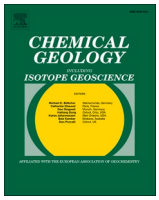
Chemical Geology 693 (2025) 122986



Contents lists available at ScienceDirect

Chemical Geology

journal homepage: www.elsevier.com/locate/chemgeo



Research Paper

Compositional evolution of the Archean subaerial crust revisited using machine learning constraints

Junteng Lyu *, Ziyi Guo, Ming Tang

SKLab-DeepMinE, MOEKLab-OBCE, School of Earth and Space Sciences, Peking University, China

ARTICLE INFO

Editor: S Aulbach

Keywords:
Archean
Continental crust
Continental composition
Machine learning

ABSTRACT

The composition of the Archean subaerial crust, whether mafic or felsic, remains debated due to tectonic uncertainties and proxy limitations. Here, we employ Random Forest Regression algorithm-based machine learning model to analyze multi-element geochemical data from terrigenous fine-grained sedimentary rocks to reconstruct the MgO content in the subaerial crust. The model is trained using major oxides (TiO_2 and Al_2O_3) and trace elements (Sc, Cr, Co, Ni, Cu, Zn, Y, Zr, Nb and REEs) of simulated mixture samples from Archean craton igneous rocks and post-Archean igneous rocks. The validity of this model is demonstrated by using various modern samples to predict MgO content in the modern subaerial crust, whose MgO content has been well constrained. We also demonstrate that the accuracy of the predictions is not sensitive to tectonic settings. We then use this model to reconstruct the chemical evolution of the subaerial crust over Earth's history and compare the differences between our model and other proxies reflecting the composition of the subaerial crust. Our model results suggest that a systematic decline of MgO content in the subaerial crust (7–10 wt% to 3–4 wt%) and a shift from mafic (70–100 % komatiite-basalt) to felsic dominance (~70–90 % TTG) of the subaerial crust in the Neoarchean period, after which the subaerial crust attained a similar composition to that of the modern subaerial crust. The results from our approach are similar to those from Ni/Co and Cr/Zn proxies, but differ from the results of Zr/TiO_2 and Al_2O_3/TiO_2 proxies. We suggest that many previously used composition proxies are sensitive to tectonic settings and applying these proxies to terrigenous sediments can lead to erroneous results on the subaerial crust composition in the Archean. In addition, we caution against the use of proxies whose correlations with MgO or SiO_2 content are highly nonlinear, which may result in large uncertainties in the projected MgO or SiO_2 content for the source rocks.

1. Introduction

The chemical composition and lithological characteristics of the Earth's crust are fundamental for understanding mechanisms of continental formation (Rudnick, 1995), the onset of plate tectonics (Dhuime et al., 2015; Greber et al., 2017; Tang et al., 2016), and the chemical evolution of Earth's surface systems, including the atmosphere (Lee et al., 2016; Smit and Mezger, 2017) and oceans (Konhauser et al., 2009). One unique feature of Earth's crust is its subaerial portion—the part of the continental crust that lies above sea level and is directly exposed to weathering and erosion (hereafter termed "subaerial crust"). At present, Earth's subaerial crust is largely represented by the thick continental crust, which has a felsic composition that contrasts with Earth's oceanic crust and the crusts on other terrestrial planets (Rudnick and Gao, 2014; Taylor and McLennan, 2009).

Although constraining SiO_2 and MgO contents is essential for understanding the composition of Archean subaerial crust, doing so remains challenging. Large-scale surface sampling is subject to sampling bias, a problem that becomes more pronounced with increasing sample age, as erosion over time preferentially removes subaerial crust materials (Cawood et al., 2022; Kamber and Ossa Ossa, 2025; Tang, 2021). Terrigenous fine-grained sedimentary rocks have also been used to constrain the average composition of large tracts of subaerial crust, but direct measurements of MgO in terrigenous fine-grained sedimentary rocks fail to accurately reflect the MgO content of the source rocks because Mg is highly mobile. As a result of its high solubility, Mg is readily leached during chemical weathering and transported to the oceans (Holland, 2020; Kamber and Ossa Ossa, 2025; Rudnick and Gao, 2014), and it is also prone to redistribution during diagenesis (Berg et al., 2019; Kamber and Ossa Ossa, 2025). Similarly, SiO_2 content in

* Corresponding author.
E-mail address: jtlv@stu.pku.edu.cn (J. Lyu).

https://doi.org/10.1016/j.chemgeo.2025.122986
Received 7 April 2025; Received in revised form 21 July 2025; Accepted 25 July 2025
Available online 26 July 2025
0009-2541/© 2025 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

terrigenous fine-grained sedimentary rocks does not provide reliable estimates of the source rock compositions due to grain-size dependent sorting during transportation and diagenetic alteration (Greber et al., 2017; Greber and Dauphas, 2019).

To overcome the challenges mentioned above, one widely used approach involves establishing empirical relationships between geochemical proxies formulated using insoluble elements and MgO or SiO_2 contents in igneous rocks, then applying these relationships to fine-grained sedimentary rocks to infer source compositions (e.g. Chen et al., 2020; Gaschnig et al., 2014; Greber et al., 2017; Greber and Dauphas, 2019; Tang et al., 2016). However different proxies often yield conflicting results. For example, the Ni/Co, Cr/Zn ratios (Tang et al., 2016) in shales and diamictites indicate that the subaerial crust was primarily composed of mafic rocks through much of the Archean and transitioned into a dominantly felsic composition near the end of the Archean. This conclusion was later supported by Cr/U ratio in shales and diamictites (Smit and Mezger, 2017) and Cu content in diamictites (Chen et al., 2020) but studies based on Sm/Nd (Garçon, 2021), Al_2O_3/TiO_2 and Zr/TiO_2 ratios (Greber and Dauphas, 2019) and Ti isotopes (Greber et al., 2017) have reached a different conclusion that the Paleoarchean subaerial crust was dominantly felsic.

We note that one difficulty with previous studies is that they all relied on proxies calibrated based on very few elements, which may not accurately capture the compositional variation of the source rocks. To overcome this difficulty, we develop a multi-proxy predictive model based on the Random Forest Regression algorithm (RFR), a machine learning tool (Breiman, 2001), to predict the chemical composition of the subaerial crust. Our RFR model incorporates multiple elements to constrain the composition of subaerial crust and highlights the most important elements in the prediction. To validate its reliability, we test our RFR model sensitivity to different tectonic setting (i.e., convergent margin and intraplate volcanics). We also compare the predicted subaerial crust MgO content with existing reference values. Finally, using this RFR model, we quantify the subaerial crust MgO content throughout geological history.

2. Method and data compilation

2.1. Random forest regression

Random Forest Regression (RFR) is a robust ensemble learning algorithm widely used for regression tasks due to its ability to handle nonlinear relationships and reduce overfitting. The method constructs many decision trees during training and combines their predictions to improve accuracy. Unlike single decision trees, Random Forest utilizes randomness in both data selection and feature selection to enhance model robustness and generalization (Breiman, 2001). The dataset is randomly sampled to create multiple subsets, each used to train an individual decision tree (Tree 1, Tree 2, …, Tree N). Each tree makes an independent prediction based on its training data, denoted as Prediction 1, Prediction 2, …, Prediction N. The final prediction of the RFR model is computed as the average of these individual predictions. This ensemble approach reduces the variance and increases the model's predictive power by leveraging the combined knowledge of all trees. The MATLAB codes used in this study are available on the authors' GitHub page (https://github.com/JuntengLyu).

2.2. Feature element selection

Regarding element selection criteria for reconstructing the subaerial crust composition through fine-grained terrigenous sedimentary rocks, previous studies have proposed the following standards (Chen et al., 2020; Greber and Dauphas, 2019; Ptáček et al., 2020; Tang et al., 2016): (1) Fluid-immobile elements that remain stable during water-rock interactions without being mobilized into aqueous solutions through chemical weathering, and (2) Biologically inert elements unaffected by biogeochemical processes. Additionally, selected features must have sufficient data in the training set (>1000 samples) to ensure RFR model robustness capability. Ultimately, we selected the following elements as optimal RFR model features for their correlation with MgO: TiO_2, Al_2O_3, Zr, Ni, Co, Cr, Zn, Cu, Y, Nb, Sc, REEs. This selection integrates previous studies that identified elemental proxies in terrigenous fine-grained sedimentary rocks to evaluate the composition of the subaerial crust.

2.3. Data compilation, model building and evaluation

The source rocks of terrigenous sedimentary rocks are typically of mixed origins. Therefore, before constructing the RFR model, we first mix the compiled igneous rock data. This dataset includes Archean craton igneous rocks (n = 1693) sourced from the GEOROC database and post-Archean igneous rocks (n = 4426) sourced from Keller et al. (2015) (Table S1). Each simulated mixture sample consists of a random combination of 2 to 100 individual samples, with the mixing weights of each sample also assigned randomly (e.g. Tang et al., 2016). A total of 10,000 simulated mixtures are created for both Archean and post-Archean samples. The simulated mixtures are randomly divided into two parts: one subset (70 % of the Archean and post-Archean samples) for modeling (Fig. 1a) and the other (the remaining 30 % of the samples) for testing (Fig. 1b). To build the RFR model, two parameters need to be set: the number of variables per tree (Number of variables) and the number of decision trees (Number of trees). To select the optimal parameters, we repeated the model construction process 10 times and recorded the impact of different numbers of variables (Fig. 1c) and different numbers of decision trees (Fig. 1d) on the prediction performance for the test set. The small variation between models from each training result indicates that our model demonstrates high robustness. Finally, we selected the optimal parameters (Number of variables = 8, Number of trees = 600) to complete the model construction.

Because magmatic differentiation follows different trends in different tectonic settings, particularly between convergent margin and intraplate settings, synthetic mixtures were generated separately for these endmember tectonic settings (Table S3). Convergent margin igneous rocks (n = 2617, data are from Chen et al. (2023) were randomly combined (2–100 samples per mixture, random weights) to produce 10,000 synthetic mixtures representing calc-alkaline magmatism. Similarly, intraplate volcanic samples (n = 16,158, data are from GEOROC) were mixed independently to generate 10,000 tholeiitic-dominated mixtures. This approach ensures that the model accounts for tectonic-specific geochemical variability, critical for interpreting sedimentary proxies derived from heterogeneous crustal sources.

We compiled the chemical composition of 1861 samples of shales (including pelites, graywackes and other fine-grained terrigenous sedimentary rocks) and glacial diamictites from 63 publications (Table S2). Due to the impact of missing feature elements on the accuracy of model predictions in RFR (Breiman, 2001), we only select samples with complete feature elements from the compiled data for prediction.

The performance of the RFR model is evaluated using two metrics: the coefficient of determination (R^2) and the root-mean-square error (RMSE). R^2 indicates the goodness of fit, measuring how effectively the model predicts unknown samples. The formula for R^2 is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (Y_i - y_i)^2}{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}$$

Where Y_i and y_i are the true and predicted values (MgO content, wt %) respectively, $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ is the average true MgO content of the samples and n is the number of samples. The best possible score for R^2 is 1.

The root-mean-square error (RMSE) represents the standard deviation of the prediction errors, providing a measure of how accurately the

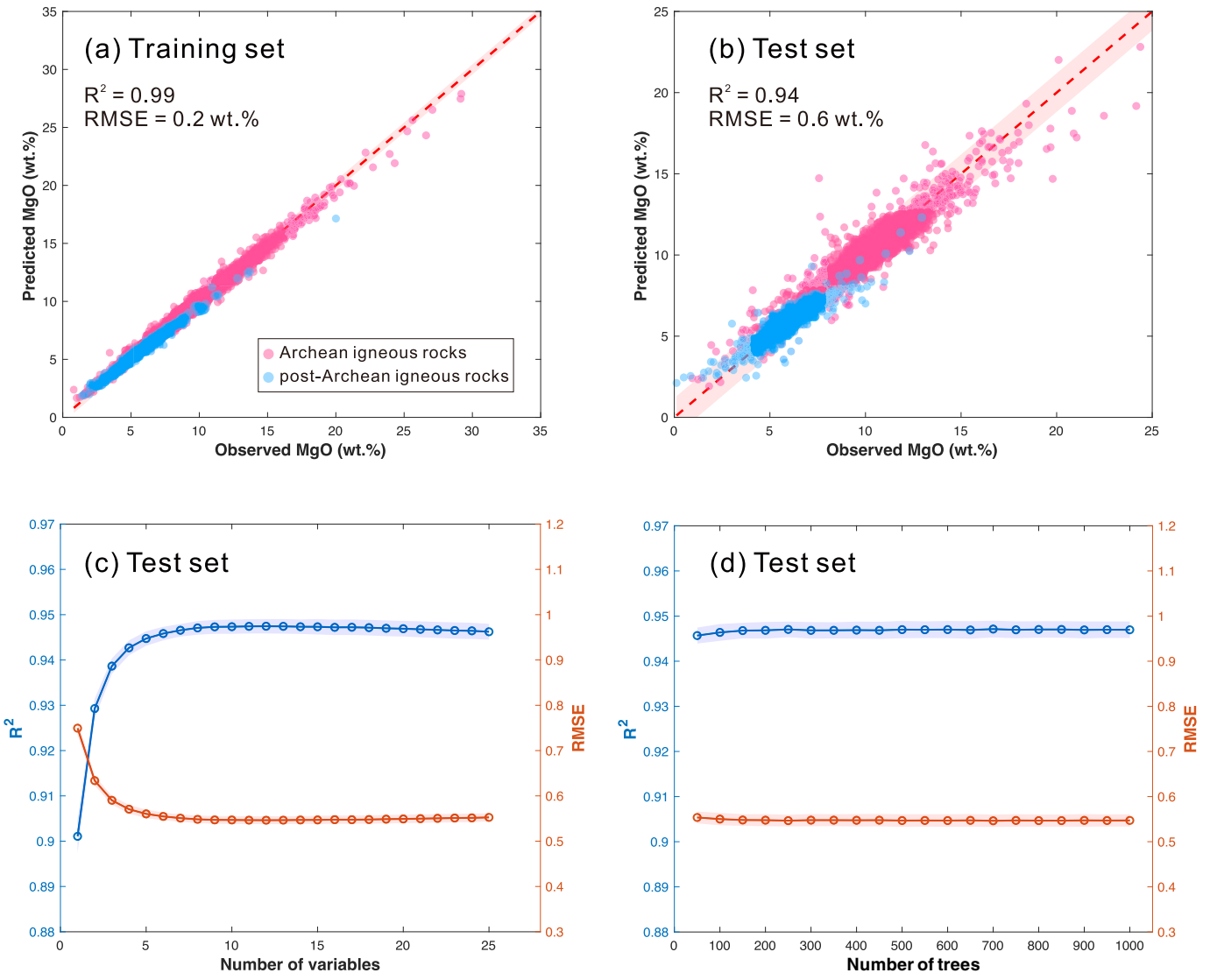


Fig. 1. Comparison of MgO content predicted by our RFR model with observed MgO content of igneous rocks in (a) training data set and (b) test data set. (c) Determination of the parameters (number of variables) for RFR model. (d) Determination of the parameters (number of trees) for RFR model. The shaded region represents 2 SE.

model predicts the target values. The closer the RMSE is to 0, the more accurate the model's predictions are. It is defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - y_i)^2}{n}}$$

3. Results

3.1. Important geochemical features for predicting the MgO content in the subaerial crust

The RFR model is interpretable, enabling the determination of feature importance during the computation, specifically identifying the contribution of each element in predicting MgO content in the subaerial crust. In this study, the feature importance was calculated using the Out-Of-Bag Permutation method in a Random Forest model (Breiman, 2001). This approach evaluates the contribution of each feature to the model's prediction accuracy by permuting the values of each feature and measuring the increase in the Out-Of-Bag error. We repeated the model construction process 10 times and recorded the values of feature importance. As shown in Fig. 2, Sc, Ni, Al_2O_3, TiO_2, Y, Cr, Co and Zr are the most significant features (above the average feature importance) for predicting the MgO content in the subaerial crust using the RFR model.

3.2. Performance of the RFR model

The RMSE and the R^2 score were used to assess the accuracy of the predicted dataset in comparison to the observed data. The performance of our final trained RFR model for predicting MgO content in the source rocks is shown in Fig. 1. In the training set, the R^2 is 0.99, and the RMSE is 0.2 wt%; in the test set, the R^2 is 0.94, and the RMSE is 0.6 wt% (Fig. 1a and b). These results indicate that the trained RFR model exhibits strong predictive capability.

We then tested the performance of the trained RFR model under convergent margin and intraplate settings. Synthetic mixtures were generated exclusively within the two endmember tectonic groups (Fig. 3). In the convergent margin setting, the RMSE is 0.3 wt%, while in the intraplate environment, the RMSE is 0.6 wt%. The model exhibits better accuracy in the convergent margin setting. But overall, the model demonstrates strong predictive accuracy in both settings, with a RMSE of less than 0.7 wt%. This indicates that the model reliably estimates the

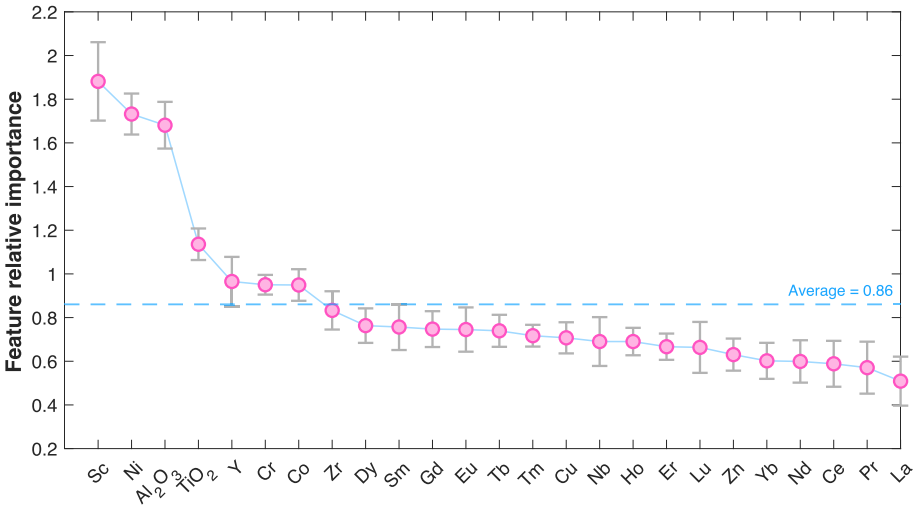


Fig. 2. The relative importance of each element during the RFR model training. Thin horizontal line represents the average value (~0.86). Error bars are 2 SE.

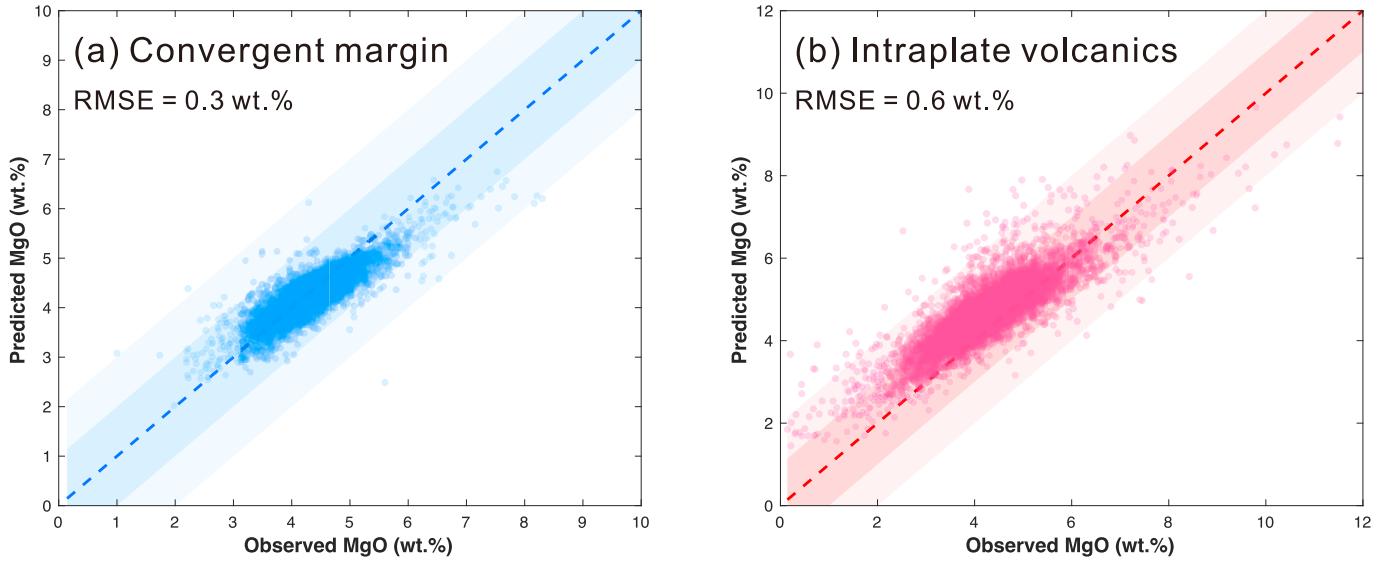


Fig. 3. Model testing across tectonic endmembers: (a) Convergent margin (calc-alkaline, data are from Chen et al. (2023) and (b) intraplate volcanic (tholeiitic, data are from GEOROC) settings. These data are provided in the Table S3. The dashed line represents Predicted MgO content = Observed MgO content. The darker and lighter shaded areas represent Predicted MgO = Observed MgO +/− 1 wt% and +/− 2 wt%, respectively.

source rock MgO content, regardless of the tectonic settings.

3.3. Predicted MgO content in the subaerial crust

After training and testing the RFR model using a global dataset of Archean and post-Archean igneous rocks, we further compared the predicted modern subaerial crust MgO content with existing estimates. Fig. 4 compares the predicted MgO content for various samples representative of the modern subaerial crust (e.g., loess, large-scale surface sampling, and river sediments) and the reference value. The consistent MgO content predicted by our model for the subaerial crust and the reference value (2 to 3 wt%, Rudnick and Gao, 2014) demonstrates that our model can accurately predict the composition of the subaerial crust.

Finally, we applied the RFR model to estimate the MgO content of the subaerial crust throughout geological history, based on the geochemical compositions of continentally derived fine-grained sedimentary rocks, such as shales and glacial diamictites (Fig. 5). The results show that before the Mesoarchean, the predicted MgO content of the subaerial crust based on glacial diamictites was approximately 7 wt%, while the prediction based on shale is around 10 wt%. When approaching the Archean-Proterozoic boundary (~2.5 Ga), the predicted MgO content of the subaerial crust, based on shales, had decreased to 3–4 wt% and remained relatively stable until the present, close to the current subaerial crust value of approximately 2–3 wt% (Rudnick and Gao, 2014). Similarly, the MgO content predicted from glacial diamictites also decreased to ~4 wt% around the same time and remained stable, consistent with the modern subaerial crust composition.

4. Discussion

4.1. Petrological interpretation of the model

As shown in Fig. 2, our RFR model identifies Sc, Ni, Al_2O_3, TiO_2, Y, Cr, Co, and Zr as important features, each having a relative importance greater than the model-wide average value for predicting MgO content in the subaerial crust. This can be readily explained by magmatic differentiation. In magmatic systems, typical compatible elements such as Sc, Ni, and Cr preferentially partition into early-crystallizing mafic

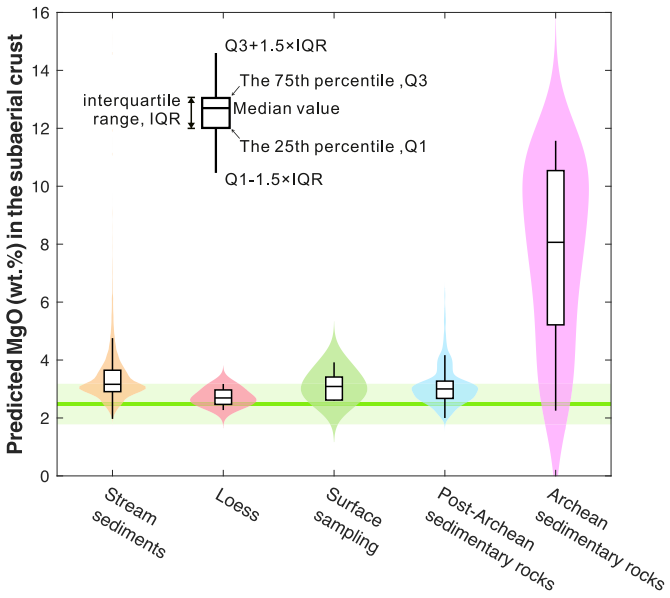


Fig. 4. Comparison between the predicted MgO contents of the subaerial crust based on the terrigenous sediments/rocks. The green bars denote the reference values for present-day subaerial crust (Rudnick and Gao, 2014). Surface sampling data are from Gao et al. (1998); Loess data are from Tang et al. (2015), and river sediments data are from Wang et al. (2024). The shaded region green represents 2 SE. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

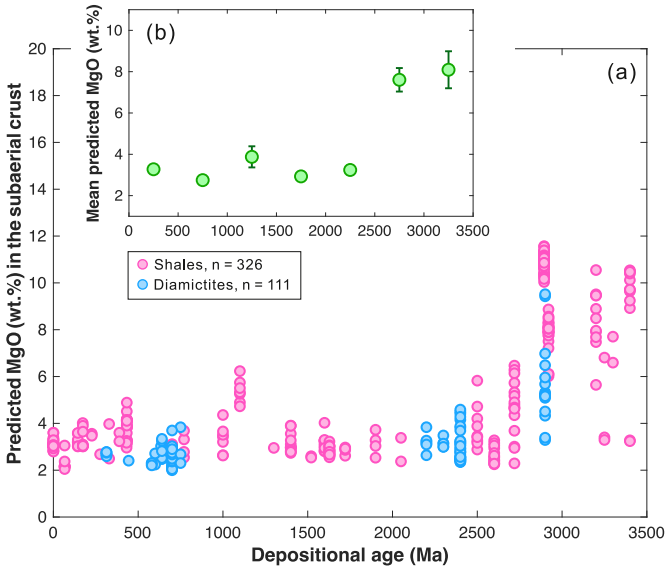


Fig. 5. (a) Predicted MgO content in the subaerial crust by terrigenous fine-grained sedimentary rocks through time. Shale and glacial diamictite data are provided in Table S2. (b) Age-binned Predicted MgO content in the subaerial crust combining shale and glacial diamictite data (bin size = 0.5 Ga). Error bars are 2 SE.

minerals, such as olivine, pyroxene, and spinel (Allègre et al., 1977; Bougault and Hekinian, 1974; Horn et al., 1994; Leeman and Scheidegger, 1977). These minerals crystallize abundantly during the early stages of magmatic evolution, leading to the enrichment of these elements in mafic rocks such as basalts and komatiites. As differentiation progresses, these elements are progressively depleted from the melt, resulting in significantly lower concentrations in the residual melt that eventually forms felsic rocks. In addition, incompatible elements such as Y and Zr are not readily incorporated into minerals during the early stages of differentiation, leading to their progressive enrichment in the residual melt and ultimately resulting in elevated concentrations in more evolved felsic rocks (Greber and Dauphas, 2019; Hayashi et al., 1997; Taylor and McLennan, 1985). This elemental behavior results in an inverse correlation with MgO content, namely, the lower the MgO content (i.e., the more evolved the magma), the higher the concentrations of these incompatible elements tend to be. While this feature selection largely reflects differentiation-controlled geochemical patterns, processes such as high-degree mantle melting may lead to higher Ni and MgO concentrations in the melt, because higher degrees of partial melting allow more Ni- and MgO-rich minerals like olivine to melt (Kamber and Ossa Ossa, 2025; Kamber and Tomlinson, 2019). Nonetheless, these complexities do not affect the robustness of our model in predicting the bulk MgO content (Fig. 1b). In summary, the features extracted by the RFR model are highly consistent with established petrological understanding (Greber et al., 2017; Greber and Dauphas, 2019; Kamber and Ossa Ossa, 2025; Kamber and Tomlinson, 2019; Ptáček et al., 2020; Tang et al., 2016). This alignment validates both the geochemical predictive capability and geological plausibility of the model developed in this study.

4.2. Model sensitivity to tectonic settings

A central challenge in reconstructing the Archean subaerial crust composition is the unresolved tectonic framework of early Earth. Critically, tectonic setting governs magmatic geochemistry: calc-alkaline differentiation dominates subduction zones, while tholeiitic differentiation typifies intraplate or plume-related settings (Foley et al., 2002). If geochemical proxies used to infer subaerial crustal composition are sensitive to tectonic setting, then uncertainties in the Archean geodynamic regime may directly affect the reliability of such predictions (e.g. Chen et al., 2020; Garçon, 2021; Greber et al., 2017; Greber and Dauphas, 2019; Keller and Harrison, 2020). For instance, if the Archean subaerial crust formed under a stagnant lid regime with localized plume activity (Bédard, 2006), the associated intraplate magmatic series (tholeiitic) would best approximate the subaerial crust composition. Conversely, if proto-subduction existed (Ge et al., 2018; Harrison et al., 2008; Hopkins et al., 2008), the convergent margin magmatic series (calc-alkaline) would be more relevant.

In the feature elements selected for our RFR model, although elements such as Ti are included, whose content may be influenced by different plate tectonic settings (Chen et al., 2023; Deng et al., 2019), the ensemble learning approach of the RFR inherently avoids reliance on any single proxy (Breiman, 2001). By leveraging the collective information from many feature elements simultaneously, the RFR model minimizes the influence of tectonic setting uncertainties. As shown in Fig. 3, the strong predictive accuracy in both settings (RMSE <0.7 wt%) demonstrates that the model reliably estimates the subaerial crust MgO content independent of tectonic setting. Therefore, we emphasize that one important advantage of the RFR model lies in its robustness to the uncertainty in tectonic settings of the source rocks.

4.3. Compositional evolution of subaerial crust

Our results reveal a discernible trend: the MgO content of the subaerial crust underwent a systematic decline during the late Archean, decreasing from approximately 7–10 wt% at 2.8 Ga to near-modern values of 3–4 wt% at 2.5 Ga (Fig. 5). In other words, by the Archean-Proterozoic boundary at the latest, the subaerial crust had transitioned to a low-MgO, felsic composition. Although some of these sediments may reflect local sources (e.g. Hofmann et al., 2003; Kamber and Ossa Ossa, 2025; Ronov, 1964), global compilations are thought to reveal robust secular evolution (Chen et al., 2020; Gaschnig et al., 2014; Greber et al., 2017; Greber and Dauphas, 2019; McLennan, 1982; Ptáček et al., 2020; Tang et al., 2016; Taylor and McLennan, 1985).

These observations are broadly consistent with the global-scale Ni/Co and Cr/Zn trends compiled by Tang et al. (2016). A similar evolution trend has also been reported in regional studies. For example, fine-grained sedimentary rocks from the ~2.7 Ga Abitibi Greenstone Belt in Canada show a systematic decrease in Ni/Co ratios, albeit over a shorter time span (~70 Ma) (Feng and Kerrich, 1990; Kamber and Tomlinson, 2019). Such rapid regional changes do not obscure the long-term global trends shown here. Greber et al. (2017) questioned the robustness of Ni/Co and Cr/Zn and proposed that the high Ni/Co and Cr/Zn in the Archean sedimentary rocks could result from higher proportions of komatiites and not necessarily reflect a more mafic crust on average. However, Chen et al. (2020) inferred the proportion of Archean basalts in the subaerial crust by analyzing the Cu content in glacial diamictite, finding that the Mesoarchean subaerial crust was primarily composed of basaltic rocks and predominantly mafic. Here we also compiled data for Cu in shales. While the Cu concentration in diamictites decreases from 70 to 80 ppm in the Mesoarchean to approximately 30 ppm at the Archean-Proterozoic boundary (Chen et al., 2020, Fig. 6b), Cu in shales does not show a clear trend. This could be due to the stronger chemical weathering, which led to the leaching of Cu and its subsequent enrichment during shale formation (e.g. Mathur et al., 2012). In contrast, glacial diamictites are primarily formed through physical weathering of their parent rocks, with less influence from chemical weathering (Chen et al., 2020; Kamber and Ossa Ossa, 2025). Therefore, using Cu content to trace the parent rock composition requires limited chemical weathering.

Our results contrast with those from Ti isotopes, Zr/TiO_2, Al_2O_3/TiO_2, and Sm/Nd ratios in sedimentary rocks instead point to a felsic subaerial crust already established by ~3.0 Ga (Garçon, 2021; Greber et al., 2017; Greber and Dauphas, 2019). Greber et al. (2017) used the correlation between Ti isotopes and SiO_2 content in magmatic differentiation to infer the bulk SiO_2 content of the source rocks of the shales through time. This approach was soon challenged by Deng et al. (2019), who found that, under different tectonic settings, Ti isotopes and SiO_2 content follow different correlations such that evolved plume rocks (tholeiitic) are more isotopically fractionated in Ti than differentiated island arc rocks (mainly calc-alkaline). This suggests that Ti isotopes in terrigenous fine-grained sedimentary rocks cannot be used to trace SiO_2 content in crustal rocks when the tectonic setting is unclear. In the data we compiled, neither Al_2O_3/TiO_2 nor Zr/TiO_2 exhibits significant changes over time, consistent with the results of Greber and Dauphas (2019), who attributed the lack of significant temporal changes in Al_2O_3/TiO_2 and Zr/TiO_2 to no significant change in the composition of the subaerial crust. However, we note that the Al_2O_3/TiO_2-SiO_2 and Zr/TiO_2-SiO_2 correlations are strongly nonlinear in igneous rocks (Fig. 7a and b). At low SiO_2 contents (40–65 wt%), Al_2O_3/TiO_2 and Zr/TiO_2 ratios increase slowly, whereas at high SiO_2 contents (65–80 wt%), these ratios increase rapidly and show significant variations. This strongly nonlinear relationship presents an issue: when these ratios are low, even slight fluctuations can lead to significant changes in the predicted SiO_2 content (Fig. 7 a and b). As a result, the changes in subaerial crust SiO_2 content could be masked by the uncertainties intrinsic to the strongly nonlinear relationships between Al_2O_3/TiO_2 and SiO_2 content and between Zr/TiO_2 and SiO_2 content. Similarly, the Sm/Nd proxy proposed by Garçon (2021) exhibits weak and inconsistent variation in the low SiO_2 range (<50 wt%), lacking a clear systematic trend with SiO_2 (Fig. 7c). This insensitivity limits its effectiveness in tracking magmatic differentiation.

Another data-driven study relevant to the present discussion is that of Ptáček et al. (2020), but their conclusions differ from ours and suggested that the subaerial crust attained a felsic composition in the early Archean. The conflicting results may stem from differences in how our method and that of Ptáček et al. (2020) handle geochemical proxies. Ptáček et al. (2020) pre-selected a set of elemental ratios that meet their proxy criteria and assigned equal weights to all proxies in their inverse modeling to estimate rock end-member proportions. By contrast, our approach does not rely on predefined proxy combinations but identifies the optimal element proxies for predicting MgO content based on machine learning. In addition, instead of assigning uniform weights to predefined proxies, our model automatically identifies and emphasizes features that contribute more to predicting MgO during training. We further note that the model of Ptáček et al. (2020) was not tested against different tectonic settings. These differences in computational frameworks may explain the contrasting conclusions.

During the Archean, the rapid decrease in MgO content over time reflects the systematic replacement of the Mg-rich mafic subaerial crust by silica-rich felsic subaerial crust, reaching a composition similar to the present-day subaerial crust by the Archean-Proterozoic boundary. From 2.5 Ga to the present, the MgO content of the subaerial crust has remained largely unchanged, indicating that its composition has been stable since ~2.5 Ga (Fig. 5). Here we follow the calculation of Chen et al. (2020) and assess the lithological proportions of the subaerial crust and their overall compositional changes over time based on the MgO content predicted by our model. The results indicate that during the Mesoarchean, the subaerial crust was predominantly composed of mafic rocks, with komatiite and basalt accounting for 70–100 % of the composition. From the Mesoarchean to the Neoarchean (3.0–2.5 Ga), the proportion of felsic components gradually increased, reaching approximately 70–90 %, a ratio similar to that of the present-day subaerial crust (Rudnick and Gao, 2014), by 2.5 Ga (Fig. 8).

4.4. Implications

The fundamental change in crustal composition has several far-reaching implications. The density of felsic crust is significantly lower than that of mafic crust. According to the principle of Airy isostasy, this density contrast allows felsic crust to achieve greater elevation at the same thickness (Tang et al., 2024). For example, a 45 km-thick felsic crust could generate elevations of 3–5 km above the Archean seafloor, compared to 1–3 km for mafic crust (Tang et al., 2024). This compositional shift near the Archean-Proterozoic boundary may have contributed to continental emergence by enhancing the lateral density contrast in the crust, although this process alone may not be sufficient (Chowdhury et al., 2025; Tang et al., 2024). Indeed, the rise of continents near the end of the Archean appears to be consistent with a number of recent observations including triple oxygen isotope compositions of shales (Bindeman et al., 2018), seawater O isotopes recorded in altered oceanic crust (Johnson and Wing, 2020), the proportion of large igneous provinces (Kump and Barley, 2007), and temporal variation of K/La ratio of volcanic rocks (Liu and He, 2021).

This fundamental shift in the composition of the subaerial crust likely reflects a transition in the dominant tectonic regime during the late Neoarchean. Alternatively, as proposed by Kamber and Ossa Ossa (2025), it may instead record the waning of episodic resurfacing by mafic-ultramafic magmas (e.g., komatiites and basalts), which had previously obscured felsic crustal components in the sedimentary record. This possibility complicates the relationship between the compositional evolution of the subaerial crust and the onset of plate tectonics. However, the rapid mantle cooling (Condie et al., 2016) that led to the cessation of resurfacing may itself have been a consequence of the transition from a stagnant-lid regime to modern-style plate tectonics. Regardless of the mechanism, the composition of the weatherable surface layer of the continents underwent a fundamental transformation during this period. The exposure of the continental surface, accompanied by a shift from mafic to felsic lithologies, would have facilitated subaerial weathering of silica-rich rocks, enhancing the drawdown of atmospheric CO_2 and increasing the flux of nutrients to the oceans (Bindeman et al., 2018; Chowdhury et al., 2025; Husson and Peters, 2017; Rey and Coltice, 2008). Because a felsic crust contains less Fe^{2+} rich minerals and sulfides, a felsic crust consumes less oxygen than a mafic crust during oxidative weathering (Lee et al., 2016). In addition, felsic minerals such as feldspar and quartz produce higher amounts of

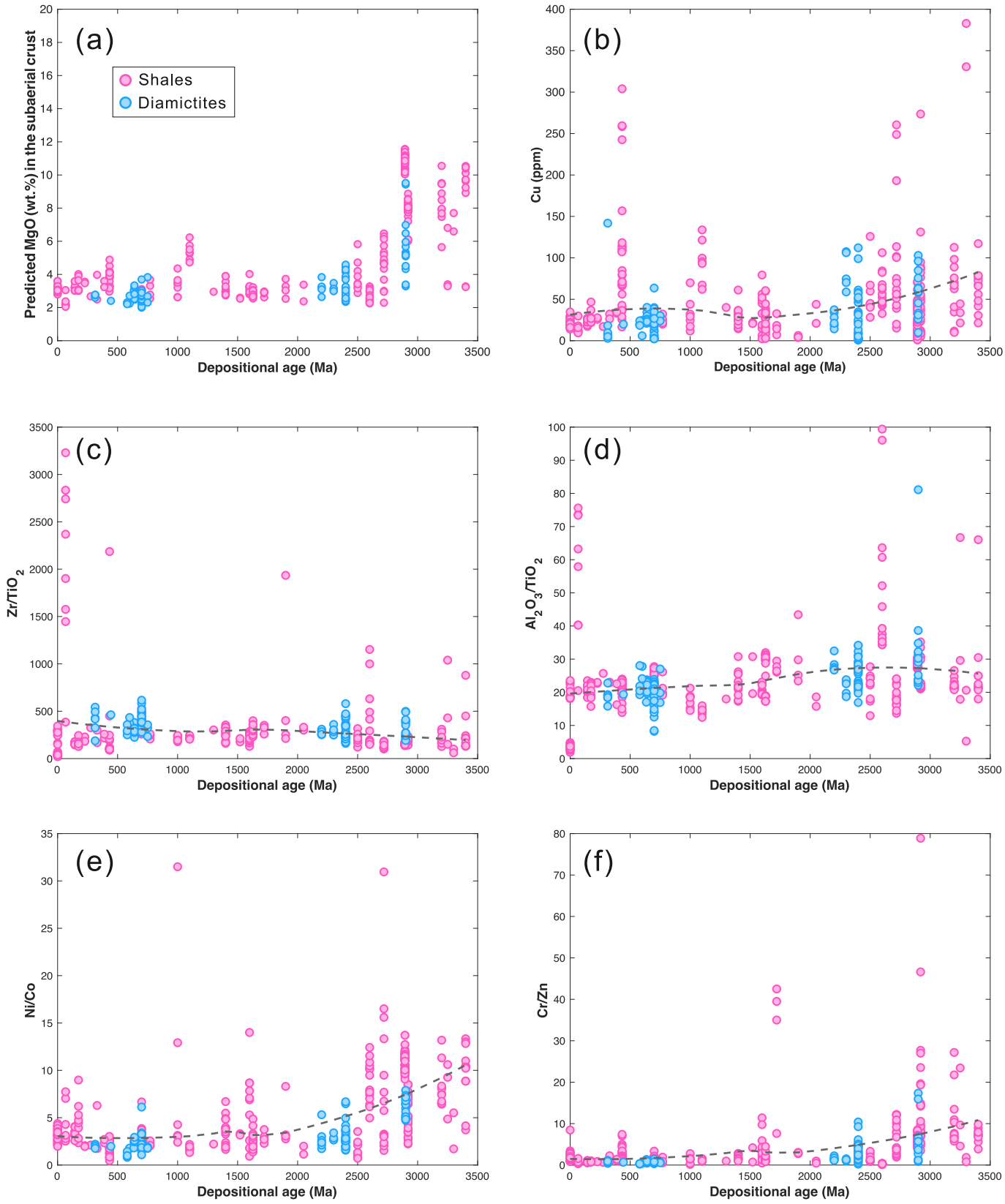


Fig. 6. Elements content or ratios of terrigenous fine-grained sediments plotted versus their age. (a) Our model predicted MgO content (wt%) in the subaerial crust, (b) Cu content (ppm), normalized to the average subaerial crust abundance of Al_2O_3 (15.26 %) as reported by Rudnick and Gao (2003), in order to eliminate the influence of dilution from quartz and carbonate, (c) Zr/TiO_2, (d) Al_2O_3/TiO_2, (e) Ni/Co, (f) Cr/Zn.

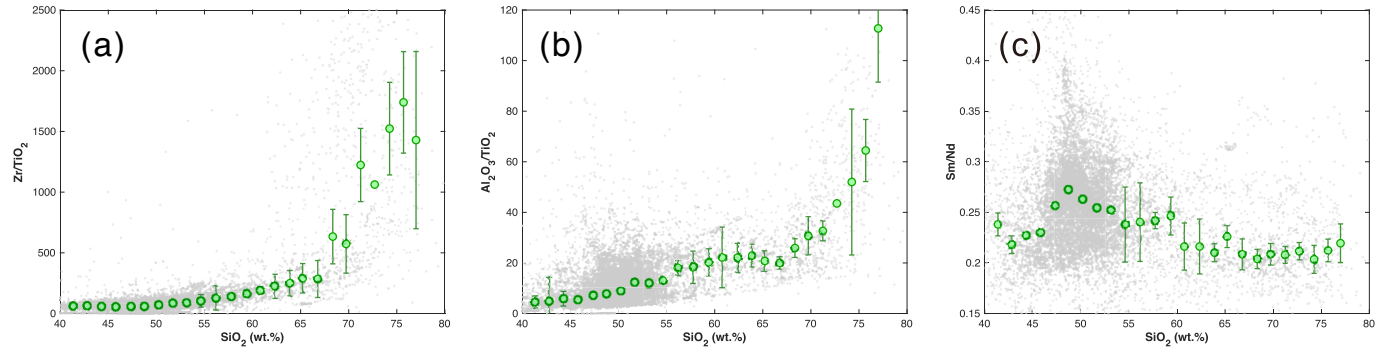


Fig. 7. Zr/TiO_2, Al_2O_3/TiO_2 and Sm/Nd ratios vs. SiO_2 content (wt%) of igneous rocks. Gray dots are from the dataset used in Greber and Dauphas (2019). Green dots represent the median values obtained by binning based on SiO_2 (bin size = 1.5 wt%). Error bars are 2 SE. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

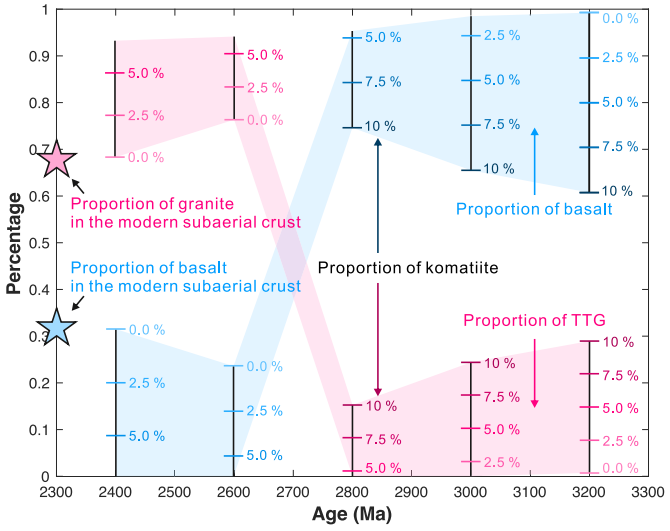


Fig. 8. Calculated mass proportions of felsic rocks (pink) and basalts (blue) in the subaerial crust over time. The vertical percentages (0 %, 2.5 %, 5 %, etc.) represent the proportions of felsic rocks and basalts under varying komatiite contributions, with komatiite mass fractions ranging from 0 % to 10 % of the total crustal mass. Komatiites have an average MgO content of 26 wt%. Archean TTGs and basalts have average MgO contents of 1.1 wt% and 7.6 wt%, respectively. Post-Archean TTGs and basalts have average MgO contents of 1.6 wt% and 6.9 wt%, respectively (Condie, 1993). The star markers correspond to the estimated proportions of basalts and granites in the modern subaerial crust (Rudnick and Gao, 2014). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

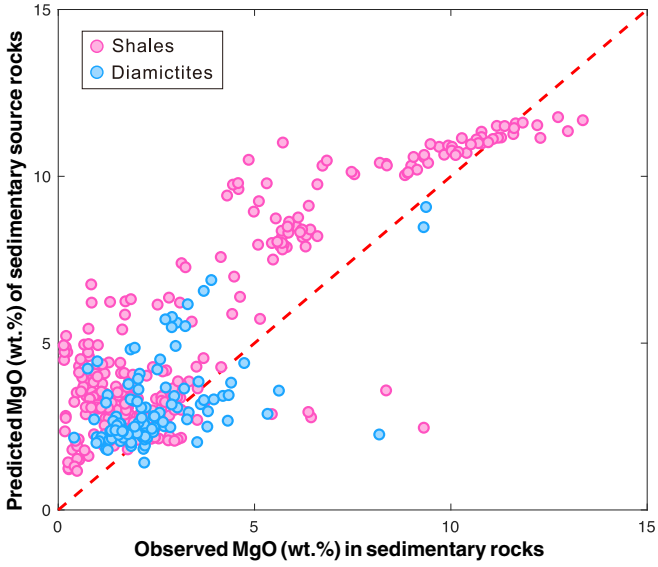


Fig. 9. Comparison between observed MgO content in fine-grained sedimentary rocks and the corresponding source rocks compositions predicted using the RFR model. Most data points lie above the 1:1 line, indicating systematic loss of Mg during weathering and sediment transport.

reactive oxygen species during physical weathering compared to mafic minerals (Wu et al., 2023), which would further increase the size of surface oxidant reservoirs. These processes together may have contributed to breaking the redox balance at the surface near the end of the Archean and eventually triggered the Great Oxidation Event (GOE) around 2.4 Ga (Holland, 2002).

One final interesting observation is that the source MgO content predicted from our RFR model positively correlates with the MgO content of fine-grained sedimentary rocks but is generally higher than the latter (Fig. 9). This offset is consistent with Mg being readily mobilized and leached during chemical weathering (Holland, 2020; Kamber and Ossa Ossa, 2025; Rudnick and Gao, 2014). Quantifying the loss of soluble elements in fine-grained sedimentary rocks (e.g., Mg, Na, and K) may provide insights into the cycling of weathering-controlled elements on early Earth.

5. Conclusions

Our study presents a novel machine learning approach using Random Forest Regression to reconstruct the MgO content in the subaerial crust through multi-element geochemical analysis of terrigenous fine-grained sedimentary rocks over Earth's history. By training the model on simulated mixtures of Archean and post-Archean igneous rocks using selected feature elements (TiO_2, Al_2O_3, Sc, Cr, Co, Ni, Cu, Zn, Y, Zr, Nb and REEs), we demonstrate its robustness in predicting MgO content for the modern subaerial crust and is independent of tectonic settings. Our results reveal a systematic decline in subaerial crust MgO content from 7 to 10 wt% to modern-like values of 3–4 wt%, accompanied by a compositional shift from mafic (70–100 % komatiite-basalt) to felsic (~70–90 % TTG) dominance in the Neoarchean period. This transition marks the stabilization of subaerial crust composition into a character resembling modern subaerial crust.

CRediT authorship contribution statement

Junteng Lyu: Writing – original draft, Visualization, Software, Methodology. **Ziyi Guo:** Writing – review & editing, Visualization. **Ming Tang:** Writing – review & editing, Resources, Methodology, Funding

acquisition, Conceptualization.

Funding

This work was supported by the National Key Research and Development Program of China (grant number 2023YFF0803801), National Natural Science Foundation of China (42125302 and 42073026).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We sincerely thank Editor Sonja Aulbach, and the two anonymous reviewers for their efficient editorial handling and for providing insightful and constructive comments. Their valuable suggestions have significantly improved the quality and clarity of this manuscript. We also thank our group members Hao Chen, Runwu Li, Jiazhen Wang, Zi Ye, Xuyang Zheng, Mimi Chen, Xi Huang and Weizhe Qin as well as others who provided valuable suggestions during the course of this work. This work was supported by the National Key Research and Development Program of China (grant number 2023YFF0803801), National Natural Science Foundation of China (42125302 and 42073026).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemgeo.2025.122986.

Data availability

I have shared the link to my data/code at the Attach File step.

References

Allègre, C.J., Treuil, M., Minster, J.-F., Minster, B., Albarède, F., 1977. Systematic use of trace element in igneous process: Part I: Fractional crystallization processes in volcanic suites. Contrib. Mineral. Petrol. 60, 57–75. https://doi.org/10.1007/BF00372851.

Bédard, J.H., 2006. A catalytic delamination-driven model for coupled genesis of Archaean crust and sub-continental lithospheric mantle. Geochim. Cosmochim. Acta 70, 1188–1214. https://doi.org/10.1016/j.gca.2005.11.008.

Berg, R.D., Solomon, E.A., Teng, F.-Z., 2019. The role of marine sediment diagenesis in the modern oceanic magnesium cycle. Nat. Commun. 10, 4371. https://doi.org/10.1038/s41467-019-12322-2.

Bindeman, I.N., Zakharov, D.O., Palandri, J., Greber, N.D., Dauphas, N., Retallack, G.J., Hofmann, A., Lackey, J.S., Bekker, A., 2018. Rapid emergence of subaerial landmasses and onset of a modern hydrologic cycle 2.5 billion years ago. Nature 557, 545–548. https://doi.org/10.1038/s41586-018-0131-1.

Bougault, H., Hekinian, R., 1974. Rift Valley in the Atlantic Ocean near 36°50′N: petrology and geochemistry of basaltic rocks. Earth Planet. Sci. Lett. 24, 249–261. https://doi.org/10.1016/0012-821X(74)90103-4.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Cawood, P.A., Chowdhury, P., Mulder, J.A., Hawkesworth, C.J., Capitanio, F.A., Gunawardana, P.M., Nebel, O., 2022. Secular evolution of continents and the Earth system. Rev. Geophys. 60, e2022RG000789. https://doi.org/10.1029/2022RG000789.

Chen, K., Rudnick, R.L., Wang, Z., Tang, M., Gaschnig, R.M., Zou, Z., He, T., Hu, Z., Liu, Y., 2020. How mafic was the Archean upper continental crust? Insights from Cu and Ag in ancient glacial diamictites. Geochim. Cosmochim. Acta 278, 16–29.

Chen, K., Tang, M., Hu, Z., Liu, Y., 2023. Generation of tholeiitic and calc-alkaline arc magmas and its implications for continental growth. Geochim. Cosmochim. Acta 355, 173–183. https://doi.org/10.1016/j.gca.2023.07.002.

Chowdhury, P., Cawood, P.A., Mulder, J.A., 2025. Subaerial emergence of continents on Archean Earth. Annu. Rev. Earth Planet. Sci. https://doi.org/10.1146/annurev-earth-040722-093345.

Condie, K.C., 1993. Chemical composition and evolution of the upper continental crust: contrasting results from surface samples and shales. Chem. Geol. 104, 1–37. https://doi.org/10.1016/0009-2541(93)90140-E.

Condie, K.C., Aster, R.C., Van Hunen, J., 2016. A great thermal divergence in the mantle beginning 2.5 Ga: Geochemical constraints from greenstone basalts and komatiites. Geosci. Front. 7, 543–553. https://doi.org/10.1016/j.gsf.2016.01.006.

Deng, Z., Chaussidon, M., Savage, P., Robert, F., Pik, R., Moynier, F., 2019. Titanium isotopes as a tracer for the plume or island arc affinity of felsic rocks. Proc. Natl. Acad. Sci. USA 116, 1132–1135. https://doi.org/10.1073/pnas.1809164116.

Dhuime, B., Wuestefeld, A., Hawkesworth, C.J., 2015. Emergence of modern continental crust about 3 billion years ago. Nat. Geosci. 8, 552–555. https://doi.org/10.1038/ngeo2466.

Feng, R., Kerrich, R., 1990. Geochemistry of fine-grained clastic sediments in the Archean Abitibi greenstone belt, Canada: Implications for provenance and tectonic setting. Geochim. Cosmochim. Acta 54, 1061–1081. https://doi.org/10.1016/0016-7037(90)90439-R.

Foley, S., Tiepolo, M., Vannucci, R., 2002. Growth of early continental crust controlled by melting of amphibolite in subduction zones. Nature 417, 837–840. https://doi.org/10.1038/nature00799.

Gao, S., Luo, T.-C., Zhang, B.-R., Zhang, H.-F., Han, Y., Zhao, Z.-D., Hu, Y.-K., 1998. Chemical composition of the continental crust as revealed by studies in East China. Geochim. Cosmochim. Acta 62, 1959–1975. https://doi.org/10.1016/S0016-7037(98)00121-5.

Garçon, M., 2021. Episodic growth of felsic continents in the past 3.7 Ga. Sci. Adv. 7, eabj1807. https://doi.org/10.1126/sciadv.abj1807.

Gaschnig, R.M., Rudnick, R.L., McDonough, W.F., Kaufman, A.J., Hu, Z., Gao, S., 2014. Onset of oxidative weathering of continents recorded in the geochemistry of ancient glacial diamictites. Earth Planet. Sci. Lett. 408, 87–99. https://doi.org/10.1016/j.epsl.2014.10.002.

Ge, R., Zhu, W., Wilde, S.A., Wu, H., 2018. Remnants of Eoarchean continental crust derived from a subducted proto-arc. Sci. Adv. 4, eaao3159. https://doi.org/10.1126/sciadv.aao3159.

Greber, N.D., Dauphas, N., 2019. The chemistry of fine-grained terrigenous sediments reveals a chemically evolved Paleoarchean emerged crust. Geochim. Cosmochim. Acta 255, 247–264. https://doi.org/10.1016/j.gca.2019.04.012.

Greber, N.D., Dauphas, N., Bekker, A., Ptáček, M.P., Bindeman, I.N., Hofmann, A., 2017. Titanium isotopic evidence for felsic crust and plate tectonics 3.5 billion years ago. Science 357, 1271–1274. https://doi.org/10.1126/science.aan8086.

Harrison, T.M., Schmitt, A.K., McCulloch, M.T., Lovera, O.M., 2008. Early (≥4.5 Ga) formation of terrestrial crust: Lu–Hf, δ18O, and Ti thermometry results for Hadean zircons. Earth Planet. Sci. Lett. 268, 476–486. https://doi.org/10.1016/j.epsl.2008.02.011.

Hayashi, K.-I., Fujisawa, H., Holland, H.D., Ohmoto, H., 1997. Geochemistry of ~1.9 Ga sedimentary rocks from northeastern Labrador, Canada. Geochim. Cosmochim. Acta 61, 4115–4137. https://doi.org/10.1016/S0016-7037(97)00214-7.

Hofmann, A., Bolhar, R., Dirks, P., Jelsma, H., 2003. The geochemistry of Archaean shales derived from a Mafic volcanic sequence, Belingwe greenstone belt, Zimbabwe: provenance, source area unroofing and submarine versus subaerial weathering. Geochim. Cosmochim. Acta 67, 421–440. https://doi.org/10.1016/S0016-7037(02)01086-4.

Holland, H.D., 2002. Volcanic gases, black smokers, and the great oxidation event. Geochim. Cosmochim. Acta 66, 3811–3826. https://doi.org/10.1016/S0016-7037(02)00950-X.

Holland, H.D., 2020. The Chemical Evolution of the Atmosphere and Oceans. Princeton University Press.

Hopkins, M., Harrison, T.M., Manning, C.E., 2008. Low heat flow inferred from >4 Gyr zircons suggests Hadean plate boundary interactions. Nature 456, 493–496. https://doi.org/10.1038/nature07465.

Horn, I., Foley, S.F., Jackson, S.E., Jenner, G.A., 1994. Experimentally determined partitioning of high field strength- and selected transition elements between spinel and basaltic melt. Chem. Geol. 117, 193–218. https://doi.org/10.1016/0009-2541(94)90128-7.

Husson, J.M., Peters, S.E., 2017. Atmospheric oxygenation driven by unsteady growth of the continental sedimentary reservoir. Earth Planet. Sci. Lett. 460, 68–75. https://doi.org/10.1016/j.epsl.2016.12.012.

Johnson, B.W., Wing, B.A., 2020. Limited Archaean continental emergence reflected in an early Archaean 18O-enriched ocean. Nat. Geosci. 13, 243–248. https://doi.org/10.1038/s41561-020-0538-9.

Kamber, B.S., Ossa Ossa, F., 2025. Evolution of continental crust and sedimentary rock chemistry through time. In: Treatise on Geochemistry. Elsevier, pp. 729–773. https://doi.org/10.1016/B978-0-323-99762-1.00007-3.

Kamber, B.S., Tomlinson, E.L., 2019. Petrological, mineralogical and geochemical peculiarities of Archaean cratons. Chem. Geol. 511, 123–151. https://doi.org/10.1016/j.chemgeo.2019.02.011.

Keller, C.B., Harrison, T.M., 2020. Constraining crustal silica on ancient Earth. Proc. Natl. Acad. Sci. USA 117, 21101–21107. https://doi.org/10.1073/pnas.2009431117.

Keller, C.B., Schoene, B., Barboni, M., Samperton, K.M., Husson, J.M., 2015. Volcanic–plutonic parity and the differentiation of the continental crust. Nature 523, 301–307. https://doi.org/10.1038/nature14584.

Konhauser, K.O., Pecoits, E., Lalonde, S.V., Papineau, D., Nisbet, E.G., Barley, M.E., Arndt, N.T., Zahnle, K., Kamber, B.S., 2009. Oceanic nickel depletion and a methanogen famine before the Great Oxidation Event. Nature 458, 750–753. https://doi.org/10.1038/nature07858.

Kump, L.R., Barley, M.E., 2007. Increased subaerial volcanism and the rise of atmospheric oxygen 2.5 billion years ago. Nature 448, 1033–1036. https://doi.org/10.1038/nature06058.

Lee, C.-T.A., Yeung, L.Y., McKenzie, N.R., Yokoyama, Y., Ozaki, K., Lenardic, A., 2016. Two-step rise of atmospheric oxygen linked to the growth of continents. Nat. Geosci. 9, 417–424. https://doi.org/10.1038/ngeo2707.

Leeman, W.P., Scheidegger, K.F., 1977. Olivine/liquid distribution coefficients and a test for crystal-liquid equilibrium. Earth Planet. Sci. Lett. 35, 247–257. https://doi.org/10.1016/0012-821X(77)90128-5.

Liu, C.-T., He, Y.-S., 2021. Rise of major subaerial landmasses about 3.0 to 2.7 billion years ago. Geochem. Persp. Let. 18, 1–5. https://doi.org/10.7185/geochemlet.2115.

Mathur, R., Jin, L., Prush, V., Paul, J., Ebersole, C., Fornadel, A., Williams, J.Z., Brantley, S., 2012. Cu isotopes and concentrations during weathering of black shale of the Marcellus Formation, Huntingdon County, Pennsylvania (USA). Chem. Geol. 304–305, 175–184. https://doi.org/10.1016/j.chemgeo.2012.02.015.

McLennan, S.M., 1982. On the geochemical evolution of sedimentary rocks. Chem. Geol. 37, 335–350. https://doi.org/10.1016/0009-2541(82)90087-0.

Ptáček, M.P., Dauphas, N., Greber, N.D., 2020. Chemical evolution of the continental crust from a data-driven inversion of terrigenous sediment compositions. Earth Planet. Sci. Lett. 539, 116090. https://doi.org/10.1016/j.epsl.2020.116090.

Rey, P.F., Coltice, N., 2008. Neoarchean lithospheric strengthening and the coupling of Earth's geochemical reservoirs. Geol 36, 635. https://doi.org/10.1130/G25031A.1.

Ronov, A.B., 1964. Common tendencies in the chemical evolution of the earth's crust, ocean and atmosphere. Geochem. Int. 1, 713–737.

Rudnick, R.L., 1995. Making continental crust. Nature 378, 571–578. https://doi.org/10.1038/378571a0.

Rudnick, R.L., Gao, S., 2003. 3.01 - Composition of the Continental Crust. In: Holland, H. D., Turekian, K.K. (Eds.), Treatise on Geochemistry. Pergamon, Oxford, pp. 1–64. https://doi.org/10.1016/B0-08-043751-6/03016-4.

Rudnick, R.L., Gao, S., 2014. Composition of the Continental Crust. In: Treatise on Geochemistry. Elsevier, pp. 1–51. https://doi.org/10.1016/B978-0-08-095975-7.00301-6.

Smit, M.A., Mezger, K., 2017. Earth's early O2 cycle suppressed by primitive continents. Nat. Geosci. 10, 788–792. https://doi.org/10.1038/ngeo3030.

Tang, M., 2021. Composition of the Earth's Crust. In: Encyclopedia of Geology. Elsevier, pp. 178–186. https://doi.org/10.1016/B978-0-08-102908-4.00044-8.

Tang, M., Rudnick, R.L., McDonough, W.F., Gaschnig, R.M., Huang, Y., 2015. Europium anomalies constrain the mass of recycled lower continental crust. Geology 43, 703–706. https://doi.org/10.1130/G36641.1.

Tang, M., Chen, K., Rudnick, R.L., 2016. Archean upper crust transition from mafic to felsic marks the onset of plate tectonics. Science 351, 372–375. https://doi.org/10.1126/science.aad5513.

Tang, M., Chen, H., Lee, C.-T.A., Cao, W., 2024. Subaerial crust emergence hindered by phase-driven lower crust densification on early Earth. Sci. Adv. 10, eadq1952. https://doi.org/10.1126/sciadv.adq1952.

Taylor, S.R., McLennan, S.M., 1985. The Continental Crust: Its Composition and Evolution.

Taylor, S.R., McLennan, S., 2009. Planetary Crusts: Their Composition, Origin and Evolution. Cambridge University Press.

Wang, B., Granitto, M., Labay, K.A., Pongratz, A.L., 2024. Selected Whole-Rock and Sediment Geochemistry Reanalysis Data Additions to the Alaska Geochemical DataBase (AGDB) from Archived Samples. U.S. Geological Survey Data Release, Alaska. https://doi.org/10.5066/P9PR1T6X.

Wu, X., Zhu, J., He, H., Xian, H., Yang, Y., Ma, L., Liang, X., Lin, X., Li, S., Konhauser, K. O., Li, Y., 2023. Geodynamic oxidation of Archean terrestrial surfaces. Commun. Earth Environ. 4, 132. https://doi.org/10.1038/s43247-023-00789-3.