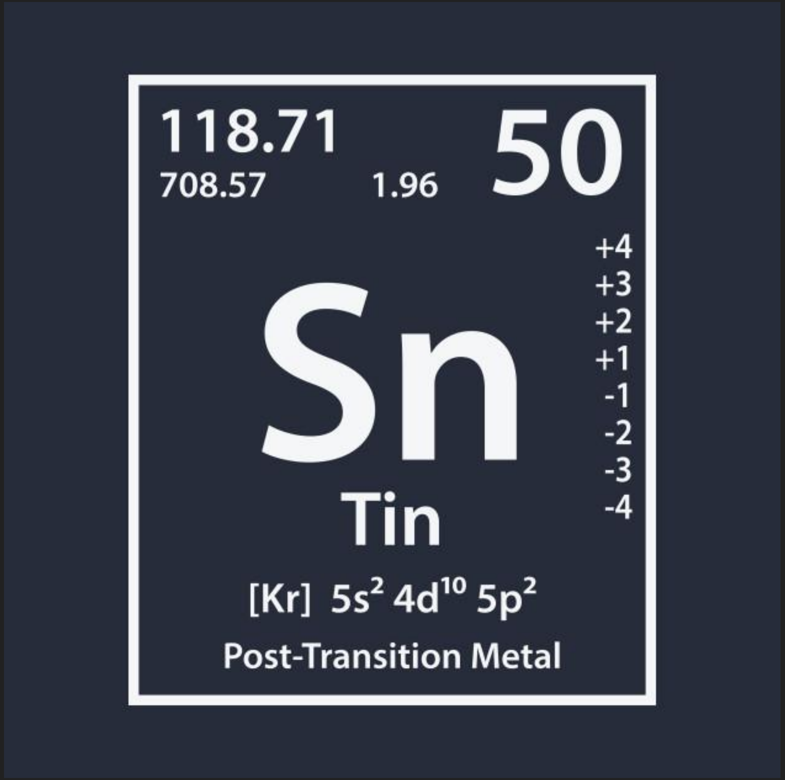
South American Tin Project

Tectonic Environments Revealed through Machine Learning

By Emily Benjamin, Anna Jamieson, Maggie Kiesow, and Ali Oswald

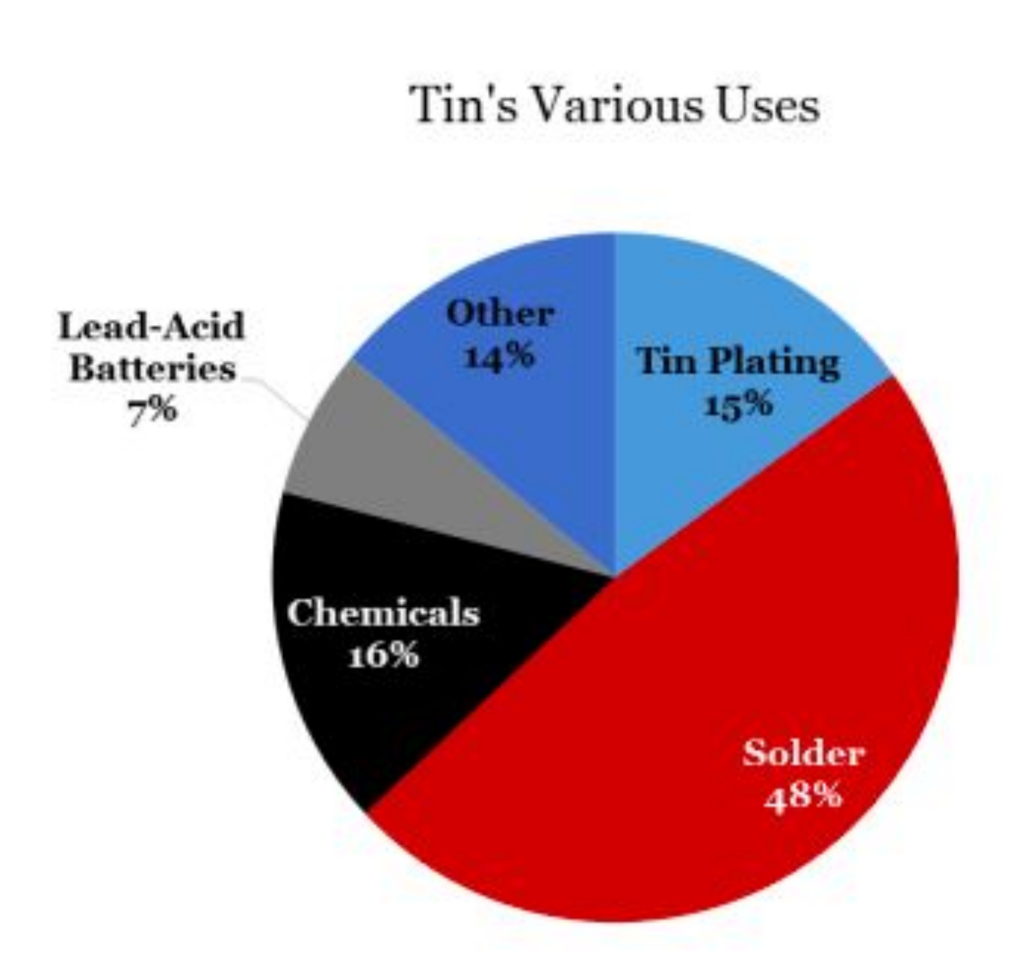
Tectonic environments of Tin in South America revealed through Machine Learning and Data Science



TeePublic, 2019

- Long mining history (~3500 B.C.)
- Accounts for 0.001% of the crust
- Largest global use is solder
- Largest world reserves in China

Anderson, n.d.; Gagnon, n.d.; USGS, 2019



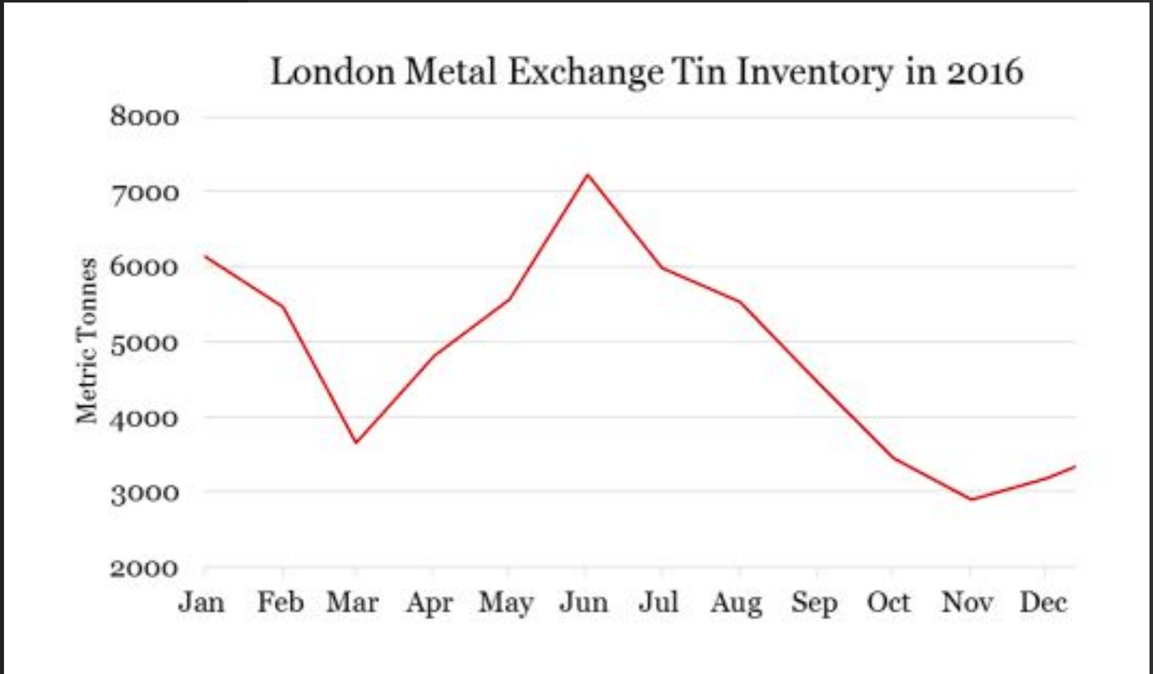
(ITA, 2018)

(ITA, 2019)

Significance

- Decline in global tin
- Uncertain future
- Increasing price

Iskyan, K. (2016)



Iskyan, K. (2016)

Mineralogy

Cassiterite

Craig and Vaughan, D, 2019

Cassiterite

Handbook of Mineralogy, 2019

Associated Minerals

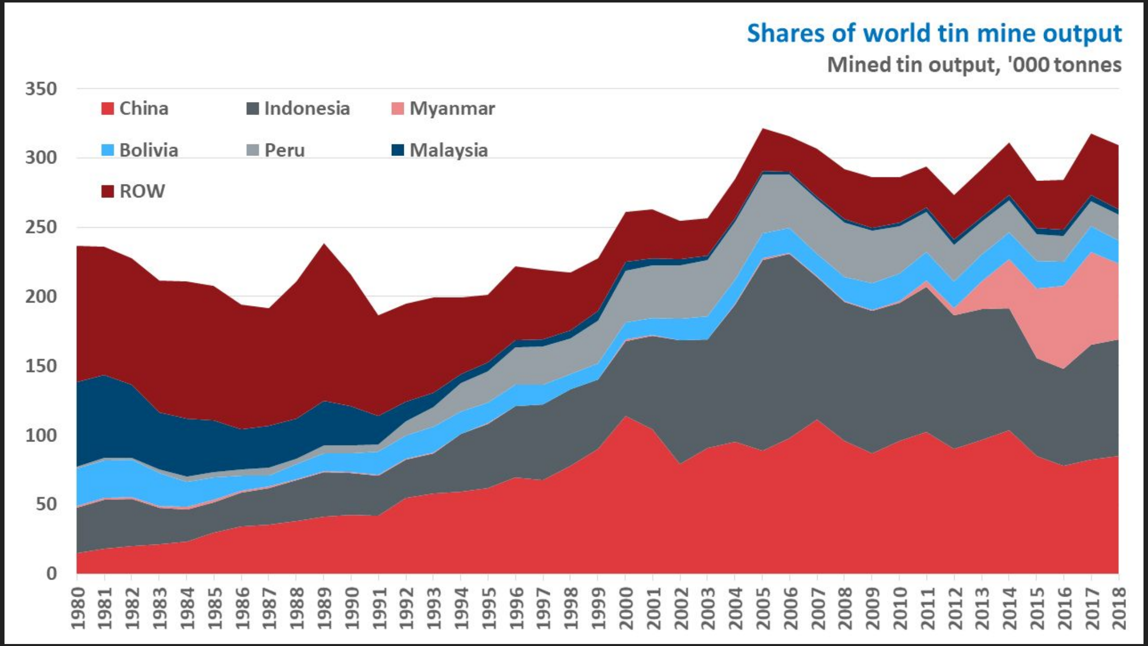
- Quartz
- Muscovite
- Wolframite
- Tourmaline
- Topaz
- Fluorite
- Scheelite
- Lepidolite
- Arsenopyrite
- Bismuth
- molybdenite

Cassiterite

Occurrence

- Medium- to high- temperature hydrothermal veins
- granite pegmatite
- Rhyolite
- Large alluvial placers

Where is it mined?

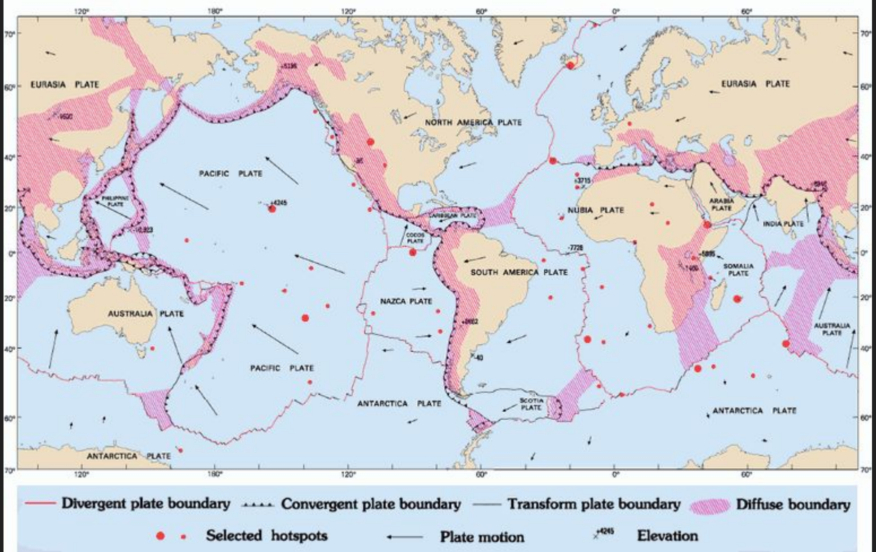


(ITA, 2019)

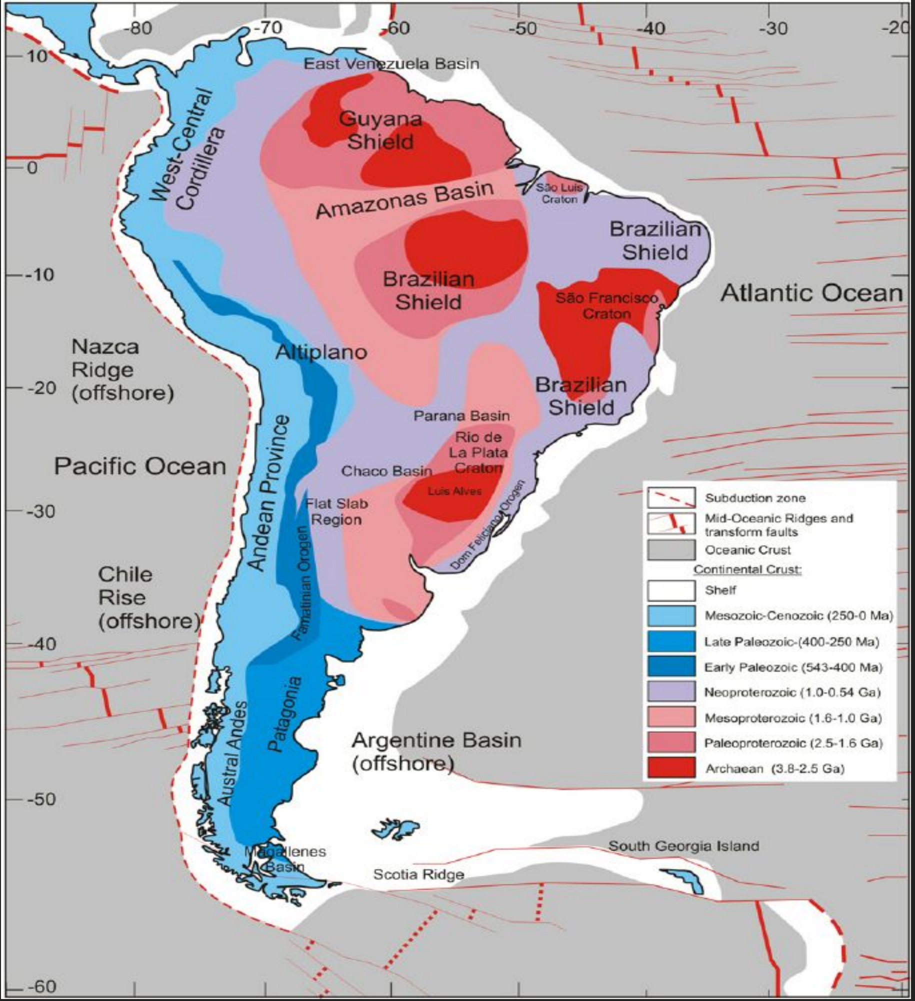
Plate Boundaries

South American Tectonics:

- Major boundary of convergence
- Major Subduction boundary



Alden, 2018



Gubanov & Mooney, 2012

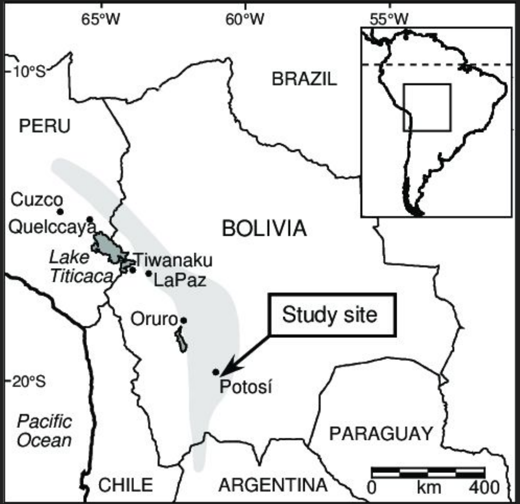
Tectonic Results

Inner Arc of the Central Andes

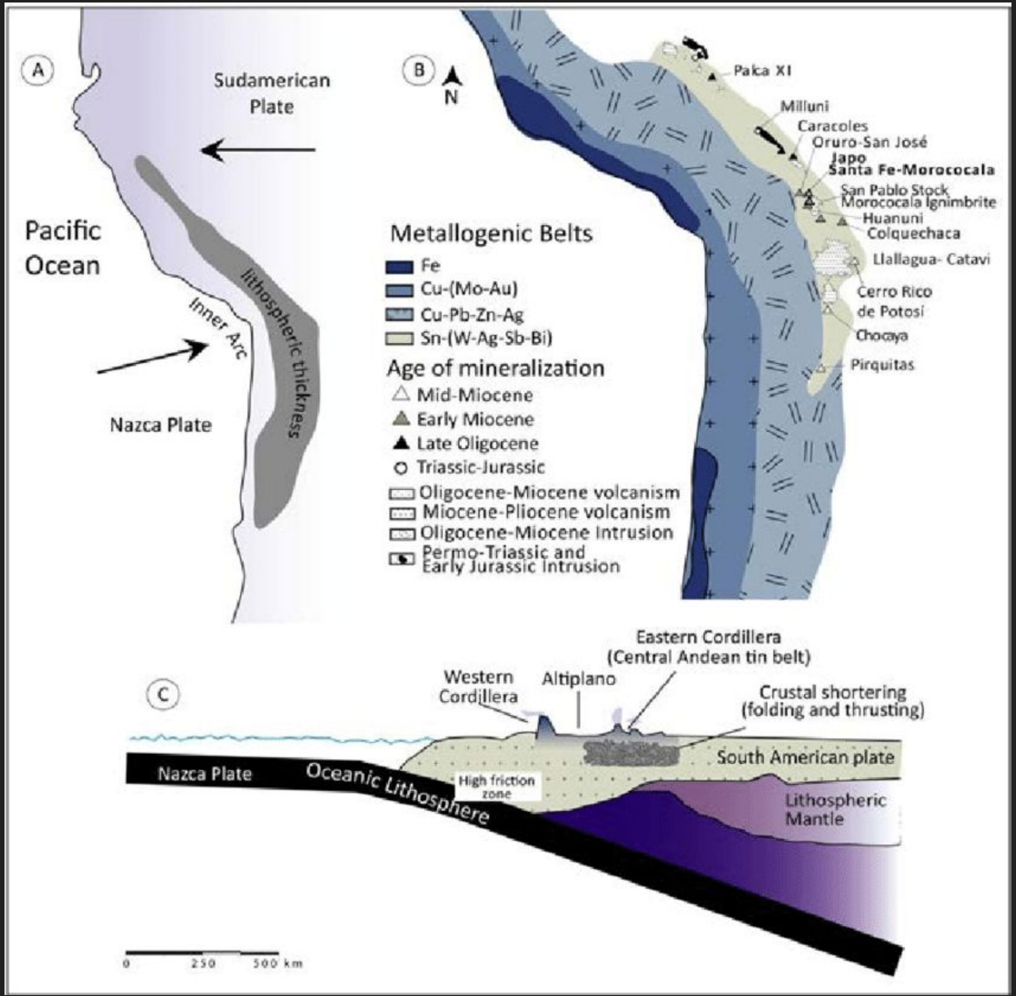
Bolivian Tin Belt

- 900 km
- 15% global tin production

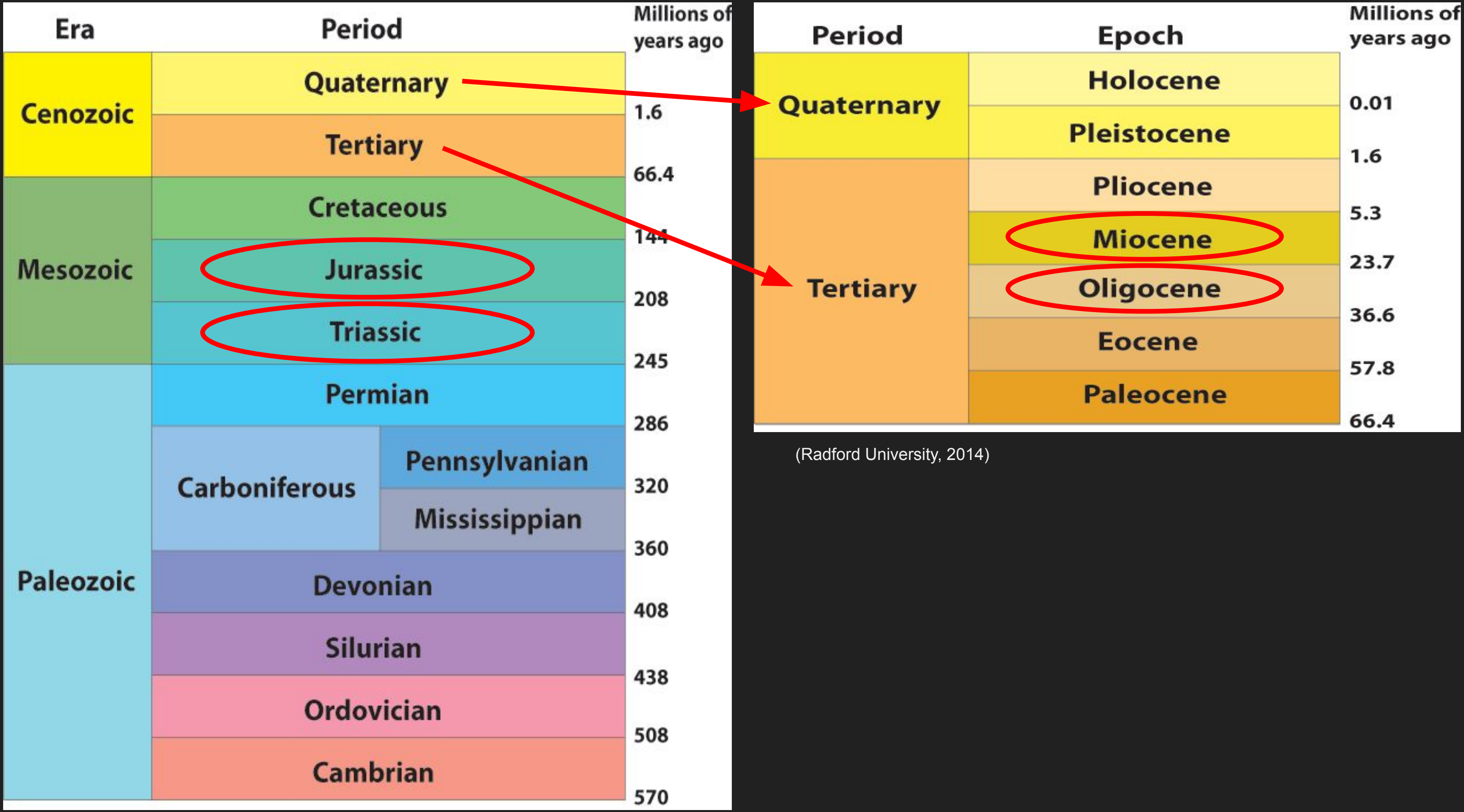
Lehmann et al., 1990



Hu, 2003



Jiménez-Franco et al., 2018

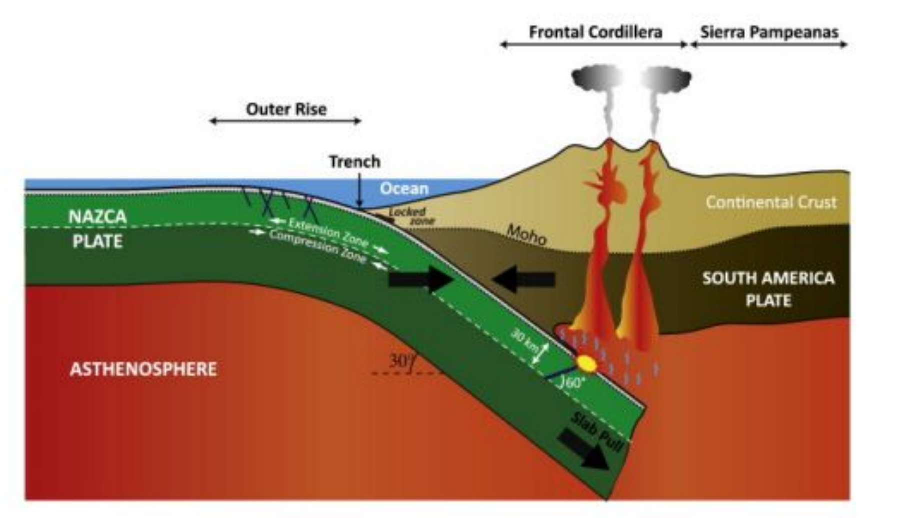


(Radford University, 2014)

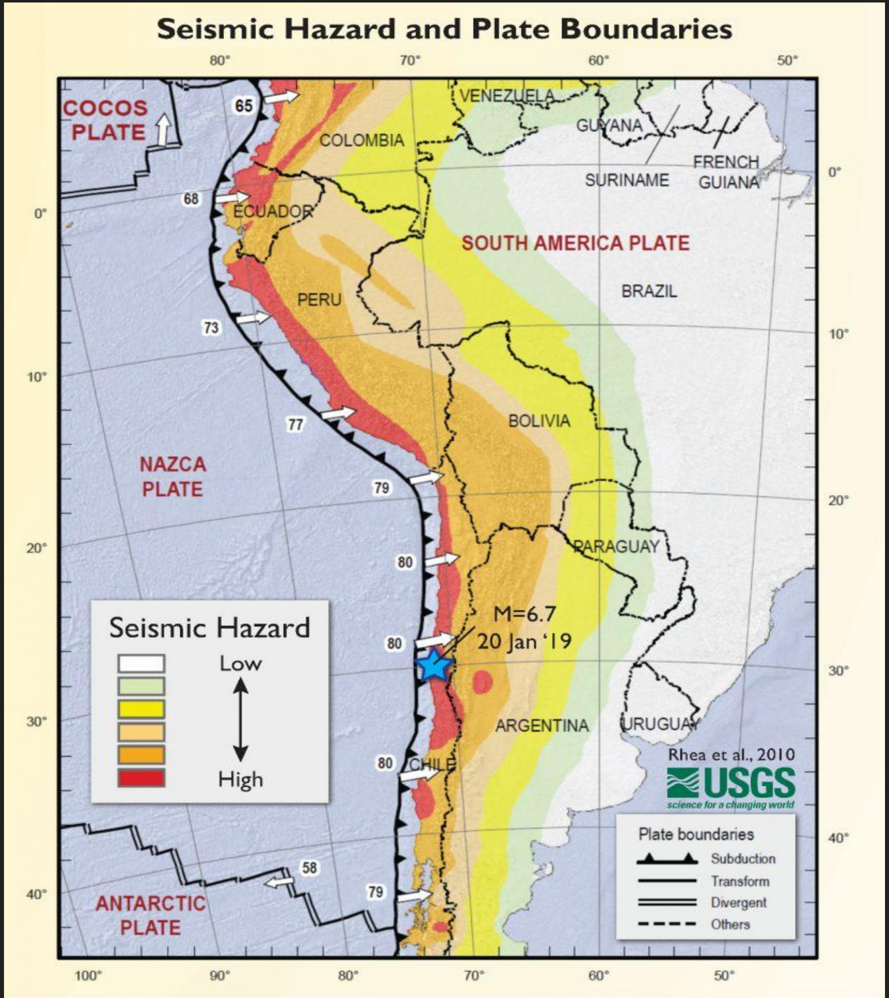
Nazca Subduction

Nazca plate subduction

- Peru-Chile boundary
- Rapid convergence
- Frequent seismic events



Lubecka, 2015



Patton & Ammirati, 2019

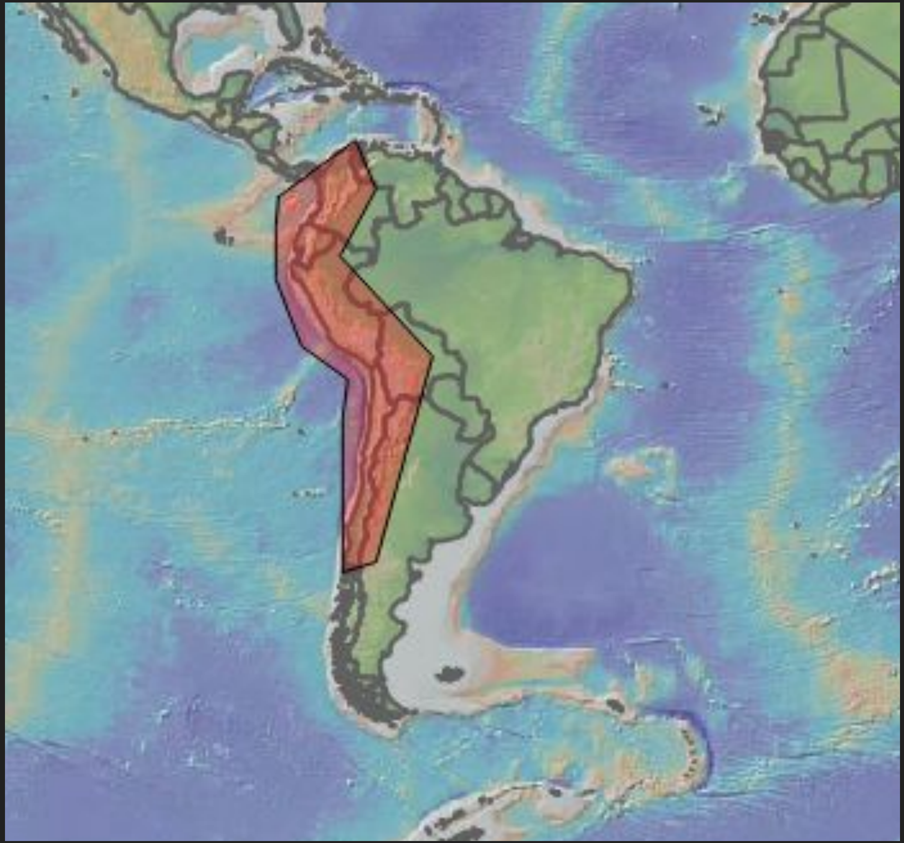
Data Review

EarthChem Data Collective

- GEOROCK
- United States Geological Survey (USGS)

Used Polymer Builder Tool

EarthChem.org, 2019



EarthChem.org, 2019

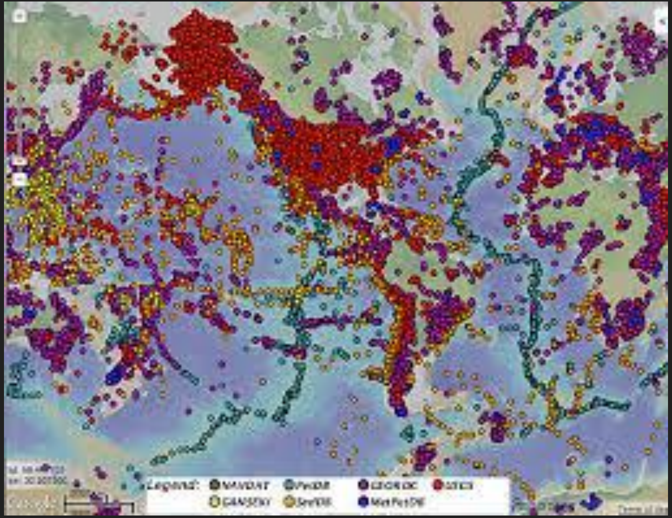
Data Review

- Four Parameters
- 313 Rows
- 230 Columns
- 18 -20 slices

```
print andesPresent.shape

print "loaded datasets"
```

```
(377, 1, 18)
(377, 1, 18)
(313, 230, 18)
loaded datasets
```



EarthChem.org, 2019

```
print "Shape of data array: ", andesData.shape
```

```
Shape of data array:  (299, 7)
```

Benjamin et al, 2019

List of Variables

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 (1 or 0)

Size and Shape

- Isolated the longitude and latitude of South America in coregLoop
- Changed the time period to 230 Ma

Wrangling Data

- Imported pygplates - reconstructed plate model with understanding of subduction kinematics
- Imported basic tools and libraries, machine learning tools and processing tools
- Cleaned up data - use average of points that did not coregister
- Added parameters

Methods through Machine Learning

List of Variables

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 (1 or 0)

Part 1: Data Wrangling

```
print andesPresent.shape

print "loaded datasets"
```

```
(377, 1, 18)
(377, 1, 18)
(313, 230, 18)
loaded datasets
```

```
#Remove data that has zeros for everyhting (or just where the longitude 0)

clean=numpy.where(andes[:,0,1]!=0)
andesClean = cleanCondition(clean,andes)

clean=numpy.where(andesRand[:,0,1]!=0)
andesRandClean = cleanCondition(clean,andesRand)

#Muller 13Ma and 41Ma have issues along the trench,
#take the average of known good values for any points that did not coregister.
for i in xrange(len(andesPresent[0,:,2])):
    for j in xrange(len(andesPresent[:,0,2])):
        if andesPresent[j,i,3]==0:
            andesPresent[j,i,:]=(andesPresent[j,i+1,:]+andesPresent[j,i-1,:])/2
        #If there are nans, just replace them with the next closest timestep
        for k in xrange(len(andesPresent[0,0,:])):
            if numpy.isnan(andesPresent[j,i,k]):
                print andesClean.shape, andes.shape
                print andesRandClean.shape, andesRand.shape
                print andesPresent.shape
```

Creating Arrays

- Array of list of points that match up with present-day trench distance
- Empty array
  - store looped data through time and distance
  - save index and distances of neighbouring points
  - Add subduction obliquity
- Random Data empty array
  - Dataset to test against parameters that do not result in ore deposits
- Combining Random and Ore Deposit datasets

Map of South America showing Tin Deposits Along Trench

Majority younger than 50 Ma

Majority lie between 2000 km and 4000 km

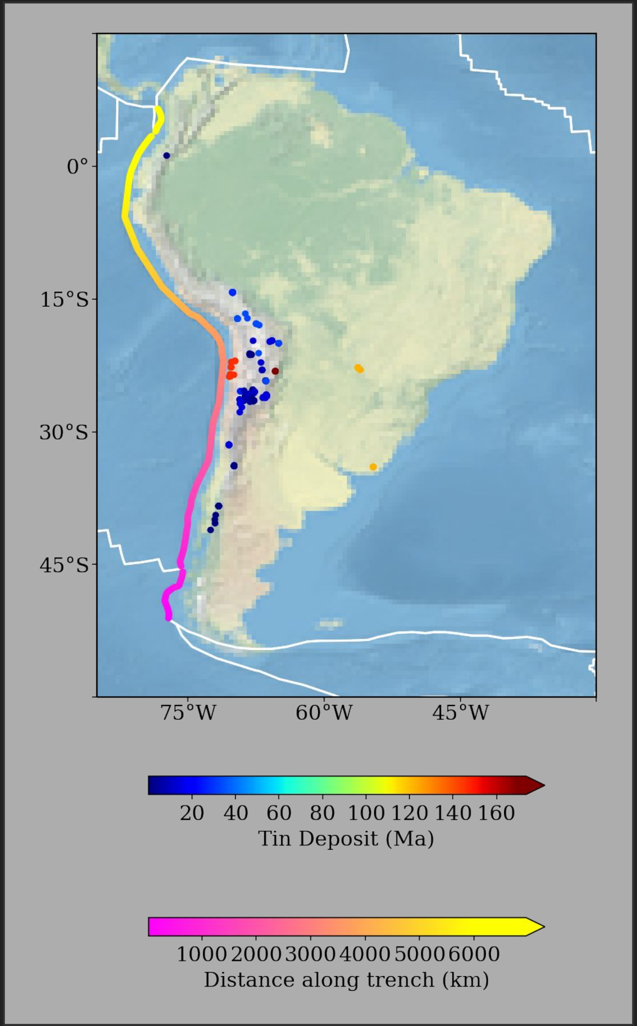
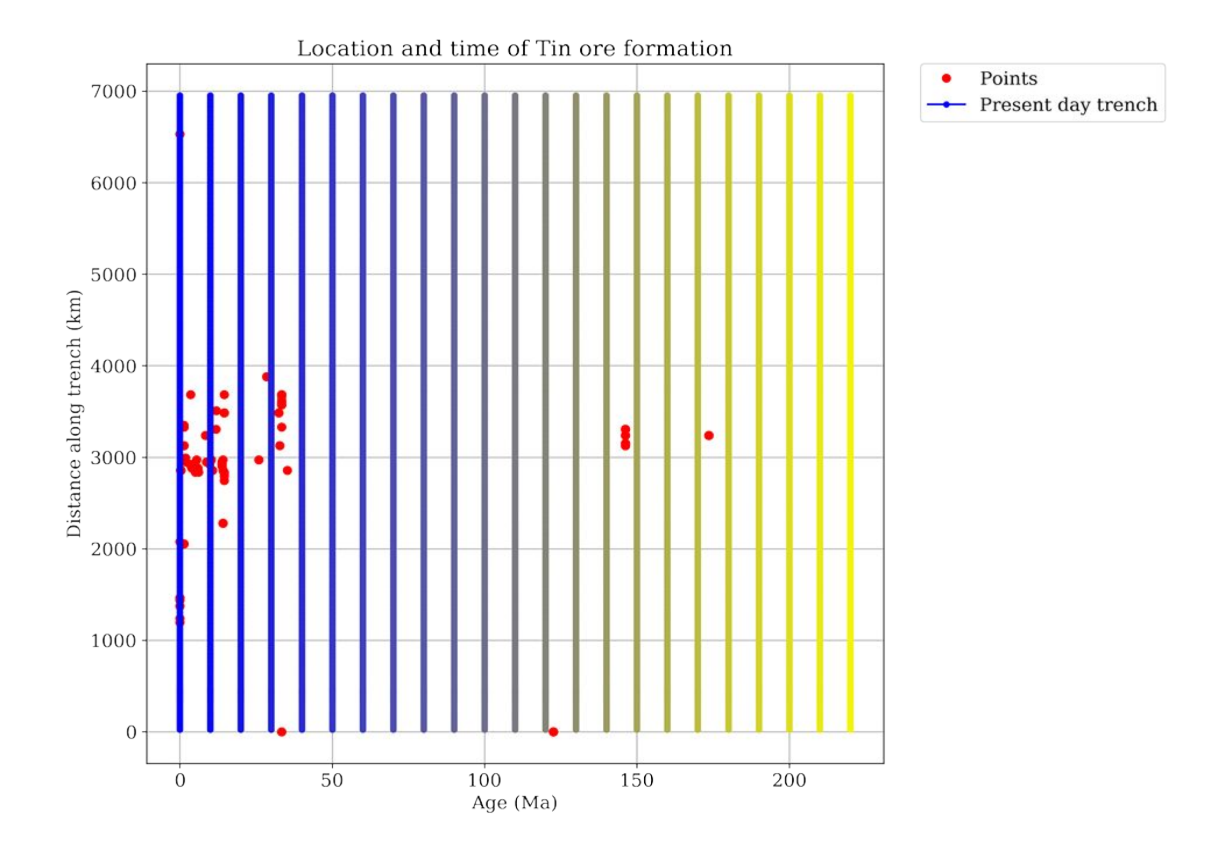


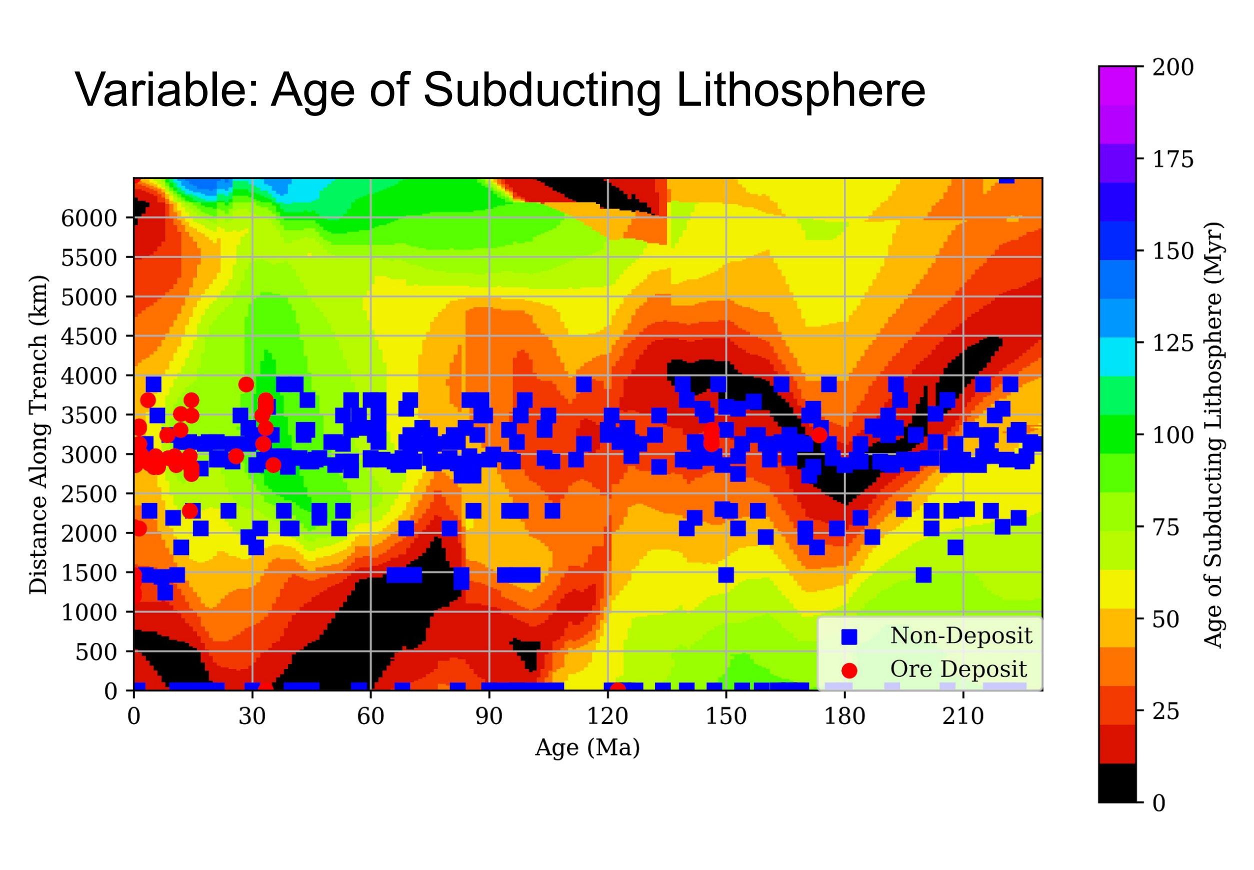
Figure demonstrating merged data set

- Gain basic understanding of the distribution and age of deposits along the trench through time



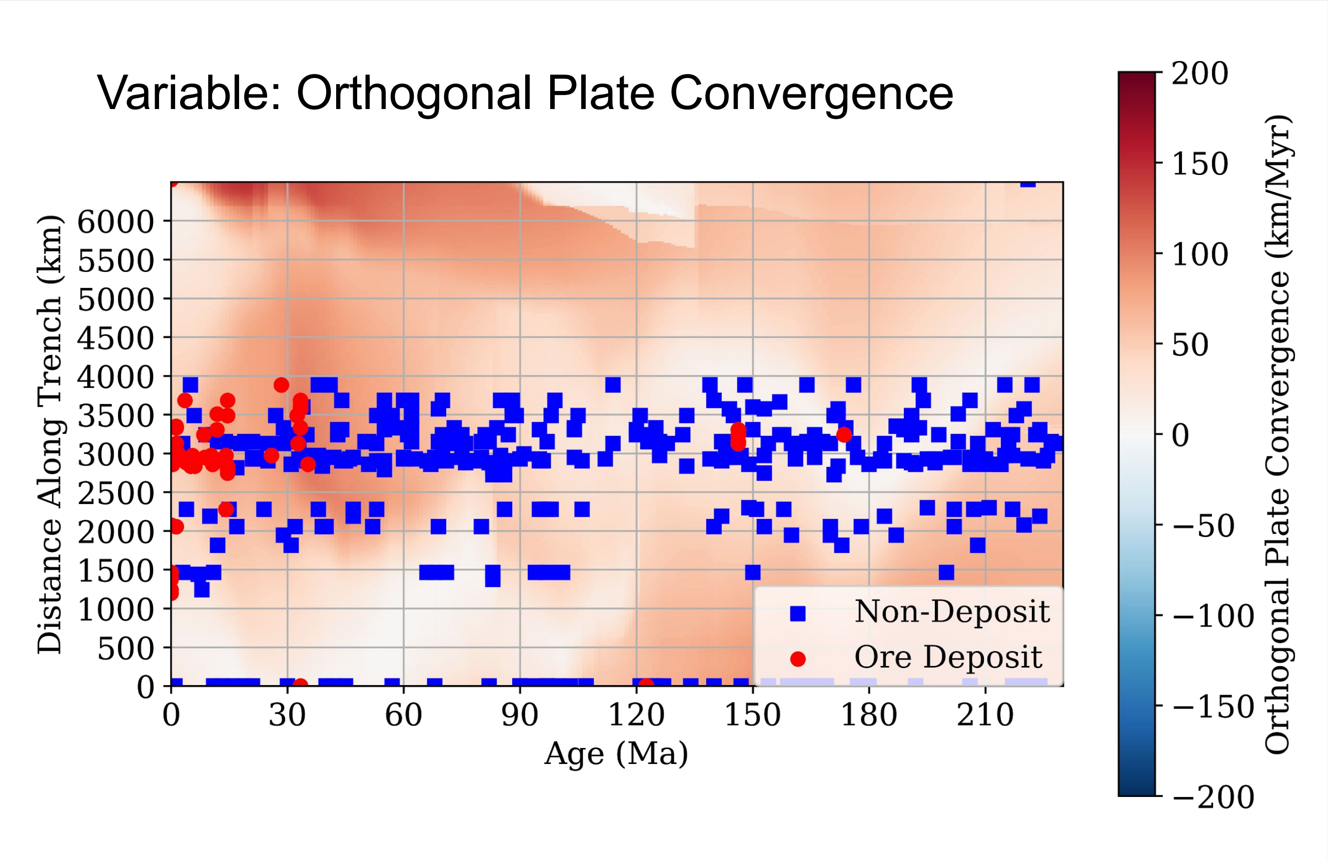
Colour map using spatio-temporal bin parameters

- Spatial domains = distance along trench
- Temporal domains = Age
- Increments set here for data separation
-



Colour map using spatio-temporal bin parameters

- Same spatio-temporal bin parameters
- Different variable



Part 2: Machine Learning

- Setting up the data: Formatting
  - Select parameters
  - Recombine the data
  - Split data into testing and training sets
    - Training shows which are important
    - Testing set test the validity of the training sets
  - Random Forest
  - Support Vector Machine

List of Variables

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)

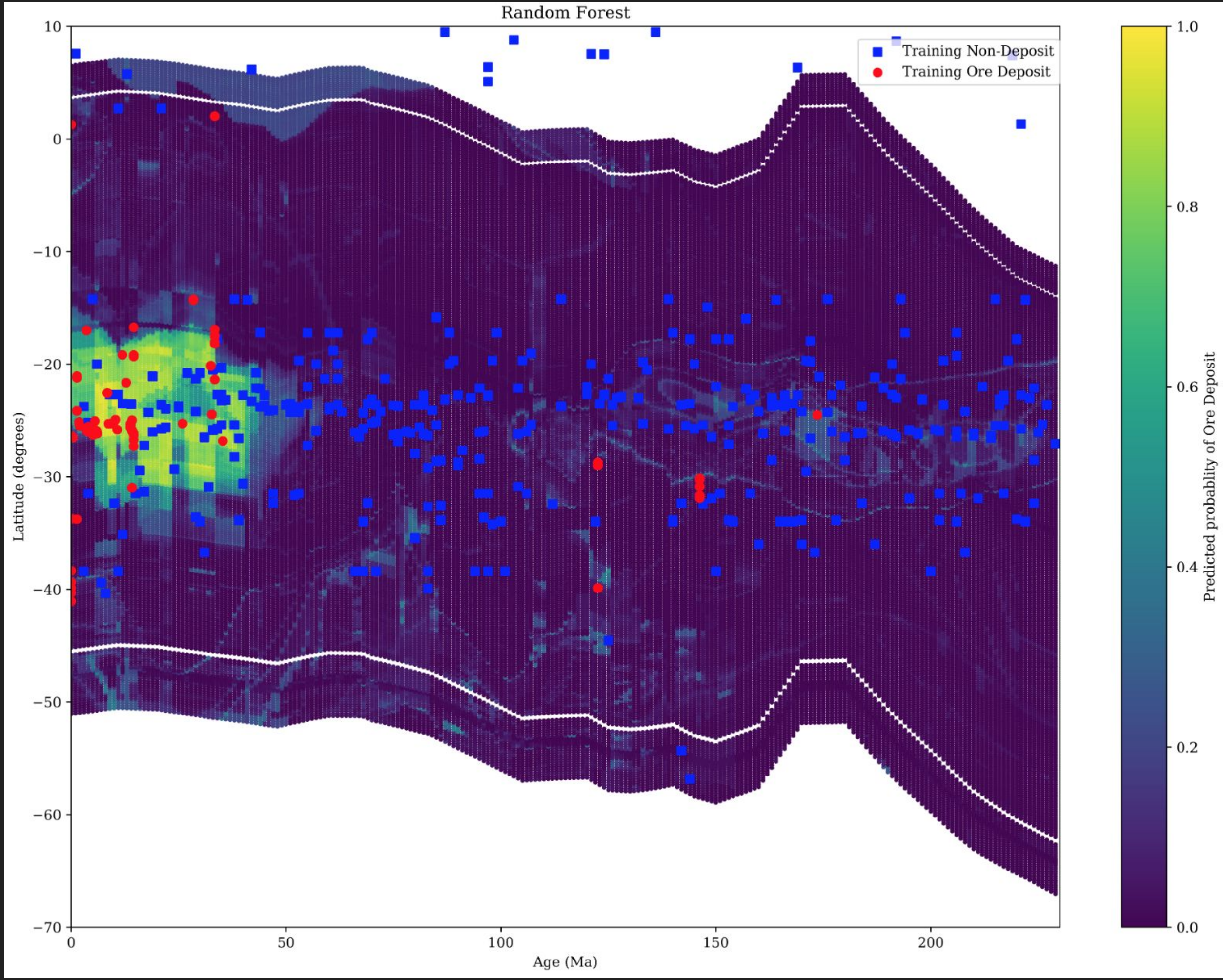
19 Subduction obliquity signed (radians)

20 Ore Deposits Binary Flag (1 or 0)

Random Forests

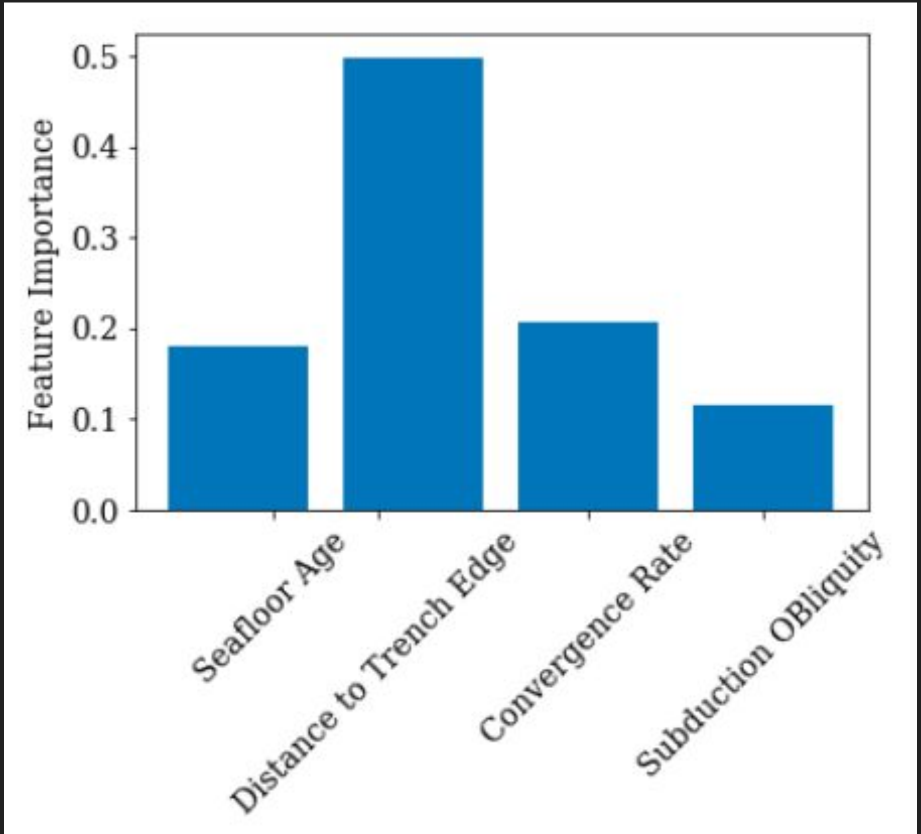
- Non-linear classifier
- Decisions trees - more decision trees the more robust
- Each tree votes for a parameter
- 128 Trees - more consistent response means the correlation is stronger between parameter and deposit
- Each tree predicts the category, assigning new record which is officially signed to the category that majority of trees predicted
- Unbiased, stable

Random Forest - trained data, illustrating the predicted probability of each point across latitude and age as being an ore deposit

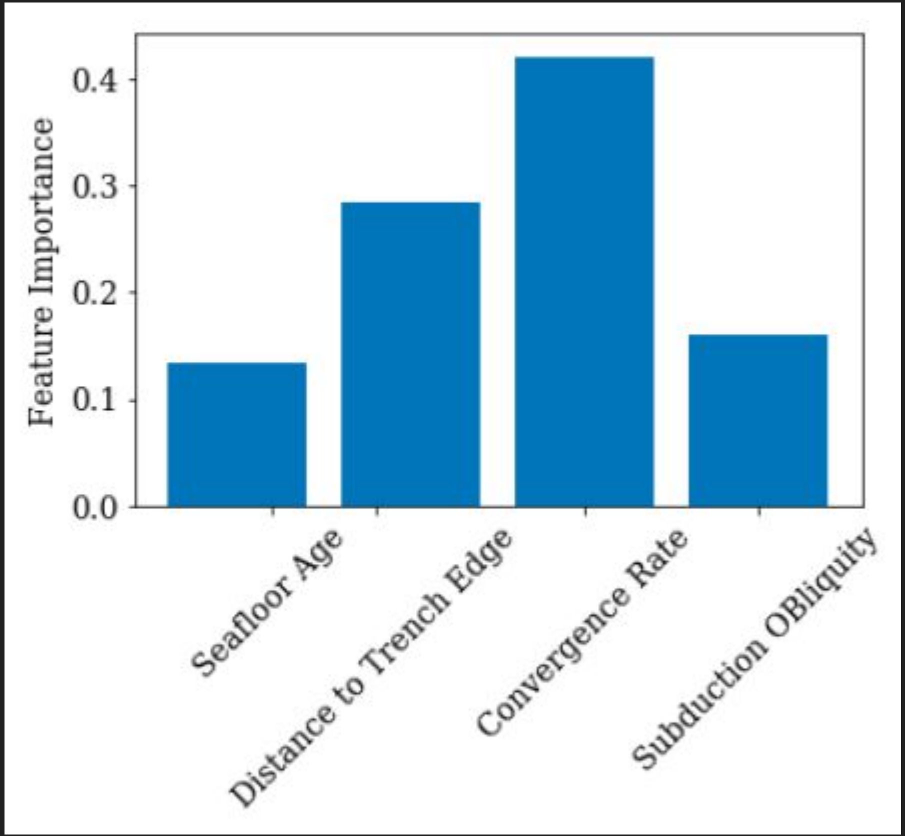


RF Feature Importance

Area of Formation



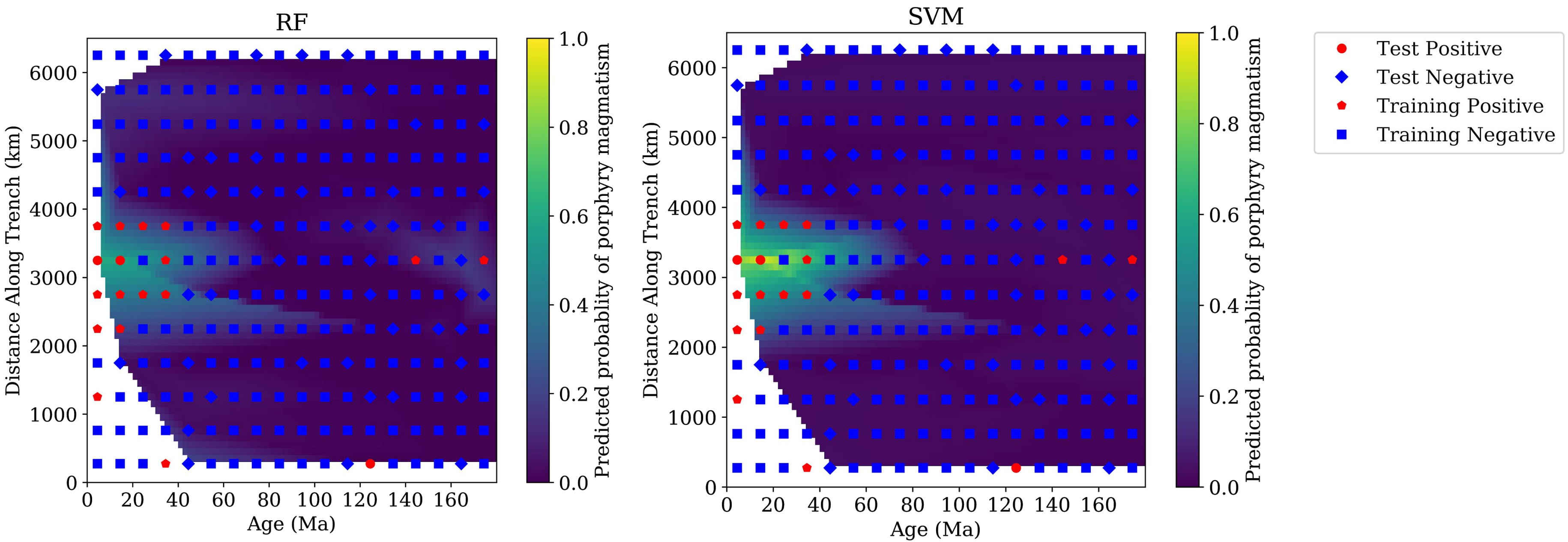
Point of Formation



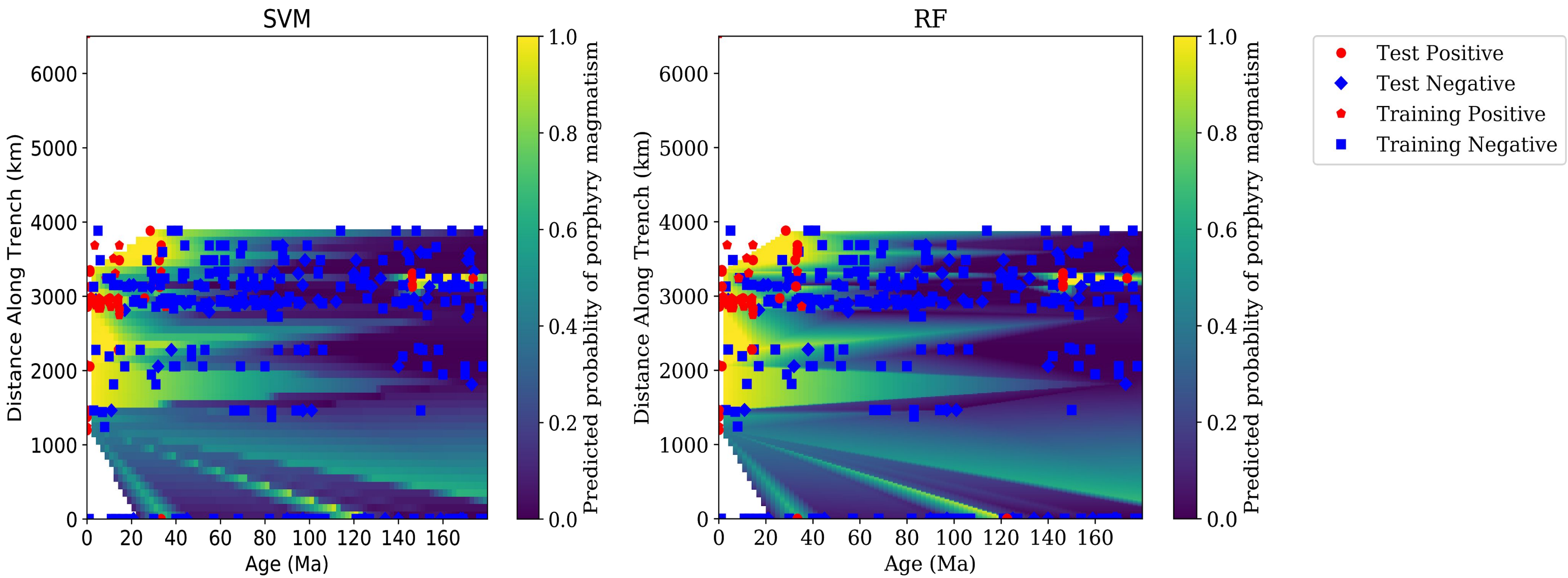
Support Vector Machines

- Non-linear classifier
- Decision boundary between parameters - hyperplane
- Kernal space separates data into positive and negative classes
- Binary class labels (deposit and nondeposit) are used to slot the tectonic parameters into
- Separate the classes with a hyperplane

Prediction Results: Partitioned Data

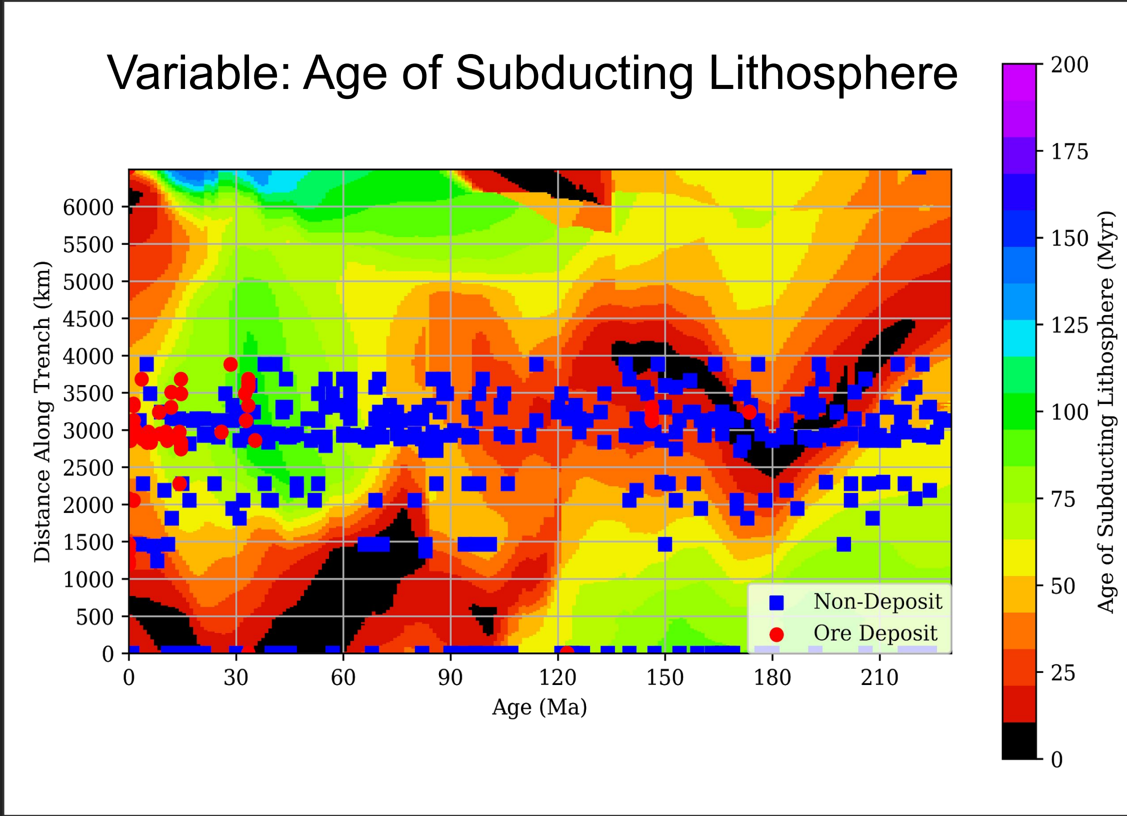


Prediction Results: Point Data



Colour Map Using Spatio-Temporal Bin Parameters

- Spatial domains = distance along trench
- Temporal domains = Age
- Increments set here for data separation



Discussion

- A majority of the deposits in the Bolivian Tin Belt are no older than 40 Ma, which would be in the oligocene and miocene time periods. Each deposits is also centrally located between 2000 km and 4000 km

- Formations of tin were young and closest to the present day trench, while only five deposits existed between the 100 Ma and 200 Ma time period.

Discussion

- The ore deposits which were aged between 0 Ma and 40 Ma were located within lithosphere aged no younger than 50 Ma and no older than 100 Ma. The younger deposits are associated with the lithospheric age associated with the oligocene time period.

- The collision resulted in a focused horizontal compressional stress that created large scale failure in the region.

- The thickening of the oceanic lithosphere as it gets older or further away from the spreading center and the relatively thick layer of sediments accumulated on the subducting plate may contribute more metals and volume to the melt generated by subduction of the oceanic lithosphere.

Summary

- Machine learning and computer science were able to help in the process of comprehending a hefty compilation of data
- Strong coupling between the tectonic events and ore deposits at the Nazca and South American Plate margin and the timestamp of the lithosphere and porphyry deposits support this analysis.
- Random forest and SVM of both the area and point related plots showed similar results overall, but had rated the parameters importance differently in regards to their effect on tectonic related ore deposit formation.
- It is clear from reviewed literature and this analysis that tin ore deposits are strongly related to tectonic events.

Literature cited

Craig, J. and Vaughan, D. (2019). *Cassiterite | Virtual Microscope*. [online] Virtualmicroscope.org. Available at: https://www.virtualmicroscope.org/content/cassiterite-1 [Accessed 29 May 2019].

Handbookofmineralogy.org. (2019). [online] Available at: http://www.handbookofmineralogy.org/pdfs/cassiterite.pdf [Accessed 30 May 2019].

Hu, F. (2003). Cyclic Variation and Solar Forcing of Holocene Climate in the Alaskan Subarctic. *Science*, 301(5641), pp.1890-1893.

Iskyan, K. (2016). *Why tin prices will continue to "solder" on - Stansberry Pacific*. [online] Stansberry Pacific. Available at: https://stansberrypacific.com/asia-wealth-investment-daily/why-tin-prices-will-continue-to-solder-on/ [Accessed 29 May 2019].

Jiménez-Franco, A., Alfonso, P., Canet, C. and Trujillo, J. (2018). Mineral chemistry of In-bearing minerals in the Santa Fe mining district, Bolivia. *Andean Geology*, 45(3), p.410.

TeePublic. (2019). *Tin Element by cerebrands*. [online] Available at: https://www.teepublic.com/t-shirt/1303040-tin-element [Accessed 29 May 2019].