Article

Mineral Content Estimation of Lunar Soil of Lunar Highland and Lunar Mare Based on Diagnostic Spectral Characteristic and Partial Least Squares Method

Ping Zhou ¹ , Zhe Zhao ², Guangyuan Wei ³ and Hongyuan Huo ^{4,*}

¹ School of Geosciences and Resources, China University of Geosciences (Beijing), Beijing 100083, China; zhoupx@cugb.edu.cn

² Hebei Bureau of Coal Geological Exploration, Shijiazhuang 072750, China; zhaozhe@hebmt.gov.cn

³ School of Geosciences, University of Edinburgh, King's Buildings, James Hutton Rd., Edinburgh EH9 3FE, UK; 2418990@dundee.ac.uk

⁴ Faculty of Architecture, Civil and Transportation Engineering, Beijing University of Technology, Beijing 100021, China

* Correspondence: huohongyuan@bjut.edu.cn

Abstract: Extraction of mineral and rock information of lunar regolith is of far-reaching significance to the study of material composition, geological structure and historical evolution of lunar regolith. Visible and near-infrared spectra can reflect mineral composition information, and can be used to extract mineral composition and distribution characteristics of lunar regolith. In this paper, the LSCC (Lunar Soil Characterization Consortium) data of lunar regolith is taken as the research object. The partial least squares (PLS) regression model is used to estimate the spectra of lunar regolith measured in RELAB laboratory of Brown University. The mineral contents of plagioclase, pyroxene, olivine, ilmenite, agglutinate and volcanic glass in lunar regolith have been optimized and retrieved. The LSCC spectra of lunar regolith have been pre-processed by multivariate scattering correction (MSC), which highlight the spectral features of lunar regolith. The optimal number of principal components has been selected by cross-validation test. The PLS regression have been established for samples from lunar highland and lunar mare respectively. Two-thirds of samples have been randomly selected as experimental group to establish the prediction relationship between the spectra of lunar regolith and mineral content. The remaining one-third of samples have been used as verification group to further validate the prediction relationship. The results show that the partial least squares regression model has high accuracy and good stability. It is of theoretical and practical significance to optimize the inversion of mineral content in lunar regolith using spectral data of lunar regolith.

Keywords: partial least squares method; lunar regolith; spectra; mineral content estimation

Citation: Zhou, P.; Zhao, Z.; Wei, G.; Huo, H. Mineral Content Estimation of Lunar Soil of Lunar Highland and Lunar Mare Based on Diagnostic Spectral Characteristic and Partial Least Squares Method. *Appl. Sci.* **2022**, *12*, 1197. <https://doi.org/10.3390/app12031197>

Academic Editors: Mario De Cesare and Yangquan Chen

Received: 26 November 2021

Accepted: 20 January 2022

Published: 24 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Lunar exploration, especially the exploration and exploitation of lunar resources, is a key step to further explore the planets of the solar system, such as Mars. It has epoch-making significance in the history of human history and the development of science and technology. The identification and exploration of mineral resources of lunar regolith, especially the extraction of mineral information from lunar regolith by various technical means, is of far-reaching significance to the study of material composition, geological structure and historical evolution of lunar regolith [1–4].

Determining the chemical and mineral composition of lunar regolith is the basis for understanding the origin and geological evolution of the moon [5]. Compared with the earth, the mineral composition of lunar regolith is simple, mainly composed of plagioclase, clinopyroxene, orthopyroxene, olivine, ilmenite, agglutinate and volcanic glass [6]. Visible and near infrared spectra are sensitive to the known lunar regolith mineral compositions

and can be used to extract the mineral composition and distribution characteristics of lunar regolith. Visible and near infrared light can reflect the composition information of minerals [7]. In the process of interaction between photons and mineral surfaces, they are either reflected or absorbed. Different wavelengths and crystal structures of minerals produce unique reflection spectra [7,8]. Plagioclase is widely distributed in the lunar sea and lunar highlands. Plagioclase in the lunar soil is poor in sodium and rich in calcium, and has relatively strong absorption near 0.65 μm and 1.25 μm , but it is not easy to observe in the lunar soil. This is because the absorption characteristics of plagioclase are weaker than other strong absorption minerals and can easily be concealed. Charge transfer between Fe^{2+} and Ti^{4+} in agglutinate and volcanic glass will lead to strong absorption in ultraviolet band (300–600 nm), and the shorter the wavelength, the stronger the absorption, which will make the spectra become more “red” (spectral reflectance increases with wavelength from ultraviolet to visible light). Fe^{2+} in agglutinate and volcanic glass will also lead to absorption features near 1000 nm. The increase of ilmenite will weaken the “red” of the spectra and make the absorption features weaker and darker.

Optical remote sensing is the most effective method to obtain the chemical and mineral composition of lunar regolith by linking the spectral features and mineral compositions of the returned samples of lunar regolith [9]. The spectral features of lunar regolith, such as reflectance, absorption location and absorption depth, can be used in mineral identification. The spectra of Apollo and Luna samples of lunar regolith is the main basis for researching material composition and spectral mechanism of lunar regolith by optical remote sensing, which plays an important role in lunar exploration and research.

In this paper, the LSCC (Lunar Soil Characterization Consortium) data of lunar regolith is taken as the research object. The partial least squares (PLS) regression model is used to estimate the spectra of lunar regolith measured in RELAB laboratory of Brown University. The mineral contents of plagioclase, pyroxene, olivine, ilmenite, agglutinate and volcanic glass in lunar regolith have been optimized and inversed.

2. Data and Preprocessing

2.1. LSCC Data

The research material is the results of measurements in the Earth’s laboratory on samples transported from the moon, and the reflectance measurements are taken at Brown University, and the dataset is publicly available, and can be freely accessed “<https://www.planetary.brown.edu/relab> (accessed on 10 January 2022)”. LSCC (Lunar Soil Characterization Consortium) lunar soil sample data consists of 10 lunar highland lunar soil samples collected by Apollo 14/16 and 9 lunar soil samples collected by Apollo 11/12/15/17. These data are widely used in the field of lunar research, including lunar soil spectrum, mineral composition and content, particle size, maturity and other information, which have a high degree of recognition and authority (<http://apps.geology.brown.edu/RELAB.php> accessed on 10 January 2022). All the spectral data of LSCC are measured in Brown University’s RELAB laboratory under the conditions of an incident angle of 30 degrees and an exit angle of 0 degrees. LSCC is the only complete “ground reality” data on the moon, including the reflectance spectra and mineral abundance information of lunar regolith [5]. LSCC consists of 10 samples of lunar highland collected by Apollo 14/16 and 9 samples of lunar mare collected by Apollo 11/12/15/17 [10–12]. All the 19 original samples have been subdivided into four groups of <10 μm , 10–20 μm , 20–45 μm and >45 μm according to the particle size, a total of 76 samples. The reflectance spectra of all samples and the mineral content of the first three particle size groups have been measured by RELAB Laboratory of Brown University. All spectra were obtained at i (incident angle) = 30 degrees, e (emergence angle) = 0 degrees, and g (phase angle) = 30 degrees. The spectral resolution is 5 nm, forming a spectral database of 460 bands from 300 to 2600 nm [9,10].

Three particle size groups < 10 μm , 10–20 μm and 20–45 μm were selected in this paper which have both spectral data and mineral abundance information. The particle size group of >45 μm lacks mineral content information and is not considered as the research

object. Plagioclase is widely distributed in the lunar sea and lunar highlands. Plagioclase in the lunar soil is poor in sodium and rich in calcium, and has relatively strong absorption near $0.65\ \mu\text{m}$ and $1.25\ \mu\text{m}$, but it is not easy to observe in the lunar soil. This is because the absorption characteristics of plagioclase are weaker than other strong absorption minerals and can easily be concealed. By decomposing and filtering data information, PLS extracts the most explanatory comprehensive variables for dependent variables, identifies the information and noise in the system, and better overcomes the adverse effects of multiple correlations of variables in system modeling. The mineral contents of lunar mare and lunar highland are different. Lunar highland contains more plagioclase, so the spectra of lunar mare (Figure 1a) and the spectra of lunar highland are quite different (Figure 1b). Besides, plagioclase is widely distributed in the lunar sea and lunar highlands. Plagioclase in the lunar soil is poor in sodium and rich in calcium, and has relatively strong absorption near $0.65\ \mu\text{m}$ and $1.25\ \mu\text{m}$, but it is not easy to observe in the lunar soil. This is because the absorption characteristics of plagioclase are weaker than other strong absorption minerals and can easily be concealed. By decomposing and filtering data information, PLS extracts the most explanatory comprehensive variables for dependent variables, identifies the information and noise in the system, and better overcomes the adverse effects of multiple correlations of variables in system modeling. Therefore, 27 samples of lunar mare and 30 samples of lunar highland were selected and modeled separately. When modeling the samples of lunar mare and lunar highland respectively, two-thirds of samples were randomly selected as the experimental group to establish the prediction relationship, while the remaining one-third of the samples were used as the verification group to further verify the prediction relationship.

Figure 1. (a) is the reflectance spectra of 10,084 sampling point of lunar mare, and (b) is the reflectance of 14,141 sampling point of lunar highland.

2.2. Data Preprocessing Using Unscrambler 9.7

The visible and near-infrared spectra of lunar samples contain noise and various external disturbance factors in the data measurement. Before using PLS method to retrieve mineral content of lunar regolith, the spectra data should be pre-processed first. By pre-processing the spectra data of lunar samples, the noise information and irrelevant variables can be effectively eliminated. The influence of various non-target factors on the spectra can be eliminated so as to highlight the spectral features and improve the accuracy of the model [13].

In this paper, the original spectral data were pre-processed with the multivariate statistical software the Unscrambler 9.7 of Camo Company, Sweden. Common preprocessing

methods include continuum-removal, natural logarithmic variation, first-order derivative, second-order derivative, multiple scattering correction (MSC) and so on.

In this paper, the PLS regression relationship between the spectral and mineral content of lunar regolith of samples of lunar highland and lunar mare for LSCC data was constructed. The two sets of samples in LSCC data were pre-processed by continuum-removal, first derivative, second derivative and multiple scattering correction (MSC) respectively. Finally, MSC was chosen as the best pretreatment method to measure the error index and the number of principal components. MSC can effectively eliminate scattering caused by uneven particle distribution and particle size difference [14]. After MSC pretreatment of LSCC data, the effects of sample size, density, surface roughness and other surface factors on the spectra were eliminated, spectral drift was effectively eliminated, and spectral features were enhanced [15]. MSC can maximize the expression of dependent variable information on the premise of fewer principal components, and at the same time ensure small error, which plays a positive optimization role for the model (Figure 2). Figure 2a,b compare the spectral changes of a typical lunar sea sampling point (10,084) and a lunar highland sampling point (14,141) before and after MSC preprocessing.

Figure 2. (a) is the reflectance spectra of 10,084 sampling point of lunar mare processed by MSC, and (b) is the 14,141 sampling point of lunar highland processed by MSC.

3. Methodology and Principle

Before modeling by partial least squares, the spectrum of lunar soil samples was preprocessed by multivariate scattering correction MSC. The cross-validity test method is used to select the optimal number of principal components. In this paper, partial least square regression models are established for the lunar soil samples of the lunar highlands and the lunar sea respectively. Two-thirds of the samples were randomly selected as the experimental group to establish the prediction relationship between lunar soil spectra and mineral content. The remaining 1/3 of the samples are used as the verification group to further verify the prediction relationship.

3.1. Partial Least-Squares Regression (PLS)

Partial least squares (PLS) method has been widely used in many fields such as data analysis, parameter estimation, system identification, regression modeling and prediction [16,17]. PLS is a new multivariate statistical method, which integrates principal component analysis, multivariate correction analysis, canonical correlation analysis and linear regression analysis. It provides a multiple linear regression modeling method, overcoming the shortcomings of traditional regression analysis method which can not describe

the multicollinear relationship of variables. PLS can achieve regression modeling, data structure optimization, correlation analysis of two groups of variables, which obviously improves the reliability of data analysis and the accuracy of calculation process [18,19]. PLS method not only pay close attention to the information on the independent variables, and pay attention to the relationship between dependent and independent variables [20]. PLS can decompose and filter the data information in the system, identify the system information and noise, eliminate the redundant information among the data, and extract the comprehensive variable with the strongest explanatory power to the dependent variable. PLS can establish the model with higher accuracy when the number of sample points is less than the number of variables, which has the advantages that traditional regression analysis methods do not have.

PLS indirectly reflects the linear relationship between independent variables and dependent variables by establishing the regression model of latent variables of independent variables about latent variables of dependent variables. In the final model, all original independent variables will be included, and the regression coefficient of each independent variable in the regression model will be easier to explain [21]. The basic principle is as follows.

The p -independent variable matrix $X = \{x_1, \dots, x_p\}$ and q -dependent variable matrix $Y = \{y_1, \dots, y_q\}$ of n samples are decomposed. Components u_1 and v_1 are extracted from X and Y . u_1 is a linear combination of $x_1, \dots, x_p$ and v_1 is a linear combination of $y_1, \dots, y_q$. When extracting these two components, u_1 and v_1 should carry as much variation information as possible in their respective data tables, and the correlation degree of u_1 and v_1 can reach the maximum, which ensures that u_1 and v_1 represent matrix X and Y as well as possible and independent variable component u_1 has the strongest explanatory power for dependent variable component v_1 . After extracting the first components u_1 and v_1 , partial least squares regression is applied to the regression of X to u_1 and Y to v_1 , respectively. When the regression equation reaches the required precision, the algorithm will be terminated. Otherwise, the residual information after X and Y are interpreted by u_1 and v_1 respectively will be used for the next round of component extraction, so that the cycle can be completed until the required precision is reached. If the r components $u_1, \dots, u_r$ are extracted from the independent variable matrix X , the partial least squares regression will establish the regression equation of the dependent variable matrix Y and $u_1, \dots, u_r$, and then be expressed as the regression equation of the dependent variable matrix Y and the independent variable matrix X .

3.2. The Optimal Number of Principal Components

The selection of principal components directly affects the actual prediction ability of PLS model [22]. Too few principal components can not fully reflect the information of independent variables, resulting in inadequate fitting and low prediction accuracy. Excessive calculation of principal components is complex, and noise information is mixed in, which leads to overfitting and reduces the prediction ability of the model [23,24]. Due to the existence of Hughes phenomenon [25], when the feature dimension increases to a certain critical point, the classification accuracy will decrease and the prediction error of the model will also increase.

PLS method usually does not need to select $u_1, \dots, u_r$ with r components to establish regression equation, but just select the first l components ($l \leq r$) like PCA to get a regression model with better prediction ability. The optimal number of principal components can be determined by cross validation, according to the cumulative variance contribution of principal components to dependent variable matrix Y and the sum of squares of predicted residuals (Press). Greater contribution of cumulative variance and smaller residuals can lead to better prediction effect of PLS method.

4. Experimental Results and Analysis

4.1. Establishment of PLS Prediction Model for Mineral Content of Lunar Regolith

In this paper, the cross-validity test method was used to select the number of principal components, that is, the data set was randomly divided into k parts, one of which was used as a test set each time, and the remaining $k-1$ parts were trained as a training set [26]. The relationship between the number of principal components and the cumulative variance contribution of corresponding Y and the sum of squares of predicted residuals was calculated, as shown in Figures 3 and 4.

Figure 3. The optimal number of principal components of lunar highland, (a) is the relationship between the number of principal components and the cumulative variance contribution of corresponding Y , (b) is the relationship between the number of principal components and the sum of squares of predicted residual.

Figure 4. The optimal number of principal components of lunar mare, (a) is the relationship between the number of principal components and the cumulative variance contribution of corresponding Y , (b) is the relationship between the number of principal components and the sum of squares of predicted residual.

Regarding to lunar highland, as can be seen from Figure 3, the variance captured by the first principal component is only about 20% in the modeling of LSCC data of lunar highland. When the principal component number is 7, its cumulative variance contribution to Y has reached 85%, while the sum of squares of predicted residuals at this time is less than 1. It shows that the selection of the 7 principal components does not lose much information, reduces the data dimension and meets the modeling requirements.

While regarding to lunar mare, as demonstrated by Figure 4, we can find that the variance captured by the first principal component is less than 10%, but the total value of variance of the first and second principal components is slightly more than 20%. The cumulative variance contribution to Y reached to more than 85% when summing the first

7 principal components, at the same time, the sum of squares of predicted residuals at this time is less than 1. This figure also shows that the first 7-principal-components does not lose much information, thus, the reducing of the data's dimension could meet the modeling requirements.

As far as the mineral content retrieval of lunar highland, 20 randomly selected samples from 30 highland samples were used to model the spectra and mineral content of lunar regolith based on PLS method, as shown in Figure 5. The mineral contents of volcanic glass (Figure 5a), agglutinate (Figure 5b), ilmenite (Figure 5c), plagioclase (Figure 5d), pyroxene (Figure 5e), and olivine (Figure 5f) in the lunar regolith were optimized and retrieved, and compared with the measured mineral contents in the lunar regolith.

Figure 5. (a–f) are respectively the prediction of mineral content of volcanic glass, agglutinate, ilmenite, plagioclase, pyroxene, and olivine of lunar highland from the experimental group and their comparative analysis with the measured mineral contents.

Figure 5 shows that the PLS model has a good simulation effect. This calculation is done for all grain size range, it is combined from all grain sizes. There are a total of 4 particle size groups in LSCC data, of which 3 groups have measured mineral content, so we use these 3 groups of data. These 3 sets of data are calculated as a whole, and the total range is 0–45 μm . The LSCC data is subdivided into four groups of <10 μm , 10–20 μm , 20–45 μm , and <45 μm according to the particle size level, a total of 76 samples. Among them, <10 μm , 10–20 μm , 20–45 μm particle size group samples contain mineral abundance information. This paper selects a total of 57 sets of data in three particle size groups of <10 μm , 10–20 μm , and 20–45 μm (30 sets of lunar highlands and 27 sets of lunar seas) as the research objects, and the overall range of mineral particle sizes is between 0–45 μm . The mineral content of lunar regolith retrieved from highland samples can be linearly fitted with the measured mineral content of lunar regolith. The predicted and measured values of six major minerals of lunar regolith are evenly distributed along the $y = x$ line.

Table 1 gives the root mean square error (RMSE) and correlation coefficient of the predicted and measured content of the six major minerals in lunar highland, and evaluates the accuracy of the PLS model. The correlation coefficients between predicted and measured

contents of the six major mineral are all greater than 0.99. The smallest RMSE is 0.0386 for ilmenite and the largest is 1.5440 for plagioclase. The overall model has a good simulation effect. Based on the PLS method and the measured spectra data, the inversion of mineral content of lunar highland can be realized.

Table 1. Evaluation of prediction accuracy of mineral content of lunar highland from the experimental group.

	Volcanic Glass	Agglutinate	Ilmenite	Plagioclase	Pyroxene	Olivine
Correlation Coefficient	0.9963	0.9928	0.9967	0.9940	0.9950	0.9986
RMSE	0.1716	1.4305	0.0386	1.5440	0.5550	0.0588

Regarding to the mineral content retrieval of lunar mare, 18 randomly selected samples from 27 mare samples were used to model the spectra and mineral content of lunar regolith by PLS method, as shown in Figure 6. The mineral contents of volcanic glass (Figure 6a), agglutinate (Figure 6b), ilmenite (Figure 6c), plagioclase (Figure 6d), pyroxene (Figure 6e), and olivine (Figure 6f) in the lunar regolith were optimized and retrieved, and compared with the measured mineral contents in the lunar regolith.

Figure 6. (a–f) are respectively the prediction of mineral content of volcanic glass, agglutinate, ilmenite, plagioclase, pyroxene, and olivine of lunar mare from the experimental group and their comparative analysis with the measured mineral contents.

Figure 6 shows that the PLS model has a good simulation effect. As described as Figure 5, this calculation is done for all grain size range, it is combined from all grain sizes. The mineral content of lunar regolith retrieved from mare samples can be linearly fitted with the measured mineral content of lunar regolith. The predicted and measured values of six major minerals of lunar regolith are evenly distributed along the $y = x$ line.

Table 2 gives the root mean square error (RMSE) and correlation coefficient of the predicted and measured content of the six major minerals in lunar mare, and evaluates

the accuracy of the PLS model. The largest correlation coefficients between predicted and measured contents of the six major mineral is 0.9964 for volcanic glass, and the smallest is 0.9340 for plagioclase. The smallest RMSE is 0.3122 for ilmenite and the largest is 1.2640 for agglutinate. The overall model has a good simulation effect. Based on the PLS method and the measured spectra data, the inversion of mineral content of lunar mare can be realized.

Table 2. Evaluation of prediction accuracy of mineral content of lunar mare from the experimental group.

	Volcanic Glass	Agglutinate	Ilmenite	Plagioclase	Pyroxene	Olivine
Correlation	0.9964	0.9879	0.9946	0.9340	0.9854	0.9887
Coefficient RMSE	0.4831	1.2640	0.3122	0.6912	0.8673	0.1495

4.2. Validation of PLS Prediction Model for Mineral Content of Lunar Regolith

In this paper, PLS regression relationship between the lunar soil spectra and mineral content were established for lunar soil samples from the lunar highlands and the lunar sea. Two-thirds of samples were randomly selected as the experimental group to establish the prediction relationship, while the remaining one-third of the samples were used as the verification group to further verify the prediction relationship.

Based on the PLS equation of 20 randomly selected samples of lunar highland, the mineral contents of volcanic glass (Figure 7a), agglutinate (Figure 7b), ilmenite (Figure 7c), plagioclase (Figure 7d), pyroxene (Figure 7e), and olivine (Figure 7f) in the remaining 10 samples of lunar highland were predicted, and compared with the measured mineral contents in the lunar regolith, as shown in Figure 7.

Figure 7. (a–f) are respectively the prediction of mineral content of volcanic glass, agglutinate, ilmenite, plagioclase, pyroxene, and olivine of lunar highland from the validation group and their comparative analysis with the measured mineral contents.

Figure 7 shows that the PLS model has a good simulation effect. The mineral content of lunar regolith retrieved from highland samples can be linearly fitted with the measured mineral content of lunar regolith. The predicted and measured values of six major minerals of lunar regolith are evenly distributed along the $y = x$ line.

Table 3 gives the root mean square error (RMSE) and correlation coefficient of the predicted and measured content of the six major minerals in lunar highland, and evaluates the accuracy of the PLS model. The largest correlation coefficients between predicted and measured contents of the six major mineral is 0.9823 for plagioclase, and the smallest is 0.7010 for olivine. The smallest RMSE is 0.9660 for olivine and the largest is 6.2112 for agglutinate. The overall model has a good simulation effect. Based on the PLS method and the measured spectra data, the inversion of mineral content of lunar highland can be realized.

Table 3. Evaluation of prediction accuracy of mineral content of lunar highland from the validation group.

	Volcanic Glass	Agglutinate	Ilmenite	Plagioclase	Pyroxene	Olivine
Correlation	0.8944	0.9390	0.7744	0.9823	0.8586	0.7010
Coefficient RMSE	1.6070	6.2112	0.3597	4.2077	4.1708	0.9660

Based on the PLS equation of 18 randomly selected samples of lunar mare, the mineral contents of plagioclase, pyroxene, olivine, ilmenite, agglutinate and volcanic glass in the remaining 10 samples of lunar mare were predicted, and compared with the measured mineral contents in the lunar regolith, as shown in Figure 8.

Figure 8. (a–f) are respectively the prediction of mineral content of volcanic glass, agglutinate, ilmenite, plagioclase, pyroxene, and olivine of lunar mare from the validation group and their comparative analysis with the measured mineral contents.

Figure 8 shows that the PLS model has a good simulation effect. The mineral content of lunar regolith retrieved from mare samples can be linearly fitted with the measured

mineral content of lunar regolith. The predicted and measured values of six major minerals of lunar regolith are evenly distributed along the $y = x$ line.

Table 4 gives the root mean square error (RMSE) and correlation coefficient of the predicted and measured content of the six major minerals in lunar mare, and evaluates the accuracy of the PLS model. The largest correlation coefficients between predicted and measured contents of the six major mineral is 0.9790 for pyroxene, and the smallest is 0.7298 for plagioclase. The smallest RMSE is 0.9611 for olivine and the largest is 4.8934 for agglutinate. The overall model has a good simulation effect. Based on the PLS method and the measured spectra data, the inversion of mineral content of lunar mare can be realized.

Table 4. Evaluation of prediction accuracy of mineral content of lunar mare from the validation group.

	Volcanic Glass	Agglutinate	Ilmenite	Plagioclase	Pyroxene	Olivine
Correlation	0.8617	0.9011	0.9331	0.7298	0.9790	0.7708
Coefficient RMSE	2.1364	4.8934	1.8143	1.4403	2.4985	0.9611

5. Conclusions

In this paper, the lunar soil characterization consortium data of lunar regolith consisting of Apollo samples is taken as the research object. The PLS regression model is used to estimate the spectra of lunar regolith measured in RELAB laboratory of Brown University. The mineral contents in lunar highland and lunar mare have been optimized and inversed.

PLS method can identify system information and noise, eliminate data redundancy, extract comprehensive variables with the strongest explanatory power to the dependent variables, and build models with high accuracy when the number of samples is less than the number of variables. This method can be used to optimize the inversion of mineral content in lunar regolith, make up for the limited number of Apollo and Luna lunar landing points and the insufficiency of a small number of collected samples of lunar regolith, and can be further applied to mineral identification of lunar regolith on a whole-moon scale.

The spectra data of lunar regolith were pre-processed with MSC to eliminate the influence of sample size, density, surface roughness and other surface factors on the spectra, so as to weaken or eliminate the influence of various non-target factors on the spectra and highlight the spectra features of lunar regolith. The cross-validity test method was used to select the best number of principal components, and the information of mineral content in lunar regolith was maximally expressed and the error was small.

In the study, PLS models have been established for samples from lunar highland and lunar mare respectively. Two-thirds of samples have been randomly selected as experimental group to establish the prediction relationship between the spectra of lunar regolith and mineral content. The remaining one-third of samples have been used as verification group to further verify the prediction relationship. The results show that the prediction of mineral content of 20 lunar highland samples and 18 lunar mare samples in the experimental group is mostly accurate, and the errors between simulated and measured mineral content of plagioclase, olivine, pyroxene, ilmenite, agglutinate and volcanic glass are small. Based on the PLS equation established by the experimental group, the mineral content of 10 lunar highland samples and 9 lunar mare samples of the validation group was further predicted, with little error from the measured values. The PLS model has high accuracy and good stability. It has good application prospects to optimize the inversion of mineral content in lunar regolith using spectral data of lunar regolith.

Author Contributions: P.Z. and Z.Z. are equally contributed to this paper, both of them are co-first authors. The specific contribution of all authors are: Conceptualization, P.Z., H.H. and Z.Z.; methodology, Z.Z.; software, Z.Z.; validation, Z.Z., P.Z. and H.H.; formal analysis, Z.Z.; investigation, Z.Z., G.W. and P.Z.; resources, Z.Z. and G.W.; data curation, G.W.; writing—original draft preparation, Z.Z., P.Z. and H.H.; writing—review and editing, H.H.; supervision, P.Z. and H.H.; project administration, P.Z. and H.H.; funding acquisition, P.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Major Projects of High-Resolution Earth Observation System of China (04-Y30B01-9001-18/20).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Hecker, C.; Dilles, J.H.; van der Meijde, M.; van der Meer, F.D. Thermal infrared spectroscopy and partial least squares regression to determine mineral modes of granitoid rocks. *Geochem. Geophys. Geosystems* **2012**, *13*, 1–15. [\[CrossRef\]](#)
- Sidike, A.; Zhao, S.; Wen, Y. Estimating soil salinity in Pingluo County of China using QuickBird data and soil reflectance spectra. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *26*, 156–175. [\[CrossRef\]](#)
- Black, M.; Riley, T.R.; Ferrier, G.; Fleming, A.H.; Fretwell, P.T. Automated lithological mapping using airborne hyperspectral thermal infrared data: A case study from Anchorage Island, Antarctica. *Remote Sens. Environ.* **2016**, *176*, 225–241. [\[CrossRef\]](#)
- Rasera, J.N.; Cilliers, J.J.; Lamamy, J.A.; Hadler, K. The beneficiation of lunar regolith for space resource utilisation: A review. *Planet. Space Sci.* **2020**, *186*, 104879. [\[CrossRef\]](#)
- Li, S.; Li, L. Radiative transfer modeling for quantifying lunar surface minerals, particle size, and submicroscopic metallic Fe. *J. Geophys. Res.* **2011**, *116*, E09001. [\[CrossRef\]](#)
- Taylor, L.A.; Pieters, C.M.; Keller, L.P.; Morris, R.V.; McKay, D.S. Lunar mare soils: Space weathering and the major effects of surface-correlated nanophase Fe. *J. Geophys. Res.* **2001**, *106*, 27985–27999. [\[CrossRef\]](#)
- Clark, R.N.; Rencz, A.N. Spectroscopy of rocks and minerals, and principles of spectroscopy. *Man. Remote Sens.* **1999**, *3*, 3–58.
- Burns, R.G. *Mineralogical Applications of Crystal Field Theory*; Cambridge University Press: New York, NY, USA, 1993.
- Li, L. Quantifying lunar soil composition with partial least squares modeling of reflectance. *J. Geophys. Res. Planets* **2008**, *42*, 267–274. [\[CrossRef\]](#)
- Taylor, L.A.; Pieters, C.M.; Morris, R.V.; McKay, D.S.; Pachen, A.; Wentworth, S. Integration of the chemical and mineralogical characteristics of lunar soils with reflectance spectroscopy. In Proceedings of the Lunar and Planetary Science Conference, 30th Annual Lunar and Planetary Science Conference, Houston, TX, USA, 15–29 March 1999.
- Taylor, L.A.; Pieters, C.M.; Patchen, A.; Taylor, D.H.; Morris, R.V.; Keller, L.P.; McKay, D.S. Mineralogical characterization of lunar highland soils. *Lunar Planet Sci.* **2003**, *34*, 1774.
- Taylor, L.A.; Pieters, C.M.; Patchen, A.; Taylor, D.H.S.; Morris, R.V.; Keller, L.P.; McKay, D.S. Mineralogical and chemical characterization of lunar highland soils: Insights into the space weathering of soils on airless bodies. *J. Geophys. Res. Planets* **2010**, *115*, 1–14. [\[CrossRef\]](#)
- Wang, D.M.; Jin, S.Z.; Chen, H.C.; Chen, X.D. Quantitative Analysis of Cotton Content in Cotton-terylene Textile by Near Infrared Technique. *Opt. Precis. Eng.* **2008**, *16*, 2052.
- Huang, Q.T.; Zhou, L.Q.; Shi, Z.; Li, Z.Y.; Gu, Q. Analysis of Lead in Soil with Partial Least Square Regression (PLS) Method and Field Portable X-Ray Fluorescence (FPXRF) Analyzer. *Spectrosc. Spectr. Anal.* **2009**, *29*, 1434–1438.
- Qiao, X.Y.; Wang, Y.J.; Li, G. Prediction of Acetamiprid Residues by Fluorescence Spectroscopy Based on PLS. *Opt. Precis. Eng.* **2010**, *18*, 2369–2374.
- Wold, S.; Albano, C.; Dunn, W.J., III; Esbensen, K.; Hellberg, S. Pattern recognition: Finding and using regularities in multivariate data. *Proc. Food Res. Data Anal.* **1983**, *3*, 147–188.
- Hu, D.G.; Zhang, R.C. *Multivariate Data Analysis Method—Pure Algebraic Processing*; Nankai University Press: Tianjin, China, 1990.
- Zhou, P.; Zhao, Z.; Wei, G.; Huo, H.Y. Two Simulated Spectral Databases of Lunar Regolith: Method, Validation, and Application. *Remote Sens.* **2022**, *14*, 277. [\[CrossRef\]](#)
- Gao, B.B.; Pan, J.Y.; Liu, Y.P.; Yuan, M.S. Study on prediction model of seam gas content based partial least squares regression. *J. Henan Polytech. Univ. (Nat. Sci.)* **2015**, *34*, 146–150.
- Ge, L.; Wang, B.; Zhang, L.M. Partial Least Squares Based Band Selection for Hyperspectral Imagery. *J. Comput.-Aided Des. Comput. Graph.* **2011**, *23*, 1844–1852.
- Li, P.C.; Zhu, J.T.; Ma, Y.D.; Wu, X. Near Infrared Spectral Analysis Based on Partial Least Squares. *J. Geomat.* **2015**, *40*, 53–56.
- Hughes, G. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* **1968**, *14*, 55–63. [\[CrossRef\]](#)
- Sun, Y.W.; Zhou, P. Research on Spectral Pretreatment Methods for Quantitative Analysis of Wet Gluten in Flour by Near Infrared Spectroscopy. *China Brew.* **2016**, *35*, 137–140.
- Chu, X.L. *Molecular Spectroscopy Analytical Echnology Combined with Chemomentrics and Its Applications*; Chemical Industry Press: Beijing, China, 2011.
- Sun, B.L.; Liu, J.L.; Cai, Y.B. Determination of Crystallinity in Neosinocalamus affinis Based on Near Infrared Spectroscopy and PLS Methods. *Spectrosc. Spectr. Anal.* **2011**, *31*, 366–370.
- Yu, J.C.; Liu, W.L.; Yan, B.K.; Dong, X.F.; Wang, Z.; Li, N. Inversion of Geochemical Compositions of Basalts Based on Field Measured Spectra. *Remote Sens. Land Resour.* **2017**, *29*, 158–163.