between  $12$  and  $25^\circ$ .

#### 4.2. Point of Formation With Restriction of Time Period

We now restrict the deposits and the nondeposits to formation between 0 and 70 Ma. This reduces the number of deposits to 121 and the number of nondeposits to 67. The trends with the time-restricted data (Figure S2 in the supporting information) are similar to the results from the previous section. Here deposit formation peaks around plate convergence rates of 30 and 100 km/Myr, with similar rates for nondeposits. If there is a

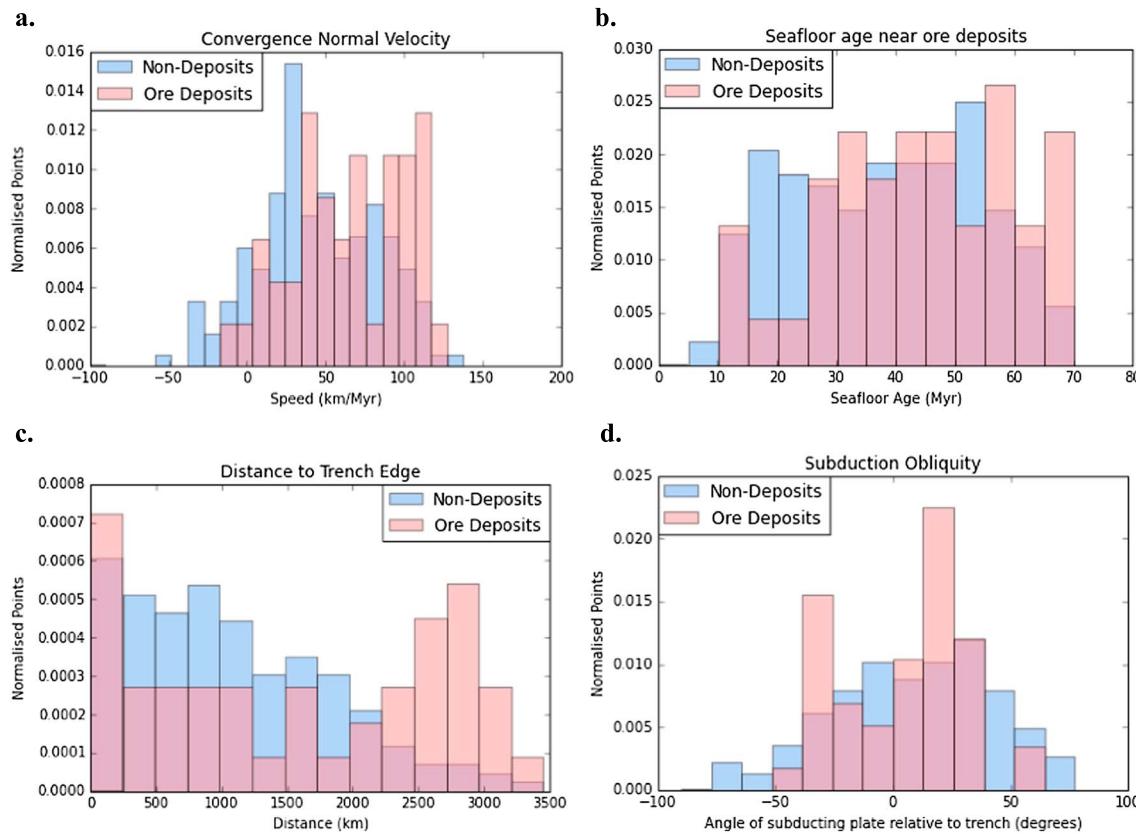


**Figure 3.** Histograms for each parameter for the point of formation data analysis. (a) Histogram of orthogonal convergence rates for nondeposits or negative cases (blue) and deposits or positive cases (red), with overlapping cases shaded purple. The results of the negative cases indicate the general trends of the subducting slab along the South American margin, contrasted with the results of the positive cases that indicate when an ore deposit is formed. If there is a significant difference between the results of the positive and negative cases, then the parameter is likely influential in determining porphyry magmatism. Convergence velocities show significant spikes around 30, 80, and 100 km/Myr. (b) Same as Figure 3a but for seafloor age. Older seafloor (~28–71 Myr old) preferentially forms porphyry deposits. Nondeposits sample seafloor age in a more uniform distribution from ~12–71 Myr. (c) Same as Figure 3a but for distance to trench edge. Deposits tend to form when they are far away (>2000 km) from an edge of the subducting trench. (d) Same as Figure 3a but for the angle of obliquity between the subducting and overriding plates. Both deposits and nondeposits have similar trends; however, deposits are more clustered around trench-normal subduction (i.e., 0° obliquity) with a peak around 12–25°.

significant difference between results for positive and negative cases, then the parameter is likely influential in determining porphyry magmatism. Seafloor age associated with both deposits and nondeposits is relatively old (>28 Myr). Older seafloor (~33–71 Myr old) is preferential for forming porphyry deposits. Nondeposits peak when seafloor age is between 38 and 50 Myr old and between 55 and 71 Myr old. Both deposits and nondeposits are related to being far from the trench edge, with formation rates increasing as the deposit moves beyond 2000 km from the closest plate boundary along the trench. Even with the time-restricted data deposits and nondeposits show similar trends and are associated with trench-perpendicular subduction (i.e., 0° obliquity), with a slight increase in samples when obliquity is approximately 12–25°.

#### 4.3. Area of Formation (Partitioned Data)

For the second data analysis approach, rather than considering each data point individually, we partition the South American margin into spatiotemporal bins and track the tectonomagmatic properties of an entire area. Spatially, we take the present day Nazca-South American trench and split it into 500 km segments, which accounts for the maximum width of a single porphyry deposit being around 100 km [Singer *et al.*, 2008]. The margin is analyzed every 0.5° (around 55 km). The plate kinematics (defined by the Euler pole configurations of each of the plates) varies at around the same resolution. The oceanic age grid [Seton *et al.*, 2012] has a resolution of 1° (around 110 km). We find that these parameters do not vary significantly over the chosen partitioning length of 500 km (Figure 2) and therefore constitute a reasonable assumption for the scale that



**Figure 4.** Histograms for each parameter for the area of formation data analysis. (a) Histogram of orthogonal convergence rates for nondeposits or negative cases (blue) and deposits or positive cases (red), for the partitioned data. When deposits and nondeposits overlap they are shaded purple. The results for the negative cases indicate the general trends of the subducting slab along the South American margin, contrasted with the results of the positive cases that indicate when an ore deposit is formed. Where a significant difference exists between positive and negative cases, the parameter is likely influential in determining porphyry magmatism. Deposits have a significant spike in formation when convergence is around 40, 70, and 100 km/Myr. (b) Same as Figure 4a but for the age of the subducting seafloor. Deposit formation is concentrated around seafloor aged ~28–71 Myr. (c) Same as Figure 4a but for distance to trench edge. Deposits tend to form when they are far away (>2000 km) from the edge of the subducting trench. (d) Same as Figure 4a but for the angle of obliquity between the subducting and overriding plates. Deposits have a peak in formation rate when they are associated with an angle of subduction obliquity of 12–25°.

kinematic parameters operate on and subsequently how they may condition the crust for hosting porphyry systems. A magma chamber of at least  $50 \text{ km}^3$  is needed for porphyry production, but a larger chamber is needed (up to  $500 \text{ km}^3$ ) to produce larger deposits [Sillitoe, 2010; Dilles, 1987].

We temporally bin the tectonomagmatic history into 10 Myr increments. Many of the deposits do not have well-constrained ages [Bertrand *et al.*, 2014; Singer *et al.*, 2008; Billa *et al.*, 2004]. Ages have a reported error range between 0 and up to 40 Myr and are occasionally representative of a given geological period only. Furthermore, the modal age for a deposit to be formed has been quantified previously as 9 Myr [Kesler and Wilkinson, 2008] and up to 20 Myr [Sillitoe, 2010], indicating that attributing a precise age to an ore deposit can be misleading. We only analyze back to 170 Ma as no known deposits are formed between 170 and 200 Ma (the extent of our plate model). Deposits are counted as when a space-time bin contains a known data point of porphyry magmatism and nondeposits when no known points fall within the bin. In this data partitioning (shown in Figure 2) there are 45 examples of deposits and 176 examples of nondeposits.

With the different definition for how porphyry formations are grouped we find similar correlations to the previous sections. Figure 4a shows peaks in porphyry formation corresponding to convergence rates of 44–55 km/Myr, 65–75 km/Myr, and 86–117 km/Myr; however, the peaks are less distinctive. The peaks in nondeposits (indicative of the South American margin in general) are clearly defined at 24–34 km/Myr and 75–86 km/Myr.

The results in Figure 4b suggest metallogenesis is associated preferentially when seafloor is 25–70 Myr old, with another peak also occurring with seafloor aged between 10 and 15 Myr. Concurrently, nondeposits tend to be broadly associated with seafloor between 10 and 55 Myr old.

Figure 4c shows a peak in deposit formation when they are over 2000 km from a plate boundary along the trench. There is also a peak in formation very close to the trench boundary. The nondeposit data show the same peak close to the trench boundary, with nondeposits becoming less frequent with distance from the trench.

Both deposits and nondeposits cluster around trench-perpendicular subduction, with the deposits forming a tighter grouping around 0° (Figure 4d). The deposits also show two remarkable peaks in formation, when the obliquity is within the 12–25° and –38 to –25° bins.

The spatial and temporal resolution of this area of formation method can be easily changed in the supporting information IPython notebooks. We find that results are consistent when the spatial partitioning is reduced to 100 km segments along the margin and with time increments reduced to 5 Myr.

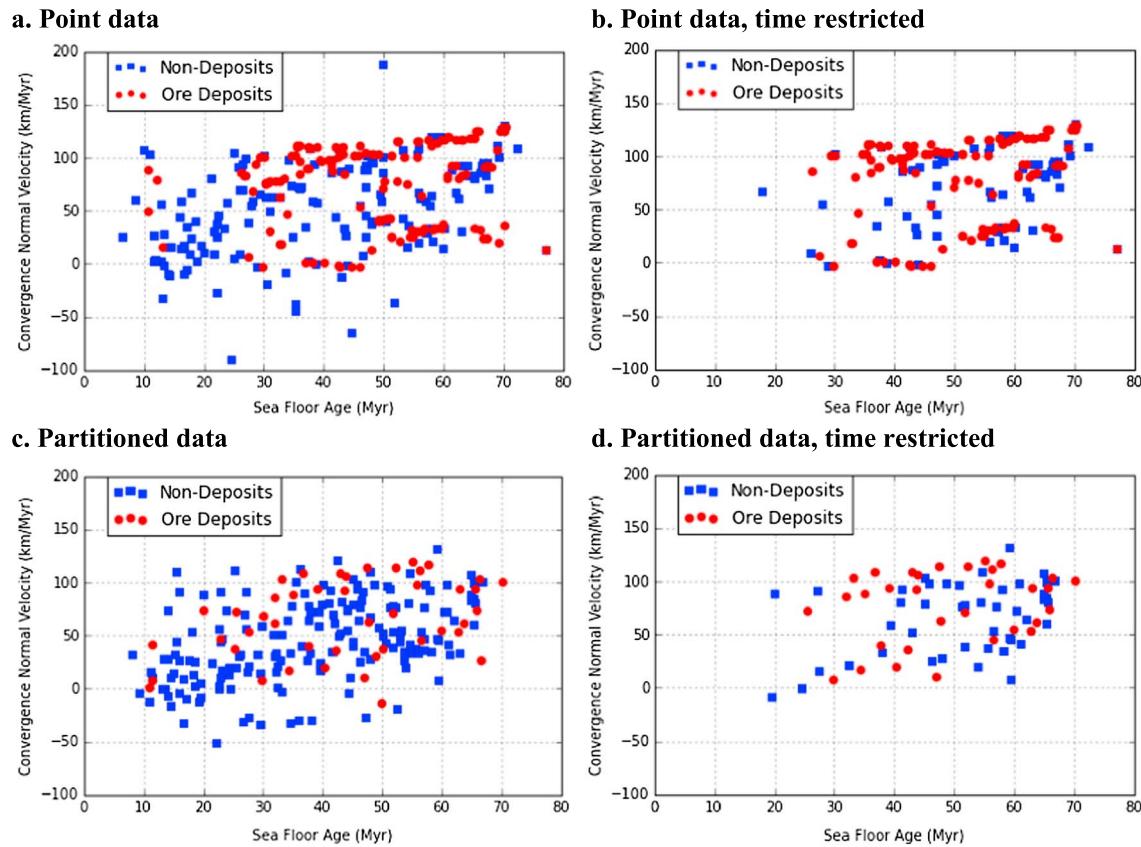
#### 4.4. Area of Formation (Partitioned Data) With Restriction of Time Period

As 70% of ore deposits have formed since 70 Ma the analysis was subsequently restricted to this time period, thereby reducing the number of deposits to 32 and the negative examples to 45. This balances the positive and negative samples. All of these deposits formed at least 1000 km north of the southern tip of the trench; thus, we apply this spatial constraint by removing this area from the analysis. The data here are relatively sparse; however, we find similar trends when compared with the non-time-restricted results (Figure S4 in the supporting information). Here convergence rates peaking between 86 and 117 km/Myr are associated with the formation of porphyry copper deposits. However, there is an overlap with the formation of nondeposits, peaking when convergence rates are between 76 and 107 km/Myr. Seafloor age is again associated with porphyry formation between 25 and 70 Myr. In this partitioning, nondeposits peak when seafloor is 55–60 Myr old. Subduction obliquity shows similar patterns to the previous results, again with a peak in deposits when obliquity is around 15°. Coherent with previous findings, the distance to the trench edge seems to influence the formation of deposits when they are relatively far from a trench boundary.

### 5. Machine Learning Analysis

Characteristics of the deposit data alone do not necessarily give evidence for causal relationships between a particular parameter and the formation of a porphyry deposit; rather they must be considered alongside a representation of the average parameters operating. If the nondeposit results indicate a similar relationship to a particular parameter compared with the deposit results, this suggests that the parameter is merely an identifying feature of the South American subduction zone. However, when the deposit and nondeposit results differ (when the tectonic parameters associated with a deposit deviate from the typical tectonics of the South American margin, as compared in Figure 5), this suggests that the parameter may be somehow related to metallogenesis. In Figure 5 the spread of nondeposits gives an indication for the tectonic parameters associated with subduction along the South American margin not necessarily associated with ore formation. This is contrasted with the distinct clusters of deposits, implying a link between these parameters and their association with ore deposit formation.

To reinforce which tectonomagmatic parameters drive metallogenesis, we perform statistical analyses using machine learning methods. We utilize two different supervised machine learning methods, random forests (RF) and support vector machines (SVM), to identify the strongest connections in the data and distinguish which parameter interactions are likely key to porphyry genesis. Each method has its strengths and weaknesses for our data types; thus, the comparison of the methods provides insight into the robustness of the predictive capabilities of the data. Both methods work by sampling our data set for training points and testing points. We train the classifiers on a subset of our data points to make a training set. Both methods use this set to learn which tectonomagmatic parameters are linked to ore deposits and which are linked to nondeposits. This leaves a different subset of the data aside to test our trained classifiers on, which is achieved by presenting the method with the tectonomagmatic parameters of the testing set and then the classifiers tell us whether to expect an ore deposit or not (for which we also know the solution). The result is a prediction to compare against the actual data (Figure 6). We utilize a fivefold cross validation, which splits



**Figure 5.** Correlation between seafloor age and plate convergence with porphyry magmatism for the four different data partitions represented by (a) point data, (b) point data time restricted to 0–70 Ma, (c) partitioned data, and (d) partitioned data time restricted to 0–70 Ma. The machine learning methods attempt to separate the deposits (positive cases; blue circles) and nondeposits (negative cases; red squares). Additional parameters showing the various correlations and deviations of deposits and nondeposits are plotted in Figures S1–S4 in the supporting information.

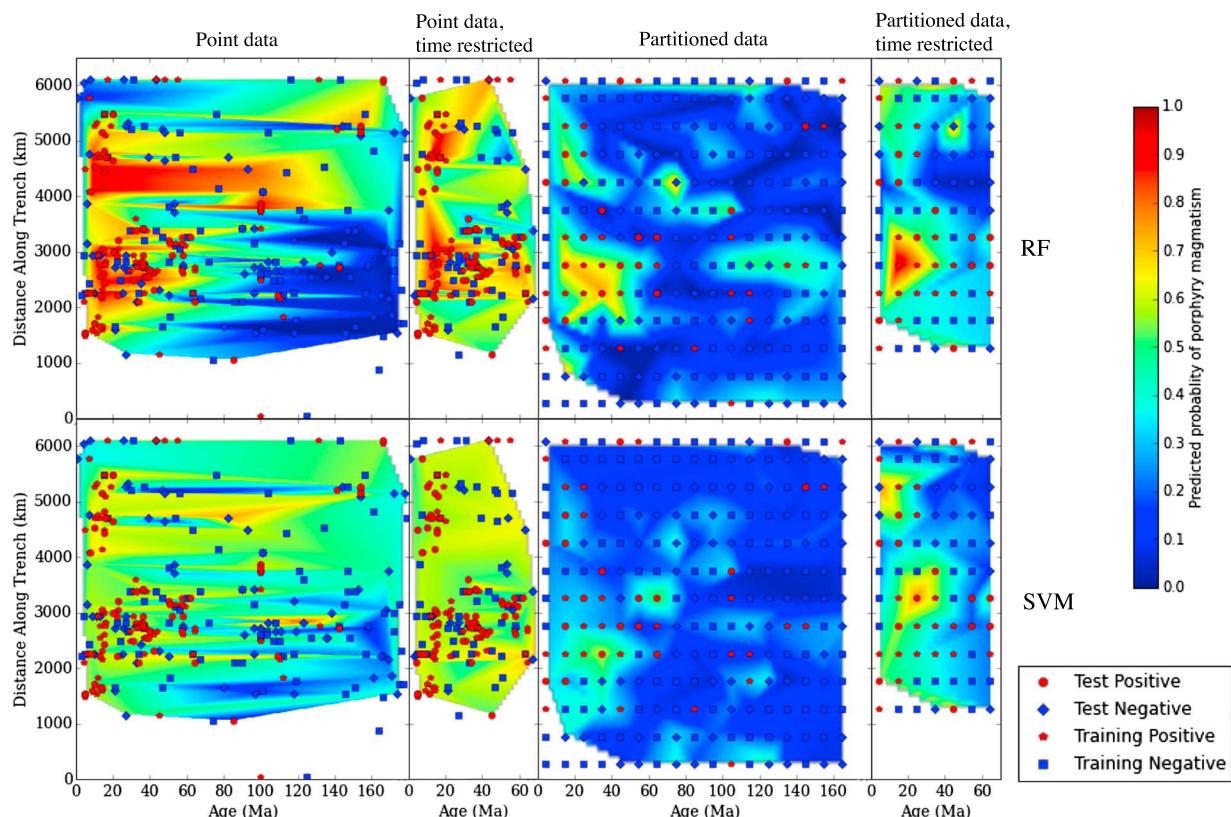
the training and testing sets into five sets with 80% for training data and 20% for testing data and provides a score for each split. The average of the five scores is presented in Table 1. The Python scripts implementing each of these steps are distributed in the supporting information.

### 5.1. Random Forest

RF are nonlinear classifiers [Breiman, 2001], and we use the implementation from the scikit-learn Python package [Pedregosa et al., 2011]. They train an ensemble of decision trees on a random subset of features (tectonic parameters) and training examples, with replacement. For prediction, the majority vote of the decision trees' predictions on a query point determines the class of the query point (Figures 6 and 7). This randomization process is often called bootstrap aggregating or “bagging” (random forests use bagging on both the data and the features). Since random forests use bagging, it is quite easy to determine which features are important for prediction (Figure 7); the exact method for this is discussed in Breiman [2001]. This method can work well with unbalanced class data such as ours [Oshiro et al., 2012] and has been used for both mineral exploration [Cracknell et al., 2014; Carranza and Laborte, 2016; McKay and Harris, 2016] and lithological classification [Cracknell and Reading 2013; Waske et al., 2009].

### 5.2. Support Vector Machines

The SVM is another nonlinear classification method—it is a kernel machine and uses sets of support vectors to determine classification [Cortes and Vapnik, 1995]. It can nonlinearly project the features (tectonic parameters) to a high (or infinite) dimensional space (the kernel space) using the so-called kernel trick, in which the data may become more “linearly” separable into the positive and negative classes. The projection and separating hyperplane are determined by the choice of kernel function (we use a radial basis function) and



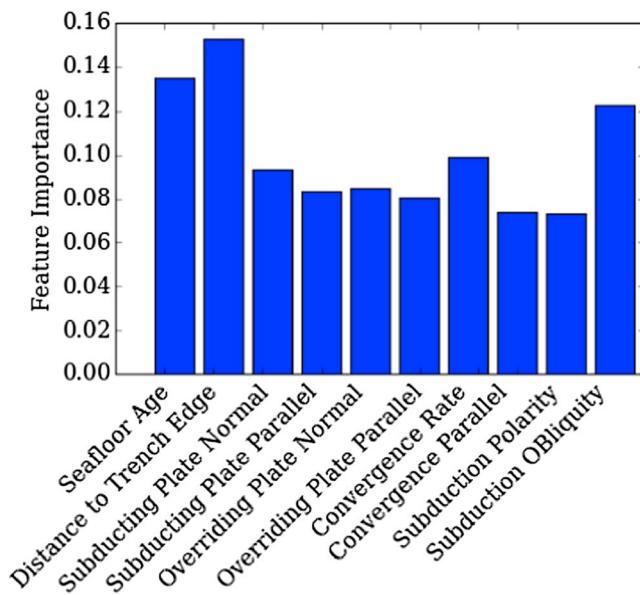
**Figure 6.** Typical time-space likelihood map of porphyry copper formation. This shows the prediction of porphyry magmatism occurring through time and space for two machine learning methods and four data sampling methods. The top row shows the RF results and the bottom row are the SVM results. Each column is labeled with its respective data analysis approach. We split the data into a training set and a testing set, as indicated in the legend. The red markers denote the deposits or positive examples, with the test set shown as circles and the training set as pentagons. The nondeposits or negative examples are shown as blue markers, with the test set shown as diamonds and the training set as squares. The prediction is interpolated between the randomly selected testing points. Here we display typical predicted probabilities based on these four primary tectonomagmatic parameters.

the learned support vectors; see Cortes and Vapnik [1995] for a theoretical presentation of the method. SVM uses known binary class labels, in this case deposits and nondeposits, for which we have several features (tectonic parameters) that optimize the data vector classification into each of these cases. The SVM scales linearly,  $O(N)$  where  $N$  is the size of the training data set, for training when a linear kernel is used (then the SVM is a linear classifier), or more than quadratically,  $O(N^2)$ , when a nonlinear kernel function is used. We only have a small training data set and so can use the more powerful, nonlinear radial basis kernel [Akbari et al., 2004]. For example, in Figure 5 the kernel would attempt to draw a line between the two data sets. However, in higher-dimensional parameter space (the results for the four tectonomagmatic parameters here are shown in Figure 6) it attempts to separate the deposits and nondeposits with a hyperplane described by a set of vectors. A full description of the method and an application to mineral prospectivity can be found in Abedi

**Table 1.** Fivefold Cross-Validation Scores for Each Data-Analysis Partitioning and Machine Learning Method<sup>a</sup>

	Points	Points Time Restrict	Partition	Partition Time Restrict
RF	$0.66 \pm 0.06$	$0.62 \pm 0.04$	$0.78 \pm 0.03$	$0.59 \pm 0.13$
SVM	$0.67 \pm 0.07$	$0.64 \pm 0.02$	$0.82 \pm 0.03$	$0.60 \pm 0.18$
Random	0.50	0.54	0.68	0.51
Pick Case	0.51	0.64	0.80	0.58

<sup>a</sup>A score of 1 indicates the testing set is predicted by the parameters learned from the training set. A score of 0.5 indicates random performance when the data sets are balanced; the expected random scores for our data sets are in the third row for each method, and the max score from always picking one case (positive or negative) are in the bottom row. The errors stated are the standard deviations of the cross-fold validation. See Breiman [2001] for how RF and see Platt [1999] for how SVM scores are determined.



**Figure 7.** Typical feature importance of each parameter for identification of robust predictors. Feature importance is determined by the relative predictability of a target variable. Features (tectonic parameters) that end up being at the top of a decision tree (after the RF classification) naturally contribute to a larger fraction of predictions and thus are considered more important. When combined with the additional parameters in the supporting information, the four parameters focused on (seafloor age, distance to trench edge, trench-orthogonal convergence, and subduction obliquity) consistently show a feature importance above ~10% (if all 10 features tested here contributed equally, they would each contribute exactly 10%) and produce significant out-of-bag accuracy scores.

parameters, *learning* what parameters are important for the prediction of porphyry deposits (positive) and nondeposits (negative). We find that every method suggests that being far from the trench edge, having an older seafloor and having a relatively fast convergence rate, are important parameters related to the formation of porphyry copper deposits, so we focus on these four primary parameters. The results from the machine learning produce a series of time-space maps (Figure 6) indicating the likelihood of porphyry magmatism occurring throughout the subduction history of the South American trench. Random Forest probabilities are determined as an “out-of-bag accuracy” [Breiman, 2001], while SVM probabilities are determined using the Platt scaling method [Platt, 1999] that assigns probabilities by comparing the trained support vectors.

The machine learning outcome of the RF method provides us with a quantitative measure of what tectonic parameters (features) are important. This is achieved in the scikit-learn Python package by following the decision trees and determining which decision trees provide the highest out-of-bag accuracy [Breiman, 2001]. We find the feature importance of each parameter to be fairly balanced in each scenario. Typical feature importance is shown in Figure 7, although these results fluctuate depending on the choice of training and testing sets (as can be done in supporting information IPython notebook). These parameters consistently show similar results between each other, and only when included with the additional parameters (supporting information) do they appear more significant. We find that testing each parameter used individually is no better than random selection in predicting other porphyry deposits, with performance scores of approximately 0.5 (Table 1).

## 6. Discussion

### 6.1. Tectonomagmatic Parameters Investigated

Seafloor age, distance to trench edge, trench-orthogonal velocity, and subduction obliquity were analyzed extensively as these four parameters emerged as dominant tectonic parameters and the most robust

*et al.* [2012]. The use of SVMs for remote sensing was reviewed by *Mountrakis et al.* [2011], and they have also been used for land cover mapping [*Huang et al.*, 2002, 2008] and mineral targeting [*Rodriguez-Galiano et al.*, 2015; *Zuo and Carranza*, 2011]. For the SVM we again utilize the scikit-learn Python package implementation [*Pedregosa et al.*, 2011] with the default radial basis function kernel and parameters.

We have analyzed each data set individually and combined with the other parameters. We find systematically that some of the data sets (subducting plate velocity, overriding plate velocity, parallel convergence velocity, subduction polarity, and calculated slab dips as shown in supporting information) lead to overinterpretations, overtraining, or create noise and are better left out of the analysis. This was done by systematically testing each parameter, plotting them as in Figure 7, and determining which features were consistently significant. The Python scripts in the supporting information can be used to reimplement each parameter. We train our model using all of the subduction

predictors of ore deposit location during initial testing (Figure 7 and supporting information). However, we also investigated the subducting plate velocity, overriding plate velocity, parallel convergence velocity, subduction polarity, and calculated slab dips. We used subduction polarity (the angle a subduction zone makes with north) to help classify what we assumed to represent a feature of no importance. The use of additional parameters (e.g., heat flow and geochemistry) may provide further insights into the tectonic conditions required for porphyry magmatism.

## 6.2. Associations With Tectonomagmatic Parameters

Any single parameter on its own does not produce a statistically significant robust predictor for porphyry formation. However, when considered simultaneously, and when the parameters pertain to certain conditions, then the likelihood of porphyry magmatism increases. Traditional porphyry formation studies have considered single parameters, and this may explain why some contradictions in formation mechanisms associated with tectonics abound [Sillitoe, 1998, 2010; Tosdal and Richards, 2001; Bertrand *et al.*, 2014]. The fact that there is no single dominant feature in the determination of the classes (e.g., Figure 7) suggests that any parameter acting alone is not necessarily enough for formation of porphyry copper deposits.

We find relatively older seafloor age, at least up to 70 Myr old (our data do not consist of older seafloor), seems to be positively correlated with increased formation of porphyry copper deposits, with peaks in formation rates when seafloor is older than 30 Myr old. It has been suggested that metals contained in porphyry ore deposits are derived from divergent margins [Sillitoe, 1972; Qu *et al.*, 2004]. Seafloor formed at spreading ridges may be associated with volcanic-hosted massive sulphide deposits [Galley *et al.*, 2007] with variable copper content in seafloor mafic and ultramafic rocks as a function of enrichment due to hydrothermal activity or source magma [Hattori and Keith, 2001]. As seafloor thickens as it ages, older oceanic lithosphere on average contains more porphyry-forming minerals. Concurrently, older seafloor is overlain by thicker sediments, which add to the melt volume upon subduction, yet the reliability of this factor is debated [Qu *et al.*, 2004]. There is minimal seafloor (less than 3%) older than 70 Myr old being subducted along the Andes for the period we observed; thus, our results can only be interpreted for seafloor younger than 70 Myr. We see the likelihood of copper metallogenesis increasing with an aging seafloor up to 25 Myr old, maintaining a consistent likelihood thereafter. Thus, it seems that once a critical mass of sediments or porphyry-forming minerals is reached, porphyry copper deposits will preferentially form.

Another possibility is the role of increased flexure of a slab consisting of relatively old seafloor, possibly due to increased sediment loading and thickness. This change would affect the angle of slab penetration, the kinematic parameters, and thus the location of where melting may occur [Isacks, 1988; Davies, 2009; Capitanio *et al.*, 2009]. The age of the subducting oceanic lithosphere influences the rate of uplift and hence the topography of the overlying continental arc [Capitanio *et al.*, 2011]. This is partially through changes to the dip angle of the subducted slab as well as the buoyancy of the downgoing plate, the strength of the plates, and the thickness of the overriding plate [Butterworth *et al.*, 2012]. The subduction angle is also dependent on the absolute plate velocities of the downgoing and overriding plate. Subduction of young, warm, and light oceanic lithosphere will tend to flatten at about 100 km depth, resulting in rapid uplift of the overlying crust, modified magma chemistry, and eventually a shutdown in magmatic activity [Shatwell, 2004; Rosenbaum and Mo, 2011]. This may result in destruction (by erosion) of existing epithermal Au-Ag deposits and exposure of underlying porphyry systems, thereby underrepresenting porphyry copper deposits associated with young and flat slabs.

A location far from the trench edge appears positively correlated with increased porphyry magmatism. With increased distance from the trench edge, the return mantle flow increases [Stegman *et al.*, 2006; Schellart, 2008] and subsequently the quantity of melt derived through convection increases. Correspondingly, the convergence rates are intrinsically related to the slab-advance/rollback variations along the extent of the trench [Schellart, 2005].

Bertrand *et al.* [2014] suggested that fast convergence rates followed by a rapid deceleration increases the likelihood of porphyry deposits forming, probably due to an increase in partial melt and the promotion of magma to the upper crust. We find deposits are most likely to form when convergence is relatively rapid. The time period captured in our analysis does not compare with the time period suggested by Bertrand *et al.* [2014] (~20 Myr). Initial time series analyses employing RF and SVM methods did not indicate statistically

significant associations, which we suggest is related to the spatial and temporal resolution of our models and the data sets as well as our methods for choosing nondeposits.

The angle of obliquity relates subduction to the depth of melt. When rapid convergence rate is coupled with an angle close to orthogonal, the melt will likely be deeper within the crust. However, if the angle is relatively obtuse, melt will be too shallow for ore deposit formation. Our analysis suggests that the optimum angle of subduction obliquity for porphyry deposits to form (associated with a particular convergence rate) is relatively close to orthogonal.

The combination of parameters suggests that older seafloor entering the subduction zone provides a key constituent to form porphyry deposits. When this is coupled with a rapid convergence (and an appropriate subduction obliquity), melt is increased and thus facilitates the percolation of porphyry melts to the near surface. Furthermore, the distance to the trench edge impacts on the location of melts and convergence rates and thus operates to hinder porphyry magma ascent or destruction. Using our understanding of these parameters acting through time, we may assess likely locations of porphyry copper deposits. The predicted probabilities of porphyry magmatism in the time-space maps of Figure 6 are essentially the reconstructed position of where (along the trench) and when (at what age) present-day ore deposits (known and unknown) have likely formed. These results may be used as an exploration tool for determining likely locations where porphyry copper deposits have formed along the South American margin through time.

### 6.3. Identifying Nondeposits and Machine Learning Quandaries

Success using machine learning relies on the ability to characterize the data set of interest. We have presented two different cases for choosing nondeposits, and our results appear to be robust for both methods. Zuo and Carranza [2011] offer a formalized method for picking nondeposits based on the proximity to known deposits, geological domains, and weighting of nondeposits and deposits. However, this sampling method for nondeposits would not be effective in our case, as this criterion requires that nondeposits be far from subduction zones, which would not help us distinguish the importance of subduction-related parameters. Other methods exist for choosing nondeposits [Carranza *et al.*, 2008] but have not been considered in geoscience time-dependent machine learning cases.

The scores indicated in Table 1 for both SVM and RF appear better than what would be expected with random performance. However, in the cases of the relatively unbalanced data sets, if we were to always pick the negative case or positive case accordingly, then classification would perform similarly to our results. The “points” data set scores the best between the various data sets, and this result suggests unequivocally our classifier can make use of the information from the input features to improve what can be learned from just the labels.

The random-time cases likely represent the best choice for negative cases because we know that the “ingredients” for porphyry formation are in the right location within the lithosphere, but the tectonic parameters have not combined to enforce the porphyry magmatism to express itself. However, in this method, we have picked nondeposits that may include areas privy to copper-ore formation, thereby adding noise to the negative examples. This does help us consider whether the geochemical preconditioning of the lithosphere enables formation of porphyry copper deposits. However, in an already noisy data set with ages poorly defined, this may add excessive noise to the results.

Selecting regions along the subduction zone that display no known signs of porphyry magmatism is acceptable, but this technique may miss regions that once contained ore deposits that have since been eroded [Kesler and Wilkinson, 2008]. To overcome this, a thorough study of each area where our methods are to be applied would be necessary and nondeposit locations would have to be manually identified based on comprehensive geological evidence (e.g., geochronology, petrology, and geochemistry). However, this contributes to the problem of overinterpreting the data before we apply machine learning methods. In summary, we believe that we have established a balance of these constraints through the comparisons of methodologies we have employed here.

We also tested different methods of weighting the deposit and nondeposit locations to successfully train and balance the data, as typically these data should be of the same size [Akbari *et al.*, 2004; Carranza and Laborte, 2015]. However, we find that rather than applying weighting to the parameters and data, we have presented a fairly balanced data set by reducing the time period of interest to 70 Ma. And although this reduces the size

of the data set, the SVM employed is known to work well when small data sets rather than groups of samples or support vectors are used to train the classifier. The methods presented here are easily scalable as more data become available.

The parameter choices and machine learning techniques we have applied for analysis of Andean ore deposits have also been tested on Southeast Asian porphyry deposits but yielded poor initial results. This is likely because Southeast Asian subduction zones have a more complex history [Metcalfe, 2011; Zahirovic *et al.*, 2014] that differs in tectonic style from the Andean continental margin and mineralization may not follow the same rules as along major continental subduction zones. Results here may also become more robust if spatiotemporal resolution of each of the parameters (plate model, age grid, temporal binning, and trench partitioning) were improved.

## 7. Conclusion

Four-dimensional (space and time) analysis of the geological record is nearly always based on sparse data. Thus, understanding the nature of time-dependent geophysical processes is inherently challenging. The robustness of our results, as tested through four different methodologies with two machine learning approaches, indicates that relatively older seafloor ( $>25$  Myr), a trench normal to slightly oblique convergence angle for the subducting slab ( $\sim 15^\circ$ ), converging relatively fast ( $\sim 100$  km/Myr), while being far away from trench edges ( $>2000$  km), optimizes porphyry copper metallogenesis in the Andes. This confluence of conditions is likely through the provision of additional primordial melt material owing to a thicker down-going lithosphere and increased sediments. The optimal tectonic regimes then allow for the necessary convective cells to aid in the formation and ascension to the near-surface of porphyry copper magmas.

### Acknowledgments

This project was supported by a Big Data Knowledge Discovery project (SIEF project RP 04-174), an ARC Laureate Fellowship (FL0992245), and the Auscope National Collaborative Research Infrastructure. We thank Martin Liu, Ben Dickinson, and Michael Rubey for many discussions; Jodie Pall for extensive proofreading of the manuscript; and Bernard Guest and anonymous reviewers for valuable feedback to improve the manuscript. All data used are included in the supporting information.

### References

- Abedi, M., G.-H. Norouzi, and A. Bahroudi (2012), Support vector machine for multi-classification of mineral prospectivity areas, *Comput. Geosci.*, 46, 272–283, doi:10.1016/j.cageo.2011.12.014.
- Akbani, R., S. Kwek, and N. Japkowicz (2004), Machine Learning ECML, Chap. *Applying Support Vector Machines to Imbalanced Datasets*, pp. 39–50, Springer, Berlin.
- Bertrand, G., L. Guillou-Frottier, and C. Loiselet (2014), Distribution of porphyry copper deposits along the western tethyan and Andean subduction zones: Insights from a paleotectonic approach, *Ore Geol. Rev.*, 60(0), 174–190, doi:10.1016/j.oregeorev.2013.12.015.
- Billa, M., D. Cassard, A. Lips, V. Bouchot, B. Tourrière, G. Stein, and L. Guillou-Frottier (2004), Predicting gold-rich epithermal and porphyry systems in the central Andes with a continental-scale metallogenic GIS, *Ore Geol. Rev.*, 25(1–2), 39–67, doi:10.1016/j.oregeorev.2004.01.002.
- Breiman, L. (2001), Random forests, *Mach. Learn.*, 45(1), 5–32, doi:10.1023/A:1010933404324.
- Butterworth, N., L. Quevedo, G. Morra, and R. Müller (2012), Influence of overriding plate geometry and rheology on subduction, *Geochim. Geophys. Geosyst.*, 13, Q06W15, doi:10.1029/2011GC003968.
- Capitanio, F., G. Morra, and S. Goes (2009), Dynamics of plate bending at the trench and slab-plate coupling, *Geochim. Geophys. Geosyst.*, 10, Q04002, doi:10.1029/2008GC002348.
- Capitanio, F. A., C. Faccenna, S. Zlotnik, and D. R. Stegman (2011), Subduction dynamics and the origin of Andean orogeny and the Bolivian orocline, *Nature*, 480(7375), 83–86, doi:10.1038/nature10596.
- Carranza, E. J. M., and A. G. Laborte (2015), Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines), *Comput. Geosci.*, 74(0), 60–70, doi:10.1016/j.cageo.2014.10.004.
- Carranza, E. J. M., and A. G. Laborte (2016), Data-driven predictive modeling of mineral prospectivity using random forests: A case study in Catanduanes Island (Philippines), *Nat. Resour. Res.*, 25(1), 35–50, doi:10.1007/s11053-015-9268-x.
- Carranza, E. J. M., M. Hale, and C. Faassen (2008), Selection of coherent deposit-type locations and their application in data-driven mineral prospectivity mapping, *Ore Geol. Rev.*, 33(3–4), 536–558, doi:10.1016/j.oregeorev.2007.07.001.
- Cooke, D. R., P. Hollings, and J. L. Walshe (2005), Giant porphyry deposits: characteristics, distribution, and tectonic controls, *Econ. Geol.*, 100(5), 801–818, doi:10.2113/gsecongeo.100.5.801.
- Cortes, C., and V. Vapnik (1995), Support-vector networks, *Mach. Learn.*, 20(3), 273–297, doi:10.1007/BF00994018.
- Cracknell, M. J., and A. M. Reading (2013), The upside of uncertainty: Identification of lithology contact zones from airborne geophysics and satellite data using random forests and support vector machines, *Geophysics*, 78(3), WB113–WB126, doi:10.1190/geo2012-0411.1.
- Cracknell, M. J., and A. M. Reading (2014), Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information, *Comput. Geosci.*, 63(0), 22–33, doi:10.1016/j.cageo.2013.10.008.
- Cracknell, M. J., A. M. Reading, and A. W. McNeill (2014), Mapping geology and volcanic-hosted massive sulfide alteration in the Hellyer–Mt Charter region, Tasmania, using Random Forests and Self-Organising Maps, *Aust. J. Earth Sci.*, 61, 287–304, doi:10.1080/08120099.2014.858081.
- Davies, G. (2009), Effect of plate bending on the Urey ratio and the thermal evolution of the mantle, *Earth Planet. Sci. Lett.*, doi:10.1016/j.epsl.2009.08.038.
- Dilles, J. H. (1987), Petrology of the Yerington Batholith, Nevada: Evidence for evolution of porphyry copper ore fluids, *Econ. Geol.*, 82(7), 1750–1789, doi:10.2113/gsecongeo.82.7.1750.
- Galley, A. G., M. D. Hannington, and I. R. Jonasson (2007), Volcanogenic massive sulphide deposits. Mineral deposits of Canada: A synthesis of major deposit-types, district metallogeny, the evolution of geological provinces, and exploration methods—Geological Association of Canada, Mineral Deposits Division, Special Publication, 5, pp. 141–161.

- Griffin, W. L., G. C. Begg, and S. Y. O'Reilly (2013), Continental-root control on the genesis of magmatic ore deposits, *Nat. Geosci.*, 6(11), 905–910.
- Hattori, K. H., and J. D. Keith (2001), Contribution of mafic melt to porphyry copper mineralization: evidence from Mount Pinatubo, Philippines, and Bingham Canyon, Utah, USA, *Miner. Deposita*, 36(8), 799–806, doi:10.1007/s001260100209.
- Huang, C., L. S. Davis, and J. R. G. Townshend (2002), An assessment of support vector machines for land cover classification, *Int. J. Remote Sens.*, 23(4), 725–749, doi:10.1080/01431160110040323.
- Huang, C., K. Song, S. Kim, J. R. Townshend, P. Davis, J. G. Masek, and S. N. Goward (2008), Use of a dark object concept and support vector machines to automate forest cover change analysis, *Remote Sens. Environ.*, 112(3), 970–985, doi:10.1016/j.rse.2007.07.023.
- Isacks, B. L. (1988), Uplift of the central Andean plateau and bending of the Bolivian orocline, *J. Geophys. Res.*, 93(B4), 3211–3231, doi:10.1029/JB093iB04p03211.
- Jarrard, R. D. (1986), Relations among subduction parameters, *Rev. Geophys.*, 24(2), 217–284, doi:10.1029/RG024i002p00217.
- Jingwen, M., C. Yanbo, C. Maohong, and F. Pirajno (2013), Major types and time-space distribution of mesozoic ore deposits in south China and their geodynamic settings, *Miner. Deposita*, 48(3), 267–294, doi:10.1007/s00126-012-0446-z.
- Kay, S. M., and C. Mpodozis (2001), Central Andean ore deposits linked to evolving shallow subduction systems and thickening crust, *GSA Today*, 11(3), 4–9.
- Kesler, S. E., and B. H. Wilkinson (2008), Earth's copper resources estimated from tectonic diffusion of porphyry copper deposits, *Geology*, 36(3), 255–258.
- Khomich, V. G., N. G. Boriskina, and M. Santosh (2014), A geodynamic perspective of world-class gold deposits in East Asia, *Gondwana Res.*, 26(3), 816–833, doi:10.1016/j.gr.2014.05.007.
- Landgrebe, T. C. W., A. S. Merdith, A. Dutkiewicz, and R. D. Müller (2013), Relationships between palaeogeography and opal occurrence in Australia: A data-mining approach, *Comput. Geosci.*, 56, 76–82, doi:10.1016/j.cageo.2013.02.002.
- Macpherson, C. G., and R. Hall (2002), Timing and tectonic controls in the evolving orogen of SE Asia and the western Pacific and some implications for ore generation, *Geol. Soc. London Spec. Publ.*, 204(1), 49–67.
- Maloney, K. T., G. L. Clarke, K. A. Klepeis, and L. Quevedo (2013), The Late Jurassic to present evolution of the Andean margin: Drivers and the geological record, *Tectonics*, 32(5), 1049–1065, doi:10.1029/tect.20067.
- McGeary, S., A. Nur, and Z. Ben-Avraham (1985), Spatial gaps in arc volcanism: The effect of collision or subduction of oceanic plateaus, *Tectonophysics*, 119(1–4), 195–221, doi:10.1016/0040-1951(85)90039-3.
- McKay, G., and J. R. Harris (2016), Comparison of the data-driven random forests model and a knowledge-driven method for mineral prospectivity mapping: A case study for gold deposits around the Huritz Group and Nueltin Suite, Nunavut, Canada, *Nat. Resour. Res.*, 25(2), 125–143, doi:10.1007/s11053-015-9274-z.
- Merdith, A. S., T. C. W. Landgrebe, A. Dutkiewicz, and R. D. Müller (2013), Towards a predictive model for opal exploration using a spatio-temporal data mining approach, *Aust. J. Earth Sci.*, 60(2), 217–229, doi:10.1080/08120099.2012.754793.
- Metcalfe, I. (2011), Tectonic framework and Phanerozoic evolution of Sundaland, *Gondwana Res.*, 19(1), 3–21, doi:10.1016/j.gr.2010.02.016.
- Mountrakis, G., J. Im, and C. Ogole (2011), Support vector machines in remote sensing: A review, *ISPRS J. Photogramm. Remote Sens.*, 66, 247–259, doi:10.1016/j.isprsjprs.2010.11.001.
- O'Neill, C., R. D. Müller, and B. Steinberger (2005), On the uncertainties in hot spot reconstructions and the significance of moving hotspot reference frames, *Geochim. Geophys. Geosyst.*, 6, Q04003, doi:10.1029/2004GC000784.
- Oshiro, T. M., P. S. Perez, and J. A. Baranauskas (2012), How many trees in a random forest? in *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pp. 154–168, Springer, Berlin, doi:10.1007/978-3-642-31537-4\_13.
- Pedregosa, F., et al. (2011), Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.*, 12, 2825–2830.
- Platt, J. (1999), Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, *Advances in Large Margin Classifiers*, 10(3), 61–74.
- Qu, X., Z. Hou, and Y. Li (2004), Melt components derived from a subducted slab in late orogenic ore-bearing porphyries in the Gangdese copper belt, southern Tibetan plateau, *Lithos*, 74(3–4), 131–148, doi:10.1016/j.lithos.2004.01.003.
- Ramos, V. A. (2010), The tectonic regime along the Andes: Present-day and mesozoic regimes, *Geol. J.*, 45(1), 2–25, doi:10.1002/gj.1193.
- Richards, J. P. (2003), Tectono-magmatic precursors for porphyry cu-(mo-au) deposit formation, *Econ. Geol.*, 98(8), 1515–1533.
- Richards, J. P. (2013), Giant ore deposits formed by optimal alignments and combinations of geological processes, *Nat. Geosci.*, 6(11), 911–916, doi:10.1038/ngeo1920.
- Rodríguez-Galiano, V., M. Sanchez-Castillo, M. Chica-Olmo, and M. Chica-Rivas (2015), Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines, *Ore Geol. Rev.*, 71, 804–818, doi:10.1016/j.oregeorev.2015.01.001.
- Rosenbaum, G., D. Giles, M. Saxon, P. Betts, R. Weinberg, and C. Duboz (2005), Subduction of the Nazca ridge and the Inca plateau: Insights into the formation of ore deposits in Peru, *Earth Planet. Sci. Lett.*, 239(1–2), 18–32.
- Rosenbaum, G., and W. Mo (2011), Tectonic and magmatic responses to the subduction of high bathymetric relief, *Gondwana Res.*, 19(3), 571–582, doi:10.1016/j.gr.2010.10.007.
- Schellart, W. P. (2005), Influence of the subducting plate velocity on the geometry of the slab and migration of the subduction hinge, *Earth Planet. Sci. Lett.*, 231, 197–219, doi:10.1016/j.epsl.2004.12.019.
- Schellart, W. P. (2008), Overriding plate shortening and extension above subduction zones: A parametric study to explain formation of the Andes mountains, *Geol. Soc. Am. Bull.*, 120(11–12), 1441.
- Schedde, R. (2011), Recent trends in gold discovery, in Keynote address to the NewGenGold Conference, Perth, November.
- Seton, M., et al. (2012), Global continental and ocean basin reconstructions since 200 Ma, *Earth Sci. Rev.*, 113, 212–270, doi:10.1016/j.earscirev.2012.03.002.
- Shatwell, D. (2004), Subducted ridges, magmas, differential uplift and gold deposits: Examples from South and central America, in *The Ishihara Symposium: Granites and Associated Metallogenesis*, pp. 115–120, Geoscience, Australia.
- Sillitoe, R. H. (1972), A plate tectonic model for the origin of porphyry copper deposits, *Econ. Geol.*, 67(2), 184–197.
- Sillitoe, R. H. (1998), Major regional factors favouring large size, high hypogene grade, elevated gold content and supergene oxidation and enrichment of porphyry copper deposits, in *Porphyry and hydrothermal Copper and gold deposits-a global perspective*. Conference proceedings (Porter), TM, pp. 21–34.
- Sillitoe, R. H. (2010), Porphyry copper systems, *Econ. Geol.*, 105(1), 3–41.
- Singer, D., V. Berger, and B. Moring (2008), Porphyry copper deposits of the world: Database and grade and tonnage models, *U.S. Geol. Surv. Open File Rep.*, 2008–1155.

- Stegman, D., J. Freeman, W. P. Schellart, L. Moresi, and D. May (2006), Influence of trench width on subduction hinge retreat rates in 3-D models of slab rollback, *Geochem. Geophys. Geosyst.*, 7, Q03012, doi:10.1029/2005GC001056.
- Thiéblemont, D., G. Stein, and J.-L. Lescuyer (1997), Gisements épithermaux et porphyriques: la connexion adakite, *Comptes Rendus de l'Académie des Sciences - Series IIA - Earth Planet. Sci.*, 325(2), 103–109, doi:10.1016/S1251-8050(97)83970-5.
- Tosdal, R., and J. Richards (2001), Magmatic and structural controls on the development of porphyry Cu±Mo±Au deposits, *Rev. Econ. Geol.*, 14, 157–181.
- Waske, B., J. A. Benediktsson, K. Arnason, and J. R. Sveinsson (2009), Mapping of hyperspectral AVIRIS data using machine learning algorithms, *Can. J. Remote Sens.*, 35, 106–116.
- Zahirovic, S., M. Seton, and R. D. Müller (2014), The Cretaceous and Cenozoic tectonic evolution of Southeast Asia, *Solid Earth*, 5(1), 227, doi:10.5194/se-5-227-2014.
- Zuo, R., and E. J. M. Carranza (2011), Support vector machine: A tool for mapping mineral prospectivity, *Comput. Geosci.*, 37(12), 1967–1975, doi:10.1016/j.cageo.2010.09.014.



*Tectonics*

Supporting Information for

**Tectonic environments of South American porphyry-copper magmatism through time revealed by spatio-temporal data mining**

N. Butterworth<sup>1</sup>, D. Steinberg<sup>2</sup>, R.D. Müller<sup>1</sup>, S. Williams<sup>1</sup>, A. Merdith<sup>1</sup>, S. Hardy<sup>2</sup>

<sup>1</sup>EarthByte, School of Geosciences, The University of Sydney, Australia.

<sup>2</sup>National ICT Australia, Eveleigh, Australia.

Butterworth, N., D. Steinberg,

*Tectonics*, 35, doi:10.1002/ 2016TC004289.

**Contents of this file**

- Introduction
- Captions for Figures S1 to S4
- Caption for Movie S1
- Captions for Datasets S1 and S2

**Additional Supporting Information (Files uploaded separately)**

- Figures S1 to S4
- Movie S1
- Dataset S1
- Dataset S2
- Software S1
- Software S2
- Software S3

**Introduction**

The supporting information presented here complements the main text of the paper.

Figures S1-S4 are extra results we obtained in our analysis but found no strong correlations with amongst the data. However, the results may be useful for comparing with future work (e.g. with alternate plate models).

The Movie S1 provides a visual representation for the methodology we have undertaken and aides in digesting the data analysis processes.

Our datasets we developed and used for the analysis, along with the Python codes we used for our methodology provide the opportunity for transparency and reproducibility of the main results in the paper. More information can be found within the scripts themselves. The two 'pickled' datasets can be read using the Python 'Pickle' Package. Both have the same data format (described below). These datasets were developed using the routines and methods in convergence.py and Utils\_coreg.py. The data are to be used with '**Butterworth\_et al\_MLresults.ipynb**' (Software S3) for Machine Learning analysis. This IPython notebook steps through the method for applying Machine Learning to the data described above. Read '**Butterworth\_et al\_MLresults.ipynb**' (Software S3) for detailed information. '**convergence.py**' (Software S1) is used to develop plate kinematic data from the plate reconstruction software GPlates ([www.GPlates.org](http://www.GPlates.org)). Read the script '**convergence.py**' (Software S1) for more information. '**Utils\_coreg.py**' (Software S2) contains several functions used by each of the scripts. Please see '**Utils\_coreg.py**' (Software S2) for more information as each function is described therein.

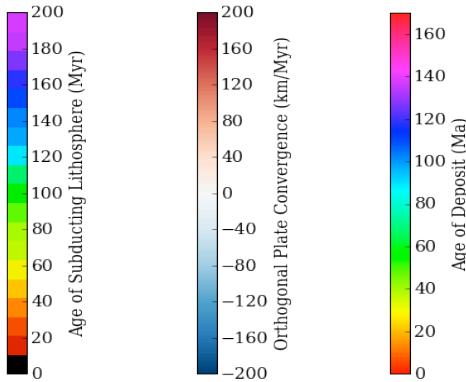
**Figure S1.** Point of formation data analysis. In this partitioning of the data there are 147 cases of "ore deposits", 154 cases of the "non-deposits". The slight disparity between the number of deposits and non-deposits is because seven of the deposit data have ages older than 200 Ma (the extent of our plate model) thus are removed from the analysis, however their locations are used to determine the non-deposit locations. Each bar graph shows the distribution of a particular parameter's values. Each scatter plot is one parameter against another as represented by its respective position in the row and column. Red colours reflect the deposit (or positive) data and blue represent the non-deposits (or negative) data. Here, the non-deposits data are created by assigning random ages to the deposit dataset. From left to right (and top to bottom) the parameters are, Calculated Slab Dip, Subducting Slab Age, Distance from Trench Edge, Subducting Plate Normal Velocity, Subducting Plate Parallel Velocity, Overriding Plate Normal Velocity, Overriding Plate Parallel Velocity, Convergence Normal Velocity, Convergence Parallel Velocity, Subduction Polarity, and Subduction Obliquity. The final analysis only utilised Subducting Slab Age, Distance from Trench Edge, Convergence Normal Velocity, and Subduction Obliquity.

**Figure S2.** Point of formation with restriction of time period data analysis. In this partitioning of the data there are 121 cases of "ore deposits", 67 cases of the "non-deposits". Each bar graph shows the distribution of a particular parameter's values. Each scatter plot is one parameter against another as represented by its respective position in the row and column. Red colours reflect the deposit (or positive) data and blue represent the non-deposits (or negative) data. Here, the non-deposits data are created by assigning random ages to the deposit dataset. From left to right (and top to bottom) the parameters are, Calculated Slab Dip, Subducting Slab Age, Distance from Trench Edge, Subducting Plate Normal Velocity, Subducting Plate Parallel Velocity, Overriding Plate Normal Velocity, Overriding Plate Parallel Velocity, Convergence Normal Velocity, Convergence Parallel Velocity, Subduction Polarity, and Subduction Obliquity. The final analysis only utilised Subducting Slab Age, Distance from Trench Edge, Convergence Normal Velocity, and Subduction Obliquity.

**Figure S3.** Area of formation data analysis. In this partitioning of the data there are 32 cases of “ore deposits”, 45 cases of the “non-deposits”. Each bar graph shows the distribution of a particular parameter’s values. Each scatter plot is one parameter against another as represented by its respective position in the row and column. Red colours reflect the deposit (or positive) data and blue represent the non-deposits (or negative) data. Here, the non-deposits data are created by assigning random ages to the deposit dataset. From left to right (and top to bottom) the parameters are, Calculated Slab Dip, Subducting Slab Age, Distance from Trench Edge, Subducting Plate Normal Velocity, Subducting Plate Parallel Velocity, Overriding Plate Normal Velocity, Overriding Plate Parallel Velocity, Convergence Normal Velocity, Convergence Parallel Velocity, Subduction Polarity, and Subduction Obliquity. The final analysis only utilised Subducting Slab Age, Distance from Trench Edge, Convergence Normal Velocity, and Subduction Obliquity.

**Figure S4.** Area of formation with restriction of time period data analysis. In this partitioning of the data there are 45 cases of “ore deposits”, 176cases of the “non-deposits”. Each bar graph shows the distribution of a particular parameter’s values. Each scatter plot is one parameter against another as represented by its respective position in the row and column. Red colours reflect the deposit (or positive) data and blue represent the non-deposits (or negative) data. Here, the non-deposits data are created by assigning random ages to the deposit dataset. From left to right (and top to bottom) the parameters are, Calculated Slab Dip, Subducting Slab Age, Distance from Trench Edge, Subducting Plate Normal Velocity, Subducting Plate Parallel Velocity, Overriding Plate Normal Velocity, Overriding Plate Parallel Velocity, Convergence Normal Velocity, Convergence Parallel Velocity, Subduction Polarity, and Subduction Obliquity. The final analysis only utilised Subducting Slab Age, Distance from Trench Edge, Convergence Normal Velocity, and Subduction Obliquity.

**Movie S1.** Animation of the South American margin through time from 200 Ma to present day. Plate motion vectors are coloured by the plate they belong to. Plate boundaries are coloured blue. Present day continents are shaded green. The age of the seafloor is coloured according to its legend. Convergence rates are shown along subducting plate boundaries and coloured from blue to red as shown in the legend. Porphyry magmatism deposits from Bertrand et al. (2014) appear through the animation at their formation location and age, their ages are coloured according to the legend.



#### **Dataset S1.** data\_coregistered\_bertrand2014.pkl

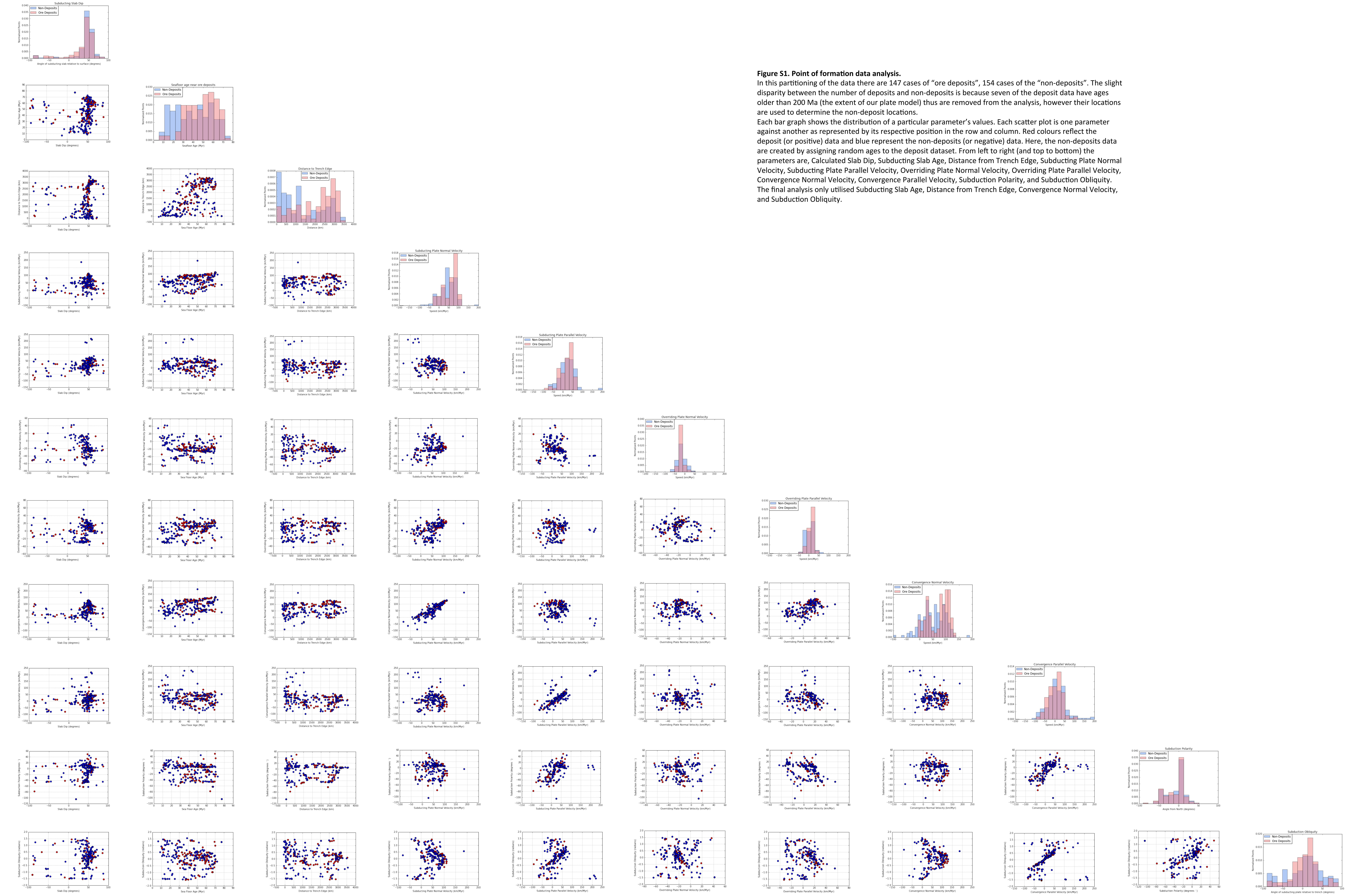
Contains 301 data, coregistered at 20 time periods, with 21 variables. In the Python ‘pickle’ format, readable with the Python Pickle package. The 301 data points contain 154 deposit data and 147 non-deposit data. The 20 time periods begin at the age of mineralisation and go back in time 20 million years in 1 million year increments. The 21 variables are:

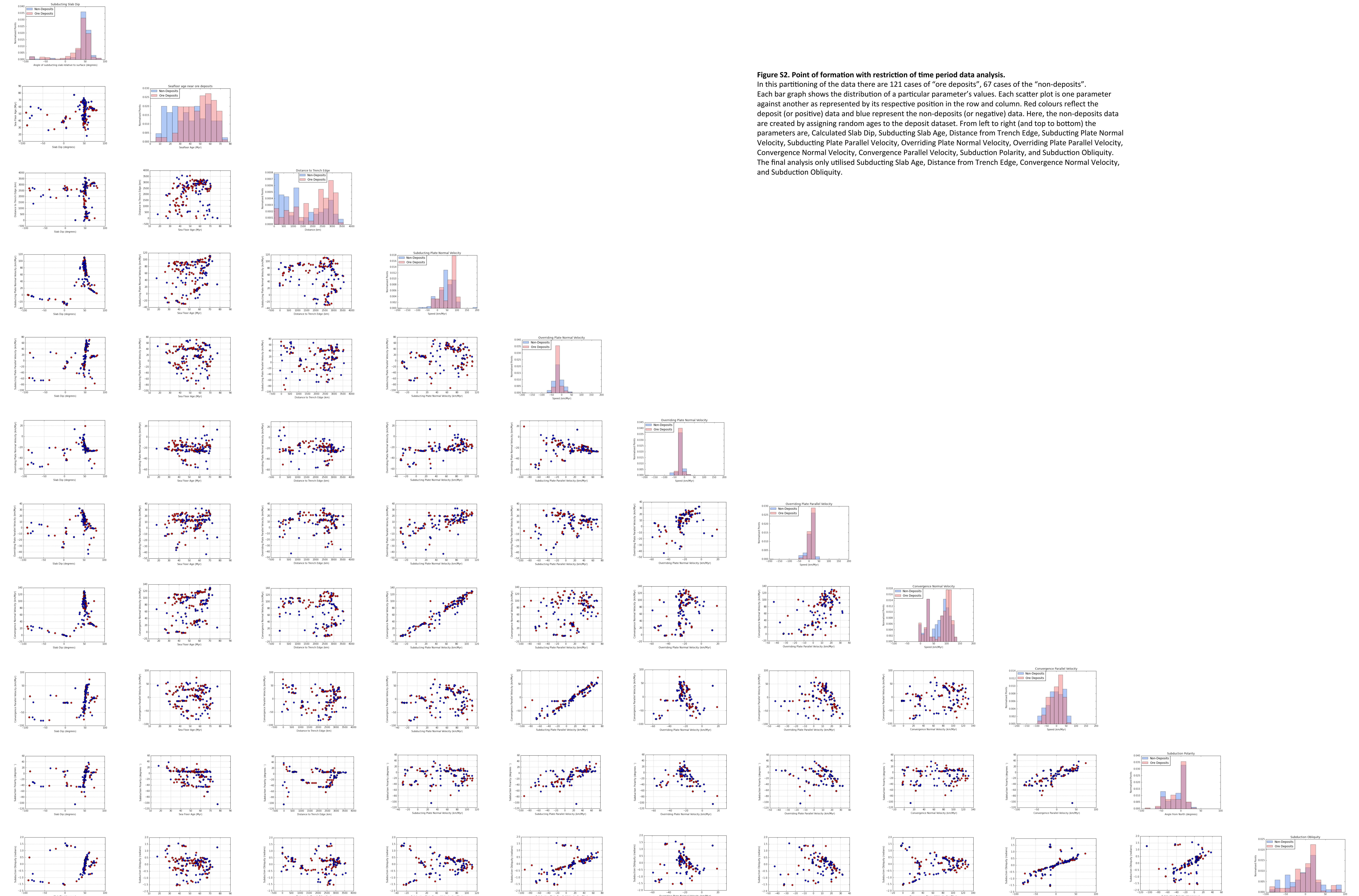
- 0 Present day longitude (degrees)
- 1 Present day latitude (degrees)
- 2 Reconstructed longitude (degrees)
- 3 Reconstructed latitude (degrees)
- 4 Age (Ma)
- 5 Time before mineralisation (Myr)
- 6 Seafloor age (Myr)
- 7 Segment length (km)
- 8 Slab length (km)
- 9 Distance to trench edge (km)
- 10 Subducting plate normal velocity (km/Myr)
- 11 Subducting plate parallel velocity (km/Myr)
- 12 Overriding plate normal velocity (km/Myr)
- 13 Overriding plate parallel velocity (km/Myr)
- 14 Convergence normal rate (km/Myr)
- 15 Convergence parallel rate (km/Myr)
- 16 Subduction polarity (degrees)
- 17 Subduction obliquity (degrees)

- 18 Distance along margin (km)
- 19 Subduction obliquity signed (radians)
- 20 Ore Deposits Binary Flag for ML (1 or 0)

**Dataset S2.** data\_coregistered\_andes.pkl

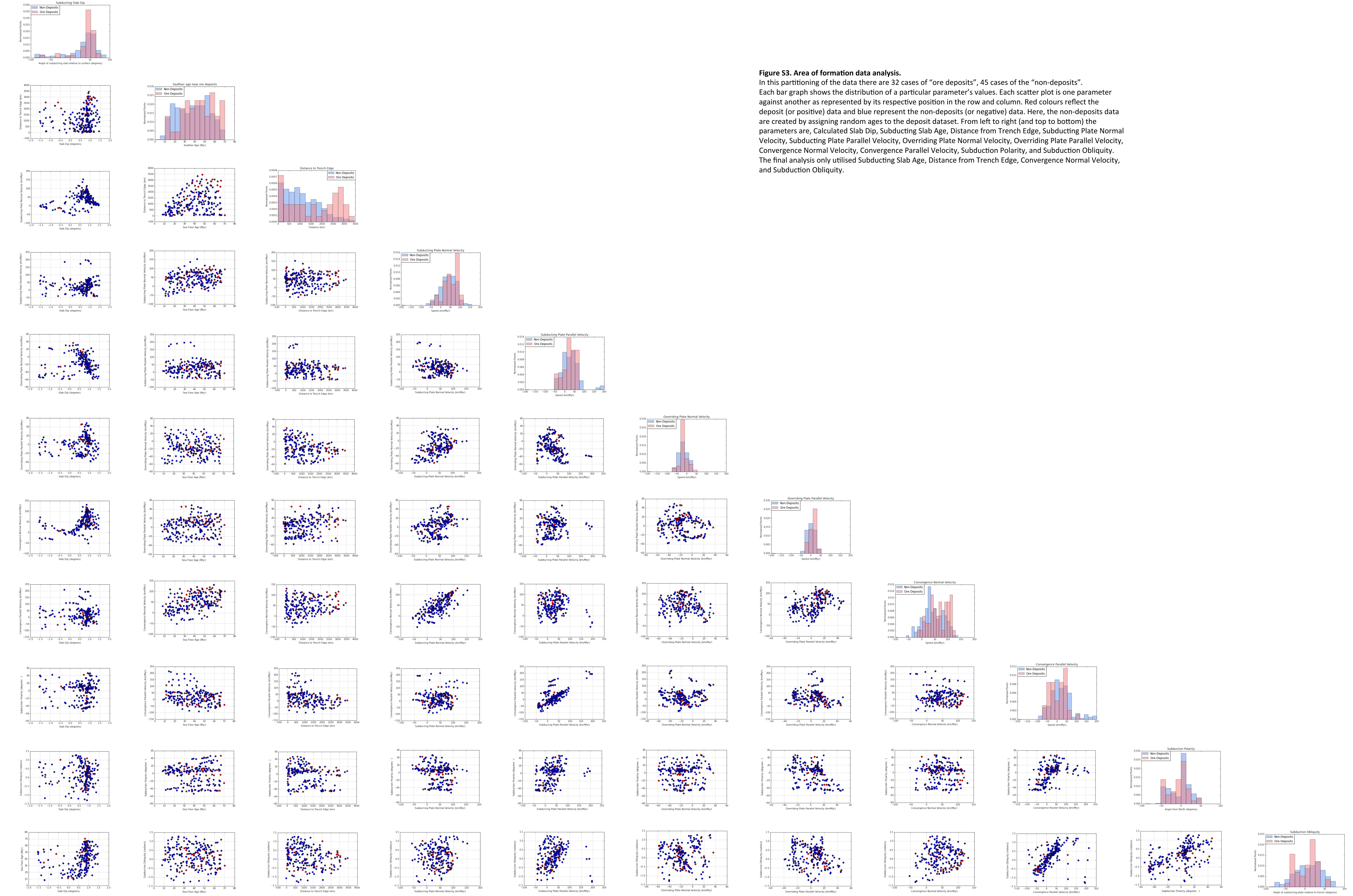
Contains 110 data, coregistered at 200 time periods, with 20 variables. In the Python ‘pickle’ format, readable with the Python Pickle package. The 110 data are the discretization of the South American trench at present day into 0.5 degree increments. The time periods begin at present day and are the reconstructed times back to 200 Ma in 1 Myr increments. The 20 variables are the same as Dataset S1, without the final Ore Deposit Binary Flag (as none of these are ore deposits).





**Figure S2. Point of formation with restriction of time period data analysis.**

In this partitioning of the data there are 121 cases of "ore deposits", 67 cases of the "non-deposits". Each bar graph shows the distribution of a particular parameter's values. Each scatter plot is one parameter against another as represented by its respective position in the row and column. Red colours reflect the deposit (or positive) data and blue represent the non-deposits (or negative) data. Here, the non-deposits data are created by assigning random ages to the deposit dataset. From left to right (and top to bottom) the parameters are, Calculated Slab Dip, Subducting Slab Age, Distance from Trench Edge, Subducting Plate Normal Velocity, Subducting Plate Parallel Velocity, Overriding Plate Normal Velocity, Overriding Plate Parallel Velocity, Convergence Normal Velocity, Convergence Parallel Velocity, Subduction Polarity, and Subduction Obliquity. The final analysis only utilised Subducting Slab Age, Distance from Trench Edge, Convergence Normal Velocity, and Subduction Obliquity.



## Figure S3. Area of formation data analysis.

**Figure S3. Area of formation data analysis.**  
In this partitioning of the data there are 32 cases of “ore deposits”, 45 cases of the “non-deposits”. Each bar graph shows the distribution of a particular parameter’s values. Each scatter plot is one parameter against another as represented by its respective position in the row and column. Red colours reflect the deposit (or positive) data and blue represent the non-deposits (or negative) data. Here, the non-deposits data are created by assigning random ages to the deposit dataset. From left to right (and top to bottom) the parameters are, Calculated Slab Dip, Subducting Slab Age, Distance from Trench Edge, Subducting Plate Normal Velocity, Subducting Plate Parallel Velocity, Overriding Plate Normal Velocity, Overriding Plate Parallel Velocity, Convergence Normal Velocity, Convergence Parallel Velocity, Subduction Polarity, and Subduction Obliquity. The final analysis only utilised Subducting Slab Age, Distance from Trench Edge, Convergence Normal Velocity, and Subduction Obliquity.



**Figure S4. Area of formation with restriction of time period data analysis.**

In this partitioning of the data there are 45 cases of “ore deposits”, 176cases of the “non-deposits”. Each bar graph shows the distribution of a particular parameter’s values. Each scatter plot is one parameter against another as represented by its respective position in the row and column. Red colours reflect the deposit (or positive) data and blue represent the non-deposits (or negative) data. Here, the non-deposits data are created by assigning random ages to the deposit dataset. From left to right (and top to bottom) the parameters are, Calculated Slab Dip, Subducting Slab Age, Distance from Trench Edge, Subducting Plate Normal Velocity, Subducting Plate Parallel Velocity, Overriding Plate Normal Velocity, Overriding Plate Parallel Velocity, Convergence Normal Velocity, Convergence Parallel Velocity, Subduction Polarity, and Subduction Obliquity. The final analysis only utilised Subducting Slab Age, Distance from Trench Edge, Convergence Normal Velocity, and Subduction Obliquity.