Silicon and Oxygen in Earth’s Core: Applications of Machine Learning to Metal-Silicate Equilibria and Core Formation

Ruben Keane

Follow this and additional works at: https://cedar.wwu.edu/wwu_honors

Part of the Geology Commons, and the Geophysics and Seismology Commons

Recommended Citation
https://cedar.wwu.edu/wwu_honors/648

This Project is brought to you for free and open access by the WWU Graduate and Undergraduate Scholarship at Western CEDAR. It has been accepted for inclusion in WWU Honors College Senior Projects by an authorized administrator of Western CEDAR. For more information, please contact westerncedar@wwu.edu.
Silicon and Oxygen in Earth’s Core: Applications of Machine Learning to Metal-Silicate Equilibria and Core Formation

Ruben Keane\textsuperscript{a,b,c,} Advised by Asmaa Boujibar\textsuperscript{a,b,d}

March 2023

\textsuperscript{a} Geology Department, Western Washington University, 516 High St., Bellingham, Washington, 98225
\textsuperscript{b} Department of Physics and Astronomy, Western Washington University, 516 High St., Bellingham, Washington, 98225
\textsuperscript{c} Honors College, Western Washington University, 516 High St., Bellingham, Washington, 98225
\textsuperscript{d} Advanced Materials Science & Engineering Center, Western Washington University, 516 High St., Bellingham, Washington, 98225

Abstract

Within Earth’s core, light elements (Si, O, C, S, N, H) are known to make up a small fraction of the total mass of the core with respect to heavy elements. The degree to which these elements exist in the cores of terrestrial planets have geophysical and geochemical implications, most notably the presence of core convection and a geodynamo, thermal conductivity within the core, and core temperature. Comparison of the composition of chondrites to Earth’s mantle composition and the Preliminary Reference Earth Model have given an estimation of about 10% light elements in Earth’s core. The concentrations of each light element have been characterized in previous literature by determining experimentally the partitioning of elements between metal and silicate phases at high pressure and temperature. Studies have constructed thermodynamic models using linear regressions, to predict the change of partition coefficients with pressure, temperature, and oxygen fugacity. However, there is a large variance among previous literature in result thermodynamic models, which is likely indicative in substantial regression errors and evidence of non-linearity ascribed to elemental depletions in partitioning experiments. Here, we used the random forest regression, a machine learning algorithm, to predict the partition coefficients of Si and O using MetSilDB, a database for metal-silicate equilibria (Boujibar et al., GSA Fall 2022 Conference) and assessed the accuracy of our models using cross-validation techniques. Using these methods, we built a model predicting elemental partitioning coefficients with a highly improved performance than previous models, finding Earth’s core being enriched in O ($5.1 \pm 1.4$ wt\%) and depleted in Si ($1.2 \pm 0.3$ wt\%) aligning with seismic constraints. Our findings will help us infer the composition of the core more accurately.
1. Introduction

The initial accretionary stages of Earth 4.56 billion years ago (Halliday & Canup, 2022) are outlined by a period of high energy collisions of planetesimals and planetary embryos until 3.8 Ga (Tera et al. 1974). The energy delivered from impact events generated a transient magma ocean, extending to about half of the current mantle depth (Andrault et al., 2011, Fischer et al. 2015, Wood et al., 2006, Drake & Righter 2002). Core-mantle differentiation ceased 30-40 Myr after Earth’s formation where 98% of Earth’s mass was accreted within this period (Kleine et al. 2002, Yu & Jacobsen, 2011). Throughout Earth’s differentiation by gravitational segregation of dense metallic phases from the lighter silicates (Rubie et al., 2003, Wood et al., 2006, Halliday & Canup, 2022), chemical reactions between these two phases lead to a quasi-irreversible change of the Earth’s core composition. The abundances of iron-loving elements, called siderophile elements became depleted in the mantle. The affinity of the elements between metal and silicate varies with the pressure and temperature conditions at which these chemical reactions occur. Numerous studies have investigated these effects on the partitioning of elements between metal and silicate by conducting laboratory experiments at high pressure and temperature. The approximate depth of the magma ocean was derived from experimental data showing that Ni and Co become less siderophile at high pressure and temperature (50 GPa & 4000K) (Siebert et al., 2011, Fischer et al. 2015).

Since the Earth’s mantle show non-negligible contents of Ni and Co (McDonough & Sun, 1995), the magma ocean should have extended to large depths allowing their retention in the mantle.

The constraints on Earth’s density profile obtained from seismic wave velocities (PREM model) show that the Earth’s core comprises 90% heavy elements (Fe and Ni) and 10% light elements (a combination of Si, O, C, N and H) (Dziewonski & Anderson, 1981). The presence of light elements in the core is critical for the geologic evolution of the Earth. They allow a decrease of the melting temperature of the core, which delay its solidification (Uchida et al., 2001, Morard et al., 2014). They also lower down the core’s thermal conductivity, which favors heat retention in the Earth’s interior (Pommier et al., 2020). These factors enable mantle convection, the generation and maintenance of the geodynamo, and the advent of plate tectonics on Earth. It is therefore important to estimate the chemical composition of the core to understand the conditions favoring these global geologic processes. The concentrations of elements in the core can be determined by studying their partitioning between metal and silicate in a wide range of pressure and temperature and with the knowledge of the Earth’s mantle composition. Elemental distribution between metal and silicate is conditioned by a number of factors, including oxygen fugacity (fO2), pressure, temperature, and the presence of other light elements in the metal (Boujibar et al. 2019, Siebert et al. 2013). Several studies have previously used experimental data to calibrate thermodynamic models predicting this elemental distribution. However, a wide variance among findings in the literature, for example in the partition of light elements such as Si and O between mantle and core (Siebert et al., 2013, Fischer et al. 2015, Badro et al. 2015), indicates limited precision in these models.

In this study, we used a newly compiled database Met-Sil-DB consisting of 2155 elemental partitioning experiments with 388 attributes from 121 peer-reviewed articles from 1993 – 2015 (Boujibar et al., GSA Fall 2022 Conference). We review the elemental partitioning of Si and O using high-pressure and temperature metal-silicate equilibria experiments from Met-Sil-DB ranging from 1400 – 5700 K and 0.1 – 115 GPa. Our aim is to compare the methodology of previous literature (Boujibar et al., 2019) with a new approach using a machine learning algorithm, a random forest regression (RFR) from scikit-learn libraries, to infer how O and Si partition in the core. Our findings can assess the validity of current thermodynamic models given the importance of O and Si partitioning on equilibria conditions in the primitive mantle (Boujibar et al., 2019). Using a continuously accreting Earth model (Bouhifd & Jephcoat, 2011, Fischer et al. 2015, Siebert et al. 2012, Andrault et al.
2. Methodology

2.1 Met-Sil-DB

A total of 625 and 1041 metal-silicate equilibria partitioning experiments were obtained from 53 and 82 different studies, respectively, from Met-Sil-DB database to evaluate the distribution of O and Si in the core. Data were derived from experiments utilizing several types of apparatuses including, laser-heated diamond anvil cells (LHDAC), multi-anvil press (MA), piston cylinder (PC), internally heated pressure vessel (IHPV), vertical furnace, and the 1 atm furnace. Analytical techniques, documented in Met-Sil-DB, were similarly varied and included electron microprobe analyzer (EMPA), laser ablation inductively coupled plasma mass spectrometry (LA-ICPMS), secondary ion mass spectrometer (SIMS), Fourier transform infrared spectroscopy (FTIR) and Raman spectroscopy. Starting samples were either natural or synthetic samples of varying compositions from chondritic to mixtures of metal alloys with pyrolytic, basaltic or other kinds of powders or glasses. These experiments encompass the broadest range of experimental pressure and temperature conditions; those selected for examining Si and O partitioning reported temperature and pressure ranges of 1473 – 5700 K and 0.1 – 111 GPa. To model the partitioning of Si and O, for each experiment, we selected data on pressure, temperature, concentrations of sulfur and carbon. For the experiments where the samples were included in a graphite capsules, carbon is known to systematically contaminate the metal phase (e.g. Boujibar et al. 2019). Therefore, we calculated metallic carbon concentrations for these experiments where its abundance was not directly measured by subtracting the total of all elemental abundances to 100%. For all experiments, we considered the elemental concentrations of Si and O in both metal and silicate phases of experimental samples and calculated the fO2 relative to the Iron-Wustite buffer following previous studies (Boujibar et al. 2019).

2.2 Traditional Linear Modeling of Elemental Partitioning

To compare the statistical validity of our approach versus previous literature, we applied similar thermodynamic models to quantify partitioning coefficients as outlined by previous studies (Boujibar et al., Siebert et al., 2013). The distribution of an element M between metal and silicate can be described by the redox reaction:

\[ \text{M}^{\text{metal}} + \text{O}_2 = \text{MO}_{n/2}^{\text{silicate}} \] (1)

Where the element M in the metal can be oxidized to form the cation \( \text{MO}_{n/2}^{\text{silicate}} \) in the silicate and vice versa. For a chemical reaction in equilibrium, an equilibrium constant \( K \) can be derived from the activities of the reactants and products \( a \), measuring the deviation from ideality. An activity \( a \) is the product of the concentration of the compound and the activity coefficient. \( fO_2 \) is the activity of O2 in the system. Taking Si as element M, the equilibrium constant of reaction (1), \( K_{eq}(1) \) is related to the change in Gibbs free energy of the system, the ideal gas constant \( R \), and the system temperature \( T \) with:

\[
\ln(K_{eq}(1)) = \ln\left(\frac{\text{d}^{\text{metal}} \times fO_2}{\text{d}^{\text{silicate}} \times fO_2} \right) = \frac{-\Delta G}{RT} = -\frac{(\Delta H - T \Delta S + \int \Delta V dP)}{RT}
\]

(2)

Where \( \Delta H \) is the enthalpy, \( \Delta S \) is the entropy, \( \Delta V \) is the volumetric change, \( dP \) is the differential change in pressure, \( T \) is the temperature, and \( R \) is the ideal gas constant. Instead of calculating the equilibrium constant using all these thermodynamic parameters which are unknown at the extreme pressures and temperatures of core formation, it is usual to utilize experimental data to predict the partition coefficients using similar types of mathematical relationship to pressure, temperature and chemical composition:

\[
\log(D_{Si}) = a + \frac{b}{T} + \frac{cP}{T} + d \times \log(fO_2) + e \times (1 - X_{\text{metal}}) + f \times (1 - X_{\text{metal}}^S) + g \times (1 - X_{\text{metal}}^C)
\]

(3)

This equation was derived by manipulating equation (2) which describes chemical equilibrium for reaction (1). The usual reaction controlling the distribution of O between core and mantle is:

\[
\text{FeO}_{\text{silicate}} = \text{Fe}_{\text{metal}} + \text{O}_{\text{metal}}
\]

(4)

Since this is not a redox reaction, the concentration of O in the metal itself is considered with similar parameters to equation (3), with the replacement of the oxygen fugacity by the FeO concentration in the silicate:

\[
\log(O_{\text{metal}}) = a + \frac{b}{T} + \frac{cP}{T} + d \times \log(fO_2) + e \times (1 - X_{\text{metal}}^S) + f \times (1 - X_{\text{metal}}^S) + g \times (1 - X_{\text{metal}}^C)
\]

(5)

Using multiple experimental data from varying P, T and chemical compositions, a linear regression can be conducted to derive the parameters a-g, by selecting results where predicted \( D_{Si} \) or \( O_{metal} \) are the closest to the observed values.
2.3 Machine Learning Applications

Here, we considered an additional approach to constrain the abundances of Si and O in the core. We used an ensemble machine learning technique called random forest regression (RFR). In particular, we used an RFR algorithm from the scikit-learn libraries in python. This approach allows to model the partitioning without making assumptions on the chemical exchange occurring between metal-silicate phases during core formation. For example, in linear regressions, the partitioning is assumed to be linearly dependent on the inverse of the temperature (see equations 3 & 5). Machine learning enables to predict the partition coefficients without fixing these dependencies to the considered variables. Independent variables (experimental conditions) and dependent variables (\(D_{Si}\) and \(O_{metal}\) wt\%) from Met-Sil-DB are separated into two separate data frames so the algorithms can compare the desired output with respect to the inputted variables. We additionally performed a cross validation, which is a resampling technique enabling to statistically improve results. It consists of randomly splitting the data into training and testing groups, where 90% of experiments drawn from Met-Sil-DB are used to train the model, and 10% of experiments are subsequently used to test the model.

The RFR generates its model predicting partitioning behavior by generating a series of decision trees (figure 2). We begin the model by separating experiments drawn from Met-Sil-DB into two data frames comprised of the equilibria conditions and the measured partition coefficients. The partition coefficients were calculated from analytical measurements of elemental concentrations by weight percentage in silicate and metal phases in literature. Secondly, we segment data into a training set used to train the model comprised of 90% of experiments and the testing set using 10% of experiments. Once data preparation is complete, the model learns by creating a series of decision trees, each tree is comprised of a randomly assigned portion of the training data. Within each tree itself, the model segments data based on a numerical qualifier, mapping input variables to the desired numerical output. In this context, a tree is comprised of a portion of the experimental data from the training set. Branches in the tree are “junctures” in which a specific experimental condition from the training set is separated, for example, in figure 2 the first branch in the decision tree separates metal-silicate partitioning experiments based on \(fO_2\) values being greater than or equal to the IW-1 value.

The model will learn by referencing analytically derived partition coefficients with the experimental conditions to learn to predict partitioning behavior as it recursively steps through the branching process using data mapping of the inputted variables to the desired numerical result. At the bottom of each branch (figure 2), based on the numerical branching the model chose, it will generate a final predicted partition coefficient and then will assess its own accuracy based on a statistical qualifier. The model will learn the relationship of equilibria conditions with respect to a measured partition coefficient over the set of decision trees comprising the model. In the training process, we gave the model the option to use 800, 900, or 1000 decision trees to adequately train itself. Once the data mapping is complete over the array of decision trees, the model will average the accuracy of all the data mapping within the decision trees themselves to create the most effective set of numerical qualifying questions to segment the equilibria conditions in the order given by an importance percent (table 1), this process is called final class voting. We can then assess the ability of the model to predict partitioning behavior by applying the trained model to the testing set, given that the model wasn’t trained on the testing set.

Each machine learning algorithm must be cross-validated to further improve upon the ability of the models to produce predicted outputs. The use of cross-validation in the application of machine learning algorithms to datasets is a common methodology to minimize validation loss or to minimize how far the output partition coefficients are from the measured value. In this study, we used k-fold cross-validation to cross-validate the models. K-fold cross-validation is the technique of assigning data into k amounts of folds, such that each fold contains the total dataset. Each k-fold will contain a different randomly assigned test-train split where each k-fold will train its own set of decision trees. However, each k-fold will test its trained model on the testing set from the other folds. Using the output root-mean-square error (RMSE) from comparing measured to predicted partition coefficients, the model generating
the smallest RMSE when comparing predicted versus measured coefficients is selected.

3. Results

Once the most effective algorithms are selected, we then apply predicted partitioning behavior on an accreting Earth model outlining metal-silicate equilibria conditions at the core-mantle boundary (CMB) along increments of a fraction of Earth accreted. These conditions include pressure at the CMB (Bouhifd and Jephcoat, 2011, Fischer et al. 2015, Siebert et al. 2012), the liquidus temperature of a bulk chondritic mantle at CMB depth (Andrault et al., 2011), FeO by weight percent in the mantle (McDounough & Sun, 1995), and light element factors (Wood et al., 2013, Palme & O’Neill, 2003). Using the accreting Earth model, we can quantify the amount of Si and O in the core by weight percent derived using both linear methods and machine learning algorithms to compare each approach.

3.1 Modeling Partitioning Behavior

Our findings suggest that the random forest regression is more effective at predicting partitioning behavior than linear regressions derived from equations 3 and 5. With respect to both Si and O metal-silicate partitioning experiments from Met-Sil-DB, our model decreased the RMSE when comparing analytically measured and predicted partition coefficients with respect to the linear model. What’s more, our model was able to predict partitioning behavior with similar accuracy in both Si and

Fig. 3. Two sets of graphs depicting measured versus predicted partition coefficients for Si and elemental concentrations for Ometal comparing linear regressions used in previous literature and the random forest regression with studies listed. Within each figure, the black regression line represents the “ideal” fit \( y=x \), such the predicted partition coefficients matched the analytically observed coefficient perfectly. (a) These figures compare the ability for linear regressions and the RFR to predict partitioning in Si, where linear methods produce an RMSE of 0.6621 and the RFR produces an RMSE of .11 among measured vs. predicted partition coefficients. (b) These figures compare the ability of linear regressions and the RFR to predict partitioning in O, where linear methods produce an RMSE of 0.38 and the RFR produces an RMSE of .10 among measured vs. predicted partitioning in Ometal concentrations. For graphs depicting the results of the RFR, the number of best estimators details how many decision trees were used. For figures a and b using experimental data from Met-Sil-DB, the shorthand reference format from the database was color coded on the figures accordingly.
O metal-silicate equilibria. As compared to the linear model, the RFR was able to produce substantially more accurate predicting behavior as it relates to the partitioning of Si between metal-silicate phases. We found a RMSE from the RFR model of predicted versus observed partition coefficients of 0.11 and 0.12, for Si and O, respectively. In contrast, linear regressions yielded an RMSE of 0.66 and 0.37, for Si and O, respectively (figure 3). What is notable specifically in these findings is the increased accuracy by which we can predict the partitioning behavior of Si and O between phases of equilibria.

### 3.2 Importance by Variable

We assessed the importance of each equilibria condition in experimental data by comparing the p-values and the importance percentages by variable outputted using linear methods and the RFR, respectively. Table 1 shows output importance percentages from the RFR, the model derives importance percentages of input variables from data mapping in the ensemble learning process during the training phase. With respect to Si partitioning behavior, our model concludes the most important factors controlling the chemical equilibrium are both the oxidation state and the pressure conditions that the phases are subjected to. The next significant parameters for Si partition are C concentration in the metal and temperature. Conversely, O partitioning behavior is more dependent on the intensive quantities of the system: pressure, followed by temperature then the FeO concentration of the silicate.

Looking at outputted p-values of regression coefficients from table 1, findings from the importance of equilibria conditions to resulting Si partitioning from the linear model show a paramount dependence on the oxidation state, similarly to RFR results. However, temperature and the oxygen concentration in the metal have similar statistical dependence being second to the previous variable. Lastly, the linear method shows the least significant factors are S and C abundances and the pressure conditions of equilibria. Using the linear approach, O partitioning will be most influenced by the pressure conditions, the S concentration, and to a lesser extent the temperature. Furthermore, the linear model has inferred less statistical importance given to the Si and C content of the metal and the FeO abundance in the silicate. Qualifying the importance by variables in the linear findings is also important. While there can be importance by variables inferred based on the p-value itself, each variable denotes a noteworthy influence on the model itself given each value is within the threshold to indicate a statistical dependence.
3.3 Continuous Core Accretion Model

Using a continuously accreting Earth model we outlined equilibria conditions, taken from literature, at 100 steps of the fraction of Earth accreted. We consider a magma ocean extending to half the depth of the current mantle (Bouhifd & Jephcoat, 2011, Fischer et al. 2015, Siebert et al. 2012), a mantle composition being primarily pyrolytic (McDonough & Sun 1995), a liquidus temperature of a chondritic mantle calculated by the pressure environment at the 50% CMB (Andrault et al., 2011), and an oxygen fugacity calculated with respect to the IW-buffer and mantle concentrations of FeO silicate. These conditions can be described using the nomenclature P-T-XFeO-fO2. Using these outlined conditions evolving over an accreting Earth model, we calculate inferred partition coefficient of Si and O elemental concentrations between the novel mantle and the core. Between the two models, the linear regression and RFR, findings for partitioned Si and O in the core do not differ between the two models, both showing a core enriched in O (5.1-5.2%) and depleted in Si (1.3-1.4%) (figure 4).

However, the RFR allowed for a significant reduction in error associated when integrating over differential steps of the fraction of Earth accreted. For Si in the core, the RFR improved upon the accuracy of the linear model decreasing integration error from ±1.9 wt% to ±0.3 wt%. Similarly, the RFR improved on the accuracy of the linear model decreasing integration error calculating O in the core from ±4.2 wt% to ±1.4 wt%.

4. Discussion and Implications of Core Formation

Using database improvements for experiments outlining metal-silicate equilibria from Met-Sil-DB in application to both a less accurate linear model taken from previous literature and the RFR, we’ve inferred a core depleted in Si and enriched in O. This inferred distribution of given light elements in the core aligns with a previous study constraining P-T-XFeO-fO2 paths with seismic considerations (Badro et al. 2015), outlining nec-
essary density constraints in the outer core from the AK135 radial seismological model (Kennett et al., 1995). These studies produce density constraints aligned with distributions of O and Si to what we have inferred in this study. Our findings thus agree with P-T-XFeO-fO2 paths outlining a hot liquidus with pressure and temperature conditions of 55-60 GPa and 4100 – 4300 K (Badro et al., 2015), respectively. Importantly, these paths produce inferred current mantle elemental concentrations (McDonough & Sun 1995). Our findings further align with core formation conditions defined by early oxidizing conditions, such that core formation is dominated by a high oxygen content followed by reduced conditions to reach the current mantle concentrations of 8% (Siebert et al., 2013). Accordingly, our findings do not align with the literature suggesting a core enriched in Si and depleted in O (Fischer et al., 2015), suggesting equilibria conditions with pressure ranges similar to our findings, but a significantly cooler temperature range of 3300-3400 K.

While our model is successful in predicting partitioning behavior given high pressure and temperature experimental conditions in metal-silicate equilibria, our inferences based on the evolving equilibria conditions within the accreting Earth and derived elemental concentrations in the core do have notable limitations. Our findings show the improved methodology of modeling partitioning behavior between metal-silicate equilibria using the RFR as opposed to linear regression. Given that the linear method is derived from geochemical assumptions based on the exchange between metal diapirs with the novel mantle, it is likely that these assumptions introduce falsehood to our thermodynamic findings. As such, the findings from our accreting Earth model do provide insights into aligning findings from previous literature on evolving thermodynamic constraints of the Earth as it gains mass. However, these findings are supported by inferences made on namely the liquidus temperature of the chondritic mantle, the pressure at the current 50% depth of the CMB, and the distribution of oxygen in FeO silicates throughout the novel mantle. Our model is thus limited to the derivation of these system constraints being based on interpretations from modeling these behaviors using linear regressions as outlined in the studies above. With this, there are some restrictions in our findings on the elemental concentrations of Si and O in the core, in addition to the equilibria conditions they are most reliant on.

5. Conclusion
In this study, we have outlined Earth’s core being enriched in O and depleted in Si using a novel approach in the application of the RFR to experimental data in Met-Sil-DB. Additionally, our findings align with seismic constraints of density in the outer core, in addition to system conditions pertaining to equilibria defined in previous literature aggregating with a higher concentration of O in Earth’s core. Future studies should compound upon the accumulation of metal-silicate equilibria experiments in Met-Sil-DB, in addition to using the methodology outlined above to apply machine learning algorithms to model partitioning behavior in metal-silicate equilibria. Experimental data in Met-Sil-DB should be assessed to determine if samples reached equilibria through the comparison of sample size given the experimental apparatus and the ideal diffusion times of Si and O at lower to medium experimental temperatures (Bouhifd et al., 2011, Hofmann, 1980). The certainty to which our findings rest is reliant on assessing findings of the oxidation state of the novel mantle, in addition to the equilibria conditions inferred during core formation based on core-depletions of Ni and Co (Bouhifd et al., 2011) and core-mantle distribution of Cr (Fischer et al., 2015, Andrault et al., 2013) using our methodology. The assessment of these findings can help future studies effectively constrain the depth of the magma ocean in addition to the general redox condition of the early Earth using the partitioning behavior of Ni and Co, and Cr.

6. References


