frontiers | Research Topics

New advances in lunar and related planetary studies

Edited by
Bojing Zhu, Caibo Hu, Changyi Xu, Yaolin Shi, Guangfei Wei,
Jinhai Zhang, Juan Li, An Yin and Tieyuan Zhu

Published in
Frontiers in Astronomy and Space Sciences

ISSN 1664-8714
ISBN 978-2-8325-4904-9
DOI 10.3389/978-2-8325-4904-9

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

New advances in lunar and related planetary studies

Topic editors
Bojing Zhu – Yunnan Observatories, National Astronomical Observatories, Chinese Academy of Sciences (CAS), China
Caibo Hu – University of Chinese Academy of Sciences, China
Changyi Xu – Institute of Geology and Geophysics, Chinese Academy of Sciences (CAS), China
Yaolin Shi – University of Chinese Academy of Sciences, China
Guangfei Wei – Deep Space Exploration Laborarory, China
Jinhai Zhang – Institute of Geology and Geophysics, Chinese Academy of Sciences (CAS), China
Juan Li – Institute of Geology and Geophysics, Chinese Academy of Sciences (CAS), China
An Yin – University of California, Los Angeles, United States
Tieyuan Zhu – The Pennsylvania State University (PSU), United States

Citation
Zhu, B., Hu, C., Xu, C., Shi, Y., Wei, G., Zhang, J., Li, J., Yin, A., Zhu, T., eds. (2024). *New advances in lunar and related planetary studies*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4904-9

Table of contents

Frontiers in Astronomy and Space Sciences

TYPE Editorial
PUBLISHED 03 May 2024
DOI 10.3389/fspas.2024.1412073

EDITED AND REVIEWED BY
Miriam Rengel,
Max Planck Institute for Solar System
Research, Germany

*CORRESPONDENCE
Yaolin Shi,
✉ shiyl@ucas.ac.cn

RECEIVED 04 April 2024
ACCEPTED 22 April 2024
PUBLISHED 03 May 2024

CITATION
Hu C, Zhu B and Shi Y (2024), Editorial: New advances in lunar and related planetary studies.
Front. Astron. Space Sci. 11:1412073.
doi: 10.3389/fspas.2024.1412073

Editorial: New advances in lunar and related planetary studies

Caibo Hu[1,2], Bojing Zhu[3] and Yaolin Shi[1,2]*

[1]College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing, China,
[2]The Key Laboratory of Computational Geodynamics, Chinese Academy of Sciences, Beijing, China,
[3]Yunnan Observatories, Chinese Academy of Sciences, Kunming, China

KEYWORDS

lunar science, martian exploration, deep space exploration, planetary science, solar system

Editorial on the Research Topic
New advances in lunar and related planetary studies

Significant strides have propelled the lunar and planetary sciences field forward in recent years, with notable achievements stemming from the booming U.S. Apollo missions. These missions stand as landmarks in our exploration of the lunar landscape, providing invaluable insights into surface conditions, evolutionary dynamics, and the internal structure of the Moon. Similarly, American Curiosity and Exploration Mars landers have left an indelible mark on Martian exploration, unraveling mysteries of the Martian terrain and atmosphere, for example.

Moreover, nations across the globe, including Europe, China, Russia, Japan, and India, for example, have redoubled their efforts in advancing our understanding of celestial bodies. Notably, China has emerged as a critical player in deep space exploration, with its Chang'e, Tianwen, and Shenzhou programs spearheading missions to the Moon and Mars. These initiatives have not only conducted lunar orbital maneuvers, surface landings, and sample retrievals but have also successfully landed on Mars, significantly expanding our knowledge of the Martian environment.

Planetary science including lunar studies needs to be based on remote sensing observations in deep space orbits, instrumented observations on planetary surfaces, patrol observations of probe landings, experiments on returned material samples, and numerical simulation studies of coupled multiphysics fields. To showcase the latest advancements in planetary science, including lunar studies, a team of nine leading experts in Earth and planetary sciences from China and the United States curated this Research Topic. Ten academic manuscripts were considered for development, undergoing rigorous review by external experts and subsequent revisions by the authors. Ultimately, eight manuscripts were accepted, representing cutting-edge research in the field. This collaborative effort makes a fundamental contribution to the establishment of an international scientific research station on the Moon and the exploitation of planetary resources in the near future. Through collaboration and innovation, we continue to push the boundaries of scientific knowledge, paving the way for future discoveries and breakthroughs in space exploration.

Research into Martian terrain features and landform identification is pivotal, encompassing Mars exploration, data acquisition, mapping techniques, and methods for classifying Martian landforms. Of particular significance are wind-sand-wind-formed, hydromorphic fluvial, and impact landforms Liu and Cheng. Quantitative analysis of rock characterization and size-frequency distribution at the Zhurong landing site on Mars can be conducted utilizing images captured by the Navigation and Terrain Cameras (NaTeCams) aboard the Zhurong rover Sun et al..

Geophysical inversions are essential for probing the Moon's internal structure. Employing wavelet multiscale analysis, Yu et al. used inverted lunar gravity data to unveil three-dimensional morphological features of two large dense masses in Mare Serenitatis. Additionally, the deeper lunar core interface morphology was quantitatively studied using characteristics of Apollo moonquake observations. Yang and Wang utilized a pseudo-spectral and finite-difference hybrid method to simulate wave propagation of deep moonquakes, considering near-surface scattering effects and analyzing the impacts of these factors on the results of superimposed reflected and converted phases of the outer lunar core.

The electrostatic migration of lunar surface dust mineral particles due to electron irradiation poses potential challenges for human settlement on the Moon. Gan et al. conducted a comparative analysis of the electrostatic migration characteristics of pyroxene, olivine, and ilmenite grains, measured using laser Doppler methods, with previous experiments on anorthosite grains. This comparison highlights distinct migration behaviors among various mineral compositions in lunar dust.

This Research Topic showcases recent advancements in numerical modeling of Earth's polar dynamics. Shi et al. demonstrate significant enhancements in prediction accuracy for Earth's polar motion across various time scales, including short-term (within 360 days) and medium-to long-term, through the innovative integration of iterative oblique singular spectrum analysis (IOSSA) with pseudo data (IOSSApd).

The lunar poles, potential sites for human research bases due to possible water ice presence, are of significant interest. Wei et al. utilized data from the Chang'e 2 microwave radiometer and Diviner observations to invert lunar polar heat flow. Their findings reveal an average heat flow of $4.9 \pm 0.2\ \mathrm{mW/m^2}$ in the permanently shadowed craters Haworth and Shoemaker at the lunar south pole.

Ground Penetrating Radar (GPR) plays a pivotal role in deciphering the shallow surface structure of the Moon and related planets. Lv and Zhang conducted a quantitative inversion of the underground permittivity of lunar clastic rocks within the Chang'e-4 landing zone utilizing isolated hyperbolic diffraction migration and local focusing analysis.

The future of planetary science including lunar studies is poised for further advancements, driven by ongoing initiatives and promising developments on multiple fronts. For instance, under Elon Musk's visionary leadership, SpaceX's starship program continues to make significant strides, propelled by its unparalleled carrying capacity and cost-effectiveness. Concurrently, China's space exploration efforts have garnered considerable attention, with the successful Magpie Bridge 2 mission and the highly anticipated Chang'e-6 mission on the horizon. These endeavors promise to expand our understanding of the Moon and pave the way for future exploration missions.

As we look ahead, we anticipate a wealth of new observation data, research methodologies, and technological applications that will deepen our knowledge of the moon and related planets. The continuous influx of data and innovative approaches will undoubtedly fuel groundbreaking discoveries and shape the trajectory of future space exploration endeavors.

In closing, we sincerely thank the handling editors, editor-in chief, reviewers, and authors whose unwavering dedication and contributions have been instrumental in the successful publication of this Research Topic. In particular, we honor the memory of Professor Yin An, one of the handling editors, whose profound insights and invaluable contributions left an indelible mark on the publication process. It is with deep appreciation and respect that we dedicate the success of this publication to Professor Yin An, whose legacy continues to inspire and guide us in the pursuit of scientific excellence.

Author contributions

CH: Conceptualization, Funding acquisition, Project administration, Supervision, Writing–original draft, Writing–review and editing. BZ: Conceptualization, Writing–original draft, Writing–review and editing. YS: Conceptualization, Investigation, Supervision, Validation, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was financially supported by the National Science Foundation of China (42374116) and supported by the Fundamental Research Funds for the Central Universities.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Frontiers in Astronomy and Space Sciences

TYPE Original Research
PUBLISHED 15 March 2023
DOI 10.3389/fspas.2023.1109714

OPEN ACCESS

EDITED BY
Changyi Xu,
Institute of Geology and Geophysics (CAS), China

REVIEWED BY
Yawen She,
Institute of Earthquake Forecasting, China Earthquake Administration, China
Guangliang Yang,
China Earthquake Administration, China

*CORRESPONDENCE
Chuang Xu,
✉ chuangxu@gdut.edu.cn

SPECIALTY SECTION
This article was submitted to Planetary Science, a section of the journal Frontiers in Astronomy and Space Sciences

RECEIVED 28 November 2022
ACCEPTED 23 January 2023
PUBLISHED 15 March 2023

CITATION
Yu H, Xu C, Wu Y, Li J, Jian G and Xu M (2023), Fine structure of the lunar crust and upper mantle in the mare serenitatis derived from gravity multi-scale analysis. *Front. Astron. Space Sci.* 10:1109714.
doi: 10.3389/fspas.2023.1109714

Fine structure of the lunar crust and upper mantle in the mare serenitatis derived from gravity multi-scale analysis

Hangtao Yu[1,2,3], Chuang Xu[4]*, Yihao Wu[5], Jinbo Li[4], Guangyu Jian[4] and Ming Xu[1,2,3]

[1]Guangzhou Marine Geological Survey, China Geological Survey, Guangzhou, China, [2]Key Laboratory of Marine Mineral Resoures, Ministry of Natural Resources, Guangzhou Marine Geological Survey, China Geological Survey, Guangzhou, China, [3]National Engineering Research Center for Gas Hydrate Exploration and Development, Guangzhou, China, [4]Department of Geodesy and Geomatics Engineering, Guangdong University of Technology, Guangzhou, China, [5]School of Earth Sciences and Engineering, Hohai University, Nanjing, China

It is significant for revealing the formation mechanism of the lunar Mascon to invert the refined 3-D lunar crust and upper mantle structure of the Mare Serenitatis. As the development of space exploration technology, lunar gravity data has advantages of high accuracy and resolution, which can be used to invert the lunar crust and upper mantle structure. However, gravity anomaly reflects all anomalous material during the whole Moon's interior, and its vertical structure recognition capability is poor. Thus, this paper adopts wavelet multi-scale analysis method to decompose the gravity anomaly in the Mare Serenitatis for enhancing vertical resolution, and the corresponding field source depths of the decomposed gravity anomalies are further estimated by power spectrum method. Subsequently, the layered densities and the crust-mantle interface depth of the Mare Serenitatis are inverted. The research results show that the 3-D morphological character of two large high-density materials in the Mare Serenitatis is clearly depicted. The southwest high-density material with its bottom center at approximately (15°E, 25°N) has the depth range of 50 km–80 km and the maximum diameter of approximately 150 km. As for the southeast high-density material, its bottom center is located at approximately (23°E, 25°N), the depth range is 30 km–60 km and the maximum diameter is approximately 100 km. Another new finding is that the crust-mantle interface uplift has obviously fallen back in the center of the Mare Serenitatis. The high-density materials and crust-mantle interface uplift may together promote the formation of the Mascon in the Mare Serenitatis.

KEYWORDS

gravity field, gravity inversion, lunar crust and mantle, mare serenitatis, MASCON

Key points

- A refined 3-D structure model of the lunar crust and upper mantle in the Mare Serenitatis was constructed.
- The 3-D morphological feature of two high-density materials beneath the study area was revealed.
- A falling phenomenon for the crust-mantle interface uplift is found in the center of the Mare Serenitatis basin.

1 Introduction

The Mare Serenitatis is one of the largest lunar mares on the Moon, with a ring structure of approximately 880 km in diameter and the area of approximately 310,000 km^2 (Solomon and Head, 1979; Watters and Konopliv, 2001). It is bordered by the western Mare Imbrium, the southeastern Mare Tranquillitatis, the northern Mare Frigoris, the northeastern Lacus Somniorum (LS) and Posidonius (PO), and the southwestern Mare Vaporum (Hiesinger et al., 2000). The Mare Serenitatis Basin is surrounded by numerous mountains, such as the Montes Caucasus (MC) in the northwest, the Mons Hadley (MHL) in the west, and the Montes Haemus (MH) in the southwest (see Figure 1). Besides, the main lunar ridges on the Mare Serenitatis surface are DA, DVC, DG, DS, SR, DO, and DAD, which form an inner ring and an outer ring (Maxwell et al., 1975). These two rings have a clear trend of northward extension and a discontinuous distribution within the Mare Serenitatis Basin, which are often considered as the surface feature of folding and retrograde faults (Maxwell et al., 1975; Plescia and Golombek, 1986; Watters, 1988; Watters and Konopliv, 2001). In addition, as a result of multiple meteorite impact events, the surface of the Mare Serenitatis is dotted with impact craters. Bessel (BE) is the largest one and the eastern basin margin is covered by a 60 km diameter impact crater, Le Monnier Crater (LM) (Pieters, 1978).

Up to now, the formation of the Mare Serenitatis is still controversial. Some scholars believe that it is a young basin type according to the Apollo 17 collection (Staudacher et al., 1978). However, the current prevailing view is that it has formed during the Nectarian Period (Wilhelms et al., 1987). Approximately 3.8 Ga ago, the Mare Serenitatis experienced an impact event from a minor celestial body and then it was filled with magma, gradually forming a special basalt-covered landscape with an average basalt thickness of 798 m (Li et al., 2018) and the main composition of low-titanium basalts (Kodama and Yamaguchi, 2003). Watters and Konopliv (2001) agreed that the Mare Serenitatis consisted of two overlapping basins (the northern basin is older than the southern basin), and suggested that the ring system of the Imbrium Basin had contributed to the topography of the Mare Serenitatis Basin, driving the terrain uplift of the western part of the Mare Serenitatis. Sharpton and James, 1982 showed that various areas of the Mare Serenitatis Basin were formed by the volcanic infilling and tectonic movements at different times. Therefore, it is of great importance to determine the refined lunar crust and upper mantle structure for understanding the formation mechanism and dynamical processes of the Mare Serenitatis (Watters and Konopliv, 2001).

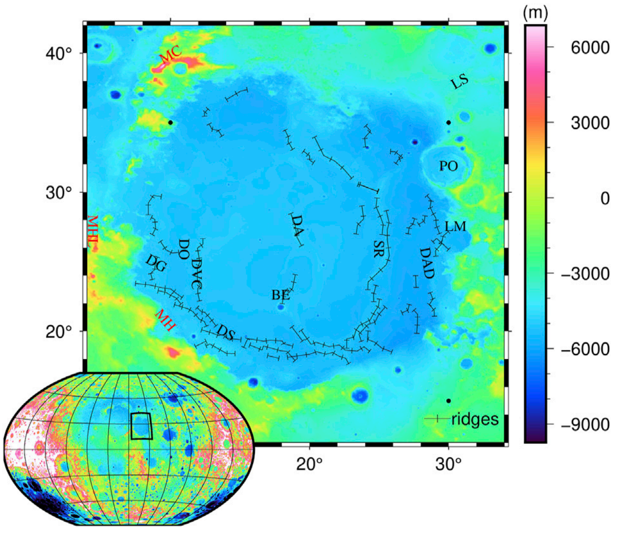


FIGURE 1
Topography of the Mare Serenitatis. MC, montes caucasus: MHL, mons hadley; MH, montes haemus; DA, dorsum azara; DVC, dorsum von cotta; DG, dorsum gast; DS, dorsa sorby; SR, serpentine ridge; DO, dorsum owen; DAD, dorsa aldrovandi; BE, bessel; LM, le monnier crater; LS, somniorum; PO, posidonius. The black lines with bars represent lunar ridges.

During recent decades, numerous scholars have studied the surface morphological character of the Mare Serenitatis (Maxwell et al., 1975; Head, 1979; Sharpton and James, 1982; Ryder et al., 1997; Kaur et al., 2013; Li et al., 2018). However, the lunar crustal and upper mantle structure of the Mare Serenitatis is still inadequate at present. With the development of space exploration technology, lunar gravity data has its advantages of high accuracy and resolution, and can directly reflect the density distribution of the Moon's interior. Thus, new progress is continuously made during the inversion of the lunar interior structure using gravity methods. Watters and Konopliv (2001) used gravity data from lunar prospectors to argue that the Mare Serenitatis contained two overlapping basins. Hikida and Wieczorek (2007) analytically calculated the external gravitational field of an arbitrarily shaped polyhedron to obtain the thickness of the whole lunar crust. Wieczorek et al. (2013) used the Gravity Recovery and Interior Laboratory (GRAIL) gravity data and the Lunar Reconnaissance Orbiter (LRO) topographic data to yield models of lunar crustal thickness under different conditions. Liang et al. (2014) used a new inversion algorithm to obtain the 3-D density distribution of the lunar crust and mantle, and found large areas of lateral density heterogeneity beneath the South Pole-Aitken Basin. Zhao et al. (2021) obtained the 3-D density structure of the lunar crust and upper mantle based on the lunar gravity field model GL1500E. However, gravity data is the comprehensive reflection for the density and volume of all materials in the lunar interior, which makes it difficult to obtain accurate density structures at different depths. Hence, it is urgent to develop more effective methods for improving the vertical identification capacity of gravity data.

At present, the methods, commonly used for separating gravity signals to enhance vertical identification capacity of subsurface structure, include trend analysis, analytic extension, wavelet multi-scale analysis, and so on. Among them, wavelet multi-scale analysis is one of the most effective methods, which can accurately extract the gravity signals corresponding to target bodies at different depths. It has been proven and widely used in the study of the Earth's crustal and upper mantle structure (Jiang et al., 2012; Xu et al., 2017; 2018). Therefore, this paper firstly adopts wavelet multi-scale analysis to decompose the Bouguer gravity anomaly in the Mare Serenitatis. And then the subsurface density structure at different depths and the crust-mantle interface relief are inverted. Lastly, the tectonic implications of the new obtained 3-D lunar crustal and upper mantle structure beneath the study area are discussed.

2 Data and methods

2.1 Data

The adopted gravity data in this paper are derived from the GRGM1200A spherical harmonic coefficient model from National

Aeronautics and Space Administration (NASA)'s GRAIL satellite (Lemoine et al., 2014; Goossens et al., 2016), which is a corrected Bouguer gravity field model with the degree of 1200 and covers the entire Moon.

2.2 Bouguer gravity anomaly calculation

The Bouguer gravity anomaly $\Delta g(\varphi,\lambda)$ in the Mare Serenitatis area can be expressed as (Wang et al., 2009; Liang, 2010):

$$\Delta g(\varphi,\lambda)=\frac{GM}{R^2}\sum_{l=2}^{\infty}\sum_{m=0}^{l}(l+1)\left(\frac{R}{r}\right)^{l+2}(\bar{C}_l^m\cos m\lambda+\bar{S}_l^m\sin m\lambda)\bar{P}_l^m(\cos\varphi) \tag{1}$$

where G is the universal gravitational constant, M is the lunar mass, R represents mean lunar radius, r is the distance from the calculation point to the mass center of the Moon, $\bar{C}_l^m$ and $\bar{S}_l^m$ are the spherical harmonic coefficients provided by gravity field model, l and m are the respective degree and order, λ and φ are the respective longitude and colatitude, and $\bar{P}_l^m(\cos\varphi)$ is the regularized Legendre function. $\Delta g(\varphi,\lambda)$ reflects the sum of all subsurface anomalous materials. In order to extract the corresponding gravity signals of targeted materials at different depths, $\Delta g(\varphi,\lambda)$ needs to be further decomposed.

2.3 Multi-scale decomposition of Bouguer gravity anomaly

According to the principle of wavelet multi-scale analysis (Mallat, 1989), $\Delta g(\varphi,\lambda)$ can be decomposed into a low-frequency part and high-frequency parts of different orders, as shown in Eq. 2.

$$\Delta \mathrm{g}(\varphi,\lambda)=A_j(\varphi,\lambda)+\sum_{j=1}^{J}D_j(\varphi,\lambda) \tag{2}$$

where $A_j(\varphi,\lambda)$ is the j th order wavelet approximation, reflecting the low frequency signal; $D_j(\varphi,\lambda)$ is the j th order wavelet detail, reflecting the high frequency signal; and J is the maximum order of decomposition.

Gravity anomalies in various frequency bands can be regarded as the signals of anomalous materials at different depths. The average field source depth H_j of $D_j(\varphi,\lambda)$ is further estimated by the radial power spectrum method (Spector and Grant, 1970)

$$H_j=\frac{\Delta\ln P_j}{4\pi\Delta k} \tag{3}$$

where $\ln P_j$ is the logarithm of the power spectrum for $D_j(\varphi,\lambda)$, k is the wave number, and Δ is the variability.

2.4 Layered density inversion

According to the mean field source depth H_j $(j=1,2,3...,J)$, the lunar crust and upper mantle of the Mare Serenitatis is firstly layered. Subsequently, each layer is further gridded by Tesseroids. The function between the decomposed gravity anomaly $D_j(\varphi,\lambda)$ and the density anomaly $\Delta\rho_j(\varphi_0,\lambda_0)$ of each Tesseroid at corresponding layer can be expressed as (Heck and Seitz, 2007):

$$D_j(\varphi,\lambda)=\sum_{\varphi_0}\sum_{\lambda_0}G\Delta\rho_j(\varphi_0,\lambda_0)\Delta r_j\Delta\varphi\Delta\lambda\left[L_{000}+\frac{1}{24}(L_{200}\Delta r^2+L_{020}\Delta\varphi^2+L_{002}\Delta\lambda^2)\right] \tag{4}$$

where φ_0 and λ_0 are the respective central latitude and longitude of the Tesseroid, $\Delta\varphi$ and $\Delta\lambda$ are the respective latitude and longitude intervals, and Δr_j is the thickness of the Tesseroid. L_{000}, L_{200}, L_{020} and L_{002} are the correlation coefficients as shown in Eqs 5–8 respectively.

$$L_{000}=\frac{r_0^2(R-r_0\cos\psi_0)\cos\varphi_0}{l_0^3} \tag{5}$$

$$\begin{aligned}L_{200}=\frac{R\cos\varphi_0}{l_0^3}\{2-\frac{3r_0}{l_0^2}[5r_0-(2R+3r_0\cos\Psi_0)\cos\Psi_0]\\+\frac{15r_0^3}{l_0^4}\sin^2\psi_0(r_0-R\cos\psi_0)\}\end{aligned} \tag{6}$$

$$\begin{aligned}L_{020}=&\left(\frac{r_0}{l_0}\right)^3\cos\varphi(1-2\sin^2\varphi_0)\cos\delta\lambda+\frac{r_0^2}{l_0^5}\{-R(R^2+r_0^2)\cos\varphi_0\\&+r_0\sin\varphi[-Rr_0(\sin\varphi\cos\varphi_0-\cos\varphi\sin\varphi_0\cos\delta\lambda)\\&+\sin\varphi_0\cos\varphi_0(2R^2+4r_0^2-3Rr_0\sin\varphi\sin\varphi_0)]\\&+r_0^2\cos\varphi\cos\delta\lambda(1-2\sin^2\varphi_0)\times[r_0+R\cos\varphi\cos\varphi_0\cos\delta\lambda]\\&+Rr_0^2\cos\varphi\sin\varphi_0\cos\varphi_0\cos\delta\lambda\times[3\sin\varphi\cos\varphi_0-4\cos\varphi\sin\varphi_0\cos\delta\lambda]\}\\&+\frac{5Rr_0^3}{l_0^7}\{-R(R^2+r_0^2)\sin\varphi_0+r_0^2\cos\varphi\sin\varphi_0\cos\varphi_0\cos\delta\lambda\times\\&(r_0+R\cos\varphi\cos\varphi_0\cos\delta\lambda)+r_0\sin\varphi[2R^2-r_0^2-Rr_0\cos\psi_0+\sin^2\varphi_0]\\&\times(R^2+2r_0^2-Rr_0\sin\varphi\sin\varphi_0)\}\times(\sin\varphi\cos\varphi_0-\cos\varphi\sin\varphi_0\cos\delta\lambda)\end{aligned} \tag{7}$$

$$\begin{aligned}L_{002}=&\left(\frac{r_0}{l_0}\right)^3\cos\varphi\cos^2\varphi_0\times\{\cos\delta\lambda-\frac{3R}{l_0^2}\Big[2r_0\cos\varphi\cos\varphi_0\sin^2\delta\lambda\\&+(R-r_0\cos\psi_0)\cos\delta\lambda\Big]+\frac{15R^2r_0}{l_0^4}\cos\varphi\cos\varphi_0(R-r_0\cos\psi_0)\sin^2\delta\lambda\}\end{aligned} \tag{8}$$

where $r_0=R-H_j$, $\delta\lambda=\lambda_0-\lambda$, $l_0=\sqrt{R^2+r_0^2-2Rr_0\cos\Psi_0}$, $\cos\Psi_0=\sin\varphi\sin\varphi_0+\cos\varphi\cos\varphi_0\cos\delta\lambda$. Simplify and write Eq. 4 into the matrix form, as shown in Eq. 9

$$D_j=B\Delta\rho_j \tag{9}$$

where B is the kernel function matrix. At last, $\Delta\rho_j(\varphi_0,\lambda_0)$ is solved by using the Tikhonov regularization method (Tikhonov and Arsenin, 1977):

$$\Delta\rho_j=(B^TB+\alpha I)^{-1}B^TD_j \tag{10}$$

where I is the unit matrix and α is the regularization factor, which is determined by L-curve method (Hansen and O'Leary, 1993).

2.5 Crust-mantle interface depth determination

It is a key factor for obtaining refined crust-mantle interface depth to accurately extract the gravity anomaly signal caused by the crust-mantle interface relief, which needs to remove the effect of sediments, consolidated crystalline lunar crust and lower lunar mantle (Wan et al., 2019). The common methods for these corrections depend on existing lunar crustal and mantle models. However, at present the lunar crustal and mantle models are not accurate, which may cause large errors during the corrections. Hence, in this paper the gravity anomaly signal g^{BA} caused by the crust-mantle interface relief of the Mare Serenitatis is extracted by combining $D_j(\varphi,\lambda)$, its corresponding field source depth H_j , and the *a priori* information on the available crust-mantle interface depth. We compared the estimated H_j with the average depth of crust-mantle interface provided by *a priori* information to determine which order of $D_j(\varphi,\lambda)$ was the gravity anomaly signal caused by the crust-mantle interface undulation.

Subsequently, the crust-mantle interface depth $h(\varphi,\lambda)$ of the study area is inverted from g^{BA} using the effective power topography method (Wieczorek and Phillips, 1998):

$$h(\varphi,\lambda) = w_l\left(\frac{g^{BA}M(2l+1)}{4\pi\Delta\delta Z^2}\right)(R/Z)^l - Z\sum_{n=2}^{l+3}\frac{h(\varphi,\lambda)^n}{Z^n n!}\frac{\prod_{s=1}^{n}(l+4-s)}{l+3}\Big) \tag{11}$$

where $\Delta\delta$ is the density contrast at the crust-mantle interface, l is the degree of spherical harmonic coefficients, Z is the distance from the mean depth of crust-mantle interface to mass center of the Moon, and w_l is the downward topological filter, as shown in Eq. 12

$$w_l = \{1+\mu[\frac{M(2l+1)}{4\pi\Delta\delta Z^2}(R/Z)^l]^2\}^{-1} \tag{12}$$

in which μ is the Lagrange multiplier. The determination of μ is a subjective process, where the larger μ is, the more the high-frequency signal will be filtered. According to Wieczorek and Phillips (1998), here we choose μ such that $w_{30} = 0.5$. The final crust-mantle interface depth $h(\varphi,\lambda)$ will be determined by the iteration of Eqs 11, 12.

2.6 Profile density inversion

Because the vertical density variation of the layered density inversion is not continuous, the profile density inversion is further conducted in order to recover a more refined 3-D structure of the lunar crust and upper mantle in the Mare Serenitatis region. The compact gravity inversion method proposed by Last and Kubik (1983) is adopted to construct the profile density model in this paper. The profile gravity anomaly g_i can be written as:

$$g_i = \sum_{j=1}^{M} a_{ij}v_j + e_i, i = 1,2,3\ldots,N \tag{13}$$

where v_j is the profile density anomaly of jth block, e_i is the noise associated with i th data point, having an initial value of $0.1g_i$. a_{ij} is the coefficient, as shown in Eq. 14

$$a_{ij} = 2G[(x_i - x_j + d/2)\log(r_2r_3/r_1r_4) + d\times\log(r_4/r_3) - (z_j + h/2)(\theta_4 - \theta_2) + (z_j - h/2)(\theta_3 - \theta_1)] \tag{14}$$

Among them, the parameters are:

$$r_1 = (z_j - h/2)^2 + (x_i - x_j + d/2)^2 \tag{15}$$

$$r_2 = (z_j + h/2)^2 + (x_i - x_j + d/2)^2 \tag{16}$$

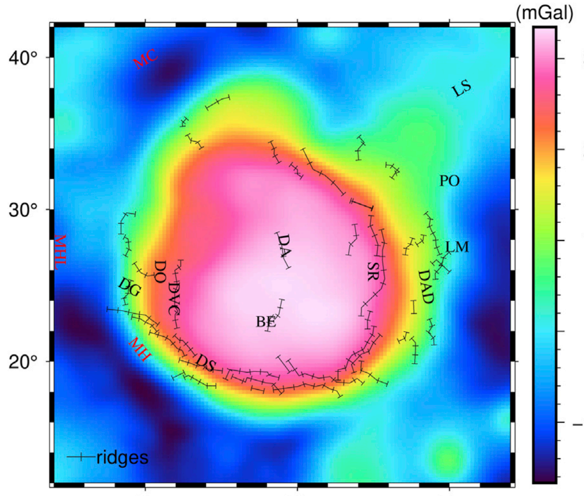


FIGURE 2
Bouguer gravity anomaly of the mare serenitatis.

$$r_3 = (z_j - h/2)^2 + (x_i - x_j - d/2)^2 \tag{17}$$

$$r_4 = (z_j + h/2)^2 + (x_i - x_j - d/2)^2 \tag{18}$$

$$\theta_1 = \arctan(x_i - x_j + d/2)/(z_j - h/2) \tag{19}$$

$$\theta_2 = \arctan(x_i - x_j + d/2)/(z_j + h/2) \tag{20}$$

$$\theta_3 = \arctan(x_i - x_j - d/2)/(z_j - h/2) \tag{21}$$

$$\theta_4 = \arctan(x_i - x_j - d/2)/(z_j + h/2) \tag{22}$$

where x_i is the coordinate of g_i, (x_j, z_j) is the coordinate of jth block, d and h are the length and width of each block. According to the least square method, the profile density anomaly vector V is solved by:

$$V = W_v^{-1}A^T\left(AW_v^{-1}A^T + W_e^{-1}\right)^{-1}g \tag{23}$$

where A is the coefficient matrix of a_{ij}, W_v is the density weighting matrix, and W_e is the noise weighting matrix. These two weight matrices are shown in Eqs 24, 25, respectively.

$$W_v^{-1} = V^2 + \varepsilon \tag{24}$$

$$W_e^{-1} = \ell_0^2 AW_v^{-1}A^T \tag{25}$$

where ε is taken as 10^{-8} and ℓ_0 is the ratio of estimated noise to signal (initial value is 0.1).

3 Results and analysis

3.1 Bouguer gravity anomaly in the mare serenitatis

Figure 2 shows the Bouguer gravity anomaly at the elevation of 0 km with the spatial resolution of $0.25^\circ \times 0.25^\circ$ in the Mare Serenitatis area using Eq. 1, where GM product is 4902.801374×10^9 km^3/s^2 and R is 1738 km (Liang, 2010).

In Figure 2, the Bouguer gravity anomalies range from −186 mGal to 336 mGal. Compared to the surrounding area, the gravity anomaly of the Mare Serenitatis Basin

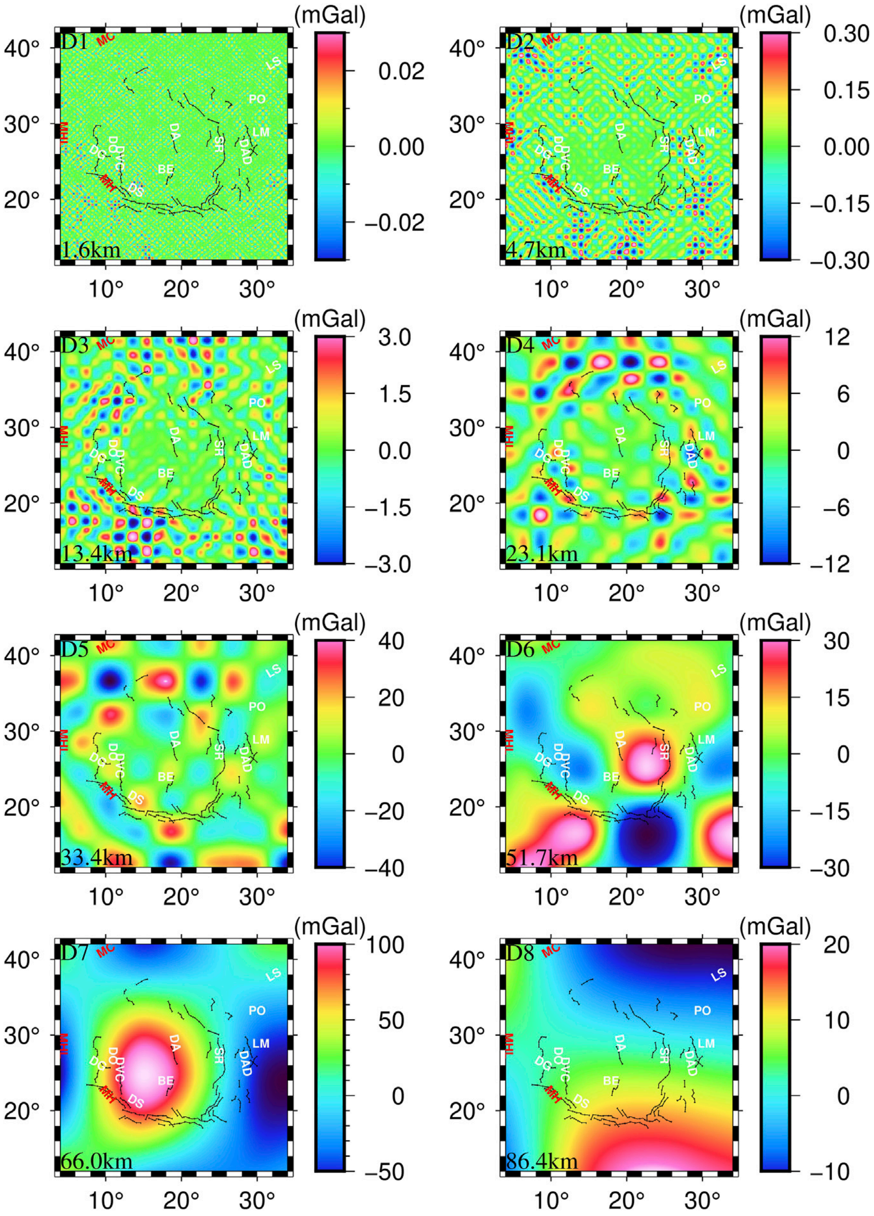


FIGURE 3
Decomposed gravity anomalies D1–D8 in the Mare Serenitatis, whose mean field source depths estimated by Eq. 3 have marked on the left bottom of each figure.

(9°E–29°E and 17°N–36°N) is positive and high, which is pear-shaped. The gravity anomaly is the highest in the center of the Mare Serenitatis Basin and gradually decreases in all directions, showing a clear Mascon character. In addition, there is a negative correlation between the Bouguer gravity anomaly and the topography (Figure 1). The gravity anomaly highs in the center of the Mare Serenitatis correspond to lowlands with the elevation of approximately −5 km. However, the gravity anomaly lows are consistent with the mountains, such as the MC, the MHL and the MH. Moreover, the ridges are primarily located at the gravity high-low transitional zones, such as the DG, the DS and the SR.

3.2 Decomposed Bouguer gravity anomaly

The Bouguer gravity anomaly is the comprehensive reflection of all subsurface anomalous materials. It cannot directly reflect the distribution of the material at different depths. Thus, the Bouguer gravity anomaly is further separated by wavelet multi-scale method. According to the previous research on the selection of the optimal wavelet basis (Xu et al., 2017), this paper adopts the “coif3” wavelet basis and performs a 2D wavelet decomposition for the Bouguer gravity anomaly of the Mare Serenitatis using Eq. 2. The decomposed results are shown in Figure 3, and their mean field source depths estimated by Eq. 3 are shown in Figure 4 and Table 1. In the subplots of

TABLE 1 The layered model in the mare serenitatis.

Layer	Range of depth (km)	Mean field source depth (km)	Thickness (km)
D1	0.0~3.2	1.6	3.2
D2	3.2~6.2	4.7	3.0
D3	6.2~20.6	13.4	14.4
D4	20.6~25.6	23.1	5.0
D5	25.6~41.2	33.4	15.6
D6	41.2~62.2	51.7	21.0
D7	62.2~69.8	66.0	7.6
D8	69.8~103.0	86.4	33.2

Figure 3, D1–D8 represent the order one to eight wavelet details, respectively. The subplots in Fig. 4 are the corresponding radial logarithmic power spectrum of wavelet detail D1–D8, respectively.

According to Figure 3, the gravity anomalies in D1 are very weak and may be interference signals. The gravity anomaly values of D2 vary slightly from −0.3 mGal to 0.3 mGal, whose average source depth is approximately 4.7 km corresponding to the distribution of shallow sediments. Gravity anomaly values of D3 with the average source depth of 13.4 km range from −3.0 mGal to 3.0 mGal. Small positive and negative alternating gravity anomaly circles are beginning to appear, obviously in the south of DS. The average source depth of D4 is approximately 23.1 km and the gravity anomaly values range from −12.0 mGal to 12.0 mGal. The gravity anomaly circles are further expanded, which are mainly distributed along the outer ring of the Mare Serenitatis, such as DO, DS and DAD. It indicates that tectonic structure begins to become complex. In addition, there are apparent gravity anomaly signals during 15°E–25°E and 35°N–42°N, which is highly consistent with the location of the center of the “D ridge system” mentioned by Maxwell et al. (1975). This could be the signals of anomalous material left over from the meteorite event that formed the “D ridge system”. Based on the lunar crust thickness model of Wieczorek et al. (2013), the mean source depth of D4 should be roughly at the interface between the lunar crust and mantle in the Mare Serenitatis region. The gravity anomaly values of D5 with the average source depth of 33.4 km range from −40 mGal to 40 mGal. The gravity anomaly circles are larger. The mean source depth of D6, with gravity anomaly values ranging from −30 mGal to 30 mGal, is approximately 51.7 km. During the 20°E–27°E and 23°N–30°N corresponding to the inner ring region between DA and SR (Maxwell et al., 1975), there is the first large anomaly in the Mare Serenitatis Basin. The center of this large anomaly is at approximately 23°E and 25°N. The D7 shows the distribution of gravity anomalies with the average source depth of 66 km. The values range from −50 mGal to 100 mGal, which is the largest variation ranges for D1–D8. There is another obvious gravity anomaly high, the second large anomaly, whose center is located at 15°E and 25°N. Compared with the first large anomaly mentioned in D6, the second large anomaly in this area is stronger, deeper and more widespread. Therefore, this paper agrees with Watters and Konopliv (2001) that the Mare Serenitatis consists of two overlapping basins. The western basin corresponds to a larger impact forming the outer ring and the eastern basin is in agreement with the inner ring structure of the Mare Serenitatis Basin. The mean source depth of D8 is approximately 86.4 km and the gravity anomaly values range from −10 mGal to 20 mGal. It can be seen that there are no longer any obvious gravity anomaly circles, and the anomaly signals gradually weaken and disappear. It indicates that the tectonic structure at this depth tends to be stable.

According to above analysis, D1–D3 primarily reflects the gravity anomaly distribution of the lunar crust in the Mare Serenitatis area. With depth increasing, the gravity anomaly circles and magnitude gradually become larger. D4—D5 show obvious positive-negative alternating gravity anomalies, indicating the complex tectonic structure. D6—D7 present the morphological characteristics of the two large gravity anomaly highs in the southeast and southwest, which are presumed to be closely related to the inner and outer ring structure of the Mare Serenitatis. D8 indicates that there are no more obvious anomalies at the depth of 86.4 km.

3.3 Inverted layered density results

Firstly, the lunar crust and upper mantle of the Mare Serenitatis is layered according to the corresponding mean field source depths of the decomposed Bouguer gravity anomalies, see Table 1. Then, each layer is gridded and modelled using Tesseroids and the size of each Tesseroid is $0.5^{\circ} \times 0.5^{\circ}$ ($\Delta\varphi = 0.5^{0}$; $\Delta\lambda = 0.5^{0}$ in Eq. 4. Lastly, the density distribution at the corresponding depth was calculated using Eq. 10 and the results are shown in Figure 5.

The results of D1 and D2 are not presented in Figure 5 because of their small density fluctuations. D3 presents the density distribution at the depth of 13.4 km underneath the Mare Serenitatis. The density values range from 2.88 g/cm3 to 2.92 g/cm3. The circles of density highs and lows begin to appear mainly around the Mare Serenitatis basin. D4 corresponds to the density distribution at the depth of 23.1 km, which is approximately at the interface between the lunar crust and mantle in the Mare Serenitatis Basin. The values of density distribution range from 2.90 g/cm^3 to 3.10 g/cm^3. Compared with D3, the circles of density highs and lows become larger. It is noteworthy that in the 15°E–25°E and 35°N–42°N (the center of the aforementioned “D ridge system”), there are several obvious density high circles, which are presumed to be meteorite impact residue. Besides, there is a small high-density circle at (30°E, 27°N), which may be caused by the meteorite impact corresponding to LS in Figure 1. It indicates that meteorite impacts may induce density heterogeneity, which has been also discussed by Kierfer et al. (2012). The density distribution of D5 reflects the materials at the

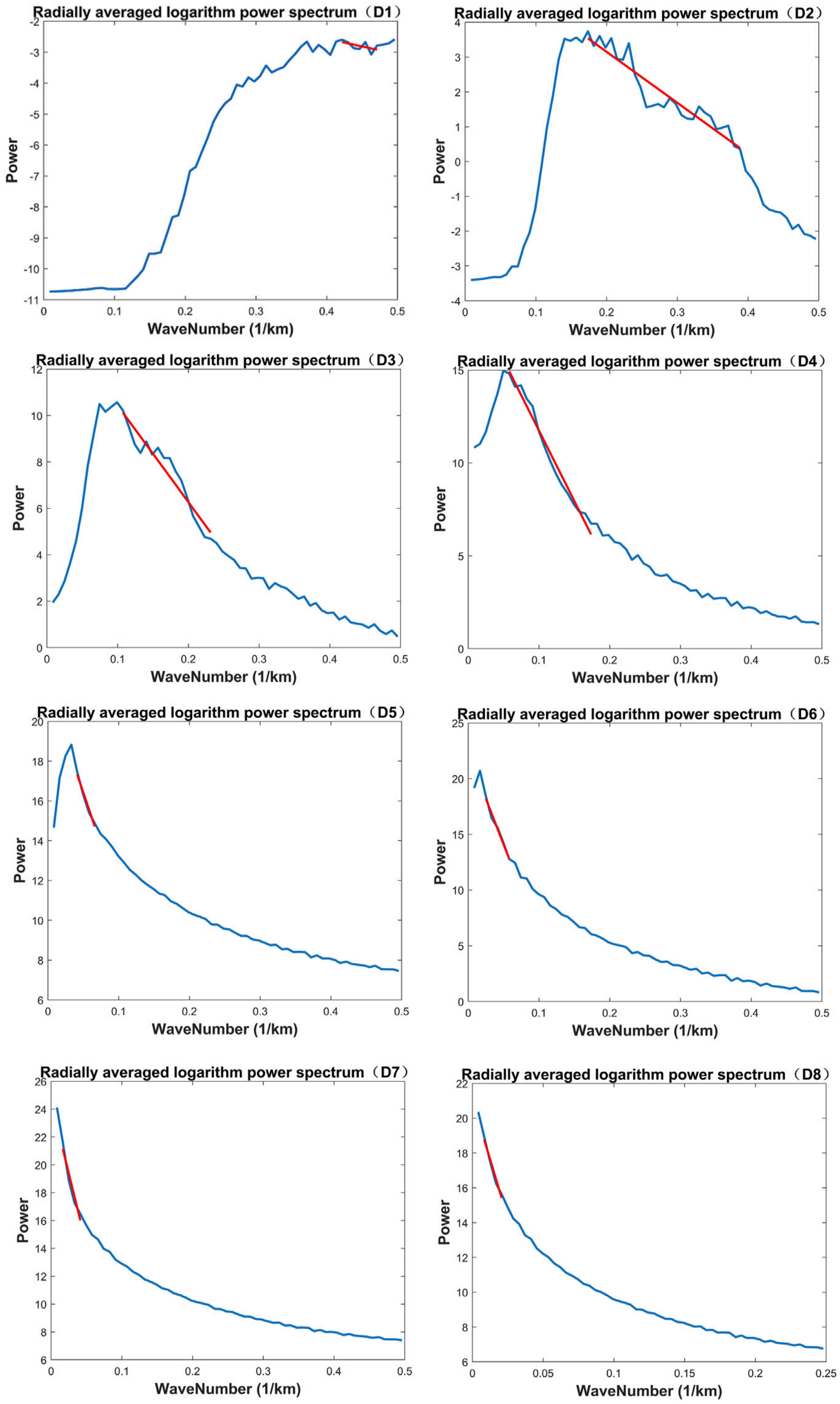


FIGURE 4
Radial logarithm power spectrum of the decomposed gravity anomalies D1—D8 in the Mare Serenitatis. The estimated mean field source depths of D1—D8 are shown in Table 1 and on the left bottom of each corresponding figure in Figure 3.

depth of 33.4 km. The high-density areas are discrete, implying the complex tectonic structure. The D6 corresponds to the density distribution at the depth of 51.7 km beneath the Mare Serenitatis. There is an obvious high density at (23°E, 25°N), which may be the result of a meteorite impact event forming the inner ring structure. D7 corresponds to the density distribution at the depth of 66 km. The whole Mare S. basin is nearly covered by a high-density material, whose center is at (15°E, 25°N). This is the second high density circles within the Mare Serenitatis, and it is presumed to be closely related to the outer ring structure of the Mare Serenitatis. The density distribution of D8 with the depth of 86.4 km has become smooth, indicating that the tectonic structure is beginning to stabilize.

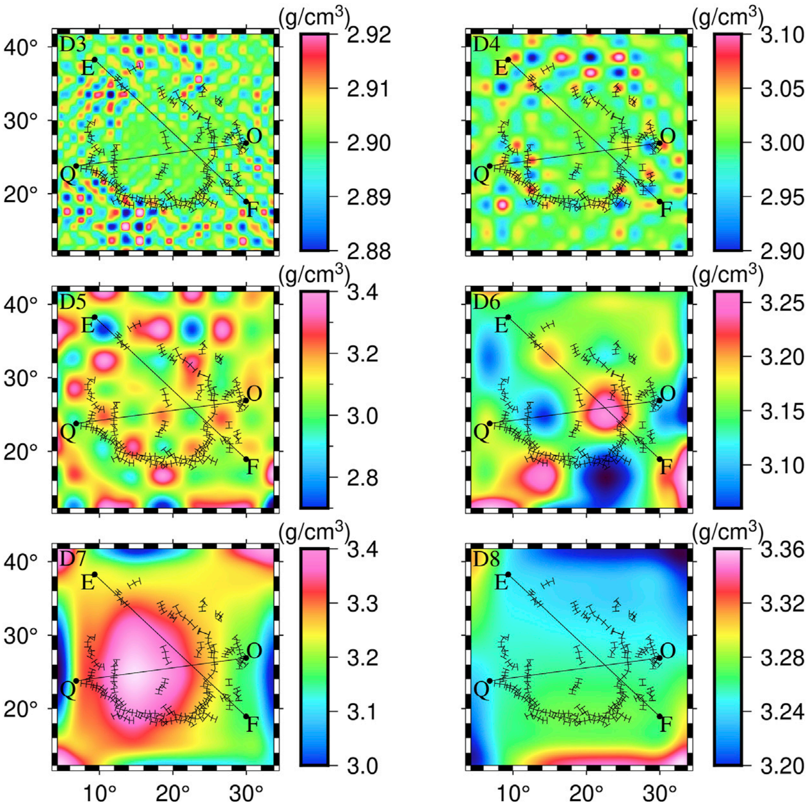


FIGURE 5
Results of the layered density inversion in the Mare Serenitatis. E-F and O-Q are the location of profile density inversion.

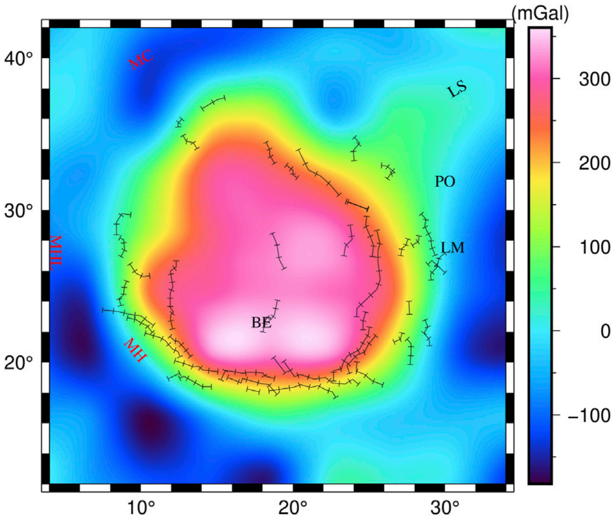


FIGURE 6
The 4th-order wavelet approximation A4 of the Bouguer gravity anomaly.

In summary, there are two high-density bodies with different locations, depths and sizes in the Mare Serenitatis Basin. The center and mean depth of the southeastern high-density body is (23°E, 25°N) and 51.7 km, respectively. The southwestern one is centered at (15°E, 25°N) and has the mean depth of 66 km. These two high-density materials are presumed to be closely related to the inner and outer ring structure of the Mare Serenitatis (Maxwell et al., 1975; Watters and Konopliv, 2001).

3.4 Determined crust-mantle interface depth

According to the a priori information (Hikida and Wieczorek, 2007; Wieczorek et al., 2013), it is known that the average crust-mantle interface depth in the Mare Serenitatis region is approximately 25 km, which is close to the average field source depth of D4 in Figure 3. Thus, the 4th-order wavelet approximation A4 (see Figure 6) of the Bouguer gravity anomaly is regarded as the signal of crust-mantle interface relief, and then the crust-mantle interface depth of the Mare S. region is determined by the iterative inversion of Eq. 11, where the average depth and density contrast of crust-mantle interface are set as 25 km and 0.56 g/cm^3, respectively (Hikida and Wieczorek, 2007; Wieczorek et al., 2013). Figure 7 shows the determined crust-mantle interface depth underneath the Mare Serenitatis.

On the whole, the crust-mantle interface depth gradually shallows from the outside to the inside of the Mare Serenitatis, see Figure 7. The depth in MC, MHL and MH is approximately 35 km, and that in LS, PO and LM is between 20 and 26 km. The depths of the marginal area (green part in Figure 7) of the Mare Serenitatis are approximately 22 km. Further, the light blue area has the average depth of approximately 15 km, and the dark blue area has the average depth

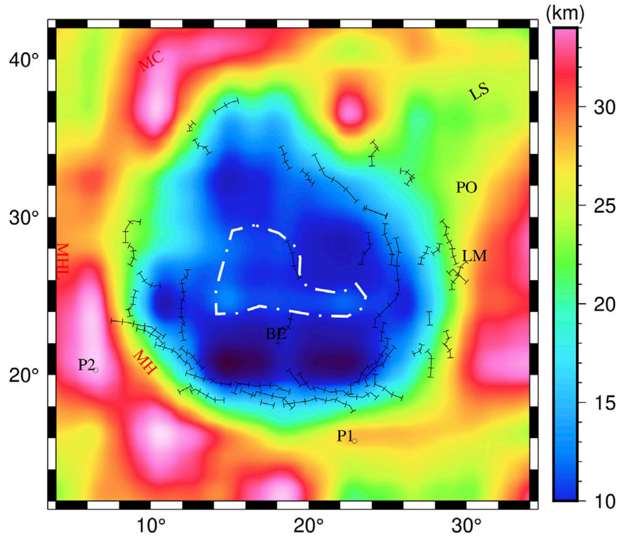


FIGURE 7
Inverted crust-mantle interface depth beneath the Mare Serenitatis. The white dashed lines represent the inferred falling *area* of the crust-mantle interface uplift. P1 and P2 are the locations of lunar seismic inversion from Chenet et al. (2006).

of about 10 km. The shallowest depth, located at the south of the Mare Serenitatis Basin (east and west of BE), is only approximately 4 km, whose positions are correspondence with the two high-density materials aforementioned.

According to above analysis, the crust-mantle interface of the Mare Serenitatis is obviously uplifted, which is in good agreement with the result of Wieczorek et al. (2013). The reason may be the post-impact rebound after a large meteorite impact event, which generated thermal energy perturbation leading to lateral temperature differences in the lunar mantle, followed by the rapid and uniform uplift of the mantle (Wieczorek and Phillips, 1999; Chen et al., 2009). In addition, there is a new finding that the uplift of crust-mantle interface falls back in the center of the Mare Serenitatis, see the white dashed lines in Figure 7. The deepest depth surrounded by the white dashed lines is approximately 17 km, which is deeper than the depth of the dark blue area. It suggests that the uplift of the crust-mantle interface is irregular. Combined with the results of layered density inversion (Figure 5), it further demonstrates that the Mascon in the Mare Serenitatis area may be formed by the combination effect of the high-density materials and crust-mantle interface uplift.

Further, we extracted the lunar seismic inversion results of P1and P2 (Figure 7) from Chenet et al. (2006) and compared them with the crust-mantle interface depth in this study, as shown in Table 2. The difference of crust-mantle interface depths between Chenet et al. (2006) and this study is 0.1 km at P1 and 3.75 km at P2, respectively. The inverted crust-mantle interface depths in this study are in agreement with those provided by Chenet et al. (2006), which verifies the correctness of our results.

3.5 Inverted profile density

Firstly, a simple synthetic test is designed to verify the correctness of the compact gravity inversion method proposed by Last and Kubik (1983), which will be applied to invert the profile density of the lunar crust and upper mantle in the Mare Serenitatis. In the synthetic test, we set up several modules with different shapes and depths, as shown in the lower subplot of Figure 8, where each module (deep red areas) has a relative density value of 0.3 g/cm^3 and other areas (white regions) have a relative density value of 0 g/cm^3. Based on this model, we calculate the gravity anomaly (Gmodel), as shown at the upper subplot (red line) of Figure 8. Then, this calculated gravity anomaly is considered as input (Gobs in the upper subplot of Figure 9), and the compact gravity inversion method is employed to inverted the profile density. The inverted results using Eq. 23 are shown in the lower subplot of Figure 9. The inverted density distribution in Figure 9 is in good agreement with the original model of Figure 8, which demonstrates that the primary density anomalous bodies can be recovered effectively by the compact gravity inversion method. In addition, the RMS of the difference between Gmodel and Gobs is 0.0004 mGal. Thus, *it indicates that t*he compact gravity inversion method is correct according to the *synthetic results*. It is also worth noting that there are some differences between the lower subplot of Figures 8, 9, such as the light blue areas in Figure 9. The reason may be that the shape of synthetic modules in Figure 8 is regular and the values of density setting are not continued, which may cause signal leakage and distortion during the inversion.

Subsequently, the compact gravity inversion method is used to invert the density distribution of two intersecting profiles, E-F and Q-O (as shown in Figure 5), in order to further reveal the detailed morphological characteristics of two high-density materials in the Mare Serenitatis. E-F and Q-O just go through the centers of two high-density materials, whose profile density anomalies (anomalies relative to the average surrounding density) are inverted as shown in Figures 10, 11. During the profile density inversion, the length and width of each cell module are set as 20 km and 5 km respectively, and the total depth of the inversion is 100 km (the density anomalies at deeper depths are no longer apparent). Two subplots are included in Figures 10, 11. The upper subplot shows the comparison between Gmodel and Gobs, and the RMS of the difference between Gmodel and Gobs for E-F and Q-O is 3.886 mGal and 2.4706 mGal, respectively. It indicates that the inverted profile density anomalies are reliable. The lower subplot shows the profile density anomaly distribution.

In Figure 10, there is an obvious large-scale high-density anomalous material, which is located at the distance of

TABLE 2 Crust-mantle interface depth comparison from various researches in the mare serenitatis.

Site	Latitude (degree)	Longitude (degree)	Crust-mantle interface depth (km)	
			Chenet et al. (2006)	This study
P1	15.8	22.9	27.9 ± 17.1	27.80
P2	20.3	6.5	32.4 ± 17.1	36.15

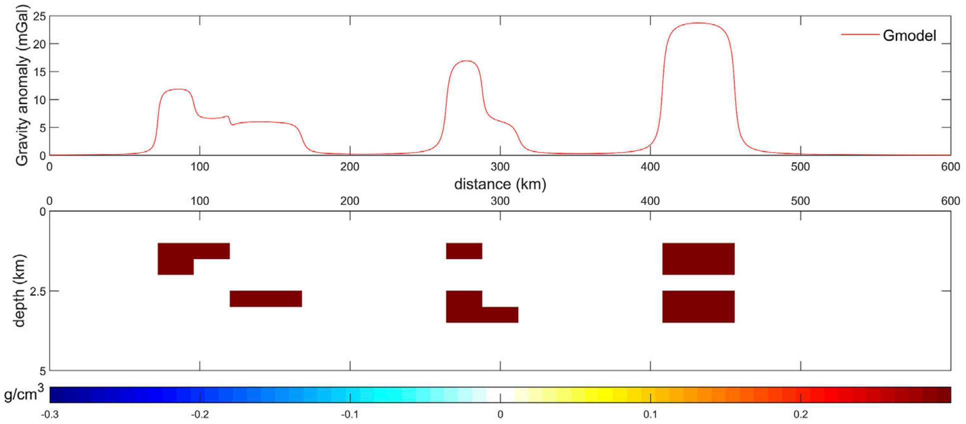


FIGURE 8
Synthetic model and forward gravity anomaly. Gmodel: calculated gravity anomaly from the synthetic model.

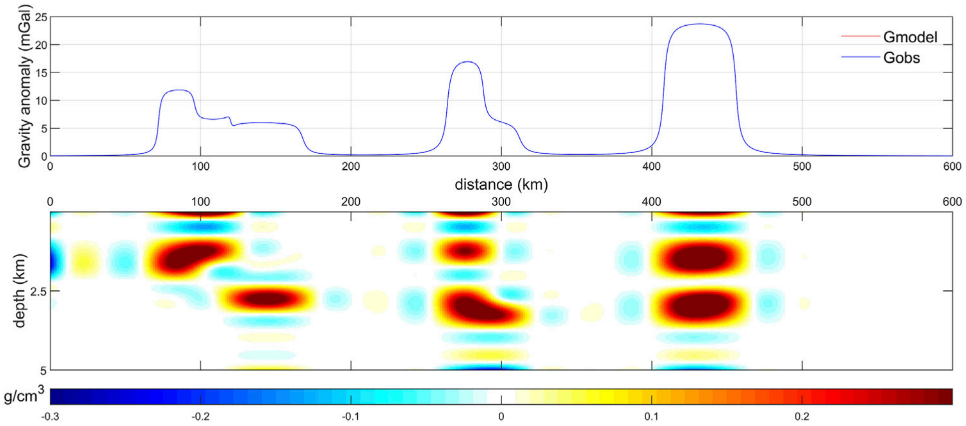


FIGURE 9
Inverted profile density distribution from the gravity anomaly. Gobs: inputted gravity anomaly.

430 km–530 km from Point E with the depth of 30 km–60 km. It is the first high-density anomalous body mentioned in the Figure 5. In Figure 11, there are two obvious larger-scale high-density anomalous materials. The left one is located at 110 km–260 km from Point Q with the depth of 50 km–80 km, which is the second high-density anomalous body mentioned in the Figure 5. The right one is consistent with the first high-density anomalous body presented in Figure 10. In addition, the falling phenomenon of the lunar crust-mantle interface uplift in the center Mare Serenitatis can be observed clearly both in the Figures 10, 11. Above the two high-density anomalous materials, there are apparent low-density materials, which may be formed by the filled magma after meteorite impacts.

Lastly, we compare the inverted profile density (Figures 10, 11) with the results of the same profiles from the inverted layered density (Figure 5) as shown in Figure 12 and Figure 13. Comparing the results in Figures 10, 12, the locations of the first high-density anomalous body are similar, which are both at the distance of approximately 500 km from Point E with the depth of approximately 50 km. Above the first high-density anomalous body, there are both low-density materials in Figures 10, 12. Comparing the results in Figures 11, 13, the locations of the second high-density anomalous body are also consistent, which are both at approximately 200 km from Point Q with the depth of approximately 70 km. Above the second high-density anomalous body, there are also both low-density materials in Figures 11, 13. Thus, on the whole, the inverted profile density (Figures 10, 11) and the inverted layered density (Figures 12, 13) can match well at the large-scale features. However, there are also some apparent differences in details between the inverted profile density (Figures 10, 11) and the inverted layered density (Figures 12, 13). For example, there is an obvious low-density anomaly at the distance of 0–110 km from Point E with the depth of 25–40 km in Figure 12, while it is not existed in Figure 10. The reason may be that the adopted methods with the assumptions of the profile density inversion and layered density inversion are all different. According to above analysis,

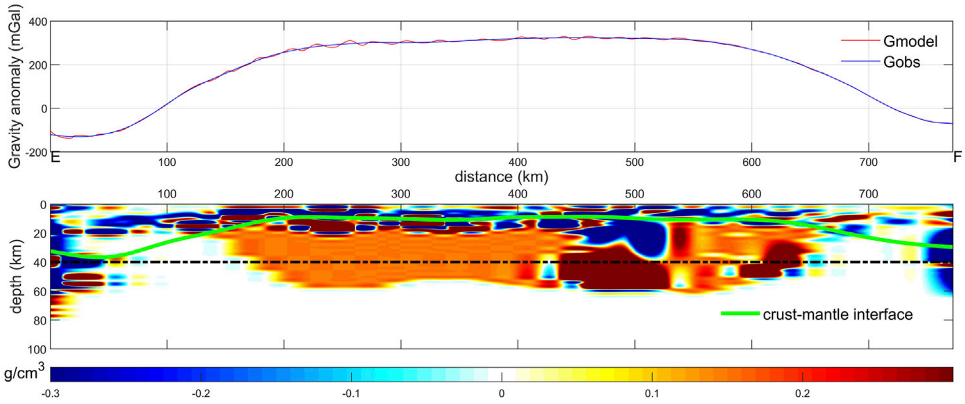


FIGURE 10
Inverted profile density anomaly distribution of E-F in the Mare Serenitatis. The green solid line is the lunar crust-mantle interface. The black dashed line represents the depth of 40 km, which is for assistant analysis.

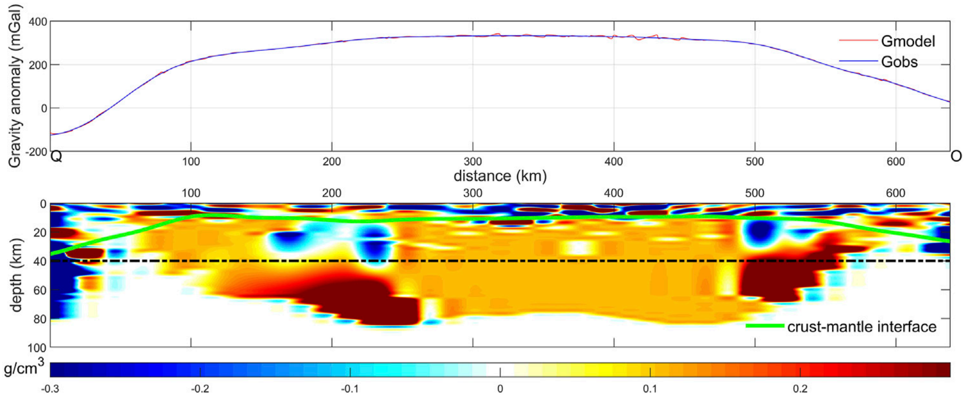


FIGURE 11
Inverted profile density anomaly distribution of Q-O in the Mare Serenitatis.

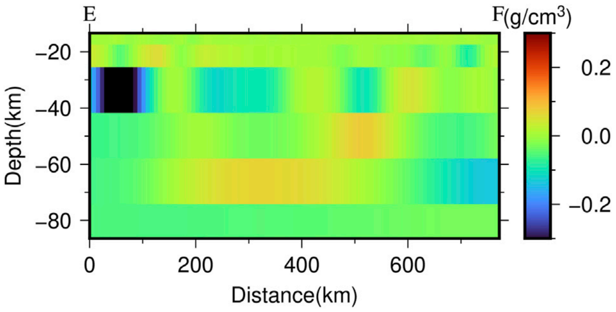


FIGURE 12
Profile density of E-F from the layered density inversion in Figure 5.

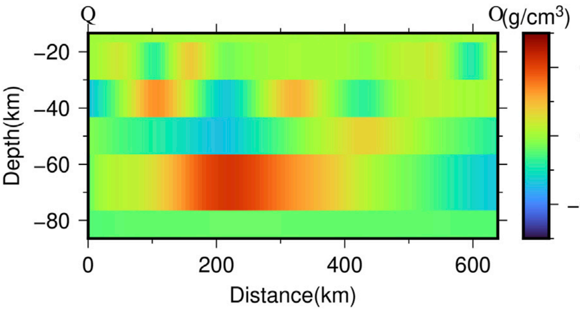


FIGURE 13
Profile density of Q-O from the layered density inversion in Figure 5.

the inverted profile density and the inverted layered density can complement each other, which may provide more information for understanding the lunar crust and upper mantle structure in the Mare Serenitatis.

4 Discussion

The layered density inversion and profile density inversion results reveal the existence of two high-density anomalous bodies with different locations and depths, which are highly related with the inner and outer ring structure of Mare Serenitatis. The first high-density anomalous body located in the southeast corner of the Mare Serenitatis Basin has an approximate depth range of 30 km–60 km and a maximum diameter of approximately 100 km, and its center is at roughly (23°E, 25°N). The second high-density anomalous body, located in the south-western corner of the Mare Serenitatis Basin with its center at approximately (15°E, 25°N), has an approximate depth range of 50 km–80 km and a maximum diameter of approximately 150 km. Liang et al. (2014) and Zhao et al. (2021) have suggested that the high-density anomalies in the interior of the Mare Serenitatis Basin are located at the depth range of 20–50 km. The difference is that we have revealed the 3-D morphological characteristics of the two high-density materials, including their central location and diameter, which have not been mentioned in the previous studies.

In addition, the depth of the crust-mantle interface reveals that there is obvious crust-mantle interface uplift in the interior of the Mare Serenitatis Basin, which is in agreement with the results of Wieczorek et al. (2013) and Hikida and Wieczorek (2007). The difference is that the results of this paper show that the trend of crust-mantle interface uplift within the basin is not smooth (the uplift in the center of the Mare Serenitatis Basin has fallen significantly back), and the most obvious crust-mantle interface uplift exists at the location of the two high-density anomalous bodies. In conjunction with this phenomenon, it is reasonable to suggest that there is a strong correlation between the high-density anomalies and the crust-mantle interface uplift.

Therefore, we propose a bold hypothesis that the Mare Serenitatis was hit by at least two major meteorite impacts. The first impact was more intense and greater, occurring mainly in the southwest corner of the Mare Serenitatis Basin, which largely contributed to the formation of the outer ring structure. The second impact occurred mainly in the southeast corner of the Mare Serenitatis Basin, which may contribute to the formation of the inner ring structure.

5 Conclusion

In this paper, gravity multi-scale analysis is applied to invert the layered density structure and determine the depth of the crust-mantle interface in the Mare Serenitatis region. Firstly, according to the decomposed Bouguer gravity anomalies and their corresponding field source depths, the distribution of gravity anomalies generated by material at different depths is various. Then, the results of the layered density inversion further reveal that the lunar upper crust in the Mare Serenitatis has some lateral density distribution inhomogeneity. It is further extended in the lunar middle and lower crust. The lunar lower crust to the upper mantle contains two high-density anomalous bodies. At 86.4 km depth, there is almost no lateral density variation. Subsequently, the results of the crust-mantle interface inversion show that the crust-mantle interface of the Mare Serenitatis Basin is significantly uplifted and the shallowest crust-mantle interface is around 4 km depth. Besides, there is an obvious falling phenomenon for the uplift in the center of the basin. Lastly, the density anomaly inversion results of the two profiles further present the characteristics of the vertical density anomaly variation. The combination of the layered density, the crust-mantle interface and the profile density anomaly results suggest that the formation of the Mascon in the Mare Serenitatis may be due to the combined effect of the internal high-density anomalous bodies and crust-mantle interface uplift.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

Conceptualization, CX; methodology, HY and CX; software, HY; validation, HY and JL; formal analysis, CX; investigation, GJ and MX; data curation, JL and HY; writing—original draft preparation, HY and CX; writing—review and editing, CX and YW; visualization, MX; supervision, CX; project administration, CX; funding acquisition, CX. All authors have read and agreed to the published version of the manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (Grant nos. 91428205, 41974014, 42274004) and the Natural Science Foundation of Guangdong Province, China (Grant no. 2022A1515010396).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Chen, C., Chen, B., Ping, J., Liang, Q., Huang, Q., Zhao, W., et al. (2009). The interpretation of gravity anomaly on lunar Apennines. *Sci. China Ser. G Phys. Mech. Astronomy* 52 (12), 1824–1832. doi:10.1007/s11433-009-0281-0

Chenet, H., Lognonné, P., Wieczorek, M., and Mizutani, H. (2006). Lateral variations of lunar crustal thickness from the Apollo seismic data set. *Earth Planet. Sci. Lett.* 243 (1-2), 1–14. doi:10.1016/j.epsl.2005.12.017

Goossens, S., Lemoine, F. G., Sabaka, T. J., Nicholas, J. B., Mazarico, E., Rowlands, D. D., et al. (2016). "A global degree and order 1200 model of the lunar gravity field using grail mission data," in *47th annual lunar and planetary science conference 1903*, 1484.

Hansen, P. C., and O'Leary, D. P. (1993). The use of the L-curve in the regularization of discrete Ill-Posed problems. *SIAM J. Sci. Comput.* 14 (6), 1487–1503. doi:10.1137/0914086

Head, J. W. (1979). Serenitatis multi-ringed basin: Regional geology and basin ring interpretation. *moon planets* 21 (4), 439–462. doi:10.1007/BF00897836

Heck, B., and Seitz, K. (2007). A comparison of the tesseroid, prism and point-mass approaches for mass reductions in gravity field modelling. *J. Geodesy* 81 (2), 121–136. doi:10.1007/s00190-006-0094-0

Hiesinger, H., Jaumann, R., Neukum, G., and Head, J. W., III (2000). Ages of mare basalts on the lunar nearside. *J. Geophys. Res. Planets* 105 (E12), 29239–29275. doi:10.1029/2000JE001244

Hikida, H., and Wieczorek, M. A. (2007). Crustal thickness of the Moon: New constraints from gravity inversions using polyhedral shape models. *Icarus* 192 (1), 150–166. doi:10.1016/j.icarus.2007.06.015

Jiang, W., Zhang, J., Tian, T., and Wang, X. (2012). Crustal structure of Chuan-Dian region derived from gravity data and its tectonic implications. *Phys. Earth Planet. Interiors* 212, 76–87. doi:10.1016/j.pepi.2012.07.001

Kaur, P., Bhattacharya, S., Chauhan, P., Kumar, A. K., and Kiran Kumar, A. (2013). Mineralogy of mare Serenitatis on the near side of the moon based on chandrayaan-1 moon mineralogy mapper (M3) observations. *Icarus* 222 (1), 137–148. doi:10.1016/j.icarus.2012.10.020

Kierfer, W. S., Macke, R. J., Britt, D. T., Irving, A. J., and Consolmagno, G. J. (2012). The density and porosity of lunar rocks. *Geophys. Res. Lett.* 39, L07201. doi:10.1029/2012GL051319

Kodama, S., and Yamaguchi, Y. (2003). Lunar mare volcanism in the eastern nearside region derived from Clementine UV/VIS data. *Meteorit. Planet. Sci.* 38 (10), 1461–1484. doi:10.1111/j.1945-5100.2003.tb00251.x

Last, B. J., and Kubik, K. (1983). Compact gravity inversion. *Geophysics* 48 (6), 713–721. doi:10.1190/1.1441501

Lemoine, F. G., Goossens, S., Sabaka, T. J., Nicholas, J. B., Mazarico, E., Rowlands, D. D., et al. (2014). GRGM900C: A degree 900 lunar gravity model from GRAIL primary and extended mission data. *Geophys. Res. Lett.* 41 (10), 3382–3389. doi:10.1002/2014GL060027

Li, B., Ling, Z., Zhang, J., Chen, J., Ni, Y., and Liu, C. (2018). Displacement-length ratios and contractional strains of lunar wrinkle ridges in Mare Serenitatis and Mare Tranquillitatis. *J. Struct. Geol.* 109, 27–37. doi:10.1016/j.jsg.2018.01.003

Liang, Q., Chen, C., and Li, Y. (2014). 3-D inversion of gravity data in spherical coordinates with application to the GRAIL data. *J. Geophys. Res. Planets* 119 (6), 1359–1373. doi:10.1002/2014JE004626

Liang, Q. (2010). *Gravity anomaly features and 3D density imaging of the Moon*. Ph. D. thesis (Wuhan: China University of Geosciences). (in Chinese).

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. pattern analysis Mach. Intell.* 11 (7), 674–693. doi:10.1109/34.192463

Maxwell, T. A., El-Baz, F., and Ward, S. H. (1975). Distribution, morphology, and origin of ridges and arches in Mare Serenitatis. *Geol. Soc. Am. Bull.* 86 (9), 1273–1278. doi:10.1130/0016-7606(1975)86<1273:DMAOOR>2.0.CO;2

Neumann, G. A. (2011). LRO MOON LASER ALTIMETER 4 GDR V1.0 [Data set]. *NASA Planet. Data Syst.* doi:10.17189/1520642

Pieters, C. M. (1978). Mare basalt types on the front side of the moon-A summary of spectral reflectance data. *Lunar Planet. Sci. Conf. Proc.* 9, 2825–2849.

Plescia, J. B., and Golombek, M. P. (1986). Origin of planetary wrinkle ridges based on the study of terrestrial analogs. *Geol. Soc. Am. Bull.* 97 (11), 1289–1299. doi:10.1130/0016-7606(1986)97<1289:OOPWRB>2.0.CO;2

Ryder, G., Norman, M. D., and Taylor, G. J. (1997). The complex stratigraphy of the highland crust in the Serenitatis region of the Moon inferred from mineral fragment chemistry. *Geochimica Cosmochimica Acta* 61 (5), 1083–1105. doi:10.1016/S0016-7037(96)00386-9

Sharpton, V. L., and James, W. H., III (1982). Stratigraphy and structural evolution of southern mare Serenitatis: A reinterpretation based on Apollo lunar sounder experiment data. *J. Geophys. Res. Solid Earth* 87 (B13), 10983–10998. doi:10.1029/JB087iB13p10983

Solomon, S. C., and Head, J. W. (1979). Vertical movement in mare basins: Relation to mare emplacement, basin tectonics, and lunar thermal history. *J. Geophys. Res. Solid Earth* 84 (B4), 1667–1682. doi:10.1029/JB084iB04p01667

Spector, A., and Grant, F. S. (1970). Statistical models for interpreting aeromagnetic data. *Geophysics* 35 (2), 293–302. doi:10.1190/1.1440092

Staudacher, T., Dominik, B., Jessberger, E. K., and Kirsten, T. (1978). Consortium breccia 73255: 40AR-39AR dating. *Lunar Planet. Sci. Conf.* 9, 1098–1100.

Tikhonov, A. N., and Arsenin, V. Y. (1977). *Solutions of ill-posed problems*. New York: John Wiley & Sons.

Wan, J., Xu, C., Luo, Z., Wu, Y., Zhou, B., and Yan, J. (2019). An approach to moho topography Recovery using the on-orbit GOCE gravity gradients and its applications in tibet. *Remote Sens.* 11, 1567. doi:10.3390/rs11131567

Wang, J. G., Zhao, G. Q., and Zhu, G. B. (2009). Contrastive analysis of common computing methods of ultra-high degree and order fully normalized associated Legendre function. *J. Geod. Geodyn.(in Chin.* 29 (2), 126–130.

Watters, T. R., and Konopliv, A. S. (2001). The topography and gravity of mare Serenitatis: Implications for subsidence of the mare surface. *Planet. Space Sci.* 49 (7), 743–748. doi:10.1016/S0032-0633(01)00007-1

Watters, T. R. (1988). Wrinkle ridge assemblages on the terrestrial planets. *J. Geophys. Res. Solid Earth* 93 (B9), 10236–10254. doi:10.1029/JB093iB09p10236

Wieczorek, M. A., Neumann, G. A., Nimmo, F., Kiefer, W. S., Taylor, G. J., Melosh, H. J., et al. (2013). The crust of the Moon as seen by GRAIL. *Science* 339 (6120), 671–675. doi:10.1126/science.1231530

Wieczorek, M. A., and Phillips, R. J. (1999). Lunar multiring basins and the cratering process. *Icarus* 139 (2), 246–259. doi:10.1006/icar.1999.6102

Wieczorek, M. A., and Phillips, R. J. (1998). Potential anomalies on a sphere: Applications to the thickness of the lunar crust. *J. Geophys. Res. Planets* 103 (E1), 1715–1724. doi:10.1029/97JE03136

Wilhelms, D. E., McCauley, J. F., and Trask, N. J. (1987). *The geologic history of the Moon*. (No. 1348).

Xu, C., Liu, Z., Luo, Z., Wu, Y., and Wang, H. (2017). Moho topography of the Tibetan Plateau using multi-scale gravity analysis and its tectonic implications. *J. Asian Earth Sci.* 138, 378–386. doi:10.1016/j.jseaes.2017.02.028

Xu, C., Luo, Z., Sun, R., Zhou, H., and Wu, Y. (2018). Multilayer densities using a wavelet-based gravity method and their tectonic implications beneath the Tibetan Plateau. *Geophys. J. Int.* 213 (3), 2085–2095. doi:10.1093/GJI/GGY110

Zhao, G., Liu, J., Chen, B., Kaban, M. K., and Du, J. (2021). 3-D density structure of the lunar Mascon basins revealed by a high-efficient gravity inversion of the GRAIL data. *J. Geophys. Res. Planets* 126 (5), e2021JE006841. doi:10.1029/2021JE006841

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Original Research
PUBLISHED 30 March 2023
DOI 10.3389/fspas.2023.1158138

OPEN ACCESS

EDITED BY
Changyi Xu,
Institute of Geology and Geophysics (CAS), China

REVIEWED BY
H. Yan,
Innovation Academy for Precision Measurement Science and Technology (CAS), China
Yu Sun,
Fuzhou University, China

*CORRESPONDENCE
Hao Ding,
✉ dhaosgg@sgg.whu.edu.cn

SPECIALTY SECTION
This article was submitted to Planetary Science, a section of the journal Frontiers in Astronomy and Space Sciences

RECEIVED 03 February 2023
ACCEPTED 16 March 2023
PUBLISHED 30 March 2023

CITATION
Shi K, Ding H, Chen T and Zou C (2023), The numerical prediction of the Earth's polar motion based on an advanced multivariate algorithm.
Front. Astron. Space Sci. 10:1158138.
doi: 10.3389/fspas.2023.1158138

The numerical prediction of the Earth's polar motion based on an advanced multivariate algorithm

Kunpeng Shi, Hao Ding*, Tao Chen and Chuanyi Zou

School of Geodesy and Geomatics, Hubei Luojia Laboratory, Wuhan University, Wuhan, China

Since there are complicated changes in the polar motion (PM) from sub-annual to decadal, precisely predicting it is challenging. Here, we provide an advanced multivariate algorithm by combining an iterative oblique singular spectrum analysis (IOSSA) with pseudo data (IOSSApd) and consider more periodic and quasi-periodic signals, especially long-period oscillations (Ding et al., Geophys. Res. Lett., 2019, 46, 13765–13774) and multi-frequency Chandler wobble (Pan, International Journal of Geosciences, 2012, 3, 930–951), than previous studies. The IOSSA in oblique coordinates, due to its weak separability conditions, has a better separation performance than general singular spectrum analysis (SSA), and the IOSSApd approach further solved the shift problem. Upon using the IOSSApd method, the PM data can be separated into deterministic and stochastic components, extrapolated by the multiple-harmonic (MH) and autoregressive integrated moving average (ARIMA) models, respectively. Based on the IERS EOPC04 PM series, we produced multiple sets of PM predictions with a 1-year leading time and reported the IERS Bulletin A predictions as a comparison. For 90-day leading time predictions, the mean absolute errors (MAEs) of the x- and y-components were 7.69 and 5.12 mas, respectively, while the corresponding MAEs obtained by IERS Bulletin A were 9.45 and 5.69 mas, respectively. For up to 360 days, our algorithm obtains the MAEs of PM slowly accumulating to 12.98 mas on average, far better than the 19.14 mas for Bulletin A's predictions (also significantly superior to the corresponding results given by previous studies). The prediction performance in the middle- and long-term prediction is further compared against the general SSA predictor. By virtue of weak periodic error, our results show that combining the IOSSApd + MH + ARIMA models improved the prediction success rate up to 75.39% and 69.58% for the x- and y-component, respectively.

KEYWORDS

polar motion prediction, iterative oblique SSA, pseudo data, multiple-harmonic model, autoregressive integrated moving average

1 Introduction

Polar motion is the variation of the orientation of the Earth's rotation axis relative to the terrestrial reference frame, which has two components x and y ($\mathbf{m} = x{-}iy$), and x along the Greenwich Meridian and y along the 90 °W longitude. There are two main periodic signals contained in the PM, i.e., the Chandler wobble (CW, Chandler (1981)) with a ~432-day period and the annual wobble (AW) with a 1-year period. The study of PM provides valuable information for studying many geophysical and meteorological phenomena. They directly link celestial and terrestrial reference frames (transformation between CRF and TRF), and an accurate PM model is a guarantee in high-precision space geodesy. Many studies have focused on the PM short-prediction (1–30 days) because their latency release hinders real-time applications (e.g., GNSS, spacecraft tracking, and astronomy observations) (Barnes et al., 1983; Buffett et al., 1991; Gross et al., 2003; Chen and Wilson 2005; Seitz and Schuh 2010; Wahr 2010). When a satellite navigation system is in autonomous navigation mode, however, the ground station cannot upload the latest EOP, and the system can only rely upon the mid- and long-term EOP prediction. The current forecast data (such as Bulletin-A, the frequently-used product from IERS) will cause an orbit determination error of over 10 m in a 1-year lead (Jia et al., 2018). Therefore, it is equally important to have such predictions available for a longer period, and accurate predictive models are much required.

PM can be present by the sum of two statistically independent parts: determination (trend and undulation) and residual (stochastic). Most of the methods either extract (or estimate) the parameters of harmonic functions and extrapolate into the future (Kosek et al., 1998; 2007; Modiri et al., 2018) or use stochastic methods (Schuh et al., 2002; Akyilmaz and Kutterer 2004; Liao et al., 2012; Yao et al., 2013). Many comparison campaigns, e.g., (EOPPCC, Oct. 2005—Mar. 2008), (WGP, Apr. 2006—Oct. 2009), and (EOPCPPP, Oct. 2010—now), aimed to investigate different strategies and techniques available for predicting EOP data from the short-term to the long-term (Kalarus et al., 2010; Kosek 2012). Among these methods, the combination of the least square (LS) and the autoregressive (AR) processes is considered to be one of the most effective for PM prediction (Kalarus et al., 2010; Xu and Zhou 2015). Unfortunately, no particular method is superior to others for all prediction intervals (Kalarus et al., 2010). A further idea was to coordinate those working on EOP predictions to compare their results using well-defined rules, which is different from many previous (individual) studies. Thus, by weighting the errors, the Bulletin-A files with contributed observations and predictive methods were produced at 7-day intervals (Stamatakos et al., 2011). However, due to the complexity of the PM excitation, it cannot reproduce the time variation of the periodic terms that influence the long-term predictive accuracy of PM (Chao and Chung 2012).

The aforementioned idea led us to discover an interesting fact: the critical factor of unsatisfactory performance in mid-term and long-term predictions may relate to the effective separation and extrapolation of deterministic components including CW, AW, and long-period oscillations. It is well known that the time-varying characteristics of the harmonics contained in PM are very complicated (a famous case is that the Chandler wobble changes from several mas to ~200 mas (Su et al., 2014)). Extracting these deterministic components accurately has become the foremost task. Other issues, including (non) linear trend hypothesis, edge effects, and high-frequency remnants, also need to be considered. We list some main disputes and their specific descriptions are presented here. First, it is unreliable and unscientific for the usual operation to consider the trend of the PM as a simple (non) linear change. Our new finding on the AR-z spectrum has revealed some long-period geophysical oscillations from interdecadal to even century-scale (see Ding et al. (2019)), such as the Markowitz wobble (~30 years) (Hinderer et al., 1987) and solar-cycle (11 years) (Currie, 1981). We have identified that after decomposition, these low-frequency signals can fit the trend terms completely in the time–frequency domain, implying the possibility of extrapolating non-stationary trends using long-period fluctuations. In addition, various spectral analyses and digital filtering have confirmed the Chandler wobble possesses multiple frequencies regardless of the data lengths, time spans, and intervals of the PM observations (Liu et al., 2000; Pan, 2007). Such Chandler frequency splits respond to the coupled oscillators in the triaxial or axially near-symmetrical Earth and are indeed equivalent to amplitude or frequency modulation (Pan, 2012). In routine practice, the CW was generally considered to have a fixed or time-varying frequency; for high-precision prediction, however, such choices may also introduce deviations (Malkin and Miller 2010; Wang G. et al., 2016). Akyilmaz et al. (2011) further showed that the quasi-periodic and irregular processes in the PM residual need to be settled for prediction. Last, although the edge effects at the end of the PM decomposition series are well known, little attention has been paid to related forecasting (Zhao and Lei 2020).

In this article, we introduce an iterative oblique SSA (IOSSA) method (Golyandina and Shlemov 2013) as a powerful separator for close-frequency components and non-stationary trends. This method performs a new decomposition of a part in the SSA, which corrected the eigenvalues to avoid their possible mixture or disjointed sets of singular values in the group corresponding to different components. It also considers orthogonality, with respect to a non-standard Euclidean inner product, which considerably weakens the separability conditions. Such features will help to distinguish different periodic signals. Through iterations, the IOSSA method can finally separate inner products converging to a stationary point. To address the phase shift issue, the classic LS + AR model is used to provide the initial prediction as pseudo data. Hence, an enhanced IOSSA with pseudo data (IOSSApd) can be established (Wang X. et al., 2016). After precisely stripping the principal components using IOSSApd, the MH (multiple-harmonic) model is introduced next to fully match and extrapolate the deterministic components in their frequency ranges or specific frequency points. Given the shorter-period tidal signals contained in the PM (Gross, 2007), the MH model will be further processed to fit the prior tidal signals from residual (stochastic) components. After those, we use the auto covariance autoregressive integrated moving average (ARIMA) to consider the residual high-frequency terms (sub-seasonal variations or other aperiodic changes) (Sun et al., 2019).

In this study, we will predict the PM for up to 1 year for multiple issues (2010–2021). We will also simultaneously predict the PM for the same time span using the IERS Bulletin A, a general SSA, and other methods for comparison, and this can confirm that our approach will obtain better results with periodic error reduction. In the following section, we will first introduce the used methods, and then use them in the prediction.

2 Methodology

In this section, we will explain how to integrate the pseudo data with the iterative oblique process. Here, the classical SSA and

IOSSApd methods will be described in detail, while their decomposed performance will be shown in Section 4.2.

2.1 Singular spectrum analysis

The singular spectrum analysis is a statistical technique that is a particular application of the empirical orthogonal function (EOF) determined from the dynamic reconstruction of a sequence (Modiri et al., 2020). The SSA method consists of embedding, singular value decomposition, grouping, and diagonal averaging (Vautard and Ghil, 1989).

2.1.1 Embedding

In each daily PM training series, $\{x_T\}$ [for example, if we choose the 1962–2010 time-span and the data-length $T = 17{,}531$ days (~48 years)] is transformed into a multi-dimensional series $X_i = (x_i, \ldots, x_{i+L-1})^{\mathrm{T}}, i = 1, \ldots, K$. Here, the window length is set as $L = 2190$ (the beat period of CW and AW oscillation) and $K = T - L + 1$. The trajectory matrix $\mathbf{X}$ with Hankel structure is then represented as follows:

$$\mathbf{X} = \left[\, X_1 \;\; X_2 \;\; \ldots \;\; X_K \,\right] = \begin{pmatrix} x_1 & x_2 & x_3 & \ldots & x_K \\ x_2 & x_3 & x_4 & \ldots & x_{K+1} \\ x_3 & x_4 & x_5 & \ldots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \ldots \\ x_L & x_{L+1} & x_{L+2} & \ldots & x_T \end{pmatrix}. \tag{1}$$

2.1.2 Singular value decomposition

In this step, the singular value decomposition (SVD) of the trajectory matrix $\mathbf{X}$ is performed as

$$\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_L, \tag{2}$$

where $\mathbf{X}_i = \sigma_i U_i V_i^{\mathrm{T}}$, $\sigma_1 \geq \sigma_2 \ldots \geq \sigma_L \geq 0$ are the singular values, and U_i and V_i are the left and right singular vectors, respectively. For rank one, the $\mathbf{X}_i$ is regarded as an elementary matrix. In addition, we named the elements (σ_i, U_i, V_i) the i_{th} eigentriple of the SVD, which will be used in the description of IOSSA.

2.1.3 Grouping

With the r_{th} largest singular value, the trajectory matrix is partitioned into separating signal and residual groups

$$\mathbf{X} = \overbrace{\mathbf{X}_1 + \cdots + \mathbf{X}_r}^{\text{Signal (Principal)}} + \overbrace{\mathbf{X}_{r+1} + \cdots + \mathbf{X}_L}^{\text{Noise (Remaining)}} = \tilde{\mathbf{X}} + \mathbf{X_0}, \tag{3}$$

where the $\tilde{\mathbf{X}}$ and $\mathbf{X_0}$ is clean and noise subspace, respectively. We divided the indices set $\{1, \ldots, r\} = \coprod_{i=1} J_i$ (for details, refer to Shen et al., 2018) and it led to decomposition into m subsets as

$$\tilde{\mathbf{X}} = \mathbf{X}_{J1} + \ldots + \mathbf{X}_{Jm}. \tag{4}$$

2.1.4 Diagonal averaging

By diagonal averaging, the resultant matrices $\mathbf{X}_{Ji}$ ($i = 1, \ldots, m$) are retransformed into a new reconstruction component (RC) series with length T. Supposing these blocks as the $\mathrm{Z}_{L\times K}$ matrix structure and z_{ij} ($1 \leq i \leq L$, $1 \leq j \leq K$) as the elements of the $\mathrm{Z}_{L\times K}$ matrix, we have $L^* = \min(L, K)$, $K^* = \max(L, K)$, and $T = L + K$—1. Let $z_{ij}{}^* = z_{ij}$ if $L < K$ and $z_{ij}{}^* = z_{ji}$; otherwise, $\mathbf{Z}$ will transfer into the series $z_1, \ldots,$ z_T according to

$$z_k = \begin{cases} \dfrac{1}{k}\displaystyle\sum_{n=1}^{k} z^*_{n,k-n+1}\,, 1 \leq k < L^* \\ \dfrac{1}{L^*}\displaystyle\sum_{n=1}^{L^*} z^*_{n,k-n+1}\,, L^* \leq k \leq K^* \\ \dfrac{1}{T-k+1}\displaystyle\sum_{n=k-k^*+1}^{T-K^*+1} z^*_{n,k-n+1}\,, K^* < k \leq T \end{cases} . \tag{5}$$

2.2 Iterative oblique SSA with pseudo data

An enhanced version of SSA and IOSSA for weakening the separation conditions is detailed in this section. The application of a standard or restricted SVD is the main distinguishing feature between both algorithms (Shaharudin et al., 2019). From the beginning, let us consider a minimal decomposition of $\mathbf{Y}\varepsilon\mathcal{M}_{L,K}$ (of rank r) in the standard SVD form

$$\mathbf{Y} = \sum_{i=1}^{r} \sigma_i P_i Q_i{}^{\mathrm{T}}, \tag{6}$$

where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$, $\{P_i\}_{i=1}^{r}$ and $\{Q_i\}_{i=1}^{r}$ are linearly independent. Eq. 6 can be expressed in the matrix form

$$\mathbf{Y} = \mathbf{P}\boldsymbol{\Sigma}\mathbf{Q}^{\mathrm{T}}, \tag{7}$$

where $\mathbf{P} = [P_1 \colon \ldots \colon P_r]$, $\mathbf{Q} = [Q_1 \colon \ldots \colon Q_r]$ and $\boldsymbol{\Sigma} = diag\,(\sigma_1, \ldots, \sigma_r)$.

Then, we extend the SVD algorithm into restricted SVD (RSVD), named $\mathbf{L}$ and $\mathbf{R}$, and the biorthogonal is given by the triple ($\mathbf{Y}, \mathbf{L}, \mathbf{R}$). Correspondingly, Cholesky decomposition for $\mathbf{L}$ and $\mathbf{R}$ is first performed as

$$\mathbf{L} = \mathbf{O}_{\mathbf{L}}^{\mathrm{T}}\mathbf{O}_{\mathbf{L}}, \tag{8}$$

$$\mathbf{R} = \mathbf{O}_{\mathbf{R}}^{\mathrm{T}}\mathbf{O}_{\mathbf{R}}. \tag{9}$$

Here, $\mathbf{O}_{\mathbf{L}}$ is an orthonormalizing matrix of $\{P_i\}_{i=1}^{r}$ and $\mathbf{O}_{\mathbf{R}}$ is an orthonormalizing matrix of $\{Q_i\}_{i=1}^{r}$, and the ($\mathbf{L}$, $\mathbf{R}$)-SVD can be finally expressed as

$$\mathbf{O}_{\mathbf{L}}\mathbf{Y}\mathbf{O}_{\mathbf{R}}^{\mathrm{T}} = \sum_{i=1}^{r} \sqrt{\lambda_i} U_i V_i^{\mathrm{T}} = \sum_{i=1}^{r} \sigma_i \left(\mathbf{O}_{\mathbf{L}} P_i\right)\left(\mathbf{O}_{\mathbf{R}} Q_i\right)^{\mathrm{T}}. \tag{10}$$

These procedures mean that a standard SVD $\mathbf{O}_{\mathbf{L}}\mathbf{Y}\mathbf{O}_{\mathbf{R}}^{\mathrm{T}}\varepsilon\mathcal{M}r, r$ is executed to decompose into left singular vectors $U_i\,(\mathbf{O}_{\mathbf{L}}P_i) \in R^r$ and right singular vectors $V_i\,(\mathbf{O}_{\mathbf{R}}Q_i) \in R^r$. Compared with the SVD of the $\mathbf{Y}$ matrix, we can obtain $\sigma_i = \sqrt{\lambda_i}$, $P_i = \mathbf{O}_{\mathbf{L}}^{\dagger}U_i$, and $Q_i = \mathbf{O}_{\mathbf{R}}^{\dagger}V_i$ where † denotes the pseudo-inverse. The result follows the fact that any biorthogonal decomposition is an SVD. Oblique SSA is the modification of SSA where the SVD step is changed to ($\boldsymbol{L}$, $\boldsymbol{R}$)-SVD. However, due to the real pair ($\boldsymbol{L}$, $\boldsymbol{R}$) being unknown, an iterative operation is necessary and the IOSSA will be processed, as shown in Figure 1.

In the chart, $\mathcal{T}^{-1}\mathcal{H}$ denotes the diagonal averaging for the submatrix $\mathbf{Y}_{Ji}$ and $\mathbb{Y}_{(i)}^{k}$ expresses the reconstruction component (RC$_{Ji}$) for the k_{th} iteration. It should be noted that we set $\mathbf{L}^0$ and $\mathbf{R}^0$ as identity matrices, meaning that oblique SSA coincides with classical

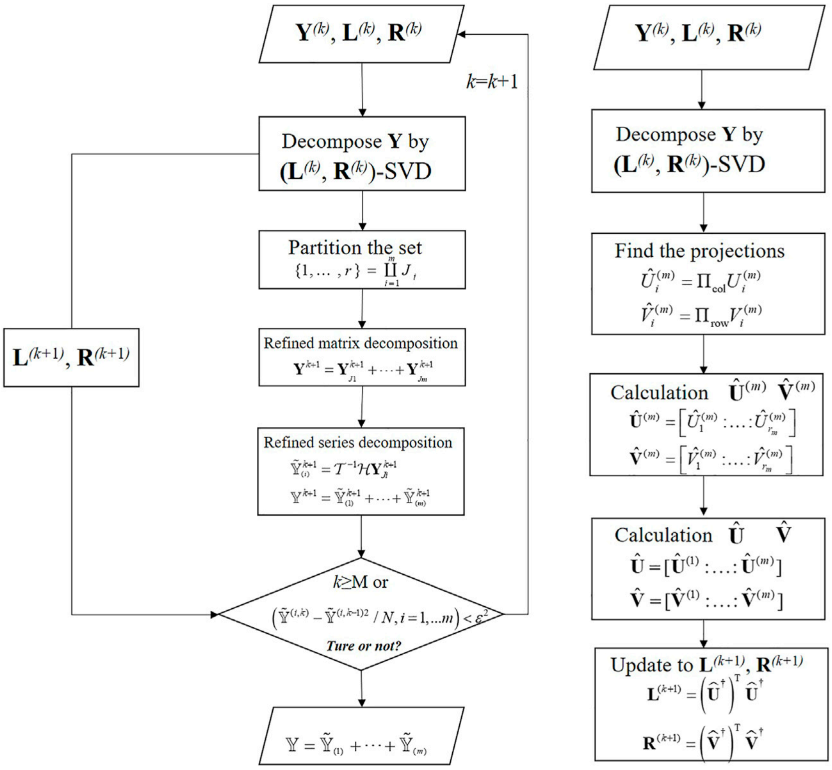


FIGURE 1
Flowchart for IOSSA. The trajectory matrix Y and the left- and right- orthogonal matrix (L, R) of the triple (Y, L, R) is updated by RSVD in left and right panel, respectively.

SSA in its first iteration calculation. In this article, three sets of eigentriple numbers in accordance with SSA (see Shen et al., 2017; 2018) are selected, and $I = \{1, 2, 3, 4, 5, 6, 7\}$ for $r = 7$ is defined. The chosen set of numbers automatically generates the partition $J_1 \sqcup J_2 \sqcup J_3$, representing the trend, Chandler, and Annual periodic components. After a matrix decomposition $X = X_{J1} + \ldots + X_{Jm}$ obtained in the SSA ***grouping step***, we denote $Y_{Ji} = X_{Ji}$ (the sub-blocks for Y). By convergence results of (L, R) SVD, the refined RC_{Ji} will be appointed as

$$\begin{cases} RC_{\text{trend}} = RC_{J1} \\ RC_{\text{Chandler}} = RC_{J2.} \\ RC_{\text{Annual}} = RC_{J3} \end{cases} \tag{11}$$

As another SVD approach, the function of IOSSA to extract the leading components like classical SVD is more appropriate for decomposing the trend and periodic subspaces. In the case of noise pollution, however, the slight edge effects (phase shift) still need to be noteworthy. Furthermore, the IOSSApd with pseudo data is used to address the phase shift issue. First, the original time series PM_{ti} $(1 \le i \le N)$ is expanded into PM_{tj} $(1 \le j \le N + l)$ by adding pseudo data with a length of $l = 365$ (day) to the end of the original time series. Expanded by applying the LS + AR model, the pseudo data are assumed to contain a trend, the CW and AW components, and residuals. Thus, the phase shift phenomenon is absorbed in the pseudo data when the expanded time series PM_{tj} is analyzed using IOSSA. The final deterministic and stochastic components with weak edge effects are obtained by intercepting the first N data (equal length to the original PM_{ti} sequence).

2.3 Experimental verification

2.3.1 Separability property (case 1)

Here, an example of a short-time series (case 1) is simulated to compare the separability of iterative orthogonal SSA and the common SSA (data length of 10 years; window length: L = 6 years) as

$$x_n = A_1 \sin(2\pi f_1 \times t) + A_2 \sin(2\pi f_2 \times t), \tag{12}$$

where $t \in [0\text{: } 10 \text{ years}]$ with the sampling interval of 0.05-year. This synthetic series consisted of two sinusoids. With the same eigenfrequency and average amplitudes of Chandler and Annual wobbles, the $f_1 = 1$ cpy and $f_2 = 0.845$ cpy, and $A_1 = 1$ and $A_2 = 1.2$ are defined. Figure 2 depicts the result of the SSA decomposition (marked in blue curves). With such close frequencies far from being orthogonal, the SSA cannot separate them for the considered window and series lengths. This completely destroys the structures of periodic signals (e.g., frequency, amplitude, and phase variations) which are clearly visible. The IOSSA algorithm is further applied using two groups, RC_{J1} and RC_{J2}. The iterative operation will be undergone until the convergence result is reached (iterations M = 55). The decomposed results (marked in

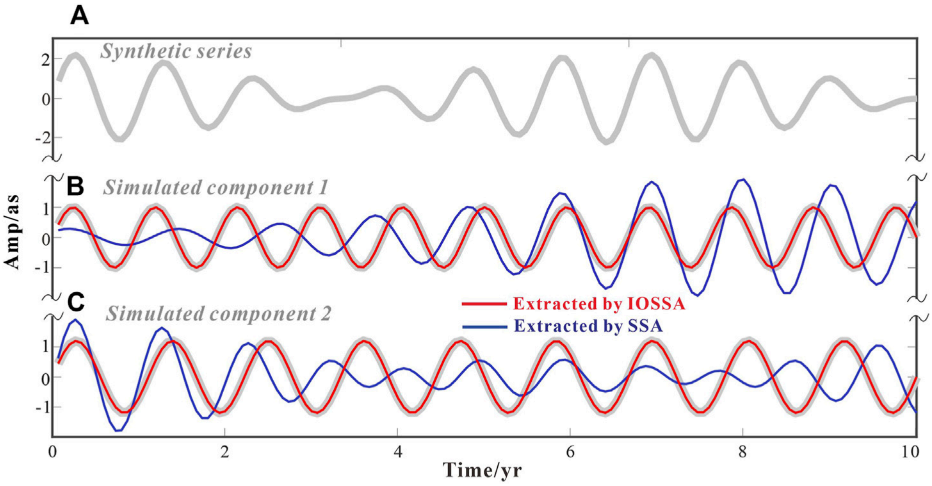


FIGURE 2
Comparison of separation performance between SSA and IOSSA when in the decomposition of sum of sinusoids. **(A)** Synthetic short-time series (of length 10 years); the extracted component 1 and 2 is shown in **(B)** and **(C)**.

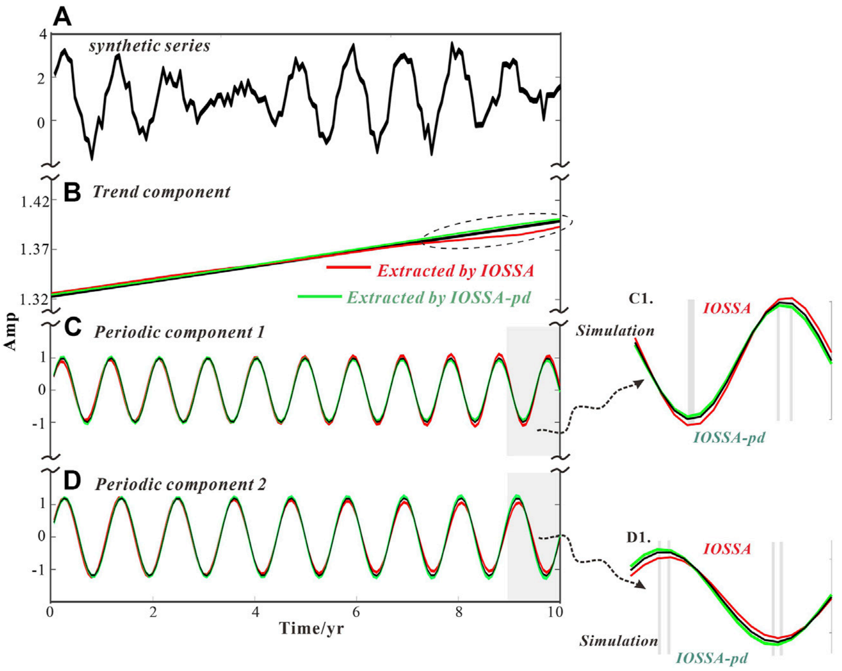


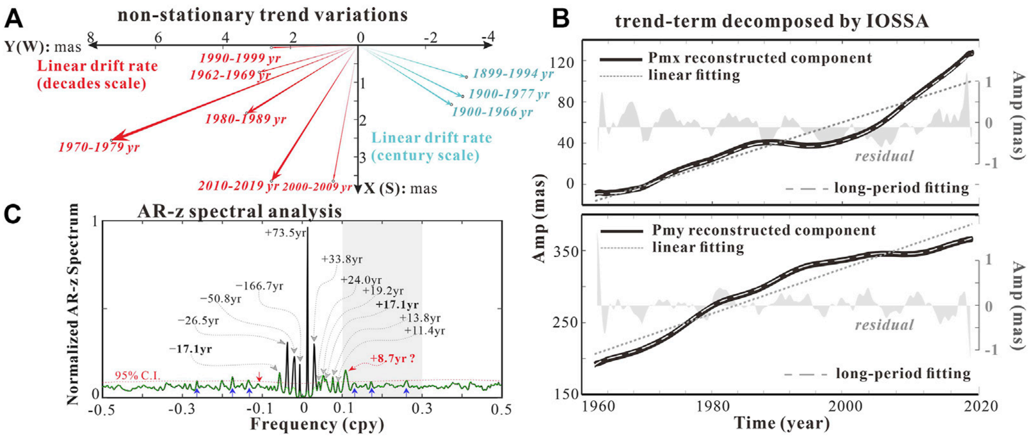
FIGURE 3
Comparison of separation performance between IOSSA and IOSSApd. **(A)** Synthetic series (of length 150) contained by Gaussian noise; the extracted trend component, periodic component 1 and 2 is shown in **(B–D)**, respectively. C1–D1 indicates the amplification to the rear end of periodic reconstruction, where the light gray domains represent the moment corresponding to the peak or trough.

red curves) confirmed that the proposed method separates harmonics exactly. Therefore, compared with the SSA which is always limited to the components disorder, the application of IOSSA means a higher possibility of such decomposition.

In reality, when the data-length and sample interval are enough, the SSA can also complete the approximate separation of trend and seasonal terms (Jin et al., 2021). The IOSSA approach, in contrast, has better separation performance even for short-length data. Under heavy noise pollution, the edge effect may become another puzzle, and thus, we designed case 2 to further verify the enhanced IOSSA decomposition with pseudo data.

2.3.2 Edge effect (case 2)

Due to the RCs aliasing, the phase shift at the front/rear ends of the reconstructed trend component will adversely affect the periodic reconstructed component. In this case, the white noise and linear term are added to the simulated case 1. Let us take the data-length as 10 years and $L = 6$ years again and then



Statement: The subgraph C is excerpted from Figure.S1 in Ding et al. (2019)*

FIGURE 4
Drift trend of polar motion. **(A)** Variations of polar motion in the decade (red arrows) and century (blue arrows) scales in previous studies, **(B)** non-stationary trend component decomposed by IOSSA, and **(C)** AR-z spectral analysis for interdecadal signals by Ding et al. (2019).

$$x_n = A_1 \sin(2\pi f_1 n) + A_2 \sin(2\pi f_2 n) + a + bn + \varepsilon_n. \qquad (13)$$

By treating sinusoidal amplitudes as $A_1 = 1$ and $A_2 = 1.2$, frequencies $f_1 = 1$ cpy and $f_2 = 0.845$ cpy, and linear terms for $a = 0.8$ and $b = 0.004$, the series is synthesized contained by noise (ε_n variance = 0.35). To show the inhibition effect of the tail effect, the ternary components in case 2 will be restored by IOSSA and IOSSApd, respectively (see Figure 3).

Figure 3B shows the performance in separating linear components. The unqualified approximation of IOSSA decomposition to the linear term implies the tail effect heavily existing in the trend components (Wang X. et al., 2016). In contrast, we suppressed the phase shift well using the IOSSApd method with extended data at the rear.

In Figure 3C, D, the influence of phase shift in the periodic components is assessed. Under the noise pollution, the offset after IOSSA decomposition and real fluctuation is recorded. At the rear part, the prominent deviation of amplitude (>0.2) or even phase delay (denoted by the gray domain) will contribute an undesirable factor for restoring time series. By contrast, the separated components by IOSSApd can accurately represent periodic reconstructions. Considering their amplitude difference (around 0.06) and full phase-matching does not matter as the fitting degree has increased to a large extent. The previous results show that a more reliable principal component can be produced by solving the phase shift problem with the IOSSApd method.

3 PM analysis

3.1 Data source

Covering 1 January 1962–31 December 2021, the EOP 14 C04 PM time series are provided by IERS. The IERS supported the 1-day solution files by radio-positioning integrated by DORIS, GNSS, VLBI, and SLR with a precision of 0.02 mas (Liu et al., 2009; Bizouard et al., 2019). Consistent with the conventional reference frame ITRF 2014, the PM dataset can be accessed at https://www.iers.org/IERS/EN/DataProducts/EarthOrientationData/eop.html.

3.2 Non-stationary trend

In the traditional LS-based forecast, the linear fitting for a non-stationary trend was reconsidered as unscientific and will introduce a predicted error of more than 10 mas. The visible secular variations appear in the polar motion measured with multi-geodetic techniques (e.g., SLR, VLBI, GNSS, and DORIS). The IOSSA analysis gives the *x* and *y* components for PM from 1962 to 2021, which departs away from zero (see Figure 4). In Figure 4A, the polar motion refers to the CIO frame (*x* for PM: south direction; *y* for PM: west direction). The trend drift (PM secular rate (per year) in polar-coordinate) rates in the decadal and centurial scales are presented in the first and second quadrants, respectively. In the decades scale, the trend rate of the PM varies from 2.56 ± 0.0714 mas to 7.73 ± 0.0286 mas per year. The linear velocity of trend drift narrates that the Earth's polar motion speeds up in the latest decade (2010–2019) (see Figure 4B), which is different from the slowing in the geologic age. It reveals the long-term fluctuation in the Earth's polar motion (Guo and Han 2009). In the century-scale, the rate and direction of PM from different authors were given (Markowitz 1968; Wilson and Vicente, 1980; McCarthy and Luzum 1996). The trend rate of pole motion is from 3.2 to 3.4 mas per year with the well-known direction of Greenland. However, the trend rates of PM in the *X* and *Y* directions in Figure 4B, covering over half of the century, are 2.01 ± 0.1819 and 3.11 ± 0.2021 mas per year. The North Pole moves to 57.1 °E in the longitude direction concerning the crust. The Barents Sea is fit in this direction. These different results also implied that the non-linear drift contains some unmistakable low-frequency signals, and the accurate separation of trend components for fitting and extrapolating is necessary.

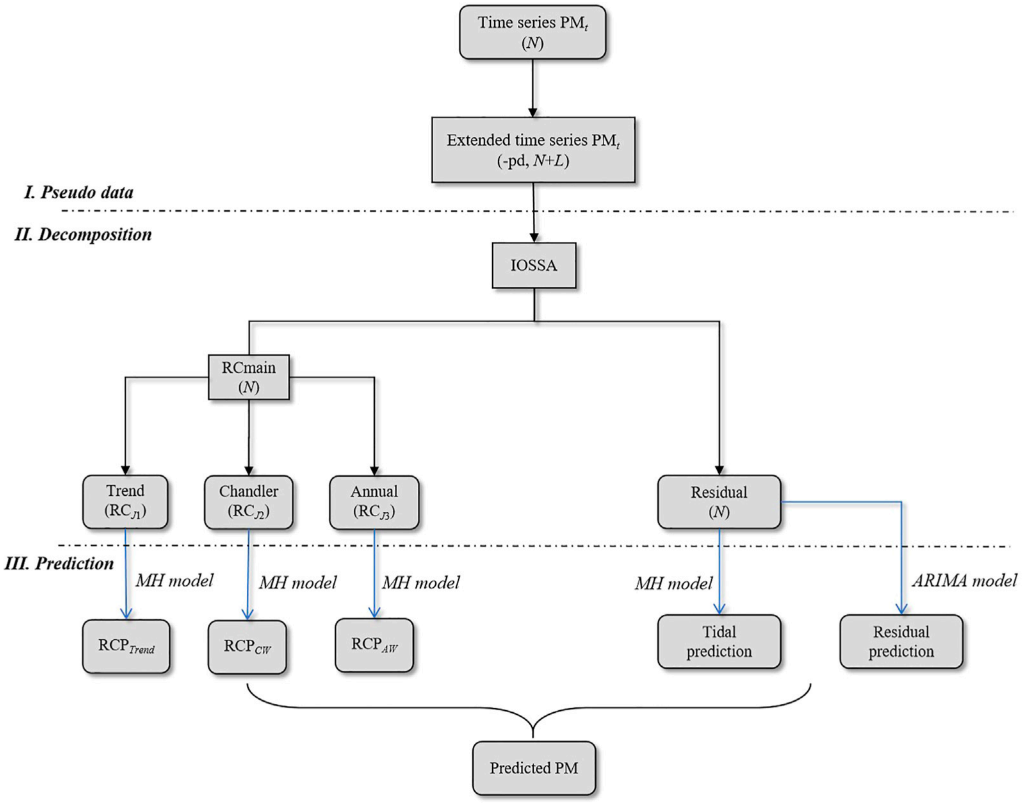


FIGURE 5
Flowchart of the IOSSApd + MH + ARIMA method for PM prediction.

The complex composition of PM oscillation from interdecadal to even century-scale has been researched by Ding et al. (2019). In Figure 4C, according to the AR-z spectrum, a robust weak signal recognizer with high spectral resolution than the conventional Fourier-based spectrum, the complex time series ($x + iy$) in a low-frequency band is explored. The nine spectral peaks in the −0.1~+0.1 cpy band, and the corresponding periods are approximately −17.1 years, −26.5 years, −50.8 years, −166.7 years, +73.5 years, +33.8 years, +17.1 years, +13.8 years, and +11.4 years, meaning the trend drifts with intense time-varying characteristics. Some of these multi-spectral peaks in the frequency domain have been identified as the consistent geophysical periodicities from, e.g., the solar cycle (~11 years) (Currie, 1981), the Earth's nutation (~33 years and ~14 years) (Ding and Chao, 2018), geomagnetic variations (~28 years, ~50 years, and ~70 years) (Dickey and Viron, 2009; Dobrica et al., 2018), and others perhaps induced from atmospheric or oceanic effects (Ding et al., 2019). Integrating with the upgraded SSA separator (iterative oblique SSA) and multiple-harmonic models, the non-stationary trend-terms are first reconstructed and then fit completely with these low-frequency signals (see Figure 4B, fitting residual <1 mas). In the following context, such long-period oscillation extrapolation is considered as the initial trial.

4 Prediction procedure

One limitation of SSA is the significant biases in the beginning and end of its model fitting to a time series called phase phenomenon (Wang X. et al., 2016). To address the phase shift phenomenon and improve the PM forecast, we proposed the IOSSApd method in the frame of a two-step strategy. By adhering to the predictions of the LS + AR (initial) model to the end of the PM series, the IOSSA (enhanced version of SSA) will give a formidable oscillation separation. After the MH + ARIMA model (terminal) triggered the second cycle forecast, the final RC predictions (RCPs) will be produced. By this proposed hybrid method, 11 years of PM predictions were made at 12-month intervals from 1 January 2010, and the main steps (see Figure 5) are as follows:

i. *Pseudo data (-pd).* In pseudo prediction, the LS + AR model is commonly used to predict the Earth's orientation parameters with simple construction and lower accuracy (Kalarus et al., 2010). Here, we fit and forecast the linear and periodic terms (L = 365 days), including Chandler wobble and annual wobble, and the AR model matches semi-annual periodic terms in PM and their remnants. The predicted series by LS + AR is attached to the end of PM observations as pseudo data with a total length of $N + L$.

ii. *IOSSA decomposition.* By high-precision observation (1962), the long-term PM series (of length $N + L$) with LS + AR predictions at the end is separated by IOSSA. According to the partition of $J_1 \sqcup J_2 \sqcup J_3$ and chosen window of 2190 points, the deterministic components of trend (RC$_{J1}$), the Chandler wobble (RC$_{J2}$), and the annual wobble (RC$_{J3}$) are obtained with the stochastic component remaining. For each component, a hybrid model of MH or ARIMA is adopted under different coverage to make a precise extrapolation.

iii. *Final prediction.* The PM series has been recognized as multifrequency modes from the interannual, short-term (e.g., seasonal) oscillations to the long-term trend. The time-varying oscillation can be precisely approximated and extrapolated by the sinusoids in frequency-amplitude modulation. A modified LS model, a multiple-harmonic (MH) model, is creatively proposed in the system of equations as

$$\begin{cases} PM_{\text{Annual}}(t) = (a_A + m_A t)\sin(2\pi f_A t + w_A), \\ PM_{\text{Chandler}}(t) = \sum_{i=1}^{M} b_{Ci}\sin(2\pi f_{Ci} t + w_{Ci}), \\ PM_{\text{Trend}}(t) = \sum_{i=1}^{N} b_{Ti}\sin(2\pi f_{Ti} t + w_{Ti}), \\ PM_{\text{Tide}}(t) = \sum_{i=1}^{P} b_{TDi}\sin(2\pi f_{TDi} t + w_{TDi}). \end{cases} \tag{14}$$

In Eq. (14), the a_A and m_A are the constants of the annual linear amplitude (Shen et al., 2018), $f_A = 1/365.25$ is the annual frequency, and w_A is the un-estimated phase. Instead of the assumption of an initial constant period (e. g., 428 days) in the biaxial rigid Earth (Becker, 1954), the later double-periods motion (e.g., 428 and 438 days) with the beat model (Colombo and Shapiro, 1968) or the recent variable period (e.g., from 392 to 441 days) under the sustained excitation (Wang G. et al., 2016), the MH fitting for CW was postulated as 5-period channels (namely, $M = 5$, Pan. 2007; Zhang et al., 1986) of T_{main} {432 days}, $T_{1,0}$ and $T_{1,1}$ {symmetrical peaks: 429 and 435 days}, and $T_{2,0}$ and $T_{2,1}${symmetrical peaks: 406 and 447 days}. These multiple Chandler frequencies correspond to the parameter f_{C1}–f_{C5}, respectively, and consist of a fundamental frequency attributable to the Earth's triaxiality, a series of small feedback frequencies arising from the Earth's instantaneous inertia, relative angular momentum, and inertia variation. In the triaxial Earth model, the Chandler wobble may behave like a multiple-coupling differential oscillation with the mantle as the major and permanent coupler. Such a multiple-frequency model can explain the Chandler amplitude and frequency variations with time (for more details about the observed CW frequency-splitting and validation, please see Pan (2007; 2012)). According to Xu (2013), the sampling for both components covers 9 years (7.6 Chandler periods or 9 annual terms). In addition, the MH model was applied for the first time in forecasting non-stationary trends as long-period oscillations ($N = 9$) and tidal harmonics as seasonal oscillations ($p = 7$). The dominant frequency (f_{T1}–f_{T9}, please refer to Figure 4) and (f_{TDi}) parameter (f_{TD1}–f_{TD7}, please refer to Chin et al. (2004)). In addition to these multiple deterministic components, the stochastic components (containing quasi-periodic signals) are also considered. In this article, for the stochastic sequence, a more stable and robust sequence is obtained by ARIMA than an AR or ARMA model in the case of the modeling of a seasonal differential operator as

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}. \tag{15}$$

The model is usually described as ARIMA (p, k, q). The ternary orders of p, k, and q offsetting the tide effect (Chin et al., 2004) are estimated using the least squares method. Combining the predictions from the multi-harmonic and ARIMA (p, q, k) model will yield the final results. When the adjacent period changes, their time-varying characteristics can be fully reflected in 1-year PM predictions.

5 Predictions of the polar motion

5.1 The predicted results and precision comparison

In this study, EOP 08 C04 PM data from 1962 to 2021 were selected for the validation of SSApd + MH + ARIMA predictions. The PM from 1 January 1964 to the predicted start time was used as the original PM series. After IOSSApd decomposition, the principal components of 1-year lead time PMs (denoted as PCP) were predicted using the MH (5-period channels CW + linear AW + non-stationary trend, see Eq. 14) and the LS model (sinusoid CW/AW + linear trend, refer to Shen et al., 2017), respectively. We produced multiple sets of predicted PMs in 1-year-leading (of 2010–2021 years). Decomposition of the original PM series concerning the prediction period was also performed by IOSSApd to obtain the real principle components (denoted by PC). The predictive errors (PC-PCP) for all prediction periods are shown in Figure 6.

Figure 6 shows that the PCP series obtained by the MH model are more consistent with the original PCs over multiple prediction periods than the conventional LS model (errors: ±40 mas in random variations. vs. ±70 mas in periodic fluctuation). The trend, annual, and Chandler terms contribute most of the energy of polar motion; therefore, the MH predictor will produce the satisfied principle components that accurately represent the original series, ARIMA. Concerning the residual IOSSApd series, we mainly used the ARIMA to extrapolate them (see Figure 5). The extrapolation results of the ARIMA model were combined with the principle component predictions of the MH (ARIMA + MH) to yield the whole 1-year PM predictions.

The predictions of Bulletin A including PM parameters for 1 year into the future were provided by IERS and are currently recognized as the official forecasts, which relied on the history observations (Stamatakos et al., 2011). To verify the reliability of the IOSSApd + MH + ARIMA method, the accuracy of the predicted series was compared with the accuracy of the IERS Bulletin A predictions (http://datacenter.iers.org/eop/-/somos/5Rgv/getTX/6). As shown in Figure 7, the systematic prediction of both methods can be seen. With a relatively high precision in the 180-day forecast, the IOSSApd + MH + ARIMA's prediction is even more accurate in the 365-day forecast than the Bulletin A. It indicated the proposed solution predicting the PM parameters precisely and effectively. As the comparable results from the IERS Bulletin A, the combination of the

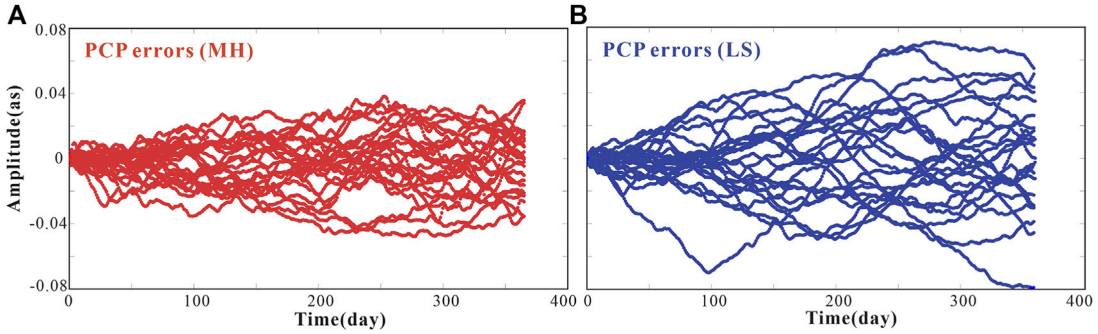


FIGURE 6
Predictive errors of PM principle components (Chandler + annual + trend) using **(A)** MH model and **(B)** LS model in both the *x/y* poles.

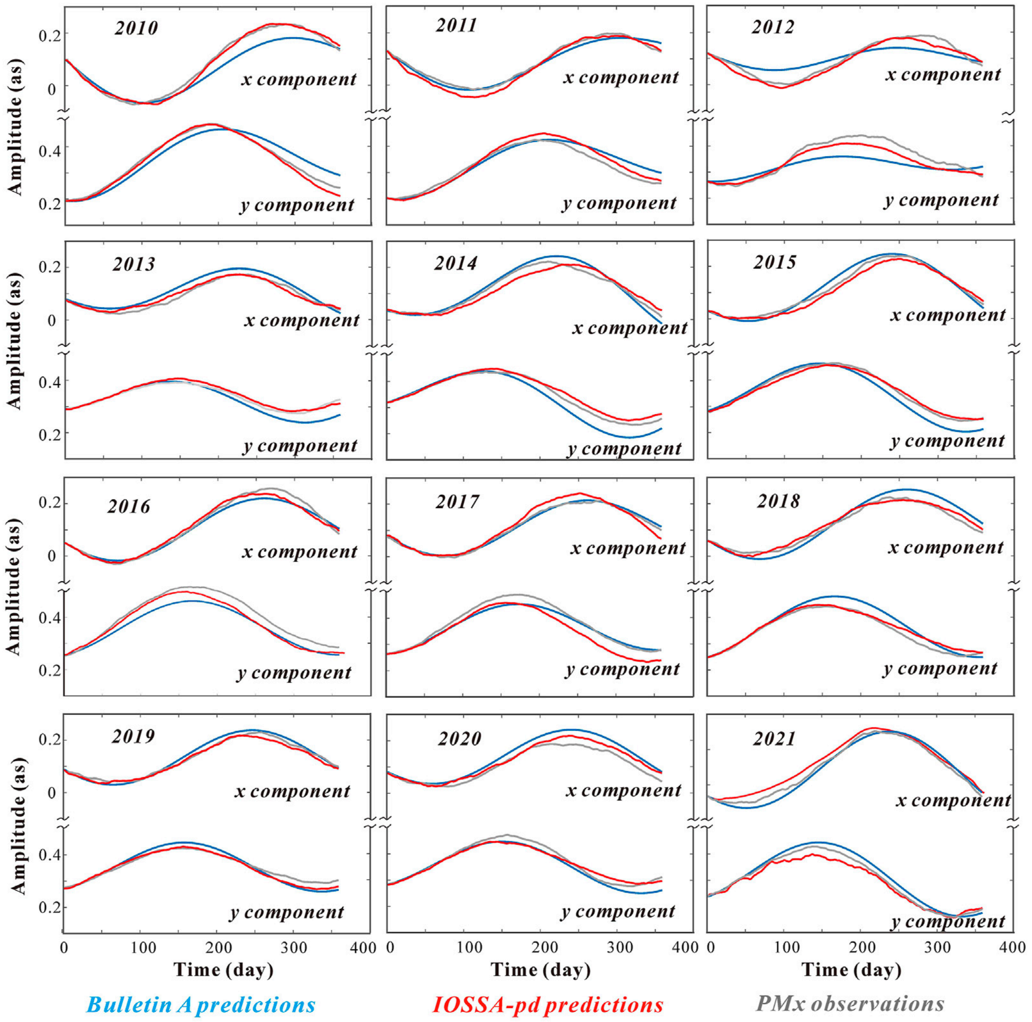


FIGURE 7
From the epoch of 2010 to 2021 years, the polar motion series (gray line), the IERS "Bulletin A" predictions (blue line), and the IOSSA + MH + ARIMA predictions (the red line) in 1-year leading.

IOSSApd and ARIMA produced PM series with a higher consistency of observations in most cases. The better performance of the IOSSApd + MH + ARIMA prediction was attributed to the modeling of frequency- and amplitude-modulation for the Chandler, the annual oscillation, and the superposition of non-stationary trend.

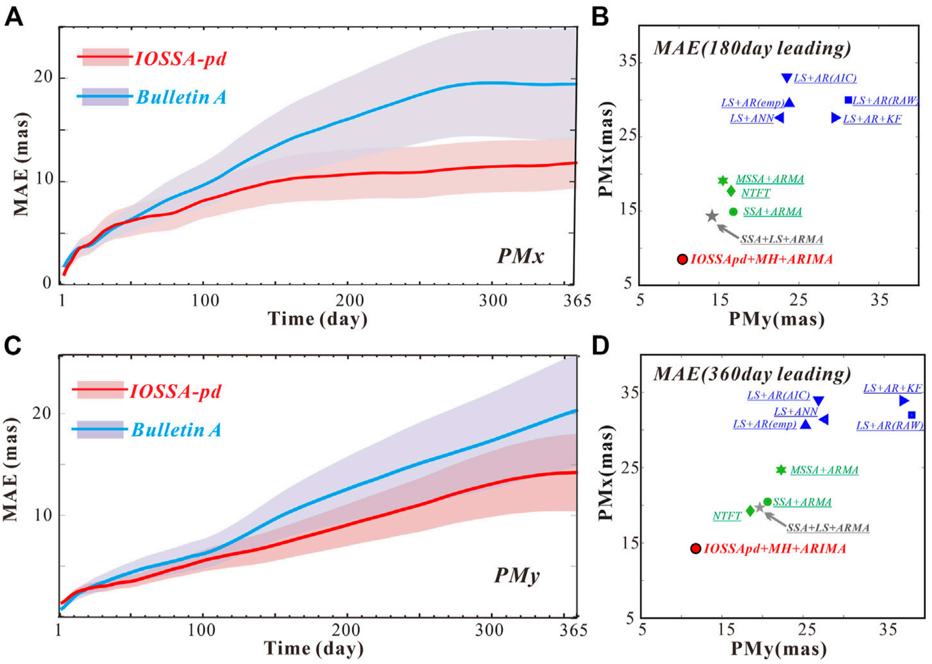


FIGURE 8
MAE of the IERS Bulletin A (red) and IOSSA + ARIMA predictions (blue) in *x*-component **(A)** and *y*-component **(C)**. MAE of the predictions by other methods in 180-day **(B)** and 360-day **(D)** leading. The shadows represent the error domain.

TABLE 1 Comparison of the MAE of IERS Bulletin A and IOSSApd + MH + ARIMA predictions.

Lead time/day	*x*-component/mas		*y*-component/mas	
	Bulletin A	IOSSApd	Bulletin A	IOSSApd
10	3.26	3.07	1.93	2.15
30	4.80	5.11	3.20	3.07
60	7.38	6.98	4.60	3.96
90	9.49	7.69	5.69	5.12
120	11.58	9.54	7.26	6.41
150	13.50	10.53	9.47	7.52
180	14.83	10.85	11.31	8.78
210	16.10	11.11	12.90	9.85
240	17.32	11.17	14.38	10.92
300	18.76	11.64	16.94	13.00
360	18.68	12.03	19.59	13.93

Figures 8A,C indicate that in the leading 1-year predictions, the mean absolute errors (MAEs) of our proposed two-step forecast were lower since the 50th day than the predictions reported routinely in Bulletin A. In particular, the IOSSApd + MH + ARIMA's predictive accuracy at forecast horizons of 180-day to 360-day was significantly higher over the Bulletin A. Corresponding to improvements of 28.89% for *y*-pole, the remaining prediction errors for *x*-pole are even more minor with the advancements of 35.60%. The MAEs are shown in Table 1. In the 180-day predictions, the MAEs of IOSSApd + ARIMA for the *x* and *y* components are 10.85 and 8.78 mas, respectively, smaller than those of 14.83 and 11.31 mas for the Bulletin A predictions. In the 360-day predictions, the MAEs for *x* and *y* components from IOSSApd + MH + ARIMA were 12.03 and 13.93 mas, respectively, much smaller than those of 18.68 and 19.59 mas for Bulletin A. The MAE of IOSSApd + MH + ARIMA increased slowly with the increase in the forecasting time, and the prediction accuracy of 360 was several mas more than that of 180 days. This implied that the forecasting errors in the 180–360 days are limited to a small range. The IOSSA can reliably reconstruct the stable principal components fitting by multispectral peak or time-varying amplitude harmonics. It enables the IOSSApd + MH + ARIMA to be particularly suitable for long-term prediction. Moreover, at the entire forecast horizons, the mean absolute error of IOSSApd + MH + ARIMA prediction was generally less than 12 mas and 15 mas for *x* and *y* components, respectively; while in the Bulletin A predictions, the time interval at equal accuracy only covers 1–130 days and 1–260 days, respectively. As for the short-term prediction within 50 days, high-frequency oscillation (or quasi-periodic oscillations) and random residual are still tricky because this article mainly refined the mid-term and long-term prediction of *x*/*y*-pole and the ultra-to fast-forecast is not discussed.

The MAE of the 180-day and the 360-day leads were also compared with the results of other methods (see Figures 8B,D). We classified these methods into two categories: the LS-based methods (marked in purple) and time-variant methods (marked in green). The former include the LS + AR, LS + AR + AF, and LS +

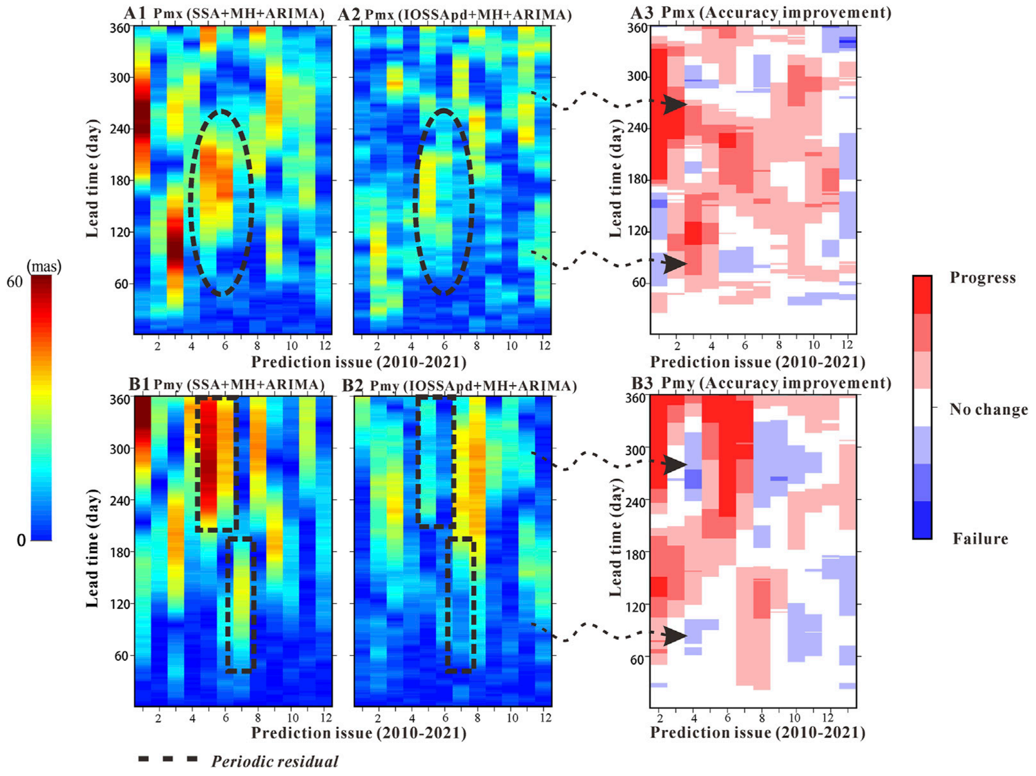


FIGURE 9
Absolute errors of the predicted x-component **(A2)** and y-component **(B2)** for the 2010–2021 years, compared with those of the SSA + MH + ARIMA model [**(A1)** and **(B1)**]. The black box represents the El Niño event as an epoch example when the error cluster is piled up with periodic residuals. The right panels show the improvement of *x*/*y*-pole prediction using the IOSSApd + MH + ARIMA model compared with the SSA + MH + ARIMA in x-component **(A3)** and y-component **(B3)**, where the progress (red), no change (white) and failure (blue) reflects the difference of both predicted error >5 mas, within ±5, and <5 mas, respectively. The darker the color, the greater their difference is.

ANN methods (Kosek et al., 2007; Liao et al., 2012; Jia et al., 2018; Sun et al., 2019), which relied upon the fixed-frequency and linear trend, and another group consisted of NTFT, MSSA + ARMA, and SSA + ARMA methods (Su et al., 2014; Shen et al., 2018; Jin et al., 2021), the algorithm dominated by decomposed non-stationary components. Evidently, the MAEs of PM with the time-variants method were significantly less than those of LS-based methods, either in the middle- or long-prediction. Notably, the SSA + LS + ARMA predictive model [marked in gray, Shen et al. (2017)] adopts a similar strategy to the IOSSApd + MH + ARIMA method, i.e., the decomposition from polar motion observations first and fitting-extrapolation (but in single Chandler harmonic at Tcw = 432 days) second to achieve a high-precise prediction (MAE<15 mas in middle-term, and <20 mas in long-term for *x*/*y*-poles). Compared with these methods, we examined the multi-frequency points/band in IOSSApd's reconstructed components and, hence, improved the predictive performance to a large extent.

5.2 The periodic errors analysis

When the PM predicted errors are accumulated with leading time, they are always clustered and manifested in the pattern of periodic errors. These periodic errors can be used as indicators for analyzing the forecast accuracy of periodic/quasi-periodic components. In this section, the absolute error (AE) comparisons against the general SSA-based method (SSA + MH + ARIMA, the same window of 6 years, and the predictive component of the first 7th RC as IOSSA) for the 1-year-leading PM prediction are further conducted in the same span of 2010–2021. In Figure 9, a typical high-error presentation with different patterns, e.g., from 2014 to 2016, which was heavily affected by the super El Niño action, is analyzed. As the NOAA National Centers for Environmental Information's State of the Climate report states, the El Niño recorded was the most prolonged in duration and the most vigorous in intensity since the comprehensive observations emerged in September 2014 and dissipated in May 2016. The circulation pattern and temperature anomaly promote the amplitude variations of the annual wobble (Coulot and Pollet, 2010). Correspondingly for SSA + MH + ARIMA, the predicted error increased with −6–65 mas (July–December 2014), −23–50 mas (September–December 2015), and −10–41 mas (February–June 2016) for the *y*-pole prediction. Advanced by ~120 days phase lead-lag, the predicted error in the *x*-pole also spread within ±50 mas, implying the prominent periodic residuals. Such error clusters covering even over half a year are a widespread occurrence in the 12-year PM predictions. Failure to trace the time-variant periodic term (annual and Chandler wobbles), here

TABLE 2 Success rate for PM (no change + progress).

Prediction year	*x*-component	*y*-component	Average
2010	70.56	96.11	83.34
2011	61.39	48.33	54.86
2012	84.45	72.78	78.62
2013	88.61	83.89	86.25
2014	88.89	91.67	90.28
2015	85.56	93.33	89.45
2016	74.72	47.78	61.25
2017	73.61	38.33	55.97
2018	80.83	72.78	76.81
2019	84.17	61.11	72.64
2020	64.17	60.83	62.50
2021	47.50	67.50	57.50
Sum (2010–2020)	75.39	69.58	72.49

for SSA, is evidenced adequately. In contrast, the IOSSApd + MH + ARIMA's reduction ability is excellent, either in error clusters timespan or its intensity. Based on the amplitude–frequency modulation, the extended conditions of the *x*/*y*-pole multi-components in high separation play a crucial role in the successful forecast.

The heat maps are drawn (Figure 9A3 and B3) to better understand such a periodic error in the mid- and long-term forecast to demonstrate the predicted improvement in the IOSSApd + MH + ARIMA solution compared to SSA + MH + ARIMA. For each prediction epoch, if the difference between predicted errors of IOSSApd + MH + ARIMA and SSA + ARMA is within ±5 mas, it cannot be regarded as an improvement in prediction. As illustrated in the white blocks in Figures 9B and D, the error amounted to the same level for both PM prediction techniques. If the difference >5 mas, the IOSSA-pd + MH + ARIMA prediction was considered to have a smaller error than SSA + ARMA and marked by red blocks. In other cases, the failure in the prediction process was denoted by blue blocks. The large improved areas (red) demonstrated that the behavior of PM prediction is progressed by IOSSA-pd + MH + ARIMA, especially in long time intervals.

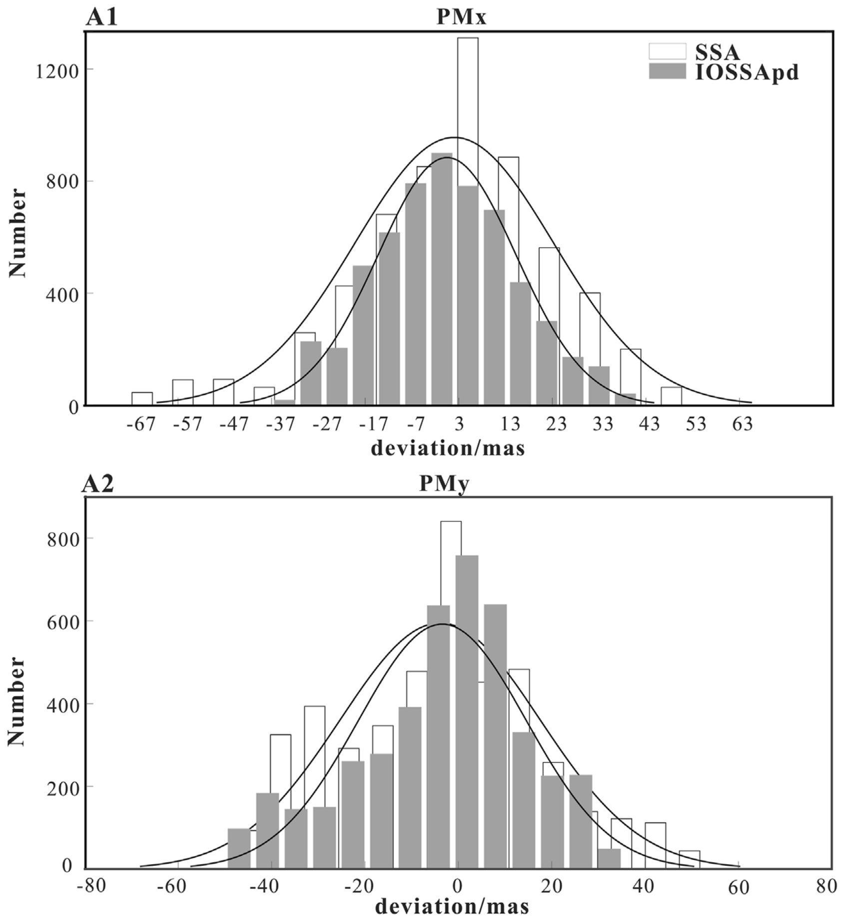


FIGURE 10
Distributions of the IOSSA + MH + ARIMA (gray) and SSA + MH + ARIMA (white) prediction residuals with the best-fitted normal distributions in *x*-pole **(A1)** and *y*-pole **(A2)**.

TABLE 3 Statistical index of Gaussian distribution in Figure 10.

	Method	Mean μ)	Std σ)	Skewness s)	Kurtosis k)
x-component	SSA	2.5	21.2	0.52	3.5
	IOSSApd	−1.5	14.7	0.01	2.8
y-component	SSA	3.9	21.4	0.15	3.3
	IOSSApd	−2.0	13.5	−0.42	3.2

By the ratio of the colored area (white + red) to the total area (white + red + blue) in each epoch, we introduced the success rate of PM prediction. Table 2 shows the success rate of 2010–2021 using the IOSSApd + MH + ARIMA algorithm. On average, the success rate of the x and y components surged to 70.82% and 74.02%, respectively, and the improvement in the PM prediction reached approximately 73%. Therefore, our predicted PM results reflect weaker and smoother periodic errors in mid- and long-term predictions, and this is largely because the IOSSApd separator performs better in reconstructing the PC series than the general SSA.

Based on the experiments covering the years 2010–2021, comprehensive prediction errors were collected by additional statistics to verify the compliance of the normal distribution. In Figure 10, the error distribution features by IOSSApd + MH + ARIMA and SSA + MH + ARMA are shown, and their error curves are also drawn.

By reducing the effect of the phase shift in each prediction epoch, the two-step solutions might be superior and then in the presentation of a Gaussian-like rule. To exactly describe the time series as reasonably Gaussian, four statistical parameters were introduced, including mean value (μ), standard deviation (σ), skewness (s), and excess kurtosis (k). As for the μ and σ parameters, they provided the quantitative reference defining the location and aggregation of the error distribution. Else, for the s and k estimators, they indicate the state or quality of asymmetry and peakedness (or flatness), describing the normal distribution error in the region about its mode. According to Tabachnick et al. (2007), these parameters are formulated as follows:

$$\hat{\mu}_i = \frac{1}{n_p}\sum_{j=1}^{n_p}\varepsilon_{i,j}, \quad \hat{\sigma}_{\hat{\mu}_i} = \frac{\hat{\sigma}_i}{\sqrt{n_p}}, \tag{17}$$

$$\hat{\sigma}_i = \sqrt{\sum_{j=1}^{n_p}\frac{\left(\varepsilon_{i,j}-\hat{\mu}_i\right)^2}{n_p-1}}, \quad \hat{\sigma}_{\hat{\sigma}_i} = \frac{\hat{\sigma}_i}{\sqrt{2n_p}}, \tag{18}$$

$$\hat{s}_i = \sum_{j=1}^{n_p}\frac{\left(\varepsilon_{i,j}-\hat{\mu}_i\right)^2}{\left(n_p-1\right)\hat{\sigma}_i^2}, \quad \hat{\sigma}_{\hat{s}_i} = \sqrt{\frac{6}{n_p}}, \tag{19}$$

$$\hat{k}_i = \sum_{j=1}^{n_p}\frac{\left(\varepsilon_{i,j}-\hat{\mu}_i\right)^4}{\left(n_p-1\right)\hat{\sigma}_i^4} - 3, \quad \hat{\sigma}_{\hat{k}_i} = \sqrt{\frac{24}{n_p}}. \tag{20}$$

Next, the prediction error distribution of the PM time series is characterized in detail. As listed in Table 3, the μ value (−1.5 mas and −2.0 mas for the x/y pole) of IOSSApd + MH + ARIMA evidenced that this algorithm offers a solution approximating the unbiased estimation, meaning the prediction errors are close to zero. The second parameter σ plays another vital role in the quality of the prediction and is related to the error range and limits. Regarding the estimator, the IOSSApd (σ = 14.1 mas on average) is still more competitive than SSA (σ = 21.3 mas on average).

Using the skewness and the kurtosis parameters, other interesting explorations can be performed. On one hand, the s parameter in both methods all have a positive value (the right tail is stronger than the left one) for the x-pole and a negative value for the y-pole (the right tail is weaker than the left one). On the other hand, the positive value of the k parameter for each pole implies that the distribution is peaked relative to a normal one. Specific to each method, the periodic signals of predicted error by IOSSApd + MH + ARIMA decreased rapidly. When compared with SSA + MH + ARIMA ($|s|$ = 0.34, k = 3.4 on average), these errors ($|s|$ = 0.21, k = 3.0 on average) are more statistically normal significant because of their smaller parameter values (except for the skewness in y-pole).

6 Conclusion

In this article, we introduce a newly improved separator (iterative oblique SSA) used in the decomposition of deterministic components. The Chandler wobble, annual wobbles, or even the non-stationary trend of the Earth's polar motion can be considered as quasi-harmonic processes, as indicated by the IOSSA. To reduce the phase shift, a two-step strategy (pseudo date in initial prediction by LS + AR) is preliminarily established, called IOSSApd. Based on the IOSSApd decomposition, many periodic/quasi oscillations in interdecadal and decadal (trend term), as well as inter-to sub-annual (CW, AW, tidal, and seasonal) scales, were involved in obtaining a routinely high-accuracy PM prediction for up to 1-year into the future. When compared to other forecast products, the main advantages of IOSSApd + MH + ARIMA predictions are listed as follows:

- The extended separation conditions than classical SVD decomposition.
- The generation of pseudo data (PD) to address the tail effect.
- Long-period superposition of interdecadal oscillations instead of conventional (non) linear trend.
- Amplitude and frequency modulation by multispectral peaks for Chandler and annual wobbles.
- Robust process for high-frequency residuals by seasonal difference operator.

Using the Bulletin A published in IERS as a reference (predictions set to 1 year into the future, 2010–2021 on experiments), multiple sets of the predicted performance were elaborated. The forecasts of IOSSApd + MH + ARIMA were almost equal to the IERS Bulletin A within the 50th-day projections, while after that, notable promotions were recorded. As forecast time increased, the prediction error accumulated slowly until the 1-year lead time. At the entire forecast horizon, the remaining prediction errors were approximately 12 mas and 14 mas for PM, respectively, corresponding to significant improvements of 35.60% and 28.89% for x and y poles over Bulletin A. Some common patterns, e.g., the El Niño effect or geomagnetic jerk events, might excite the time-variant Chandler and annual wobbles and cause high cluster errors. Interestingly, whether it is the time span or the intensity of the periodic remaining, the inhibition capability of IOSSApd + MH + ARIMA is pretty impressive. In terms of the daily improvement for the x and y poles, our multivariate algorithm was found to reach up to 75.39% and 69.58% in the 1-year leading predictions, respectively. These results exhibit the excellent consistency of the predicted and original PM periods' periodic and trend terms. Based on the collected prediction error, the PM predictions were finally evaluated by normal distribution statistics. The Gaussian-like distribution further implies by the IOSSApd + MH + ARIMA method; the PM principal components can be sufficiently extracted and extrapolated with tiny oscillatory residuals (interannual or interdecadal). Therefore, our proposed method has good reliability and high precision for this PM prediction. Finally, although the IOSSApd + MH + ARIMA method developed in the study is most suitable for middle- and long-term PM forecasts, one can also extend our multivariate algorithm to forecast other EOP parameters by considering the different multiple periods in it.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

Author contributions

KS was involved in the advanced multivariate algorithm, drawing images, and conducting the manuscript writing. HD contributed to the PM predictions design and manuscript proof reading. CZ and TC helped in collecting PM and Bulletin A datasets. All authors read and approved the final manuscript.

Funding

This study is supported by the National Natural Science Foundation of China (Grants: 41974022, 41721003, and 42192531), Educational Commission of Hubei Province of China (Grant: 2020CFA109), and the Project Supported by the Special Fund of Hubei Luojia Laboratory (Grant: 220100002).

Acknowledgments

The authors thank IERS for providing IERS 14 C04 polar motion series and Bulletin A, which can be obtained for free by downloading from https://www.iers.org/IERS/EN/DataProducts/EarthOrientationData/eop.html. The IOSSA method was performed using the R Language, available at https://rdrr.io/cran/Rssa/man/iossa.html.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Akyilmaz, O., and Kutterer, H. (2004). Prediction of Earth rotation parameters by fuzzy inference systems. *J. Geod.* 78, 82–93. doi:10.1007/s00190-004-0374-5

Akyilmaz, O., Kutterer, H., Shum, C. K., and Ayan, T. (2011). Fuzzy-wavelet based prediction of Earth rotation parameters. *Appl. Soft Comput.* 11, 837–841. doi:10.1016/j.asoc.2010.01.003

Barnes, R., Hide, R., and Wilson, A. (1983). Atmospheric angular momentum fluctuations, length-of-day changes and polar motion. *Proc. R. Soc. Lond. B. Biol. Sci.* 387, 31–73. doi:10.1098/rspa.1983.0050

Becker, R. A. (1954). *Introduction to theoretical mechanics*. McGraw-Hill, New York.

Bizouard, C., Lambert, S., Gattano, C., Becker, O., and Richard, J. Y. (2019). The IERS EOP 14C04 solution for Earth orientation parameters consistent with ITRF 2014. *J. Geod.* 93, 621–633. doi:10.1007/s00190-018-1186-3

Buffett, B., Mathews, P. M., Herring, T. A., and Shapiro, I. I. (1991). Forced nutations of the Earth: Influence of inner core dynamics, 4, Elastic deformation. *J. Geophys. Res. Atmos.* 96, 8258–8274.

Chandler, S. C. (1891). On the variation of latitude, I. *Astron. J.* 11, 59–61. doi:10.1086/101603

Chao, B. F., and Chung, W. Y. (2012). Amplitude and phase variations of Earth's Chandler wobble under continual excitation. *J. Geodyn.* 62, 35–39. doi:10.1016/j.jog.2011.11.009

Chen, J. L., and Wilson, C. R. (2005). Hydrological excitations of polar motion, 1993–2002. *Geophys. J. Int.* 160, 833–839. doi:10.1111/j.1365-246X.2005.02522.x

Chin, T. M., Gross, R. S., and Dickey, J. O. (2004). Modeling and forecast of the polar motion excitation functions for short-term polar motion prediction. *J. Geod.* 78, 343–353. doi:10.1007/s00190-004-0411-4

Colombo, G., and Shapiro, I. I. (1968). Theoretical model for the Chandler wobble. *Nature* 217, 156–157. doi:10.1038/217156a0

Coulot, D., Pollet, A., Collilieux, X., Berio, P., Collilieux, X., and Berio, P. (2010). Global optimization of core station networks for space geodesy: Application to the referencing of the SLR EOP with respect to ITRF. *J. Geod.* 84, 31–50. doi:10.1007/s00190-009-0342-1

Currie, R. G. (1981). Solar cycle signal in Earth rotation: Nonstationary behavior. *Science* 211, 386–389. doi:10.1126/science.211.4480.386

Dicke, J. O., and Viron, O. (2009). Leading modes of torsional oscillations within the Earth's core. *Geophys. Res. Lett.* 36, L15302. doi:10.1029/2009GL038386

Ding, H., and Chao, B. F. (2018). Application of stabilized AR-z spectrum in harmonic analysis for geophysics. *J. Geophys. Res. Solid Earth* 123, 8249–8259. doi:10.1029/2018JB015890

Ding, H., Pan, Y. J., Xu, X. Y., and Li, M. (2019). Application of the AR-z spectrum to polar motion: A possible first detection of the inner core wobble and its implications for the density of Earth's core. *Geophys. Res. Lett.* 46, 13765–13774. doi:10.1029/2019GL085268

Dobrica, V., Demetrescu, C., and Mandea, M. (2018). Geomagnetic field declination: From decadal to centennial scales. *J. Geophys. Res. Solid Earth* 9, 491–503. doi:10.5194/se-9-491-2018

Golyandina, N., and Shlemov, A. (2013). Variations of singular spectrum analysis for separability improvement: Non-orthogonal decompositions of time series. *Stat. Interface* 8, 277–294. doi:10.4310/SII.2015.v8.n3.a3

Gross, R. S. (2007). Earth roation variations-long period, in *Physical geodesy*, edited by T. A. Herring, Elsevier, Amsterdam, in press.

Gross, R. S., Fukumori, I., and Menemenlis, D. (2003). Atmospheric and oceanic excitation of the Earth's wobbles during 1980–2000. *J. Geophys. Res. Solid Earth* 108, 2370–2385. doi:10.1029/2002JB002143

Guo, J., and Han, Y. (2009). Seasonal and inter-annual variations of length of day and polar motion observed by SLR in 1993–2006. *Chin. Sci. Bull.* 54, 46–52. doi:10.1007/s11434-008-0504-1

Hinderer, J., Legros, H., Gire, C., and Le Mouël, J. L. (1987). Geomagnetic secular variation, core motions and implications for the Earth's wobbles. *Phys. Earth. Planet.* 49, 121–132. doi:10.1016/0031-9201(87)90136-1

Jia, S., Xu, T., and Yang, H. (2018). Two improved algorithms for LS+AR prediction model of the polar motion. *Acta Geod. Cartogr. Sinica* 47, 71–77. doi:10.11947/j.AGCS.2018.20180296

Jin, X., Liu, X., Guo, J., and Shen, Y. (2021). Analysis and prediction of polar motion using MSSA method. *Earth Planets Space* 73, 147. doi:10.1186/s40623-021-01477-2

Kalarus, M., Schuh, H., Kosek, W., Bizouard, C., and Gambis, D. (2010). Achievements of the Earth orientation parameters prediction comparison campaign. *J. Geod.* 84, 587–596. doi:10.1007/s00190-010-0387-1

Kosek, W. (2012). "Future improvements in EOP prediction," in *Geodesy for planet earth. International association of geodesy symposia vol 136*. Editors S. Kenyon, M. Pacino, and U. Marti (Berlin, Heidelberg: Springer). doi:10.1007/978-3-642-20338-1_62

Kosek, W., Kalarus, M., and Niedzielski, T. (2007). "Forecasting of the earth orientation parameters: Comparison of different algorithms," in Proceedings of the Journèes systems de dereference spatio-temporels, Observatoire de Paris, Paris, France, 17–19 Sept 2007. pp155–158.

Kosek, W., Mccarthy, D. D., and Luzum, B. J. (1998). Possible improvement of Earth orientation forecast using autocovariance prediction procedures. *J. Geod.* 72, 189–199. doi:10.1007/s001900050160

Liao, D. C., Wang, Q. J., Zhou, Y. H., Liao, X. H., and Huang, C. L. (2012). Long-term prediction of the Earth orientation parameters by the artificial neural network technique. *J. Geodyn.* 62, 87–92. doi:10.1016/j.jog.2011.12.004

Liu, L., Hsu, H., Gao, B., and Wu, B. (2000). Wavelet analysis of the variable Chandler wobble. *Geophys. Res. Lett.* 27, 3001–3004. doi:10.1029/1999GL011094

Liu, W., Li, Z., Liu, W., and Wang, F. X. (2009). Influence of EOP prediction errors on orbit prediction of navigation satellites. *GNSS World China* 34, 17–22. doi:10.13442/j.gnss.2009.06.010

Malkin, Z., and Miller, N. (2010). Chandler wobble: Two more large phase jumps revealed. *Earth Planets Space* 62, 943–947. doi:10.5047/eps.2010.11.002

Markowitz, W. (1968). *Concurrent astronomical observations for studying continental drift, polar motion, and the rotation of the Earth*. Netherlands: Springer.

McCarthy, D. D., and Luzum, B. J. (1996). Path of the mean rotational pole from 1899 to 1994. *Geophys. J. Int.* 125, 623–629. doi:10.1111/j.1365-246X.1996.tb00024.x

Modiri, S., Belda, S., Heinkelmann, R., Hoseini, M., Ferrandiz, J. M., Hoseini, M., et al. (2018). Polar motion prediction using the combination of SSA and Copula-based analysis. *Earth Planets Space* 70, 115–118. doi:10.1186/s40623-018-0888-3

Modiri, S., Belda, S., Hoseini, M., Heinkelmann, R., Ferrandiz, J. M., Heinkelmann, R., et al. (2020). A new hybrid method to improve the ultra-short-term prediction of LOD. *J. Geod.* 94, 23–14. doi:10.1007/s00190-020-01354-y

Pan, C. (2012). Linearization of the liouville equation, multiple splits of the chandler frequency, Markowitz wobbles, and error analysis. *Int. J. Geosciences* 3, 930–951. doi:10.4236/ijg.2012.325095

Pan, C. (2007). Observed multiple frequencies of the Chandler wobble. *J. Geodyn.* 44, 47–65. doi:10.1016/j.jog.2006.12.004

Schuh, H., Ulrich, M., Egger, D., Muller, J., and Schwegmann, W. (2002). Prediction of Earth orientation parameters by artificial neural networks. *J. Geodyn.* 76, 247–258. doi:10.1007/s00190-001-0242-5

Seitz, F., and Schuh, H. (2010). *Earth rotation*. Berlin: Springer.

Shaharudin, S. M., Ahmad, N., and Zainuddin, N. H. (2019). Modified singular spectrum analysis in identifying rainfall trend over peninsular Malaysia. *Indonesian J. Electr. Eng. Comput. Sci.* 15, 283–293. doi:10.11591/ijeecs.v15.i1.pp283-293

Shen, Y., Guo, J., Liu, X., Kong, Q., Guo, L., and Li, W. (2018). Long-term prediction of polar motion using a combined SSA and ARMA model. *J. Geod.* 92, 333–343. doi:10.1007/s00190-017-1065-3

Shen, Y., Guo, J., Liu, X., Wei, X., and Li, W. (2017). One hybrid model combining singular spectrum analysis and LS + ARMA for polar motion prediction. *Adv. Space Res.* 59, 513–523. doi:10.1016/j.asr.2016.10.023

Stamatakos, N., Luzum, B., Stetzler, B., Shumate, N., and Carter, M. S. (2011). "Recent improvements in IERS rapid service/prediction center products," in Proceedings of the Journées Systèmes de reference spatio-temporels, Vienna, Austria, 19-21 September 2011. pp184–187.

Su, X., Liu, L., Houtse, H., and Wang, G. (2014). Long-term polar motion prediction using normal time–frequency transform. *J. Geod.* 88, 145–155. doi:10.1007/s00190-013-0675-7

Sun, Z., Xu, T. H., Jiang, C., Yang, Y., and Jiang, N. (2019). An improved prediction algorithm for Earth's polar motion with considering the retrograde annual and semi-annual wobbles based on least squares and autoregressive model. *Acta Geod. geophys.* 54, 499–511. doi:10.1007/s40328-019-00274-4

Tabachnick, B. G., Fidell, L. S., and Ullman, J. B. (2007). *Using multivariate statistics*. Boston: Pearson.

Vautard, R., and Ghil, M. (1989). Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Phys. D.* 35, 395–424. doi:10.1016/0167-2789(89)90077-8

Wahr, J. M. (2010). The effects of the atmosphere and oceans on the Earth's wobble — I. Theory. *Geophys. J. R. Astronomical Soc.* 70, 349–372. doi:10.1111/j.1365-246X.1982.tb04972.x

Wang, G., Liu, L., Su, X., Liang, X., Yan, H., Tu, Y., et al. (2016). Variable chandler and annual wobbles in Earth's polar motion during 1900–2015. *Surv. Geophys.* 37, 1075–1093. doi:10.1007/s10712-016-9384-0

Wang, X., Cheng, Y., Wu, S., and Zhan, K. (2016). An enhanced singular spectrum analysis method for constructing nonsecular model of GPS site movement. *J. Geophys. Res. Solid Earth.* 121, 2193–2211. doi:10.1002/2015JB012573

Wilson, C. R., and Vicente, R. O. (1980). An analysis of the homogeneous ILS polar motion series. *Geophys. J. Int.* 62, 605–616. doi:10.1111/j.1365-246X.1980.tb02594.x

Xu, X. Q. (2013). *High precision prediction method of earth orientation parameters*. Beijing, China: Doctoral dissertation, Chinese Academy of Sciences.

Xu, X. Q., and Zhou, Y. H. (2015). EOP prediction using least square fitting and autoregressive filter over optimized data intervals. *Adv. Space Res.* 56, 2248–2253. doi:10.1016/j.asr.2015.08.007

Yao, Y. B., Yue, S. Q., and Chen, P. (2013). A new LS+ AR model with additional error correction for polar motion forecast. *Sci. China Earth Sci.* 56, 818–828. doi:10.1007/s11430-012-4572-3

Zhang, H., Han, Y., and Li, Z. (1986). Checking of the double-frequency feature of Chandler main peak for different periods. *Chin. J. Geophys-CH* 1, 16–27.

Zhao, D., and Lei, Y. (2020). A technique to reduce the edge effect in least squares extrapolation for enhanced Earth orientation prediction. *Stud. Geophys. Geod.* 64, 293–305. doi:10.1007/s11200-021-0546-2

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Original Research
PUBLISHED 25 May 2023
DOI 10.3389/fspas.2023.1179558

EDITED BY
Josep M. Trigo-Rodríguez,
Spanish National Research Council
(CSIC), Spain

REVIEWED BY
Akos Kereszturi,
Hungarian Academy of Sciences (MTA),
Hungary
Jianguo Yan,
Wuhan University, China

*CORRESPONDENCE
Guangfei Wei,
✉ gfwei0554@gmail.com
Hong Gan,
✉ ganhong06@gmail.com

RECEIVED 04 March 2023
ACCEPTED 12 May 2023
PUBLISHED 25 May 2023

CITATION
Wei G, Li X, Gan H and Shi Y (2023), Retrieval of lunar polar heat flow from Chang'E-2 microwave radiometer and Diviner observations.
Front. Astron. Space Sci. 10:1179558.
doi: 10.3389/fspas.2023.1179558

Retrieval of lunar polar heat flow from Chang'E-2 microwave radiometer and Diviner observations

Guangfei Wei[1,2]*, Xiongyao Li[2,3], Hong Gan[1]* and Yaolin Shi[4]

[1]Deep Space Exploration Laboratory, Hefei, China, [2]Center for Excellence in Comparative Planetology, Chinese Academy of Sciences (CAS), Hefei, China, [3]Center for Lunar and Planetary Sciences, Institute of Geochemistry CAS, Guiyang, China, [4]Key Laboratory of Computational Geodynamics, College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing, China

The internal heat flow related to the Moon's composition, interior structure, and evolution history is not well-constrained and understood on a global scale. Up to now, only two *in situ* heat flow experiments, Apollo 15 and 17 were deployed nearly 50 years ago. The measured high values of heat flow might be influenced by lateral heat at highland/mare boundaries and enhanced by heat production from radioactive elements enriched unit, and may also be disturbed by astronauts' activities. In this study, we proposed a new method to retrieve heat flows at two permanently shadowed craters, Haworth and Shoemaker of the Moon's south pole, from Chang'E-2 microwave radiometer data and Diviner observations. Our results show that the average heat flow is 4.9 ± 0.2 mW/m^2. This provides a constraint for the bulk concentration of Thorium within the lunar south polar crust 656 ± 54 ppb, which helps us understand the Moon's thermal evolution and differentiation.

KEYWORDS

Chang'E-2, microwave radiometer, Diviner, Moon, pole, heat flow, thermal model

1 Introduction

The interior heat flow provides a direct method for constraining the bulk abundance of radiogenic elements (e.g., U, Th, and K) which helps us to understand the differentiation and thermal history of the Moon (Langseth et al., 1976; Warren and Rasmussen, 1987; Wieczorek and Phillips, 2000; Siegler and Smrekar, 2014). The accurate interpretation of internal heat flow is also important to understand the volcanic history of the Moon such as the occurrence of moderately fresh "Irregular Mare Patches" (Qiao et al., 2021). Nearly 50 years ago, only two *in situ* heat flow experiments (HFEs) at the Hadley Rille and Taurus-Littrow during Apollo 15 and 17 missions obtained values of 21 ± 3 and 14 ± 2 mW/m^2, respectively (Langseth et al., 1976). However, the radiogenic elements enriched unit, Procellarum potassium, rare earth element, and phosphorus (KREEP) Terrane (PKT) (Jolliff et al., 2000), which extends to Apollo 15 landing site that likely contributes 5 ± 2 mW/m^2 to the heat flow value (Wieczorek and Phillips, 2000) and even might affect Apollo 17 heat flow measurement (Siegler and Smrekar, 2014). In addition, the two landing sites located at the boundary between the highlands and maria could also cause an additional 15%–20% heat flow (Warren and Rasmussen, 1987) and the uncertainty of the regolith thermal conductivity (Grott et al., 2010). It is worth noting that astronaut-induced surface regolith disturbance may also lead to an increase in its temperature due to the darkened surface in addition to altering the regolith texture and surface photometric properties (Nagihara et al., 2018). This might need 5–7 years to reach the equilibrium state (Langseth et al., 1976).

Langseth et al. (1976) estimated the global average heat flow value as 18 ± 2 mW/m^2 based on Th concentration, but they still cautioned that further data should be obtained to estimate the global average value. Warren and Rasmussen (1987) considered the effects of insulation of mega regolith on lunar thermal evolution and revised the global average value to 12 ± 2 mW/m^2. Nevertheless, some researchers proposed that the measured Apollo heat flows might not be representative which overestimates the global average heat flow by up to a factor of 3 due to thermal pathways and heat-producing material of the top ~100 km of the Moon (Warren and Rasmussen, 1987; Wieczorek and Phillips, 2000; Hagermann and Tanaka, 2006; Saito, 2008). Therefore, it has been concluded that the current heat flow data is not enough to constrain the thermal state of the Moon with a 100% confidence (Hagermann and Tanaka, 2006), and the limited location of Apollo HFEs emphasizes the importance of additional, geophysical different observations to better constrain the Moon's heat flow, which has been recognized as a high priority lunar science objective (Kiefer, 2012).

After Apollo, planetary heat flow mapping was focused on the potential of orbiting microwave radiometer experiments (Keihm and Langseth, 1975; Schloerb et al., 1976; Keihm, 1984; Siegler and Smrekar, 2014). Some efforts of spectral gradients and regolith physical properties are combined to interpret ground-based microwave observations for deriving heat flow (Keihm and Langseth, 1975), which unfortunately, no orbital exploration mission has been reported as of yet. Since the long-term subsurface temperature drift of Apollo HFEs was thought to stem from the astronaut activity (Langseth et al., 1976), the regolith-uninterrupted orbital microwave observations have the advantage and capability to extend the HFE to a global scale remotely. Because the temperature in the upper 1.5 m of the sunlit regolith is strongly affected by the diurnal or annual thermal waves (Kiefer, 2012), permanently shadowed regions (PSRs) of lunar poles are supposed to be ideal places to map lunar heat flow from orbit (Paige et al., 2010; Paige and Siegler, 2016). Since there is no direct solar heat perturbation within PSRs (but could be warmed up by reflected sunlight and thermal infrared radiation from surrounding crater walls), the heat flow plays an important role in the near-surface temperature (Paige et al., 2010; Paige and Siegler, 2016). This can be sensed by microwave observations which help us to characterize the lunar near-surface thermal environment and retrieve interior heat flow. Additionally, because of the large penetration depth of microwaves, the plausible buried water ice at PSRs could also be revealed by subsurface thermal regime measurements (Meng et al., 2010), which provides landing site selection for future polar missions (e.g., De Rosa et al., 2012; Wei et al., 2023).

In this study, we propose a new method to combine the Diviner and Chang'E-2 (CE-2) microwave data to retrieve heat flow at PSRs of the Moon's south pole. In Section 2, we first use nearly 10 years (from 5 July 2009, to 17 Feb 2019) of the Diviner measurements (Paige et al., 2010; Elder et al., 2019) as surface temperature constraints to improve the thermal model. Then we model microwave emission with different heat flows to match CE-2's Microwave Radiometer (MRM) observations (Zheng et al., 2019). Lastly, we can search for the best heat flow value from different MRM channels' data. In Section 3, we present the retrieved heat flow at two permanently shadowed craters. A detailed discussion of possible factors that influence our inversion results is presented in Section 4 and a conclusion is followed in Section 5.

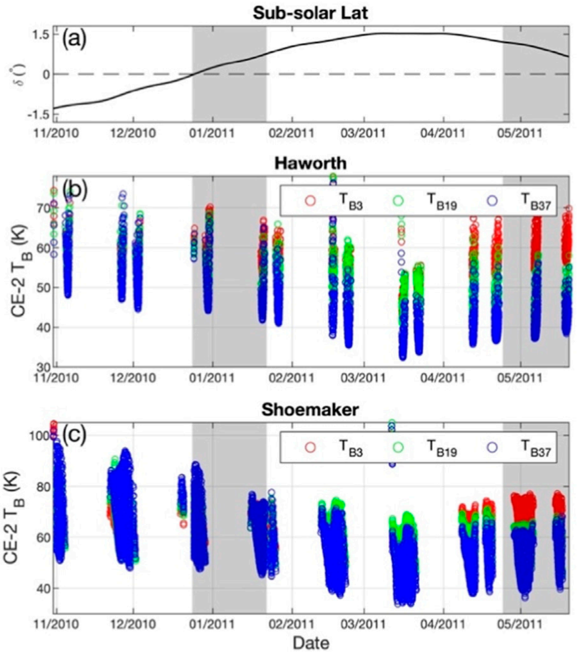


FIGURE 1
Seasonal variations of CE-2 observations at permanently shadowed craters of the Moon's south pole. **(A)** is the variation of the sub-solar latitude during the CE-2 mission. The dashed line denotes the equator. **(B)** and **(C)** are variations of T_{B3}, T_{B19} and T_{B37} near the bottom of Haworth and Shoemaker, respectively. The time span covered by gray areas indicates the observations used in this study.

2 Dataset and methodology

2.1 Chang'E-2 microwave radiometer data

Microwave Radiometer onboard CE-2 measured lunar surface with four channels ($c_i, i = 1, 2, 3, 4$), i.e., 3, 7.8, 19.35 and 37 GHz, in a polar orbit at an altitude of ~100 km (Zheng et al., 2019). Channel 1 observed the lunar surface with a spatial resolution of 25 km and the others with 17.5 km. More details about MRM have been introduced in Zheng et al. (2019). Figure 1A shows the variation of sub-solar latitude (δ) during the CE-2 mission, which covers part of the southern summer ($\delta < 0$) and most of the southern winter ($\delta > 0$). Figures 1B, C shows seasonal fluctuations of CE-2's MRM observations at 3 (T_{B3}), 19.35 (T_{B19}), and 37 GHz (T_{B37}) near the bottom of Haworth and Shoemaker craters. It can be seen that the MRM observation behaves as a seasonal variation during the mission, which has good agreement with the Diviner measurements (Williams et al., 2019). This seasonal effect is caused by the scattered sunlight and infrared radiation from ambient crater walls. The different amplitude of T_B at the same frequency between Haworth and Shoemaker also indicates different

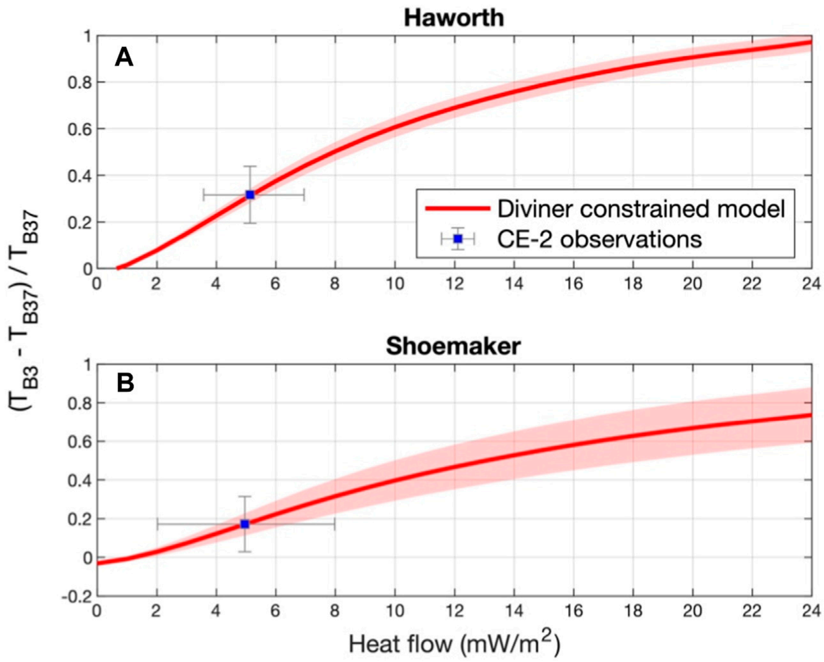


FIGURE 2
Comparison of $r_{1,4} = (c_1 - c_4)/c_4$ between CE-2 observations and Diviner constrained model at **(A)** Haworth and **(B)** Shoemaker. The vertical bar of CE-2 observations is the standard deviation, and the horizontal bar is the corresponding heat flow range. The red-shaded area is the standard deviation of modeled values.

thermal environments, thermophysical properties of regolith, and possible heat flow.

Considering the relatively small seasonal effect and more MRM data coverage during southern winter, both Diviner and MRM data obtained within this season are selected. Although the MRM observations present an apparent seasonal variation, the potential calibration issues due to CE-2 reorientation on the cold horn (a set of calibration antennas pointing to cold space for reference) for terminator orbits (Feng et al., 2020) between 22 Jan 2011, and 24 April 2011, were excluded. Thus, the data points near the bottom of Haworth and Shoemaker during southern winter are selected which are shown in the gray areas of Figure 1.

2.2 Diviner data

The Diviner Lunar Radiometer (Diviner (Paige et al., 2010)) onboard the Lunar Reconnaissance Orbiter systematically maps lunar surface thermal environment for more than 10 years since July 2009 (Williams et al., 2019). Diviner including 9 channels was designed to measure reflected solar radiation (each of the first two channels 0.35–2.8 μm) and infrared emission (other seven channels range from 7.55 to 400 μm) with a spatial resolution of ~200 m globally (Paige et al., 2010). Recently, the bolometric brightness temperature (T_{bol}) covering polar regions generated from nearly 10 years of Diviner measurements have been compiled into diurnal temperatures at southern summer/winter according to the variation of sub-solar latitudes (Williams et al., 2019). In our study, the T_{bol} data within southern winter corresponding to CE-2 observations were used as surface thermal constraints for the thermal model. And the southern summer T_{bol} data were used to investigate the maximum seasonal effect on the thermal stability of water ice in the lunar south polar region.

2.3 Improved thermal model

We use the one-dimensional thermal model to calculate the lunar polar subsurface temperatures (Vasavada, 1999; Paige et al., 2010; Vasavada et al., 2012; Hayne et al., 2017).

$$\rho(z)C(T)\frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left[K(T,z)\frac{\partial T}{\partial z}\right] \tag{1}$$

where T is the temperature, z is the depth, $\rho(z)$ is the bulk density which is dependent upon depth, $C(T)$ is the specific heat capacity which is a function of temperature, and $K(T,z)$ is the thermal conductivity of lunar regolith which depends on both temperature and depth. Vasavada et al. (2012) proposed to use the H parameter in an exponential form to characterize bulk density and thermal conductivity based on Diviner measurements; Hayne et al. (2017) derived the global distribution of the H parameter based on Vasavada et al.'s model and rock-free regolith temperature (Bandfield et al., 2011). The global mean H value 0.07 m was adopted in this study to characterize the density profile of the soil. Note that the thermophysical property of soils at extremely low temperatures (as low as 20 K) becomes complex which is a lack of comprehensive laboratory measurements. Recently, Woods-Robinson et al., 2019 derived a semi-empirical

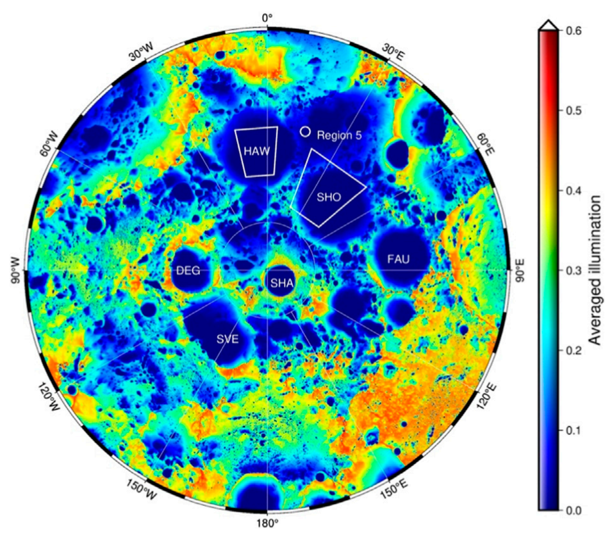


FIGURE 3
Annually averaged Sun illumination of the Moon's south pole (>85°S). The white trapezoids indicate study areas. HAW and SHO denote Haworth and Shoemaker, respectively. Other listed permanently shadowed craters are Shackleton (SHA), Faustini (FAU), De Gerlache (DEG), and Sverdrup (SVE). The white circle indicates "Region 5" which is supposed to have an upper limit heat flow value of 6 mW/m^2 (Paige and Siegler, 2016). The spatial resolution of the illumination map is 60 m/pixel, which is available at http://imbrium.mit.edu/.

model of *C* and *K* in the full range of 20–400 K by fitting lunar regolith simulant based on solid-state theory. Here, we adopt their model to calculate specific heat and thermal conductivity with the combination of Hayne et al.'s density model.

In order to solve the thermal model, we also need to input the surface temperature and interior heat flow. The thermal environment in polar regions is complex due to the complex illuminating conditions (Bussey et al., 2010; Mazarico et al., 2011; Gläser et al., 2014). Paige et al. (2010) modeled the Sun, the Earth, and the Moon as triangular meshes and used the ray tracing method to calculate lunar polar surface/subsurface temperatures at 500 m resolution. However, the simulation still presented about 15 K lower within PSRs than that of Diviner measurements during the daytime. The discrepancy is supposed to be caused by a directionally anisotropic thermal emission (Paige et al., 2010). In this study, we use diurnal variations of T_{bol} as input to constrain surface temperature instead of theoretical calculating to derive the subsurface temperatures (Wei et al., 2019; Schorghofer and Williams, 2020).

2.4 Scheme of heat flow inversion

The upper limit heat flow at "Region 5," an unnamed 3 km diameter and extremely cold crater (87.0°S, 15.3°E) which is located to the east of Haworth crater (see Figure 3) was retrieved from Diviner data (Paige and Siegler, 2016). In order to calculate the heat flow-dependent (lower boundary of Eq. 1) subsurface temperature with an unknown heat flow value, we set the upper limit heat flow as 24 mW/m^2 based on the maximum value of Apollo 15 HFE instead of Region 5's 6 mW/m^2. Thus, the heat flow can be inverted in the following steps. 1) Computing annual averaged (southern winter) subsurface temperatures with heat flows from 0 to 24 mW/m^2. 2) Modeling T_B based on the above temperature profiles (Wei et al., 2019; Feng et al., 2020; Siegler et al., 2020). 3) Match the modeled T_B and CE-2 observations and search for the best-fit heat flow value. Noting that although both the MRM observations and Diviner constrained simulations cover the same study area, they have different spatial resolutions and observation time ranges. Therefore, it could cause uncertainty by matching the modeled T_B and CE-2 observations directly. Here, we propose the following equation (Eq. 2) to minimize the discrepancy of T_B matching

$$r_{i,j} = \frac{c_i - c_j}{c_j}, \; (i = 1, j = 3, 4) \tag{2}$$

where the T_B-dependent ratio ($r_{i,j}$) is sensitive to heat flow because c_1 (3 GHz-channel) senses greater depths than the other two channels (19.35 and 37 GHz). Therefore, the heat flow can be inverted by comparing $r_{i,j}$ between MRM observations and modeled values. Figure 2 shows an example of the inversion scheme at Haworth and Shoemaker. It can be seen that the mean value of modeled $r_{1,4}$ (red lines) increases with heat flow at both craters gradually. The standard deviation of $r_{1,4}$ ratios (red shade) increase apparently with the increasing of heat flows, especially at Shoemaker. This is caused by the dominance of internal heat flow compared to the surface thermal state. The CE-2 data derived $r_{1,4}$ mean values (blue squares) then can be matched for searching the best heat flow values.

Note that the interior heat flow plays a part in near-surface temperature at PSRs and can be recorded by infrared measurements and microwave observations. However, compared to Diviner's infrared measurements at PSRs (Paige and Siegler, 2016), the longer wavelength of the microwave radiometer can sense greater depths for interior heat flow. Additionally, the discrepancy of spatial resolution between 3 GHz (25 km) and 19.35 GHz or 37 GHz (17.5 km) channels can be neglected because 1) the thermal environment including heat flow is supposed to be homogeneous for each MRM's field of view, and 2) the diurnal averaged T_B including all the data points covering the study area were derived for calculating $r_{1,3}$ and $r_{1,4}$ ratios.

3 Results

3.1 Lunar polar heat flow

To minimize the scattering effect, we select the data points near the bottom of each crater after filtering the southern winter data set. Figure 3 shows the modeled annually averaged Sun illumination of the lunar south polar region. The study area of the Haworth crater is 15–5°W, 87–88°S, and the Shoemaker's is 20–50°E, 87.3–88.6°S. It can be seen that the study areas enclosed by trapezoids at Haworth and Shoemaker present an extremely low illumination rate. Which indicates ideal places for heat flow inversion. The location of "Region 5" investigated by Paige and Siegler (2016) is also labeled in Figure 3 (white circle) which is near both of our study areas.

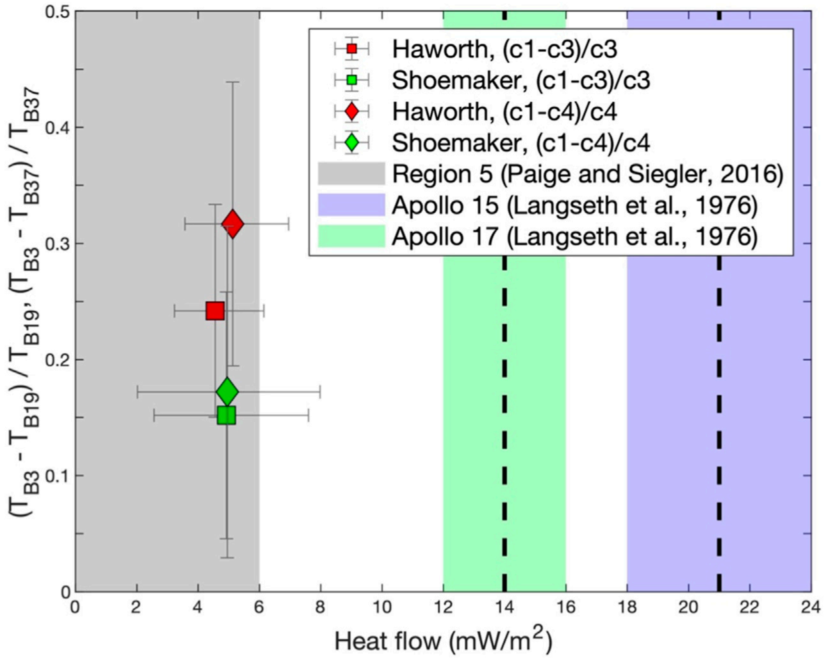


FIGURE 4
Retrieved heat flow at Haworth and Shoemaker. Other heat flow values at Region 5, Apollo 15, and 17 are presented in gray, green, and purple shades for comparison, respectively. The dashed line denotes the mean value. The vertical bar is the standard deviation of $r_{i,j}$, while the horizontal bar is the corresponding heat flow range.

Based on Eq. 2, ratios of $r_{1,3}$ and $r_{1,4}$ at study areas can be derived from observed (T_{B3}, T_{B19}) and (T_{B3}, T_{B37}), respectively. In addition, these ratios can also be derived from modeled T_B values, which are used for matching CE-2 observations (e.g., Figure 2). Figure 4 shows the retrieved heat flow at Haworth and Shoemaker using different T_B ratios. It can be seen that the averaged heat flows retrieved from $r_{1,3}$ and $r_{1,4}$ within the study area of Haworth crater (Figure 3, HAW) are ~4.6 and ~5.1 mW/m^2, respectively (Figure 4, red blocks). While the averaged heat flows within the study area of the Shoemaker crater (Figure 3, SHO) retrieved from $r_{1,3}$ and $r_{1,4}$ both are near to 4.9 mW/m^2 (Figure 4, green blocks). Note that each of the four values is averaged from thousands of data points covering the study areas based on different T_B ratios.

Although the two study areas present apparently different $r_{i,j}$ values, they have excellent consistency within their standard deviations (horizontal bars). It is worth noting that the uncertainty of inverted heat flow at the two craters varies from 1.3 to 3.0 mW/m^2 which is dependent upon the standard deviation of $r_{i,j}$. However, the good agreement of heat flow retrieved from different T_B ratios at each crater indicates the robustness of our inversion scheme. Furthermore, the averaged heat flow of the two craters is 4.9 ± 0.2 mW/m^2 which consists well with the upper limit of Region 5 (Figure 4 gray area), 6 mW/m^2 (Paige and Siegler, 2016). The similar heat flow in the two study areas suggests a relatively homogeneous interior thermal state of the Moon's south pole. Additionally, our retrieved heat flow is obviously less than that of Apollo 15 and 17 *in situ* measurements (Figure 4 green and purple shades) which are near the PKT unit.

Figure 5 shows the locations of Apollo 15 and 17 *in situ* heat flow experiments and our study areas on the Lunar Prospector gamma-ray spectrometer derived Th map (in a unit of ppm). The typical landing sites indicate that the measured heat flows might be greatly altered by deep subsurface radiogenic elements (Laneuville et al., 2013; Siegler and Smrekar, 2014) and lateral heat flow through highland/mare boundaries (Warren and Rasmussen, 1987). The study areas of PSRs are far from the PKT and highland/mare boundaries that will not enhance the sub-crust heat production and interior heat flow. In addition, the low content of surface Th at the south pole corresponds to low heat flow values while the larger Apollo 15 and 17 heat flow values correspond to highly concentrated Th. NASA's Gravity Recovery and Interior Laboratory observations show that the south pole has a roughly 40-km crustal thickness with feldspathic highland terrane associated with the low crustal thickness of the south pole Aitken basin (Wieczorek et al., 2013). The Urey ratio (the ratio between total internal heat production and total heat loss through the surface) assumed to be 0.5 is a good and representative of terrestrial planets and the Moon (Spohn and Breuer, 2002). Here, assuming the Th/U = 3.7 and K/U = 2,500 (Warren and Wasson, 1979), crustal density 2,550 kg/m^3 (Wieczorek et al., 2013), the Th bulk abundance of the lunar polar crust is constrained by 656 ± 54 ppb based on our retrieved mean heat flow (Haenel et al., 1988).

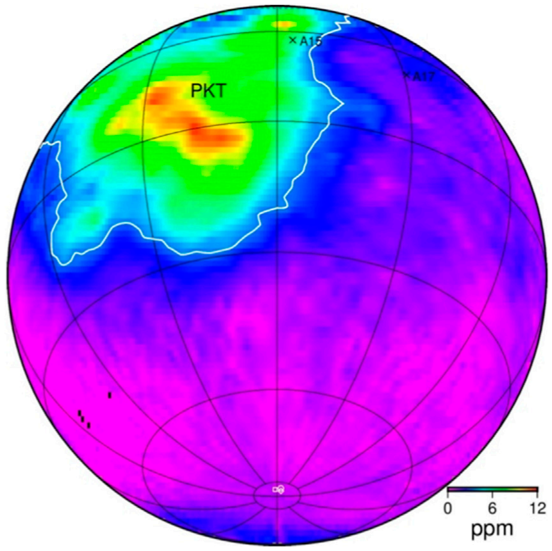


FIGURE 5
Locations of heat flow measurements on the Th map in a unit of ppm (Prettyman et al., 2006). The map is in orthographic projection centered at 0°E, 35°S. The crosses labeled A15 and A17 denote the Apollo 15 and 17 landing sites, respectively. The PKT area enclosed by the white line is constrained by the 3.5 ppm Th contour (Prettyman et al., 2006). The square, diamond, and circle denote the study areas of Haworth, Shoemaker, and Region 5 of the Moon's south pole, respectively.

3.2 Water ice effect on heat flow retrieval

Water ice is probably one of the most interesting and important elements in PSRs. The Lunar Crater Observation and Sensing Satellite (LCROSS) impact experiment confirmed that there is 5.6 ± 2.9wt% water ice at the impact site (Colaprete et al., 2010), and radar detections indicate that this water ice might present as small (<10 cm) and discrete pieces mixed with regolith rather than thick deposits of nearly pure water ice within a few meters of lunar surface (Neish et al., 2011). Furthermore, the regolith-ice mixture can behave differently with "dry" regolith in thermal emissions. Although water ice has been detected in PSRs and could be cold-trapped for billions of years (Siegler et al., 2016; Li et al., 2018), we still know less about the exact amount of water ice that appears in PSRs and the vertical distribution in the subsurface. This might lead to heat flow estimation uncertainty if water ice can influence the near-surface thermal regime obviously.

Assuming a mean regolith density of 1700 kg/m^3 with porosity 0.45 and ice density of 1,100 kg/m^3, the volumetric fraction of water ice (i.e., pore filling fraction, f_{ice}) at the LCROSS impact site is approximately 0.2. And the thermal property of ice-bearing regolith (TMP_{mix}) including density, thermal conductivity, and specific heat are calculated using the volumetric mixing model (Siegler et al., 2012), $TMP_{mix} = TMP_{reg} + p \cdot f_{ice} \cdot TMP_{ice}$, where p is the porosity of regolith which can be calculated from the bulk density of regolith, TMP_{reg} and TMP_{ice} are thermal property of regolith and ice, respectively. The dielectric constant of the mixture can be calculated based on a two-phase, three-dimensional medium mixing model at each layer (Bergman, 1978). To simplify the mixture thermal model, we suppose that water ice exists at the study area of Cabeus crater and diffuses homogeneously to different depths (z_{ice}) with the identical pore filling fraction (Figure 6).

Figure 7 shows the comparison of the T_{B3}/T_{B37} ratio between dry regolith (black line) and ice-bearing regolith with different diffusion depths (color-coded lines). It can be seen that the greater depths of water ice diffuse the less sensitive the T_{B3}/T_{B37} ratio is to the heat flow. This is mainly caused by increased lossy ice within 3 GHz sensed depths. For shallowly diffused depths, $z_{ice} < 0.2$ m, for example, the increasing of ice diffusion depth from 0.02 m (red line) to 0.2 m (green line) has relatively small influences in T_{B3}/T_{B37} ratio. However, it appears a large gap in the T_B ratio between the "dry" regolith and even small diffused ice within the top layer of 0.02 m. For greater diffusion depths, 0.5 (blue line) and 1 m (cyan line), for example, the T_{B3}/T_{B37} ratio presents a large discrepancy with respect to the "dry" regolith with increasing heat flow.

4 Discussion

4.1 Heat flow retrieved from icy regolith

Since it is unable to determine the exact amount of water ice that mixes with regolith and distributes in PSRs in what form, the *de facto* thermal emission might be more complex than we can model here. However, recent studies of near-infrared observations (Li et al., 2018) showed that only ~3.5% of cold traps of the south-polar region contained exposed water ice patches. It can be inferred that not the whole area of the study area can trap water ice and diffuse downward to a certain depth. A detailed water ice investigation in the future lunar mission such as China's unmanned Chang'E-7 lunar south polar exploration mission can help us to improve our thermal and microwave model, which could reveal more details of the subsurface thermal state.

Here, we use the scheme described in Section 2.4 to retrieve heat flow with scenarios of different ice diffusion depths that are discussed in Section 4.2. As shown in Table 1, compared to the "dry" regolith, the retrieved heat flow presents different ranges when ice diffused to the depth of 0.02, 0.2, 0.5, and 1 m at a constant pore filling fraction. The lower limit of heat flow (i.e., 0.6 mW/m^2) with different ice diffusion depths agrees well with the "dry" regolith except for the greater diffusion depth ($z_{ice} = 1$ m) which is 1.2 mW/m^2. The upper limit varies obviously (i.e., 5.9, 5.7, 3.3, 5.9 mW/m^2) with respect to "dry" regolith (4.6 mW/m^2) but is no more than 1.3 mW/m^2 in absolute difference. That is, water ice-bearing regolith has a small influence on heat flow inversion based on our ice diffusion model and retrieval scheme. It is worth noting that the retrieved heat flow values also have a good agreement with the upper limit of 6 mW/m^2 of Paige and Siegler (2016).

4.2 Rock effect

Radar observations at lunar polar regions suggested that the enhanced circular polarization ratio values might be caused by rocks rather than water ice deposits (Fa and Cai, 2013). Similar to the

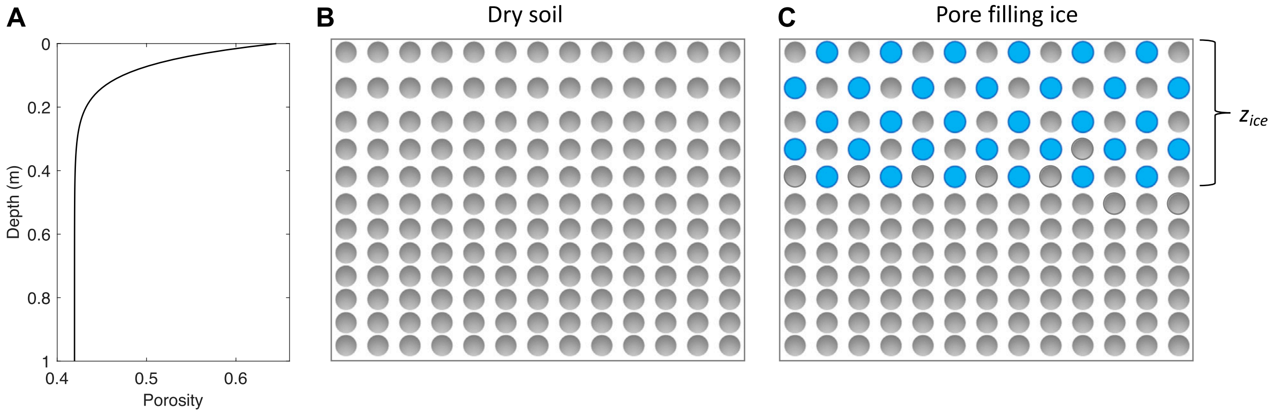


FIGURE 6
Profile of dry regolith porosity and ice-filling model. **(A)** is the porosity profile of dry lunar soil. **(B)** is the diagram of inhomogeneous dry soil from surface to a greater depth. **(C)** is the diagram with water ice that has been diffused to a certain depth of Z_{ice}.

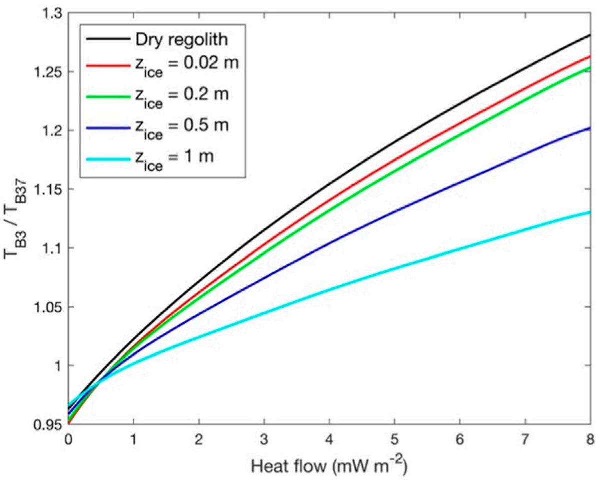


FIGURE 7
Variation of T_{B3}/T_{B37} ratio under different ice diffusion depths and interior heat flow at the Cabeus crater. The water ice pore filling fraction (f_{ice} = 0.2) is supposed to be homogeneously within each diffusion depth.

TABLE 1 Comparison of retrieved heat flows with different water ice diffusion depths at Cabeus crater. Water ice diffuses downward with a constant pore-filling fraction, f_{ice} = 0.2. Note that the cases of "dry" regolith and ice diffusion features correspond to Figure 7.

z_{ice} (m)	Dry regolith	0.02	0.2	0.5	1
HF (mW/m^2)	0.6–4.6	0.6–5.9	0.6–5.7	0.6–3.3	1.2–5.9

water ice effect, the perched and/or buried rocks can cause T_B anomaly due to their higher lossy property. The global average rock abundance (RA) within 60°N/S latitudes retrieved from Diviner measurements is 0.4% (Bandfield et al., 2011). The rock abundance referenced here is the area percentage within each pixel, and the diameter of the rock is larger than ~1 m. Since there is no quantity investigation of rock abundance at PSRs, we suppose that the polar region has a global averaged RA value. Following Bandfield et al.'s model that the regolith-rock mixture can be simplified as columns of regolith and rock within each pixel, we assume that both regolith and rocks observed by Diviner behave the same surface temperature within each bin (200 m × 200 m). Thus, we modeled the regolith-rock mixture T_B by weighting the RA value (0.4%) within each pixel. We found that the modeled $r_{i,j}$ varies slightly with heat flow 4.9 mW/m^2 compared to the rock-free (pure "dry" regolith) model, which suggests that the rock effect on heat flow inversion can be neglected based on our method.

4.3 Application for constraining water ice loss rate

Lunar polar volatiles, including water, are important for understanding the depletion, delivery, weathering, and orbital evolution of the Moon (Paige et al., 2010; Siegler et al., 2016). The water ice distribution is also crucial for *in situ* resource utilization of future lunar exploration (Cannon and Britt, 2020). Here, we discuss the thermal stability of water ice in polar regions which is controlled by the near-surface thermal environment rather than the accumulation from water-bearing meteoroids and/or the punctuation by impacts (Cannon et al., 2020). Note that the southern summer Diviner data are used here to calculate the maximum subsurface temperature and constrain the maximum sublimation rate of water ice. Neglecting the influence of water ice in regolith thermal property, we can derive subsurface temperatures using Diviner measurements and our retrieved heat flow as boundary inputs (Eq. 1). Following the work of Paige et al. (2010), Schorghofer and Taylor (2007), we calculate the loss rate of water ice at different depths and search a certain depth at which the sublimation rate equals 1 kg/(m^2 · Ga) at the lunar south pole.

Figure 8 shows the distribution of water ice at different depths with a constant sublimation rate. It can be seen apparently that water ice distributes from small to large depressions/craters. And most water ice is cold trapped in shallow depths (<10 cm) or even exposes

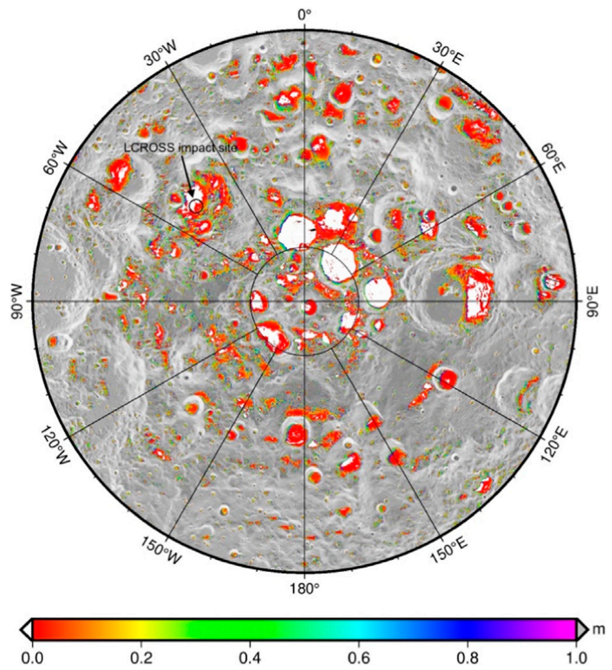


FIGURE 8
Diffusion depths of water ice with sublimation rate 1 kg/(m^2 · Ga) which is constrained by Diviner measurements and our retrieved heat flow at the lunar south polar region. The white color indicates that water ice is exposed on the surface, while the gray region indicates that the water ice can be cold trapped in a depth greater than 1 m. The circle indicates the LCROSS impact site. The outmost circle is 80°S latitude.

on the surface, for example, the LCROSS impact site (Figure 8 black circle). Compared to Paige et al.'s model which has not considered specific constraints from seasonal variation of surface T_{bol} and retrieved interior heat flow, our result provides an additional thermal constraint for the thermal stability of water ice at cold traps. For example, our model suggests that the partially sunlit Amundsen crater floor (84.5°S, 82.8°E) may harbor water ice with a loss rate of 1 mm/Ga at greater depths (>1 m) while Paige et al.'s model indicates <0.5 m depth. Thus, accurate constraints of subsurface temperatures can help us to evaluate water ice thermal stability, which is significant for investigating the depletion history of volatiles and even landing site selection for mining water ice in the future lunar exploration (Cannon and Britt, 2020).

5 Conclusion

In this study, we retrieved interior heat flow at Haworth and Shoemaker from ~10 years of Diviner measurements and CE-2 observations. The retrieved heat flows at the two craters have excellent consistency, which also agrees well with Diviner-derived Region 5 (Paige and Siegler, 2016). The average value is 4.9 ± 0.2 mW/m^2 which suggests a homogeneous interior heat flux. Assuming a uniform heat flow at the lunar south pole, the bulk content of Th of the lunar polar crust is constrained by 656 ± 54 ppb based on the Urey ratio of the Moon. In addition, we also estimated the thermal stability of water ice which is constrained by Diviner observations and retrieved heat flow. We found that water ice with a constant loss rate of 1 kg/(m^2·Ga) can be cold-trapped from the surface to shallow depths (<0.1 m) within most parts of the PSRs of the Moon. Finally, our study provides a new method to constrain the lunar subsurface thermal state at the non-PKT area by reconciling the unprecedented MRM data and longevous Diviner observations. The retrieved heat flow helps us to understand the subsurface thermal state and thermal history of the Moon.

Data availability statement

The Diviner Tbol data set was obtained from the NASA Planetary Data System Geosciences Node, available at http://pds-geosciences.wustl.edu/missions/lro/diviner.htm. CE-2 Microwave Radiometer data were provided by the Ground Research and Application System of the Chinese Lunar Exploration Program and can be downloaded from https://doi.org/10.5281/zenodo.5164204.

Author contributions

GW conceptualized this study and processed the data. XL, HG, and YS provided data interpretations. GW prepared figures and tables and wrote the draft. HG discussed the results and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

GW was supported by the National Key Research and Development Program of China (2022YFF0711400), the B-type Strategic Priority Program of the Chinese Academy of Sciences (XDB 41000000), the National Natural Science Foundation of China (42241154), and Science and Technology Program of Guizhou Province [QKHJC-ZK(2023)-476]. HG was supported by the National Natural Science Foundation of China (41903058), Frontier Science Research Program of Deep Space Exploration Laboratory (2022-QYKYJH-HXYF023), and Science and Technology Program of Guizhou Province [QKHJC-ZK(2023)-131].

Acknowledgments

We thank Stephen J. Keihm for his constructive comments and discussion of the early version of the manuscript. We also thank Jean-Pierre Williams for the helpful discussion about using Diviner T_{bol} data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Bandfield, J. L., Ghent, R. R., Vasavada, A. R., Paige, D. A., Lawrence, S. J., and Robinson, M. S. (2011). Lunar surface rock abundance and regolith fines temperatures derived from LRO Diviner Radiometer data. *J. Geophys. Res.* 116, E00H02. doi:10.1029/2011JE003866

Bergman, D. J. (1978). The dielectric constant of a composite material—a problem in classical physics. *Phys. Rep.* 43, 377–407. doi:10.1016/0370-1573(78)90009-1

Bussey, D. B. J., McGovern, J. A., Spudis, P. D., Neish, C. D., Noda, H., Ishihara, Y., et al. (2010). Illumination conditions of the south pole of the Moon derived using Kaguya topography. *Icarus* 208, 558–564. doi:10.1016/j.icarus.2010.03.028

Cannon, K. M., and Britt, D. T. (2020). A geologic model for lunar ice deposits at mining scales. *Icarus* 347, 113778. doi:10.1016/j.icarus.2020.113778

Cannon, K. M., Deutsch, A. N., Head, J. W., and Britt, D. T. (2020). Stratigraphy of ice and ejecta deposits at the lunar Poles. *Geophys. Res. Lett.* 47. doi:10.1029/2020GL088920

Colaprete, A., Schultz, P., Heldmann, J., Wooden, D., Shirley, M., Ennico, K., et al. (2010). Detection of water in the LCROSS ejecta plume. *Science* 330, 463–468. doi:10.1126/science.1186986

De Rosa, D., Bussey, B., Cahill, J. T., Lutz, T., Crawford, I. A., Hackwill, T., et al. (2012). Characterisation of potential landing sites for the European Space Agency's Lunar Lander project. *Planet. Space Sci.* 74, 224–246. doi:10.1016/j.pss.2012.08.002

Elder, C. M., Douglass, B., Ghent, R. R., Hayne, P. O., Williams, J. -P., Bandfield, J. L., et al. (2019). The subsurface coherent rock content of the moon as revealed by cold-spot craters. *J. Geophys. Res. Planets* 124, 3373–3384. doi:10.1029/2019JE006128

Fa, W., and Cai, Y. (2013). Circular polarization ratio characteristics of impact craters from mini-RF observations and implications for ice detection at the polar regions of the moon: Lunar cpr properties for ice detection. *J. Geophys. Res. Planets* 118, 1582–1608. doi:10.1002/jgre.20110

Feng, J., Siegler, M. A., and Hayne, P. O. (2020). New constraints on thermal and dielectric properties of lunar regolith from LRO diviner and CE-2 microwave radiometer. *J. Geophys. Res. Planets* 125. doi:10.1029/2019JE006130

Gläser, P., Scholten, F., De Rosa, D., Marco Figuera, R., Oberst, J., Mazarico, E., et al. (2014). Illumination conditions at the lunar south pole using high resolution Digital Terrain Models from LOLA. *Icarus* 243, 78–90. doi:10.1016/j.icarus.2014.08.013

Grott, M., Knollenberg, J., and Krause, C. (2010). Apollo lunar heat flow experiment revisited: A critical reassessment of the *in situ* thermal conductivity determination. *J. Geophys. Res.* 115, E11005. doi:10.1029/2010JE003612

Haenel, R., Rybach, L., and Stegena, L. (1988). *Handbook of terrestrial heat-flow density determination*. Springer.Berlin, Germany.

Hagermann, A., and Tanaka, S. (2006). Ejecta deposit thickness, heat flow, and a critical ambiguity on the Moon. *Geophys. Res. Lett.* 33, L19203. doi:10.1029/2006GL027030

Hayne, P. O., Bandfield, J. L., Siegler, M. A., Vasavada, A. R., Ghent, R. R., Williams, J.-P., et al. (2017). Global regolith thermophysical properties of the moon from the diviner lunar radiometer experiment: Lunar regolith thermophysical properties. *J. Geophys. Res. Planets* 122, 2371–2400. doi:10.1002/2017JE005387

Jolliff, B. L., Gillis, J. J., Haskin, L. A., Korotev, R. L., and Wieczorek, M. A. (2000). Major lunar crustal terranes: Surface expressions and crust-mantle origins. *J. Geophys. Res.* 105, 4197–4216. doi:10.1029/1999JE001103

Keihm, S. J. (1984). Interpretation of the lunar microwave brightness temperature spectrum: Feasibility of orbital heat flow mapping. *Icarus* 60, 568–589. doi:10.1016/0019-1035(84)90165-9

Keihm, S. J., and Langseth, M. G. (1975). Lunar microwave brightness temperature observations reevaluated in the light of Apollo program findings. *Icarus* 24, 211–230. doi:10.1016/0019-1035(75)90100-1

Kiefer, W. S. (2012). Lunar heat flow experiments: Science objectives and a strategy for minimizing the effects of lander-induced perturbations. *Planet. Space Sci.* 60, 155–165. doi:10.1016/j.pss.2011.07.016

Laneuville, M., Wieczorek, M. A., Breuer, D., and Tosi, N. (2013). Asymmetric thermal evolution of the moon: Lunar evolution. *J. Geophys. Res. Planets* 118, 1435–1452. doi:10.1002/jgre.20103

Langseth, M. G., Keihm, S. J., and Peters, K. (1976). Revised lunar heat-flow values. *Present. A. T. Proc. Lunar Sci. Conf.* 7th, 3143–3171.

Li, S., Lucey, P. G., Milliken, R. E., Hayne, P. O., Fisher, E., Williams, J.-P., et al. (2018). Direct evidence of surface exposed water ice in the lunar polar regions. *Proc. Natl. Acad. Sci. U. S. A.* 115, 8907–8912. doi:10.1073/pnas.1802345115

Mazarico, E., Neumann, G. A., Smith, D. E., Zuber, M. T., and Torrence, M. H. (2011). Illumination conditions of the lunar polar regions using LOLA topography. *Icarus* 211, 1066–1081. doi:10.1016/j.icarus.2010.10.030

Meng, Z., Chen, S., Osei, E. M., Wang, Z., and Cui, T. (2010). Research on water ice content in Cabeus crater using the data from the microwave radiometer onboard Chang'e-1 satellite. *Sci. China Phys. Mech. Astron.* 53, 2172–2178. doi:10.1007/s11433-010-4159-y

Nagihara, S., Kiefer, W. S., Taylor, P. T., Williams, D. R., and Nakamura, Y. (2018). Examination of the long-term subsurface warming observed at the Apollo 15 and 17 sites utilizing the newly restored heat flow experiment data from 1975 to 1977. *J. Geophys. Res. Planets* 123, 1125–1139. doi:10.1029/2018JE005579

Neish, C. D., Bussey, D. B. J., Spudis, P., Marshall, W., Thomson, B. J., Patterson, G. W., et al. (2011). The nature of lunar volatiles as revealed by Mini-RF observations of the LCROSS impact site. *J. Geophys. Res.* 116, E01005. doi:10.1029/2010JE003647

Paige, D. A., and Siegler, M. A. 2016"New constraints on lunar heat flow rates from LRO diviner lunar radiometer experiment polar observations," in Proceedings of the 47th Lunar and Planetary Science Conference. Presented at the 47th Lunar and Planetary Science Conference, Houston, TX, USA, January 2016, 2753.

Paige, D. A., Siegler, M. A., Zhang, J. A., Hayne, P. O., Foote, E. J., Bennett, K. A., et al. (2010). Diviner lunar radiometer observations of cold traps in the moon's south polar region. *Science* 330, 479–482. doi:10.1126/science.1187726

Prettyman, T. H., Hagerty, J. J., Elphic, R. C., Feldman, W. C., Lawrence, D. J., McKinney, G. W., et al. (2006). Elemental composition of the lunar surface: Analysis of gamma ray spectroscopy data from lunar prospector: Lunar elemental composition. *J. Geophys. Res.* 111.2656 doi:10.1029/2005je002656

Qiao, L., Head, J. W., Wilson, L., and Ling, Z. (2021). Ina lunar irregular mare patch mission concepts: Distinguishing between ancient and modern volcanism models. *Planet. Sci. J.* 2, 66. doi:10.3847/PSJ/abeaa0

Saito, Y. 2008"The long term temperature variation in the lunar subsurface," in Proceedings of the Presented at the Lunar and Planetary Science Conference 39th (2008), Houston, TX, USA, March 2008, 1663.

Schloerb, F. P., Muhleman, D. O., and Berge, G. L. (1976). Lunar heat flow and regolith structure inferred from interferometric observations at a wavelength of 49.3 cm. *Icarus* 29, 329–341. doi:10.1016/0019-1035(76)90137-8

Schorghofer, N., and Taylor, G. J. (2007). Subsurface migration of H_2O at lunar cold traps. *J. Geophys. Res.* 112, E02010. doi:10.1029/2006JE002779

Schorghofer, N., and Williams, J.-P. (2020). Mapping of ice storage processes on the moon with time-dependent temperatures. *Planet. Sci. J.* 1, 54. doi:10.3847/PSJ/abb6ff

Siegler, M. A., Feng, J., Lucey, P. G., Ghent, R. R., Hayne, P. O., and White, M. N. (2020). Lunar titanium and frequency-dependent microwave loss tangent as constrained by the Chang'E-2 MRM and LRO diviner lunar radiometers. *J. Geophys. Res. Planets* 125. doi:10.1029/2020JE006405

Siegler, M., Aharonson, O., Carey, E., Choukroun, M., Hudson, T., Schorghofer, N., et al. (2012). Measurements of thermal properties of icy mars regolith analogs: Thermal properties OF ICY mars regolith. *J. Geophys. Res.* 117. doi:10.1029/2011je003938

Siegler, M. A., Miller, R. S., Keane, J. T., Laneuville, M., Paige, D. A., Matsuyama, I., et al. (2016). Lunar true polar wander inferred from polar hydrogen. *Nature* 531, 480–484. doi:10.1038/nature17166

Siegler, M. A., and Smrekar, S. E. (2014). Lunar heat flow: Regional prospective of the Apollo landing sites: Regional lunar heat flow. *J. Geophys. Res. Planets* 119, 47–63. doi:10.1002/2013JE004453

Spohn, T., and Breuer, D. 2002"Surface heat flow, radiogenic heating, and the evolution of the moon," in Proceedings of the EGS 27th General Assembly, Nice, France, April 2002, 6000.

Vasavada, A. (1999). Near-surface temperatures on mercury and the moon and the stability of polar ice deposits. *Icarus* 141, 179–193. doi:10.1006/icar.1999.6175

Vasavada, A. R., Bandfield, J. L., Greenhagen, B. T., Hayne, P. O., Siegler, M. A., Williams, J.-P., et al. (2012). Lunar equatorial surface temperatures and regolith properties from the diviner lunar radiometer experiment: Lunar equatorial regolith properties. *J. Geophys. Res.* 117.3987 doi:10.1029/2011JE003987

Warren, P. H., and Rasmussen, K. L. (1987). Megaregolith insulation, internal temperatures, and bulk uranium content of the moon. *J. Geophys. Res.* 92, 3453. doi:10.1029/JB092iB05p03453

Warren, P. H., and Wasson, J. T. (1979). The origin of KREEP. *Rev. Geophys.* 17, 73. doi:10.1029/RG017i001p00073

Wei, G., Li, X., Gan, H., Blewett, D. T., Neish, C. D., and Greenhagen, B. T. (2019). A new method for simulation of lunar microwave brightness temperatures and evaluation of Chang'E-2 MRM data using thermal constraints from diviner. *J. Geophys. Res. Planets*.29 5858, doi:10.1029/2018JE005858

Wei, G., Li, X., Zhang, W., Tian, Y., Jiang, S., Wang, C., et al. (2023). Illumination conditions near the Moon's south pole: Implication for a concept design of China's Chang'E−7 lunar polar exploration. *Acta Astronaut.* 208, 74–81. doi:10.1016/j.actaastro.2023.03.022

Wieczorek, M. A., Neumann, G. A., Nimmo, F., Kiefer, W. S., Taylor, G. J., Melosh, H. J., et al. (2013). The crust of the moon as seen by GRAIL. *Science* 339, 671–675. doi:10.1126/science.1231530

Wieczorek, M. A., and Phillips, R. J. (2000). The "Procellarum KREEP Terrane": Implications for mare volcanism and lunar evolution. *J. Geophys. Res.* 105, 20417–20430. doi:10.1029/1999JE001092

Williams, J. -P., Greenhagen, B. T., Paige, D. A., Schorghofer, N., Sefton-Nash, E., Hayne, P. O., et al. (2019). Seasonal polar temperatures on the moon. *J. Geophys. Res. Planets* 124, 2505–2521. doi:10.1029/2019JE006028

Woods-Robinson, R., Siegler, M. A., and Paige, D. A. (2019). A model for the thermophysical properties of lunar regolith at low temperatures. *J. Geophys. Res. Planets* 124, 1989–2011. doi:10.1029/2019JE005955

Zheng, Y.-C., Chan, K. L., Tsang, K. T., Zhu, Y.-C., Hu, G. P., Blewett, D. T., et al. (2019). Analysis of Chang'E-2 brightness temperature data and production of high spatial resolution microwave maps of the Moon. *Icarus* 319, 627–644. doi:10.1016/j.icarus.2018.09.036

TYPE Original Research
PUBLISHED 06 June 2023
DOI 10.3389/fspas.2023.1213294

*CORRESPONDENCE
Hong Gan,
✉ ganhong06@gmail.com
Guangfei Wei,
✉ gfwei0554@gmail.com

RECEIVED 27 April 2023
ACCEPTED 24 May 2023
PUBLISHED 06 June 2023

CITATION
Gan H, Wei G, Zhang X, Xia G and Shi J (2023), Experimental study on electrostatic migration of different mineral particles composing lunar dust under electron irradiation.
Front. Astron. Space Sci. 10:1213294.
doi: 10.3389/fspas.2023.1213294

Experimental study on electrostatic migration of different mineral particles composing lunar dust under electron irradiation

Hong Gan[1]*, Guangfei Wei[1,2]*, Xiao Zhang[1,3], Guojun Xia[1,3] and Jingjing Shi[1]

[1]Deep Space Exploration Laboratory, Hefei, China, [2]Center for Excellence in Comparative Planetology, Chinese Academy of Sciences, Hefei, China, [3]Shanghai Institute of Satellite Engineering, Shanghai, China

Dust electrostatic migration is one of the primary causes of dust events on the surface of airless celestial bodies, which can result in multiple dust phenomena, including lunar horizon glow and dust pools, and also offer crucial insights into the evolution of the surface materials of airless celestial bodies. To date, the characteristics and laws of dust electrostatic migration are still not well understood. Here, we report the electrostatic migration characteristics of pyroxene, olivine, and ilmenite particles measured by the laser Doppler method and compare them with the findings of anorthite particles in previous experiments to recognize the differential migration characteristics of different mineral components in the lunar dust. The results demonstrate that the particle sizes of most moving pyroxene and olivine particles are in the range of 0–10 μm, and their vertical velocities are typically less than 2 m s^{-1}, which are consistent with the previous studies. The cohesive force between the dust particles is probably what causes the difference in the migration rates of various insulating minerals. Note that no moving ilmenite particles were detected probably due to the material's good conductivity. It can be speculated that the concentration of fugitive dust in the lunar mare is affected by the ilmenite content, but further research is needed to determine how low ilmenite content can cause dust migration.

KEYWORDS

lunar surface, lunar dust, mineral particles, electron irradiation, electrostatic migration

1 Introduction

In the 1960s, the lunar horizon glow was observed several times by the Surveyor 5, 6, and 7 missions (Criswell, 1973; Rennilson and Criswell, 1974; Colwell et al., 2007). This is the earliest space phenomenon that may be related to electrostatic dust migration. Subsequently, the astronauts of Apollo 17 described high-altitude lunar horizon glow that was visible above the lunar surface (McCoy, 1976; Zook and McCoy, 1991). Recently, five dust enhancement events during the Lunar Atmosphere and Dust Environment Explorer mission were recognized (Xie et al., 2020), which provides concrete evidence of electrostatic dust migration above the lunar surface. This phenomenon was also confirmed by the dust deposition phenomenon at a height of 190 cm above the surface of the Chang' E-3 landing site (Li et al., 2019; Zhang et al., 2020) and the reflectance feature difference of lunar rocks and regolith observed by the Chang' E-3 mission due to long-term dust activity (Yan et al., 2019).

Meanwhile, lunar dust is a very important space environmental factor on the lunar surface. The electrostatic levitation or migration of lunar dust will cause potential hazards to spacecraft and engineering systems (Gaier and Jaworske, 2007; Kawamoto et al., 2011; Zakharov et al., 2020), such as the malfunction of seals for on/off mechanisms or space suits and the functional degradation of the heat rejection systems. It is closely related to the achievement of space exploration scientific goals and even the success or failure of the entire exploration mission. Therefore, the study of the charging and migration mechanism of lunar dust in the lunar space environment not only provides scientific support for the research on the causes of the lunar horizon glow observed during the Surveyor and Apollo missions, but also provides theoretical guidance for the design of lunar dust removal strategies on optics, mechanisms, and human space flight hardware in subsequent lunar exploration missions.

In the last two decades, a series of experiments have been carried out to study the causes, conditions, and characteristics underlying the migration behavior of dust particles. The interaction of electrons and dust particles was believed to play an important role in dust charging and migration. For the individual micron-/submicron-size dust particles, the charging characteristics of dust particles from Apollo 11 and 17 dust samples (0.2–13 μm) (Abbas et al., 2010) and Lunar Highlands Type regolith simulants (0.3–3 μm) (Němeček et al., 2011) were measured under electron irradiation. The results showed that the secondary emission yield increases with the primary beam energy, reaching a maximum of approximately 3 at 350 eV, and the surface potentials are independent of the grain mass, shape, and dimensions for the grains with sizes between 0.3 and 3 μm and electron energies lower than 200 eV (Němeček et al., 2011), while for larger particles, the surface potentials of dust grains are a function of the particle size, electron energy, and electron flux density (Abbas et al., 2010). The results could be used to explain the charging of the isolated suspended dust particle. However, the lunar dust layer is accumulated by the countless dust particles. There are substantial differences between the charging and migration characteristics of individual small-size dust grains and the corresponding bulk materials. Lunar Prospector measurements of surface potentials and electron fluxes from the lunar regolith suggest that the secondary electron yield from lunar regolith is close to 1.0–1.1 (Halekas et al., 2009a), which is about three times smaller than the yield in the previous studies. This is consistent with the observation that the lunar surface potential usually remains negative in the shadow regions under various solar wind conditions (Halekas et al., 2009b).

It has been proved that dust particles can be released from the bulk samples and then migrate under electron irradiation (Wang et al., 2010). The particle size, vertical velocity, transport rate, and launch angle distribution of lofted dust particles were experimentally studied. The results showed that the diameters of most dust particles are below 10 μm and the verticle velocities are constrained by diameters (Carroll et al., 2020; Gan et al., 2022). The micron-sized insulating dust particles jump to centimetres or tens of centimetres in height with an initial speed below approximately 2 m s^{-1} exposure to electron irradiation with or without plasmas, resulting in an equivalent height below approximately **1** m on the lunar surface (Wang et al., 2016; Örger et al., 2019; Gan et al., 2022). The height is comparable to that of the so-called lunar horizon glow. The transport rate is a function of time (Hood et al., 2018; Gan et al., 2022). Based on the experimental results, the patched charge model was proposed to explain that the intense particle-particle repulsive forces are due to the re-absorption of secondary electrons at the walls of microcavities formed between neighboring dust particles (Wang et al., 2016; Schwan et al., 2017). This means that the net charge of dust particles lofted on the surfaces of airless planetary bodies is negative. Due to the increased packing density, the contact area between particles increases and the dust activity weakens, since the packing density is related to the number and size of the microcavities (Örger et al., 2019). In addition, most particles were launched from the sample surface in a direction closer to the surface normal. However, the peak value of the launch angles is around 43.5°–45° from the surface normal, and the launching angle distribution of dust particles in two horizontal directions (both ± x-axis) is quite similar (Örger et al., 2021).

Lunar dust is made of fine minerals, glass, and debris on the lunar surface. The charging properties and migration characteristics of each component of lunar dust are very different, but have not been comprehensively studied. Previous studies focused on the migration characteristics of lunar dust samples (mixture), lunar dust simulants (mixture), and silica dust. The results are not conducive to explaining the differences in migration characteristics of each component. Therefore, based on the previous measurements of the electrostatic migration characteristics of anorthite particles (Gan et al., 2022), here we present the laboratory results on the electrostatic migration characteristics of pyroxene, olivine, and ilmenite particles. The differences in the electrostatic migration characteristics of each component are compared to confirm the main influencing factors.

2 Experimental setup

The Lunar (Planetary) Dust Environment Simulation Platform, as shown in **Figure 1**, utilizes a spherical stainless-steel vacuum chamber for conducting experiments. The platform was designed to experimentally simulate the migration process of charged dust particles, and then the influence of dust components on its charging and transport characteristics will be discussed. The main testing chamber is 50 cm in diameter and has a base pressure of 10^{-6} Pa. On the chamber's ceiling, an electron gun is mounted and electrons are injected into the sample surface at a 45° angle. The energies of electrons (E_0) range from 0 to 500 eV and the emission currents (I_0) range from 1 to 500 μA. The incident electrons are denser and more energetic than the typical solar wind, which can represent the condition parameters in the special plasma environment, such as in the lunar crater wake, plasma sheet, and magnetosheath regions, as well as during the solar energetic particle (SEP) events. While the current is dramatically enhanced to obtain a larger dust flux.

Under electron irradiation, dust particles are charged and released from the surface. Phase Doppler Particle Analyzer (PDPA)/Laser Doppler Velocimeter (LDV) system from TSI Incorporated (Saint Paul, MN 55126, United States,

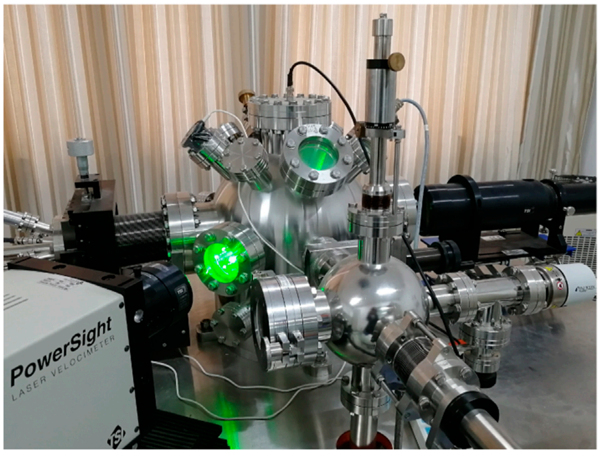


FIGURE 1
Pictorial view of the lunar (Planetary) dust environment simulation platform.

https://www.tsi.com/products/fluid-mechanics-systems/) has been employed to detect the moving dust particles. The intersection of two laser beams creates a series of light and dark fringes. Since the moving particles scatter light when crossing the bright fringe but scatter no light when passing the dark fringe, a fluctuating pattern of scattered light intensity with a frequency (f_D) is generated. The velocities of moving particles (V) can be calculated by multiplying the frequency by the known fringe spacing (δ_f):

$$V = \delta_f \cdot f_D, \tag{1}$$

where δ_f is determined by the laser wavelength (λ) and the angle (2κ) between beams:

$$\delta_f = \frac{\lambda}{2\sin(\kappa)}. \tag{2}$$

The particle diameter (D) can be obtained by the phase Doppler technique, which is an extension of LDV, allowing the size measurements of spherical particles to be obtained. The spatial frequency (spacing between the scattered fringes at the light collecting optics) of the scattered fringes can be measured as a phase shift (Φ) between two electrical signals resulting from the scattered light. The diameter is correlated to the phase shift between two laser beams from two corresponding detectors at different positions:

$$D = \frac{L_f \delta_f \Phi}{2\pi \Delta l} K, \tag{3}$$

where K is the optical constant, Δl is the space between two detectors, and L_f is the lens focal length. The detectable diameters of the moving dust particles are between 0.5 and 90 μm and velocities are between −90 and 283 m s^{-1}.

In this work, we first crushed and sieved the pyroxene and olivine samples from Damaping, Hebei Province, China and the ilmenite sample from Panzhihua, Sichuan Province, China. Second, the dust particles with diameters smaller than 40 μm were loosely placed in a square stainless-steel box (8 mm wide and 1 mm deep) for testing. The mean particle sizes of pyroxene, olivine, and ilmenite particles are 10.46, 11.48, and 4.57 μm, respectively. An insulating plate is placed between the sample box and the grounded sample stage, and the sample box is directly connected to a voltmeter. The beam spot of incident electrons was adjusted to 10 mm to prevent the electron beam impact/shadow boundaries from affecting particle transport (Wang et al., 2011). Third, before the electron beam incident, the pyroxene particles were baked at 110°C for 24 h in atmospheric conditions and deaerated for 12 h in a vacuum chamber with a pressure of 10^{-6} Pa to eliminate adsorbed gases.

Finally, under electron irradiation, the dust particles obtain charges that can be maintained for a long time. After overcoming the gravity and cohesive force, the dust particles migrate. The diameters and velocities of the moving dust particles were measured by PDPA/LDV under different electron irradiation conditions. Note that the height of the measured point (h_0), which refers to the distance between the sample surface and the measured point, was set to 4 mm here, which can be adjusted by moving the sample stage vertically.

3 Experimental results

PDPA/LDV system did not detect the moving ilmenite particles. However, both pyroxene and olivine particles had massive electrostatic migration and showed similar movement characteristics. Here, we take the electrostatic migration of pyroxene particles as an example to elaborate on its motion characteristics.

3.1 Diameter and velocity characteristics of moving pyroxene particles

We recorded 19 groups of experimental results on pyroxene migration, including the dust migration characteristics under the incident electron beam with different energies and currents. Though each experiment only lasted approximately 3–7 min, up to more than 10,000 particles were detected. However, most of the particle size and vertical velocity are still in the range of 0–10 μm and 0–2 m s^{-1}, respectively, similar to the migration characteristics of anorthite particles (Gan et al., 2022). **Figure 2**, for example, presents the migration characteristics of pyroxene particles under electron irradiation with an energy of 350 eV and a current of 500 μA, including the distributions of diameters (**Figure 2A**) and vertical velocities (**Figure 2B**), and the relationship of diameters and vertical velocities (**Figure 2C**). The red line in **Figures 2A, B** represents the cumulative particle counts. The electron beam was continuously injected for 411 s. A total of 7,760 diameter signals and 11,517 vertical velocity signals were recorded. There are 7,605 moving particles with a diameter below 10 μm (accounting for 98%), while 6,769 particles smaller than 5 μm (accounting for 87%). The moving particles with velocities in the 0–2 m s^{-1} and 0–1 m s^{-1} ranges consist of 11,447 and 10,071 particles, respectively, accounting for 99% and 87% of the total. Note that during the first few seconds, the moving dust particles increased significantly due to the extremely loose top surface layer. This layer was either the topmost layer of the bulk sample or was formed by the sedimentation of moving particles in the previous set of experiments. **Figure 2C** indicates that the vertical velocities of the moving particles are constrained by the

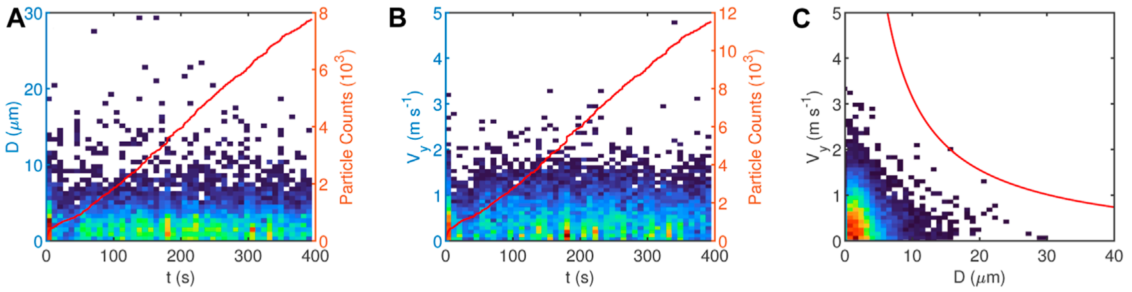


FIGURE 2
Distributions of diameters **(A)** and vertical velocities **(B)**, and the relationship between diameters and vertical velocities **(C)**. The red lines in **(A)** and **(B)** represent cumulative counts of particle size and velocity signals of moving pyroxene particles, and the red line in **(C)** is the envelope curve based on the constraint relation between vertical velocity and particle size.

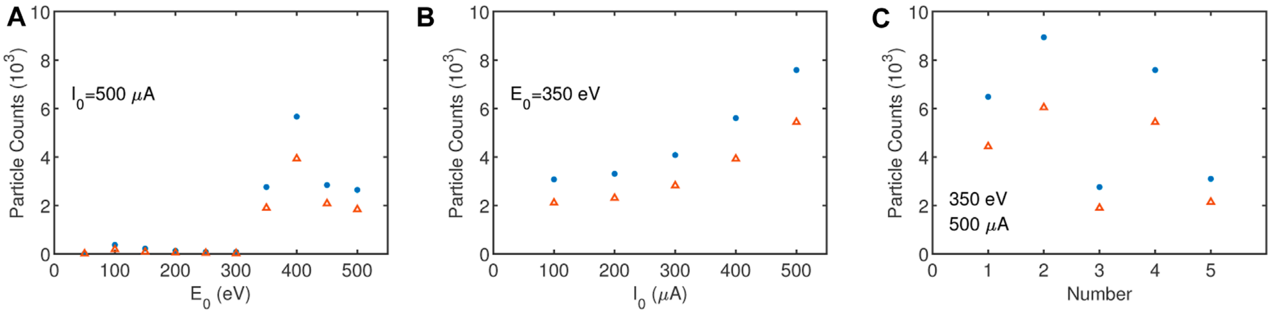


FIGURE 3
Relationships of dust counts with the electron energy **(A)** and electron current **(B)**. **(C)** shows the results of multiple parallel experiments under electron irradiation with the same energy and current conditions. The blue dot and the red triangle indicate the signal counts of vertical velocities and diameters of the moving particles, respectively.

particle size. The red envelope curve is given based on the energy conservation law.

3.2 Influencing factor of pyroxene migration characteristics

To analyze the main factors affecting the motion characteristics of pyroxene particles, Figure 3 presents the relationships of dust counts with the electron energy (a) and current (b), as well as the results (c) of multiple parallel experiments under electron irradiation with an energy of 350 eV and a current of 500 μA. Particle counts are the number of moving particles within the first 5 min of each experiment, excluding the first few seconds when the dust flux increases significantly. Figure 3A shows that the number of moving pyroxene particles changes with an energy of incident electrons when the current of incident electrons was 500 μA. As shown in Figure 3A, when the energy of incident electrons is less than 300 eV, the total number of lofted pyroxene particles is less than 400 (based on the verticle velocity signals), while the total number of pyroxene particles is more than 2,500 (based on the verticle velocity signals) when the electron energy is higher than 350 eV. This means the total number of the moving pyroxene particles has a dividing line at the electron energy of 300–350 eV. The reason may be that the secondary electron yield of pyroxene particles varies with an energy of the incident electrons, which leads to a change in its charging characteristics and surface potential distribution.

Figure 3B shows that the number of moving pyroxene particles varies with the current of incident electrons when the electron energy was set to 350 eV. It can be found that the number of pyroxene particles is always higher than 3,000 (based on the verticle velocity signals) for the incident electron current in the range of 100–500 μA, and the number increases as the electron current increases. Furthermore, we consider the variation in the number of moving pyroxene particles with depth under certain electron energy and current conditions. Therefore, multiple experiments under electron irradiation with an energy of 350 eV and a current of 500 μA were interspersed throughout the experiments, as shown in Figure 3C. The number of moving pyroxene particles is between 2,700 and 9,000. The possible reason for the large variation range of dust counts is the uneven density of dust accumulation. For naturally loose dust with a thickness below 1 mm, the local stacking density is not uniform. As the rising dust phenomenon continues to occur, the dust on the top surface and then the subsurface is stripped in succession. With the increase in depth, the local packing density varies randomly, resulting in a significant influence of the local porosity or compactness on the amount of lofted dust particles.

4 Discussions

The migration rate is defined here as the number of migrating particles per second to further analyze the migration characteristics of the moving pyroxene, olivine, and ilmenite particles. We compare our results with the previous studies on the anorthite transport rates derived from (Gan et al., 2022). **Figure 4** presents the migration rates of anorthite (An), olivine (Ol), pyroxene (Py), and ilmenite (Ilm) particles under electron irradiation with an energy of 350 eV and a current of 500 μA. Note that the height of the measured point is 6 mm in the anorthite migration experiments, while the height is 4 mm in the migration experiments of three other minerals. Based on the height difference of only 2 mm between the two measured points, we assume that there is no dramatic variation in the migration fluxes of the identical mineral particles at observed heights above. A qualitative comparison study is therefore effective. The results show that the transport rates of olivine and pyroxene particles are in the range of 6–21 and 9–30 particles per second, higher than the rates of anorthite particles (0.1–2 particles per second). The rate of ilmenite particles is zero. The above results demonstrate that there is a significant difference in the electrostatic transport characteristics of different minerals in the lunar dust.

Ilmenite is a titanium-iron oxide mineral with weak magnetism and good conductivity, while anorthite, olivine, and pyroxene are insulating silicate minerals. Due to the weak magnetism of ilmenite particles, some low-energy incident electrons are deflected or returned, so that the incident electron current slightly decreases in varying degrees. What's more, due to the good conductivity of ilmenite particles, the charge on the particle surface was transferred, resulting in the lack of charge accumulation. Therefore, ilmenite particles could not be released from the sample surface, which is consistent with our experimental findings that no ilmenite particles were detected. Whereas, the charge accumulated by the insulating silicate mineral particles, such as anorthite, olivine, and pyroxene, is difficult to transfer, which caused the dust movement to occur when the three insulating mineral particles were exposed to electron irradiation.

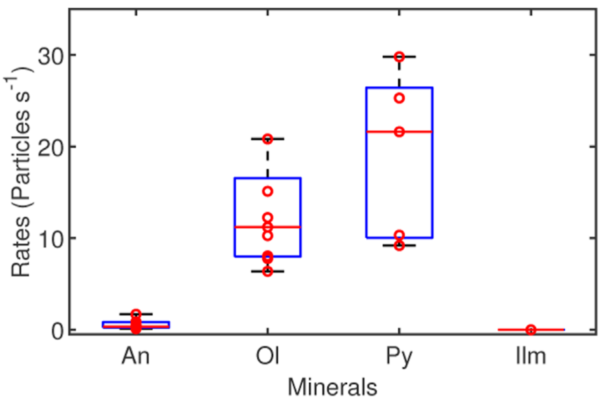


FIGURE 4
Transport rates of anorthite (An), olivine (Ol), pyroxene (Py), and ilmenite (Ilm) particles. Note that the height of the measured point is 4 mm for all experiments, except anorthite migration experiments (6 mm).

It can be concluded that the difference in electrostatic migration characteristics of different minerals is closely related to their electrical properties, such as conductivity or permittivity. Assuming that the lunar dust in different regions is composed of four minerals mixed in different proportions. For example, a small amount of ilmenite particles is dispersed in a large amount of insulating silicate mineral particles, so that the ilmenite particles cannot contact each other. In this case, the ilmenite particles may also migrate. The effect of the different percentages of ilmenite particles on lunar dust migration will be further investigated to understand the dust migration characteristic differences between the highlands and the maria, including the migration characteristic differences between the regions with high-Ti and low-Ti basalts. The results can also be inferred that low-titanium lunar soil may be more prone to migration than high-titanium lunar soil. It is worth noting that the proportion of the four kinds of mineral particles changes as the differential migration of each mineral dust. When the dust migration stops may depend on the percentage of ilmenite particles in the lunar soil.

The migration rates of three insulating silicate minerals are also different, possibly due to the distinct chemical compositions, crystal structures, and physical properties. Anorthite, olivine, and pyroxene are framework silicate, nesosilicate, and chain silicate minerals, respectively. Permittivity refers to a substance's ability to retain an electric charge. However, their permittivities are similar and range from 6.5 to 8. Density and diameter determine the magnitude of gravity that a particle has to overcome to leave the sample surface. In theory, anorthite is easier to migrate because of its lower density. However, in practice, it has the lowest migration rate compared to other silicate minerals. This may be attributed to the stronger cohesive force and higher measuring point. In particular, the effect of the cohesive force on dust migration rate is one of the key contents of subsequent work.

For large dust particles with a diameter of tens of microns, it is mainly the gravity force that needs to be overcome during the process of dust particles being released, while the cohesive force needs to be overcome for micron-sized dust particles (Hartzell and Scheeres, 2011). The cohesive force is mainly derived from the contribution of electrostatic force and van der Waals force related to surface energy. That is, the cohesive force depends on the contact area, effective surface energy, etc. The contact area is related to the distance between particles (namely, packing density), particle size, and particle shape. To exclude the influence of the particle size and shape on dust transport rate, the particle size distributions and shape characteristics of olivine, pyroxene, and ilmenite are further presented, as shown in **Figures 5**, **6**, respectively. Both olivine and pyroxene particles were directly ground in the mortar, and have similar particle size distribution characteristics (below 40 μm). Due to the higher hardness of ilmenite particles, the wet grinding method with alcohol was used. This method is more effective, and the particle size of ilmenite particles is concentrated below 20 μm. In addition, the grain shapes of olivine, pyroxene, and ilmenite minerals are angular. This means that the particle size and shape are not the main reasons for the difference in the transport rates of olivine and pyroxene particles.

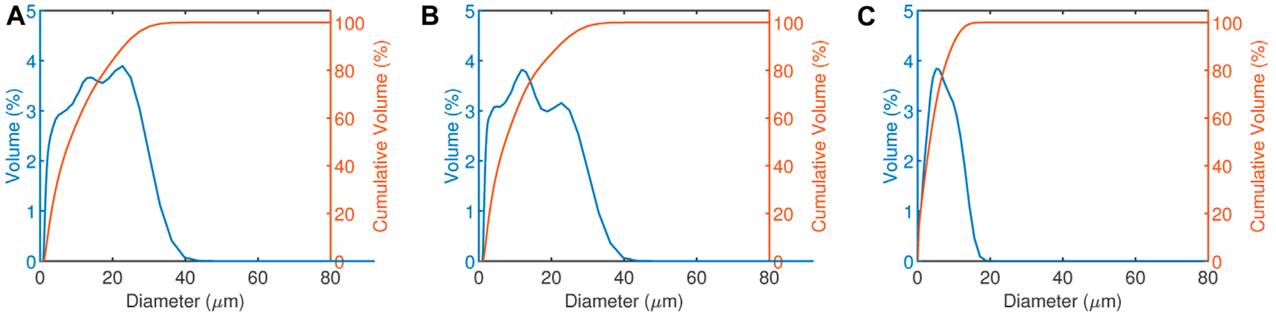


FIGURE 5
The blue lines represent the particle size distributions of olivine **(A)**, pyroxene **(B)**, and ilmenite **(C)** particles, and the red lines represent the accumulative volume of three mineral particles (unit: vol%).

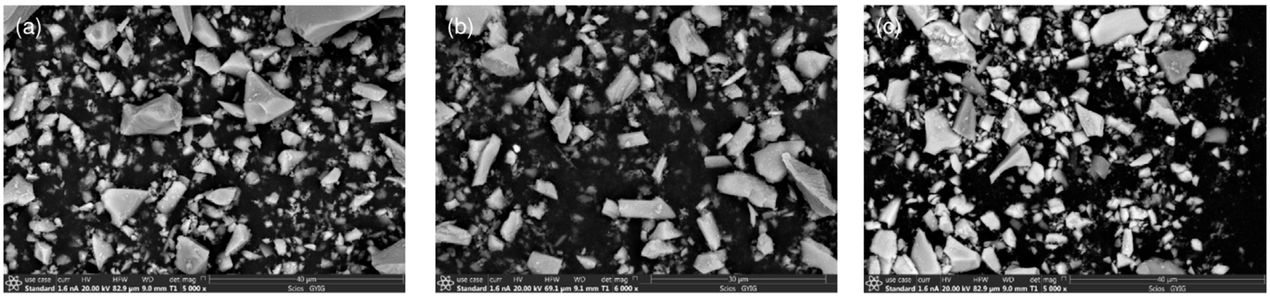


FIGURE 6
Shape characteristics of olivine **(A)**, pyroxene **(B)**, and ilmenite **(C)** particles.

Undoubtedly, the packing density is a major influencing factor. Figure 3C has demonstrated that the influence of the porosity or compactness on dust migration is prominent. However, it is speculated that the difference in the transport rate ranges between olivine and pyroxene may be caused by the difference in the surface energy, since the surface energy depends on the composition of dust particles. Moreover, in the hard vacuum conditions of the lunar environment, the effective surface energy is up to two orders of magnitude greater than that in humid air at one atmosphere pressure (Walton, 2007). Especially for fine lunar dust, the surface energy per unit mass available for cohesive forces to act in the bulk material increases as the surface area per unit mass increases, which is also the reason for the strong adhesion of fine lunar dust.

It should be emphasized that the composition of the lunar dust is much more complex than that of the experimental sample. Lunar dust is the fine fraction of lunar soil. Due to the long-term interaction of lunar soil with the space environment, lunar soil is composed of various types of particles, including rock fragments, mono-mineralic fragments, various kinds of glasses, and agglutinates. The lithic fragments always feather ultrathin amorphous coatings and nanophase iron residing in them. Agglutinates are aggregates of crystalline grains and lithic fragments bonded together by glass. The changes in material composition and crystal structure caused by space weathering can alter the effective surface energy of dust particles. This means the electrostatic migration and adhesion of dust on the lunar surface are more complicated.

5 Conclusion

To analyze the influence of lunar dust composition on the characteristics of dust electrostatic migration, we focus on the simulation of the electrostatic migration processes of pyroxene, olivine, and ilmenite. The results demonstrate that pyroxene and olivine particles can migrate, while ilmenite particles can not. The moving pyroxene and olivine particles are statistically smaller than 10 μm in size, and their verticle velocities are typically less than 2 m s^{-1}, which are consistent with the migration characteristics of anorthite particles in the previous studies. The verticle velocities are also constrained by the particle size.

The particle counts at the measured point vary with the energy and a current of incident electrons, as well as the packing density of dust particles. The energy threshold of massive dust migration is between 300 and 350 eV. The number of migrating dust particles is fewer than 400 within 5 min when the energy of the incident electron is less than 300 eV and the current is 500 μA. However, when the energy of the incident electron is higher than 350 eV, the number is higher than 2,500. The particle counts increase with the increase in the electron current. The dust particles accumulate loosely, and the local stacking density in the loose dust layer is unpredictable, so that the counts of the moving particles do not follow the law of monotonous change. It can be seen from the parallel experiments of pyroxene migration process simulations under electron irradiation with an energy of 350 eV and a current of 500 μA, the migration rates are between 9 and 30 particles per second due to the different stacking compactness. On the lunar surface, lunar regolith continues to churn under meteorite/micrometeorite bombardment, forming an extremely loose lunar soil layer, which will considerably facilitate lunar dust migration.

Meanwhile, to confirm that the material composition affects dust migration characteristics, we compared the migration rates of anorthite, olivine, pyroxene, and ilmenite particles at the measured point. As mentioned above, insulating silicate minerals like anorthite, olivine, and pyroxene are more likely to migrate than ilmenite particles with stronger conductivity. Secondly, the cohesive force may have a sizable impact on dust migration. Pyroxene and olivine particles, for example, have comparable densities, particle sizes, shapes, and dielectric constants. The difference in migration characteristics may be mostly caused by the cohesive force associated with the surface energy, packing density, and other factors. Further analysis of the influence of packing density of lunar dust on its electrostatic migration rate can offer scientific guidance for dust mitigation in the future lunar surface exploration and lunar base construction, such as whether the lunar dust can be by merely tamping the lunar soil rather than consolidating it through microwave sintering.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

HG was involved in the experimental design, experimental testing, data analysis, and manuscript writing and revision. GW supervised the experimental design and the writing of the manuscript and participated in the revision of the manuscript. XZ, GX, and JS contributed to the manuscript's writing and revision. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the National Key Research and Development Program of China (2022YFF0711400), the National Natural Science Foundation of China (42241154 and 41903058), the Frontier Science Research Program of Deep Space Exploration Laboratory (2022-QYKYJH-HXYF-023), and the Science and Technology Program of Guizhou Province (QKHJC-ZK[2023]-131 and QKHJC-ZK[2023]-476).

Acknowledgments

The authors thank the Center for Lunar and Planetary Sciences, Institute of Geochemistry, Chinese Academy of Sciences for providing testing services.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fspas.2023.1213294/full#supplementary-material

References

Abbas, M., Tankosic, D., Craven, P., LeClair, A., and Spann, J. (2010). Lunar dust grain charging by electron impact: Complex role of secondary electron emissions in space environments. *Astrophysical J.* 718, 795–809. doi:10.1088/0004-637x/718/2/795

Carroll, A., Hood, N., Mike, R., Wang, X., Hsu, H.-W., and Horányi, M. (2020). Laboratory measurements of initial launch velocities of electrostatically lofted dust on airless planetary bodies. *Icarus* 352, 113972. doi:10.1016/j.icarus.2020.113972

Colwell, J., Batiste, S., Horányi, M., Robertson, S., and Sture, S. (2007). Lunar surface: Dust dynamics and regolith mechanics. *Rev. Geophys.* 45. doi:10.1029/2005rg000184

Criswell, D. R. (1973). "Horizon-glow and the motion of lunar dust," in *Photon and particle interactions with surfaces in space: Proceedings of the 6th eslab symposium, held at noordwijk, The Netherlands, 26–29 september, 1972* (Springer), 545–556.

Gaier, J. R., and Jaworske, D. A. (2007). "Lunar dust on heat rejection system surfaces: Problems and prospects," in *AIP conference proceedings* (American Institute of Physics), 880, 27–34.

Gan, H., Zhang, X., Li, X., Jin, H., Xie, L., and Zou, Y. (2022). Experiments on the electrostatic transport of charged anorthite particles under electron beam irradiation. *Astrophysical J.* 930, 42. doi:10.3847/1538-4357/ac5d48

Halekas, J., Delory, G., Lin, R., Stubbs, T., and Farrell, W. (2009a). Lunar prospector measurements of secondary electron emission from lunar regolith. *Planet. Space Sci.* 57, 78–82. doi:10.1016/j.pss.2008.11.009

Halekas, J., Delory, G., Lin, R., Stubbs, T., and Farrell, W. (2009b). Lunar surface charging during solar energetic particle events: Measurement and prediction. *J. Geophys. Res. Space Phys.* 114. doi:10.1029/2009ja014113

Hartzell, C. M., and Scheeres, D. J. (2011). The role of cohesive forces in particle launching on the moon and asteroids. *Planet. Space Sci.* 59, 1758–1768. doi:10.1016/j.pss.2011.04.017

Hood, N., Carroll, A., Mike, R., Wang, X., Schwan, J., Hsu, H.-W., et al. (2018). Laboratory investigation of rate of electrostatic dust lofting over time on airless planetary bodies. *Geophys. Res. Lett.* 45, 13–206. doi:10.1029/2018gl080527

Kawamoto, H., Uchiyama, M., Cooper, B., and McKay, D. (2011). Mitigation of lunar dust on solar panels and optical elements utilizing electrostatic traveling-wave. *J. Electrost.* 69, 370–379. doi:10.1016/j.elstat.2011.04.016

Li, D., Wang, Y., Zhang, H., Zhuang, J., Wang, X., Wang, Y., et al. (2019). *In situ* measurements of lunar dust at the chang'e-3 landing site in the northern mare imbrium. *J. Geophys. Res. Planets* 124, 2168–2177. doi:10.1029/2019je006054

McCoy, J. E. (1976). Photometric studies of light scattering above the lunar terminator from apollo solar corona photography. *Lunar Planet. Sci. Conf. Proc.* 7, 1087–1112.

Němeček, Z., Pavlů, J., Šafránková, J., Beránek, M., Richterová, I., Vaverka, J., et al. (2011). Lunar dust grain charging by electron impact: Dependence of the surface potential on the grain size. *Astrophysical J.* 738, 14. doi:10.1088/0004-637x/738/1/14

Örger, N. C., Toyoda, K., Masui, H., and Cho, M. (2021). Experimental investigation on particle size and launch angle distribution of lofted dust particles by electrostatic forces. *Adv. Space Res.* 68, 1568–1581. doi:10.1016/j.asr.2021.03.037

Örger, N. C., Toyoda, K., Masui, H., and Cho, M. (2019). Experimental investigation on silica dust lofting due to charging within micro-cavities and surface electric field in the vacuum chamber. *Adv. Space Res.* 63, 3270–3288. doi:10.1016/j.asr.2019.01.045

Rennilson, J., and Criswell, D. R. (1974). Surveyor observations of lunar horizon-glow. *moon* 10, 121–142. doi:10.1007/bf00655715

Schwan, J., Wang, X., Hsu, H.-W., Grün, E., and Horányi, M. (2017). The charge state of electrostatically transported dust on regolith surfaces. *Geophys. Res. Lett.* 44, 3059–3065. doi:10.1002/2017gl072909

Walton, O. R. (2007). *Adhesion of lunar dust. NASA contractor report NASA/CR-2007-214685.*

Wang, X., Horányi, M., and Robertson, S. (2011). Dust transport near electron beam impact and shadow boundaries. *Planet. Space Sci.* 59, 1791–1794. doi:10.1016/j.pss.2010.12.005

Wang, X., Horányi, M., and Robertson, S. (2010). Investigation of dust transport on the lunar surface in a laboratory plasma with an electron beam. *J. Geophys. Res. Space Phys.* 115. doi:10.1029/2010ja015465

Wang, X., Schwan, J., Hsu, H.-W., Grün, E., and Horányi, M. (2016). Dust charging and transport on airless planetary bodies. *Geophys. Res. Lett.* 43, 6103–6110. doi:10.1002/2016gl069491

Xie, L., Zhang, X., Li, L., Zhou, B., Zhang, Y., Yan, Q., et al. (2020). Lunar dust fountain observed near twilight craters. *Geophys. Res. Lett.* 47, e2020GL089593. doi:10.1029/2020gl089593

Yan, Q., Zhang, X., Xie, L., Guo, D., Li, Y., Xu, Y., et al. (2019). Weak dust activity near a geologically young surface revealed by chang'e-3 mission. *Geophys. Res. Lett.* 46, 9405–9413. doi:10.1029/2019gl083611

Zakharov, A., Zelenyi, L., and Popel', S. (2020). Lunar dust: Properties and potential hazards. *Sol. Syst. Res.* 54, 455–476. doi:10.1134/s0038094620060076

Zhang, H., Wang, Y., Chen, L., Zhang, H., Li, C., Zhuang, J., et al. (2020). *In-situ* lunar dust deposition amount induced by lander landing in chang'e-3 mission. *Sci. China Technol. Sci.* 63, 520–527. doi:10.1007/s11431-019-1434-y

Zook, H. A., and McCoy, J. E. (1991). Large scale lunar horizon glow and a high altitude lunar dust exosphere. *Geophys. Res. Lett.* 18, 2117–2120. doi:10.1029/91gl02235

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Original Research
PUBLISHED 14 June 2023
DOI 10.3389/fspas.2023.1217990

Analysis of reliability and accuracy of lunar core detection based on Apollo moonquake observation

Biao Yang and Yanbin Wang*

Department of Geophysics, School of Earth and Space Sciences, Peking University, Beijing, China

OPEN ACCESS

EDITED BY
Jinhai Zhang,
Chinese Academy of Sciences (CAS),
China

REVIEWED BY
Xiang Zhang,
UMR7154 Institut de Physique du Globe de Paris (IPGP), France
Wei Zhang,
Southern University of Science and Technology, China

*CORRESPONDENCE
Yanbin Wang,
✉ ybwang@pku.edu.cn

RECEIVED 06 May 2023
ACCEPTED 30 May 2023
PUBLISHED 14 June 2023

CITATION
Yang B and Wang Y (2023), Analysis of reliability and accuracy of lunar core detection based on Apollo moonquake observation.
Front. Astron. Space Sci. 10:1217990.
doi: 10.3389/fspas.2023.1217990

Introduction: In the 1960s, the Apollo missions deployed seismometers on the near side of the Moon, and the obtained moonquake data enabled the seismic study of internal structure of the Moon. The lunar seismic waveforms are dominated by strong coda waves caused by scattering in the near surface, which masks the reflected and converted waves from the discontinuities inside the Moon. The Double Array Stacking method is an important technique to enhance the reflected and converted waves. However, its performance would be affected by many factors including signal to noise ratio, arrival time errors, differences in amplitude and polarity of waveforms caused by different station azimuth, and source mechanisms, thus needs to be further analyzed.

Method: In this study, we use the pseudo-spectral and finite-difference hybrid method on staggered grid to simulate wave propagation of deep moonquakes, considering near surface scattering, to analyze the influence of the above factors on the stacking results of reflected and converted phases from outer core of the Moon.

Results and discussion: Our results indicate that coda and noises would greatly reduce the stacking energy but have no significant impact on the peak position of the stacked waveforms. The arrival time errors have a significant influence on the vertical component of ScP phase and the radial component of the ScS phase, but have little influence on the transverse component of ScS phase. The difference in amplitude also leads to differences in stacking energy, but the stacking image shows good consistency in the shape. When the polarities of seismic phases are relatively consistent, the peak of the stacking energy comes from reflected phases from the core mantle boundary of the Moon. This research may be helpful to improve the detection accuracy of the internal structure of the Moon and provide a solid reference for the deployment of future seismic stations, which is of great significance for understanding the formation and evolution process of the Moon.

KEYWORDS

lunar core, double array stacking, scattering, polarization filtering, numerical simulation

1 Introduction

The Apollo missions, launched in the 1960s and 1970s, successfully deployed the first array of seismometers on extraterrestrial body. Many researches have been conducted to detect the interior structure of the Moon based on Apollo moonquake waveforms. However, the late phases are almost completely buried among long-last strong coda waves caused by endless scattering in the shallow subsurface of the Moon (Frohlich and Nakamura, 2009). The Moon has experienced strong volcanic activity and impact modification without hydraulic cementation, resulting in the accumulation of broken basalt layers and impact splashes on the shallow surface of the Moon (Blanchette Guertin et al., 2012; Zhang et al., 2022), leading to strong lunar seismic wave scattering and long duration of lunar seismic coda waves far beyond that of earthquakes, even reaching several hours (Garcia et al., 2019). The strong coda submerges the reflected and converted waves from the discontinuities inside the Moon, causing great difficulties in picking up the arrival times of lunar seismic phases. The large error in the arrival times results in insufficient constraints on the interior structures, including significant uncertainty of the lunar crust and core detection (Lognonné et al., 2003; Garcia et al., 2019). By modeling the propagation of lunar seismic waves inside the Moon model with lateral heterogeneous upper crust, several researches had reproduced the main characteristics of coda waves and estimated the velocity perturbation in the lunar regolith (Jiang et al., 2015a; Onodera et al., 2021; Zhang et al., 2022).

It is important to detect the core of the Moon (Garcia et al., 2019) thus a series of methods had been developed, such as arrival time inversion (Nakamura et al., 1974), electromagnetic sounding inversion (Hood et al., 1999), laser ranging inversion (Williams et al., 2001; Williams et al., 2014), joint inversion of Love number, average mass and moment of inertia (Khan et al., 2004; Khan and Mosegaard, 2005; Khan et al., 2006; Krorod and Kuskov, 2011; Khan et al., 2014; Matsumoto et al., 2015) and molecular dynamics simulation of Fe-S alloy composition inside the lunar core (Kuskov and Belashchenko, 2016; Morard et al., 2018). By seismic inversion based on arrival times of P and S waves, Garcia et al. (2019) proposed that the average density of the outer core is between 4–5 g/cm3, indicating a core composed of iron and a large number of light elements (with a radius of approximately 350 km). Meanwhile, the poor fitting of arrival times of S waves indicates that there is a very low S-wave velocity at the bottom of the mantle, which may indicate a partially melted layer in the lower mantle.

The Double Array Stacking method (Weber et al., 2011) has unique advantages in searching for reflected and converted seismic phases from lunar core. Before stacking, waveforms of deep moonquake clusters are filtered by polarization filter to enhance the signal of effective seismic phases. Then the waveforms after polarization filtering are enveloped to prevent potential reverse-polarity arrivals from negating each other in the stack. Weber et al. (2011) estimated a radius of inner core as 240 ± 10 km and a radius of outer core radius as 330 ± 20 km using Double Array Stacking of PcP, ScS, ScP, and PcS phases. The radius of outer core of the Moon in the VPREMOON model (Garcia et al., 2011) is also constrained by the reflected and converted waves from the core-mantle boundary. Before stacking, they analyzed the effect of enhancing amplitude by polarization filtering used in Weber et al. (2011) and amplitudes of different seismic phases. They believe that the amplitudes of PcP, ScP, and PcS phases are much weaker than those of ScSH phase. Under the interference of noises and coda, polarization filtering does not enhance the amplitude of those phases. Finally, Garcia et al. (2011) only use ScSH phase for Double Array Stacking and obtained a radius of outer core as 380 ± 40 km. Yuan (2018) developed Velocity Scanning Stacking based on previous studies (Lin et al., 2011; Weber et al., 2011), which can scan and analyze the velocities of P and S waves at different depths while stacking seismic phases, constraining the depth of the discontinuity and velocity structure of the Moon at the same time. The depths of the discontinuity are determined by multiplying the stacking results of four seismic phases under the optimal velocity structure, giving an inner core radius of 230 ± 20 km and an outer core radius of 310 ± 10 km. Besides, they believed that there is partial melting at a depth of 1300 km.

In the above research (Garcia et al., 2011; Weber et al., 2011; Yuan. 2018), there are differences in the data selection, data processing and seismic phase selection. In data selection, Weber et al. (2011) and Yuan (2018) choose the stacking waveforms of deep moonquake clusters while Garcia et al. (2011) choose the original waveforms of moonquakes. In data processing, Weber et al. (2011) and Yuan (2018) use polarization filtering to enhance the amplitude of seismic phases and envelope to prevent potential influence of opposite-polarity arrivals while Garcia et al. (2011) question the effectiveness of polarization filtering, only performing broadband filtering. In seismic phase selection, Weber et al. (2011) and Yuan (2018) choose ScS, ScP, PcS, and PcP phases for stacking while Garcia et al. (2011) believe that the stacking results of weak seismic phases are unreliable, so only choose ScSH phase for stacking.

The radius of outer core of the Moon obtained by Weber et al. (2011) and Garcia (2011) differs greatly, but it isn't yet clear which one is more reliable. Therefore, in-depth analysis and comparative research are needed to confirm the effectiveness of polarization filtering and envelope and analyze the performance of different seismic phases, to improve the reliability of the detection results. Besides data processing and seismic phase selection, several other factors may affect the stacking results in Apollo data and need to be considered. Significant errors in picking first arrivals, masked by noises and coda, result in significant uncertainty in source location. Although the Apollo observation station formed a huge triangular network, the limited instrument sensitivity, dispersive moonquakes and unconfirmed focal mechanism, make it difficult to study the interior structure of the Moon (Zhang et al., 2021). Therefore, evaluating the reliability and accuracy of Double Array Stacking through numerical simulation tests is of great significance for reducing the uncertainty of current models and conducting future moonquake data analysis.

In this study, we use the pseudo-spectral and finite-difference hybrid method based on staggered grid (Wang et al., 2011; Jiang et al., 2015b) and Moon model (Wang et al., 2013; Jiang et al., 2015a) with velocity perturbation in upper crust (Jiang et al., 2015a) to simulate the waveforms of deep moonquakes with scattering. We first analyze the influence of polarization filtering and envelope on the stacking results of waveforms with strong noises and coda by stacking the simulated waveforms after polarization filtering and envelope processing. Then, we analyze the influence of coda and noises, arrival time errors, and differences in amplitude and polarity of waveforms caused by different station azimuth and source mechanisms on the results of stacking. Finally, we explore the reliability of results of stacking using reflected and converted phases to detect lunar core under these factors and discuss how to reduce the influence and improve the reliability and accuracy of stacking.

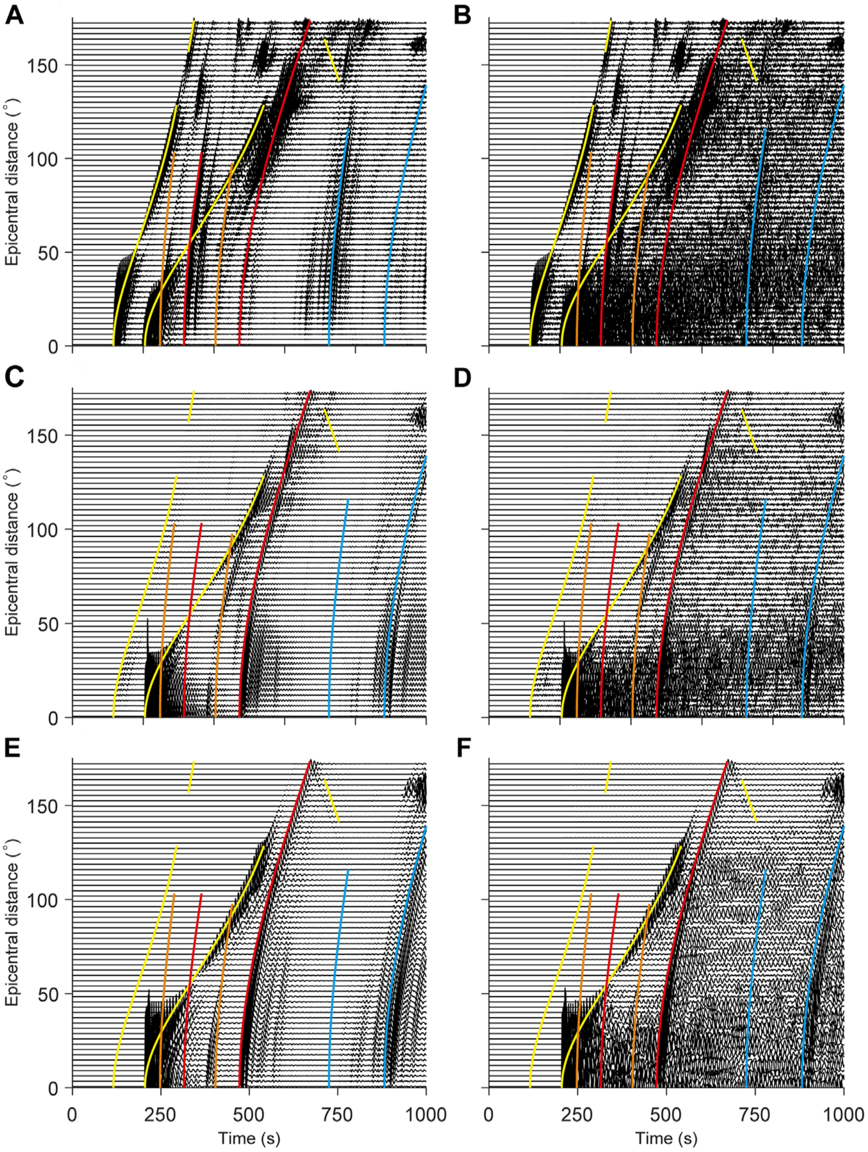


FIGURE 1
Synthetic waveforms of moonquakes with source time of 5 s **(A, C, E)** are the V, R, and T components of the moonquake waveforms without scattering. **(B, D, F)** are the V, R, and T components of the scattered moonquake waveforms. The yellow lines represent direct P and S waves, the orange lines represent PcP and PcS waves, the red lines represent ScP and ScS waves, and the blue lines represent sScP and sScS waves, respectively.

2 Numerical modeling and processing of lunar seismic waveforms

2.1 Numerical modeling of scattered waves of moonquakes

We use the pseudo-spectral and finite-difference hybrid method (Wang et al., 2011; Jiang et al., 2015b) to model seismic wave propagation of deep moonquakes with strong scattering effects in the upper crust. The model is a two-dimensional great circle cross-section of the Moon passing through the source and station defined in a two-dimensional cylindrical coordinate system, extending from the surface to a depth of 1650 km radially, and from 0° to 180° laterally. The source of moonquake is located close to the edge of the model (r = 867 km, θ = 5°), and the core-mantle boundary is included in the model allowing the reflection and conversion of seismic waves from the core mantle boundary can be calculated and recorded by the station. We add velocity perturbation in the upper crust of the Moon (Jiang et al., 2015a) to calculate the strong scattering of lunar seismic waves.

The peak frequency of the Apollo moonquake data is 0.45 Hz and the design frequency band of seismometer for the Chang'e-7 mission is 120 s to 100 Hz (Zhang et al., 2021). Under this circumstance, it is of great significance to explore the variation of amplitudes of reflected and converted seismic phases from lunar core-mantle boundary at various frequencies for the selection of frequency bands in future research of lunar core detection.

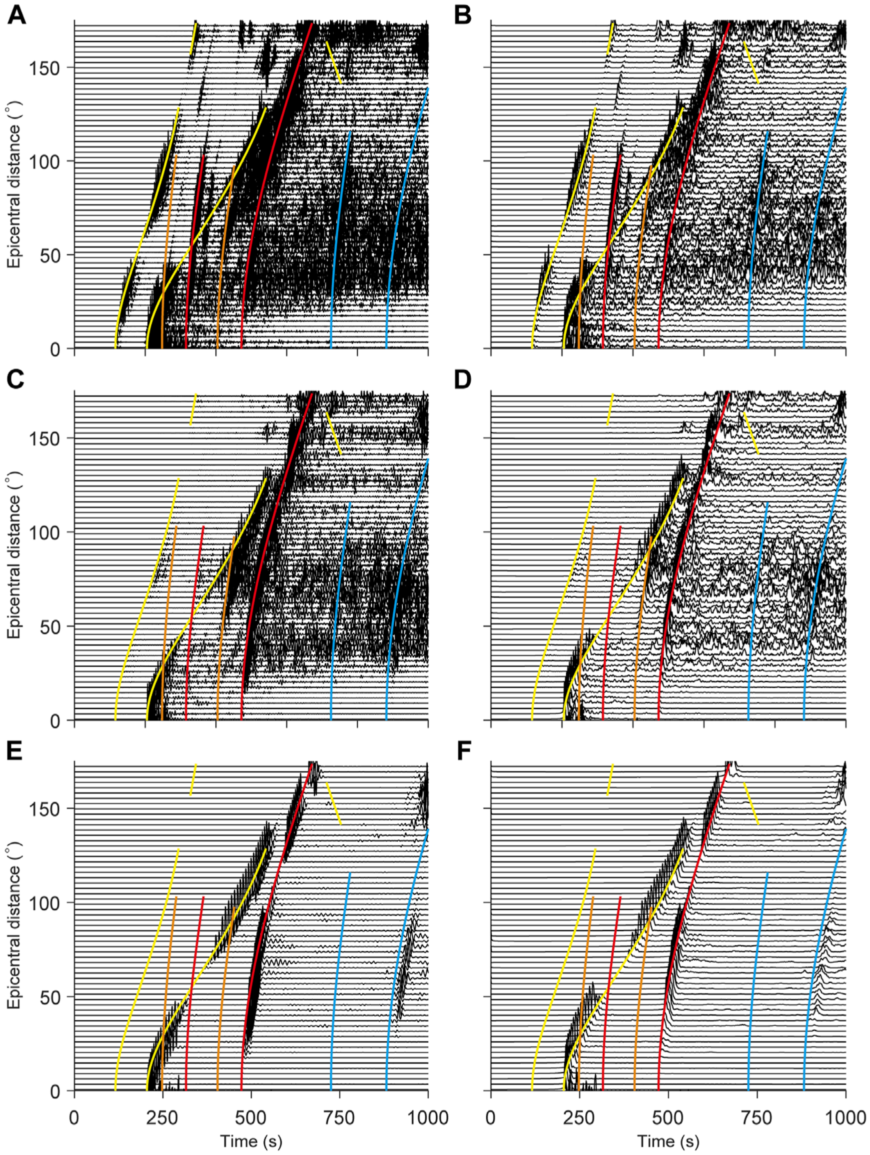


FIGURE 2
Scattering waveforms after polorization filtering **(A, C, E)** and its envelopes **(B, D, F)**. The colored lines show travel time curves of seismic phases as indicated in **Figure 1**.

Therefore, we adopted 2 kinds of Herrmann pseudo δ function (Herrmann, 1979) as the source time function with the width of 3.0 s and 5.0 s, respectively.

The model is discretized into 1650 grid points in the r direction, 4096 and 2048 grid points in the θ direction for source time function of 3.0 s and 5.0 s, respectively. The maximum grid spacing on the surface in θ direction is 1.33 km and 2.66 km for the two discretized models, and the grid spacing in r direction is fixed to 1.0 km. We adopt the velocity structure model used by Jiang et al. (2015a) with a minimum velocity of 1.8 km/s for S-waves in the crust, corresponding to the minimum wavelengths of 5.4 km and 9.0 km for source time function of 3.0 s and 5.0 s, respectively. The time interval is determined as 0.015 s from the stability condition defined by the minimum grid interval and the maximum P-wave velocity in the whole model. We calculate for 80,000 time steps, corresponding to a propagation time of 1200 s of seismic wave in Moon model. The moment tensor components of the source are $M_{rr} = 0$, $M_{\theta\theta} = 0$, $M_{r\theta} = M_{\theta r} = 1$ for the focal mechanism of deep moonquake, corresponding to a horizontal sliding fault (Wang et al., 2013).

Since our modeling are performed in a 2-D cross-section of the Moon, the geometric spreading and source radiation of synthetic waveforms slightly differ from waveforms of deep moonquake observation. Our synthetic waveforms are corrected to compensate for the differences between 2-D and 3-D geometric spreading and 2-D line and 3-D point source radiation (Helmerger and Vidalee, 1988; Wang et al., 2001; Wang et al., 2013):

$$u_{3D}(t) = \frac{1}{\sqrt{R}}\frac{1}{\pi}\frac{1}{\sqrt{t}} * \frac{d}{dt}u_{2D}(t) \tag{1}$$

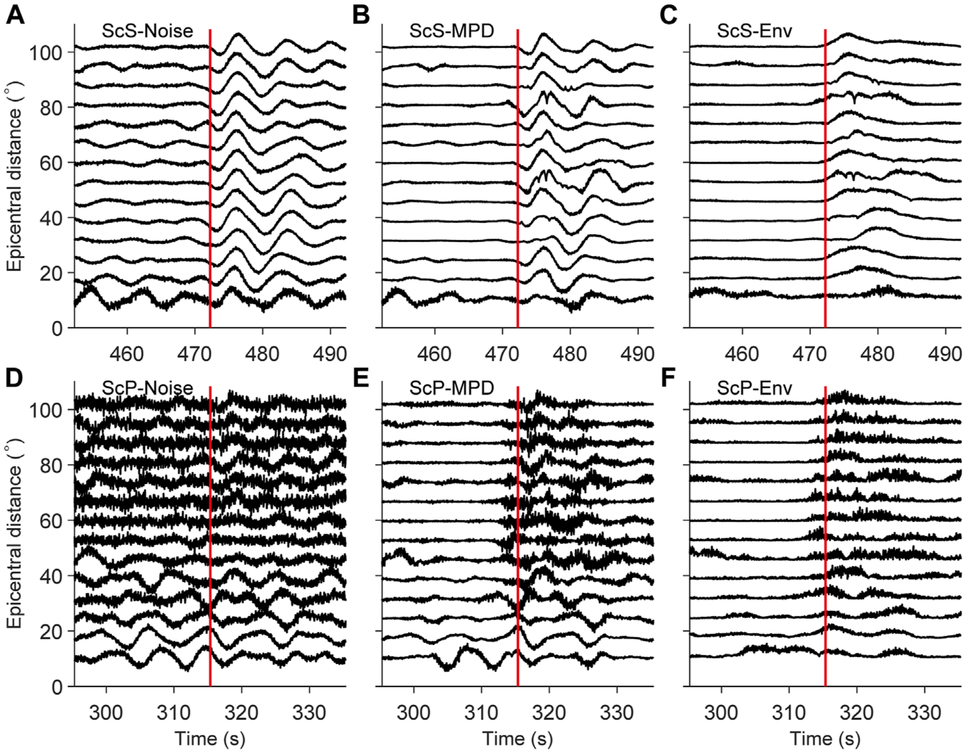


FIGURE 3
R component of scattering waveforms adding radom noises **(A, D)**, and waveforms after polorized filtering **(B, E)** and their envelopes **(C, F)**. **(A–C)** are results for ScS, and **(D–F)** are those for ScP. The red lines represent in-phase axes after alignment.

where $\frac{1}{\sqrt{R}}$ is for the difference in geometrical spreading between 2-D and 3-D wave propagation and $\frac{1}{\pi}\frac{1}{\sqrt{t}}*\frac{d}{dt}$ is for the difference in pulse shape between the "line" and the "point" source solution, $u_{2D}(t)$ is the waveform obtained by our 2-D modeling and $u_{3D}(t)$ is the converted waveform after correction.

Intrinsic attenuation of Moon medium causes waveform decaying during propagation. The anelastic attenuation is quantified by the attenuation coefficient (Graves et al., 1996), which can be expressed as:

$$A(r,\theta) = e^{\frac{-\pi f_0 \Delta t}{Q(r,\theta)}} \tag{2}$$

where f_0 and Q (r,θ) are reference frequency and quality factor, respectively. At each time step, we calculate values of stress and velocity and attenuation coefficient on each node and multiply them to consider the attenuation. Because of its high efficiency and less demand of computer memory, this numeric method has been widely applied for seismic wave propagation of earthquakes and moonquakes (Wang et al., 2011; Wang et al., 2013; Jiang et al., 2015a; Jiang et al., 2015b). We apply free surface boundary conditions at the lunar surface and absorbing boundary conditions are used for other boundaries to reduce artificial reflections caused by artificial boundaries (Cerjan et al., 1985).

2.2 Polarization filtering

The seismic wavefield is generated by the interference and superposition of different types of vibrations with different polarization characteristics. Considering strong scattering of lunar seismic wave, it is necessary to use the nonlinear filtering technique to suppress coda and noises to obtain waveforms with clearer seismic phases. White (1964) proposed a time-averaged product of vertical and radial components as a filter and referred to it as the Motion Product Detector (MPD):

$$M_j = \sum_{i=-n}^{n} Z_{j+i} R_{j+i} \tag{3}$$

where j is the time step, n determines the length of the average window, and the output (OZ, OR) after filtering is the product of M and Z, M, and R, respectively:

$$OZ_j = Z_j M_j, OR_j = R_j M_j \tag{4}$$

White (1964) theoretically analyzed the effects of MPD and found it can effectively enhance the amplitudes of body waves and suppress surface waves, scattered waves and noises. Shimsoni and Smith (1964) applied MPD to actual data and verified its effectiveness on seismic phase enhancement. For the first time, Jarosch (1977) used MPD in processing seismic data of the Apollo artificial impacts, picked up the arrival time of body wave and finally obtained the shallow structure of the subsurface of the Moon.

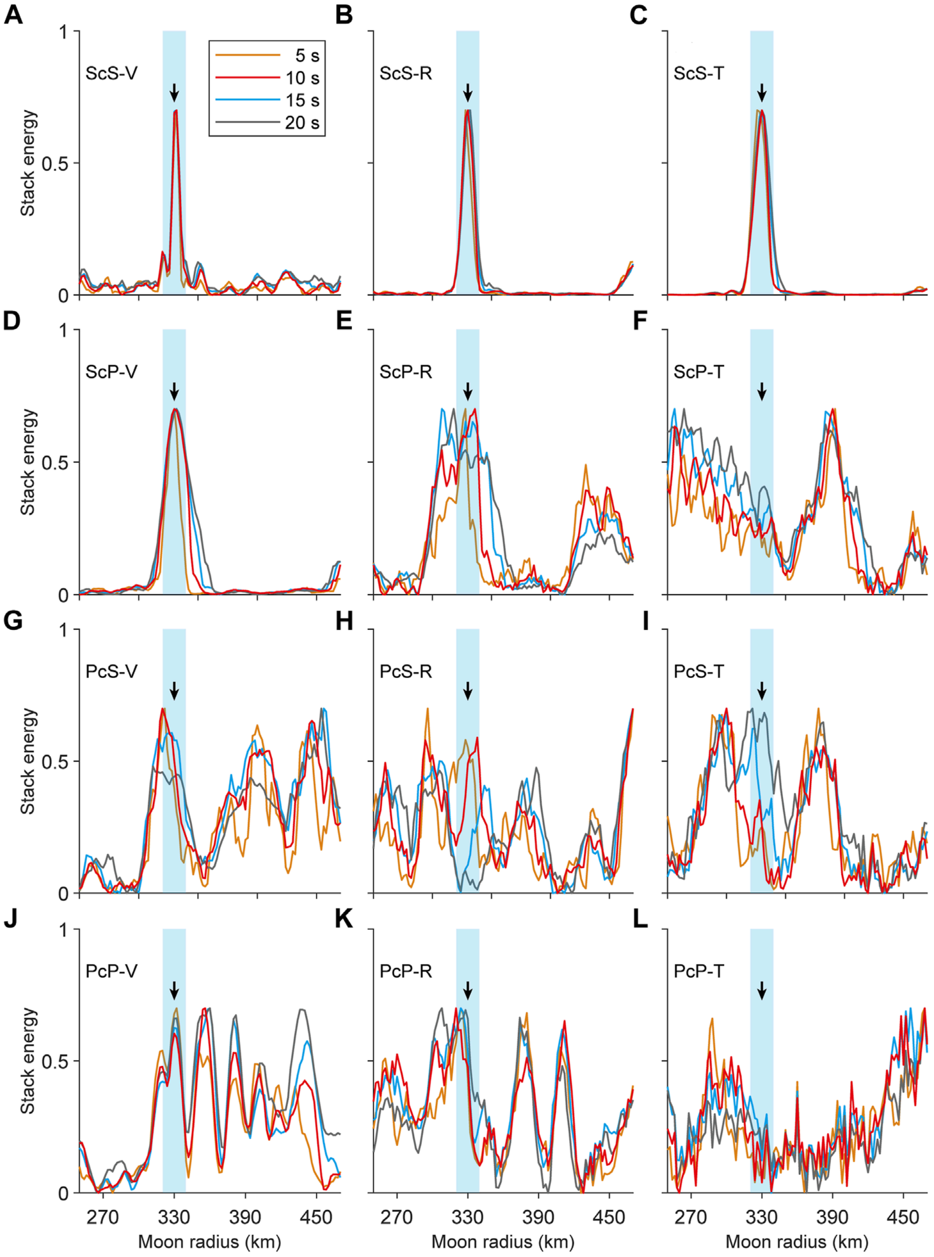


FIGURE 4
Array stacking results of PcP **(J–L)**, PcS **(G–I)**, ScP **(D–F)** and ScS **(A–C)** phases. The black arrows represent the outer core radius (330 km) in the model, and the light blue strips represent an error range of 10 km. The four colored curves represent the stacking results of different time windows.

Weber et al. (2011) used MPD to filter moonquake waveforms, which were used then for Double Array Stacking to obtain the radius of lunar core. For horizontally polarized shear waves (SH), the energy is expected to only appear on the transverse component (T). To process waveforms on T component, Weber et al. (2011) proposed the following formula:

$$OT_j = T_j \sum_{i=-n}^{n} T_{j+i} T_{j+i} \tag{5}$$

where j is the time step, n determines the length of the average window, and the output (OT) after filtering is the product of T.

2.3 Double Array Stacking

Since the 1960s, seismic array technology has brought new impetus to seismology. In addition to improving signal-to-noise ratio, seismic array technology can study the fine structure of the Earth's interior (Kárason and van der Hilst, 2001; Rost and Thomas, 2002), including regional fault and mantle (Arlit et al., 1999; Castle and Creager, 1999; Krüger et al., 2001; Ritter et al., 2001), and the heterogeneity of the outer and inner core (Vidale and Earle, 2000;Rost and Revenaugh, 2001; Thomas et al., 2002).

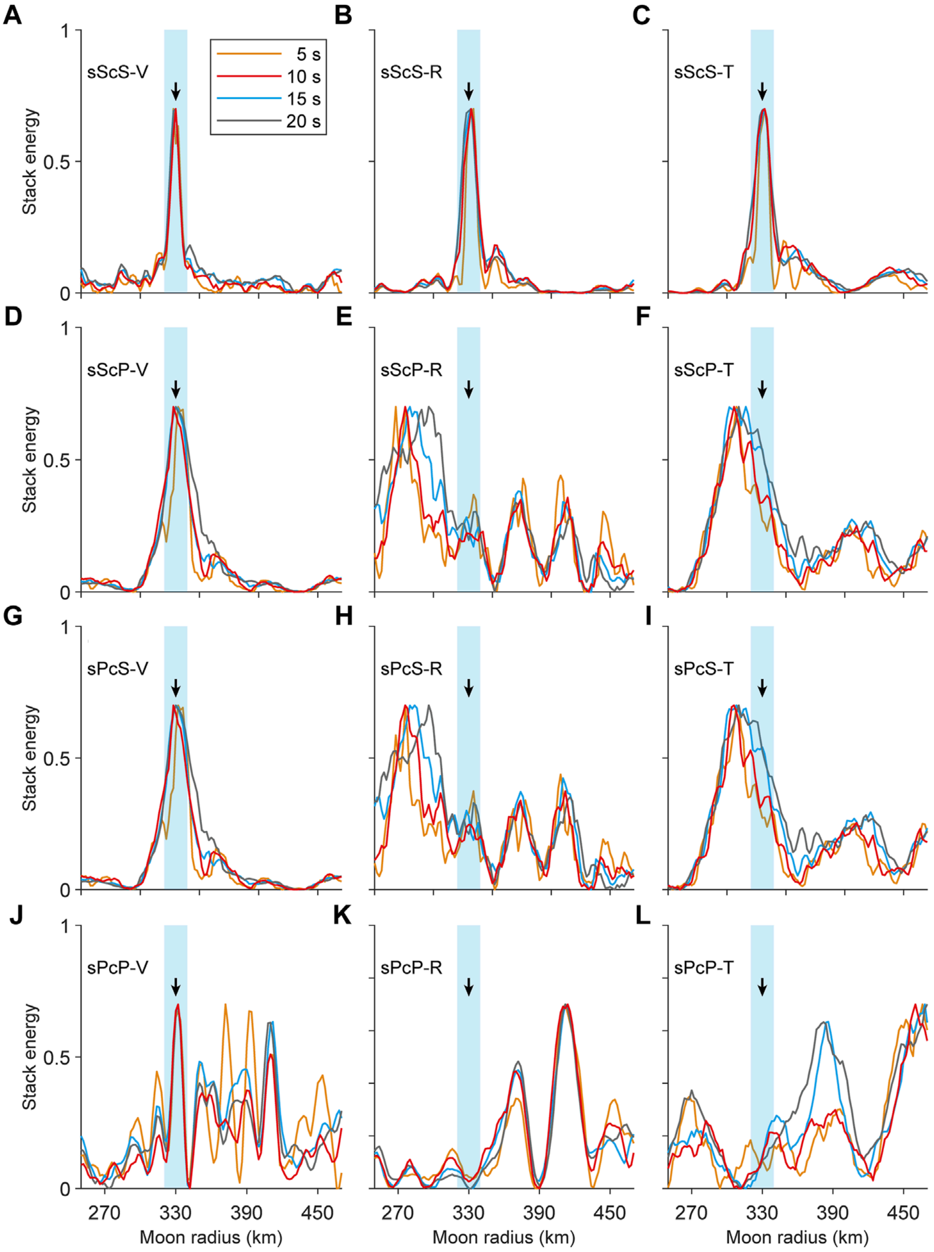


FIGURE 5
Array stacking results of sPcP **(J–L)**, sPcS **(G–I)**, sScP **(D–F)** and sScS **(A–C)** phases. The black arrows represent the outer core radius in the model (330 km), and the light blue strips represent an error range of 10 km. The four colored curves represent the stacking results of different time windows.

Migration is a method originally used in geophysics exploration. In global seismology, migration is used to image mantle heterogeneity in many regions (Lynnes and Lay, 1989; Revenaugh, 1995; Boston and Rondenay, 2002). In previous studies, migration is used for teleseismic arrays to study the discontinuity in the Earth's interior (Neal and Pavlis, 1999; Neal and Pavlis, 2001). The specific procedure is to shift the waveforms to the same zero point according to the theoretical arrival times of reflected seismic phases and stack these waveforms. The depth of discontinuity in the velocity model, which corresponds to the strongest amplitude of the stacked waveforms, is considered as the depth of the discontinuity in the Earth's interior.

Weber et al. (2011) extended this method to Double Array Stacking method and applied it to the processing of observation data from different moonquakes and stations. If the estimated radius in the velocity model deviates from the given core radius of the Moon, the shifted waveforms would be chaotic and the stacked amplitudes would be tiny. Otherwise, waveforms after stacking would maintain the original waveforms with a relatively large amplitude. Various time windows corresponding to different

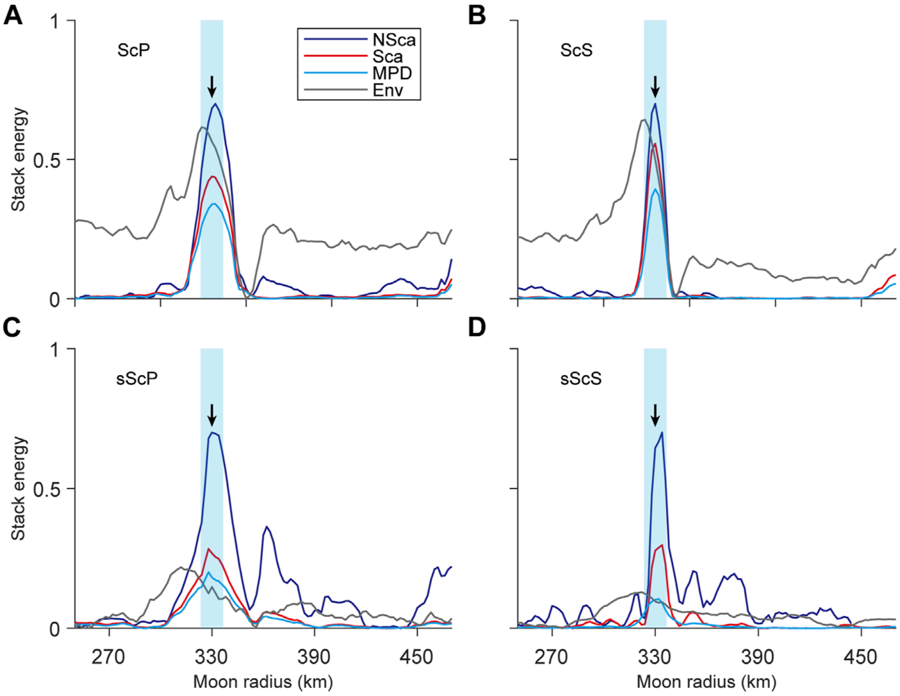


FIGURE 6
Comparison of four stacking results of ScP **(A)**, ScS **(B)**, sScP **(C)** and sScS **(D)** phases after normalization. Purple, red, blue, and gray lines represent the stacking results of waveforms without scattering, with scattering and noises, after polarization filtering and their envelopes, respectively. The black arrows represent the outer core radius in the model (330 km), and the light blue strips represent an error range of 10 km.

reflected and converted phases are selected for stacking, and then the energy of the waveforms within the window are calculated by:

$$E = \frac{1}{N}\sum_{i=1}^{N} Amp(i)^2 dt \qquad (6)$$

where E and N represent the average energy of the waveform within the window and the window length, respectively. $Amp(i)$ represents the amplitude at point i and dt is the time step. Before stacking, waveforms will be normalized. In our study, waveforms in different windows are normalized separately to ensure consistent contribution of each moonquake waveform to the stacking results.

3 Synthetic lunar seismic waveforms

3.1 Synthetic waveforms without and with scattering

We use two Herrmann function with width of 3.0 s and 5.0 s as the source time function and a lateral heterogeneous Moon model to simulate the propagation of seismic wave from deep moonquake. **Figures 1A, C, E** show waveforms on vertical (V), radial (R), and transverse (T) components calculated with model without velocity perturbation in the upper crust. In the case of high value of Q in the Moon, ScP and sScP phases on V component and ScS and sScS phases on R and T components are very obvious. **Figures 1B, D, F** show scattered waveforms on V, R, and T components. Compared with **Figures 1A, C, E** amplitudes of the seismic phases reflected and converted from core mantle boundary are weaker and strong coda appear.

However, the in-phase axes of ScP and sScP phases on V component, as well as ScS and sScS phases on R and T components, can still be seen, indicating that the scattering waves affect more on picking PcP and PcS seismic phases, less on picking ScS, ScP, sScS, and sScP seismic phases. Therefore, it can be predicted that for Double Array Stacking of Apollo moonquake waveforms, these seismic phases (ScS, ScP, sScS, and sScP) are more reliable than that of PcP and PcS.

The amplitude of seismic phases of the synthetic waveforms at high frequency (source time function of 3 s, **Supplementary Figure S1**) is relatively weaker than that at low frequency (source time of 5 s, **Figure 1**). ScP, ScS, sScP, and sScS seismic phases are submerged in coda at high frequency. The most sensitive detection frequency of the Apollo seismometer in peak mode is 0.45 Hz. After processing (Garcia et al., 2011), frequency band of Apollo waveforms is broadened to 0.3–0.9 Hz, higher than that in our modeling. No doubt that seismic phases in Apollo data are masked by the coda and noises. Garcia et al. (2011) found that polarization filtering will enhance noises rather than seismic phases in Apollo data and the energy peak of stacking results of PcP, ScP, and PcS phases comes from scattered waves and noises rather than seismic phases.

Therefore, they argued that the method of combining results of four seismic phases (Weber et al., 2011; Yuan, 2018) to estimate the radius of lunar core is unreliable. Therefore, it is important to

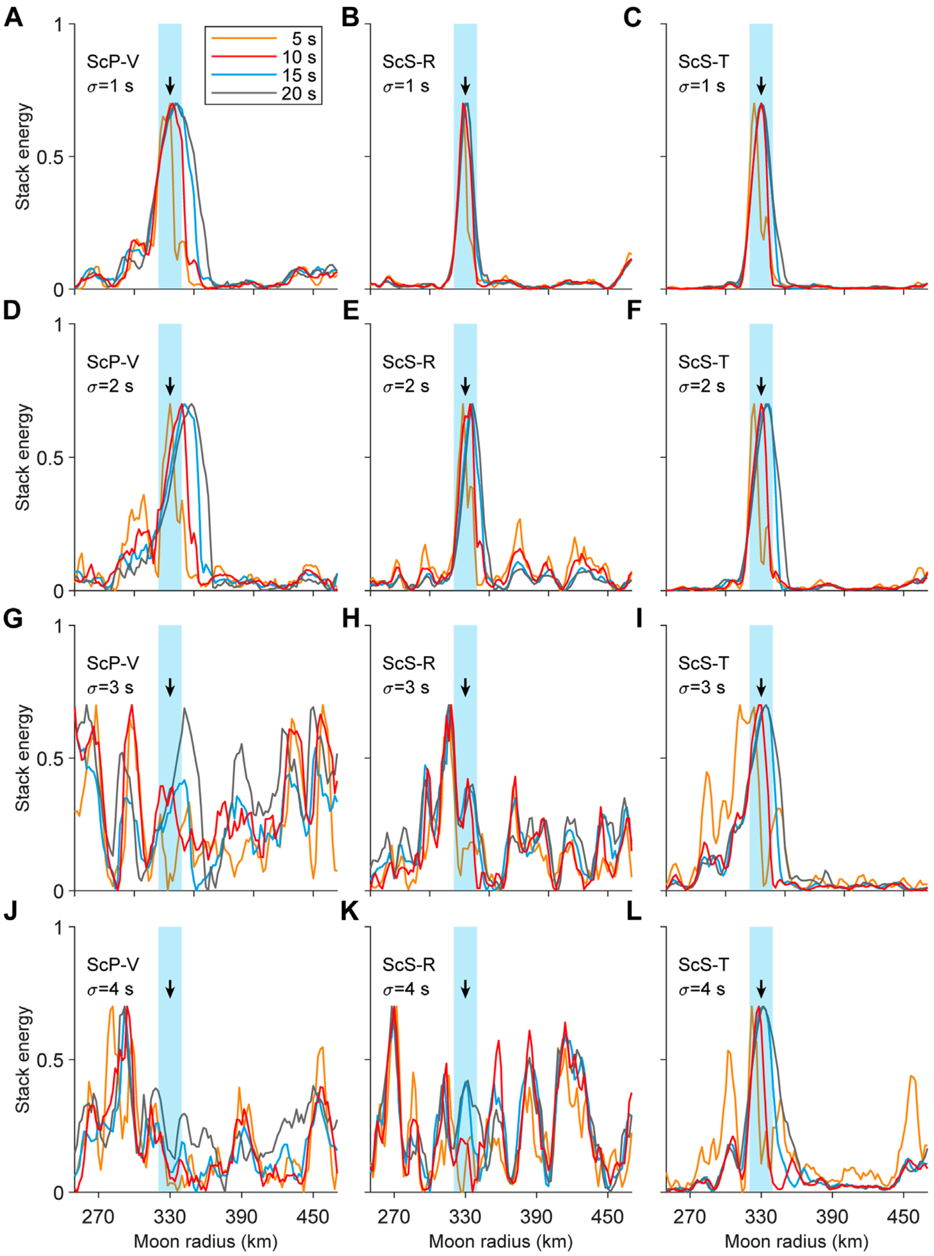


FIGURE 7
Stacking results of ScP phases on V components **(A, D, G, J)** and ScS phases on R **(B, E, H, K)** and T **(C, F, I, L)** components with random arrival time errors with standard deviation of 1, 2, 3, and 4 s, respectively. The black arrows represent the outer core radius in the model (330 km), and the light blue strips represent an error range of 10 km. The four colored curves represent the stacking results of different time windows.

analyze the reliability of stacking results of weak seismic phases. In our modeling, amplitudes of PcP and PcS phases are weaker than that of ScP and ScS. Seismic phases with different amplitudes are used to analyze the reliability of stacking.

3.2 Polarization filtering of synthetic waveforms

In order to eliminate the influence of reverse polarity of seismic phases, Weber et al. (2011) conducted Hilbert transform on the polarized waveforms and took the absolute value to obtain the envelopes. In order to verify the direct impact of the two operations on synthetic waveforms, we perform the same polarization filtering (time window of 2.4 s) on the simulated scattered waveforms (source time of 5 s) and take the envelopes.

Figures 2A, C, E show the scattered waveforms after polarization filtering on V, R, and T components. Compared to the original waveforms (Figures 1B, D, F), the seismic phases of the polarized waveforms are more prominent, especially on T component. After polarization filtering, the in-phase axes of the three seismic phases (S, ScS, and sScS) are very obvious.

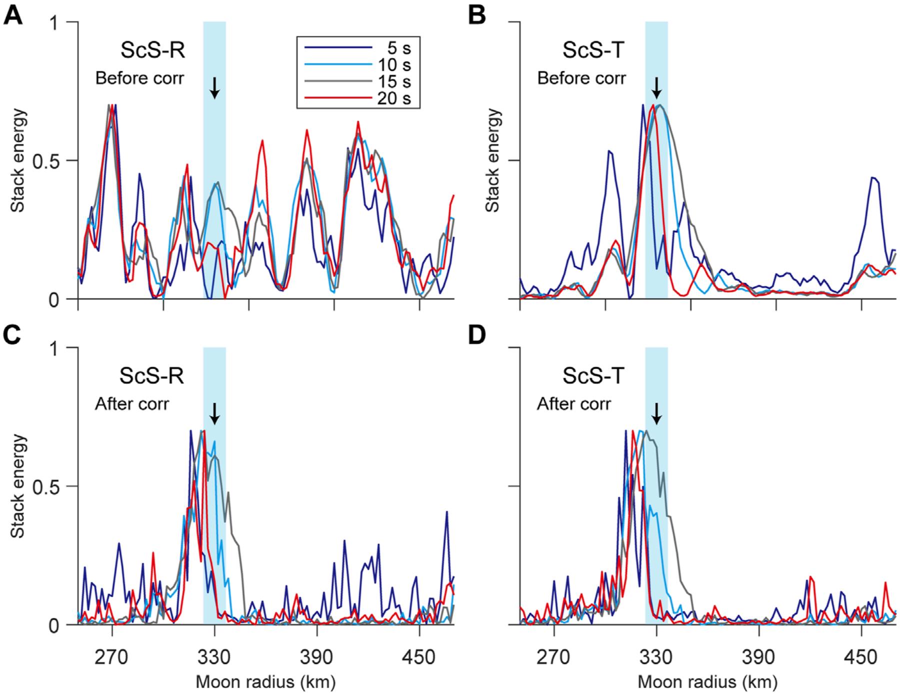


FIGURE 8
The stacking results of ScS phases on R and T components with arrival time errors of standard deviation of 4 s **(A, B)** before cross-correlation alignment **(C, D)** after cross-correlation alignment. The black arrows represent the outer core radius in the model (330 km), and the light blue strips represent an error range of 10 km. The four colored curves represent the stacking results of different time windows.

Figures 2B, D, F are envelopes of polarized waveforms, which show more obvious in-phase axis of the seismic phase.

As shown in Figure 2, the in-phase axes of four seismic phases (ScS, ScP, sScS, and sScP) are the most obvious and other phases are masked by coda. There are stronger coda and noises in Apollo moonquake observation. Therefore, Garcia et al. (2011) believe polarization filtering will enhance the amplitude of noises rather than that of seismic phases used in the processing of Apollo seismic data. To analyze the effect of polarization filtering on scattered waveforms with strong noises, we added Gaussian noises (SNR = 10) to synthetic moonquake waveforms. Then we align ScS and ScP phases by their arrival times and select waveforms within 20 s time window before and after arrivals of seismic phases (i.e., with a window of 40 s).

Figure 3 shows scattered waveforms with noises, waveforms after polarization filtering and their envelopes on R component. ScS seismic phases with strong amplitudes (Figure 3A), whose in-phase axis are still obvious, are less affected by noises, while ScP seismic phases with weak amplitudes (Figure 3D) are submerged in noises. However, polarization filtering enhances the amplitudes of ScP phases and restore an obvious in-phase axis (Figure 3E). In conclusion, polarization filtering can suppress scattered waves and noises and enhance amplitudes of seismic phases, which is suitable to process the scattered and noisy data before stacking.

4 Array stacking of synthetic waveforms

4.1 Performances of stacking by different seismic phases

Weber et al. (2011) synthesized the stacking results of four seismic phases (PcP, PcS, ScP, and ScS) to estimate the radius of the lunar core. According to the analysis of simulated moonquake waveforms, the in-phase axes of PcP and PcS phases are not as clear as that of ScP and ScS phases. However, the in-phase axes of sScP and sScS phases can be seen in the waveforms, which means inconsistent reliability of seismic phases on Double Array Stacking and requires further analysis.

We picked up waveforms of 8 seismic phases reflected and converted from lunar core by their arrival times (i.e., PcP, PcS, ScP, ScS, sPcP, sPcS, sScP, and sScS). Besides, the width of time window also affects the stacking results. Stacking results of short windows are less affected by scattered waveforms, but may not contain the complete waveform of seismic phases, while stacking results of long windows will be affected by the scattered waveforms and other seismic phases. Therefore, we choose four time window lengths (5 s, 10 s, 15 s, and 20 s) for each seismic phase stacking to find the appropriate window length. Figures 4, 5 show the stacking results of waveforms without scattering of PcP, PcS, ScP, and ScS phases and sPcP, sPcS, sScP, and sScS phases, respectively.

The peaks of the stacking results of PcP and sPcP phases on V component (Figures 4J, 5J) are relatively stable and doesn't change

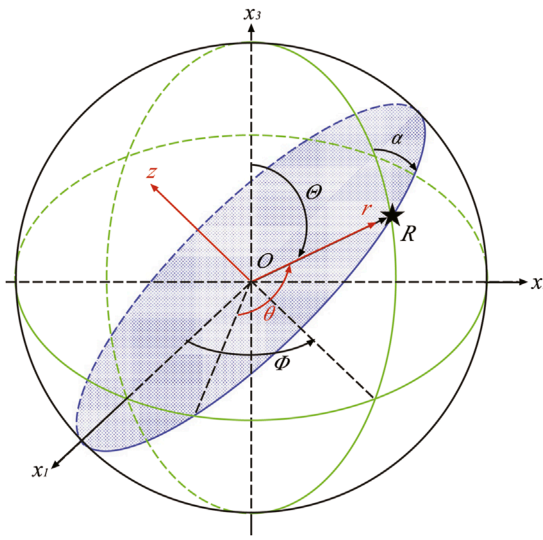


FIGURE 9
Coordinate transformation from the spherical coordinate system (R, Θ, and Φ) to the cylindrical coordinate system (r, θ, and z) (Wang et al., 2014). The blue shaded plane is the two-dimensional lunar model we used, which is a greta circle plane passing through the source and station. α is the azimuth of the source-station path at the epicenter.

TABLE 1 **The moment tensors of simulated moonquakes.**

	$M_{\theta\theta}$	$M_{r\theta}$	M_{rr}
1	1	0	−1
2	1	0	0
3	0	1	0
4	0	0	1

with the time window while peaks on T component (Figure 4I) change significantly, indicating that energy on V component comes from the seismic phases and energy on T component mainly comes from scattered waves and noises. However, energies of PcP and sPcP phases on V component appear multiple peaks (Figures 4J, 5J), with one peak at 330 km, indicating that PcP and sPcP phases are mixed with other seismic phases. Multiple energy peaks make PcP and sPcP phases not suitable for estimating the core radius. The energy peak of PcS and sPcS phases on R and T components changes with window length, indicating strong interference from scattered waves, and noises. Although peak of PcS on V component (Figure 4G) is stable but deviated from 330 km. Travel times of sPcS and sScP phases are very close, so the peak (Figure 5G) at 330 km in stacking results of sPcS on V component is from sScP phases.

The energy peak of ScP and sScP phases on V component (Figures 4D, 5D) is clear at 330 km but becomes broad as the length of window increases. Although energy peak of ScP on R component is also at around 330 km but the energy peak on V component is more stable than that on R component (Figure 4E), because the amplitude of ScP on V component is stronger. The ScS and sScS phases are the best and most suitable for estimating the radius of lunar core. The energy peaks of ScS and sScS on three components are at 330 km with narrow width. Compared to stacking energy of ScS phase, stacking energy of sScS phase appears several secondary peaks, corresponding to other seismic phases before and after the main seismic phase.

In summary, the stacking energy of P-wave seismic phases on V component (PcP, sPcP, ScP, and sScP) and of the S-wave seismic phases (PcS, sPcS, ScS, and sScS) on T component appear peaks at core radius. The stronger the amplitudes of seismic phases, the weaker the interference from scattered waves, noises and other seismic phases, and the narrower the energy peak is. Due to the strong interference from scattered waves and other seismic phases, four seismic phases (PcP, sPcP, PcS, and sPcS) aren't suitable for Double Array Stacking. Meanwhile, the other four seismic phases (ScS, sScS, ScP, and sScP) are suitable for estimating the lunar core radius.

4.2 Influence of coda and noises, polarization filtering and envelope on stacking

To analyze the influence of scattered waves and noises, polarization filtering and envelope on stacking results, we compare the results of ScP and sScP phases on V component and ScS and sScS phases on R component. Figure 6 shows the comparison of stacking results after normalization. The purple, red, blue, and gray solid lines represent the stacking results of waveforms without scattering, waveforms with scattered waves and noises, waveforms after polarization filtering, and their envelopes, respectively. It can be seen that coda and noises directly reduce the energy but the position of peaks is consistent with the stacking results of waveforms without scattering. In addition, polarization filtering effectively suppresses secondary peaks in stacking results of ScP, sScP, and sScS phases, which is most obvious in stacking of sScP and sScS phases. However, compared with other factors, envelope broadens and shifts the peaks in the stacking results and even produces false peaks. It will be more obvious after normalization separately (Supplementary Figure S2). Broadened peaks can be attributed to lower main frequency of waveforms reduced by envelope while false peaks can be attributed to the loss of phase information, indicating that envelopes aren't suitable for estimation of the core radius.

4.3 Influence of random arrival time errors on stacking

The arrivals of seismic phases in synthetic waveforms are in good agreement with the theoretical travel time curves. While in observations, there are arrival time errors due to source parameter error, structure difference between Moon model and actual Moon and errors in seismic phase picking caused by coda and noises. Because we align the waveforms by theoretical travel time curves before stacking, arrival time errors will definitely have a significant impact on the stacking results, so it is necessary to analyze the performance of stacking results of different seismic phases with different degrees of arrival time errors. We generate 4 groups of

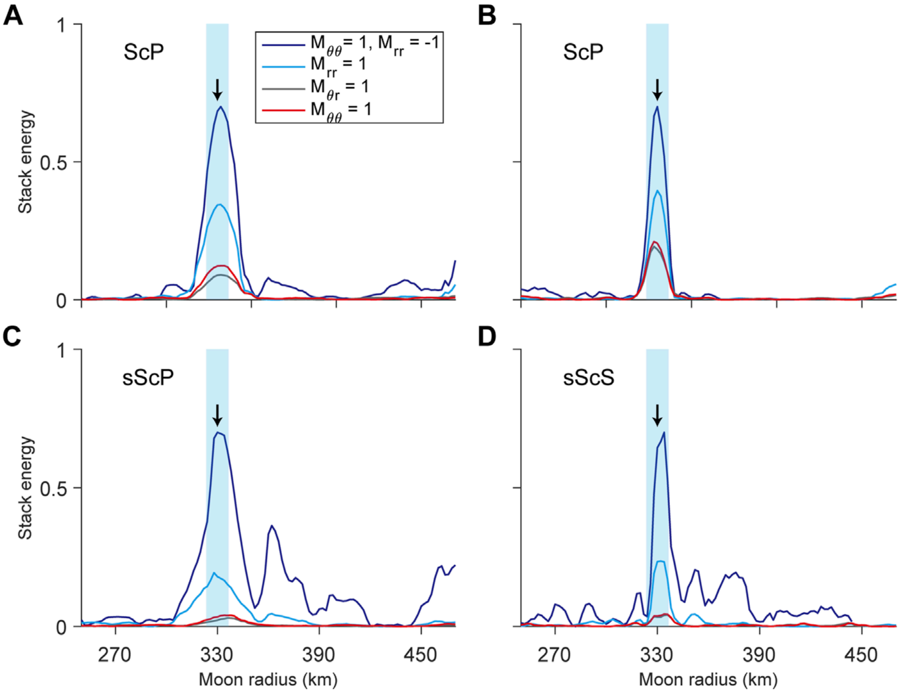


FIGURE 10
Comparison of stacking results of wavefroms of four moonquakes with different focal mechanisms [ScP **(A)**, ScS **(A)**, sScP **(C)** and sScS **(D)** phases]. Purple, red, blue, and gray lines represent the stakcing results of the four focal mechanisms in **Table 1** without normalization, respectively. The black arrows represent the outer core radius in the model (330 km), and the light blue strips represent an error range of 10 km.

normal random numbers with a standard deviation of 1, 2, 3, and 4 s and add them into theoretical travel time curves. We use the travel time curves with errors to cut waveform windows of two seismic phases (ScP and ScS) and analyze their stacking results.

Figure 7 shows the stacking results of ScP phases on V component, ScS phases on R and T components. It can be seen that arrival time errors obviously affect the stacking results of ScP phases. As shown in **Figure 7D**, the peak deviate from 330 km, the core radius in the model when standard deviation of arrival time error is 2 s while the peaks in stacking results of ScS phase on R and T components are still consistent with core radius in the model (**Figures 7E, F**). When the arrival time error is larger (σ = 3 s, 4 s), ScP phase stacking can't obtain ideal results. Multiple peaks appear and deviate from core radius (330 km) in the model. At the same time, stacking results of ScS phase on R component become unreliable, showing peak deviation and multiple peaks. It is worth noting that the stacking results of the ScSH phase on T component are minimally affected by arrival time errors and can obtain the core radius of the Moon even when the arrival time error is large, which can be attributed to the lack of interference from the P waves on T component.

From our simulation tests, it can be seen that ScS phase on T component is the best for Double Array Stacking, superior to ScS phase on R component and ScP phase on V component. For existing Apollo moonquake data, the inspiration is that we can improve the reliability and accuracy of results of Double Array Stacking by optimizing P and S wave picking, accuracy of source location, and eliminating data with significant arrival time errors. For moonquake observation data with better signal-to-noise ratio and wider frequency band in the future, we can use cross-correlation to reduce the impact of arrival time errors. We conducted a simple test on the effect of cross-correlation using synthetic waveforms. Before stacking, we align the waveforms based on time shifts calculated from cross-correlation. The stacking results of ScS phases on R component (**Figures 8A–C**) and T component (**Figures 8B–D**) are improved obviously, showing the effectiveness of cross-correlation.

4.4 Influence of focal mechanisms and azimuth on stacking

4.4.1 Variation of amplitude of seismic phases

The focal mechanisms of moonquakes used for Double Array Stacking are unknown, the distribution of sources and stations is relatively scattered, and the stations aren't high-density arrays. These factors can lead to inconsistent polarity and amplitude of seismic phases in different observed moonquake waveforms. In order to analyze the influence of inconsistent amplitude and polarity on Double Array Stacking, we need synthetic waveforms with multiple focal mechanisms. In this study, we use a great circle cross-section (**Figure 9**), passing through stations and seismic sources, as the two-dimensional model. The projection of moment tensor of a point source of moonquakes from spherical (R, Θ, and Φ) to a line source in cylindrical (r, θ, and z) coordinates (Wang et al., 2014) can be

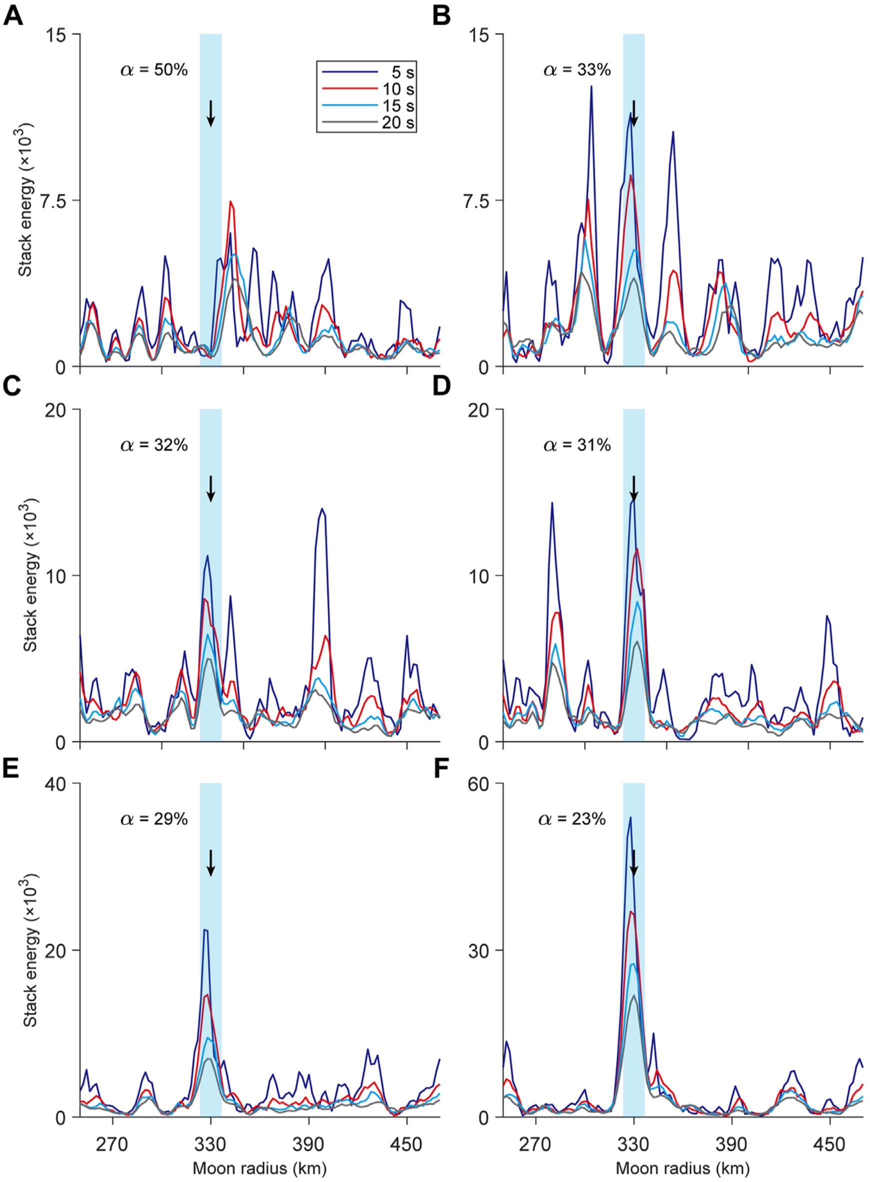


FIGURE 11
The stacking results of the waveforms of different α [50% **(A)**, 33% **(B)**, 32% **(C)**, 31% **(D)**, 29% **(E)**, 23% **(F)**]. The black arrows represent the outer core radius in the model (330 km), and the light blue strips represent an error range of 10 km. The four colored curves represent the stacking results of different time windows.

given as:

$$\begin{cases} M_{rr} = M_{RR} \\ M_{r\theta} = M_{\theta r} = -M_{R\Theta}\cos\alpha + M_{R\Phi}\sin\alpha \\ M_{\theta\theta} = M_{\Theta\Theta}\cos^2\alpha - 2M_{\Theta\Phi}\sin\alpha\cos\alpha + M_{\Phi\Phi}\sin^2\alpha \\ M_{z\theta} = M_{\theta z} = (M_{\Theta\Theta} - M_{\Phi\Phi})\sin\alpha\cos\alpha + M_{\Theta\Phi}\cos 2\alpha \\ M_{rz} = M_{zr} = -M_{\Theta R}\sin\alpha - M_{\Phi R}\cos\alpha \\ M_{zz} = M_{\Theta\Theta}\sin^2\alpha + 2M_{\Theta\Phi}\sin\alpha\cos\alpha + M_{\Phi\Phi}\cos^2\alpha \end{cases} \quad (7)$$

where M_{rr}, $M_{\theta\theta}$, $M_{r\theta}$, $M_{z\theta}$, M_{rz}, M_{zz} are moment tensors for a line source in the cylindrical coordinate systems while M_{RR}, $M_{\Theta\Theta}$, $M_{R\Theta}$, $M_{R\Phi}$, $M_{\Phi\Phi}$, $M_{\Theta\Phi}$ are moment tensors for a point source in the spherical coordinate system, respectively. Transformation of source mechanism is related with azimuth α, which means that different source mechanisms and azimuths in the spherical coordinate system exhibit different source mechanisms in the cylindrical coordinate system.

We simulate several moonquakes of different focal mechanisms in cylindrical coordinate systems (Table 1), and discuss performance of their stacking results. There are significant differences in the waveforms of moonquakes of different focal mechanisms. In some moonquakes, reflected and converted phases are obvious, while in other moonquakes, ScP and ScS phases are weak and masked by coda. Even after polarization filtering, those seismic phases can't be clearly seen.

In order to analyze the influence from the difference of seismic phase amplitudes, we stack four phases (ScP, ScS, sScP, and sScS) of these four moonquakes, separately. **Figure 10** shows the comparison of stacking images after separate normalization. It can be seen that the variation of amplitudes directly affects the energy of stacking results, but positions of the peaks are still consistent. After normalization separately, the stacking energy images of waveforms of different focal mechanism almost overlap (**Supplementary Figure S3**), indicating that there is little influence of amplitude variation of seismic phases on lunar core estimation.

4.4.2 Waveform polarities

The difference in focal mechanisms and azimuths not only affects the amplitude of waveforms, but also affects the polarity of seismic phases. Stacking waveforms with reverse polarity will reduce the amplitude of seismic phases reflected and converted from core, leaving energy of noises and scattered waves behind. Weber et al. (2011) used envelope to prevent potential influence of waveforms with reverse polarity. However, our stacking tests show envelopes aren't suitable for stacking. Due to the strong coda and noises, it is difficult to analyze the polarity of the seismic phases from core. Therefore, it is necessary to analyze the stacking results of waveforms with different proportion of reverse polarity to show the effects of polarity variation on stacking.

We linearly stack the waveforms of two different source mechanisms ($M_{\theta\theta} = 1$, $M_{rr} = 1$) to obtain enough waveforms of different focal mechanisms, and control the proportion of waveforms with opposite polarity:

$$\begin{cases} w_{\theta\theta} = \dfrac{1}{1-\alpha} \cdot \mathrm{Rand}(N) - 1 \\ w_{rr} = -\mathrm{sign}(w_{\theta\theta}) \cdot \left(1 - {w_{\theta\theta}}^{2}\right)^{\frac{1}{2}} \\ W_i = w_{\theta\theta}^{i} W_{\theta\theta}^{i} + w_{rr}^{i} W_{rr}^{i} \end{cases} \tag{8}$$

where $w_{\theta\theta}, w_{rr}$ are the coefficients of two types of waveforms, α represents the proportion of waveforms with reverse polarity and N is the number of waveforms. Rand(N) generates N random numbers between (0, 1).

As shown in **Figure 11**, stacking results of ScS phases vary when the portion factor (α) changes. When the proportion of waveforms with reverse polarity reaches 50%, the energy of the stacking waveforms is weak with peaks come from noises and scattered waves. When α is around 31%–33%, the stacking energy is several times stronger than the energy when α is 50%. The peak corresponds to the radius of the lunar core, while the local peaks come from scattered waves and noises. When α is less than 30%, the stacking energy is very strong, and the peak is located near 330 km, the given radius of the core. If the stacking results of Apollo data present strong energy and obvious peak, it indicates that they can be used to obtain the radius of the lunar core. Conversely, it indicates that proportion of waveforms with reverse polarity is high so the stacking results are unreliable.

5 Discussion and conclusion

We modeled the propagation of P-SV and SH waves in the Moon and calculated moonquake waveforms with major frequencies of 0.2 and 0.3 Hz. By comparison, it was found that the relative amplitude of the reflected phases from core mantle boundary in low frequency are stronger and therefore are more suitable for studying the radius of lunar core. After polarization filtering, the reflection phases are more prominent, especially for ScS and sScS phases on T component. Envelopes with lower frequency also enhance the events but reduce the sensitivity of the estimated core size.

Through stacking tests of multiple phases, it is found that without other interference factors, the stacking results on V component of P-wave phases (PcP, sPcP, ScP, and sScP) and T component of S-wave phases (PcS, sPcS, ScS, and sScS) can obtain core radius in the model correctly. The amplitudes of ScS, sScS, ScP, and sScP phases are stronger so their stacking results are less affected by coda and noises and obtain more accurate core radius.

By comparing the stacking results of waveforms without scattering, waveforms with scattering and noises, and waveforms after polarization filtering and their envelopes, it is found that coda and noises affect the stacking energy but don't change the energy peak position, therefore don't affect obtaining radius of the core. After polarization filtering, secondary peaks are suppressed which is beneficial for estimating radius of the lunar core. However, false peaks appear in the stacking results of envelopes, which can be attributed to loss of the information of seismic phase.

After considering arrival time errors, there is a significant difference in the reliability of stacking results of different seismic phases, with ScSH phases being the most reliable, indicating that the analysis of moonquake waveforms should be based on the results of ScSH phases. In the future, the analysis and collection of moonquake data should consider how to reduce arrival time errors such as deploying instruments underground to reduce the impact of strong scattering in the regolith.

By comparing the stacking results of waveforms from different focal mechanisms, we analyze the influence of differences in amplitude and polarity of waveforms caused by different station azimuths and focal mechanisms. The difference of amplitude of seismic phases only affects the energy of the stacking results but doesn't affect their morphology obviously, so the estimation of the core radius isn't much affected. When the proportion of seismic phases with reverse polarity involved in stacking is less than 30%, the peak can directly reflect the core radius of the Moon.

In this study, we clarify the influences of scattered waves, noises, arrival time errors, variation of amplitude, and polarity of waveforms on the identification of lunar core through numerical modeling of moonquake waveforms with scattering. This study confirms the feasibility of the Double Array Stacking method, and the enhancement on seismic phases and their stacking results of polarization filtering, and finds that the envelopes (Weber et al., 2011; Yuan, 2018) aren't suitable for studying the radius of the lunar core. Results of this study may provide an important reference for subsequent data mining of moonquakes, and help to study the depth of interior discontinuities of the Moon and other planets in the future.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

Author contributions

BY and YW contributed to conception and design of the study. BY performed numerical modeling and the statistical analysis. BY wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work has been supported by the National Natural Science Foundation of China (Grant No. 41930103).

Conflict of interest

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fspas.2023.1217990/full#supplementary-material

References

Arlitt, R., Kissling, E., Ansorge, J., and Tor Working Group, (1999). Three-dimensional crustal structure beneath the TOR array and effects on teleseismic wavefronts. *Tectonophysics* 314, 309–319. doi:10.1016/s0040-1951(99)00250-4

Blanchette-Guertin, J. F., Johnson, C. L., and Lawrence, J. F. (2012). Investigation of scattering in lunar seismic coda. *J. Geophys. Res.* 117, E06003. doi:10.1029/2011je004042

Bostock, M. G., and Rondenay, S. (2002). Migration of scattered teleseismic body waves. *Geophys. J. Int.* 137, 732–746. doi:10.1046/j.1365-246x.1999.00813.x

Castle, J. C., and Creager, K. C. (1999). A steeply dipping discontinuity in the lower mantle beneath Izu-Bonin. *J. Geophys. Res.* 04, 7279–7292. doi:10.1029/1999jb900011

Cerjan, C., Kosloff, D., Kosloff, R., and Reshef, M. (1985). A nonreflecting boundary condition for discrete acoustic and elastic wave equations. *Geophysics* 50 (4), 705–708. doi:10.1190/1.1441945

Frohlich, C., and Nakamura, Y. (2009). The physical mechanisms of deep moonquakes and intermediate-depth earthquakes: How similar and how different. *Phys. Earth Planet. Interiors* 173 (3-4), 365–374. doi:10.1016/j.pepi.2009.02.004

Garcia, R. F., Gagnepain-Beyneix, J., Chevrot, S., and Lognonné, P. (2011). Very preliminary reference Moon model. *Phys. Earth Planet. Interiors* 188, 96–113. doi:10.1016/j.pepi.2011.06.015

Garcia, R. F., Khan, A., Drilleau, M., Margerin, L., Kawamura, T., Sun, D., et al. (2019). Lunar seismology: An update on interior structure models. *Space Sci. Rev.* 215 (8), 50. doi:10.1007/s11214-019-0613-y

Graves, R. W. (1996). Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences. *Bull. Seismol. Soc. Am.* 86 (4), 1091–1106.

Helmberger, D. V., and Vidale, J. E. (1988). Modeling strong motions produced by earthquakes with two-dimensional numerical codes. *Bull. Seismol. Soc. Am.* 78 (1), 109–121. doi:10.1785/bssa0780010109

Herrmann, R. B. (1979). SH-Wave generation by dislocation sources—a numerical study. *Bull. Seismol. Soc. Am.* 69 (1), 1–15. doi:10.1785/bssa0690010001

Hood, L. L., Mitchell, D. L., Lin, R. P., Acuna, M. H., and Binder, A. B. (1999). Initial measurements of the lunar induced magnetic dipole moment using Lunar Prospector magnetometer data. *Geophys. Res. Lett.* 26, 2327–2330. doi:10.1029/1999gl900487

Jarosch, H. S. (1977). The use of surface reflections in lunar seismograms. *Bull. Seismol. Soc. Am.* 67, 1647–1659. doi:10.1785/bssa0670061647

Jiang, X. H., Wang, Y. B., and Takashi, F. (2015). Numerical simulation of lunar seismic coda caused by scattering in upper Moon crust. *Chin. J. Geophys. (in Chinese)* 58 (5), 1675–1691.

Jiang, X. H., Wang, Y. B., Qin, Y., and Takenaka, H. (2015). Global SH-wave propagation in a 2D whole Moon model using the parallel hybrid PSM/FDM method. *Earthquake Science* 28, 163–174. doi:10.1007/s11589-015-0121-4

Kárason, H., and van der Hilst, R. D. (2001). Tomographic imaging of the lowermost mantle with differential times of refracted and diffracted core phases (PKP, Pdiff). *Journal of Geophysical Research* 106 (B4), 6569–6587. doi:10.1029/2000jb900380

Khan, A., Connolly, J. A. D., Pommier, A., and Noir, J. (2014). Geophysical evidence for melt in the deep lunar interior and implications for lunar evolution. *Journal of Geophysical Research Planets* 119 (10), 2197–2221. doi:10.1002/2014je004661

Khan, A., Maclennan, J., Taylor, S. R., and Connolly, J. A. D. (2006). Are the earth and the moon compositionally alike? Inferences on lunar composition and implications for lunar origin and evolution from geophysical modeling. *Journal of Geophysical Research Planets* 111 (E5), E05005. doi:10.1029/2005je002608

Khan, A., and Mosegaard, K. (2005). Further constraints on the deep lunar interior. *Geophysical Research Letters* 32, 22203. doi:10.1029/2005gl023985

Khan, A., Mosegaard, K., Williams, J. G., and Lognonné, P. (2004). Does the moon possess a molten core? Probing the deep lunar interior using results from LLR and lunar prospector. *Journal of Geophysical Research Planets* 109, 09007. doi:10.1029/2004je002294

Kronrod, V. A., and Kuskov, O. L. (2011). Inversion of seismic and gravity data for the composition and core sizes of the Moon. *Izvestiya, Physics of the Solid Earth* 47, 711–730. doi:10.1134/s1069351311070044

Krüger, F., Baumann, M., Scherbaum, F., and Weber, M. (2001). Mid-mantle scatterers near the Mariana slab detected with a double array method. *Geophysical Research Letters* 28, 667–670. doi:10.1029/2000gl011570

Kuskov, O. L., and Belashchenko, D. K. (2016). Thermodynamic properties of Fe–S alloys from molecular dynamics modeling: Implications for the lunar fluid core. *Physics of the Earth and Planetary Interiors* 258, 43–50. doi:10.1016/j.pepi.2016.07.006

Lin, P. Y. P., Weber, R. C., Garnero, E. J., and Schmerr, N. C. (2011). "Further constraints and uncertainties on the deep seismic structure of the moon," in Proceedings of the AGU fall meeting 2011, December 2011. San Francisco, CA, USA.

Lognonné, P., Gagnepain-Beyneix, J., and Chenet, H. (2003). A new seismic model of the moon: Implications for structure, thermal evolution and formation of the moon. *Earth and Planetary Science Letters* 211 (1–2), 27–44. doi:10.1016/s0012-821x(03)00172-9

Lynnes, C. S., and Lay, T. (1989). Inversion of P coda for isotropic scatterers at the Yucca Flat test site. *Bulletin of the Seismological Society of America* 79, 790–804.

Matsumoto, K., Yamada, R., Kikuchi, F., Kamata, S., Ishihara, Y., Iwata, T., et al. (2015). Internal structure of the Moon inferred from Apollo seismic data and selenodetic data from GRAIL and LLR. *Geophysical Research Letters* 42, 7351–7358. doi:10.1002/2015gl065335

Morard, G., Bouchet, J., Rivoldini, A., Antonangeli, D., et al. (2018). Liquid properties in the Fe–FeS system under moderate pressure: Tool box to model small planetary cores. *American Mineralogist* 103 (11), 1770–1779.

Nakamura, Y., Latham, G., Lammlein, D., Ewing, M., Duennebier, F., and Dorman, J. (1974). Deep lunar interior inferred from recent seismic data. *Geophysical Research Letters* 1, 137–140. doi:10.1029/gl001i003p00137

Neal, S. L., and Pavlis, G. L. (2001). Imaging P- to S conversions with broad-band seismic arrays using multichannel time-domain deconvolution. *Geophysical Journal International* 146, 57–67. doi:10.1046/j.1365-246x.2001.00507.x

Neal, S. L., and Pavlis, G. L. (1999). Imaging P- to-S conversions with multichannel receiver functions. *Geophysical Research Letters* 26, 2581–2584. doi:10.1029/1999gl900566

Onodera, K., Kawamura, T., Tanaka, S., Ishihara, Y., and Maeda, T. (2021). Numerical simulation of lunar seismic wave propagation: Investigation of subsurface scattering properties near Apollo 12 landing site. *Journal of Geophysical Research Planets* 126, e2020JE006406. doi:10.1029/2020je006406

Revenaugh, J. (1995). The contribution of topographic scattering to teleseismic coda in southern California. *Geophysical Research Letters* 22, 543–546. doi:10.1029/95gl00162

Ritter, J. R. R., Jordan, M., Christensen, U. R., and Achauer, U. (2001). A mantle plume below the Eifel volcanic fields, Germany. *Earth and Planetary Science Letters* 186 (1), 7–14. doi:10.1016/s0012-821x(01)00226-6

Rost, S., and Revenaugh, J. (2001). Seismic detection of rigid zones at the top of the core. *Science* 294, 1911–1914. doi:10.1126/science.1065617

Rost, S., and Thomas, C. (2002). Array seismology: Methods and applications. *Review of Geophysics* 40 (3), 2-1–2-27. doi:10.1029/2000rg000100

Shimshoni, M., and Smith, S. W. (1964). Seismic signal enhancement with three-component detectors. *Geophysics* 29, 664–671. doi:10.1190/1.1439402

Thomas, C., Kendall, J. M., and Weber, M. (2002). The lowermost mantle beneath northern Asia—I. Multi-Azimuth studies of a D$''$ heterogeneity. *Geophysical Journal International* 151 (1), 279–295. doi:10.1046/j.1365-246x.2002.01759.x

Vidale, J. E., and Earle, P. S. (2000). Fine-scale heterogeneity in the Earth's inner core. *Nature* 404, 273–275. doi:10.1038/35005059

Wang, Y. B., Luo, Y., Qin, Y. F., and Zhang, X. B. (2014). Global SH-wavefield calculation for a two-dimensional whole-Earth model with the parallel hybrid PSM/FDM algorithm. *Earthquake Science* 27, 277–284. doi:10.1007/s11589-014-0085-9

Wang, Y. B., Takenaka, H., and Furumura, T. (2001). Modelling seismic wave propagation in a two dimensional cylindrical whole Earth model using the pseudospectral method. *Geophysical Journal International* 145 (3), 689–708. doi:10.1046/j.1365-246x.2001.01413.x

Wang, Y. B., Takenaka, H., Jiang, X. H., and Lei, J. (2013). Modelling two-dimensional global seismic wave propagation in a laterally heterogeneous whole-moon model. *Geophysical Journal International* 192 (3), 1271–1287. doi:10.1093/gji/ggs094

Wang, Y. B., and Takenaka, H. (2011). SH-wavefield simulation for a laterally heterogeneous whole-Earth model using the pseudospectral method. *Science China Earth Sciences* 54 (12), 1940–1947. doi:10.1007/s11430-011-4244-8

Weber, R. C., Lin, P., Garnero, E. J., Williams, Q., and Lognonne, P. (2011). Seismic detection of the lunar core. *Science* 331, 309–312. doi:10.1126/science.1199375

White, J. E. (1964). Motion product seismograms. *Geophysics* 29, 288–298. doi:10.1190/1.1439360

Williams, J. G., Boggs, D. H., Yoder, C. F., Ratcliff, J. T., and Dickey, J. O. (2001). Lunar rotational dissipation in solid body and molten core. *Journal of Geophysical Research* 106, 27933–27968. doi:10.1029/2000je001396

Williams, J. G., Konopliv, A. S., Boggs, D. H., Park, R. S., Yuan, D. N., Lemoine, F. G., et al. (2014). Lunar interior properties from the GRAIL mission. *Journal of Geophysical Research Planets* 119, 1546–1578. doi:10.1002/2013je004559

Yuan, Y. F. (2018). *The analysis of the lunar core structure based on lunar seismic data*. Wuhan, China: Department of Geophysics, School of Geophysics and Spatial Information, China University of Geosciences.

Zhang, X., and Zhang, J. H. (2021). Research progress and prospect of moonquakes. *Reviews of Geophysics and Planetary Physics (in Chinese)* 52 (4), 391–401.

Zhang, X., Zhang, L., Zhang, J., and Mitchell, R. N. (2022). Strong heterogeneity in shallow lunar subsurface detected by Apollo seismic data. *Journal of Geophysical Research Planets* 127, e2022JE007222. doi:10.1029/2022je007222

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Original Research
PUBLISHED 28 June 2023
DOI 10.3389/fspas.2023.1188232

OPEN ACCESS

EDITED BY
Francesca Altieri,
Institute for Space Astrophysics and Planetology (INAF), Italy

REVIEWED BY
Alice Le Gall,
Université de Versailles Saint-Quentin-en-Yvelines, France
Jianguo Yan,
Wuhan University, China

*CORRESPONDENCE
Jinhai Zhang,
✉ zjh@mail.iggcas.ac.cn

RECEIVED 17 March 2023
ACCEPTED 14 June 2023
PUBLISHED 28 June 2023

CITATION
Lv W and Zhang J (2023), High-resolution permittivity estimation of ground penetrating radar data by migration with isolated hyperbolic diffractions and local focusing analyses. *Front. Astron. Space Sci.* 10:1188232. doi: 10.3389/fspas.2023.1188232

High-resolution permittivity estimation of ground penetrating radar data by migration with isolated hyperbolic diffractions and local focusing analyses

Wenmin Lv[1,2,3,4] and Jinhai Zhang[1,2,3,4]*

[1]Engineering Laboratory for Deep Resources Equipment and Technology, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China, [2]Key Laboratory of Earth and Planetary Physics, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China, [3]Innovation Academy for Earth Science, Chinese Academy of Sciences, Beijing, China, [4]College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing, China

Ground penetrating radar (GPR) is important for detecting shallow subsurface structures, which has been successfully used on the Earth, Moon, and Mars. It is difficult to analyze the underground permittivity from GPR data because its observation system is almost zero-offset. Traditional velocity analysis methods can work well with separable diffractions but fail with strong-interfered diffractions. However, in most situations, especially for lunar or Martian exploration, the diffractions are highly interfered, or even buried in reflections. Here, we proposed a new method to estimate the underground permittivity and apply it to lunar penetrating radar data. First, we isolate a group of diffractions with a hyperbolic time window determined by a given velocity. Then, we perform migration using the given velocity and evaluate the focusing effects of migration results. Next, we find the most focused results after scanning a series of velocities and regard the corresponding velocity as the best estimation. Finally, we assemble all locally focused points and derive the best velocity model. Tests show that our method has high spatial resolution and can handle strong noises, thus can achieve velocity analyses with high accuracy, especially for complex materials. The permittivity of lunar regolith at Chang'E-4 landing area is estimated to be ~4 within 12 m, ranging from 3.5 to 4.2 with a local perturbation of ~2.3%, consistent with ~3% obtained by numerical simulations using self-organization random models. This suggests that the lunar regolith at Chang'E-4 landing area is mature and can be well described by self-organization random models.

KEYWORDS

ground penentrating radar, diffraction isolation, migration, Moon, chang'E-4

1 Introduction

Ground penetrating radar (GPR) is important for detecting shallow subsurface structures, which has been successfully used on the Earth, Moon, and Mars (Boisson et al., 2009; Jordan et al., 2009; Fang et al., 2014; Xiao et al., 2015; Zhang et al., 2015; Schroeder et al., 2019; Hamran et al., 2020; Zhou et al., 2020; Oudart et al., 2021; Zhang X et al., 2021; Li et al., 2022). However, it is difficult to analyze the underground velocity because the observation system of the GPR is almost zero-offset (i.e., the transmitter and the receiver are close to each other in position), which has no enough number of multiple offsets (Su et al., 2014; Zhang et al., 2014; Fa et al., 2015; Feng et al., 2019). Consequently, we can only estimate the velocity using methods originally developed for zero-offset data (Economou et al., 2020; Leong and Zhu, 2021). Fortunately, the zero-offset radar profiles are similar to the post-stack seismic profiles; thus, we can take the benefits of the velocity estimation methods that were developed in seismic exploration field for post-stack seismic profiles (Claerbout, 1985; Zhu et al., 1998; Hu et al., 2001; Fomel et al., 2007; Reshef and Landa, 2009; Carpentier et al., 2010).

The simplest method of velocity estimation is to assign a constant velocity for each identifiable layer, assuming that there is no lateral variation in each layer (Dai et al., 2014; Li et al., 2017). Undoubtedly, this method would encounter great uncertainty because it is difficult to evaluate the accuracy of velocity estimations. Another velocity estimation method is to fit hyperbolic-shape diffractions in the GPR profiles by velocity scanning (Yilmaz, 2001; Ristic et al., 2009; Dong et al., 2020; Giannakis et al., 2021). However, this method requires that the diffractions have a relatively high signal-to-noise ratio and have little interference from adjacent hyperbolas. In the presence of strong scattering, the diffractions from different objects mostly gather together rather than staying isolated; thus, it is difficult to separate them with existing methods.

Another method of velocity analysis is to focus the diffractions by migration (Sava et al., 2005; Yuan et al., 2019). The migration-based methods can numerically collect the weak diffractions into smaller areas and undoubtedly enhance the local signal-to-noise ratio in the migration domain (Li and Zhang, 2022a); however, the available amount of diffractions for this method is relatively limited; consequently, we can not cover the whole model and leave some empty regions, especially in the presence of strongly cluttered diffractions. Unfortunately, this is the usual case for the Moon and Mars, because the extraterrestrial bodies had experienced long-term impact. The abundant ejecta in subsurface layers would cause strong interference of diffractions in the GPR data. Therefore, many existing velocity analysis methods can achieve good results for theoretical models with sparse scattering objects but are almost powerless for models with vast strong-scattering objects.

The diffraction separation technique was proposed in seismic exploration field for zero-offset data (Fomel, 2003a; Fomel, 2003b). This method performs plane wave decomposition-based local curvature analysis to separate diffractions from reflections. However, as a data-driven method, it relies on curvature analysis in the data domain (Fomel et al., 2007; Li et al., 2021a; Song et al., 2021; Li and Zhang, 2022b) and is vulnerable to the influence of the original signal-to-noise ratio. Additionally, all rapid local variations of curvature tend to be regarded as diffractions, while the others are regarded as reflections. Thus, the reflections and diffractions can not be fully separated purely by local variations of curvature, since some reflections may also have relatively rapid local variations of curvature and the diffractions may also have relatively slow local variations of curvature. Therefore, new methods on GPR velocity analyses are needed to identify weak diffractions precisely and avoid the interferences from both adjacent diffractions and strong reflections.

In this paper, we proposed a new method to estimate the permittivity (converted from velocity) of subsurface materials based on migration and diffraction isolation. First, we isolate a group of diffractions in the un-migrated domain with hyperbolic windows determined by a given velocity. Then, we perform migration on the isolated data using the given velocity and evaluate the focusing effects of the migration results of isolated diffractions. Next, we pick up the most focused results after scanning a series of velocities and regard the corresponding velocity as the best estimation. Finally, we obtain a dense set of velocity estimation points at all locally focused positions, and the best estimation of the velocity model can be derived after assembling all available points.

2 Methodology

In this section, we show the details of the proposed method for estimating the permittivity of subsurface materials (Figure 1). There are three steps: 1) isolating weak diffractions with a set of hyperbolic windows determined by different velocities; 2) collecting the isolated weak diffractions by migration; 3) focusing analysis by evaluating the ratio of amplitudes between the focused point and around the focused point in the migrated domain.

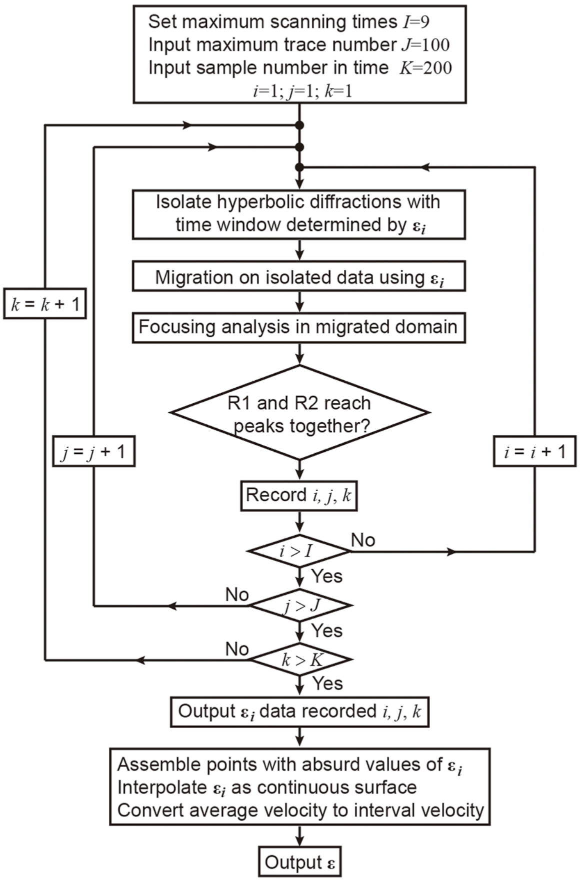


FIGURE 1
The flow chart of the proposed velocity estimation method.

The permittivity of lunar regolith and basalt blocks generally ranges from 3 to 7 (Olhoeft et al., 1975a; Heiken et al., 1991). Within this range, we apply grid searching on nine possible values. In the migrated domain, we define two parameters to evaluate the focusing effects under a given velocity, and the diffractions can be easily identified from cluttered diffractions and reflections.

2.1 Formation of hyperbolic-shape diffractions in GPR profile

The electromagnetic wave can propagate in the air and underground media. We can observe back-propagating waves of reflections on the boundary of diffractors with discontinuous dielectric properties, such as basalt blocks (Hiesinger and Head, 2006; Jolliff et al., 2006; Fa et al., 2015); similarly, we can also observe diffractions when the size of diffractor is comparable to the wavelength of electromagnetic wave. The energy of diffractions is usually smaller than that of the reflections.

We assume that each block of ejecta or rocks can be considered as a point-like diffractor since the influence of scattering body decreases with the increase of depth h_0(m); thus, the travel time of diffractions presents as a hyperbola in the GPR profile (Capineri et al., 2008; Ristic et al., 2009; Kouemou, 2010; Soldovieri and Solimene, 2010; Ding et al., 2020; Lv et al., 2020), when the GPR rover moving from x to x_0 over the diffractor at (x_0, h_0). Figure 2A illustrates the forming principle and geometric relationship of corresponding diffractions as

$$(x_0 - x)^2 + h_0{}^2 = v^2 t^2, \tag{1}$$

which can be transformed into a hyperbolic form as

$$\frac{t^2}{\left(\frac{h_0}{v}\right)^2} - \frac{(x_0 - x)^2}{h_0{}^2} = 1, \tag{2}$$

where t represents the one-way travel time, v represents the average propagating velocity of electromagnetic waves. Given the location of a scatter at (x_0, h_0), the only parameter controlling the shape of hyperbola will be v. The relative magnetic permeability of most subsurface material approximately equals to 1 and the conductivity affects little during the wave propagation for high-frequency radar waves; thus, the media could be regarded as ideal dielectric and v can be calculated by (Heiken et al., 1991)

$$v = \frac{c}{\sqrt{\varepsilon_r}}, \tag{3}$$

where c is the speed of light in vacuum and ε_r is the relative dielectric permittivity (Gold et al., 1972; Olhoeft et al., 1974; Heiken et al., 1991; Rochette et al., 2010). At the same location, for a smaller permittivity, the propagating velocity will be higher and the corresponding hyperbola will be sharper and narrower; in contrast, for a larger permittivity, the velocity will be lower and the corresponding hyperbola will be flatter and broader (Figure 2B).

2.2 Improving signal-to-noise using migration

Migration is a common technique for imaging underground structures (Claerbout, 1985; Leuschen and Plumb, 2001; Stuart, 2003; Özdemir et al., 2014). It can numerically collect a group of diffractions into a small region around the apex if the permittivity used is close to the true one; as a result, the diffractions originally hidden by strong noises can be highlighted in the migrated domain (Li and Zhang, 2022a; Li and Zhang, 2022b). Here, we use F-K migration (Stolt, 1978), which is a direct method that is the fastest known migration technique.

We test on two simple models with strong random noises to illustrate the necessity of migration on improving signal-to-noise in the migration domain (Figure 3). We can see that the hyperbolic-shape diffractions are almost completely hidden in background noises before using migration. In contrast, the results after using migration (Stolt, 1978) show a clear peak in the wiggle along the apex of diffractions. This indicates that the weak diffractions can be extracted from the strong background noises using migration. In other words, the diffractions can be

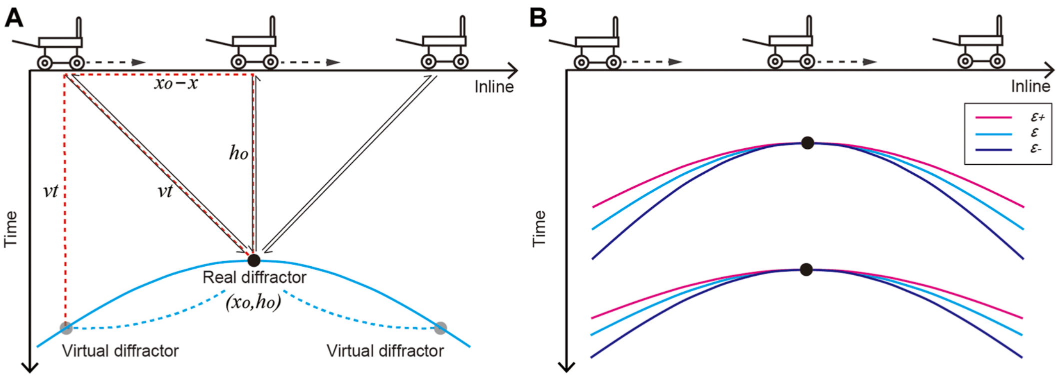


FIGURE 2
Illustration of the generation of diffractions and their isolations. **(A)** Mechanism of the radar observation and the forming principle of diffractions. The black lines with arrows indicate the propagation tracks of electromagnetic waves, the blue hyperbola represents the corresponding echo from the point-like diffractor in the radargram. **(B)** The center hyperbolas of isolating windows determined by different permittivities.

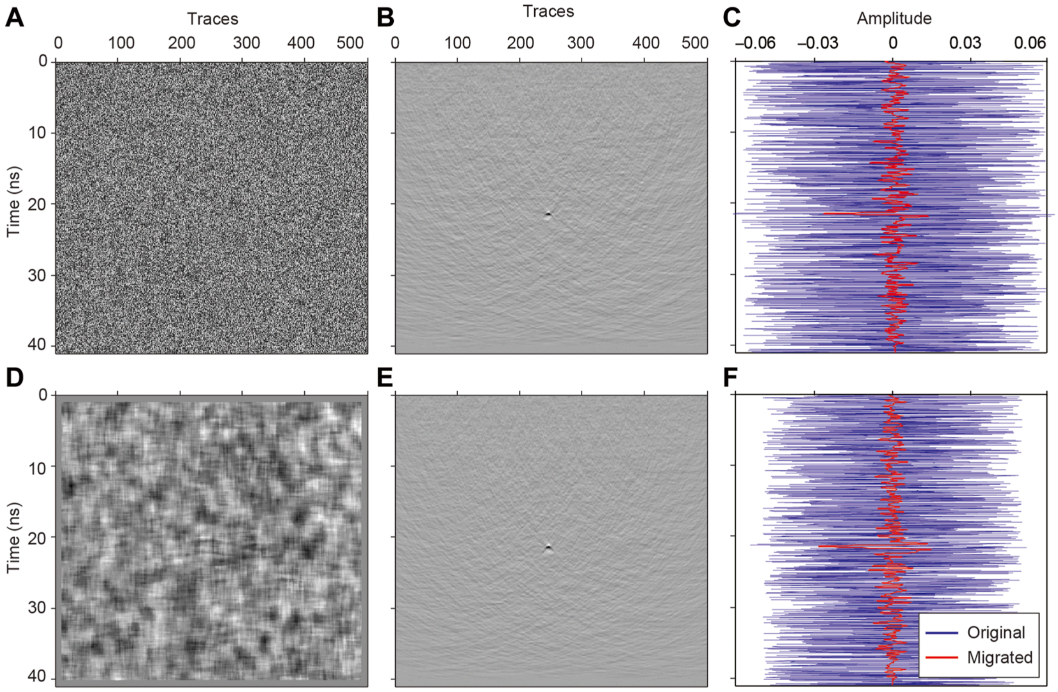


FIGURE 3
Illustration of improving the signal-to-noise ratio using migration. **(A)** Weak hyperbolic-shape diffractions buried in strong random noises. **(B)** The migration results of **(A)**. **(C)** Comparison of wiggles of the one trace along the apex of diffractions: the blue curve denotes the original data shown in **(A)** and the red curve denotes the migrated profile shown in **(B)**. **(D–F)** Another case similar to **(A–C)** but with a different kind of random noises shown in **(D)**.

highlighted after migration and the energy peak can well indicate the location of the diffractor. With migration, we can achieve the best velocity estimation after comparing the focusing effects under different velocities. However, the actual media is far more complex with a large number of scattering objects; additionally, their diffractions mostly gather together and interfere with each other, especially for the widely existing ejecta layer on the Moon and Mars. Thus, the velocity analyses after directly using migration could not well separate these interferences. We propose to generate a series of hyperbola windows within the possible range of permittivity to isolate each group of diffractions and suppress both random noises and the interferences. The velocity analyses are performed on these isolated data thus can well handle cluttered diffractions.

2.3 Diffraction isolation

The isolation of diffractions and the migration had been proven effective in reducing the above-surface diffractions (Li et al., 2021a). Traditional methods perform migration to the whole data set without isolating each group of diffractions; thus, the cluttered diffractions are still strongly interfered in the migrated domain (Carpentier et al., 2010). In contrast, we isolate the target diffractions from others in the un-migrated domain by some proper windows, which can largely avoid the influence from both adjacent diffractions and reflections, since the input data for the migration are dominated by the target diffractions and only small parts of adjacent diffractions or some segments of reflections exist. This can greatly improve the signal-to-noise ratio after applying migration to such windowed input data, compared with applying migration to the whole data; additionally, it can also improve the spatial resolution by separately analyzing each group of diffractions at a time.

According to the forming principle of diffractions, the isolation window can be automatically built up with the widen of the theoretical hyperbola. Therefore, we can build a series of hyperbola windows in different shapes at different depths using various permittivities (i.e., velocities). After analyzing some typical GPR data, we set the temporal width of the window to be 10–20 samples. These windows would be applied to isolate the diffractions from the GPR profile independently, and the surrounding noises and interferences can be well suppressed (Figure 4). In this way, the input data d can be divided into two parts: $d_{w(\varepsilon i)}$ is the isolated part within the hyperbolic window region, the shape of the hyperbolic window is under the control of permittivity ε_i, and the rest part is out of the window d_r = d—$d_{w(\varepsilon i)}$.

The main drawback of the proposed method is the huge computational cost, due to dealing with a vast number of potential diffractions independently. Fortunately, the total computational cost becomes affordable after applying a series of optimization steps. First, we apply grid searching on every 5 time steps, this can improve the efficiency greatly while preserving the shapes of diffractions; then, we apply migration within a small region instead of the

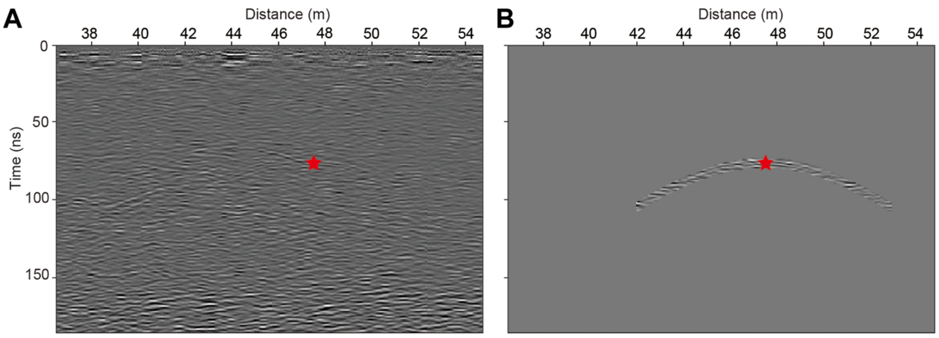


FIGURE 4
Illustration of diffraction isolation in Chang'E-4 LPR profile. **(A, B)** are the radar profile before and after diffraction isolation, respectively.

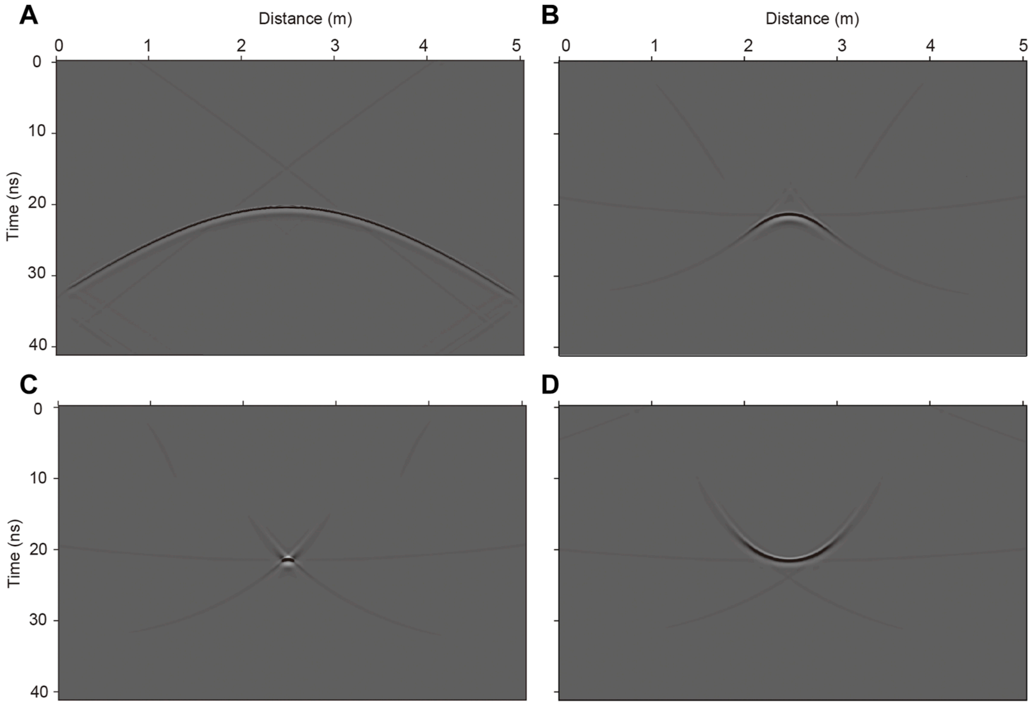


FIGURE 5
Migration results under different permittivities. **(A)** Hyperbolic-shape diffraction. **(B–D)** The corresponding migration results with the permittivity of ε+, ε, and ε-, respectively.

whole profile, which can further reduce the computational cost; finally, we adopt a fast implementation of migration (e.g., Stolt, 1978).

2.4 Focusing analyses in migrated domain

For each group of diffractions, permittivity is the key parameter controlling the focusing effects of migration results. If the permittivity is true, the diffractions can be perfectly focused within a small region close to the true location of the diffractor; if too large, the diffractions will be under-migrated and display as a downward-opening curve with reduced spatial scale; otherwise, if too small, the diffractions will be over-migrated and display as an upward-opening curve (Figure 5) (Claerbout, 1985; Zhu et al., 1998; Li et al., 2021b). For a given group of diffractions, there must be a permittivity model with which the diffractions can be best tailored by a hyperbola window, and the migration results would be highly focused with the highest signal-to-noise ratio. In other words, for a properly-tailored group of diffractions, the permittivity can be determined once we obtain the best-focused migration results. This allows us to finely tune the velocity for each potential diffractor independently.

We assume that one group of diffractions is already tailored properly with a group of given windows determined by the permittivity εi, and A is the migrated result as below

$$A = FKMig(d_{w,\varepsilon i}). \tag{4}$$

Then we calculate the partial derivatives H_t and H_x of A in time and horizontal direction,

$$[H_t, H_x] = gradient(A). \tag{5}$$

We then use the sum of H_t and H_x to represent the amplitude variation at point (t_0, x_0) in A, which can be calculated as

$$H(t_0, x_0) = |H_t(t_0, x_0)| + |H_x(t_0, x_0)|. \tag{6}$$

To further quantify the amplitude variation (i.e., the focusing effects around this point under a specific permittivity), we design a gradient template around the target. We bring in a parameter P to express the focusing energy around the target:

$$P = \sum_{t=t_0-u}^{t_0+u} \sum_{x=x_0-v}^{x_0+v} H(t,x), \tag{7}$$

where u and v ($u > 1, v > 1$) represent the side lengths of the template in time and horizontal direction, respectively. Usually, P would be greatly higher for diffractors, compared with non-diffractors. When P exceeds a certain threshold, we consider that there may be diffractors there, which require further focusing analysis. Then we introduce A,B,C,D to represent the focusing energy around the point (t_0, x_0) (Figure 6),

$$A = \sum_{t=t_0-2u}^{t_0-u-1} \sum_{x=x_0-v}^{x_0+v} H(t,x), \tag{8}$$

$$B = \sum_{t=t_0+u+1}^{t_0+2u} \sum_{x=x_0-v}^{x_0+v} H(t,x), \tag{9}$$

$$C = \sum_{t=t_0-u}^{t_0+u} \sum_{x=x_0-2v}^{x_0-v-1} H(t,x), \tag{10}$$

$$D = \sum_{t=t_0-u}^{t_0+u} \sum_{x=x_0+v+1}^{x_0+2v} H(t,x), \tag{11}$$

In different radar profiles with different time step and spatial interval, we need to set different u and v. Figure 6 shows the template under $u = 4$, $v = 2$. For high-frequency radar profile, u is usually greater than 10 and there will be a multiple difference between u and v. Therefore, we can ignore the effects of signal energy within the gray area in Figure 6 during focus evaluation.

Then we introduce two parameters R_1 and R_2:

$$R_1 = \frac{P}{P+A+B+C+D}, \tag{12}$$

which is the gradient ratio between the amplitudes of target area (blue region in Figure 6) and all points within the template.

$$R_2 = \frac{P}{A+B+C+D}, \tag{13}$$

which is the gradient ratio between the amplitudes of target area and area around the target (pink, yellow, green and purple region in Figure 6). These two ratios (i.e., R_1 and R_2) can well represent the focusing effects in the migrated domain.

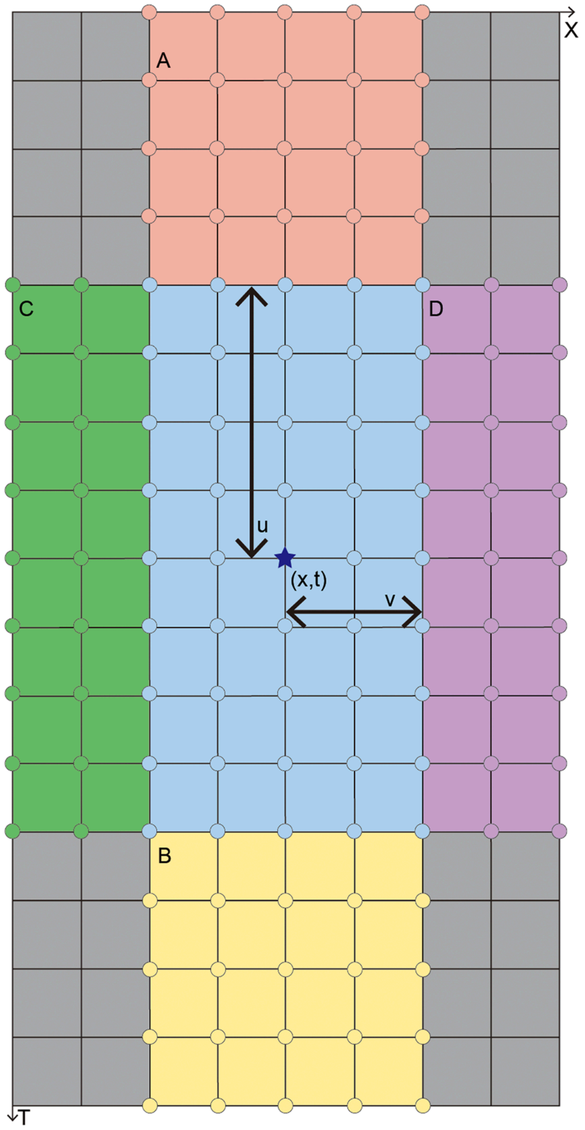


FIGURE 6
The template of energy distribution around the target (marked with the star) in migrated domain. P is the signal energy within the blue area; **(A–D)** represent the surrounding signal energy, drawn in pink, yellow, green and purple respectively.

Here we show the R1 and R2 (Figure 7) calculated from Figures 5B–D. It is obvious that the focusing energy is the strongest under an accurate velocity model.

We argue that if there is a diffractor, there must be two maximums of R1 and R2 after comparing a series of these two ratios obtained under different permittivities. In other words, when R1 and R2 reach their maximums under the same permittivity simultaneously, this permittivity can be regarded as the best estimation; otherwise, we suppose that there is no diffractor at current point. In this way, we can locate the diffractor and estimate the corresponding permittivity simultaneously.

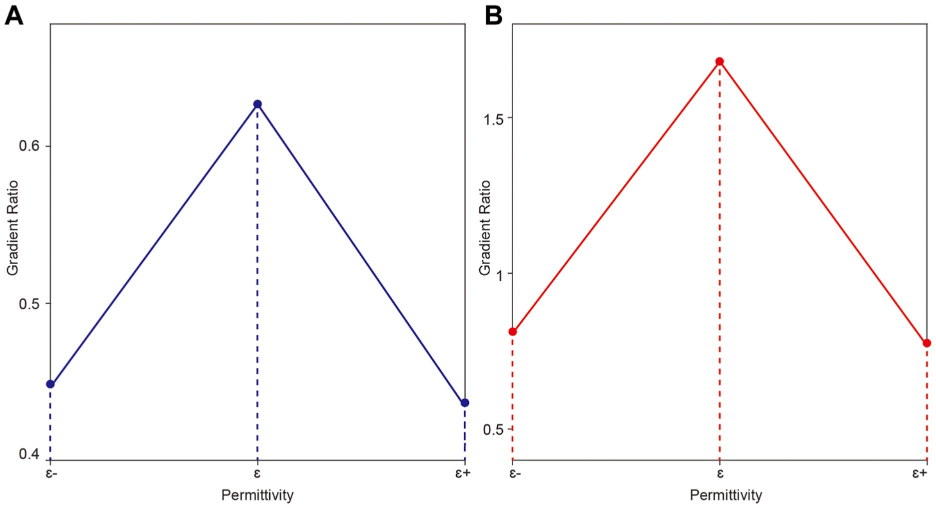


FIGURE 7
Histograms of R1[blue line in 7**(A)**] and R2[red line in 7**(B)**]. ε-, ε and ε+ corresponds to the over-migrated, perfectly migrated and under-migrated situation, respectively.

3 Application in lunar penetrating radar data

We hope to employ our method to a local lunar penetrating radar (LPR) profile from Chang’E-4 mission (Zhang J et al., 2021), so that we can acquire the corresponding permittivity profiles at the landing site. This can help us to better understand the structure and property of lunar subsurface material.

The LPR onboard Yutu-2 rover has two channels, with the dominant frequency of 60 MHz (Channel-1) and 500 MHz (Channel-2), respectively. We only consider the high-frequency channel, which is for detecting shallow structures. Under the dominant frequency of 500 Mhz, the diffractors with diameter over ~15 cm would be mostly identified (Jol., 2009). The Channel-2 antennas are 30 cm over the surface, in Bowtie shape. The two receiving antennas (A and B) are close to the emitter with a distance of 15.4 cm and 31.7 cm, respectively (Fang et al., 2014). The time step of channel 2B is 0.3125 ns and spatial interval is 3.65 cm.

We first test our method with a numerical model to show its feasibility. We build a simple scattering model and perform numerical simulation with the finite-difference method (Figures 8A, B). Based on the simulated radar records, we further carry out the F-K migration and focusing analysis in the migrated domain. The scanning region (the box in Figure 8B) is between 13 ns and 17.7 ns in time, from $x = 0.5$ m to $x = 4.25$ m. Figure 8C shows that our proposed method can well locate the diffractors and derive the corresponding permittivities from the radar record.

3.1 Application in a typical group of diffractions

After confirming the feasibility of this method, we then employ it to a typical group of diffractions from the LPR profile and compare the focusing effects in the migrated domain before and after diffraction isolation. The location of the group of diffractions is marked with the red star in Figure 9.

After nine rounds of diffraction isolation and migration, we can obtain nine groups of gradient ratios R1 and R2 (Figures 10A, C). It can be seen that R1 and R2 reach their maximums simultaneously with the permittivity of 4. To be more precise, we refine the permittivity from 3.5 to 4.5 by a one-fifth interval and find that the largest ratios appear at permittivity of 3.7 (Figures 10B, D). Therefore, the best estimation of the permittivity at this point should be 3.7. Under this permittivity, we compare the focusing effects in the migrated domain with or without diffraction isolation (Figure 11). Obviously, the signal energy is better focused after using diffraction isolation, although the peak is slightly reduced compared with that before using diffraction isolation. This indicates that the diffraction isolation can well suppress the interference from noises and surrounding diffractions, which can also explain why the energy peak falls slightly. Besides, the location of the diffractor can be more accurate after diffraction isolation, as shown by the wiggles in Figure 11D.

3.2 Application in local profile

We further apply our method to the local LPR profile. The location of the scanning region is indicated with the red dotted box in Figure 9. The original LPR data have a time step of 0.3125 ns and a spatial interval of 3.65 cm. We pick up the data every 5 time steps from traces No. 1150 to No. 1350 to save the computational cost. Comparing the gradient ratios of R1 and R2, we can pick up the diffractors while recording their corresponding permittivities. Ultimately, we can acquire a two-dimensional map of permittivity distribution using the proposed method (Figure 12).

The depth of the region (red box in Figure 9) is within 20 m, the permittivity is generally between 3 and 5 (Li et al., 2021b).

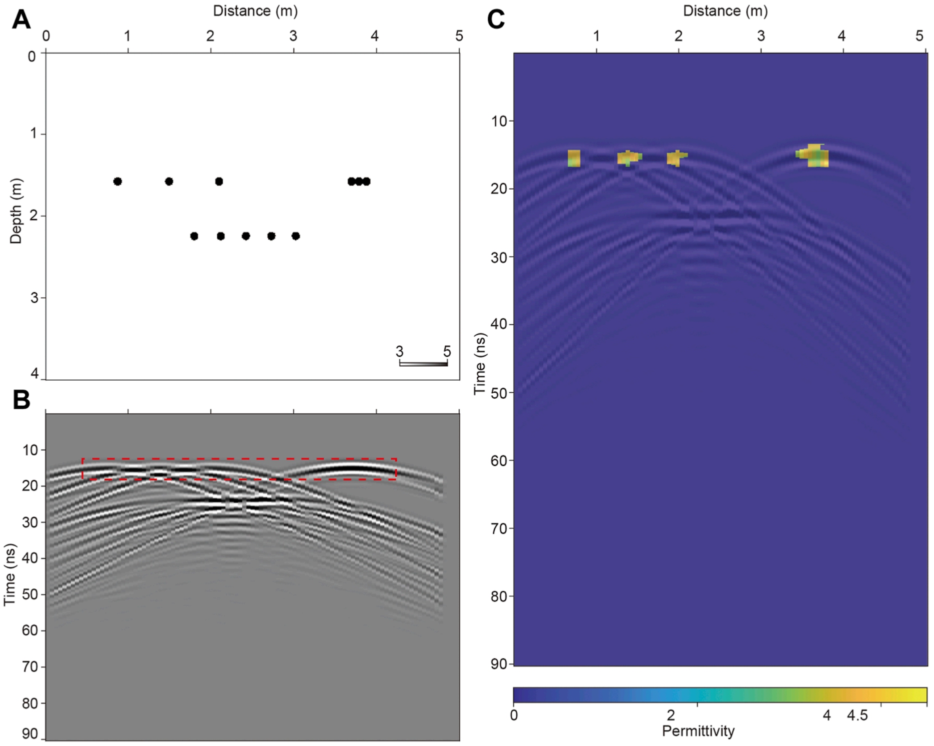


FIGURE 8
Test on a scattering model. **(A)** Numerical permittivity model. **(B)** The corresponding simulation result. **(C)** The result of velocity estimation.

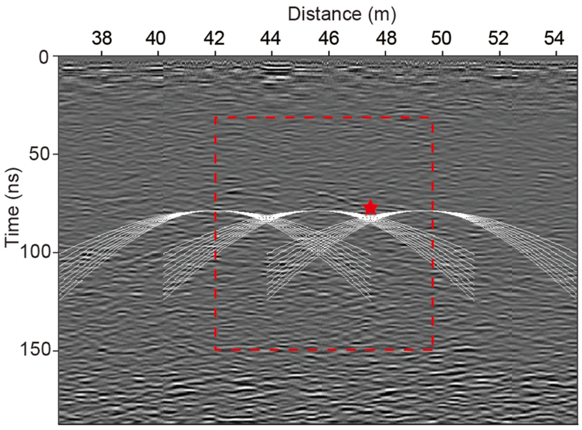


FIGURE 9
Illustration of diffraction isolation. The base map is the local LPR profile obtained by the Yutu-2 lunar rover (Zhang X et al., 2021). The star marks the location of typical diffractions used in 3.2.1. The box indicates the scanning region in 3.2.2. The hyperbolas illustrate nine rounds of diffraction isolation at three coordinate points using different permittivities.

Therefore, the permittivity beyond the range [3,5] would be absurd. We eliminate the absurd values of permittivity to avoid potential influence from few wild points (Figure 12A). Then we interpolate the permittivity profile to generate continuous velocity estimation along the whole LPR profile then apply two rounds of median filtering to smooth the interpolated results. It should be noted that the permittivity calculated above represents the depth-averaged permittivity between the surface and the diffractor, instead of the interval permittivity at a specific depth. Thus, we need to transfer it into the interval velocity with Dix inversion as follows (Dix, 1955; Bradford and Harper, 2005; Sato and Feng, 2005)

$$v_{int,t_n} = \sqrt{\frac{{v_{ave,t_n}}^2 t_n - {v_{ave,t_{n-1}}}^2 t_{n-1}}{t_n - t_{n-1}}}, \tag{14}$$

where t_n represents the time samples, $v_{int'tn}$ represents the interval velocity at the time point of t_n, and $v_{ave,tn}$ represents the average velocity within time t_n, which is the average of the first n temporal pixels. Thus, the interval permittivity can be calculated point by point. Finally, the two-dimensional map of permittivity distribution can be obtained, as shown in Figure 12D. The spatial resolution can be roughly up to 30 cm, comparable to the spatial resolution of the LPR (~30 cm). From Figures 12A, D clear increasing tendency of permittivity *versus* time can be seen, revealed as an expanded yellow region.

Based on this local map of permittivity distribution, we can gather more information on dielectric property by analyzing the statistical characteristics of lunar regolith. Figure 13A shows the relative permittivity scatterplot derived from the map. The permittivities vary in a similar tendency between different traces, which indicates that there is sight lateral variation along the LPR profile. Additionally, the permittivities on all points are generally ~4 from 30 to 150 ns. The linear fitting results show a tiny increasing

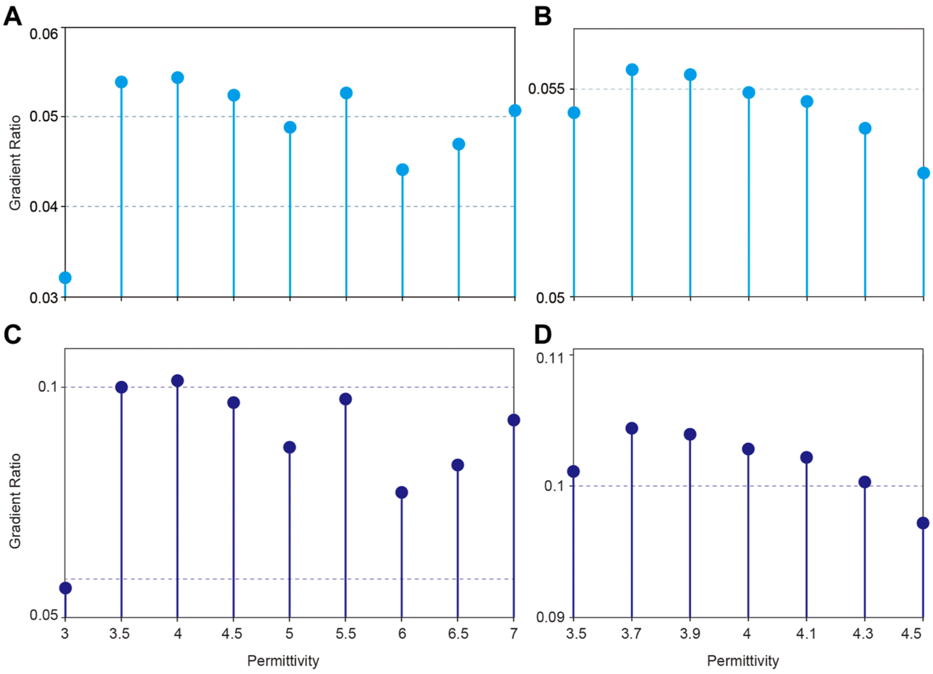


FIGURE 10
The gradient ratios histograms in typical diffraction tests. **(A)** and **(C)** list R1 and R2 calculated using different permittivities ranging from 3 to 7, respectively. **(B, D)** are the locally refined results between 3.5 and 4.5 of **(A, C)**, respectively.

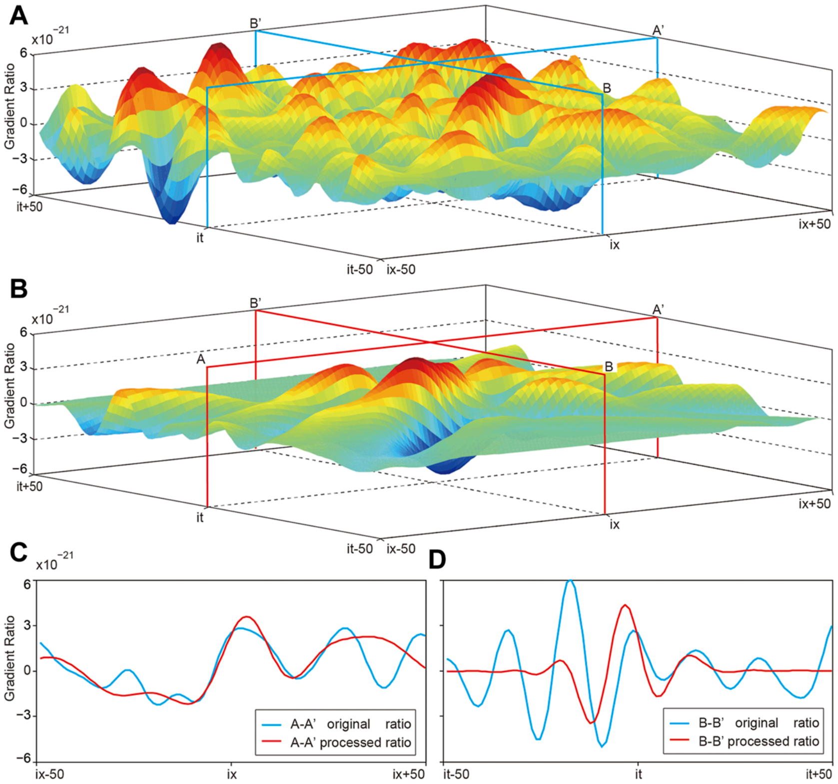


FIGURE 11
Local energy maps and slices of migration. **(A, B)** are energy maps with or without diffraction isolation, respectively. (*it, ix*) is the coordinate of the diffractor. **(C)** The energy slices along A-A′ shown in **(A, B)**. **(D)** The energy slices along B-B′ shown in **(A, B)**.

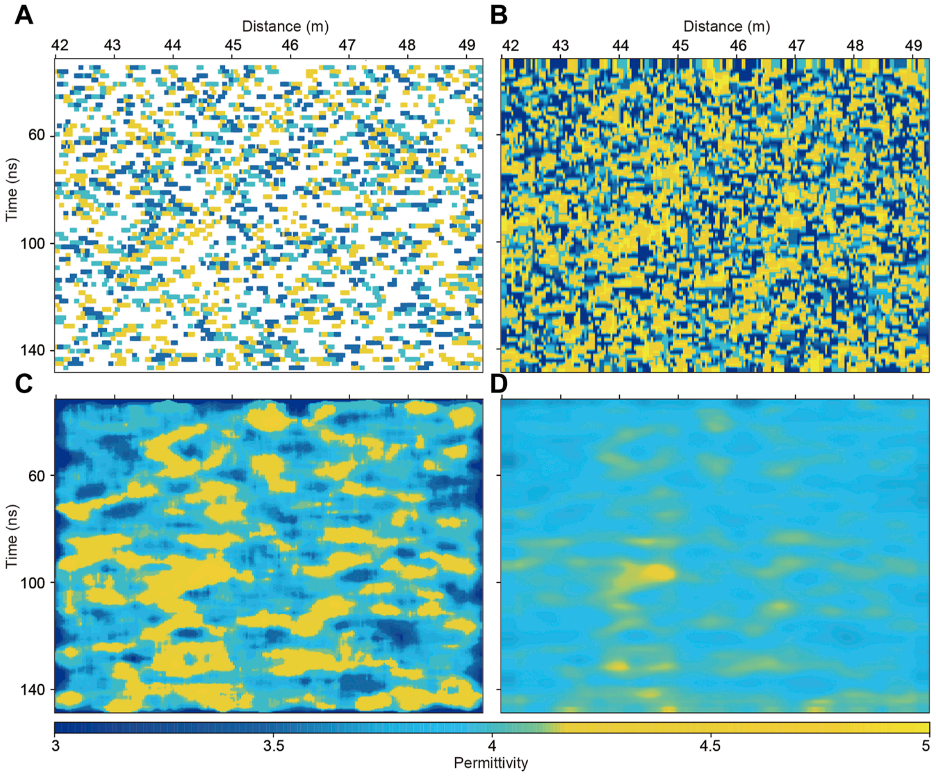


FIGURE 12
Permittivity map obtained from scattering points. **(A)** The scatter plot of average permittivity. **(B)**The interpolation result of scattering points shown in **(A)**. **(C)** The median filtering result of **(B)**. **(D)** The smoothed map of interval permittivity converted from **(C)** by Dix inversion.

linear trend from 3.85 to 4.1 between 30 and 150 ns, depicted with the red line in Figure 13A. This indicates that there is almost no rapid variation of permittivity with depth. Furthermore, we calculate the mean and median of permittivity at every moment; after the reservation of only one decimal, we calculate the mode average. As shown in Figure 13B, the statistical curves reemphasize the increasing trend of permittivity *versus* time, and the upward tendency indicted by the mean and median permittivity are basically the same, but the curve of mean value changes more gently. It is also worth noticing that with the increasing of depth, the permittivity varies more violently. This phenomenon indicates that the heterogeneity of lunar regolith would be more obvious in deeper materials, which is generally consistent with the gardening process of shallow lunar materials (Zhang, J, et al., 2021).

Another statistics and error analysis is carried out by averaging the estimated permittivities every 5 time steps, as illustrated in Figure 14A. The time points are marked with red dots, and the one standard deviation (1σ) errors within each subsection are identified with dark bars, indicting the confidence bound of our results. For instance, the average permittivity at time point t_0 is calculated with

$$\varepsilon_{r,t_0} = \frac{\sum_{i=t_0-4}^{t_0+4} \varepsilon_{r,i}}{9},\ t_0 = 5, 10, 15\ldots \tag{15}$$

This graph provides a new sight for analyzing the permittivity distribution. Averaging the permittivity helps to limit the influence of some singular permittivity values while retaining the local characteristics on the overall trend, which is beneficial to summarize the relationship between permittivity and travel time.

With the estimated permittivity at a certain travel time, we can extract a relationship by curve fitting as

$$\varepsilon_r = \frac{4.1t + 16.41}{t + 6.884}, \tag{16}$$

where t is the travel time (ns), ε_r is the relative permittivity. Consistent with previous studies, the statistical regulation shows an increasing trend of permittivity *versus* depth, indicating that the relative permittivity of lunar subsurface materials is in strong positive correlations with depth. The red curve in Figure 14B represents the averaged permittivity at every moment, and the blue curve is the fitted relationship. With these fitting results, we can analyze the degree of permittivity variation within this region. The peaks and valleys at time point t_j are collected and the root mean square local permittivity perturbation ξ can be calculated with

$$\xi = \sum_{j=1}^{N} \frac{\left|\varepsilon_{r,t_j} - \varepsilon_{fit,t_j}\right|}{N \bullet \varepsilon_{fit,t_j}}, \tag{17}$$

where N is the number of peaks and valleys, ε_{r,t_j} and ε_{fit,t_j} is the averaged relative permittivity and the fitted one, respectively. The local permittivity perturbation ξ is estimated to be ~2.3%, which is consistent with 3% obtained by the simulated results (Zhang, X., et al., 2021). This verified that the lunar regolith at the Chang'E-4 landing area can be well described by self-organization random models.

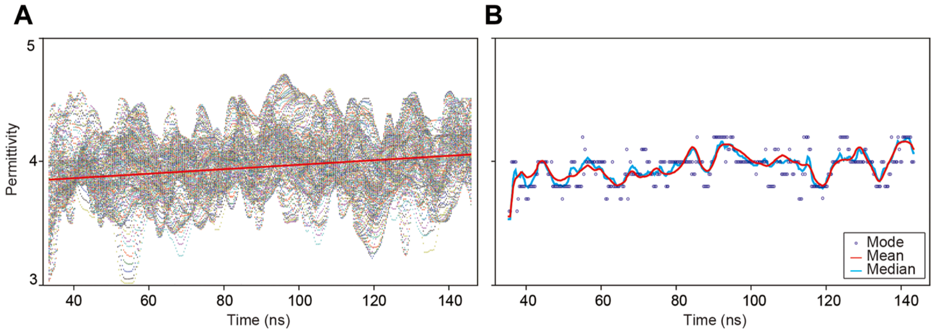


FIGURE 13
Statistical characteristics of permittivity distribution. **(A)** The scatterplot of permittivity, where the red line is the linear fitting result. **(B)** The mode, mean and median permittivities at every moment.

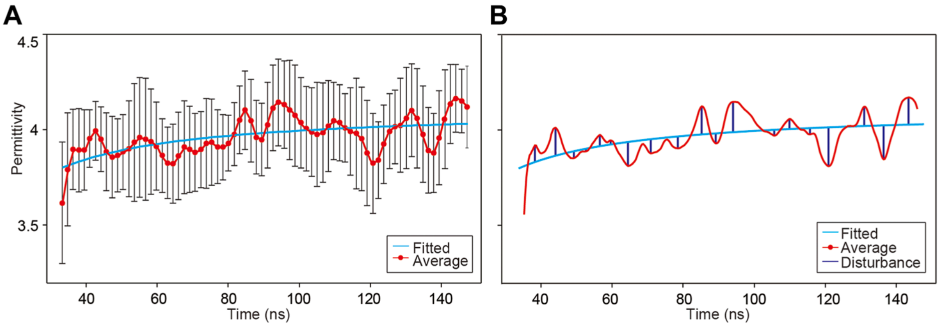


FIGURE 14
Permittivity distribution along time. **(A)** The red line with dots and the black bars represent the average permittivity and one standard deviation every 5 time steps, respectively. **(B)** The red line is the average permittivity at every moment, and the blue curve represents the fitting result. The vertical bars represent the perturbation of peaks and valleys.

4 Discussion

Compared with the existing methods, our method has some unique advantages. The hyperbolic-fitting method requires that the diffractions should be in well-formed and distinguishable hyperbolic shape, which is difficult to achieve in practice; additionally, the traditional migration method can not suppress the surrounding noises and interferences in the focusing analysis due to working on the whole profile, which may lead to large errors. Our method combines the advantage of hyperbolic fitting and migration; thus, it can help to estimate the relative permittivity more precisely and effectively, especially for complex materials. Besides, the spatial resolution of our method is to that of the LPR payload, which can help to better recognize and delineate the abnormal bodies in subsurface materials.

Limited by the computation cost, our method is only applied in shallow part within 12 m in depth and within ~7 m in distance; however, the statistical characteristics extracted from this local area should be universal to the Chang'E-4 landing area along the rover's track. The map of permittivity distribution shows that there is an abnormal high-permittivity region existing in the map at ~100 ns? We suppose that small basaltic or ejecta blocks may gather there. Besides, we can notice that the permittivity becomes larger and varies severer with the increasing depth. This indicates that there is a strong positive correlation between the permittivity of lunar subsurface materials and depth, and the heterogeneity of lunar regolith is more obvious in deeper materials. The increasing tendency of permittivity with depth is consistent with former studies (Heiken et al., 1991). We suppose this is probably because of more blocks or larger grain size of deeper materials.

5 Conclusion

We propose a migration-based method to estimate the relative permittivity of subsurface materials from the ground penetrating radar data, including three steps: diffraction isolation, migration, and focusing analysis. Single point test shows that diffraction isolation can help to improve the reliability of migration and is a prerequisite for accurate focusing analysis in the migrated domain. Numerical experiments show that the proposed method is robust and effective for permittivity estimation, especially in the presence of strong noises and cluttered diffractions. The spatial resolution is higher than that of the existing methods due to the application of diffraction isolation and migration; thus, we can pick up a great number of diffractors, which is helpful to construct a continuous velocity profile in depth. The permittivity at the Chang'E-4 landing area is estimated to be ~4 within 12 m, ranging from 3.5 to 4.2. Additionally, the local perturbation of the permittivity is ~2.3%, which is consistent with that (~3%) obtained by numerical simulations using self-organization random models. This suggests that the lunar regolith at Chang'E-4 landing area is mature and can be well described by the self-organization random models.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://moon.bao.ac.cn/ce5web / searchOrder_dataSearchData.search.

Author contributions

Conceptualization: WL and JZ; Methodology: WL; Supervision: JZ; Writing—Original draft: WL and JZ; Writing—Review and editing: WL and JZ. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the National Natural Science Foundation of China (41941002), Key Research Program of the Institute of Geology and Geophysics, CAS (IGGCAS-202203), and Key Research Program of the Chinese Academy of Sciences (ZDBS-SSWTLC001). JZ was also supported by Foundation for Excellent Member of the Youth Innovation Promotion Association, Chinese Academy of Sciences (2016).

Acknowledgments

We thank the Supercomputing Laboratory of Institute of Geology and Geophysics, Chinese Academy of Sciences (IGGCAS) for providing computing resources.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Boisson, J., Heggy, E., Clifford, S. M., Frigeri, A., Plaut, J. J., Farrell, W. M., et al. (2009). Sounding the subsurface of Athabasca Valles using MARSIS radar data: Exploring the volcanic and fluvial hypotheses for the origin of the rafted plate terrain. *J. Geophys. Res. Planets* 114 (E8). doi:10.1029/2008JE003299

Bradford, J. H., and Harper, J. T. (2005). Wave field migration as a tool for estimating spatially continuous radar velocity and water content in glaciers. *Geophys. Res. Lett.* 32 (8), L08502. doi:10.1029/2004GL021770

Capineri, L., Daniels, D. J., Falorni, P., Lopera, O. L., and Windsor, C. G. (2008). Estimation of relative permittivity of shallow soils by using the ground penetrating radar response from different buried targets. *Prog. Electromagn. Res. Lett.* 2, 63–71. doi:10.2528/PIERL07122803

Carpentier, S. F., Horstmeyer, H., Green, A. G., Doetsch, J., and Coscia, I. (2010). Semiauto-mated suppression of above-surface diffractions in GPR data. *Geophysics* 75 (6), J43–J50. doi:10.1190/1.3497360

Carrier, W. D., Olhoeft, G. R., and Mendell, W. (1991). Physical properties of the lunar surface. *Lunar Sourceb.*, 475–594.

Claerbout, J. F. (1985). *Imaging the earth's interior*. Oxford: Blackwell scientific publications.

Dai, S., Su, Y., Yuan, X., Feng, J., and Xing, S. (2014). *Lunar regolith structure model and echo simulation for Lunar Penetrating Radar*. IEEE, Proceedings of the 15th International Conference on Ground Penetrating Radar. July 4, 2014, Belgium, IEEE.

Ding, C., Li, C., Xiao, Z., Su, Y., Xing, S., Wang, Y., et al. (2020). Layering structures in the porous material beneath the Chang'e-3 landing site. *Earth Space Sci.* 7, e2019EA000862. doi:10.1029/2019EA000862

Dix, C. H. (1955). Seismic velocities from surface measurements. *Geophysics* 20 (1), 68–86. doi:10.1190/1.1438126

Dong, Z., Fang, G., Zhao, D., Zhou, B., Gao, Y., and Ji, Y. (2020). Dielectric properties of lunar subsurface materials. *Geophys. Res. Lett.* 47 (22), e2020GL089264. doi:10.1029/2020GL089264

Economou, N., Vafidis, A., Bano, M., Hamdan, H., and Ortega-Ramirez, J. (2020). Ground-penetrating radar data diffraction focusing without a velocity model. *Geophysics* 85 (3), H13–H24. doi:10.1190/geo2019-0101.1

Fa, W., Zhu, M., Liu, T., and Plescia, J. (2015). Regolith stratigraphy at the Chang'E-3 landing site as seen by lunar penetrating radar. *Geophys. Res. Lett.* 42 (23), 10179–10187. doi:10.1002/2015gl066537

Fang, G., Zhou, B., Ji, Y., Zhang, Q., Shen, S., Li, Y., et al. (2014). Lunar penetrating radar onboard the change-3 mission. *Res. astronomy astrophysics* 14 (12), 1607–1622. doi:10.1088/1674-4527/14/12/009

Feng, J., Su, Y., Li, C., Dai, S., Xing, S., and Xiao, Y. (2019). An imaging method for Change5 Lunar Regolith Penetrating Radar. *Planet. Space Sci.* 167 (MAR.), 9–16. doi:10.1016/j.pss.2019.01.008

Fomel, S., Landa, E., and Taner, M. T. (2007). Poststack velocity analysis by separation and imaging of seismic diffractions. *Geophysics* 72 (6), U89–U94. doi:10.1190/1.2781533

Fomel, S. (2003a). Time-migration velocity analysis by velocity continuation. *Geophysics* 68 (5), 1662–1672. doi:10.1190/1.1620640

Fomel, S. (2003b). Velocity continuation and the anatomy of residual prestack time migration. *Geophysics* 68 (5), 1650–1661. doi:10.1190/1.1620639

Giannakis, I., Zhou, F., Warren, C., and Giannopoulos, A. (2021). Inferring the shallow layered structure at the change-4 landing site: A novel interpretation approach using lunar penetrating radar. *Geophys. Res. Lett* doi:10.1002/essoar.10506249.1

Gold, T., Bilson, E., and Yerbury, M. (1972). Grain size analysis, optical reflectivity measurements, and determination of high-frequency electrical properties for Apollo 14 lunar samples. *Lunar Planet. Sci. Conf. Proc.*

Hamran, S. E., Paige, D. A., Amundsen, H., Berger, T., Yan, M. J., Carter, L., et al. (2020). Radar imager for Mars' subsurface experiment—rimfax. *Space Sci. Rev.* 216 (128), 128. doi:10.1007/s11214-020-00740-4

Heiken, G. H., Vaniman, D. T., and French, B. M. (1991). *Lunar Sourcebook, a user's guide to the Moon*. Houston, Texas. Cambridge University Press.

Hiesinger, H., and Head, J. W., III (2006). New views of lunar geoscience: An introduction and overview. *Rev. mineralogy Geochem.* 60 (1), 1–81. doi:10.2138/rmg.2006.60.1

Hu, J., Schuster, G. T., and Valasek, P. A. (2001). Poststack migration deconvolution. *Geophysics* 66 (3), 939–952. doi:10.1190/1.1444984

Jol, H. (2009). *Ground penetrating radar: Theory and applications*. Germany: Elsevier Science.

Jolliff, B. L., Wieczorek, M. A., Shearer, C. K., and Neal, C. R. (2006). *New views of the Moon*. Beijing: Walter de Gruyter GmbH and Co KG. doi:10.2138/rmg.2006.60.0

Jordan, R., Picardi, G., Plaut, J., Wheeler, K., Kirchner, D., Safaeinili, A., et al. (2009). The Mars express MARSIS sounder instrument. *Planet. Space Sci.* 57 (14–15), 1975–1986. doi:10.1016/j.pss.2009.09.016

Kouemou, G. (2010). *Radar technology*.Germany BoD–Books on Demand.

Leong, Z. X., and Zhu, T. (2021). Direct velocity inversion of ground penetrating radar data using GPRNet. *J. Geophys. Res. Solid Earth* 126, e2020JB021047. doi:10.1029/2020JB021047

Leuschen, C. J., and Plumb, R. G. (2001). A matched-filter-based reverse-time migration algorithm for ground-penetrating radar data. *IEEE Trans. Geoscience Remote Sens.* 39 (5), 929–936. doi:10.1109/36.921410

Li, C., Lin, Y., Lv, W., and Zhang, J. (2021b). Eliminating above-surface diffractions from ground-penetrating radar data using iterative Stolt migration. *Geophysics* 86 (1), H1–H11. H1–H11. doi:10.1190/GEO2019-0796.1

Li, C., Zhang, J., Loxas, M., and Sharma, P. (2021a). Hemosiderotic dermatofibroma mimicking melanoma: A case report and review of the literature. *Remote Sens.* 13, 1387–1392. doi:10.3390/rs13071387

Li, C., and Zhang, J. (2022a). Preserving signal during random noise attenuation through migration enhancement and local orthogonalization. *Geophysics* 87, V451–V466. doi:10.1190/geo2021-0385.1

Li, C., and Zhang, J. (2022b). Wavefield separation using irreversible-migration filtering. *Geophysics* 87 (3), A43–A48. **A43–A48**. doi:10.1190/GEO2021-0607.1

Li, C., Zheng, Y., Wang, X., Zhang, J., Wang, Y., Chen, L., et al. (2022). Layered subsurface in utopia basin of Mars revealed by zhurong rover radar. *Nature* 610, 308–312. doi:10.1038/s41586-022-05147-5

Li, J., Zeng, Z., Liu, C., Huai, N., and Wang, K. (2017). A study on lunar regolith quantitative random model and lunar penetrating radar parameter inversion. *IEEE Geoscience Remote Sens. Lett.* 14 (11), 1953–1957. doi:10.1109/LGRS.-2017.2743618

Lv, W., Li, C., Song, H., Zhang, J., and Lin, Y. (2020). Comparative analysis of reflection characteristics of lunar penetrating radar data using numerical simulations. *Icarus* 350, 113896. doi:10.1016/j.icarus.2020.113896

Olhoeft, G., Frisillo, A., Strangway, D., and Sharpe, H. (1974). Temperature dependence of electrical conductivity and lunar temperatures. *moon* 9 (1-2), 79–87. doi:10.1007/bf00565394

Olhoeft, G. R., and Strangway, D. W. (1975b). Dielectric properties of the first 100 meters of the Moon. *Earth Planet. Sci. Lett.* 24, 394–404. doi:10.1016/0012821x(75)90146-6

Olhoeft, G., Strangway, D., and Pearce, G. (1975a). Effects of water on electrical properties of lunar fines. *Proc. Lunar Sci. Conf. 6th*, 3333–3342.

Oudart, N., Ciarletti, V., Le Gall, A., Mastrogiuseppe, M., Hervé, Y., Benedix, W. S., et al. (2021). Range resolution enhancement of WISDOM/ExoMars radar soundings by the Bandwidth Extrapolation technique: Validation and application to field campaign measurements. *Planet. Space Sci.* 197, 105173. doi:10.1016/j.pss.2021.105173

Özdemir, C., Demirci, Ş., Yiğit, E., and Yilmaz, B. (2014). A review on migration methods in B-scan ground penetrating radar imaging. *Math. Problems Eng.* 2014, 1–16. doi:10.1155/2014/280738

Reshef, M., and Landa, E. (2009). Post-stack velocity analysis in the dip-angle domain using diffractions. *Geophys. Prospect.* 57 (5), 811–821. doi:10.1111/j.1365-2478.2008.00773.x

Ristic, A. V., Petrovacki, D., and Govedarica, M. (2009). A new method to simultaneously estimate the radius of a cylindrical object and the wave propagation velocity from GPR data. *ComputersandGeosciences* 35 (8), 1620–1630. doi:10.1016/j.cageo.2009.01.003

Rochette, P., Gattacceca, J., Ivanov, A., Nazarov, M., and Bezaeva, N. (2010). Magnetic properties of lunar materials: Meteorites, Luna and Apollo returned samples. *Earth Planet. Sci. Lett.* 292 (3–4), 383–391. doi:10.1016/j.epsl.2010.02.007

Sato, M., and Feng, X. (2005). *GPR migration algorithm for landmines buried in inhomogeneous soil. IEEE Antennas and Propagation Society International Symposium.*

Sava, P. C., Biondi, B., and Etgen, J. (2005). Wave-equation migration velocity analysis by focusing diffractions and reflections. *Geophysics*, 70(3), U19–U27. doi:10.1190/1.1925749

Schroeder, D. M., Dowdeswell, J. A., Siegert, M. J., Bingham, R. G., Chu, W., MacKie, E. J., et al. (2019). Multidecadal observations of the Antarctic ice sheet from restored analog radar records. *Proc. Natl. Acad. Sci.* 116 (38), 18867–18873. doi:10.1073/pnas.1821646116

Soldovieri, F., and Solimene, R. (2010). Ground penetrating radar subsurface imaging of buried objects. *Tech.*

Song, H., Li, C., Zhang, J., Wu, X., Liu, Y., and Zou, Y. (2021). Rock location and property analysis of lunar regolith at Chang'E-4 landing site based on local correlation and semblance analysis. *Remote Sens.* 13 (1), 48. doi:10.3390/rs13010048

Stolt, R. H. (1978). Migration by fourier transform. *Geophysics* 43 (1), 23–48. doi:10.1190/1.1440826

Stuart, G. (2003). Characterization of englacial channels by ground-penetrating radar: An example from austre Brøggerbreen, Svalbard. *J. Geophys. Res. Solid Earth* 108 (B11), 2525. doi:10.1029/2003JB002435

Su, Y., Fang, G. Y., Feng, J. Q., Xing, S. G., Ji, Y. C., Zhou, B., et al. (2014). Data processing and initial results of Chang'e-3 lunar penetrating radar. *Res. Astronomy Astrophysics* 14 (12), 1623–1632. doi:10.1088/1674–4527/14/12/010

Xiao, L., Zhu, P., Fang, G., Xiao, Z., Zou, Y., Zhao, J., et al. (2015). A young multilayered terrane of the northern Mare Imbrium revealed by Chang'E-3 mission. *Science* 347 (6227), 1226–1229. doi:10.1126/science.1259866

Yilmaz, Ö. (2001). *Seismic data analysis: Processing, inversion, and interpretation of seismic data*. fm: Society of exploration geophysicists. doi:10.1190/1.9781560801580

Yuan, H., Montazeri, M., Looms, M. C., and Nielsen, L. (2019). Diffraction imaging of ground-penetrating radar data. *Geophysics* 84 (3), H1–H12. doi:10.1190/geo2018-0269.1

Zhang, H., Zheng, L., Su, Y., Fang, G., Zhou, B., Feng, J., et al. (2014). Performance evaluation of lunar penetrating radar onboard the rover of CE-3 probe based on results from ground experiments. *Res. Astronomy Astrophysics* 14 (12), 1633–1641. doi:10.1088/1674–4527/14/12/011

Zhang, J., Yang, W., Hu, S., Lin, Y., Fang, G., Li, C., et al. (2015). Volcanic history of the imbrium basin: A close-up view from the lunar rover Yutu. *Proc. Natl. Acad. Sci.* 112 (17), 5342–5347. doi:10.1073/pnas.1503082112

Zhang, J., Zhou, B., Lin, Y., Zhu, M., Song, H., Dong, Z., et al. (2021). Lunar regolith and substructure at Chang'E-4 landing site in South Pole-Aitken basin. *Nat. Astron.* 5, 25–30. doi:10.1038/s41550-020-1197-x

Zhang, X., Lv, W., Zhang, l., Zhang, J., Lin, Y., and Yao, Z. (2021). Self-organization characteristics of lunar regolith inferred by Yutu-2 lunar penetrating radar. *Remote Sens.* 13, 3017. doi:10.3390/rs13153017

Zhou, B., Shen, S., Lu, W., Liu, Q., Tang, C., Li, S., et al. (2020). The Mars rover subsurface penetrating radar onboard China's Mars 2020 mission. *Earth Planet. Phys.* 4 (4), 1–10. doi:10.26464/epp2020054

Zhu, J., Lines, L., and Gray, S. (1998). Smiles and frowns in migration/velocity analysis. *Geophysics* 63 (4), 1200–1209. doi:10.1190/1.1444420

Frontiers in Astronomy and Space Sciences

TYPE Review
PUBLISHED 02 November 2023
DOI 10.3389/fspas.2023.1275516

Progress and prospects for research on Martian topographic features and typical landform identification

Danyang Liu[1,2] and Weiming Cheng[1,2,3,4]*

[1]State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China, [2]University of Chinese Academy of Sciences, Beijing, China, [3]Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, China, [4]Chinese Academy of Sciences, Center for Excellence in Comparative Planetology, Hefei, China

OPEN ACCESS

EDITED BY
Guangfei Wei,
Deep Space Exploration Laborarory,
China

REVIEWED BY
Yuyan Zhao,
Chengdu University of Technology,
China
Jiannan Zhao,
China University of Geosciences Wuhan,
China

*CORRESPONDENCE
Weiming Cheng,
✉ chengwm@lreis.ac.cn

RECEIVED 10 August 2023
ACCEPTED 13 October 2023
PUBLISHED 02 November 2023

CITATION
Liu D and Cheng W (2023), Progress and prospects for research on Martian topographic features and typical landform identification.
Front. Astron. Space Sci. 10:1275516.
doi: 10.3389/fspas.2023.1275516

The study of Martian surface topography is important for understanding the geological evolution of Mars and revealing the spatial differentiation of the Martian landscape. Identifying typical landform units is a fundamental task when studying the origin and evolution of Mars and provides important information for landing on and exploring Mars, as well as estimating the age of the Martian surface and inferring the evolution of the Earth's environment. In this paper, we first investigate Mars exploration, data acquisition and mapping, and the classification methods of Martian landforms. Then, the identification of several typical Martian landform types, such as aeolian landforms, fluvial landforms, and impact landforms, is shown in detail. Finally, the prospects of Mars data acquisition, landform mapping, and the construction and identification of the Martian landform classification system are presented. The construction of the Martian landform classification system and the identification of typical Martian landforms using deep learning are important development directions in planetary science.

KEYWORDS

Martian topography, landform classification, target recognition, Mars science, Martian mapping

1 Introduction

Mars is the most Earth-like planet in the Solar System (Ouyang and Xiao, 2011). Following studies of the Moon, Mars is a frontier area for space agency competition. As the planet that is closest to the Earth in the Solar System, Mars has attracted the attention of the United States, the European Union, India, and the former Soviet Union. Mars exploration began in October 1960, and vehicles that have successfully reached Mars to carry out missions include Mariner 9, the Viking series, the Mars Global Surveyor, Mars Odyssey, Mars Express, and the Mars Reconnaissance Orbiter. With the successful return of data from China's "Tianwen-1" probe, new research on Mars is well underway. Mars exploration is the first and crucial step from lunar exploration to planetary exploration. The study of the evolution of Mars is helpful for further exploration of possible life on Mars. Moreover, it has an important role in promoting evolutionary simulations and future climate predictions of the Earth.

The surface environment of Mars resembles those of both the Earth and the Moon. Mars has a lunar-like impact landform and volcanic landforms. However, the current state of the Martian landscape is more complex than that of the moon. Mars has sand dunes, gullies, alluvial fans and canyons, which have a similar landscape to that of Earth. Mars has a thin atmosphere and four seasons. Thus, Mars can be considered a composite version of the Moon and the Earth, which is a combination of many types of landforms. Explorations of the geological evolution of Mars and the conditions needed for the existence of life have attracted the attention of many researchers. Scientists have begun to search for areas on Earth that are similar to the Martian landscape to simulate and study the Martian environment. Mars-like regions are called "Mars analogs", such as the Antarctic Dry Valleys, the Atacama Desert in Chile, the Mars Desert Experiment Station in Utah, and the Tarim and Qaidam Basins in China (Xiao et al., 2017; Gou et al., 2018). Mars has many similarities to Earth; therefore, the study of the morphological characteristics of the Martian landscape has implications for the study of future evolutionary processes on Earth.

Martian landforms result from a combination of internal and external forces, revealing important geological events on Mars. Martian landforms reflect the highs and lows on the Martian surface and record the global distribution of water, heat and atmospheric motion on Mars. Currently, although many countries and institutions have launched probes to Mars, it remains impossible to bring back Martian samples to be studied, such as through dating and analysis (Yue et al., 2022). Therefore, it is important for Martian geomorphology research to focus on the surface topographic features of Mars based on remote sensing. With the continuous acquisition of higher-resolution and higher-accuracy Mars data, the quantification of Martian surface topography is becoming increasingly urgent. The quantitative analysis of Mars topographic features can provide a research basis for the exploration of Mars, landing on Mars, and the exploitation of Martian resources. Martian geomorphology research deepens the understanding of geomorphology research on Earth and is valuable as an important reference for estimating the age of the Martian surface, selecting of rover landing areas, and the assessing the evolution of the Martian geological landscape. However, there are few review studies of Martian surface topography. In this paper, we reviewed Mars exploration, data acquisition and mapping, as well as Mars geomorphological classification and identification methods, to reveal the characteristics of Martian topography. Additionally, the current research status of typical geomorphic unit identification was explored, and a reference and basis for both Mars science and related fields was provided.

2 Progress in Mars data acquisition and mapping

Humans have been exploring Mars since the 1960s, and the launched probes are shown in Table 1. Between 1960 and 2022, more than 40 probe missions were conducted, and a large amount of scientific data was obtained (Liu et al., 2006). Prior to the 1990s, Mars exploration projects mainly ended in failure. However, following the 1990s, the success rate of Mars exploration projects increased significantly and the requirements for exploration missions became increasingly high, evolving from flybys to orbits to landing and rover missions (Di et al., 2018). "Tianwen-1" was the first Mars probe launched in China, and "orbiting, landing and patrolling" were completed simultaneously (Li et al., 2015), which was historically significant.

Exploring the geomorphology, material composition, and atmospheric activity of Mars was among the main reasons to launch the probe. The successful launch of the rover has enabled the human understanding of Mars to evolve from "long-range exploration" (astronomical telescopes) to "close-range exploration" (orbiters, Mars rovers, etc.). The image data for Mars include both visible and thermal infrared images as well as topographic data. Currently, the above data are available for download on the internet at various spatial resolutions, as shown in Table 2.

NASA's Mars Survey/Mapping Working Group (MGCWG) defines the Mars coordinate system in polar and right-handed systems with specific parameters (Seidelmann, 2002). The stellar solid coordinate system (Di et al., 2021) used for Mars positioning and mapping includes the stellar solid Cartesian coordinate system (Vaucouleurs et al., 1973). The stellar solid geodetic coordinate system (Duxbury et al., 2002; Archinal et al., 2011), with an origin and a base plane, is similar to that of the Earth. Kim and Muller, 2008 developed a processing workflow by combining digital terrain models (DTMs) of 0.5–4 m networks extracted from the High-Resolution Imaging Science Experiment (HiRISE) and DTMs of 12–18 m grids extracted from stereo image pairs. Yan et al. (2022) produced a topographic dataset of the "Tianwen-1" landing area based on the HiRIC stereo photogrammetry processing scheme, with a ground sampling distance of 0.7 m, a digital orthophoto map (DOM) resolution of 3.5 m, and a digital elevation model (DEM) resolution of 3.5 m. Lakdawalla (2005) used Mars Orbiter Laser Altimeter (MOLA) data to produce a global topographic map of Mars.

Mars topographic mapping is the most direct means of studying Martian landforms, which provides a solid foundation for the selection of the landing area of Mars exploration projects and the study of the geological evolution of Mars. Currently, the accepted method of Mars mapping is to divide Mars into 30 panels (MC1-30), whose latitude and longitude are shown in Table 3. The USGS published a geologic map of Mars by Tanaka et al. (2014), which is geologically zoned to map Mars in a geologic age combining geomorphic unit code, which contains 44 geologic units at a scale of 1:20,000,000. In terms of a single type of landform mapping, Piqueux et al. (2019) mapped the location of near-surface ice across Mars based on the location of ice found under thin sandy soils. Liu et al. (2020) mapped the global distribution of centroids of Martian yardangs based on multisource Mars remote sensing data and provided a presumed time of yardang landform formation. Zhao (2017) mapped different types of paleolakes in the southern highlands of Mars, including open-system paleolakes. Dong and his team (Dong, 2020a) created the world's first aeolian geomorphologic map of Mars, classifying wind erosion and wind accretion landforms and filling the gap in the thematic map of Mars wind and sand. Carr (2006), Alemanno et al. (2018), Hynek et al. (2010) and Luo and Stepinski (2009) mapped the distribution of the global valley networks on Mars and formed a vector database, among which Luo

TABLE 1 Mars exploration projects since the 1960s.

Launch time	Detector name	Nation	Result
1960.10	Mars 1960A/1960B	Soviet Union	Launch failure
1962.10-11	Mars 1962A/1962B	Soviet Union	Launch failure
1962.11	Mariner 1	Soviet Union	Lost midstream
1964.11	Mariner 3	United States	Launch failure
1964.11	Detector 2	Soviet Union	Lost midstream
1965.7	Mariner 4	United States	Successful flyby, first photos of Martian surface
1969.7-8	Mariner 6/7	United States	Flyby success, flyby analysis of the Martian atmosphere and surface
1969-1971	Mars 1969A/1969B/Universe 419	Soviet Union	Launch failure
1971.5	Mariner 8	United States	Launch failure
1971.11	Mariner 9	United States	Surrounding success
1971.11	Mars 2	Soviet Union	Surrounding success, landing failure
1971.12	Mars 3	Soviet Union	Surrounding success, landing failure
1974.2-3	Mars 4/5/6/7	Soviet Union	Detector failure
1976.7-9	Viking 1/2	United States	Successful orbit, successful landing, first human photos of the surface of Mars
1988-1989	Phobos 1/2	Soviet Union	Lost midstream
1992.9	Mars Observer	United States	Lost midstream
1996.11	Mars 96	Russia	Launch failure
1997.7	Pathfinder/Sojourner	United States	Successful landing, first human rover on Mars
1997.9	Mars Global Surveyor	United States	Successful orbiting and global mapping of Mars completed
1998.7	Hope	Japan	Detector failure
1998.12	Mars Climate Orbiter	United States	Lost midstream
1999.1	Mars Polar Lander	United States	Landing failure
2001.10	Mars Odyssey	United States	Surrounding Success
2003.12	Mars Express	Europe	Surrounding success, landing failure
2004.1	Mars Exploration Rover	United States	Launch success
2006.3	Mars Reconnaissance Orbiter	United States	Surrounding success
2007.2	Rosetta Philae	Europe	Successful fly by
2008.5	Phoenix Mars Lander	United States	Successful landing, first landing on Martian North Pole
2009.2	Dawn	United States	Successful fly by
2011.11	Phobos/Ground	Russia/China	Launch failure
2012.8	Curiosity	United States	Landing success
2014.9	MAVEN	United States	Surrounding Success

(Continued on the following page)

TABLE 1 (*Continued*) Mars exploration projects since the 1960s.

Launch time	Detector name	Nation	Result
2014.9	Mangalyaan	India	Surrounding Success
2016.10	ExoMars2016	Europe/Russia	Surrounding success, landing failure
2018.11	InSight/MarCO	United States	Successful landing, successful flyby
2020.7	Perseverance/Ingenuity	United States	First unmanned Martian helicopter
2021.2	Hope	United Arab Emirates	Surrounding Success
2021.2	Tianwen-1	China	The first time "surrounding, landing, and patrol" were completed at once

TABLE 2 Landform and image data for the Martian surface (available for download online).

Data source	Resolution (m)	Scale	Download address
HiRISE DTM	1	Local	https://www.uahirise.org/dtm/index.php?page=2
THEMIS optical image	100	Global	http://themis.asu.edu/feature
MOLA DEM	463	Global	https://astrogeology.usgs.gov/search/details/Mars/GlobalSurveyor/MOLA/Mars_MGS_MOLA_DEM_mosaic_global_463m
HRSC DEM, relief	200	Global	https://astrogeology.usgs.gov/search/map/Mars/Topography/HRSC_MOLA_Blend/Mars_HRSC_MOLA_BlendShade_Global_200mp
MRO CTX optical image	6	South pole	https://planetarymaps.usgs.gov/mosaic/Mars/Mars_MRO_CTX_SPole_Mosaic_Robbins/
MRO CTX optical image	5	Global	https://murray-lab.caltech.edu/CTX/tiles/

and Stepinski (2009) also calculated the volume, length and other attributes of valley networks.

3 Research progress on the classification of Martian landform types

3.1 Macroscopic analysis of the surface topographic features of Mars

The north-south dichotomy is the most obvious geomorphic feature of Mars. The Southern Hemisphere is highly topographically variable with complex landform types, and includes impact craters, highlands, canyons, dry rivers, sand dunes, yardangs, volcanoes, and other landform types of various sizes. There are three clearly visible impact basins on Mars: the Argyre and Hellas basins in the south and the Isidis basin near the equator. Figure 1 shows the global topography of Mars. On the other hand, the Northern Hemisphere is dominated by low-relief plains. The main northern plain consists of Vastitas Borealis. The Tharsis bulge is near the equator, and its northern edge includes three volcanic regions: Olympus Mons, Alba Mons, and Tempe Terra. The presence of liquid water on Mars has been a mystery pursued by scientists. Orosei et al. (2018), through the Mars Express orbiter mission, found a 20-km-wide body of liquid water at a depth of 1.5 km below the ice cap.

3.2 The basis of Martian landform classification

The landform types mentioned above have been influenced by endogenic and exogenic forces. The geological lifetime of Mars ended when the magma activity in the interior of Mars ceased. Subsequently, the external magnetic field of Mars disappeared with the disappearance of nuclear convection. The disappearance of the magnetic field left Mars completely exposed to solar wind and cosmic high-energy rays. The combination of endogenic forces, such as early volcanic eruptions and tectonic movements, and exogenic forces, including meteorite impacts, wind accumulation and erosion, and water scouring, have resulted in the formation of a variety of landforms.

The common landforms included impact craters, volcanic landforms, glacial landforms, valley networks, and sand dunes on the surface of Mars and Earth (Wang, 2018). Typical flowing water features, such as alluvial fans and canyons, were preserved on the Martian surface, indicating the possibility of liquid water on

TABLE 3 Mars Quadrangles (modified from https://marspedia.org/Mars_Quadrangles).

Number	Name	Latitudes	Longitudes
MC-01	Mare Boreum	65°–90° N	180°W-180°E
MC-02	Diacria	30°–65° N	120°–180° W
MC-03	Arcadia	30°–65° N	60°–120° W
MC-04	Mare Acidalium	30°–65° N	0°–60° W
MC-05	Ismenius Lacus	30°–65° N	0°–60° E
MC-06	Casius	30°–65° N	60°–120° E
MC-07	Cebrenia	30°–65° N	120°–180° E
MC-08	Amazonis	0°–30° N	135°–180° W
MC-09	Tharsis	0°–30° N	90°–135° W
MC-10	Lunae Palus	0°–30° N	45°–90° W
MC-11	Oxia Palus	0°–30° N	0°–45° W
MC-12	Arabia	0°–30° N	0°–45° E
MC-13	Syrtis Major	0°–30° N	45°–90° E
MC-14	Amenthes	0°–30° N	90°–135° E
MC-15	Elysium	0°–30° N	135°–180° E
MC-16	Memnonia	0°–30° S	135°–180° W
MC-17	Phoenicis Lacus	0°–30° S	90°–135° W
MC-18	Coprates	0°–30° S	45°–90° W
MC-19	Margaritifer Sinus	0°–30° S	0°–45° W
MC-20	Sinus Sabaeus	0°–30° S	0°–45° E
MC-21	Iapygia	0°–30° S	45°–90° E
MC-22	Mare Tyrrhenum	0°–30° S	90°–135° E
MC-23	Aeolis	0°–30° S	135°–180° E
MC-24	Phaethontis	30°–65° S	120°–180° W
MC-25	Thaumasia	30°–65° S	60°–120° W
MC-26	Argyre	30°–65° S	0°–60° W
MC-27	Noachis	30°–65° S	0°–60° E
MC-28	Hellas	30°–65° S	60°–120° E
MC-29	Eridania	30°–65° S	120°–180° E
MC-30	Mare Australe	65°–90° S	180° W—180° E

Mars. Atmospheric movement and the high diurnal temperature differences have contributed to the formation of typical aeolian landforms, such as sand dunes and yardangs. The University of Western Ontario (Canada) initiated the Interactive Mapping of Mars (https://imars.uwo.ca/tutorial/) project, which classified Martian landforms into aeolian, water, glacial/periglacial, impact cratering, mass movement and volcanic landforms. Di et al. (2021) classified the landforms according to the causes of formation, including aeolian landforms, fluvial landforms, and tectonic landforms (referring to impact geomorphology and volcanic geomorphology). The above macroscopic landform classifications were further subdivided based on morphology and formation. OuYang and Zou, 2015 proposed classifying Martian landforms into aeolian, fluvial, canyon, impact, volcanic, and glacial landforms. This classification, combined with the traditional method of Martian landform classification, was used in this paper according to the relevant formation mechanisms, as shown in Figure 2.

In addition, the geomorphological classification systems of the Earth and Moon can aid in the geomorphological classification of Mars. The three-level and nine-class classification system of Earth (Institute of Geography Chinese Academy of Sciences, 1987; Geomorphic Map Editorial Committee of the People's Republic of China, 2009; Zhou et al., 2009) includes the geomorphological category, geomorphological class, geomorphological shape (three levels), and macromorphological type subclass, land elevation and seafloor bathymetry subclass, maincrop force type subclass, maincrop force mode of action subclass, material composition and lithology subclass, geomorphic age subclass, combined morphological subclass, micromorphological subclass, and slope morphological subclass (nine classes). The three-level and eight-class classification system of the Moon (Cheng et al., 2018; Liu et al., 2022) does not include the maincrop force mode of action compared with the three-level and nine-class classification systems of Earth.

3.3 Method of classifying Martian landforms

In terms of overall classification, Martian landform classification methods can be divided into two categories.

One category is mainly based on surface elevation, relief, slope and other features, which can be used to parameterize the surface morphology or extract surface features to classify landforms with clustering or machine learning methods. Bue and Stepinski, 2007 classified Martian landforms into highlands, impact craters, lowlands, high-relief landforms, and channels based on six topographic parameters (elevation, flood, slope, flooded slope, contributing area, and flooded contributing area), using a self-organizing mapping method and Ward clustering. However, a limitation was misclassification for certain features, such as craters. Wang et al. (2017) classified lunar landforms into high-relief, highlands, lowlands, impact craters and other landform types. Comparably, Wang et al. (2017) added relief to topographic parameters and replaced Ward clustering with ISO clustering to achieve an overall accuracy of 83.34% and a kappa coefficient of up to 0.77, despite certain limitations influenced by aggradation, degradation, and complex landform types.

Additionally, with the rapid development of machine learning, deep learning and other methods to segment the content of images at the pixel level, image segmentation has been used in classification tasks. The related steps include creating training sets, designing

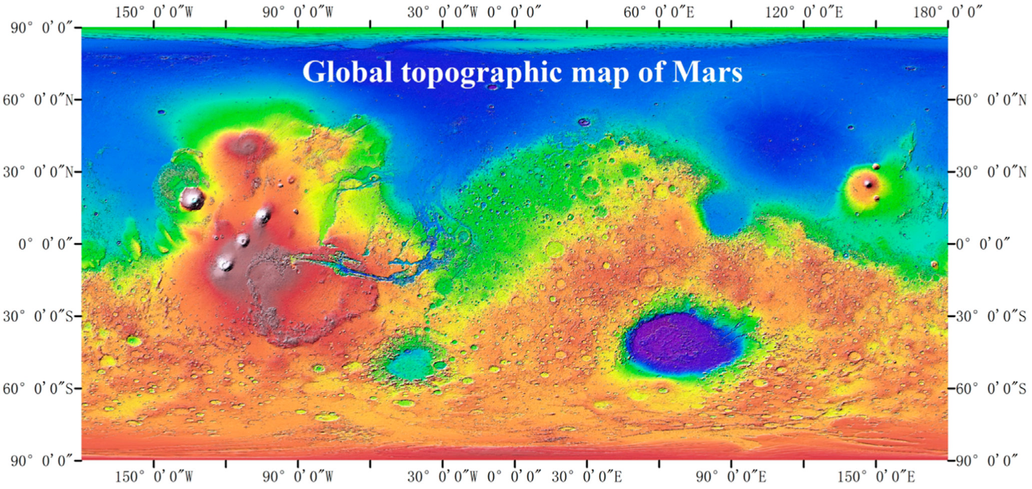


FIGURE 1
Global topographic map of Mars (MOLA DEM; Resolution: 463 m per pixel.) (https://astrogeology.usgs.gov/search/map/Mars/GlobalSurveyor/MOLA/Mars_MGS_MOLA_ClrShade_merge_global_463m).

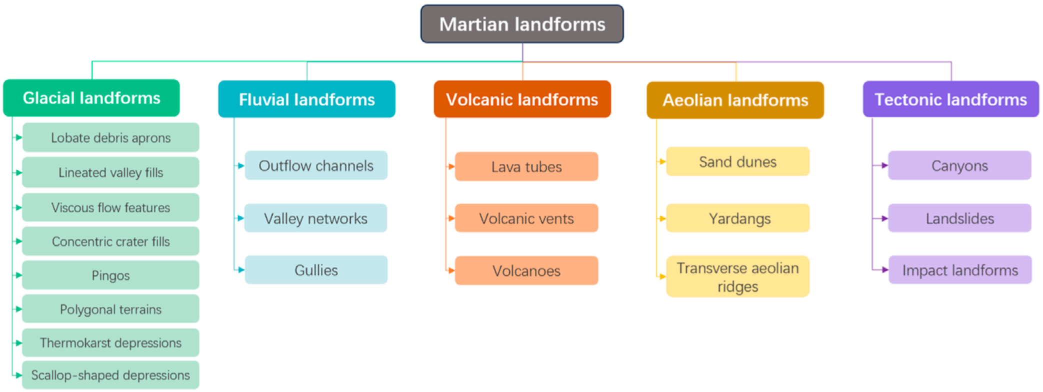


FIGURE 2
Types of Martian landforms.

network model structures, and adjusting parameters, among others, for the binary/multiclass segmentation of landform elements. Shang and Barnes, 2013 used the fuzzy rough feature selection method combined with a support vector machine (SVM) to classify landform types in a single Mars image, and the comparative experimental results showed that the method outperformed decision trees and K-nearest neighbor classification. Jiang et al. (2021) developed a new end-to-end deep learning framework, rotated SSD, to localize and identify different Martian landforms at the same time by using a rotatable-based anchor box mechanism and introducing unsupervised training based on autoencoders. Barrett et al. (2022) used HiRISE images of Oxia Planum and Mawrth Vallis and developed a deep learning terrain classification system (NOAH-H) based on semantic segmentation with deep neural networks. Wright et al. (2022) used NOAH-H to classify four HiRISE images of the terrain near the 2020 Perseverance Rover landing site at Jezero based on impact crater images, and the classification results agreed with the manually created geological maps. Rothrock et al. (2016) developed an algorithm (SPOC) based on deep convolutional neural networks (CNNs) that can identify terrain elements, such as sand, bedrock, wrinkled ridges, and steep slopes, and the algorithm was successfully applied for terrain identification in the Mars 2020 rover landing zone and MSL mission sliding prediction. Wang et al. (2021) designed a new framework for Mars rover image classification with semisupervised contrast learning; it ignores intraclass pairs in labeled data and counterexample pairs in unlabeled data, thus substantially improving the classification model.

With the rapid development of computer vision, image segmentation, deep learning and other fields, Martian landform classification research has also been rapidly developed, and an increasing number of types of landform elements can now be recognized, thus providing the basis for research on landform recognition and the evolution pattern of Mars landings.

4 Progress in the identification of typical landform types on Mars

4.1 Aeolian landforms

Aeolian landforms are the most prevalent and active geomorphic features on Mars, and the ancient Chinese name for Mars "Yinghuo," is in part due to the sand phenomenon that causes Mars to be bright and pale (Dong et al., 2020b). Li et al. (2022) developed a similar visual degradation process based on the remote sensing images of "Tianwen-1" to synthesize real dust images and used these real dust images to train a deep learning model to identify dust-free images, inspired by the fog formation process on Earth. Yao, 2021 identified 882 dust storms with a diameter of 4,000 km in the southern part of Utopia based on Tianwen-1 landing area image data and analyzed their causes and distribution characteristics.

Sand dunes are the most typical manifestation of aeolian landforms. The methods for identifying sand dunes on Mars can be classified into four types.

1) Visual observation: Hayward et al. (2007) produced a database of medium and large sand dunes with areas larger than 1 km^2, covering 550 dunes between 65°N and 65°S; however, many dunes are still not included in the database. "Visual observation" could obtain sand dunes intuitively and qualitative information, while it is inefficient for analyzing large-scale datasets.
2) Extracting gradient and grayscale statistical features: Carrera et al. (2019) used gradient and statistical features to structure a probabilistic classifier based on the R-vine distribution to identify Martian dunes and compared it with a centralized advanced classification algorithm, and superior results were obtained. "Extracting gradient and grayscale statistical features" can efficiently provide quantitative data related to texture and shape efficiently. Moreover, it could be used as input features for machine learning models. However, it is sensitive to variations in lighting and image quality and may miss complex patterns.
3) Machine learning and deep learning: Rubanenko et al. (2021) used Mask R-CNN based on the MRO's background camera (CTX) for the automatic extraction of Martian dunes and then mapped the distribution of the full Martian dune field, discovering that dunes were more abundant in the Northern Hemisphere than in the Southern Hemisphere, which might be attributed to latitudinally dependent wind regimes, sediment supply, or sediment availability. "Machine learning and deep learning" can be used to process large-scale image analysis and reduce human subjectivity. Nevertheless, it requires substantial training data and depends on the data quality.
4) Combining machine learning with feature extraction: Bandeira et al. (2010) extracted gradient and grayscale histogram features from images aggregated into larger areas, which constituted detection units, and combined boosting and SVM classifiers to extract Martian surface dunes with a 98.7% recognition rate. This method combines the strengths of both approaches, while it requires expert knowledge in both feature extraction and machine/deep learning. This high complexity makes it difficult to interpret subsequent results.

Machine/deep learning requires a complete training dataset. The diverse shapes (Li, 2018) and sizes (Mandt and LeoneYardang, 2015) of yardang landforms make it difficult to establish a complete training dataset or a unified identification standard. Therefore, the identification of yardang landforms is primarily performed manually. Researchers have focused on the color, morphology, and formation conditions of yardang landforms to characterize the Martian atmospheric environment and prevailing wind direction. Ward (1979), Bridges et al. (2007), and Zimbelman and Griffin, 2010 conducted analyses of the aspect ratio, scale distribution, and material properties of large-scale yardang landforms in the Martian Amazonian plain and Medusa trough layer, respectively. Additionally, some scholars have focused on dating yardang landforms (Kerber et al., 2011; Zimbelman and Scheidt, 2012; Liu et al., 2021a).

The identification of transverse aeolian ridges (TARs) is largely based on their sediment composition (Fenton et al., 2003; Balme and Bourke, 2005; Balme et al., 2008; Bourke et al., 2003) and the difference in albedo (Bouke et al., 2008; Gou et al., 2022) using visual interpretation. Lu et al. (2022) identified four crescent-shaped lateral sand ridges based on Zhurong rover exploration data and analyzed the erosion process of these beds. The orientation of these beds is related to the angle between the bed crest and the wind direction. Several typical aeolian landforms are shown in Figure 3.

4.2 Fluvial landforms

Fluvial landforms have long been regarded as the best evidence for the existence of liquid water on the Martian surface, which is essential for the existence of life and an indication of the warm and humid climate that once existed on Mars. Fluvial landforms can be categorized into narrowly defined fluvial landforms and broadly defined fluvial landforms. Narrowly, Ouyang and Zou (2015) further classified the fluvial landforms of Mars into outflow channels, valley networks, and gullies according to their size scale, as shown in Figure 4. Broadly, Zhao et al. (2021) classified the fluvial landforms of Mars into valley networks, outflow channels, paleolake basins, and alluvial fans and deltas. In this article, fluvial landforms were discussed from a narrow perspective, namely, outflow channels, valley networks, and gullies. These three landforms were distinguished by their sizes, patterns, and formation mechanisms.

The outflow channels on Mars formed during catastrophic floods (Cutts and Blasius, 1981), with a width of 1 km to several hundred km. Consequently, the age of their formation was deduced to be Hesperian (Liu et al., 2021b). Outflow channels are mostly found in fracture zones or near canyons. There are distinct tear-drop islands in the channels. In addition to their size, another obvious difference between valley networks and outflow channels is dendritic branching. Valley networks are mostly found in the old southern hemisphere and are very rare in the younger northern hemisphere, with a width of several kilometers. This suggested that valley network formation occurred on early Mars (Tanaka et al.,

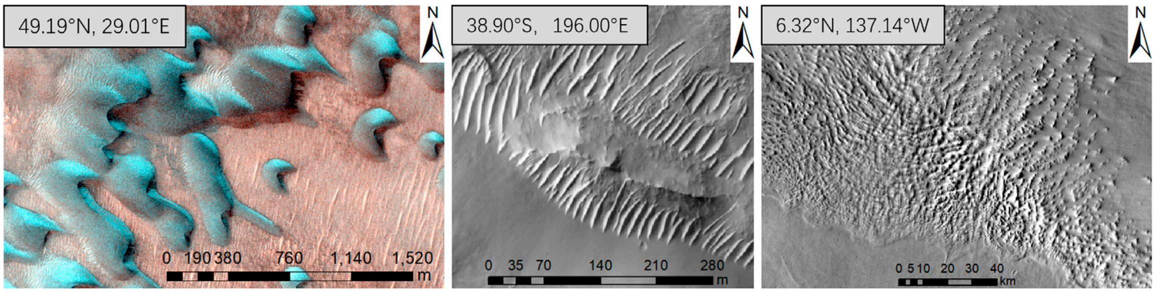


FIGURE 3
Several typical aeolian landforms (left, dunes, HiRISE; middle, TARs, HiRISE; right, yardangs, Themis day IR; modified after Balme et al., 2008).

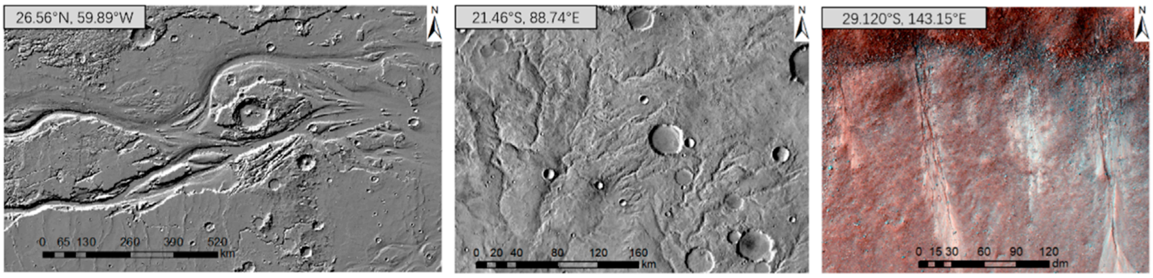


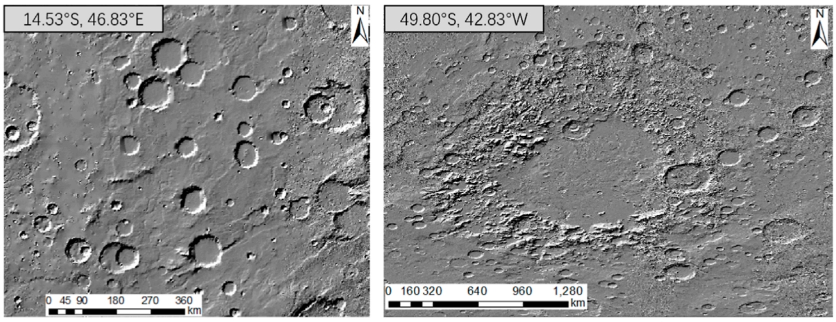
FIGURE 4
Several typical fluvial landform types (left, outflow channels, MOLA hillshade; middle, valley networks, Themis day IR; right, gullies, HiRISE).

2014; Ouyang and Zou, 2015; Liu et al., 2021b). The traditional view holds that valleys resulted from surface runoff on early Mars, which climate was warm and wet (Carr, 2006; Hynek et al., 2010). Another view holds that valleys resulted from groundwater loss or glaciation. Gullies are distributed along slopes, with a width of several meters. Their formation is controversial, and may have included melting snow or CO_2 ice. Generally, the identification of gullies is based on high-resolution data due to their small-scale features.

The methods used to identify valley networks can be divided into 1) Visual interpretation: Carr (2006) mapped the distribution of Martian valley networks and outflow channels. Hynek et al. (2010) mapped a Mars-wide network of canyons based on visible, infrared, and topographic data and analyzed the reasons for the presence of these canyons. Alemanno et al. (2018) modified the approach of Carr (2006), and the results showed that the valley network of Mars is predominantly located in the Southern Hemisphere, with small distributions in the Northern Hemisphere on the rim of the plateau and near Elysium volcano. "Visual interpretation" relies on the level of experts and costs a lot of manpower and time, which is similar with the pros and cons of visual interpretation for aeolian landforms. 2) Hydrological analysis: Stepinski and Collier (2004) proposed a method for extracting drainage networks from lower basins using a contributing area threshold, which was validated for 28 Noachian-age regions on Mars, with high efficiency and accuracy. Gou et al. (2018) extracted valley networks based on the DEM data from the Evros Vallis basin using traditional hydrological analysis methods, such as filling, flow direction analysis, catchment accumulation, and catchment area calculation. "Hydrological analysis" relies on quantitative algorithms and can provide more objective results compared to visual interpretation, reducing the potential for human bias. However, the accuracy of hydrological analysis methods depends on the validity of certain assumptions, which may not hold true in all situations. Furthermore, it is limited by the resolution of DEM. 3) Extraction of elevation change characteristics: Molly and Stepinski (2007) proposed a DEM-based method for the identification of landforms characterized by curvature and separated valleys from other landforms with convex shapes and then reconnected the segments along the drainage networks. The research about identification of gullies is rare due to the data resolution limitations (Li et al., 2015). Li et al. (2015) extracted Martian gullies in six regions using HiRISE images based on mathematical morphology, the bottom-hat transform and path opening and closing, with detection rates reaching 76%–94%. "Extraction of elevation change characteristics" is suitable for regional and global studies, while choosing appropriate parameters for elevation change analysis can be challenging.

4.3 Impact landforms

Impact landforms include impact craters and impact basins, such as the geomorphic units shown in Figure 5. As tracers of



Impact crater

Impact basin

FIGURE 5
Impact crater and impact basin. (MOLA hillshade; resolution: 463 m/pixel).

surface processes, impact craters are informative for studying subsurface minerals and useful for dating the planetary surface (Neukum et al., 1975; Wilhelms et al., 1987; Hartmann and Neukum, 2001). Thus, related studies are important for assessing planetary geomorphology. Furthermore, impact basins are often regarded as impact craters when they are identified. Mars and the Moon are largely similar in their morphological characteristics regarding impact crater landforms, and therefore, this paper combines the progression of impact crater studies on Mars and the Moon.

The methods for identifying impact craters can be divided into manual and automatic identification methods. Manual identification aims to identify accurate boundaries based on the visual interpretation of impact craters in images, topographic data and the subsequent derived data. However, manual identification is time-consuming and laborious, relying heavily on expert knowledge of individual identifiers, and the identification standards are inconsistent. Currently, the published Mars impact crater databases include the Barlow (1988) database, MA132843GT (Salamuniccar et al., 2012; Robbins and Hynek, 2010; Robbins and Hynek, 2012) database. The Barlow database manually extracted 25,826 impact craters with diameters greater than or equal to 8 km and relatively young ages based on Viking images, of which 60% were formed by heavy bombardment. The MA132843GT catalog is based on the MA130301GT catalog (Salamunićcar et al., 2011b), which combines manually mapped impact craters and 72,668 impact craters extracted by automatic identification algorithms. The Robbins database contains 384,343 impact craters with diameters greater than or equal to 1 km, which were manually identified from images.

The automatic recognition algorithm saves time and effort compared with manual recognition, and the recognition standard is uniform. Automatic recognition algorithms can be roughly divided into four categories: 1) Traditional edge detection and circle fitting algorithms. Edge extraction is achieved by using grayscale mutation on both sides of the impact crater rim (Kim et al., 2005), such as via the Sobel operator (Lu et al., 2013), the Robert operator (Yuan et al., 2013), the Prewitt operator, the Canny algorithm (Salamunićcar et al., 2010; Jiang et al., 2013), the region growing method (Luo et al., 2014), and the Hough transform. In addition, the bright and dark areas formed by impact craters in optical images (Urbach and Stepinski, 2009) can be paired to identify impact craters. This method can be easily understood, while the results were fitting circles, instead of accurate boundaries. 2) Digital terrain analysis method. Based on the changes in terrain factors, such as elevation and slope at the rim of a crater, watershed analysis, contour lines, and the extraction of terrain features were used to identify impact craters from a three-dimensional perspective (Bue and Stepinski, 2007; Luo et al., 2013; Xie et al., 2013; Liu et al., 2017; Chen et al., 2018). This method obtained changes from elevation and variations in crater boundaries. However, it may provide low efficiency and be influenced easily by minor terrain mutations. 3) Traditional machine learning algorithms. These methods tend to first express the features of impact craters using feature descriptors, such as Haar and PHOG, and then combine them with traditional machine learning models, such as SVM and AdaBoost (Ding et al., 2013; Bandeira et al., 2014), to perform impact crater detection. The quality of the training results depends on the quality of feature selection (Liu et al., 2023). 4) Deep learning algorithms. Deep learning is data-driven and is used to delegate feature selection to a machine learning algorithm. Deep learning can be used to extract features automatically from images, which is comparable to the idea of a "black box". Nevertheless, it is highly dependent on the quality of the data and uneasily interpretative (Hsu et al., 2021; Silburt et al., 2019; Yang et al., 2020; Zheng et al., 2020; Gao et al., 2022). Combining digital terrain analysis and deep learning may be a further development direction, which could consider geographical information. The results could provide accurate boundaries over large-scale regions.

4.4 Glacial landforms

Martian glacial landforms are closely related to the presence of water, the stable presence of which is necessary to support life. The topography of Mars has a distinct north-south dichotomy, with high south and low north topography. The large topographic height difference redistributes water between the poles, resulting in glacial

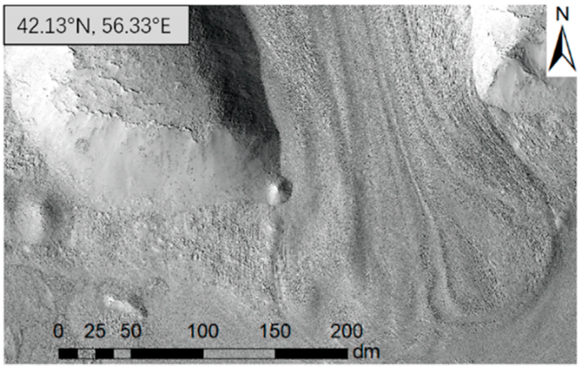


FIGURE 6
Glacier-like features on Mars (HiRISE).

landforms in the mid-latitudes (Hepburn et al., 2020), as shown in Figure 6.

Souness and Hubbard, 2012 identified and cataloged Martian glacier-like landforms from 8058 CTX images, 1,309 of which were concentrated at mid-latitudes of 39.3°N and 40.7°S, and inferred the response mechanisms of Martian glacial landforms based on latitude and altitude. There are various types of Martian glacial landforms, including concentric crater fills (CCFs), lobed valley fills (LVFs), lobate debris aprons (LDAs), and viscous flow features (VFFs) (Levy et al., 2010; Pedersen and Head, 2010; Liu et al., 2021c). Milliken et al. (2003) used 13,000 Mars orbiting camera images distributed globally to identify 146 images containing VFFs and marked the locations of these VFFs as points. Galofre et al. (2022) combined data from CaSSIS, SHARAD, CTX, and HiRISE to explore the deformation history of tongue-shaped rocky debris slopes, their internal structure, and their influence on regional climate change from surface morphology, which can be used to infer past glaciation on Mars. Petersen et al. (2018) used SHARAD radar data to estimate the water ice content (greater than 80%) in LDAs. Furthermore, only a small fraction of cases that matched typical surface water erosion indicated extensive subglacial erosion on the Martian surface after analyzing more than 10,000 Martian valleys.

In addition to Martian glacial landforms (beyond CCFs, LVFs, LDAs, and VFFs), periglacial landforms, such as pingos (Dundas et al., 2008; Richard et al., 2021), scallop-shaped depressions (Lefort et al., 2010), polygonal terrains (Soare et al., 2021), and thermokarst depressions (Bernhardt et al., 2016), are also present. Especially, some glacial-relevant landforms were found in the mid-latitudes of the Utopia Planitia. Wang et al. (2021) applied 2D-PCA and identified potential water ice in Utopia Planitia using FP-SPR data from the Zhurong rover. Bina and Osinski, 2021 found a new landform "decameter-scale rimmed depressions" in Utopia Plantita using SHARAD, which represents a further marker for the presence of ground ice in the northern plains of Mars.

Although glacial landforms, such as VFFs, CCFs, LVFs, and LDAs, have been discussed, studies have focused more on the formation, evolution, material composition, and structure of the above landforms and less on their identification. Additionally, the relationship of periglacial and glacial landforms between Mars climate evolution has been a focus of glacial landform research.

5 Prospects for studying Martian surface topography

Since Martian samples cannot be returned to Earth at present, the study of the surface topographic features of Mars based on remote sensing is still a very important direction of research, which has promoted a better understanding of the geological development and evolution of Mars, the dating of typical geomorphic units and the inference of its genesis. The Mars exploration project, the production of Mars optical image data and derivative products, and the study of Martian aeolian landforms and impact landforms have received more attention than other topics. However, more focus and deeper research are still needed in the directions of Martian landform classification and mapping, fluvial and glacier landform identification, and comparative planetary studies.

5.1 Mars data acquisition and mapping

Topographic data are characterized without the limitation of optical light and darkness and reflect the real terrain. Global-scale Martian data are usually characterized by low resolution, and localized data have a high resolution but a limited range. In low-resolution data, some fine-textured geomorphic units, such as gullies, miniature yardangs, and secondary impact craters, cannot be identified. Data resolution is an important factor affecting the recognition accuracy. However, the current Mars global DEM data resolution is 463 m/pixel (the blended DEM from MOLA and HRSC data resolution has a resolution of 200 m/pixel), which is not sufficient for the recognition and analysis of fine features. The production of higher-resolution global-scale Mars topographic data would be beneficial for the identification of finer geomorphic units.

In Mars geomorphology mapping research, some specific geomorphological units and Mars geological maps have been produced, but there is a lack of global geomorphological regionalization maps for Mars. Therefore, the digital geomorphological regionalization methods of the Earth and Moon could be used as references to draw the Mars geomorphological regionalization classification map based on the topographic indexes of Mars, such as relief, slope and elevation. This approach could aid in exploring the spatial differentiation characteristics of the Mars surface evolution process.

5.2 Construction of a classification system for Martian landforms

Mars has landforms that are typical of both the Earth and the Moon. The "three classes and nine levels" classification system for Earth's landforms and the "three classes and eight levels" classification system for lunar landforms are useful for establishing a complete classification system of Martian landforms. However, the applicability of these classification systems to Martian landforms needs to be further explored. At present, studies on impact craters (Lei, 2017), sand dunes (Li et al., 2020), and yardangs (Liu, 2021) have established a detailed morphological index system, but index systems for describing other geomorphic types has been less studied. Most of the existing geomorphological classification studies (Bue and Stepinski, 2006; Wang et al., 2017; Deng et al., 2022; Liu et al., 2022) have divided geomorphological units into subtypes based on morphology. However, there have been few studies of descriptive indexes reflecting the formation or material composition. Combining the cause of formation and morphology is the key to geomorphological classification (Shen et al., 1982; Cheng et al., 2011). Combined with multisource remote sensing data, including visible images, thermal infrared images and topographic data, descriptive indexes considering the formation, material composition and geological age was introduced to build a complete index system and a database of corresponding geomorphic units. The construction of an index system for Martian geomorphic classification needs to be further explored.

5.3 Identification of typical geomorphic units of Mars

In terms of recognizable objects, automatic recognition has been rapidly developed in the field of impact crater recognition; however, to recognize more complex geomorphic units, such as sand dunes, yardangs, canyons, valley networks, gullies and degraded impact craters, manual recognition is still the main identification method. Manual identification could represent the real morphology accurately, but it is time-consuming and laborious and depends on the professional knowledge level of the experts. That results in heterogeneous mapping standards and heterogeneous mapping quality. Therefore, in research on the recognition of geomorphological units with complex morphology, automatic recognition methods need to be studied further.

In terms of the applicable scope of the recognition methods, most existing terrain classification methods are applicable to small-scale or local-scale terrain classification, while there are fewer automatic recognition methods applicable to large-scale or even global-scale classification. Although large-scale catalogs of impact craters (Robbins and Hynek, 2012), sand dunes (Hayward et al., 2007), yardangs (Liu, 2021), valley networks (Alemanno et al., 2018), and glacial-like landforms (Souness and Hubbard, 2012) were proposed, these features were identified manually rather than automatically. Because the accuracy, efficiency and robustness of these automatic methods for the global-scale regions were unsatisfactory, the automatic methods were typically applied in local regions. The methods that are applicable only for local small-scale target recognition lose the essential advantages of automatic recognition methods. Therefore, automatic recognition methods that are applicable at large and global scales require further research.

In terms of the requirements of recognition methods based on the users' level of knowledge, the current Mars morphology recognition are mostly based on traditional machine learning, with a lack of end-to-end deep learning methods. On the one hand, the effect of machine learning depends on feature selection (Guyon and Elisseeff, 2003; Domingos, 2012), which requires a high level of expertise for the user. Deep learning is data-driven with the ability of automatic representation learning (Qian et al., 2022; Chen et al., 2023). On the other hand, traditional machine learning provides interpretable models, while deep learning is often seen as a "black box". It is difficult to interpret the modeling process. Furthermore, deep learning requires more computational resources and larger training datasets than machine learning. Currently, end-to-end deep learning models are widely used in various professional fields, and it is hoped that they can be applied to intelligent terrain type recognition in the case of Mars. Broadly applicable training datasets and well-performing models are expected to be developed to identify multiple Martian terrain units.

Author contributions

DL: Investigation, Writing–original draft, Writing–review and editing, Conceptualization. WC: Conceptualization, Funding acquisition, Project administration, Supervision, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This article was funded by the B-type Strategic Priority Program of the Chinese Academy of Sciences, Grant No. XDB41000000 and National Natural Science Foundation of China, No. 42130110.

Acknowledgments

We appreciate the detailed suggestions and constructive comments from the editor and the reviewers.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alemanno, G., Orofino, V., and Mancarella, F. (2018). Global map of Martian fluvial systems, age and total eroded volume estimations. *Earth Space Sci.* 5 (10), 560–577. doi:10.1029/2018ea000362

Archinal, B. A., A'Hearn, M. F., Bowell, E., Conrad, A., Consolmagno, G. J., Courtin, R., et al. (2011). Report of the IAU working group on cartographic coordinates and rotational elements: 2009. *Celest. Mech. Dyn. Astron.* 109 (2), 101–135. doi:10.1007/s10569-010-9320-4

Balme, M., Berman, D. C., Bourke, M. C., and Zimbelman, J. R. (2008). Transverse aeolian ridges (TARs) on Mars. *Geomorphology* 101 (4), 703–720. doi:10.1016/j.geomorph.2008.03.011

Balme, M. R., and Bourke, M. C. (2005). *Preliminary results from a new study of transverse aeolian ridges (TARS) on mars//36th lunar and planetary science conference*. Houston, Texas: The Open University 1892.

Bandeira, L., Machado, M., and Pina, P. (2014). Automatic detection of sub-km craters on the Moon. *Lunar Planet. Sci. Conf.*

Bandeira, L., Marques, J. S., and Saraiva, J. (2010). Automated detection of sand dunes on Mars//7th international conference on image analysis and recognition. *Image Analysis Recognit.* 6112 (1), 306–315. doi:10.1007/978-3-642-13775-4_31

Barlow, N. G. (1988). Crater size-frequency distributions and a revised Martian relative chronology. *Icarus* 75 (2), 285–305. doi:10.1016/0019-1035(88)90006-1

Barrett, M. A., Balme, R. M., Woods, M., Karachalios, S., Petrocelli, D., Joudrier, L., et al. (2022). NOAH-H, a deep-learning, terrain classification system for Mars: results for the ExoMars Rover candidate landing sites. *Icarus* 371 (1), 114701. doi:10.1016/j.icarus.2021.114701

Bernhardt, H., Reiss, D., Hiesinger, H., and Ivanov, M. A. (2016). The honeycomb terrain on the Hellas basin floor, Mars: a case for salt or ice diapirism. *J. Geophys. Res. Planets* 121 (4), 714–738. doi:10.1002/2016je005007

Bina, A., and Osinski, R. G. (2021). Decameter-scale rimmed depressions in Utopia Planitia: insight into the glacial and periglacial history of Mars. *Planet. Space Sci.* 204, 105253. doi:10.1016/j.pss.2021.105253

Bourke, M. C., Wilson, S. A., and Zimbelman, J. R. (2003). *The variability of transverse aeolian ridges in troughs on Mars//34th Lunar and Planetary Science Conference*.Texas: League, 2090.

Bridges, N. T., Geissler, P. E., Mcewen, A. S., Thomson, B. J., Chuang, F. C., Herkenhoff, K. E., et al. (2007). Windy Mars: a dynamic planet as seen by the HiRISE camera. *Geophys. Res. Lett.* 34 (23), 497–507. doi:10.1029/2007gl031445

Bue, B. D., and Stepinski, T. F. (2006). Automated classification of landforms on Mars. *Comput. Geosci.* 32 (5), 604–614. doi:10.1016/j.cageo.2005.09.004

Bue, B. D., and Stepinski, T. F. (2007). Machine detection of Martian impact craters from digital topography data. *IEEE Trans. Geosci. Remote Sens.* 45 (1), 265–274. doi:10.1109/tgrs.2006.885402

Carr, M. H. (2006). *The surface of mars*. Cambridge, UK: Cambridge University Press.

Carrera, D., Bandeira, L., Santana, R., and Lozano, J. A. (2019). Detection of sand dunes on Mars using a regular vine-based classification approach. *Knowl. Based Syst.* 163 (1), 858–874. doi:10.1016/j.knosys.2018.10.011

Chen, M., Liu, D. Y., Qian, K. J., Li, J., Lei, M., and Zhou, Y. (2018). Lunar crater detection based on terrain analysis and mathematical morphology methods using digital elevation models. *IEEE Trans. Geosci. Remote Sens.* 56 (7), 3681–3692. doi:10.1109/tgrs.2018.2806371

Chen, M., Qian, Z., Boers, N., Jakeman, A. J., Kettner, A. J., Brandt, M., et al. (2023). Iterative integration of deep learning in hybrid Earth surface system modelling. *Nat. Rev. Earth Environ.* 4, 568–581. doi:10.1038/s43017-023-00452-7

Cheng, W. M., Liu, Q. Y., and Wang, J. (2018). A preliminary study of classification method on lunar topography and landforms. *Adv. Earth Sci.* 33 (9), 885–897. doi:10.11867/j.issn.1001-8166.2018.09.0885

Cheng, W. M., Zhou, C. H., Li, B. Y., Shen, Y., and Zhang, B. (2011). Structure and contents of layered classification system of digital geomorphology for China. *J. Geogr. Sci.* 21 (5), 771–790. doi:10.1007/s11442-011-0879-9

Cutts, J. A., and Blasius, K. R. (1981). Origin of Martian outflow channels: the Eolian hypothesis. *J. Geophys. Res.* 86 (NB6), 5075–5102. doi:10.1029/jb086ib06p05075

Deng, J. Y., Cheng, W. M., Liu, Q. Y., Jiao, Y. M., and Liu, J. Z. (2022). Morphological differentiation characteristics and classification criteria of lunar surface relief amplitude. *J. Geogr. Sci.* 32 (11), 2365–2378. doi:10.1007/s11442-022-2052-z

Di, K. C., Liu, B., and Liu, Z. Q. (2018). Review and prospect of Mars mapping technique using remote sensing data. *Spacecr. Eng.* 27 (1), 10–24. doi:10.3969/j.issn.1673-8748.2018.01.002

Di, K. C., Ye, L. J., Wang, R. Z., and Wang, Y. (2021). Advances in planetary target detection and classification using remote sensing data. *Nat. Remote Sens. Bull.* 25 (1), 365–380. doi:10.11834/jrs.20210231

Ding, M., Cao, Y. F., and Wu, Q. X. (2013). Novel approach of crater detection by crater candidate region selection and matrix-pattern-oriented least squares support vector machine. *Chin. J. Aeronaut.* 26 (2), 385–393. doi:10.1016/j.cja.2013.02.016

Domingos, P. (2012). A few useful things to know about machine learning. *Commun. ACM* 55 (10), 78–87. doi:10.1145/2347736.2347755

Dong, Z. B. (2020a). *Aeolian geomorphologic map of mars*. Scale 1:12,000,000. Xi'an Map Press.

Dong, Z. B., Lu, P., and Li, C. (2020b). Research methodology of Martian aeolian geomorphology. *Adv. Earth Sci.* 35 (8), 771–788.

Dundas, C. M., Mellon, M. T., McEwen, A. S., Lefort, A., Keszthelyi, L. P., and Thomas, N. (2008). HiRISE observations of fractured mounds: possible Martian pingos. *Geophys. Res. Lett.* 35, L04201. doi:10.1029/2007GL031798

Duxbury, T. C., Krik, R. L., and Archinal, B. A. (2002). *Mars geodesy/cartography working group recommendations on mars cartographic constants and coordinate systems//Proceeding of the ISPRS Commission IV Symposium "Geospatial Theory, Processing and applications"*. Ottawa: ISPRS Comission IV.

Fenton, L. K., Bandfield, J. L., and Ward, A. W. (2003). Aeolian processes in Proctor Crater on Mars:Sedimentary history as analyzed from multiple data sets. *J. Geophys. Res. Planets* 108 (E12), 5129. doi:10.1029/2002JE002015

Galofre, G. A., Whipple, K. X., Christensen, P. R., and Conway, S. J. (2022). Valley networks and the record of glaciation on ancient Mars. *Geophys. Res. Lett.* 49, e2022GL097974. doi:10.1029/2022GL097974

Gao, A., Zhou, Y. J., and Wang, J. W. (2022). Lightweight deep learning method for Lunar surface crater detection. *J. Astron.* 43 (6), 830–838. doi:10.3873/j.issn.1000-1328.2022.06.014

Geomorphic Map Editorial Committee of the People's Republic of China (2009). *Geomorphic atlas of the People's Republic of China (1:100000)*. Science Press.

Gou, S., Yue, Z. Y., Di, K. C., and Xu, Y. (2018). Quantitative comparison of morphometric and hydrological characteristics of valley networks between Evros Vallis on Mars and Kaidu River in Tarim Basin as terrestrial analog. *J. Remote Sens.* 22 (2), 313–323. doi:10.11834/jrs.20187014

Gou, S., Yue, Z. Y., Di, K. C., Zhao, C., Bugiolacchi, R., Xiao, J., et al. (2022). Transverse aeolian ridges in the landing area of the tianwen-1 Zhurong rover on utopia Planitia, Mars. *Earth Planet. Sci. Lett.* 595 (10), 117764. doi:10.1016/j.epsl.2022.117764

Guyon, I., and Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3, 1157–1182. doi:10.5555/944919.944968

Hartmann, W. K., and Neukum, G. (2001). Cratering chronology and the evolution of Mars. *Space Sci. Rev.* 96 (1-4), 165–194. doi:10.1023/a:1011945222010

Hayward, R. K., Mullins, K. F., Fenton, L. K., Hare, T. M., Titus, T. N., Bourke, M. C., et al. (2007). Mars global digital dune database and initial science results. *J. Geophys. Res. Planets* 112 (E11), E11007. doi:10.1029/2007je002943

Hepburn, A. J., Ng, F. S. L., Holt, T. O., and Hubbard, B. (2020). Late amazonian ice survival in kasei valles, Mars. *J. Geophys. Res. Planets* 125, e2020JE006531. doi:10.1029/2020JE006531

Hsu, C. Y., Li, W., and Wang, S. (2021). Knowledge-Driven GeoAI: integrating spatial knowledge into multi-scale deep learning for Mars crater detection. *Remote Sens.* 13 (11), 2116. doi:10.3390/rs13112116

Hynek, M. B., Beach, M., and Hoke, T. M. (2010). Updated global map of Martian valley networks and implications for climate and hydrologic processes. *J. Geophys. Res. Planets* 115 (E9), E09008. doi:10.1029/2009je003548

Institute of Geography, Chinese Academy of Sciences (1987). *China's 1:100000 topographic map mapping specification[S]*. Science Press.

Jiang, H. K., Tian, X. L., and Xu, A. (2013). An automatic algorithm for detecting lunar impact craters in a defined feature space. *Sci. Sin-Phys Mech. As* 43 (11), 1430–1437. doi:10.1360/132013-321

Jiang, S. C., Wu, F., Yung, K. L., Yang, Y., Ip, W., Gao, M., et al. (2021). A robust end-to-end deep learning framework for detecting Martian landforms with arbitrary orientations. *Knowl. Based. Syst.* 12, 107562. doi:10.1016/j.knosys.2021.107562

Kerber, L., Head, J. W., Madeleine, J.-B., Forget, F., and Wilson, L. (2011). The dispersal of pyroclasts from apollinaris patera, Mars: implications for the origin of the medusae fossae formation. *Icarus* 216 (1), 212–220. doi:10.1016/j.icarus.2011.07.035

Kim, J. R., Muller, J. P., Van Gasselt, S., Morley, J. G., and Neukum, G. (2005). Automated Crater detection, A new tool for Mars cartography and chronology. *Photogramm. Eng. Remote Sens.* 71 (10), 1205–1217. doi:10.14358/pers.71.10.1205

Kim, J. R., and Muller, J. P. (2008). Very high resolution stereo DTM extraction and its application to surface roughness estimation over Martian surface. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 37 (B4), 993–998. doi:10.14358/PERS.71.10.1205

Lakdawalla, E. (2005). *The planetary society. Map of Mars with major regions labeled.* Available at: https://www.planetary.org/space-images/map-mars-major-features.

Lefort, A., Russell, P. S., and Thomas, N. (2010). Scalloped terrains in the peneus and amphitrites paterae region of Mars as observed by HiRISE. *Icarus* 205 (1), 259–268. doi:10.1016/j.icarus.2009.06.005

Lei, M. L. (2017). *Indexes of lunar impact craters based on the morphology characteristics*. Nanjing: Nanjing Normal University.

Levy, J., Head, J. W., and Marchant, D. R. (2010). Concentric crater fill in the northern mid-latitudes of Mars: formation processes and relationships to similar landforms of glacial origin. *Icarus* 209 (2), 390–404. doi:10.1016/j.icarus.2010.03.036

Li, C., Dong, Z. B., Lü, P., Zhao, J., Fu, S., Feng, M., et al. (2020). A morphological insight into the Martian dune geomorphology. *Chin. Sci. Bull.* 65, 80–90. doi:10.1360/TB-2019-0168

Li, H. L., Li, J., and Ren, X. (2022). Deep learning eliminates massive dust storms from images of tianwen-1. *Comput. Vis. Pattern Recognit.* ArXiv. Preprint. Available at: https://arxiv.org/abs/2206.10145.

Li, L. L. (2018). *A study of Martian Yardang landforms*. Shaanxi Normal University.

Li, W., Di, K. C., Yue, Z. Y., Liu, Y., and Sun, S. (2015). Automated detection of martian gullies from HiRISE imagery. *Photogramm. Eng. Remote Sens.* 81 (12), 913–920. doi:10.14358/pers.81.12.913

Liu, D. Y., Chen, M., Qian, K. J., Lei, M., and Zhou, Y. (2017). Boundary detection of dispersal impact craters based on morphological characteristics using lunar digital elevation model. *IEEE J-STARS* 10 (12), 5632–5646. doi:10.1109/jstars.2017.2749403

Liu, D. Y., Cheng, W. M., Qian, Z., Deng, J. Y., Liu, J. Z., and Wang, X. M. (2023). Boundary delineator for martian crater instances with geographic information and deep learning. *Remote Sens.* 15, 4036. doi:10.3390/rs15164036

Liu, J., Di, K. C., and Gou, S. (2020). Mapping and spatial statistical analysis of Mars Yardangs. *Planet. Space Sci.* 192, 105035. doi:10.1016/j.pss.2020.105035

Liu, J. (2021). *Global mapping and formation mechanism study of Mars yardangs*. Beijing: University of Chinese Academy of Sciences.

Liu, J., Yue, Z. Y., Di, K. C., Gou, S., and Niu, S. (2021a). A study about the temporal constraints on the martian yardangs' development in medusae fossae formation. *Remote Sens.* 13, 1316. doi:10.3390/rs13071316

Liu, J. Z., Ouyang, Z. Y., and Li, C. L. (2006). *A preliminary study on the scientific objectives and optimization principles of mars exploration//improve the scientific quality of the whole people and build an innovative country – proceedings of the 2006 annual conference of the Chinese association for science and Technology*, Volume II.

Liu, Q. Y., Cheng, W. M., and Yan, G. J. (2022). Distribution characteristics and classification schemes of lunar surface elevation. *Acta Geogr. Sin.* 76 (1), 106–119. doi:10.1016/j.epsl.2022.117785

Liu, Y., Wu, X., Liu, Z. H., Zhou, Q., and Chen, X. (2021b). Geological evolution and habitable environment of Mars: progress and prospects. *Rev. Geophys. Planet. Phys.* 52 (4), 416–436. doi:10.1038/d41573-021-00080-0

Liu, Y., Liu, Z. H., Wu, X., Qin, L., Wu, Y. H., Zhang, C. L., et al. (2021c). Evolution of water environment on Mars. *Acta Geologica Sinica* 95 (9), 2725–2741. doi:10.19762/j.nki.Izhixuebao.2021270

Lu, Y. H., Miao, F., and Du, J. (2013). An automatic detection algorithm of lunar craters based on feature matching. *Sci. Surv. Map.* 38 (5), 5.

Lu, Y., KennethEdgettWu, S. B., Wang, Y., Li, Z., Michael, G. G., et al. (2022). Aeolian disruption and reworking of TARs at the Zhurong rover field site, southern Utopia Planitia, Mars. *Earth Planet Sci. Lett.* 595 (10), 117785. doi:10.1016/j.epsl.2022.117785

Luo, L., Mu, L. L., Wang, X. Y., Li, C., Ji, W., Zhao, J., et al. (2013). Global detection of large lunar craters based on the CE-1 digital elevation model. *Front. Earth Sci.* 7 (4), 456–464. doi:10.1007/s11707-013-0361-3

Luo, W., and Stepinski, T. F. (2009). Computer-generated global map of valley networks on Mars. *J. Geophys. Res.* 114, E11010. doi:10.1029/2009JE003357

Luo, Z. F., Kang, Z. Z., and Liu, X. Y. (2014). The automatic extraction and recognition of lunar impact craters fusing CCD images and DEM data of Chang'e-1. *Acta Geod. Cartogr. Sin.* 43 (9), 924–930. doi:10.13485/j.cnki.11-2089.2014.0137

Mandt, K., and LeoneYardang, G. (2015). *Encyclopedia of planetary landforms*. Editors H. Hargitai, and A. Kereszturi (New York: Springer), 2340–2347.

Milliken, R. E., Mustard, J. F., and Goldsby, D. L. (2003). Viscous flow features on the surface of Mars: observations from high-resolution Mars Orbiter Camera (MOC) images. *J.Geophys. Res.* 108 (E6), 5057. doi:10.1029/2002je002005

Molly, I., and Stepinski, T. F. (2007). Automatic mapping of valley networks on Mars. *Comput. Geosci.* 33, 728–738. doi:10.1016/j.cageo.2006.09.009

Neukum, G., König, B., and Arkani-Hamed, J. (1975). A study of lunar impact crater size-distributions. *moon* 12 (2), 201–229. doi:10.1007/bf00577878

Orosei, R., Lauro, S. E., Pettinelli, E., Cicchetti, A., Coradini, M., Cosciotti, B., et al. (2018). Radar evidence of subglacial liquid water on Mars. *Science* 361 (6401), 490–493. doi:10.1126/science.aar7268

Ouyang, Z. Y., and Xiao, F. G. (2011). Major scientific issues involved in Mars exploration. *Spacecr. Environ. Eng.* 28 (3), 205–217. doi:10.3969/j.issn.1673-1379.2011.03.001

OuYang, Z. Y., and Zou, Y. L. (2015). *Introduction to martian science*. Shanghai Science and Technology Education Press.

Pedersen, G. B. M., and Head, J. W. (2010). Evidence of widespread degraded Amazonian-aged ice-rich deposits in the transition between Elysium Rise and Utopia Planitia, Mars: guidelines for the recognition of degraded ice-rich materials. *Planet. Space Sci.* 58 (14-15), 1953–1970. doi:10.1016/j.pss.2010.09.019

Petersen, L. E., Holt, W. J., and Levy, S. J. (2018). High ice purity of Martian lobate debris aprons at the regional scale: evidence from an orbital radar sounding survey in Deuteronilus and Protonilus Mensae. *Geophys. Res. Lett.* 45, 11595–11604. doi:10.1029/2018gl079759

Piqueux, S., Buz, J., Edwards, C. S., Bandfield, J. L., Kleinböhl, A., Kass, D. M., et al. (2019). Widespread shallow water ice on Mars at high latitudesand midlatitudes. *Geophys. Res. Lett.* 46 (24), 14290–14298. doi:10.1029/2019GL083947

Qian, Z., Min, C., Teng, Z., Fan, Z., Rui, Z., Zhang, Z. X., et al. (2022). Deep Roof Refiner: a detail-oriented deep learning network for refined delineation of roof structure lines using satellite imagery. *Int. J. Appl. Earth Obs.* 107, 102680. doi:10.1016/j.jag.2022.102680

Robbins, S. J., and Hynek, B. M. (2012). A new global database of Mars impact craters≥ 1 km: 1. Database creation, properties, and parameters. *J. Geophys. Res. Planets* 117 (E5). 1991 –2012. E05004. doi:10.1029/2011je003966

Robbins, S. J., and Hynek, B. M. (2010). *Progress towards a new global catalog of Martian craters and layered ejecta properties, complete to 1.5 km//Proceedings of the 41st Lunar and Planetary Science Conference*.

Rothrock, B., Kennedy, R., and Cunningham, C. (2016). *SPOC: deep learning-based terrain classification for mars rover missions*, 9. AIAA Space.

Rubanenko, L., Perez-Lopez, S., Schull, J., and Lapotre, M. G. A. (2021). Automatic detection and segmentation of barchan dunes on Mars and Earth using a convolutional neural network. *IEEE J-STARS.* 14 (11), 9364–9371. doi:10.1109/jstars.2021.3109900

Salamunićcar, G., Loncaric, S., and Lončarić, S. (2010). Method for crater detection from Martian digital topography data using gradient value/orientation, morphometry, vote analysis, slip tuning, and calibration. *IEEE Trans. Geosci. Remote Sens.* 48 (5), 2317–2329. doi:10.1109/tgrs.2009.2037750

Salamuniccar, G., Loncaric, S., and Mazarico, E. M. (2012). LU60645GT and MA132843GT catalogues of Lunar and Martian impact craters developed using a Crater Shape-based interpolation crater detection algorithm for topography data. *Planet. Space Sci.* 60, 236–247. doi:10.1016/j.pss.2011.09.003

Salamunićcar, G., Loncaric, S., Pina, P., Bandeira, L., and Saraiva, J. (2011b). MA130301GT catalogue of Martian impact craters and advanced evaluation of crater detection algorithms using diverse topography and image datasets. *Planet. Space Sci.* 59 (1), 111–131. doi:10.1016/j.pss.2010.11.003

Seidelmann, P. K. (2002). Report of the IAU/IAG working group on cartographic coordinates and rotational elements of the planets and satellites: 2000. *Celest. Mech. Dyn. Astron.* 82, 83–110. doi:10.1007/s10569-007-9072-y

Shang, C., and Barnes, D. (2013). Fuzzy-rough feature selection aided support vector machines for Mars image classification. *Comput. Vis. Image Underst.* 117 (3), 202–213. doi:10.1016/j.cviu.2012.12.002

Shen, Y. C., Su, S. Y., and Yin, Z. S. (1982). Retrospect and prospect of the research work on the classification, regionalization and mapping of the geomorphology of China. *Sci. Geol. Sin.* 2 (2), 97–105. doi:10.13249/j.cnki.sgs.1982.02.97

Silburt, A., Ali-Dib, M., Zhu, C. C., Jackson, A., Valencia, D., Kissin, Y., et al. (2019). Lunar crater identification via deep learning. *Icarus* 317 (1), 27–38. doi:10.1016/j.icarus.2018.06.022

Soare, J. R., Williams, J. P., Conway, J. S., and El-Maarry, M. R. (2021). Pingo-like mounds and possible polyphase periglaciation/glaciation at/adjacent to the Moreux impact crater. *Mars Geol. Enigmas*, 407–435. doi:10.1016/B978-0-12-820245-6.00014-8

Souness, C., and Hubbard, B. (2012). Mid-latitude glaciation on Mars. *Prog. Phys. Geogr. Earth Env.* 36 (2), 238–261. doi:10.1177/0309133312436570

Stepinski, T. F., and Collier, M. L. (2004). Extraction of Martian valley networks from digital topography. *J. Geophys. Res. Planets.* 109 (E11): E11005, doi:10.1029/2004je002269

Tanaka, L. K., Skinner, A. J., and Dohm, M. J. (2014). Geologic map of Mars: U.S. *Geol. Surv. Sci. Investig. Map* 43. 3292, scale 1:20,000,000, pamphlet. doi:10.3133/sim3292

Urbach, E. R., and Stepinski, T. F. (2009). Automatic detection of sub-km craters in high resolution planetary images. *Planet. Space Sci.* 57 (7), 880–887. doi:10.1016/j.pss.2009.03.009

Vaucouleurs, G. D., Davies, M. E., and Sturms, F. M., Jr. (1973). Mariner 9 areographic coordinate system. *J. Geophys. Res.* 78 (20), 4395–4404. doi:10.1029/jb078i020p04395

Wang, J. (2018). *Geologic characteristics of yardangs on Mars and their implications for paleo-environments: constraints from analog study between the Qaidam Basin and Aeolis-Zephyria region*. Wuhan: China University of Geosciences.

Wang, J., Cheng, W. M., Zhou, C. H., and Zheng, X. (2017). Automatic mapping of lunar landforms using DEM-derived geomorphometric parameters. *J. Geogr. Sci.* 27 (11), 1413–1427. doi:10.1007/s11442-017-1443-z

Wang, W. J., Lin, L. L., Fan, Z. J., and Liu, J. (2021). Semi-supervised learning for Mars imagery classification. *IEEE international conference on image processing (ICIP)*. doi:10.1109/ICIP42928.2021.9506533

Wang, Y., Feng, X., Zhou, H., Dong, Z., Liang, W., Xue, C., et al. (2021). Water ice detection research in utopia Planitia based on simulation of Mars rover full-polarimetric subsurface penetrating radar. *Remote Sens.* 13, 2685. doi:10.3390/rs13142685

Ward, A. W. (1979). Yardangs on Mars: evidence of recent wind erosion. *J. Geophys. Res.* 84 (B14), 8147–8166. doi:10.1029/jb084ib14p08147

Wilhelms, D. E., Mccauley, J. F., and Trask, N. J. (1987). *The geologic history of the Moon*. Washington DC: US Government Printing Office.

Wright, J., Barrett, M. A., Fawdon, P., Favaro, E. A., Balme, M. R., Woods, M. J., et al. (2022). Jezero crater, Mars: application of the deep learning NOAH-H terrain classification system. *J. Maps.* 18 (2), 484–496. doi:10.1080/17445647.2022.2095935

Xiao, L., Wang, J., Dang, Y. N., Cheng, Z. Y., Huang, T., Zhao, J. N., et al. (2017). A new terrestrial analogue site for Mars research: the Qaidam Basin, Tibetan Plateau (NW China). *Earth-Science Rev.* 164, 84–101. doi:10.1016/j.earscirev.2016.11.003

Xie, Y. Q., Tang, G. A., Yan, S. J., and Hui, L. (2013). Crater detection using the morphological characteristics of Chang'E-1 digital elevation models. *IEEE Geosci. Remote Sens. Lett.* 10 (4), 885–889. doi:10.1109/lgrs.2012.2226432

Yan, W., Ren, X., Liu, J. J., Zhang, L., Chen, W., Wang, D., et al. (2022). Topographic reconstruction of the "tianwen-1" landing area on the Mars using high resolution imaging camera images. *IEEE Trans. Geosci. Remote Sens.* 60 (9), 1–14. doi:10.1109/tgrs.2022.3206961

Yang, C., Zhao, H., Bruzzone, L., Benediktsson, J. A., Liang, Y., Liu, B., et al. (2020). Lunar impact crater identification and age estimation with Chang'E data by deep and transfer learning. *Nat. Commun.* 11 (12), 6358. doi:10.1038/s41467-020-20215-y

Yao, P. W. (2021). *Spatiotemporal distribution of dust storm activity in Tianwen-1 landing area and Mars non-polar region based on Mars remote sensing images*. Jinan: Shandong University.

Yuan, Y. F., Zhu, P. M., and Zhao, N. (2013). Automated identification of circular mare craters based on mathematical morphology. *Sci. Sin-Phys Mech. As* 43 (3), 324–332. doi:10.1360/132012-425

Yue, Z. Y., Di, K. C., Gregory, M., Gou, S., Lin, Y., and Liu, J. (2022). Martian surface dating model refinement based on Chang'E-5 updated lunar chronology function. *Earth Planet. Sci. Lett.* 595, 117765. doi:10.1016/j.epsl.2022.117765

Zhao, J. N. (2017). *Geologic characteristics of the paleolakes in Martian southern highland: implications for Martian paleo-climate and paleo-environment*. China University of Geosciences.

Zhao, J. N., Shi, Y. T., and Zhang, M. J. (2021). Advances in Martian water-related landforms. *Acta Geol. Sin.* 95 (9), 2755–2768. doi:10.19762/j.cnki.dizhixuebao.2021267

Zheng, L., Hu, W. D., and Liu, C. (2020). Large crater identification method based on deep learning. *J. B. Univ. Aeron. Astron.* 46 (5), 994–1004. doi:10.13700/j.bh.1001-5965.2019.0342

Zhou, C. H., Cheng, W. M., and Qian, J. K. (2009). *Digital geomorphical interpretation and mapping from remote sensing/*. Science Press.

Zimbelman, J. R., and Griffin, L. J. (2010). HiRISE images of yardangs and sinuous ridges in the lower member of the Medusae Fossae Formation, Mars. *Icarus* 205 (1), 198–210. doi:10.1016/j.icarus.2009.04.003

Zimbelman, J. R., and Scheidt, S. P. (2012). Hesperian age for western medusae fossae formation, Mars. *Science* 336 (6089), 1683. doi:10.1126/science.1221094

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Original Research
PUBLISHED 22 January 2024
DOI 10.3389/fspas.2024.1270079

The characteristic and size−frequency distribution of rocks at the Zhurong landing site, Mars

Xiukuo Sun[1], Shouding Li[1,2]*, Juan Li[2,3]*, Yanfang Wu[1,4], Shuo Zhang[5], Bo Zheng[1,2], Zhaobin Zhang[1,2], Tao Xu[1,2], Xinshuo Chen[1,2] and Yiming Diao[1,2]

[1]Key Laboratory of Shale Gas and Geoengineering, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China, [2]College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing, China, [3]Key Laboratory of Earth and Planetary Physics, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China, [4]School of Mathematics, North University of China, Taiyuan, China, [5]Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China

OPEN ACCESS

EDITED BY
Josep M. Trigo-Rodríguez,
Spanish National Research Council
(CSIC), Spain

REVIEWED BY
Akos Kereszturi,
Hungarian Academy of Sciences (MTA),
Hungary
Jianguo Yan,
Wuhan University, China

*CORRESPONDENCE
Shouding Li,
✉ lsdlyh@mail.iggcas.ac.cn
Juan Li,
✉ juanli@mail.iggcas.ac.cn

RECEIVED 31 July 2023
ACCEPTED 08 January 2024
PUBLISHED 22 January 2024

CITATION
Sun X, Li S, Li J, Wu Y, Zhang S, Zheng B, Zhang Z, Xu T, Chen X and Diao Y (2024), The characteristic and size−frequency distribution of rocks at the Zhurong landing site, Mars. *Front. Astron. Space Sci.* 11:1270079.
doi: 10.3389/fspas.2024.1270079

The rock characteristic and size−frequency distribution (SFD) on Mars are important for understanding the geologic and geomorphic history of the surface, for evaluating the trafficability of roving, and for planning the potential infrastructure construction. Tianwen-1, China's first autonomous Mars exploration mission, formed an excavated depression during touchdown, which has been the deepest depression on the Martian surface so far compared with others. According to the images captured using the Navigation and Terrain Cameras (NaTeCams) onboard the rover, Zhurong, the SFD of rocks is calculated and compared inside the excavated depression, within and out of the blast zone. For the first time, the rock size distribution inside the excavated depression is obtained, exposing the geological features of the shallow subsurface on Mars at a depth of tens of centimeters, which will surely be important for future drilling missions. It is found that the rock abundance in the depression is smaller than the original abundance on the surface, and the distribution of rocks in the blast zone on the surface is greatly influenced by the touchdown. In addition, based on the fractal dimension of rock sizes, the rocks (>10 mm) at the shallow subsurface of the Zhurong landing site may experience two different geological processes.

KEYWORDS

rock, size−frequency distribution, Zhurong rover, excavated depression, Mars

1 Introduction

The size–frequency distribution (SFD) of rocks has been widely investigated on the previous Mars exploration missions (Golombek et al., 1997; Golombek and Rapp, 1997; Golombek et al., 2003; Ward et al., 2005; Golombek et al., 2006; Grant et al., 2006; Golombek et al., 2008; Grant et al., 2011; Craddock and Golombek, 2016; Golombek et al., 2021; Wu et al., 2021; Chen et al., 2022; Wu et al., 2022). Since rock distributions result from the geological process, including impacts, volcanic activities, aeolian processes, and even water transportation, the SFD of rocks could be an important target for the Martian geological history (Craddock and Golombek, 2016; Chen et al., 2022) and past climate analysis (Kereszturi, 2012). In addition, it plays an important role in quantifying the hazards of spacecraft landing and evaluating the risks for rover moving (Golombek et al., 2003; Golombek et al., 2021). Furthermore, it is significant for planning the potential infrastructure construction on Mars for the future.

The Tianwen-1 probe is China's first autonomous Mars exploration mission, which was launched on 23 July 2020 and then entered the Mars orbit on 10 February 2021. After circling Mars for more than 3 months, the Zhurong rover, which was onboard the Tianwen-1 lander, finally successfully landed in the lowland area of southern Utopia Planitia on Mars on 15 May 2021. Among all the six scientific instruments mounted on the Zhurong rover, the Navigation and Terrain Cameras (NaTeCams) are used to take photographs of the Martian surface to provide support for the guidance, navigation, and control of the rover. In addition, these photographs could also be used for scientific observations. The rock SFD analysis in this paper is all based on the pictures captured using NaTeCams.

During the landing process, due to the impact of the retro-rocket engine on the Martian surface, an excavated depression was formed just below the lander. Compared with other depressions caused by the plume surface interaction (PSI), this depression is the deepest with an excavated depth of ~ 40 cm (Xu et al., 2023), exposing the shallow depth of the soil here and the geological information beneath the surface. Since the high-resolution imaging camera (HiRIC) on the orbit and NaTeCams on the ground both provide images of the surface, the high-frequency and low-frequency radar detect the zones with a depth of several meters to tens of meters below the surface, the depth range of tens of centimeters has not been detected yet, and the investigation on this exposed depression just makes up for the research gaps in this depth. Most importantly, the structure of soil and rocks at the subsurface will definitely be more important for future drilling missions (Baker and Carter, 2019; Altieri et al., 2023), such as the future ExoMars rover with 2-m-deep subsurface access (Kereszturi et al., 2016). It is inferred that subsurface sampling will also be significant in the prospective Mars exploration, for example, according to the China National Space Administration, the collection and return of samples on Mars will be accomplished around 2030, and the preliminary research has already begun (Wei, et al., 2018; Hou et al., 2019). Thus, the geological structure at this depth can provide an important basis for the selection of future landing sites (Kereszturi et al., 2016) and drilling missions.

Apart from the excavated depression, the region around the lander was also disturbed during Zhurong touchdown, which was called the blast zone. It is known that the blast zone is brighter, having extremely different photometric properties from the surroundings, which may result from the smoothing of the surface and the redistribution of fine particles (Clegg et al., 2012). This means that small particles in the blast zone might be blown away by the plume, which affected the rock distribution here.

Before Zhurong went dormant, it had driven 1,921 m in total for the last 361 Sols and kept taking photographs along its route. These photographs enlarge our view of the southern Utopia Planitia and provide more information about the Martian surface, including the rock characteristics and distributions. Therefore, in this paper, the rock SFD is calculated and compared inside the excavated depression, within and out of the blast zone, and the influence of touchdown is also analyzed. Finally, we attempt to discuss the fractal dimension of rock sizes and its potential indication briefly and preliminarily.

2 Methods

Below the lander, there is a depression excavated by the retrorocket during touchdown. The NaTeCams mounted on the Zhurong rover are binocular stereo cameras, which can capture a pair of photographs of the observed object through their two camera lenses, CamB and CamA, from slightly different perspectives at the same time. Such a pair of photographs of the depression (shot on sol 12) is shown in Figures 1A, B. After the adjustment of contrast, the depression of the pictures is enlarged so that it can be seen more clearly (Figures 1C, D). It is worth noting that only one large columnar-like rock block with a height > 15 cm (Xu et al., 2023) stands separately in the deep center of the depression and that other small rocks are all embedded in the wall.

The calculation of the rock size and area includes four steps. The first step is the rock identification. Here, the regions of the excavated depression and inside rocks are outlined manually by brown and red lines (Figure 1C) for preparation. Second, based on the digital image correlation (DIC) algorithm (Blaber, 2023), the same depression region and rocks in the CamA image are matched automatically (Figure 1D). The size of the whole picture (Figures 1A, B) is 2048 × 2048 pixel. During automatic matching, a smaller area shown in Figure 1A, which is 300 × 300 pixel covering the depression, is selected. After tracking the features of this small area using the DIC algorithm, the matching area is shown in Figure 1B, and the displacement between them is calculated. According to this displacement of the small area, the displacement of each pixel within this area is also obtained, and the matching pixels in Figure 1B are found one by one. Based on the automatic matching of pixels, the third step is followed subsequently. According to the stereo vision 3D reconstruction method (Zhang et al., 2017; Zhang et al., 2020; Zhang et al., 2022), the 3D coordinates (X, Y, and Z) in the CamB coordinate system of each pixel with the corresponding root mean square error are calculated. To ensure the accuracy of the calculation, the pixel coordinates whose root mean square errors are larger than 1.0 mm were deleted and replaced based on the interpolation method. Furthermore, if the distance between the pixel and the CamB projection center along the optical axis (optical axis distance, represented by coordinate Z) is larger than 8.0 m, then this pixel was also excluded from the calculation. At last, the areas and sizes of rocks are computed. Since there is no notable evidence proving that the rocks are arranged in a specific direction, the observed apparent width is considered an average sample of the actual rock diameter (Golombek et al., 2021), which is the length in the horizontal direction of each rock. We chose this length as the rock size. According to the 3D coordinates, the areas of the three-dimensional surface of the outlined rocks and depression are also calculated. Thus, an important parameter reflecting the richness of rocks—rock abundance—is obtained, which equals the ratio of total rock areas to the depression area.

The blast zone (Figure 2A) is the region around the lander that was disturbed during descent (Clegg et al., 2012). For Tianwen-1, it was approximately 88.21 m wide and 129.60 m long, according to a rough estimate based on the threshold segmentation of the blast

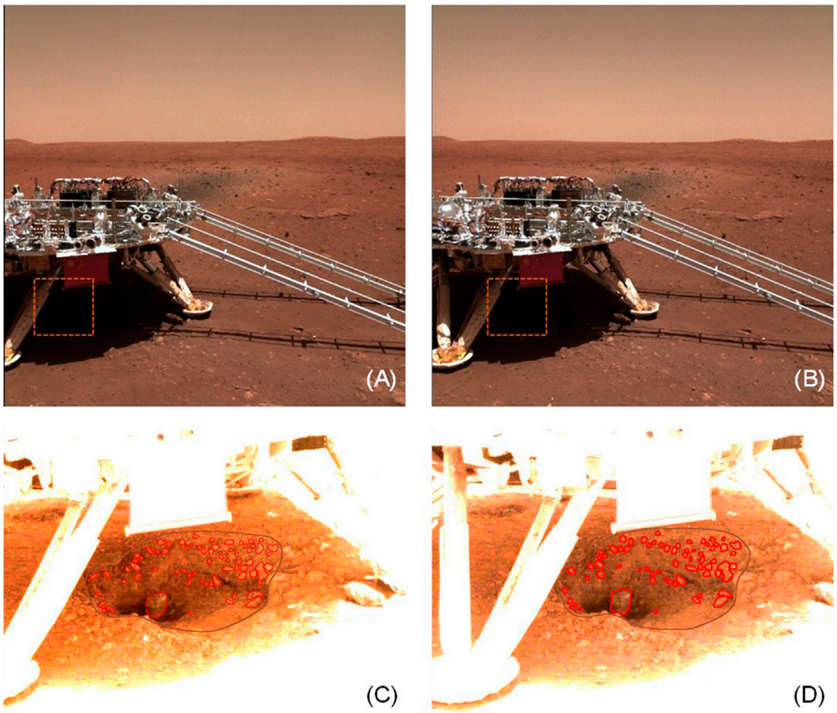


FIGURE 1
Rocks in the excavated depression of Zhurong. **(A)** Image shot by CamB; **(B)** image shot by CamA; **(C)** depression enlarged from **(A)**; and **(D)** depression enlarged from **(B)**.

zone image (Figure 2B), captured by the High Resolution Imaging Science Experiment (HiRISE). After the touchdown, Zhurong started its tour mainly toward the south on Mars. This journey began from the center of the blast zone (Site A, Figures 2A, 3A), passing through its wing (sites B and C, Figures 2A, 3B, C), and reached the edge (Site D, Figures 2A, 3D) on 13–15 June (sol 30–32). After that, the rover kept walking, passing through a sand dune (Site S, Figures 2A, 4A) and also observing some normal and typical scenes (Site N, Figures 2A, 4B). Affected by the touchdown, the characteristics of rocks at sites A, B, C, and D in the blast zone varied. For comparison, the rock characteristics in unaffected regions, for example, at sites S and N out of the blast zone, were also analyzed. The number of these image pairs captured using the NaTeCams of each site is shown in Supplementary Table S1.

The calculation of rock sizes and areas out of the depression also takes the similar four steps. The difference is that when identifying the rocks at the first step, the Cascade region-based convolutional neural network (Cascade R-CNN) (Cai and Vasconcelos, 2018) is used instead of outlining manually, given the large number of rocks in the pictures (Figures 3, 4). This is a multi-stage object detection algorithm. In this study, ResNet-50 is used as the backbone of the model, which is augmented with a feature pyramid network for robust feature extraction (Lin et al., 2017). Then, a region proposal network (RPN) and a differentiable region of interest (RoI) pooling layer (Girshick, 2015) are used to generate proposal feature maps. For training and validation purposes, a dataset comprising 4,007 images with annotated bounding boxes highlighting Martian rocks is curated. We used weights pretrained on ImageNet (Deng et al., 2009) as the initial model and fine-tuned on our datasets to obtain the final model. Based on this method (Wang et al., 2022), the rocks are identified automatically. To identify as many rocks with different diameters as possible, only the rocks in the lower half of the picture are used for recognition and calculation because the upper half is far away from the camera, and small-sized rocks are easy to be ignored. The identification results of rocks are shown in Supplementary Figures S1, S2. We checked the recognition results of the rocks in the five images by a human, and the accuracy turns out to be 94.6% on average. The picture size in Figures 3, 4 is 2048 × 2048 pixel. Among these six pictures, the longest optical axis distance from the center of the picture to the projection center of the CamB is 7.97 m in Figure 3A, and the shortest distance is 3.62 m in Figure 4A.

As the distance from the camera increases, the actual size represented by each pixel (S_p) changes regularly, as well as the size of the smallest rock (S_r) identified at each distance. For example, at the optical axis distance (D) of 3.0 m, S_p and S_r are 1.58 mm and 9.98 mm on average, respectively; while at the D value of 8.0 m, S_p and S_r both increase to 10.81 mm and 107.89 mm on average, respectively. This means that at the optical axis distance D, the rocks smaller than S_r will not be identified in this research. In the log–log plot, S_p and S_r both show linear fits with D (Equations 1, 2 and Figure 5):

$$\boldsymbol{logS_p} = \mathbf{1.963}\,\boldsymbol{logD} - \mathbf{0.739}, \tag{1}$$

$$\boldsymbol{logS_r} = \mathbf{2.427}\,\boldsymbol{logD} - \mathbf{0.159}. \tag{2}$$

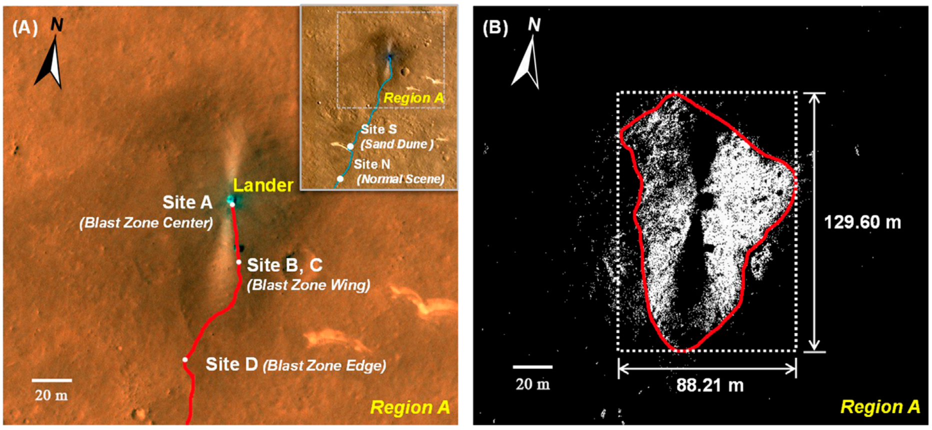


FIGURE 2
Blast zone with the Zhurong route (modified from HiRISE image "ESP_069665_2055-1"). **(A)** Blast zone with the Zhurong route projected on it. **(B)** Range of the blast zone.

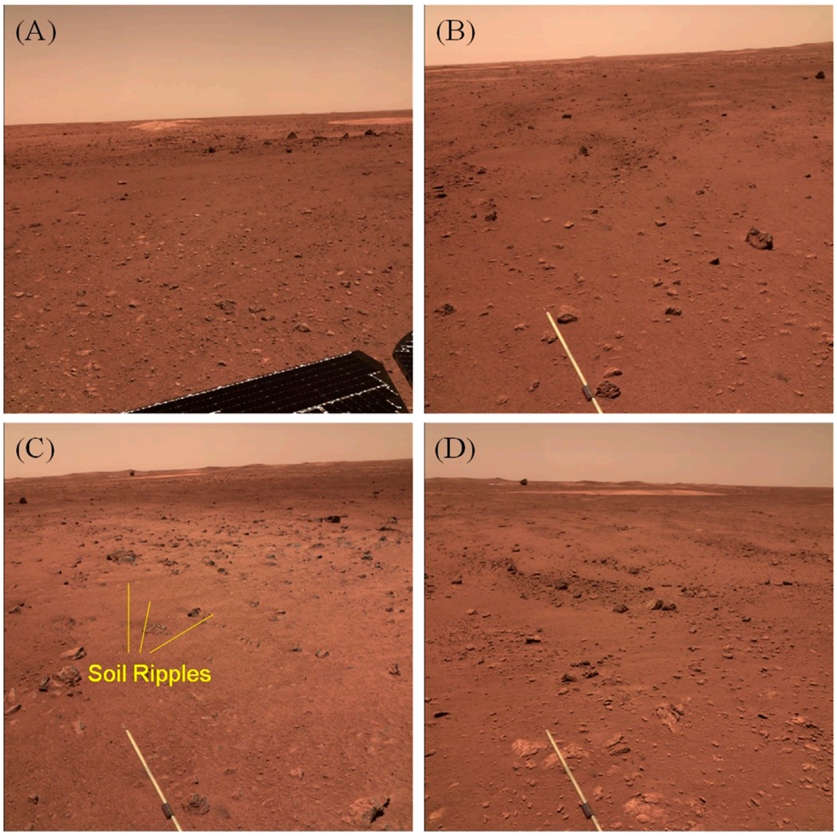


FIGURE 3
Rocks in the blast zone. **(A)** Rocks at Site A, the center of the blast zone; **(B, C)** rocks at sites B and C, the wing of the blast zone; and **(D)** rocks at Site D, the edge of the blast zone.

When calculating the areas of rocks out of the depression, we follow the hypothesis: the undersurfaces of the rocks are considered to be circles, and their diameters are the observed apparent width, as described previously. In addition, the area of the whole region is also calculated as a flat plane, and the rock abundance equals the ratio of the area of all rocks here to the area of this region.

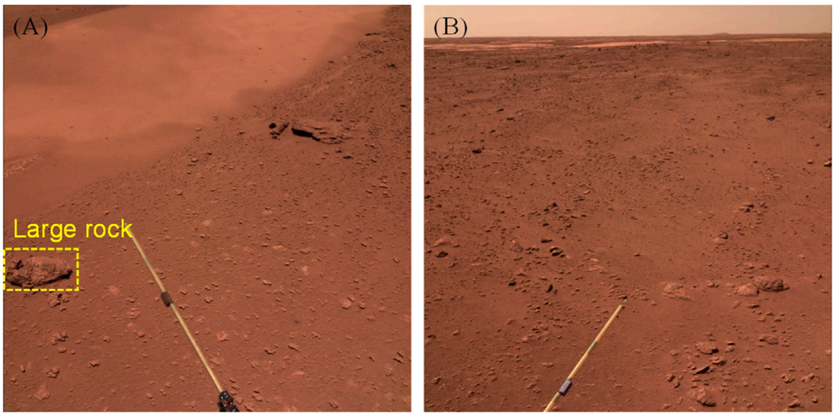


FIGURE 4
Rocks out of the blast zone. **(A)** Rocks at Site S, near a sand dune. **(B)** Rocks at Site N, a normal scene along the route.

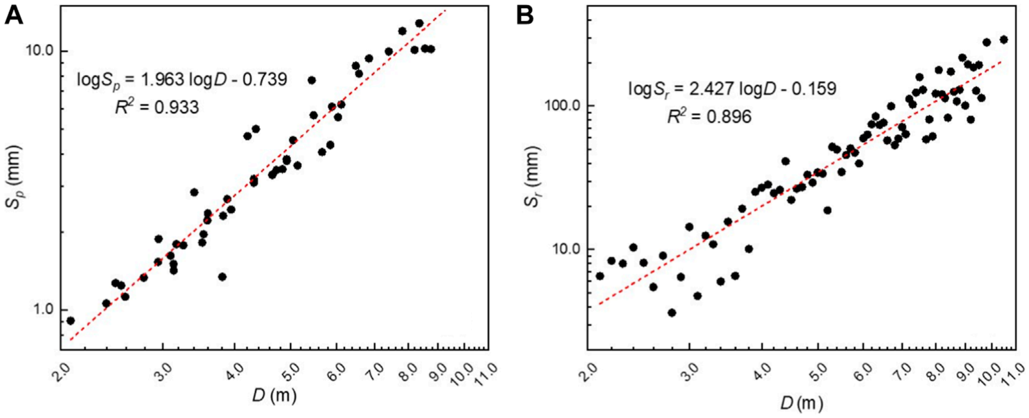


FIGURE 5
Pixel size (S_p) and the identified smallest rock size (S_r) at different optical axis distances (D). **(A)** S_p *versus* D. **(B)** S_r *versus* D.

There are some inevitable uncertainties in the calculation of rock sizes and areas. To start with, because of the limit of the image resolution, the small and more distant rocks are ignored. Second, some small rocks may be sheltered by the bigger ones. In addition, some rocks are partly buried or hidden in the depression wall or under the surface, and only the sides of exposed rocks facing the camera can be seen, which certainly affected the size and area calculation. Furthermore, the hypothesis that the undersurfaces of rocks are round may lead to inaccuracies in the calculation of the rock areas. The rock sizes and areas in different sites are shown in Table 1.

3 Results

3.1 The SFD of rocks in the excavated depression

The SFD of rocks in the excavated depression of Zhurong is shown in Figure 6 as the dark red curves. If the rock size is larger than the value shown on the horizontal axis, the cumulative fractional area (CFA) of these rocks is indicated in Figure 6A, and the cumulative number per m^2 (CN) of these rocks is indicated in Figure 6B. In order to compare and describe the distributions quantitatively, the CFA of rocks measured from the Martian surface is usually fitted as an exponential model, according to Equations 3, 4 (Golombek and Rapp, 1997; Wu et al., 2022):

$$\boldsymbol{F_k(D) = k\ exp\{-q(k)D\}}, \tag{3}$$

$$\boldsymbol{q(k) = (A + B/k)}, \tag{4}$$

where $F_k(D)$ is the CFA, meaning the cumulative fractional area of rocks whose diameter is larger than D; k is the rock abundance; $q(k)$ is the exponential factor, which defines the rate of drop-off in the exponential function at large rock diameters; and A and B are the fitting constants fitted by the data from Viking 1 and Viking 2, respectively, and $A = 1.79$ and $B = 0.152$. Therefore, the exponential model curves shown in Figure 6A reflect the CFA of different rock abundances.

TABLE 1 Rock sizes and areas in different sites.

Region	Depression	Site A	Site B	Site D	Site S	Site N
Pixel size range (mm)	3.73–5.91	1.62–10.81	1.09–6.40	1.08–8.17	0.74–2.33	0.98–4.67
Region area (m^2)	3.90	18.40	10.10	13.60	3.73	6.56
Rock number	63	428	212	372	276	360
Rock abundance (%)	4.15	4.87	4.13	4.98	7.66	6.95
Smallest rock size (mm)	12.83	10.05	10.80	10.06	10.38	10.23
Largest rock size (mm)	130.61	223.80	170.86	243.78	395.93	223.59

*The "pixel size range" refers to the variation range of the actual size represented by each pixel at different optical axis distances in the calculation area of each site.

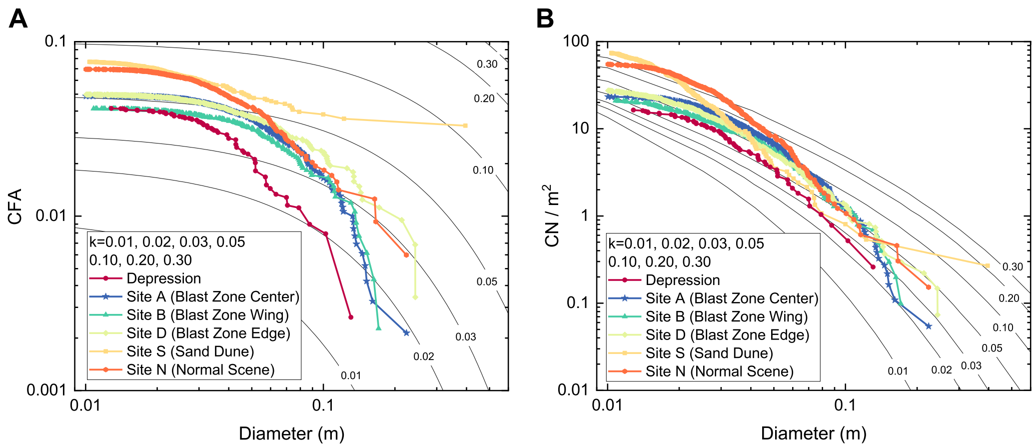


FIGURE 6
Rock size–frequency distribution (SFD) of six sites along with exponential model curves for different total CFA or k of 1%, 2%, 3%, 5%, 10%, 20%, and 30%. **(A)** Cumulative fractional area (CFA) of rocks larger than any given diameter *versus* diameter. **(B)** Cumulative number of rocks per m^2 larger than any given diameter (CN) *versus* diameter.

The rock abundance in the depression is 4.15%. The size of the rocks here ranges from 12.83 to 130.61 mm, and the largest rock block is at the bottom of the depression, whose width and height are 130.61 and 145.16 mm, respectively, with an area of 20,354.27 mm^2 (0.02 m^2). The SFD *versus* diameter in the depression conforms to the 1.4%–4.2% CFA model curve (Figure 6A). It is almost parallel to the model of 4% at relatively small rocks, while starts to gradually drop off for a rock diameter of >25.54 mm. Then, at 51.43 and 102.37 mm, this curve meets the 3% and 2% exponential models, respectively. Finally, it reaches 1.4% model at the largest diameter of 130.61 mm.

Similarly, the SFD curve in the depression also follows the CN model. The difference is that the cumulative number of SFDs is parallel to CN curves with slightly higher k than the CFA plots (Golombek et al., 2021). The smallest rock falls to the 3.0% CN model, and then, with the increase in the rock size, this curve reaches up to the 5%–10% model at the diameter of approximately 30 mm, and finally, drops to the 2%–3% model at the largest rock.

3.2 The SFD of rocks in the blast zone

The scenes of the Blast Zone Center, Wing, and Edge are shown in Figure 3. The rocks at the center of the blast zone (Figure 3A) are relatively uniform in size and spatial distribution compared with the other two sites. Since the plume surface interactions during touchdown were so strong that it dug a deep depression below the lander, the original small rocks on the surface might be blown away, and those previously in the depression were blown out and left at the Blast Zone Center. They formed the SFD features here together with the original half-buried or large rocks. However, the region of the Blast Zone Wing (Figures 3B, C) shows completely different characteristics, lacking the separate small rocks. The reason is that the plume kept blowing when the lander was adjusting its posture while selecting the suitable landing area, and during this process, the plume blew away the small original isolated rocks, leaving the large or half-buried ones. In Figure 3C, some "soil ripples" can be seen clearly on the surface, proving the influence resulted from the plume. This image shows the wing of the blast zone vividly, but the rocks are too far away from the camera, causing large errors in rock recognition and coordinate calculation, so it is only used for illustration rather than the following SFD calculation. At the edge of the blast zone (Figure 3D), the rock features are also different. Isolated small rocks could be observed again. Some of them may be the ones blown out of the blast zone, while the others are *in situ*. In summary, among these three zones, due to the remolding effect on the Martian surface of the touchdown, the characteristics of rocks change prominently, further resulting in differences in their distribution features.

The rock abundances of the Blast Zone Center, Blast Zone Wing, and Blast Zone Edge are 4.87%, 4.13%, and 4.98%, respectively, which is consistent with the previously discussed rock distribution variation in the three locations: the rocks at the Blast Zone Center contain the original large or half-buried rocks, plus the small rocks splashed away from the depression; while the rocks at the Blast Zone Wing lack the splashed ones compared with the Blast Zone Center, which shows the lowest rock abundance; and as for the Blast Zone Edge, the rocks here include the original ones and the blown-out ones from the blast zone, allowing this place the highest rock abundance.

The same variation trend of the rock SFD in these three regions is also shown in Figure 6. It can be seen from the curves that the CFA of the Blast Zone Edge is consistently larger than that of the other two regions (Figure 6A). For small rocks (the diameter of < 50 mm), the CFA of the Blast Zone Center is close to the Blast Zone Edge; for rocks with a diameter between 50 and 90 mm, it decreases, approaching the CFA of the Blast Zone Wing gradually; and then, for large rocks (the diameter of > 90 mm), the CFA of the Blast Zone Wing is a little higher than that of the Blast Zone Center, meaning that the abundance of large rocks here is higher than that of the Blast Zone Center. The cumulative number of rocks in the Blast Zone Center, Wing, and Edge is 23.3/m^2, 21.0/m^2, and 27.4/m^2, respectively, which owns the same variation as the rock abundance. Furthermore, the SFD of CN curves (Figure 6B) shows similar trends as the CFA of rocks (Figure 6A).

3.3 The SFD of rocks out of the blast zone

After completely passing through the blast zone, the Zhurong rover kept moving south, and the next regions it captured were not affected by the Tianwen-1 touchdown. In other words, the rock characteristics and SFD there can be regarded as undisturbed and original features on southern Utopia Planitia on Mars.

Two classes of typical regions are selected to analyze (Figure 4). The first one contains some part of a sand dune (Site S, Figure 4A); the rock distribution here shows a sudden transformation from the ground surface to the dune, where the rocks which can be distinguished by the naked eye abruptly disappear, replaced by tiny sand grains. The second region shows a normal scene along the Zhurong rover's route (Site N, Figure 4B). It seems that in the natural environment of Mars, there are many blocks on the ground, whose albedo is normally brighter than the soil matrix.

The rock abundance of sites S and N is 7.66% and 6.95%, respectively, and the SFD curves of these two sites are also shown in Figure 6. The CFA curve of Site N (Figure 6A) gradually drops off to a smaller exponential model, which is from 6.6% to 2.9%, with the diameter increases from 10.23 mm to 223.59 mm. However, unlike this situation, the final point of the Site S curve does not descend to a smaller model curve but goes up to meet the 10% model. This is because at the left part of Site S, there is an extremely large rock (Figure 4A), whose size is 395.93 mm, dominantly larger than other blocks, which makes the final point shift. Apart from that, the two CN curves (Figure 6B) where the rocks are smaller than 30 mm are a lot higher than other CN curves, which may illustrate that the number of these small rocks here is a lot more than that of the other four sites. This proves that the plume during touchdown does blow away small rocks and affects the rock distribution in the blast zone.

In addition to Mars, humans have also explored the Moon and asteroids and calculated the SFD there. An interesting comparison can be made about the rock distribution of them. For the rocks with a size of >50 mm, the number density is 3.0–11.5 per m^2, according to the statistics of the landing sites of Zhurong, Pathfinder, Spirit, Phoenix, and InSight on Mars (Lorenz, 2023). As for the Moon, it is 0.4–8.9 per m^2, a little lower than that on Mars, based on the statistics of the landing sites of Surveyor I, III, V, VI, VII, the Apollo 11, and Chang-E-3 (Lorenz, 2023). However, it is rather different for asteroids. Since the vast majority of asteroids between ~0.2 and 10 km in size are rubble piles, the rocks there are extremely large. For example, the largest boulder on Itokawa is more than 1/10 the diameter of the asteroid, which makes the rock density (>2.0 m) approximately 3,000 per km^2 (Pajola et al., 2022).

4 The fractal dimension of rock sizes

Fractal geometry can well-describe the regularity of irregular things in nature quantitatively, for example, the fractal characteristic of rock sizes reflects its fragmentation mode, in other words,

TABLE 2 Fractal dimension of rock sizes in the Zhurong landing site.

Calculation site	D_1	D_2	R^2_1	R^2_2	S_d (mm)
In the depression	0.43	2.06	0.963	0.976	21.72
Blast Zone Center (Site A)	1.00	2.92	0.923	0.967	52.63
Blast Zone Wing (Site B)	0.86	2.54	0.971	0.968	44.64
Blast Zone Edge (Site D)	0.42	1.72	0.989	0.965	16.83
Sand dune (Site S)	0.80	2.09	0.992	0.979	14.82
Normal scene (Site N)	0.16	1.90	0.975	0.950	13.17

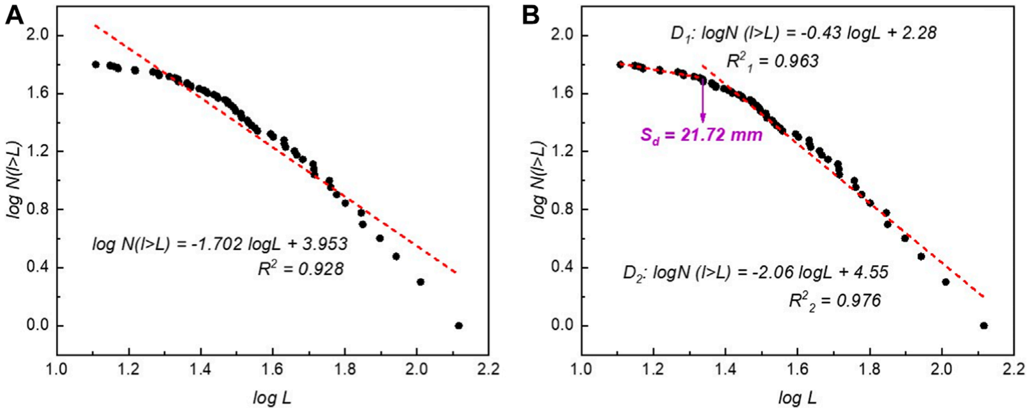


FIGURE 7
Fractal dimensions of rock sizes in the depression. **(A)** A single dimension for all rocks. **(B)** Dimensions for small and large rocks respectively.

the physical mechanism of rock evolution (Mandelbrot, 1998; Badge et al., 2002; Tu et al., 2005). This may illustrate that different fractal dimensions represent different geological processes during the formation of the rock blocks.

The fractal dimension of rock sizes could be calculated as follows: if rock size $l > L$, where L is a specific value, then the cumulative number of rocks $N(l > L)$ with the characteristic size L varies as (Turcotte, 1986; Mandelbrot, 1998; Badge et al., 2002; Tu et al., 2005)

$$\boldsymbol{N\,(l > L) \sim L^{-D}}, \tag{5}$$

where D is the fractal dimension of the size distribution of the rocks. Performing a logarithmic operation to both sides of Equation 5, it can be obtained:

$$\boldsymbol{\log N(l > L) \sim -D \log L}. \tag{6}$$

According to Equation 6, the least square method is used for linear fitting based on the rock sizes computed previously; then, the fractal dimension of rocks (D) and the corresponding adjusted determination coefficient (R^2) are obtained (Table 2). Figure 7 shows the cumulative rock number in the depression *versus* the rock size in a log–log plot. It is notable that we cannot use a single power-law exponent to characterize the particle distribution, for in that case, both the relatively smaller and larger rocks are excluded (Figure 7A). Therefore, we roughly divide the rocks at each site into small rocks and large rocks to calculate the fractal dimensions (Figure 7B). The dimension of small rocks is called D_1 with an adjusted determination coefficient $R^2{}_1$ and that of large rocks is called D_2 and $R^2{}_2$. The size that distinguishes these two kinds of rocks with different sizes is called the demarcation size (S_d). We develop a scheme to divide the rocks—that $R^2{}_1$ and $R^2{}_2$ for each place are set to be as high as possible (>0.92)—with the highest average of all possible attempts. The fractal dimensions are shown in Table 2. For the rocks inside the depression, D_1 and D_2 are 0.43 and 2.06, respectively. The difference in fractal dimensions may reflect different fragmentation processes (Charalambous, 2015), in other words, different geological processes during the formation of rocks. This means that in the depth of tens of centimeters of the Zhurong landing site, the small rocks (12.83–21.72 mm) may experience different geological processes from the large rocks (>21.72 mm).

Through the fractal dimensions of rock sizes, we can infer that different geological processes have been involved; however, it is difficult to confine what process it exactly is. On the Earth, based on the particle size distribution of soil, the depositional environments could be inferred, according to the judgment formula established by Sahu (1964). However, this work is built on a large number of statistical results of particle sizes and the corresponding known environments. Unfortunately, the depositional environments on Mars are still unclear for now, and the particle size distribution is confined to only several landing regions. Thus, it is extremely hard to determine the geological process just through Mars's rock distribution. This requires much other geological research studies and evidence, which is very important work in the future.

The fitting curves of the other five regions out of the depression are shown in Supplementary Figures S3–S7, and the fractal dimensions are also presented in Table 2. A similar phenomenon is observed. Since at each distance from the camera, the rocks smaller than a certain size cannot be identified (Figure 5), such small rocks are excluded in the calculation of fractal dimension D_1, which makes this value not precise. Thus, only the dimension D_2 is used to analyze. For the touchdown-disturbed Blast Zone Wing, D_2 is 2.54, and in the center of it, where the rocks had a greater influence by the touchdown, D_2 is as high as 2.92. For the undisturbed regions like sites S and N, D_2 is 2.00 on average, and it is similar for rocks in the depression. However, although the Blast Zone Edge is also affected by the descent, considering that it has largely maintained the original distribution of rocks, D_2 is closer to that of sites S and N.

Similarly, S_d shows the same regulation as D_2. For the Blast Zone Center and Wing, where the rock distribution is greatly influenced by the touchdown, S_d of small and large rocks is 44.64–52.63 mm; however, for sites S and N where the rock distribution is undisturbed, this size is 13.17–14.82 mm, which is a lot smaller. With regard to the Blast Zone Edge, S_d is also much closer to that of sites S and N, which is 16.83 mm. This means that the rocks in the Blast Zone Center smaller than 52.63 mm may be disturbed by the touchdown, as well as the rocks smaller than 44.64 mm in the Blast Zone Wing. In addition, the fractal dimension for large rocks at the disturbed regions is usually higher than that of the undisturbed regions.

5 Conclusion

In this paper, the rock size–frequency distributions are calculated in six regions of the Zhurong landing site, Mars, including regions inside the depression, within and out of the blast zone. Among them, the distribution of rocks in the depression is the first time to be obtained, before which the shallow subsurface on Mars with a depth of tens of centimeters has not been investigated. Apart from that, the fractal dimension is calculated to quantitatively describe the size distribution and reflect the geological process experienced by them in an attempt. The following conclusions are drawn:

1) The rocks in the depression and out of the blast zone are original, without disturbance by touchdown. The rock abundance in the depression is 4.15%, smaller than that of the surface out of the blast zone (the rock abundance at sites S and N is 7.66% and 6.95%, respectively).
2) The rocks in the blast zone are affected by the touchdown. Some small rocks are splashed from the depression to the Blast Zone Center, and some are blown away at the Blast Zone Wing; at the edge, *in situ* rocks and small blown-out rocks coexist. This makes the rock abundance first decrease then increase as getting further away from the lander.
3) At the shallow subsurface of the Zhurong landing site, the large rocks may experience different geological processes from the small rocks. The rocks in the Blast Zone Center and Wing smaller than 52.63 mm and 44.64 mm, respectively, may be disturbed by the touchdown, and the fractal dimension for large rocks in the regions affected by the touchdown is usually higher than that of the unaffected regions.

For future Mars exploration missions, the analysis of the rock SFD in this research will provide important reference, especially for landing, drilling, and sampling. According to Golombek et al. (2021), rocks whose sizes are larger than 700–900 mm are considered to pose a potential threat to landing. Based on this research, on the surface we investigate, only one rock reaches a size of ~400 mm; others are all < 250 mm. To be more conservative, even if we consider any rock larger than 400 mm to be a landing hazard, this place is still relatively safe. This could also be proved by the low density of larger rocks. The density of rocks with a size of > 100 mm is 0.5–1.3/m^2, and that for rocks > 200 mm is 0–0.3/m^2, which means the probability that the spacecraft will encounter a rock larger than 100 mm is relatively low. However, the excavated depression under the Tianwen-1 lander exposes a large rock, which might be *in situ* (Xu et al., 2023). It shows that the rock size of the shallow subsurface is not uniform, and it is likely that there are large blocks buried in it, which may cause deflection and obstruction of the drill bit during drilling and sampling.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

XS: conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, writing–original draft, and writing–review and editing. SL: conceptualization, funding acquisition, methodology, project administration, resources, supervision, and writing–review and editing. JL: conceptualization, formal analysis, investigation, methodology, project administration, supervision, and writing–review and editing. YW: data curation, methodology, and writing–review and editing. SZ: data curation, methodology, and writing–review and editing. BZ: conceptualization, investigation, and writing–review and editing. ZZ: supervision and writing–review and editing. TX: investigation, validation, and writing–review and editing. XC: validation and writing–review and editing. YD: validation and writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the National Natural Science Foundation of China (Grant 42230111), the Key Research Program of the Institute of Geology and Geophysics, CAS (Mars Mission; Grant IGGCAS-202102), the Key Research Program of the Institute of Geology and Geophysics, CAS (Grant IGGCAS-201904), and the CAS Key Technology Talent Program.

Acknowledgments

The authors thank all the scientists and engineers who were involved in the Tianwen-1 mission, including the China National Space Administration (CNSA), the National Astronomical Observatories of the Chinese Academy of Sciences (NAOC), China Academy of Space Technology (CAST), Beijing Aerospace Control Center (BACC), and other critical institutes. The data set used in this article is processed and produced by "Ground Research and Application System (GRAS) of China's Lunar and Planetary Exploration Program, provided by CNSA (http://moon.bao.ac.cn)." The TensorFlow implementation of the Cascade R-CNN model is based on Wang et al. (2022), and the authors thank Zhuowei Xiao for his assistance in training the model.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fspas.2024.1270079/full#supplementary-material

References

Altieri, F., Frigeri, A., Lavagna, M., Le, G. A., Yuryevich, N. S., Stoker, C., et al. (2023). Investigating the oxia planum subsurface with the ExoMars rover and drill. *Adv. Space Res.* 71 (11), 4895–4903. doi:10.1016/j.asr.2023.01.044

Bagde, M. N., Raina, A. K., Chakraborty, A. K., and Jethwa, J. L. (2002). Rock mass characterization by fractal dimension. *Eng. Geol.* 63, 141–155. doi:10.1016/s0013-7952(01)00078-3

Baker, D. M. H., and Carter, L. M. (2019). Probing supraglacial debris on Mars 1: sources, thickness, and stratigraphy. *Icarus Int. J. Sol. Syst. Stud.* 319, 745–769. doi:10.1016/j.icarus.2018.09.001

Blaber, J. (2023). GitHub. Avaliable at: https://github.com/justinblaber/ncorr_2D_matlab.

Cai, Z., and Vasconcelos, N. (2018). "Cascade R-CNN: delving into high quality object detection," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18-23 June 2018 (IEEE), 6154–6162.

Charalambous, C. (2015). *On the evolution of particle fragmentation with applications to planetary surfaces*. London, United Kingdom: Imperial College London.

Chen, Z., Wu, B., Wang, Y. R., Liu, S., Li, Z. J., Yang, C. Y., et al. (2022). Rock abundance and erosion rate at the Zhurong landing site in southern utopia Planitia on Mars. *Earth Space Sci.* 9, e2022EA002252. doi:10.1029/2022ea002252

Clegg, R. N., Jolliff, B. L., and Metzger, P. T. (2012). Photometric analysis of the Apollo landing sites. *Earth Space* 2012, 218–227. doi:10.1061/9780784412190.025

Craddocka, R. A., and Golombek, M. P. (2016). Characteristics of terrestrial basaltic rock populations: implications for Mars lander and rover science and safety. *Icarus* 274, 50–72. doi:10.1016/j.icarus.2016.02.042

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., and Li, F. F. (2009). "ImageNet: a large-scale hierarchical image database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20-25 June 2009 (IEEE), 248–255.

Girshick, R. (2015). "Fast R-CNN," in Proceedings of the IEEE International Conference on Computer Vision, 2015 Inter, Santiago, Chile, 07-13 December 2015 (IEEE), 1440–1448.

Golombek, M. P., Cook, R. A., Moore, H. J., and Parker, T. J. (1997). Selection of the Mars pathfinder landing site. *J. Geophys. Res. Planets* 102 (E2), 3967–3988. doi:10.1029/96je03318

Golombek, M. P., Crumpler, L. S., Grant, J. A., Greeley, R., Cabrol, N. A., Parker, T. J., et al. (2006). Geology of the Gusev cratered plains from the Spirit rover transverse. *J. Geophys. Res. Planets* 111, E02S07. doi:10.1029/2005je002503

Golombek, M. P., Haldemann, A. F. C., Forsberg-Taylor, N. K., DiMaggio, E. N., Schroeder, R. D., Jakosky, B. M., et al. (2003). Rock size-frequency distributions on Mars and implications for Mars Exploration Rover landing safety and operations. *J. Geophys. Res. Planets* 108 (E12), 8086. doi:10.1029/2002je002035

Golombek, M. P., Huertas, A., Marlow, J., McGrane, B., Klein, C., Martinez, M., et al. (2008). Size-frequency distributions of rocks on the northern plains of Mars with special reference to Phoenix landing surfaces. *J. Geophys. Res.* 113, E00A09. doi:10.1029/2007je003065

Golombek, M. P., and Rapp, D. (1997). Size-frequency distributions of rocks on Mars and Earth analog sites: implications for future landed missions. *J. Geophys. Res. Planets* 102 (E2), 4117–4129. doi:10.1029/96je03319

Golombek, M. P., Trussell, A., Williams, N., Charalambous, C., Abarca, H., Warner, N. H., et al. (2021). Rock size-frequency distributions at the InSight landing site, Mars. *Earth Space Sci.* 8, e2021EA001959. doi:10.1029/2021ea001959

Grant, J. A., Golombek, M. P., Grotzinger, J. P., Wilson, S. A., Watkins, M. M., Vasavada, A. R., et al. (2011). The science process for selecting the landing site for the 2011 Mars Science Laboratory. *Planet. Space Sci.* 59, 1114–1127. doi:10.1016/j.pss.2010.06.016

Grant, J. A., Wilson, S. A., Ruff, S. W., Golombek, M. P., and Koestler, D. L. (2006). Distribution of rocks on the gusev plains and on husband hill, Mars. *Geophys. Res. Lett.* 33, L16202. doi:10.1029/2006gl026964

Hou, X. Y., Ding, T. X., Cao, K. R., Chen, T., Li, L., Yu, Z. J., et al. (2019). Research on multi-pipe drilling and pneumatic sampling technology for deep Martian soil. *Adv. Space Res.* 64, 211–222. doi:10.1016/j.asr.2019.03.019

Kereszturi, A. (2012). Landing site rationality scaling for subsurface sampling on Mars—case study for ExoMars Rover-like missions. *Planet. Space Sci.* 72 (1), 78–90. doi:10.1016/j.pss.2012.07.007

Kereszturi, A., Bradak, B., Chatzitheodoridis, E., and Ujvari, G. (2016). Indicators and methods to understand past environments from ExoMars rover drills. *Orig. Life Evol. Biosphere* 46 (4), 435–454. doi:10.1007/s11084-016-9492-3

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). "Feature pyramid networks for object detection," in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21-26 July 2017 (IEEE), 936–944.

Lorenz, R. D. (2023). Analytic rock abundance distributions and their application to spacecraft landing hazards. *Planet. space Sci.* 234, 105720. doi:10.1016/j.pss.2023.105720

Mandelbrot, B. B., and Wheeler, J. A. (1998). The fractal geometry of nature. *Am. J. Phys.* 51, 286–287. doi:10.1119/1.13295

Pajola, M., Barnouin, O. S., Lucchetti, A., Hirabayashi, M., Ballouz, R. L., Asphaug, E., et al. (2022). Anticipated geological assessment of the (65803) didymos–dimorphos system, target of the DART–LICIACube mission. *Planet. Sci. J.* 3, 210. doi:10.3847/psj/ac880d

Sahu, B. K. (1964). Depositional mechanisms from the size analysis of clastic sediments. *J. Sediment. Petrology* 34, 73–83. doi:10.1306/74d70fce-2b21-11d7-8648000102c1865d

Tu, X. B., Wang, S. J., and Yue, Z. Q. (2005). Fractal fragmentation of weathered rock and its application in engineering Geology (in Chinese). *Chin. J. Rock Mech. Eng.* 24 (4), 587–595.

Turcotte, D. L. (1986). Fractals and fragmentation. *J. Geophys. Res.* 91 (B2), 1921–1926. doi:10.1029/jb091ib02p01921

Wang, Y., Li, G., and Xiao, Z. (2022). The local time and altitude dependences of long-lived nonspecular meteor echoes with weak magnetic aspect sensitivity at low latitudes. *J. Geophys. Res. Space Phys.* 127, e2022JA030458. doi:10.1029/2022ja030458

Ward, J. G., Arvidson, R. E., and Golombek, M. P. (2005). The size-frequency and areal distribution of rock clasts at the Spirit landing site, Gusev Crater, Mars. *Geophys. Res. Lett.* 32, L11203. doi:10.1029/2005gl022705

Wei, Y., Yao, Z., and Wan, W. (2018). China's roadmap for planetary exploration. *Nat. Astron.* 2 (5), 346–348. doi:10.1038/s41550-018-0456-6

Wu, B., Dong, J., Wang, Y. R., Li, Z. J., Chen, Z. Y., Liu, W. C., et al. (2021). Characterization of the candidate landing region for tianwen-1—China's first mission to Mars. *Earth Space Sci.* 8, e2021EA001670. doi:10.1029/2021ea001670

Wu, B., Dong, J., Wang, Y. R., Rao, W., Sun, Z. Z., Li, Z. J., et al. (2022). Landing site selection and characterization of tianwen-1 (Zhurong rover) on Mars. *J. Geophys. Res. Planets* 127, e2021JE007137. doi:10.1029/2021je007137

Xu, T., Zheng, B., Zhang, Z. B., Li, J., Li, S. D., Chen, X. S., et al. (2023). Phenomenology of plume–surface interactions and preliminary results from the Tianwen-1 landing crater on Mars. *Earth Planet. Phys.* 7, 311–330. doi:10.26464/epp2023044

Zhang, S., Jia, Y., Peng, S., Wen, B., Ma, Y. Q., Qi, C., et al. (2020). Self-Calibration of the stereo vision system of the chang'e-4 lunar rover based on the points and lines combined adjustment. *Photogrammetric Eng. Remote Sens.* 86 (3), 169–176. doi:10.14358/pers.86.3.169

Zhang, S., Liu, S. C., Ma, Y. Q., Qi, C., Ma, H., and Yang, H. (2017). Self calibration of the stereo vision system of the Chang'e-3 lunar rover based on the bundle block adjustment. *ISPRS J. Photogrammetry Remote Sens.* 128, 287–297. doi:10.1016/j.isprsjprs.2017.04.004

Zhang, S., Peng, S., Wu, Y. J., Wen, B., Zhang, J. L., Jia, Y., et al. (2022). Positioning accuracy of the Zhurong Mars rover's hand–eye system. *Photogrammetric Rec.* 37 (177), 61–86. doi:10.1111/phor.12401

Frontiers in Astronomy and Space Sciences

TYPE Correction
PUBLISHED 29 February 2024
DOI 10.3389/fspas.2024.1385545

Check for updates

OPEN ACCESS

APPROVED BY
Frontiers Editorial Office,
Frontiers Media SA, Switzerland

*CORRESPONDENCE
Shouding Li,
✉ lsdlyh@mail.iggcas.ac.cn
Juan Li,
✉ juanli@mail.iggcas.ac.cn

RECEIVED 13 February 2024
ACCEPTED 14 February 2024
PUBLISHED 29 February 2024

CITATION
Sun X, Li S, Li J, Wu Y, Zhang S, Zheng B, Zhang Z, Xu T, Chen X and Diao Y (2024), Corrigendum: The characteristic and size–frequency distribution of rocks at the Zhurong landing site, Mars.
Front. Astron. Space Sci. 11:1385545.
doi: 10.3389/fspas.2024.1385545

Corrigendum: The characteristic and size–frequency distribution of rocks at the Zhurong landing site, Mars

Xiukuo Sun[1], Shouding Li[1,2]*, Juan Li[2,3]*, Yanfang Wu[1,4], Shuo Zhang[5], Bo Zheng[1,2], Zhaobin Zhang[1,2], Tao Xu[1,2], Xinshuo Chen[1,2] and Yiming Diao[1,2]

[1]Key Laboratory of Shale Gas and Geoengineering, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China, [2]College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing, China, [3]Key Laboratory of Earth and Planetary Physics, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China, [4]School of Mathematics, North University of China, Taiyuan, China, [5]Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China

KEYWORDS

rock, size–frequency distribution, Zhurong rover, excavated depression, Mars

A Corrigendum on
The characteristic and size–frequency distribution of rocks at the Zhurong landing site, Mars

by Sun X, Li S, Li J, Wu Y, Zhang S, Zheng B, Zhang Z, Xu T, Chen X and Diao Y (2024). Front. Astron. Space Sci. 11:1270079. doi: 10.3389/fspas.2024.1270079

In the published article, there was an error in the **Acknowledgments**. The authors wish to add their **Acknowledgments** of the whole Tianwen-1 team and their data provider. The correct statement appears below:

"The authors thank all the scientists and engineers who were involved in the Tianwen-1 mission, including the China National Space Administration (CNSA), the National Astronomical Observatories of the Chinese Academy of Sciences (NAOC), China Academy of Space Technology (CAST), Beijing Aerospace Control Center (BACC), and other critical institutes. The data set used in this article is processed and produced by "Ground Research and Application System (GRAS) of China's Lunar and Planetary Exploration Program, provided by CNSA (http://moon.bao.ac.cn)." The TensorFlow implementation of the Cascade R-CNN model is based on Wang et al. (2022), and the authors thank Zhuowei Xiao for his assistance in training the model."

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

Publisher's note

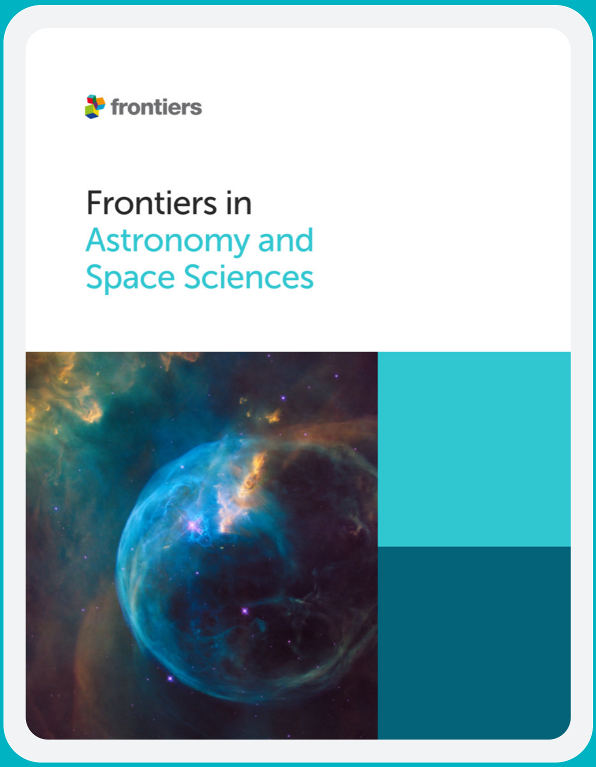
Frontiers in
Astronomy and Space Sciences

Explores planetary science and extragalactic astronomy in all wavelengths

Advances the understanding of our universe - from planetary science to extragalactic astronomy, to high-energy and astroparticle physics.

Discover the latest Research Topics

See more →


frontiers

Frontiers in
Astronomy and
Space Sciences

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact