RESEARCH ARTICLE
PLANETARY SCIENCES

Earth and Planetary Physics
8: 854–867, 2024
doi: 10.26464/epp2024011

Quantifying the chemical composition of weathering products of Hainan basalts with reflectance spectroscopy and its implications for Mars

Xing Wu[1†], JiaCheng Liu[2†], WeiChao Sun[3], Yang Liu[1,4,5*], Joseph Michalski[2], Wei Tan[6], XiaoRong Qin[6], and YongLiao Zou[1]

[1]State Key Laboratory of Space Weather, National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China;

[2]Department of Earth Sciences and Laboratory for Space Research, The University of Hong Kong, Hong Kong, China;

[3]Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;

[4]Center for Excellence in Comparative Planetology, Chinese Academy of Sciences, Hefei 230026, China;

[5]University of Chinese Academy of Science, Beijing 100049, China;

[6]Key Laboratory of Mineralogy and Metallogeny, Guangdong Provincial Key Laboratory of Mineral Physics and Materials, Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou 510640, China

Key Points:

- We established a spectral quantitative model to derive the chemical concentrations of Hainan weathered basalts.
- The effects of various spectral preprocessing methods on the prediction accuracy were evaluated.
- The spectral subsets selected by our model offer guidelines for the design of spectral channels for future spectrometers.
- The proposed model has the potential to estimate chemical concentrations in basaltic weathering profiles on Mars.

Citation: Wu, X., Liu, J. C., Sun, W. C., Liu, Y., Michalski, J., Tan, W., Qin, X. R., and Zou, Y. L. (2024). Quantifying the chemical composition of weathering products of Hainan basalts with reflectance spectroscopy and its implications for Mars. *Earth Planet. Phys.*, *8*(6), 854–867. http://doi.org/10.26464/epp2024011

Abstract: With the development of the hyperspectral remote sensing technique, extensive chemical weathering profiles have been identified on Mars. These weathering sequences, formed through precipitation-driven leaching processes, can reflect the paleoenvironments and paleoclimates during pedogenic processes. The specific composition and stratigraphic profiles mirror the mineralogical and chemical trends observed in weathered basalts on Hainan Island in south China. In this study, we investigated the laboratory reflectance spectra of a 53-m-long drilling core of a thick basaltic weathering profile collected from Hainan Island. We established a quantitative spectral model by combining the genetic algorithm and partial least squares regression (GA-PLSR) to predict the chemical properties (SiO_2, Al_2O_3, Fe_2O_3) and index of laterization (IOL). The entire sample set was divided into a calibration set of 25 samples and a validation set of 12 samples. Specifically, the GA was used to select the spectral subsets for each composition, which were then input into the PLSR model to derive the chemical concentration. The coefficient of determination (R^2) values on the validation set for SiO_2, Al_2O_3, Fe_2O_3, and the IOL were greater than 0.9. In addition, the effects of various spectral preprocessing techniques on the model accuracy were evaluated. We found that the spectral derivative treatment boosted the prediction accuracy of the GA-PLSR model. The improvement achieved with the second derivative was more pronounced than when using the first derivative. The quantitative model developed in this work has the potential to estimate the contents of similar weathering basalt products, and thus infer the degree of alteration and provide insights into paleoclimatic conditions. Moreover, the informative bands selected by the GA can serve as a guideline for designing spectral channels for the next generation of spectrometers.

Keywords: reflectance spectroscopy; weathered basalts; terrestrial analog; quantitative retrieval; Mars

1. Introduction

Although modern Mars is cold and hyperarid, abundant morphological records, such as fluvial and lacustrine features, indicate liquid water flowed on the surface of early Mars (Fassett and Head, 2008; Hynek et al., 2010; Goudge et al., 2015; Michalski et al., 2022). The diverse and widespread clay minerals detected in ancient terrains further support the aqueous weathering that existed on Mars (e.g., Poulet et al., 2005; Mustard et al., 2008; Ehlmann et al., 2009; Du PX et al., 2023). These minerals are regarded as tracers of aqueous episodes as well as major constitu-

First author: X. Wu, wuxing@nssc.ac.cn
[†] These authors contributed equally to this work.
Correspondence to: Y. Liu, yangliu@nssc.ac.cn
Received 10 OCT 2023; Accepted 19 JAN 2024.
First Published online 06 FEB 2024.

tions of exobiological targets (Ehlmann et al., 2008). The Martian weathering profiles identified from orbit show the Al-clays stratigraphically above Fe/Mg-clays, which favors a precipitation-driven top-down leaching (pedogenic weathering) process (Gaudin et al., 2011; Carter et al., 2015; Ye BL and Michalski, 2022). Specifically, as weathering proceeds, mobile and soluble elements are leached, whereas immobile elements are enriched, resulting in different weathering products (Liu JC et al., 2021a). Therefore, the nature of the weathering profile is associated with the environmental and climatic conditions that prevailed during pedogenesis (Sheldon and Tabor, 2009; Liu JC et al., 2021b).

Several studies have been conducted to survey the distribution and chronology of weathering profiles on Mars. The specific Al-clay over Fe/Mg-clay profile mirrors the mineralogical and chemical trends observed in weathered basalts in Hainan Island in south China (He Y et al., 2008; Liu ZF et al., 2017), which has direct implications for weathering sequences on Mars (Ehlmann et al., 2012; Carter et al., 2015; Liu JC et al., 2021b), the crustal protolith of which is dominated by basalt (McSween et al., 2009). Investigations of terrestrial analogs are crucial for advancing our knowledge of weathering on Mars (Gaudin et al., 2011; Broz et al., 2022). Various mineralogic and geochemical proxies determined via laboratory analyses have been utilized as indicators of the weathering intensity. For instance, the chemical index of alteration (CIA) and index of laterization (IOL) are widely used to interpret the degree of alteration (Wang P et al., 2020). Although accurate, geochemical analysis is unsuitable for large-scale investigation owing to the costly and time-consuming nature of sample preparation and analysis (Rossel and Behrens, 2010). Moreover, wet chemical methods are not environmentally friendly because they produce hazardous wastes (Rossel and Behrens, 2010).

Visible and near-infrared (VNIR) reflectance spectroscopy has attracted increasing attention over the past decades because of its advantages of portability, rapidity, and nondestruction (Rossel and Behrens, 2010; Wang JJ et al., 2014; Zhao LL et al., 2018; Sun WC et al., 2022a). The absorption features of soil spectra in the VNIR region are dominated by spectrally active components, including organic matter, iron oxides, clay minerals, and water, owing to the electronic transitions, bending, and stretching vibrations of chemical bonds (Wu YZ et al., 2007). Therefore, various soil compositions can be inferred from a single spectral measurement. Visible and near-infrared reflectance spectroscopy has been regarded as an alternative to rapidly predict soil properties. Zhao LL et al. (2018) assessed the relationship between VNIR spectra and geochemical proxies from red earth sediments in south China. They found that the degree of weathering was closely associated with the band depth ratio of spectral absorption values around 2200 nm and 1900 nm (BD2200/BD1900). Liu JC et al. (2021a) argued that the band depth ratio BD1400/BD1900 is a more suitable way to evaluate the relative leaching and hydrolysis intensity of weathered basalts. Nonetheless, the application of spectral parameters can be greatly affected by the mineral compositions and protolith types (Liu JC et al., 2021a).

Numerous qualitative methods have been proposed to predict soil properties by VNIR reflectance. The partial least squares regression (PLSR) is an extensively used multivariate technique for the prediction of soil properties (Wu YZ et al., 2007; Rossel and Behrens, 2010; Zhao LL et al., 2018). Proper spectral pretreatments with a PLSR have been reported to produce predictions comparable with those using more complex algorithms (Wang JJ et al., 2014; Hong YS et al., 2019). The widely used spectral preprocessing techniques include spectral derivatives and spectral transformations (Hong YS et al., 2019). For example, the first and second derivatives eliminate baseline effects and isolate overlapping peaks (Nawar et al., 2016). To reduce the nonlinearities and scattering effects, some researchers have transformed the VNIR reflectance into absorbance (Wang JJ et al., 2014). This dimensional reduction is essential in model calibration, especially when the number of bands far exceeds the number of samples. The genetic algorithm (GA) is an optimization strategy that facilitates a random and global search within a high-dimensional space, and it is extensively used for band selection in spectroscopic multivariate calibration (Durand et al., 2007). The combination of GA and PLSR significantly promotes the performance of the PLSR and has been successfully used to estimate soil properties (Vohland et al., 2011; Wang JJ et al., 2014; Sun WC et al., 2022a).

In this study, we investigated the reflectance spectra of a 53-m-long drilling core of a thick basaltic weathering sequence on Hainan Island. The spectral study of basaltic weathering products provides an alternative perspective on the paleoclimate and paleoenvironment (Liu JC et al., 2021a). We used the GA and PLSR to predict weathering-related chemical concentrations. The objectives of this work were threefold: (1) to establish a quantitative model between the chemical compositions and VNIR spectra of weathered basalt; (2) to evaluate the effects of various spectral preprocessing methods on their prediction accuracy; and (3) to select informative spectral bands critical to the chemical properties. The established model has the potential to estimate chemical concentrations of similar weathering products. Moreover, the informative spectral bands could offer guidelines for the design of spectral channels for future spectrometers.

2. Dataset and Methodology

2.1 Dataset

2.1.1 The basaltic sequence on Hainan Island

The Chengmai basaltic weathering profile is situated to the north of Chengmai County in Hainan Province, south China (Figures 1a and 1b). The basalts have been deeply weathered over millions of years under a tropical climate (He Y et al., 2008). A 53-m-long drilling core was studied in this work. From bottom to top, the examined basaltic weathering profile at Chengmai consists of a humic zone (Zone A: 5.25–0.00 m), a totally weathered zone (Zone B: 39.00–5.25 m), and a semi-weathered zone (Zone C: 53.00–39.00 m) (Liu JC et al., 2021a). The core samples were collected at intervals of 0.5 m in the topmost profile, and lithologic contacts were 0.5 m. For homogeneous sections, samples were taken at intervals of 1.0 to 2.0 m. A total of 42 core samples were collected. The representative core samples are shown in Figure 1c. The dark tone in Zone A indicates the elevated total organic carbon concentrations. The core samples between 20.0 and 31.5 m contain veins and have textures resembling almonds that are filled with white materials. The Zone B samples are reddish in

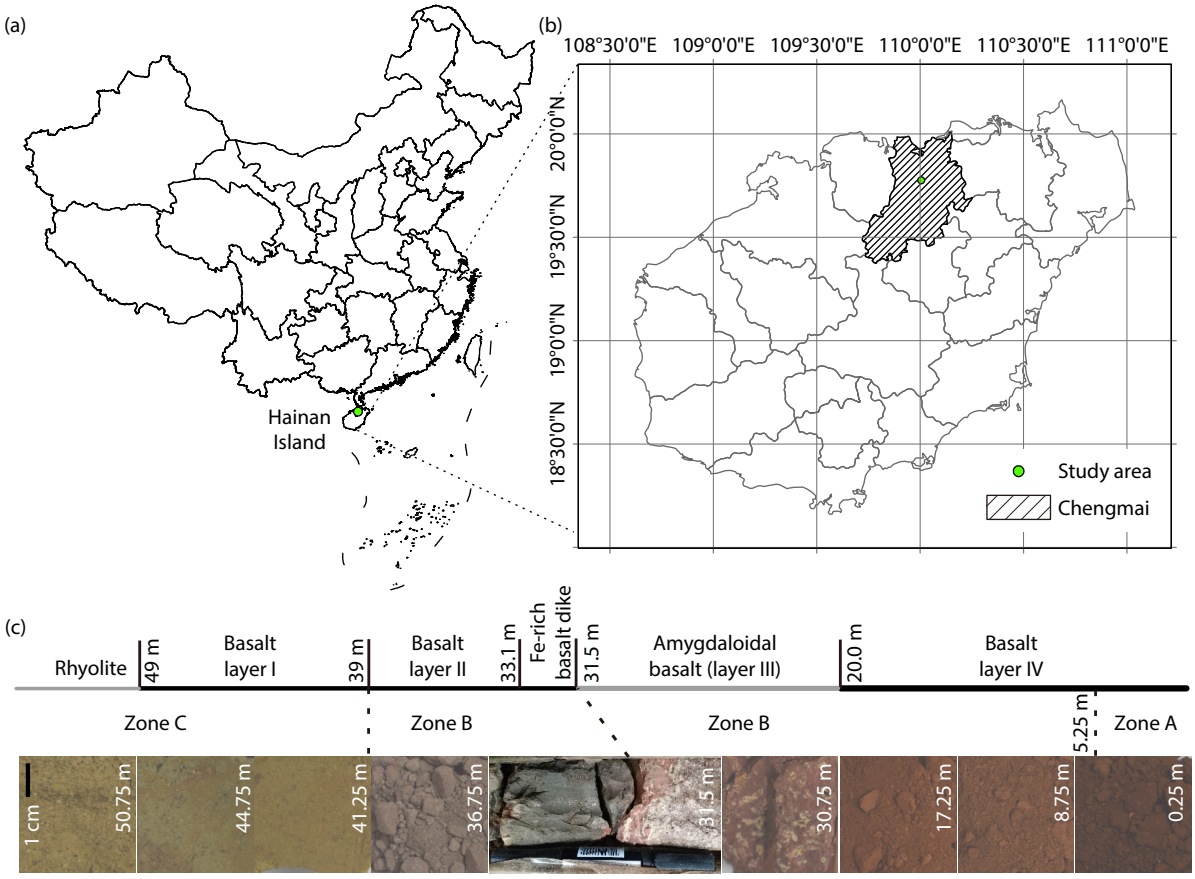

Figure 1. (a) The map of China displays the location of the basalt weathering profile in Hainan Island, south China. (b) The map enlargement shows the location of the study area (19°46′34″N, 110°00′23″E) in Chengmai County. (c) Divisions and representative photographs of the 53-m-long drilling core, modified from Liu JC et al. (2021a).

hue. In Zone C, the weathered rocks are consolidated and exhibit a yellow coloration, marking a sudden shift in both color and texture when transitioning from Zone B to Zone C.

2.1.2 Bulk sample chemical compositions

The collected samples were sieved to remove large debris and then ground into fine particles (<74 μm). The powder of each sample was baked at 105 °C to remove absorbed water and then cooled in a closed glass dish before measurements (Liu JC et al., 2021a). The major oxides (e.g., SiO_2, Al_2O_3, and Fe_2O_3) were identified by wavelength-dispersive X-ray fluorescence spectroscopy of fused glass beads. They were measured with an ARL Perform'X 4200 X-ray fluorescence spectrometer on glass disks, which were created by melting an 8 g mixture of $Li_2B_4O_7$ (66%) and $LiBO_2$ (34%), together with 400 mg of sample powder, at 1050 °C. We also measured the certified reference materials SARM-4, SAEM-45, and GBW07105 as standards. The major element detection limits were better than 30 ppm, and the analytical accuracy was better than 1%. The mineralogical analysis was conducted with a Rigaku D/max-1200 diffractometer with Co Kα1 radiation (15 mA and 40 kV). Silica was used to calibrate the reflection peak positions of the XRD patterns (Liu JC et al., 2021a).

Although the CIA is a more widely used indicator than the IOL, most samples have CIA values exceeding 90% (Liu JC et al., 2021a), making it challenging to differentiate the degree of alteration. Therefore, we used the IOL [IOL = $(Al_2O_3 + Fe_2O_3)/(SiO_2 + Al_2O_3 + Fe_2O_3)$] to measure the weathering intensity in this work. In Zone A, the samples exhibited the highest IOL values, ranging from 50 to 60, indicating a highly weathered zone. These samples consisted of quartz, kaolinite, gibbsite, hematite, goethite, and anatase. Correspondingly, this zone had higher concentrations of Al_2O_3 (25–28 wt%) and Fe_2O_3 (18–22 wt%) compared with Zone B but a lower content of SiO_2 (34–41 wt%). Additionally, the concentration of Al_2O_3 increased upward, whereas the concentrations of Fe_2O_3 declined with increasing total organic carbon. In Zone B, the samples exhibited increasing IOL values, ranging from 30 to 65, corresponding to a highly weathered zone. Notably, samples collected between 22.25 and 24.25 m exhibited high SiO_2 contents (56–60 wt%) and low IOL values (32–40). The sample collected from the weathered mafic dike at 32.75 m was extremely rich in Fe (31.80 wt%) and poor in SiO_2 (29.55 wt%). In Zone C, the IOL values gradually increased from 33 to 40, and this region corresponded to a moderately weathered zone. The slightly weathered samples were composed of plagioclase, nontronite, kaolinite, and almost no quartz or Fe (oxyhydr)oxides. They had inhomogeneous chemical compositions and contained 9–20 wt% Al_2O_3, 41–68 wt% SiO_2, and 7–16 wt% Fe_2O_3. Specifically, Al_2O_3 and Fe_2O_3 gradually increased from the bottom up, whereas SiO_2 displayed the opposite trend.

2.1.3 VNIR spectra of basalt weathering profiles

The VNIR spectra were measured using an Analytical Spectral Devices (ASD) FieldSpec-3 spectroradiometer. The ASD spectrometer covers a spectral range of 350–2500 nm and offers a spectral resolution of 3 nm at 350–1000 nm and 10 nm at 1000–2500 nm. The spectral sampling intervals were 1.4 nm at 350–1000 nm and 2.0 nm at 1000–2500 nm. The samples were illuminated by a quartz halogen lamp in a dark room under ambient conditions. The incidence and emergence angles were set as 30° and 0°, respectively. A standard white panel (Spectralon) was also measured to calibrate the spectral measurements. To improve the signal-to-noise ratio, 50 scans were acquired for each sample (Liu JC et al., 2021a). Ten bands around the 1000 nm artifact (from 996 to 1005 nm) were discarded because of the splice of two wavelength ranges. The VNIR absorptions in the basaltic weathering profile at Chengmai originate from overtones and combination tones of the fundamental stretching and bending vibrations of hydroxyl in the structure of clay minerals, as well as electron transitions of Fe in Fe oxides (e.g., hematite and goethite) and Fe clay minerals (e.g., nontronite; Liu JC et al., 2021a). To be specific, all samples exhibited absorptions at 1.4, 1.9, and 2.2 μm contributed by the OH^- and H_2O of kaolinite and nontronite. Additionally, the Fe^{3+}-related absorption feature around 0.9 μm appeared in all samples.

2.2 Methodology

2.2.1 Division of the calibration and validation sets

To provide unbiased mineralogical, chemical, and spectral data, we excluded five samples from granitic material (49–53 m). We adopted the systematic sampling strategy (Sun WC et al., 2018), which is more suitable than the random sampling manner for the case of a small calibration set. Specifically, for each chemical composition, samples were sorted in ascending order and then divided into calibration and validation sets in a ratio of 2:1. Starting from the second sample, validation samples were selected every two samples thereafter, with the remaining samples designated as calibration samples. The 37 samples were divided into 25 calibration and 12 validation samples, as shown in Figure 2. Generally, the calibration and validation sets for each chemical property exhibited a similar distribution in terms of mean, median, and standard deviation (Table 1). Samples with extreme values located at boundaries (e.g., Fe-rich basalt dike, low IOL and high SiO_2) are also denoted in the histograms.

2.2.2 Spectral pretreatments

The VNIR spectra were initially smoothed by the Savitzky–Golay filter with a window size of 11 bands and a polynomial order of 2. Subsequently, we used five popular spectral preprocessing techniques to evaluate the effects of different pretreatments on the prediction model. Specifically, we conducted continuum removal (Clark and Roush, 1984); reflectance transformed into apparent absorbance (Clark and Roush, 1984); discarding of the water absorption bands around 1400 nm (1300–1500 nm) and 1900 nm (1800–2000 nm); and application of the first and second derivative operations. Among these techniques, continuum removal is widely used to enhance the absorption features by eliminating the baseline representing the general trend of the spectrum, which can vary because of factors such as environmental conditions (Clark and Roush, 1984). As a proxy for absorption intensity, apparent absorbance was formulated as $-\log_{10}R$, where R indicates reflectance (Clark and Roush, 1984). This operation could reduce the nonlinearities and scattering effects (Gomez et al., 2008). The soil moisture and atmospheric water vapor significantly degrade the ~1400 and ~1900 nm absorption values in field and orbital acquired spectra. To explore the potential of the prediction model on possible field and orbital acquired hyperspectral data, we removed the water absorption bands around 1400 nm (1300–1500 nm) and 1900 nm (1800–2000 nm). The first derivative

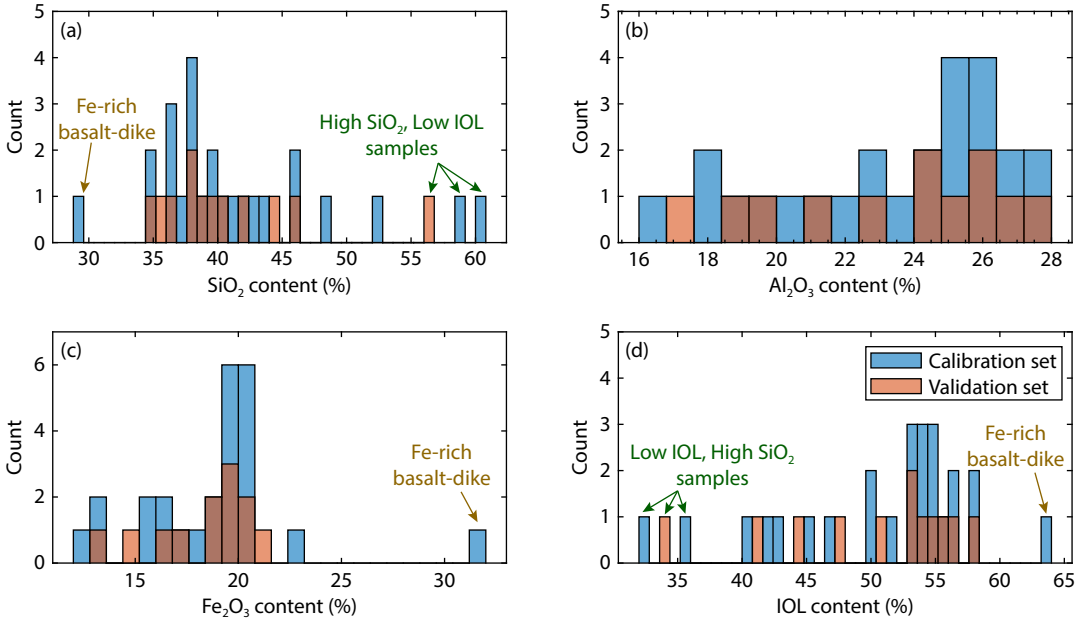

Figure 2. Histograms of the calibration and validation sets for (a) SiO_2, (b) Al_2O_3, (c) Fe_2O_3, and (d) the IOL. The Fe-rich basalt dike and the high SiO_2 and low IOL samples are denoted by brown and green arrows, respectively.

Table 1. Chemical concentrations for the Hainan weathered basalt samples.

Chemical composition	Min	Max	Mean	Median	Standard deviation
SiO_2 (wt%)					
All samples (n = 37)	29.55	60.06	41.24	39.43	6.75
Calibration set (n = 25)	29.55	60.06	41.43	39.43	7.19
Validation set (n = 12)	34.47	56.20	40.85	39.11	5.99
Al_2O_3 (wt%)					
All samples (n = 37)	16.21	27.69	23.36	24.35	3.30
Calibration set (n = 25)	16.21	27.69	23.41	24.35	3.35
Validation set (n = 12)	17.38	27.42	23.26	24.47	3.36
Fe_2O_3 (wt%)					
All samples (n = 37)	12.13	31.79	18.68	19.30	3.41
Calibration Set (n = 25)	12.13	31.79	18.83	19.30	3.80
Validation set (n = 12)	12.95	21.47	18.35	19.27	2.55
IOL					
All samples (n = 37)	32.69	63.27	50.57	53.33	7.19
Calibration set (n = 25)	32.69	63.27	50.77	53.33	7.34
Validation set (n = 12)	34.17	57.83	50.14	53.22	7.16

accentuates the rate of change in reflectance, which aids in identifying inflection points and slope changes. The second derivative, focusing on the curvature of the spectrum, is particularly effective in resolving overlapping absorption features and reducing background (Nawar et al., 2016). To simplify, we designated the reflectance spectra as Dataset 1, the continuum removal spectra as Dataset 2, the apparent absorbance data as Dataset 3, reflectance with the water absorption bands removed as Dataset 4, the first derivative spectra as Dataset 5, and the second derivative spectra as Dataset 6. The VNIR reflectance spectra of all samples and the associated preprocessing dataset are shown in Figure 3.

2.2.3 Model construction

In practice, highly redundant spectral bands often degrade the performance of multivariate calibration (Vohland et al., 2011). We adopted two steps to select informative bands to reduce model

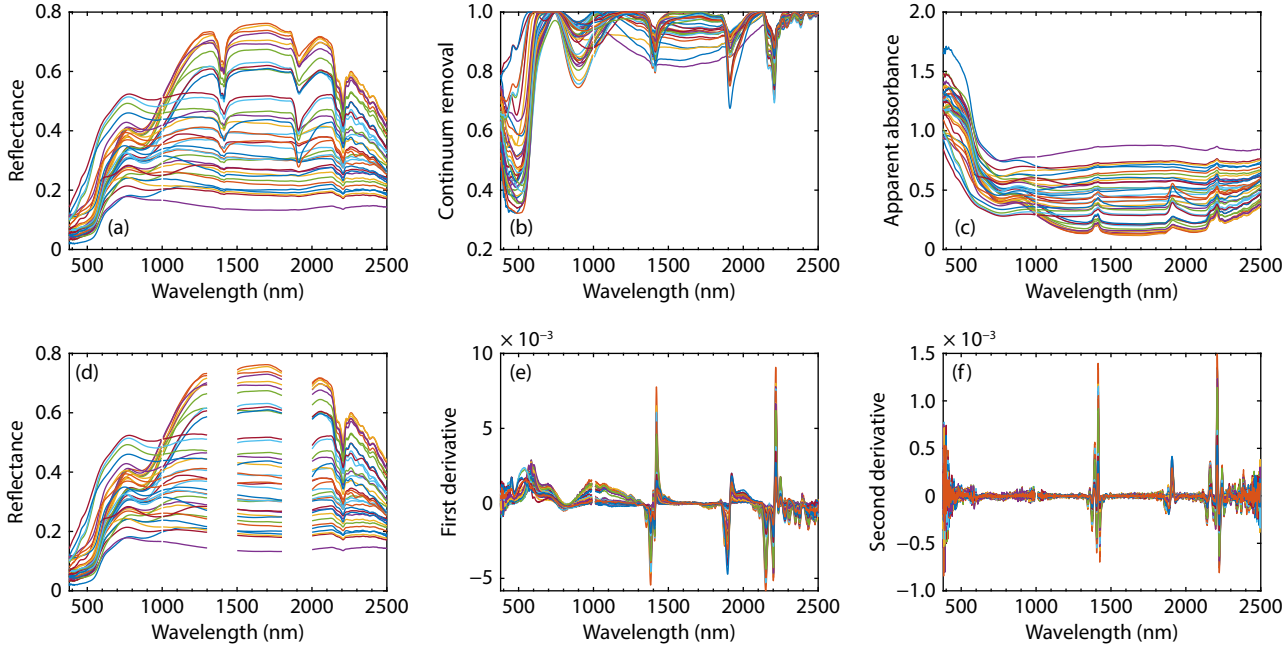

Figure 3. The VNIR reflectance spectra of all samples and the associated preprocessing dataset. (a) Reflectance. (b) Spectra with continuum removed. (c) Apparent absorbance. (d) Reflectance with ~1400 and ~1900 nm absorption bands removed. (e) First derivative. (f) Second derivative.

complexity and improve the predictive ability. A Pearson correlation analysis between the spectral band and chemical concentration was first conducted to remove irrelevant spectral regions. The GA was then used to search for the optimized spectral subsets by mimicking the process of natural evolution (Durand et al., 2007). The GA consists of five steps, namely, the coding of variables, initiation of the population, evaluation of responses, reproduction, and mutation (Leardi and Lupiáñez González, 1998). In this work, we used the GA toolbox released by the University of Sheffield (Sheffield, UK; https://uos-codem.github.io/GA-Toolbox/).

The GA population comprises 20 randomly generated chromosomes, each serving as a binary-coded gene determining variable selection. The reproductive process generates a new population of chromosomes, forming the subsequent generation. The generation gap, representing the fractional difference between the new and old population sizes, is set at 90%. A gene mutation is used to create potential chromosomes, with a mutation rate fixed at 10% for the population (Sun WC et al., 2018, 2022a, b). Every chromosome was evaluated with a fitness function, which was the root mean square error (RMSE) of cross-validation in this work. Those individuals demonstrating a good fit within the function were incorporated into the population to replace less optimal counterparts. Consequently, the optimized spectral subsets were input into the PLSR for estimation of the chemical composition.

The PLSR addresses challenges related to high dimensionality and multicollinearity by utilizing statistical rotations. Unlike the approach of decomposing spectra into a set of eigenvectors and scores and conducting a regression with chemical attributes in distinct steps, the PLSR algorithm chooses to select successive orthogonal factors that maximize the covariance between predictor and response variables. The input parameters to perform the PLSR are the predictor variable (spectral data), response variable (chemical composition), and number of latent variables (LVs). The optimal number of LVs in the PLSR model is determined by the leave-one-out cross-validation method (Sun WC et al., 2018). Considering the dependence of the result of the GA on the randomly generated first generation, we ran the model 20 times independently. For each chemical property, the median prediction among the 20 individual runs analyzed for each dataset was illustrated with scatterplots. In addition, the statistics of model performance (e.g., R^2 and RMSE) across 20 runs were reported for comparison. The flowchart of the chemical concentration estimation with VNIR reflectance spectroscopy is illustrated in Figure 4.

2.2.4 Evaluation metrics

The goodness of fit (R^2), RMSE, and residual prediction deviation (RPD) are widely used metrics to evaluate the accuracy of a model (Nawar et al., 2016). The greater the R^2 and RPD, the smaller the RMSE in the validation set, indicating the model has higher prediction accuracy. The calculation formulas for the three metrics are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_m^i - y_p^i)^2}{\sum_{i=1}^{N}(y_m^i - \bar{y}_m)^2}, \quad (1)$$

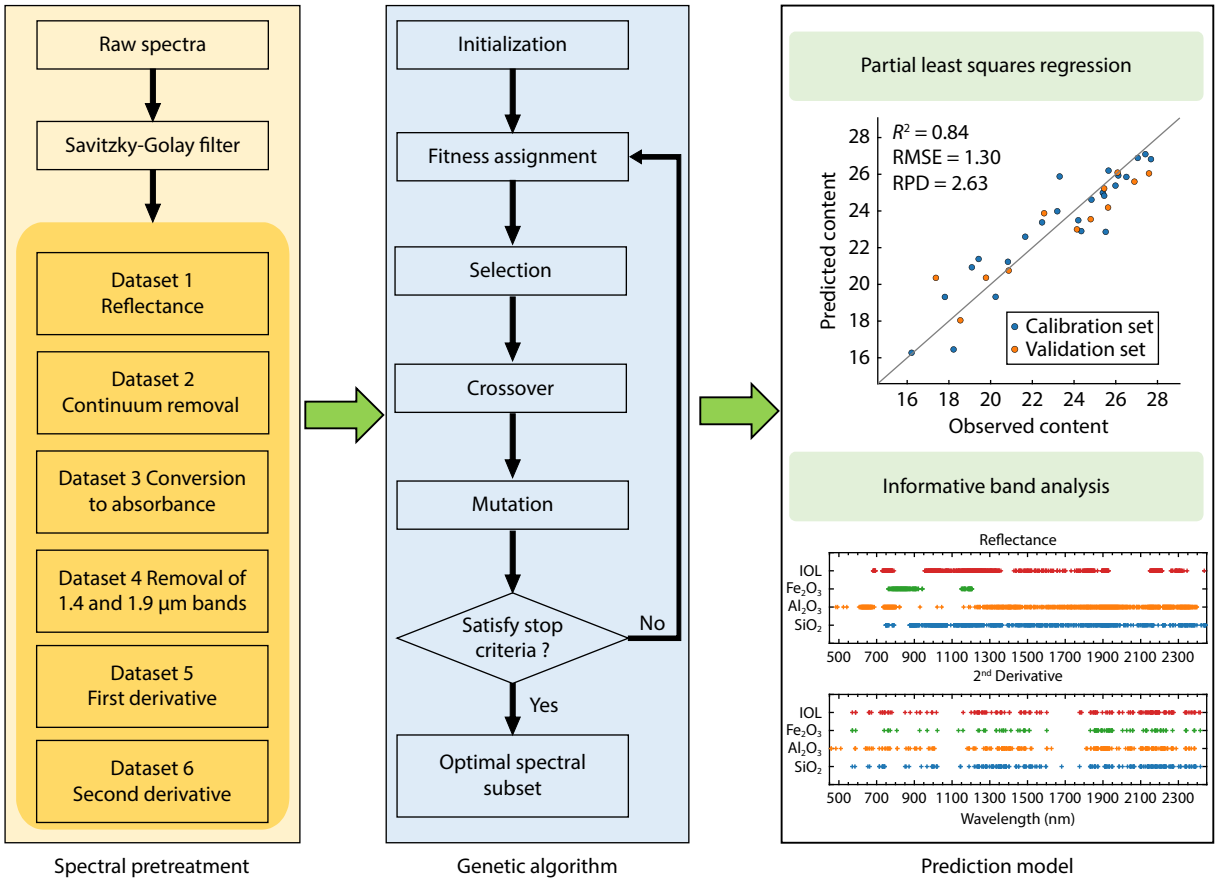

Figure 4. The flowchart of chemical concentration estimations with VNIR spectra.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_m^i - y_p^i)^2}{N}}, \tag{2}$$

$$\mathrm{RPD} = \frac{\mathrm{std}}{\mathrm{RMSE}}, \tag{3}$$

where y_m^i is the measured elemental content of the i-th sample; y_p^i is the predicted elemental content of the i-th sample; $\bar{y}_m$ is the average of the measured value; and N is the number of samples.

3. Results

3.1 Correlation Analysis

The Pearson correlation coefficient (r) between various chemical compositions was first analyzed. Overall, SiO_2 was negatively correlated with the other properties, with the largest r of −0.98 for the IOL. A weak correlation was found between Al_2O_3 and Fe_2O_3 (r = 0.42). The IOL was better correlated with Fe_2O_3 (r = 0.86) than with Al_2O_3 (r = 0.80). The correlation coefficients between major chemical properties and the preprocessed VNIR data are shown in Figure 5. The orange dots highlight bands significantly correlating with chemical properties at the calculated probability level less than 0.01 ($p < 0.01$) level, which were retained as inputs for the GA-PLSR model. For the reflectance data (Figure 5a), the concentration of SiO_2 had a high positive correlation with the reflectance in the shortwave infrared range. In contrast, the reflectance was negatively correlated with the Fe, Al, and IOL concentrations. The highly negative correlation between Fe and reflectance in the near-infrared range can be explained by the broad and strong absorption of Fe^{3+} between 750 and 1100 nm. For the spectra with the continuum removed (Figure 5b), the shape of the correlation coefficient curves for Al_2O_3, Fe_2O_3, and the IOL were similar in the shortwave infrared range. Both chemical properties were significantly correlated with absorption features around 1400, 1900, 2200, and 2300–2400 nm. In addition, Al_2O_3 and the IOL were positively correlated, whereas SiO_2 was negatively correlated with a broad absorption between 780 and 1000 nm. The correlation coefficients for the apparent absorbance data were expected to be reversed in relation to those of the reflectance data because the absorbance equals $-\log_{10}R$ (Figure 5c). The spectral derivatives increase correlations at some specific wavelengths. The first derivative highlighted continuous spectral regions (Figure 5d), whereas the second derivative accentuated isolated bands (Figure 5e).

3.2 Estimation of Chemical Concentrations

3.2.1 Estimation of SiO_2 concentrations

The model-predicted SiO_2 concentration was plotted against the measured concentration, as shown in Figure 6. The 1:1 line is shown in gray for reference. Evaluation metrics (R^2, RMSE, RPD) illustrating the prediction accuracy of the validation set are presented in the upper left. A marginal increase in R^2 values from 0.82 to 0.86 was observed with continuum removal in the reflectance dataset (Figures 6a and 6b), further reaching 0.89 with the transformation of the reflectance dataset into apparent absorbance (Figure 6c). A minimal impact on the prediction accuracy was observed with the removal of the water absorption bands (Figure 6d). The R^2 value increased to 0.91 with the first derivative dataset (Figure 6e) and reached 0.99 with the second derivative dataset (Figure 6f). The RMSE values declined with the use of the derivative dataset. The brown arrow denotes the Fe-rich basalt dike sample with extremely low SiO_2, which is slightly above the 1:1 line (Figure 6f). The experimental results suggest that the pretreatments, such as continuum removal, transforma-

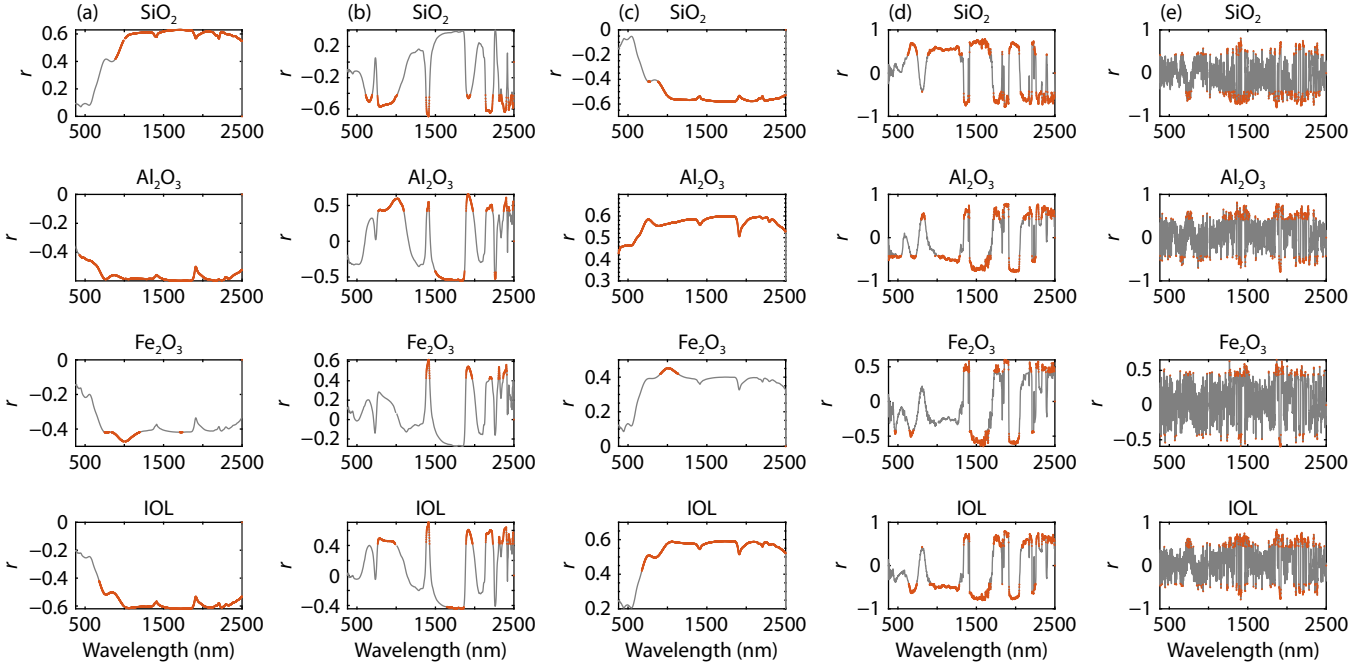

Figure 5. Pearson correlation coefficients (r) between chemical properties and (a) reflectance spectra, (b) spectra with the continuum removed, (c) apparent absorbance spectra, (d) first derivative spectra, and (e) second derivative spectra. The orange dots highlight bands significantly correlated with each composition at the $p < 0.01$ level.

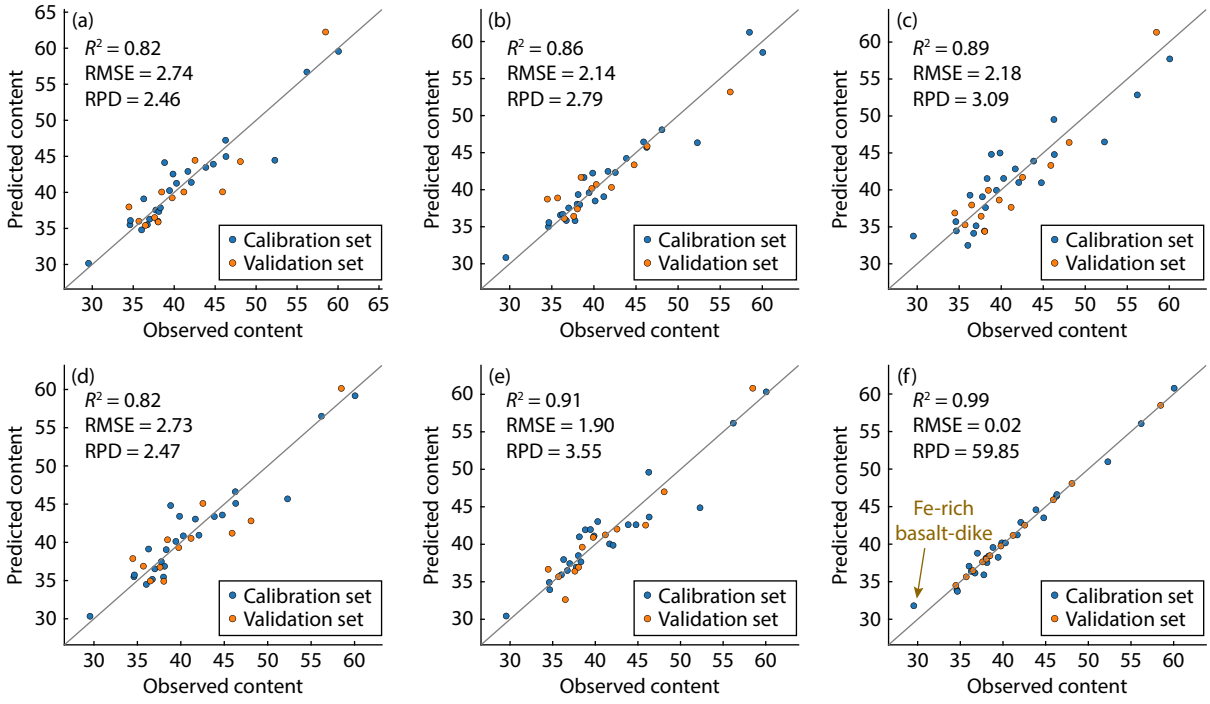

Figure 6. Scatterplots of the measured SiO_2 concentrations against the predicted concentrations using the (a) reflectance, (b) spectra with the continuum removed, (c) apparent absorbance, (d) reflectance with the water absorption bands removed, (e) first derivative, and (f) second derivative datasets, respectively. The 1:1 line is shown in gray for reference.

tion into apparent absorbance, and removal of the water absorption bands, had minimal effects on the prediction of the SiO_2 content.

3.2.2 Estimation of Al_2O_3 concentrations

Scatterplots of the measured Al_2O_3 concentrations against the predicted concentrations are shown in Figure 7. A sharp decline in the R^2 value from 0.84 to 0.68, accompanied by a reduction in the RPD value from 2.63 to 1.85 was observed with continuum removal in the reflectance dataset (Figures 7a and 7b). Minimal changes in prediction accuracy were noted for the apparent absorbance (Figure 7c) and water band removal datasets (Figure 7d). Notably, the R^2 underwent an increase from 0.84 to 0.97, coupled with a decrease in the RMSE from 1.30 to 0.57 for the first derivative dataset (Figure 7e). A marginal decrease in model performance occurred for the second derivative dataset (Figure 7f) when compared with the first derivative dataset. Specifically, both the R^2 and RPD decreased from 0.97 and 6.03 to 0.95 and 4.47, respectively, accompanied by an increase in the RMSE from 0.57 to 0.76.

3.2.3 Estimation of Fe_2O_3 concentrations

Scatterplots of the measured concentrations against the predicted Fe_2O_3 concentrations are shown in Figure 8. A marginal decline in prediction accuracy was observed upon removal of the continuum from the reflectance data (Figures 8a and 8b). The R^2 value increased from 0.61 to 0.81, coupled with an increase in the RPD value from 1.67 to 2.38, following the transformation of reflectance into apparent absorbance (Figure 8c). The R^2 value decreased from 0.61 to 0.54 when the water bands were removed (Figure 8d). Notably, the introduction of the first derivative (Figure 8e) and second derivative datasets (Figure 8f) resulted in substantial gains in the R^2 value, amounting to 0.37 and 0.38, respectively, when contrasted with the reflectance dataset. The brown arrow denotes the Fe-rich basalt dike sample, which deviates from the 1:1 line. Similarly, the spectral derivatives boosted the prediction accuracy for Fe_2O_3.

3.2.4 Estimation of IOL concentrations

Scatterplots of the measured against predicted IOL are shown in Figure 9. A noticeable decline of R^2 in prediction accuracy was observed, dropping from 0.87 to 0.78 when the continuum was removed from the reflectance (Figures 9a and 9b). Prediction accuracy marginally decreased when reflectance was converted to apparent absorbance (Figure 9c). Before and after removing the water absorption bands, the R^2, RPD, and RMSE values remained comparable (Figure 9d). Consistent with earlier experiments, the spectral derivative pretreatments significantly enhanced the performance of the model. For instance, the R^2 value grew from 0.87 (Figure 9a) to 0.96 (Figure 9e) and reached an impressive 0.99 with the second derivative dataset (Figure 9f), signifying its superiority among the pretreatments. The brown arrow denotes the Fe-rich basalt dike sample with the highest IOL value, which is below the 1:1 line.

4. Discussion

4.1 Effects of Spectral Pretreatments on Modeling

The initial assessment of how the spectral pretreatments affected the performance of the model is given in Section 3.2. Considering the dependence of the GA result on the randomly generated first generation, we ran the model 20 times independently to achieve

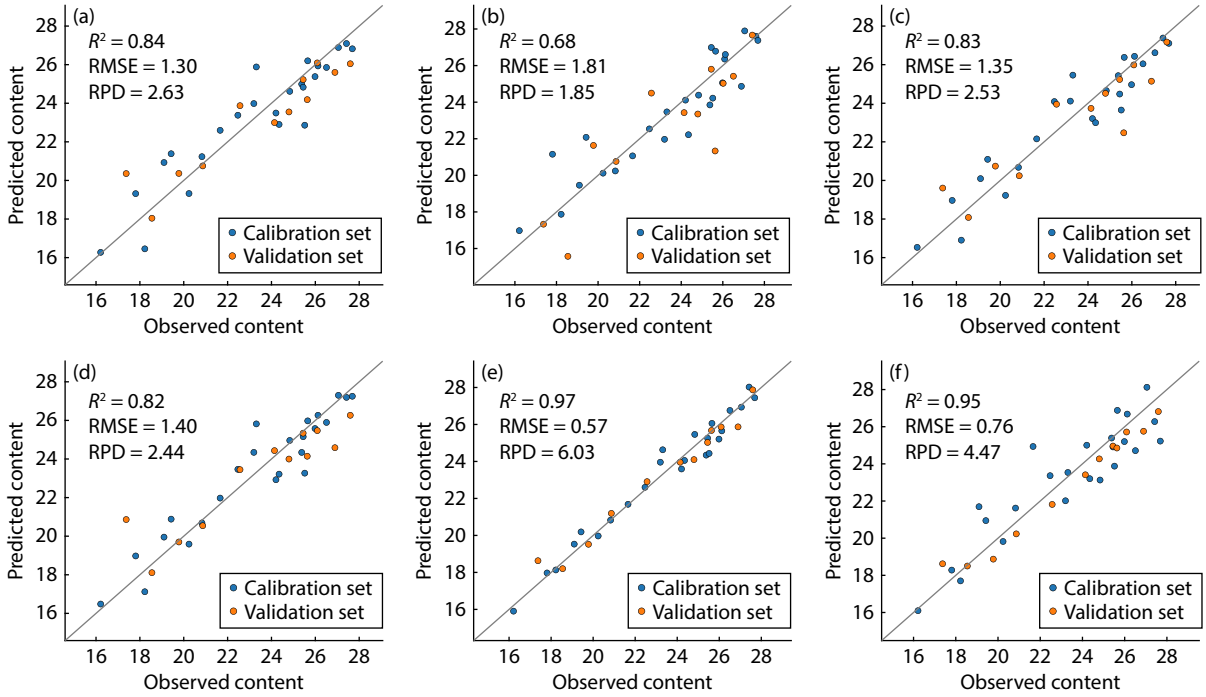

Figure 7. Scatterplots of the measured Al_2O_3 concentrations against the predicted concentrations using the (a) reflectance, (b) spectra with the continuum removed, (c) apparent absorbance, (d) reflectance with the water absorption bands removed, (e) first derivative, and (f) second derivative datasets, respectively. The 1:1 line is shown in gray for reference.

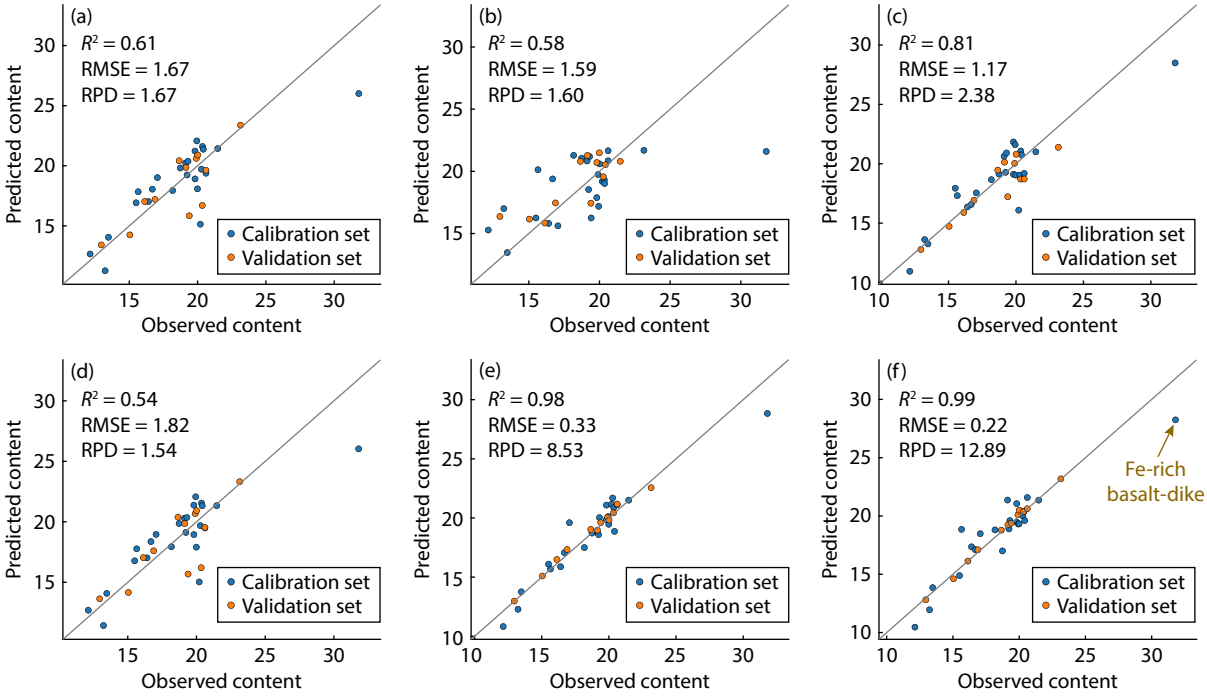

Figure 8. Scatterplots of the measured Fe_2O_3 concentrations against the predicted concentrations using (a) reflectance, (b) spectra with the continuum removed, (c) apparent absorbance, (d) reflectance with the water absorption bands removed, (e) first derivative, and (f) second derivative datasets, respectively. The 1:1 line is shown in gray for reference.

a robust evaluation. The best prediction results for each spectral dataset are tabulated in Table 2. Typically, a superior model exhibits a higher R^2 value and a lower RMSE value (Rossel and Behrens, 2010; Nawar et al., 2016). For visual comparison, the boxplots of R^2 and RMSE for different datasets are illustrated in Figure 10. The labels D1 to D6 on the *x*-axis denote Dataset 1 to Dataset 6, respectively. The upper and lower edges of the boxplot represent the 25th and 75th percentiles, respectively. A larger gap between these edges indicates more noticeable variations attributed to GA initialization. The central thick black line signifies

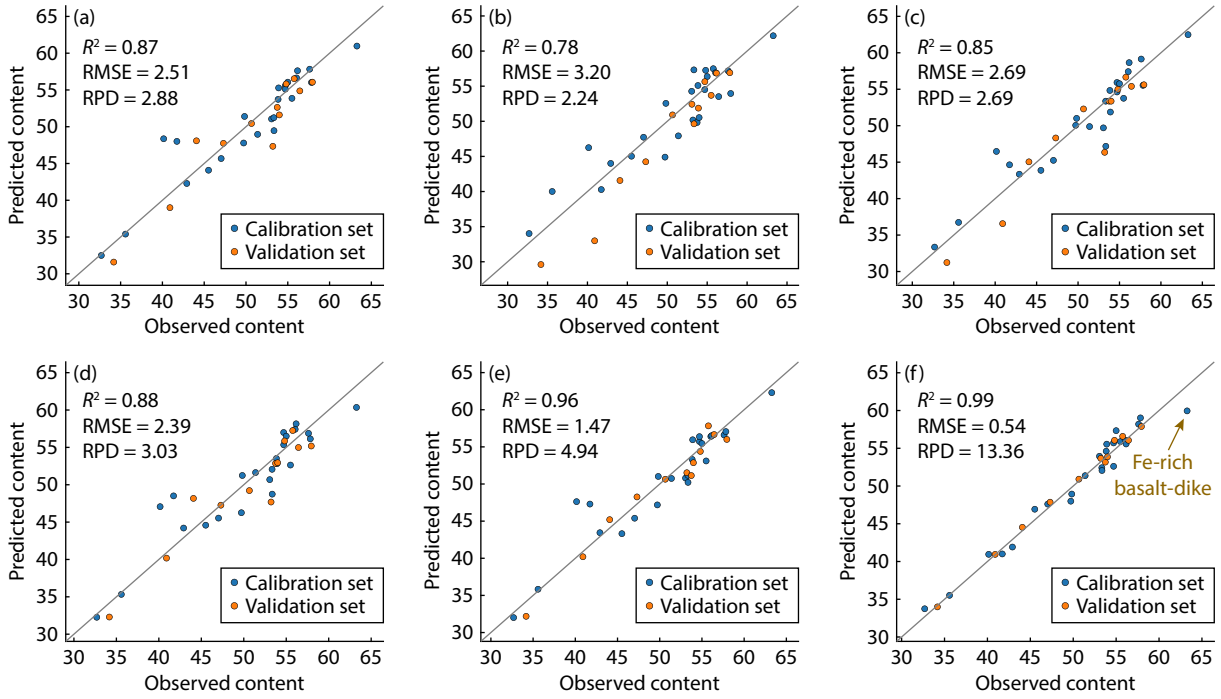

Figure 9. Scatterplots of the measured IOL concentrations against the predicted concentrations using (a) reflectance, (b) spectra with the continuum removed, (c) apparent absorbance, (d) reflectance with the water absorption bands removed, (e) first derivative, and (f) second derivative datasets, respectively. The 1:1 line is shown in gray for reference.

Table 2. Validation results of the GA-PLSR models for SiO_2, Al_2O_3, Fe_2O_3, and the IOL with various spectral datasets.[a]

Chemical composition	Metric	Spectral dataset					
		Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6
SiO_2	LVs	7	9	6	7	4	5
	R^2	0.856	0.874	0.899	0.838	0.931	**0.999**
	RMSE	2.443	2.028	2.043	2.592	1.705	0.004
Al_2O_3	LVs	6	7	8	7	6	7
	R^2	0.856	0.741	0.889	0.828	0.983	**1.000**
	RMSE	1.239	1.635	1.090	1.356	0.417	0.003
Fe_2O_3	LVs	7	8	6	7	6	6
	R^2	0.746	0.910	0.847	0.724	0.995	**0.999**
	RMSE	1.347	0.732	1.044	1.404	0.183	0.011
IOL	LVs	6	6	7	6	3	6
	R^2	0.878	0.856	0.857	0.888	0.962	**1.000**
	RMSE	2.415	2.596	2.618	2.317	1.342	0.020

[a]The highest R^2 values are emphasized in bold.

the median, whereas outliers are individually marked with solid circles.

Several interesting phenomena were observed, as shown in Figure 10:

(1) Compared with results for the original reflectance dataset, continuum removal did not improve the model accuracy except for SiO_2.

(2) Transforming reflectance into apparent absorbance improved the prediction accuracy for SiO_2 and Fe_2O_3.

(3) The prediction accuracy of SiO_2, Al_2O_3, and the IOL showed subtle variations by removing the water absorption bands. This pretreatment decreased the prediction accuracy for Al_2O_3 but enhanced the IOL prediction model.

(4) Generally, the spectral derivative treatment rapidly improved the prediction accuracy of all the chemical properties (Nawar et al., 2016; Hong YS et al., 2019). The enhanced prediction accuracy achieved with the second derivative dataset could be ascribed to two effects. First, the second derivative treatment isolated over-

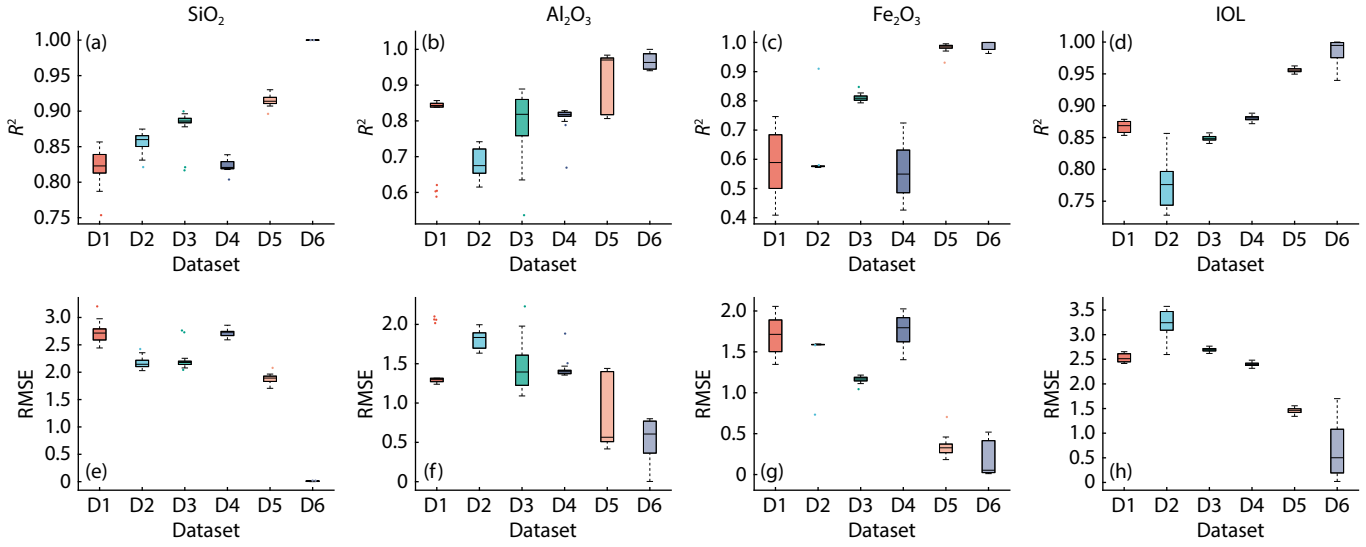

Figure 10. Boxplots of R^2 for (a) SiO_2, (b) Al_2O_3, (c) Fe_2O_3, and (d) the IOL, and boxplots of the RMSE for (e) SiO_2, (f) Al_2O_3, (g) Fe_2O_3, and (h) the IOL. The top and bottom edges of the boxplot denote the 25th and 75th percentiles. The central thick black line signifies the median. The labels D1 to D6 on the *x*-axis denote the reflectance (D1), spectra with the continuum removed (D2), apparent absorbance (D3), reflectance with the water bands (1300–1500 nm and 1800–2000 nm) removed (D4), first derivative (D5), and second derivative (D6) datasets, respectively.

lapping absorptions and amplified subtle signals (Tan W et al., 2022). Second, using fewer bands reduced the model complexity and thus avoided the issue of overfitting.

(5) Slight variations were observed in boxplots for SiO_2 (Figures 10a and 10e) and the IOL (Figures 10d and 10h), suggesting that the random initialization of the GA had a minimal influence on prediction of the SiO_2 and IOL contents.

(6) The relatively poor prediction outcome for the Fe-rich dike sample could be attributed to its chemical composition deviating significantly from the distribution of other samples (Figure 2). Consequently, the model performance may have degraded for samples whose chemical compositions were far from the range covered by weathered basalts in this work.

4.2 Informative Band Analysis

The sensitive bands for major oxides and the IOL were initially evaluated with the Pearson correlation coefficient, as described in Section 3.1. In the first three datasets, the bands that exhibited significant correlations generally aligned with the diagnostic absorption values of ferric and clay minerals in the drilling samples (Figures 5a–5c). The application of spectral derivatives enhanced correlations at specific wavelengths. The first derivative highlighted continuous spectral regions around 650–750 nm, 900–1200 nm, 1350–1650 nm, 1850–2050 nm, and 2100–2500 nm (Figure 5d). In contrast, the second derivative accentuated isolated bands across the entire spectral range (Figure 5e).

The GA was used to further refine the selection of optimal spectral subsets. The selected subsets for each chemical composition were recorded for every individual run. If a specific wavelength was chosen in more than half of the individual runs (10 times in this study), it was considered an informative band. Figure 11 depicts plots of informative bands identified by the GA from both the original reflectance (Figure 11a) and the second derivative datasets (Figure 11b). Crosses denote important bands related to different chemical properties, and densely populated cross regions indicate frequent selection in model construction. For the reflectance dataset, the selected spectral subsets were consistent with the absorption of dominant minerals in the drilling samples. Notably, important wavelengths for Al_2O_3 and SiO_2 overlapped at approximately 1400, 1900, 2200, and 2300 nm. For Fe_2O_3, informative bands spanned the range from 750 to 950 nm and 1150 to 1200 nm. Informative bands for the IOL spanned multiple wavelength regions because the IOL was determined by the other three oxides. The derived informative bands have the potential to guide the design of spectral channels (e.g., the band center and bandwidth) for future spectrometers.

The number of informative bands significantly decreased when using the second derivative dataset. The frequently selected bands from the second derivative dataset were sparsely distributed across the entire wavelength range, except for the regions of 1000–1150 nm and 1600–1800 nm (Figure 11b). For instance, the inflection points and absorption values related to Fe^{3+} were located at ~530, ~600, ~900, ~2290, and ~2350 nm. Bands related to Al-OH occurred at ~1390, ~1410, ~1910, ~2165, and ~2207 nm owing to the dominance of kaolinite in the weathering products (Liu JC et al., 2021a). Nonetheless, the physical interpretation of some selected bands remained challenging. Further work is required to elucidate the isolated features in the derivative spectra.

4.3 Implications for Mars

The composition and thickness of basaltic weathering profiles are largely influenced by climate and time (Carter et al., 2015; Ye BL and Michalski, 2022). The removal of mobile elements and the preservation of immobile elements lead to significant variations in weathering products. Chemical weathering in reducing or anoxic conditions mobilizes Fe^{2+}, leaving a persistent geological record of Fe loss (Liu JC et al., 2021b). The proposed spectral model accurately predicts the contents of major oxides (i.e., SiO_2, Fe_2O_3, and Al_2O_3) and the IOL in Hainan weathered products. The derived

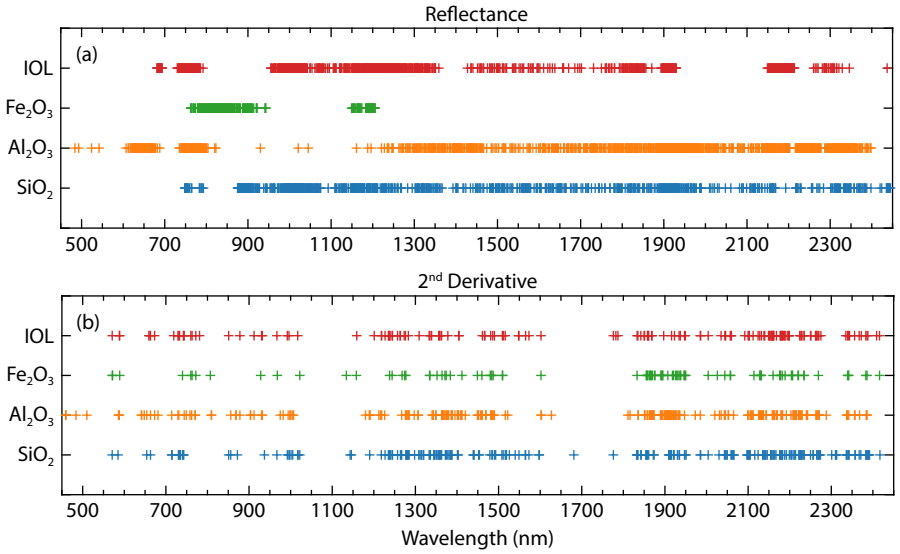

Figure 11. Plots of informative wavelengths based on the GA over 20 individual runs. Crosses denote informative bands related to the chemical properties. (a) Reflectance dataset. (b) Second derivative dataset.

concentrations of Fe_2O_3 and Al_2O_3 exhibit the separation of Fe from Al along the weathering profile, resembling the mineralogical and spectral trends observed in weathered basalts on Mars (Liu JC et al., 2021b). Consequently, monitoring their distributions emerges as a valuable tool for constraining paleoclimates and the redox state of early Mars.

At present, there is an absence of a dedicated payload tailored to measuring chemical compositions within Martian weathering profiles. The proposed model offers an alternative and promising method for assessing chemical concentrations on Mars by using VNIR spectra. To assess the potential for Mars application, we used the JSC Mars-1 soil simulant, chosen for its spectral resemblance to the Martian surface and its well-established analog status (Allen et al. 1998). This selection was also driven by the lack of Martian basalt weathering product samples or corresponding simulants. According to studies from the Viking and Mars Pathfinder missions, JSC Mars-1 closely mimics Martian regolith and is ideal for testing the transferability of the model. The spectrum of JSC Mars-1 exhibits a relatively featureless ferric absorption edge in the visible region, a ferric absorption band around 800–1000 nm, and OH^-- and H_2O-related absorption values at 1400 nm and 1900 nm (Figure 12). The predictions for chemical compositions are shown in the inset table in Figure 12. The discrepancies between the predicted and observed contents were 0.71, −4.56, and −2.73 for SiO_2, Al_2O_3, and Fe_2O_3, respectively. The predicted content for SiO_2 was the most accurate, followed by Fe_2O_3, whereas Al_2O_3 was the least accurately predicted. The poor accu-

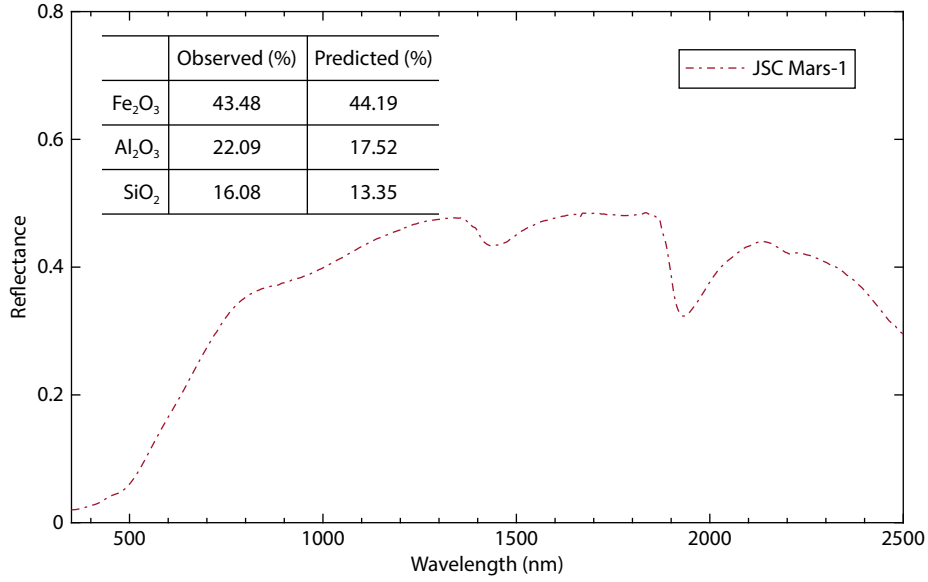

Figure 12. The reflectance of JSC Mars-1, obtained from the Reflectance Experiment Laboratory, Brown University (Providence, Rhode Island, USA). The spectral ID is cjcc18. The inset table shows the observed and predicted chemical concentrations. The composition of JSC Mars-1 is from Allen et al. (1998).

racy for Al_2O_3 could be attributed to the limited chemical alteration in Martian soil when compared with moderately to highly weathered Hainan basalts. It was also evidenced by the absence of the ~2200 nm asymmetry absorption (Figure 12) attributable to the weathering product kaolinite, which is dominant in the Hainan basalts.

Therefore, it is essential to acknowledge the complexities associated with the diverse compositions and weathering processes on Mars. Additionally, the presence of atmospheric correction residuals and instrumental artifacts could impair the model performance. Several fundamental issues should be addressed before the model can be confidently applied on Mars:

(1) Acquiring more representative weathered basalt products. Our model was established by using moderately and highly weathered basalts (20 < IOL < 60) from Hainan Island. To enhance the applicability of the model, it is essential to acquire a more diverse set of weathered basalt samples. This set should include samples from various geological and climatic regions, as well as those exhibiting different degrees of weathering. Expanding the sample diversity will improve the ability of the model to transfer across various geological settings.

(2) Addressing spectral variability. Spectral variability arises from intrinsic factors such as the grain size and crystalline properties of minerals, whereas additional variations stem from imaging conditions, calibration methods, the surface relief, and spatial resolutions (Wu X et al., 2021). Using the laboratory spectral model directly on orbital data may yield unrealistic results (Sun WC et al., 2022a, b). To mitigate this issue, it is imperative to develop advanced techniques for atmospheric correction and spectral denoising (Itoh and Parente, 2021). These techniques will help reduce spectral discrepancies between laboratory and orbital data, thus enhancing the robustness and accuracy of the model when applied on Mars.

5. Conclusions

In this study, we developed a quantitative spectral model (GA-PLSR) to predict the chemical concentrations along a 53-m-long basaltic weathering profile. We assessed the impact of various spectral pretreatments on the model accuracy. Notably, spectral derivatives significantly enhanced the performance of the model in predicting chemical properties. In addition, we conducted an analysis of informative bands based on their selection frequency, which directly informed the designation of spectrometer channels. In summary, the GA-PLSR model exhibited superior performance when utilizing the second derivative dataset.

The inclusion of weathering samples from diverse regions and of various weathering degrees could further enhance the applicability of the model across different geological settings. The proposed model has the potential to estimate chemical concentrations in basaltic weathering profiles on Mars, thereby enabling assessment of the degree of alteration and offering valuable insights into paleoclimatic conditions. Using the proposed model to predict chemical concentrations from *in situ* and orbital data will be the focus of future research.

Acknowledgments

This work was supported by the National Key Research and Development Project (Grant No. 2019YFE0123300), the National Natural Science Foundation of China (Grant Nos. 42072337, 42241111, and 42241129), and Pandeng Program of National Space Science Center, Chinese Academy of Sciences. Xing Wu also acknowledges support from the Young Elite Scientists Sponsorship Program by the China Association for Science and Technology (Grant No. 2022QNRC001) and the China Postdoctoral Science Foundation (Grant No. 2021M700149).

References

Allen, C. C., Morris, R. V., Jager, K. M., Golden, D. C., Lindstrom, D. J., Lindstrom, M. M., and Lockwood, J. P. (1998). Martian regolith simulant JSC Mars-1. In *29th Annual Lunar and Planetary Science Conference*. Houston, abstract no. 1690.

Broz, A. P., Clark, J., Sutter, B., Ming, D. W., Tu, V., Horgan, B., and Silva, L. C. R. (2022). Mineralogy and diagenesis of Mars-analog paleosols from eastern Oregon, USA. *Icarus*, *380*, 114965. https://doi.org/10.1016/j.icarus.2022.114965

Carter, J., Loizeau, D., Mangold, N., Poulet, F., and Bibring, J. P. (2015). Widespread surface weathering on early Mars: A case for a warmer and wetter climate. *Icarus*, *248*, 373–382. https://doi.org/10.1016/j.icarus.2014.11.011

Clark, R. N., and Roush, T. L. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *J. Geophys. Res.: Solid Earth*, *89*(B7), 6329–6340. https://doi.org/10.1029/JB089iB07p06329

Du, P. X., Yuan, P., Liu, J. C., and Ye, B. L. (2023). Clay minerals on Mars: An up-to-date review with future perspectives. *Earth-Sci. Rev.*, *243*, 104491. https://doi.org/10.1016/j.earscirev.2023.104491

Durand, A., Devos, O., Ruckebusch, C., and Huvenne, J. P. (2007). Genetic algorithm optimisation combined with partial least squares regression and mutual information variable selection procedures in near-infrared quantitative analysis of cotton-viscose textiles. *Anal. Chim. Acta*, *595*(1-2), 72–79. https://doi.org/10.1016/j.aca.2007.03.024

Ehlmann, B. L., Mustard, J. F., Fassett, C. I., Schon, S. C., Head, J. W., III, Des Marais, D. J., Grant, J. A., and Murchie, S. L. (2008). Clay minerals in delta deposits and organic preservation potential on Mars. *Nat. Geosci.*, *1*(6), 355–358. https://doi.org/10.1038/ngeo207

Ehlmann, B. L., Mustard, J. F., Swayze, G. A., Clark, R. N., Bishop, J. L., Poulet, F., Des Marais, D. J., Roach, L. H., Milliken, R. E., … Murchie, S. L. (2009). Identification of hydrated silicate minerals on Mars using MRO-CRISM: Geologic context near Nili Fossae and implications for aqueous alteration. *J. Geophys. Res.: Planets*, *114*(e2), E00D08. https://doi.org/10.1029/2009JE003339

Ehlmann, B. L., Bish, D. L., Ruff, S. W., and Mustard, J. F. (2012). Mineralogy and chemistry of altered Icelandic basalts: Application to clay mineral detection and understanding aqueous environments on Mars. *J. Geophys. Res.: Planets*, *117*(E11), E00J16. https://doi.org/10.1029/2012JE004156

Fassett, C. I., and Head, J. W. (2008). Valley network-fed, open-basin lakes on Mars: Distribution and implications for Noachian surface and subsurface hydrology. *Icarus*, *198*(1), 37–56. https://doi.org/10.1016/j.icarus.2008.06.016

Gaudin, A., Dehouck, E., and Mangold, N. (2011). Evidence for weathering on early Mars from a comparison with terrestrial weathering profiles. *Icarus*, *216*(1), 257–268. https://doi.org/10.1016/j.icarus.2011.09.004

Gomez, C., Lagacherie, P., and Coulouma, G. (2008). Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. *Geoderma*, *148*(2), 141–148. https://doi.org/10.1016/j.geoderma.2008.09.016

Goudge, T. A., Aureli, K. L., Head, J. W., Fassett, C. I., and Mustard, J. F. (2015). Classification and analysis of candidate impact crater-hosted closed-basin lakes on Mars. *Icarus*, *260*, 346–367. https://doi.org/10.1016/j.icarus.2015.07.

026

He, Y., Li, D. C., Velde, B., Yang, Y. F., Huang, C. M., Gong, Z. T., and Zhang, G. L. (2008). Clay minerals in a soil chronosequence derived from basalt on Hainan Island, China and its implication for pedogenesis. *Geoderma*, *148*(2), 206–212. https://doi.org/10.1016/j.geoderma.2008.10.007

Hong, Y. S., Chen, S. C., Liu, Y. L., Zhang, Y., Yu, L., Chen, Y. Y., Liu, Y. F., Cheng, H., and Liu, Y. (2019). Combination of fractional order derivative and memory-based learning algorithm to improve the estimation accuracy of soil organic matter by visible and near-infrared spectroscopy. *CATENA*, *174*, 104–116. https://doi.org/10.1016/j.catena.2018.10.051

Hynek, B. M., Beach, M., and Hoke, M. R. T. (2010). Updated global map of Martian valley networks and implications for climate and hydrologic processes. *J. Geophys. Res.: Planets*, *115*(E9), E09008. https://doi.org/10.1029/2009JE003548

Itoh, Y., and Parente, M. (2021). A new method for atmospheric correction and de-noising of CRISM hyperspectral data. *Icarus*, *354*, 114024. https://doi.org/10.1016/j.icarus.2020.114024

Leardi, R., and González, A. L. (1998). Genetic algorithms applied to feature selection in PLS regression: How and when to use them. *Chemometr. Intell. Lab. Syst.*, *41*(2), 195–207. https://doi.org/10.1016/S0169-7439(98)00051-3

Liu, J. C., He, H. P., Michalski, J., Cuadros, J., Yao, Y. Z., Tan, W., Qin, X. R., Li, S. Y., and Wei, G. J. (2021a). Reflectance spectroscopy applied to clay mineralogy and alteration intensity of a thick basaltic weathering sequence in Hainan Island, South China. *Appl. Clay Sci.*, *201*, 105923. https://doi.org/10.1016/j.clay.2020.105923

Liu, J. C., Michalski, J. R., Tan, W., He, H., Ye, B., and Xiao, L. (2021b). Anoxic chemical weathering under a reducing greenhouse on early Mars. *Nat. Astron.*, *5*(5), 503–509. https://doi.org/10.1038/s41550-021-01303-5

Liu, Z. F., Ma, J. L., Wei, G. J., Liu, Q. S., Jiang, Z. X., Ding, X., Peng, S. S., Zeng, T., and Ouyang, T. P. (2017). Magnetism of a red soil core derived from basalt, northern Hainan Island, China: Volcanic ash versus pedogenesis. *J. Geophys. Res.: Solid Earth*, *122*(3), 1677–1696. https://doi.org/10.1002/2016JB013834

McSween, H. Y., Jr., Taylor, G. J., and Wyatt, M. B. (2009). Elemental composition of the Martian crust. *Science*, *324*(5928), 736–739. https://doi.org/10.1126/science.1165871

Michalski, J. R., Goudge, T. A., Crowe, S. A., Cuadros, J., Mustard, J. F., and Johnson, S. S. (2022). Geological diversity and microbiological potential of lakes on Mars. *Nat. Astron.*, *6*(10), 1133–1141. https://doi.org/10.1038/s41550-022-01743-7

Mustard, J. F., Murchie, S. L., Pelkey, S. M., Ehlmann, B. L., Milliken, R. E., Grant, J. A., Bibring, J. P., Poulet, F., Bishop, J., … Wolff, M. (2008). Hydrated silicate minerals on Mars observed by the Mars Reconnaissance Orbiter CRISM instrument. *Nature*, *454*(7202), 305–309. https://doi.org/10.1038/nature07097

Nawar, S., Buddenbaum, H., Hill, J., Kozak, J., and Mouazen, A. M. (2016). Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. *Soil Tillage Res.*, *155*, 510–522. https://doi.org/10.1016/j.still.2015.07.021

Poulet, F., Bibring, J. P., Mustard, J. F., Gendrin, A., Mangold, N., Langevin, Y., Arvidson, R. E., Gondet, B., and Gomez, C. (2005). Phyllosilicates on Mars and implications for early martian climate. *Nature*, *438*(7068), 623–627. https://doi.org/10.1038/nature04274

Rossel, R. A. V., and Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, *158*(1-2), 46–54. https://doi.org/10.1016/j.geoderma.2009.12.025

Sheldon, N. D., and Tabor, N. J. (2009). Quantitative paleoenvironmental and paleoclimatic reconstruction using paleosols. *Earth-Sci. Rev.*, *95*(1-2), 1–52. https://doi.org/10.1016/j.earscirev.2009.03.004

Sun, W. C., Zhang, X., Sun, X. J., Sun, Y. L., and Cen, Y. (2018). Predicting nickel concentration in soil using reflectance spectroscopy associated with organic matter and clay minerals. *Geoderma*, *327*, 25–35. https://doi.org/10.1016/j.geoderma.2018.04.019

Sun, W. C., Liu, S., Zhang, X., and Li, Y. (2022a). Estimation of soil organic matter content using selected spectral subset of hyperspectral data. *Geoderma*, *409*, 115653. https://doi.org/10.1016/j.geoderma.2021.115653

Sun, W. C., Liu, S., Zhang, X., and Zhu, H. T. (2022b). Performance of hyperspectral data in predicting and mapping zinc concentration in soil. *Sci. Total Environ.*, *824*, 153766. https://doi.org/10.1016/j.scitotenv.2022.153766

Tan, W., Qin, X. R., Liu, J. C., Zhou, M. F., He, H. P., Wang, C. Y., Huang, J., Zhu, J. X., Yao, Y. Z., and Cudahy, T. (2022). Feasibility of visible short-wave infrared reflectance spectroscopy to characterize regolith-hosted rare earth element mineralization. *Econom. Geol.*, *117*(2), 495–508. https://doi.org/10.5382/econgeo.4877

Vohland, M., Besold, J., Hill, J., and Fründ, H. C. (2011). Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy. *Geoderma*, *166*(1), 198–205. https://doi.org/10.1016/j.geoderma.2011.08.001

Wang, J. J., Cui, L. J., Gao, W. X., Shi, T. Z., Chen, Y. Y., and Gao, Y. (2014). Prediction of low heavy metal concentrations in agricultural soils using visible and near-infrared reflectance spectroscopy. *Geoderma*, *216*, 1–9. https://doi.org/10.1016/j.geoderma.2013.10.024

Wang, P., Du, Y. S., Yu, W. C., Algeo, T. J., Zhou, Q., Xu, Y., Qi, L., Yuan, L. J., and Pan, W. (2020). The chemical index of alteration (CIA) as a proxy for climate change during glacial–interglacial transitions in Earth history. *Earth-Sci. Rev.*, *201*, 103032. https://doi.org/10.1016/j.earscirev.2019.103032

Wu, X., Mustard, J. F., Tarnas, J. D., Zhang, X., Das, E., and Liu, Y. (2021). Imaging Mars analog minerals' reflectance spectra and testing mineral detection algorithms. *Icarus*, *369*, 114644. https://doi.org/10.1016/j.icarus.2021.114644

Wu, Y. Z., Chen, J., Ji, J. F., Gong, P., Liao, Q. L., Tian, Q. J., and Ma, H. R. (2007). A mechanism study of reflectance spectroscopy for investigating heavy metals in soils. *Soil Sci. Soc. Am. J.*, *71*, 918–926. https://doi.org/10.2136/sssaj2006.0285

Ye, B. L., and Michalski, J. R. (2022). Chemical weathering over hundreds of millions of years of greenhouse conditions on Mars. *Commun. Earth Environ.*, *3*, 266. https://doi.org/10.1038/s43247-022-00602-7

Zhao, L. L., Hong, H. J., Liu, J. C., Fang, Q., Yao, Y. Z., Tan, W., Yin, K., Wang, C. W., Chen, M., and Algeo, T. J. (2018). Assessing the utility of visible-to-shortwave infrared reflectance spectroscopy for analysis of soil weathering intensity and paleoclimate reconstruction. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, *512*, 80–94. https://doi.org/10.1016/j.palaeo.2017.07.007