Deep Learning driven interpretation of Chang'E4 Lunar Penetrating Radar

Roncoroni G.[1], Forte E.[1], Santin I.[1], Černok A.[1], Rajšić A.[2], Frigeri A.[3], Zhao W.[4], Pipan M.[1]

[1] Department of Mathematics and Geosciences, University of Trieste, Italy
[2] Department of Earth, Atmospheric and Planetary Sciences, Purdue University, West Lafayette, Indiana, USA
[3] Istituto di Astrofisica e Planetologia Spaziali (IAPS), Istituto Nazionale di Astrofisica (INAF), Rome, Italy
[4] Key Laboratory of Geoscience Big Data and Deep Resource of Zhejiang Province, School of Earth Sciences, Zhejiang University, Hangzhou 310058, China

Corresponding author: Giacomo Roncoroni (groncoroni@units.it)

Key Points

- Processing and analysis of a more than 1400 m long Chang'E4 Lunar Penetrating Radar profile collected on the farside of the Moon.
- For the first time a Deep Learning based algorithm is exploited on Lunar radar data to automatically extract the subsurface horizon probability.
- Improved subsurface geometry was obtained and new elements were detected, including craterform structures and related deposits.

Plain Language Summary

We provide a new analysis and interpretation of Chang'E4 Lunar Penetrating Radar collected in the Chinese mission on the farside of the Moon. Radar waves penetrated into the ground revealed new subsurface structures down to a maximum depth of about 50 m. To extract the information contained in the radar signal, we exploit a new approach based on a Deep Learning algorithm, integrated and cross-validated with some calculated signal attributes, thus limiting the subjectivity of the interpretation while providing more affordable and constrained information. In particular, more than 20 shallow craterform structures and four deeper craters have been detected for the first time. By integrating radar results with satellite-derived information, and specifically with surface photographs and detailed elevation models, we discovered that some of the observed subsurface geological units are well correlated with the present-day topography. Our new findings proved the importance of integrated analysis of Lunar data for subsurface structures identification and characterization, which is mandatory for resources evaluation and their possible future exploitation.

Abstract

We reprocessed and interpreted Chang'E-4 Lunar Penetrating Radar (LPR) data collected until 14th February 2023, exploiting a new Deep Learning-based algorithm to automatically extract reflectors from a processed radar dataset. The results are in terms of *horizon probability* and have been interpreted by integrating signal attribute analysis with orbital imagery. The approach provides more objective results by minimizing the subjectivity of data interpretation allowing to link radar reflectors to their geological context and surface structures. For the first time, we imaged dipping layers and at least 20 shallow buried crateriform structures within the regolith using LPR data. We further recognized four deeper structures similar to craters, locating ejecta deposits related to a crater rim crossed by the rover path and visible in satellite image data.

1 Introduction

The aim of the Chinese lunar landing mission Chang'E-4 (CE-4) is to unravel the causes of irregular volcanic products and regolith between the near and far side of the Moon. As a part of this mission, the Yutu-2 rover landed on 3rd January 2019, on the lunar far side, in the ancient Van Kármán crater (diameter D = 185 km; 177.5991°E, 45.4446 °S), located within the South Pole-Aitken Basin (SPA), the largest and likely the oldest impact structure on the Moon (Byrne, 2008), Fig. Sup. S1, S2. The two major scientific targets of the Yutu-2 rover are: 1) to study the mineralogy of the SPA by collecting in situ reflectance spectra; and 2) image the subsurface shallow geology using a subsurface penetrating radar system. The LPR on Yutu-2 is the first radar moving directly on the surface of the Moon's far side (Dong et al., 2021). As in the Chang'E-3 (CE-3) mission, the fundamental goal of the LPR surveys in CE-4 was the exploration of the lunar subsurface structures along the rover's path down to several tens or even hundreds of meters (Fang et al., 2014; Jia et al., 2018; Wu et al, 2019). For these reasons, in addition to reflectance spectra and several other sensors, the Yutu-2 rover is equipped with a dual frequency Lunar Penetrating Radar (LPR) with central frequencies centered at 60 and 500 MHz (CH-1 and CH-2, respectively). This is the first instrument traveling on the Moon's surface capable of radar sounding at such depths, with horizontal and vertical spatial resolutions up to about 0.1 meters.

Since it landed, the rover has been moving along an irregular path (Fig. 1, Fig. Sup. S4), segmented by many stops and turnarounds points. The initial studies focused on the first hundreds of meters of the path by applying further analysis, processing, and inversion algorithms (Giannakis et al., 2021; Wang et al., 2021; Zhou et al., 2021) before data interpretation (Dong et al., 2021; Dong et al., 2020; Lai et al., 2020; Li et al., 2021). These early studies revealed a horizontally layered subsurface with an almost constant regolith thickness of ~10–12 meters and several ejecta layers just below it, as well as deeper basalt layers (Lai et al., 2020; Li et al., 2021).

As the mission progressed, new datasets were released, and new evidence of buried structures emerged from the data, such as a paleo-crater from a meteorite impact (Zhang et al., 2021), dipping features (Feng et al., 2022), a "sandwich structure" within a paleo-crater (Zhou et al., 2022), and faults (Chen et al., 2022).

Up to now, most of the studies used visual interpretation to detect horizons by only considering the reflection amplitude, while just in two cases, single and straightforward signal attributes such as the instantaneous amplitude (Zhou et al., 2022) and signal central frequency (Feng et al., 2023) were exploited. In this way, an unavoidable subjectivity was introduced into the interpretation process, and other analyses were additionally needed to support the interpretation, in particular numerical simulation and velocity analysis (Giannakis et al., 2021; Wang et al., 2021; Li et al., 2022; Zhou et al., 2022; Chen et al., 2022). Diffraction hyperbolas analysis can be effective in estimating the EM velocity field, from which properties such as dielectric permittivity and mean density can be derived. However, there are intrinsic problems in addition to the limited number of diffractions (most of them concentrated in the shallower part of the profile, Fig. 2), their interference and their often-irregular shape. A noteworthy issue is related to the not-rectilinear travel path of the rover on the Moon's surface which features abrupt changes of direction along a highly irregular route, as well as varying speeds (Figs. Sup. S8, S10-S13.). Other studies estimated the dielectric constant from the reflection amplitude (e.g. Dong et al., 2020; Feng et al., 2023). This inversion approach is undoubtedly effective in some cases (Forte et al., 2014).

The low-frequency data of the LPR system are affected by interference phenomena first described for the CE-3 mission (Li et al., 2018) and then reported also for the CE-4 mission (Pettinelli et al., 2021). The debate is still open (Zhang et al., 2021) and some recent studies continued to exploit the low frequency dataset (Cao et al., 2023). Our work focuses on the high-frequency LPR dataset due to its high quality and potential information content emerged from the preliminary analysis.

We first address the problem of reflectors extraction by applying a new automated method based on Deep Learning techniques, which provides objective and reliable results and has proven its effectiveness in different environments, datasets, and signal-to-noise ratios (Roncoroni et al., 2022a) (Fig. 2 bottom). We then applied a combination of signal attributes that have already been successfully exploited on GPR datasets, e.g. refs. Sénéchal et al., 2000; Forte et al., 2012; Zhao et al., 2018., to further constrain and improve data interpretation.

For the first time, we here show the high-frequency LPR (CH-2) data recorded until 27th March 2023, representing the longest dataset openly available at the time of writing, adding more than 700 m to the longest high frequency profile published so far (Chen et al., 2022). We show new structures previously not considered or imaged, while summarizing or partially re-interpreting the ones already described.

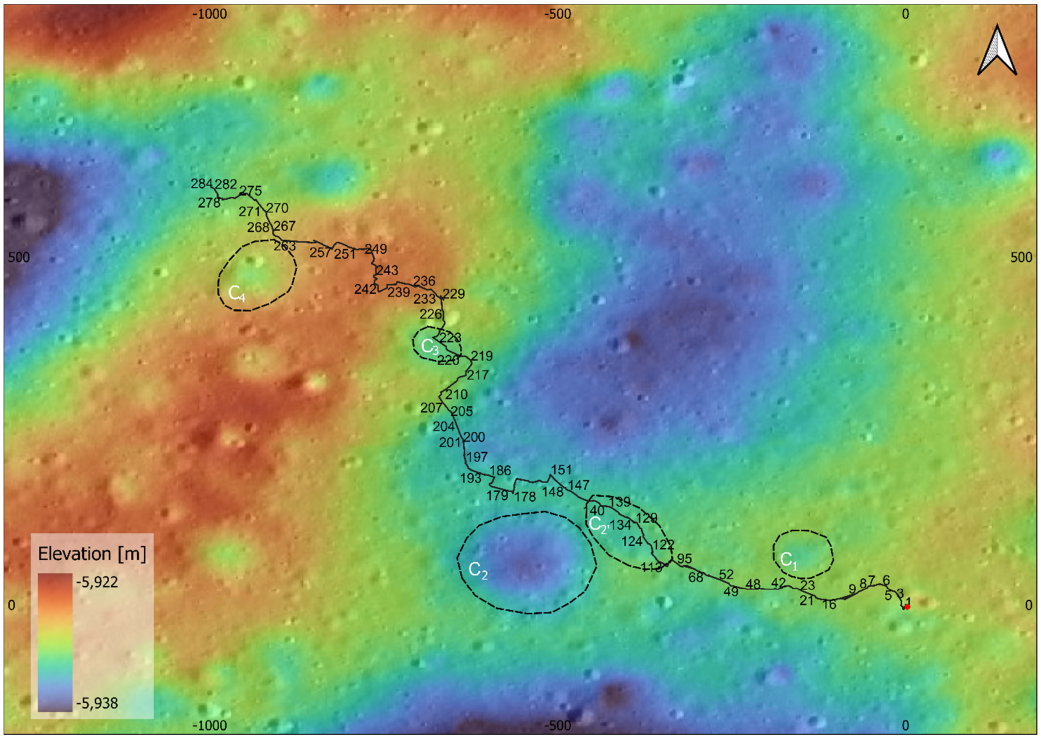

Figure 1. Rover path with waypoint numbers superimposed on a fusion of topography and satellite image. C_n refer to craters and related structures described in the text. The red dot marks the landing point. Coordinates are in meters relative to the landing point (0,0). Data from: https://quickmap.lroc.asu.edu/.

1.1 Overview of the landing site and geological context

The entire CE-4 landing region exhibits a superposition of complex impact morphologies spanning from the pre-Nectarian to Copernican epochs. The oldest structure is the SPA Basin, interpreted as one of the oldest, if not the oldest recognizable lunar basin (~4.3 Ga, Fernandes et al., 2013; White et al., 2020). The Von Kármán crater was predominantly dated as pre-Nectarian, i.e. ~4.2 Ga (Lu et al., 2021) though other references describe it as Nectarian (~4 Ga) (Feng et a., 2022). The neighbouring impacts, notably Finsen, Alder, Leibniz, Maksutov, and Von Kármán L and L' (Fig. Sup. S1), produced ejecta materials that filled in and affected the bottom of the Von Kármán crater (Lu et al., 2021; Huang et al., 2018; Chang et al., 2021). The northern and eastern parts of Von Kármán are covered by ejecta from Lebnitz and Finsen craters, respectively, while the western part is flooded by mare basalts (Lu et al., 2021; Huang et al., 2018) (Fig. Sup. S1, S2). The timing and relative sequence of these ejecta depositions and basalt flows are relevant for interpreting the local stratigraphy at the CE-4 landing site (Lai et al., 2020). However, the studies up to date have shown a persistent inconsistency in the interpretation of the local stratigraphy (Chang et al., 2021). The Yutu-2 rover LPR profiles have been interpreted to show that the post-mare deposits at the CE-4 landing site are up to ~45 m thick, while a recent article (Feng et al., 2023) suggests a shallow basaltic lava layers starting at ~10 m depth. The Finsen crater has been unequivocally described as the dominant source of ejecta that covers the landing site (Lu et al., 2021; Huang et al., 2018; Xiao et al., 2021; Xu et al., 2021). Less agreement has been reached on the exact age of the Finsen crater, as it was reported to be either Eratosthenian

(Fortezzo et al., 2020) (~3.0–3.1 Ga (Lu et al., 2021; Chang et al., 2021)) or Late Imbrian (~3.5 Ga (Gou et al., 2021) or ~3.6 Ga (Ivanov, 2018)). Further inconsistencies include the significance of the Alder crater (Imbrian (Lu et al., 2021) or Nectarian (Chang et al., 2021)) ejecta in the topmost (> 45 m) layer. While the interpretation of the early (and therefore shorter) Yutu-2 data considered it a prominent component (Lai et al., 2020), subsequent studies found Alder crater ejecta to be negligible in the topmost layer, in agreement with remote sensing interpretations (Huang et al., 2018), and to possibly only occur beneath the youngest mare basalts at greater depths (more than 50 m (Lu et al., 2021; Chang et al., 2021; Xu et al., 2021)). Most recently, four craters have been identified as principal sources of primary ejecta at the CE-4 landing site, and their most likely emplacement sequence from older to younger (Xu et al., 2021) is: Maksutov, Von Kármán L′, Von Kármán L (all late-Imbrian), and then Finsen. The ejecta delivered by larger and older impacts like Leibnitz and Schrödinger, or as distant as Imbrium or Orientale, are expected at greater depth, not accessible by CH-2, beneath mare basalts (Xiao et al., 2021). The mare basalts flooded the floor of the Von Kármán in several episodes, namely between ~3.15 and 3.75 Ga (Ling et al., 2019). Those deeper structures and stratigraphy, for instance, the oldest basalt flows, that occur at depths greater than ~50 m have been assessed using the lower-frequency CH-1 (Lai et al., 2020; Cao et al., 2023). However, their reliability is still debated (Cao et al., 2023).

2 Methods

Radar data pre-processing is a crucial step before data analysis and interpretation of any subsurface structure. In addition to the normal processing flow, that is performed also on the earth GPR data, we observed problems related to duplicated traces and data file stitching (Lai et al., 2021). Importantly, removal of redundant data is a critical step due to the acquisition system, since the rover stops to acquire other measurements like panoramic cam or visible near infrared spectra without interrupting the acquisition of LPR data. This process generates raw data with local redundancies that need to be removed. We have designed an algorithm capable of performing this removal automatically and minimizing the subjectivity of the procedure, saving time, and avoiding residual duplications (Fig. Sup. S13). The entire algorithm is available at https://figshare.com/s/36c46ad26ab1aadcfcd7.

Moreover, data acquired on different days are stored separately in different files (SOL) and need to be merged to get the full dataset. 634,419 A-scans (i.e. traces) for a total length of the path equal to ~1440 m within the SOL range (Lunar days) between 01 (4th January 2019) and 286 (27th March 2023) have been released at the moment of writing (August 2023) and are downloadable at https://moon.bao.ac.cn/ce5web/moonGisMap.search (Table Sup. S1 provides the list of all the used original files).

Beside standard processing steps, one of the commonly applied GPR processing algorithms is migration: its purpose is to correct for the distortions that can occur in the recorded signals due to both subsurface dipping reflectors, and diffraction of the electromagnetic waves (scattering).

The migration changes the reflector dip, location and length only if they are not horizontal, while in the latter case they are not modified anymore (Yilmaz, 2001). Since for migration the EM velocity model is the most crucial parameter, we chose not to apply it due to the rover's non-linear path (Fig. Sup. S8, S9) and also because the out-of-plane hyperbolas (Jiao et al., 2000) did not allow for retrieval of a trustful velocity model. On the other hand, migration

can surely focus diffraction hyperbolas, but such a peculiar shape is very helpful in localizing scatterers.
Therefore, the migration procedure would not allow us to retrieve better resolution on the horizon and would potentially heavily degrade the imaging of deeper horizons, for which there are no reliable constraints on the velocity model and signal degradation is expected due to border effects (Yilmaz, 2001).

For similar reasons, time-to-depth conversion was done using a constant EM velocity equal to 0.16 m/ns. It is certainly true that a more detailed velocity field could be reconstructed exploiting diffraction hyperbolas, but as previously pointed out, not without relevant and insuperable issues and limitations.

2.1 LPR Horizon extraction

For the automatic horizon extraction, Fig. 2c, we modified the workflow proposed in Roncoroni et al., 2022a and 2022b, for GPR measurements, implementing and exploiting a Neural Network that takes both the data amplitude and the cosine of the instantaneous phase, as input. The entire train model and codes can be found in https://github.com/Giacomo-Roncoroni/CE4-HrEx.

The algorithm utilizes a Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997) architecture to maintain the causality of the data and take advantage of its ability to better fit the physics behind wave propagation. The use of Bi-Directional LSTM is also employed to improve the accuracy of NN classification. The output of the NN is driven by a dense layer with two neurons and a SoftMax activation function (Mannor et al., 2005) that outputs a probability value indicating the presence of reflections as a function of time.

We trained the neural network (NN) using a synthetic dataset to eliminate potential biases arising from the field dataset and to exert full control over the NN performance through the known subsurface model that generated the training data. The reference output was represented by a binary indicator (1,0) labelling each sample as either reflection or no reflection, respectively. The first prediction output is given as a probability set, where each point is associated with a probability value indicating its likelihood of belonging to a reflecting surface. Then, to obtain the binary indicator we set a threshold. The optimum threshold is estimated by evaluating the number of points classified as reflectors at various threshold values, and selecting the sharp inflection point visible in the resulting curve. This method minimizes the subjectivity of the choice and is applied as a constant on all data.

Since we are working with a 1-D methodology, to reduce the noise effect we trained the NN to predict the whole wave package and not only its maximum phase, as performed in Roncoroni et al., 2022a. To mitigate the uncertainty in predictions, an ensemble learning strategy was employed. This strategy leverages multiple learning algorithms to achieve improved predictive outcomes (Mendes-Moreira et al., 2009). This methodology resulted in two separate predictions, which were then combined using their geometric mean, as it provided better results compared to the arithmetic mean. This can be attributed to the nature of the prediction, where probabilities in the range [0-1] are being predicted.

2.2 LPR attributes analysis

Attribute analysis is a technique used to extract features and information from GPR data to support interpretation and data analysis and at first exploited for reflection seismic data (Chopra and Marfurt, 2007). In this paper we used several attributes to get a more detailed and constrained LPR interpretation and to verify and validate the results obtained with the automated horizon extraction. In particular, we calculated:

- *Cosine of the instantaneous phase* (Chopra et Marfurt, 2007) (a.k.a. cosine of phase): it is a complex and amplitude independent attribute that clearly displays bedding (Fig. S16).
- *Dominant Frequency* (Chopra et Marfurt, 2007) it is a complex attribute, commonly used for highlighting specific events, such as abnormal attenuation and thin bed tuning (Fig. S18, S19, S20).
- *Sweetness* (Oliveros et al., 1997): It is an attribute computed by dividing the trace envelope by the square root of the instantaneous frequency (Fig. S17, S19, S20). It is able to characterize and emphasize differences between various facies.

3 Results

3.1 **Topography and DTM analysis of the landing site**

The topography of the terrain where the rover landed is dominated by sub-parallel ejecta rays interpreted to originate mainly from the Finsen crater (Fig. Sup. S3). This is reflected in the distribution of alternating ~400–500 m wide topographic low and high zones, occasionally connected by lower-lying bridging material (Fig. 1, Fig. Sup. S5). The rover has landed on a relatively high zone and the first 400 m of the path covered these ejecta-rich strata. The path then continues across a lower zone ~500 m wide that is followed by another high zone ~1,000–1,400 m along the path (Fig. 1). The final ~100 m of the path, undisclosed until now, are placed towards another low zone. This topography plays essential role in ejecta distribution, but earlier studies have not considered it.

3.2. **LPR profile interpretation and stratigraphy**

We describe different stratigraphic units interpreted by exploiting LPR data of the ~1,440 m long rover path, focusing on the first ~50 m depth. The rover path is presented in Fig. 1 (the landing site is marked by the red dot), and the radar profile in Fig. 2 (the landing site is on the left side). In discussing the results, we outline the general stratigraphic units as obtained by automated Neural Network extraction integrated by radar attributes evaluation (see Methods).

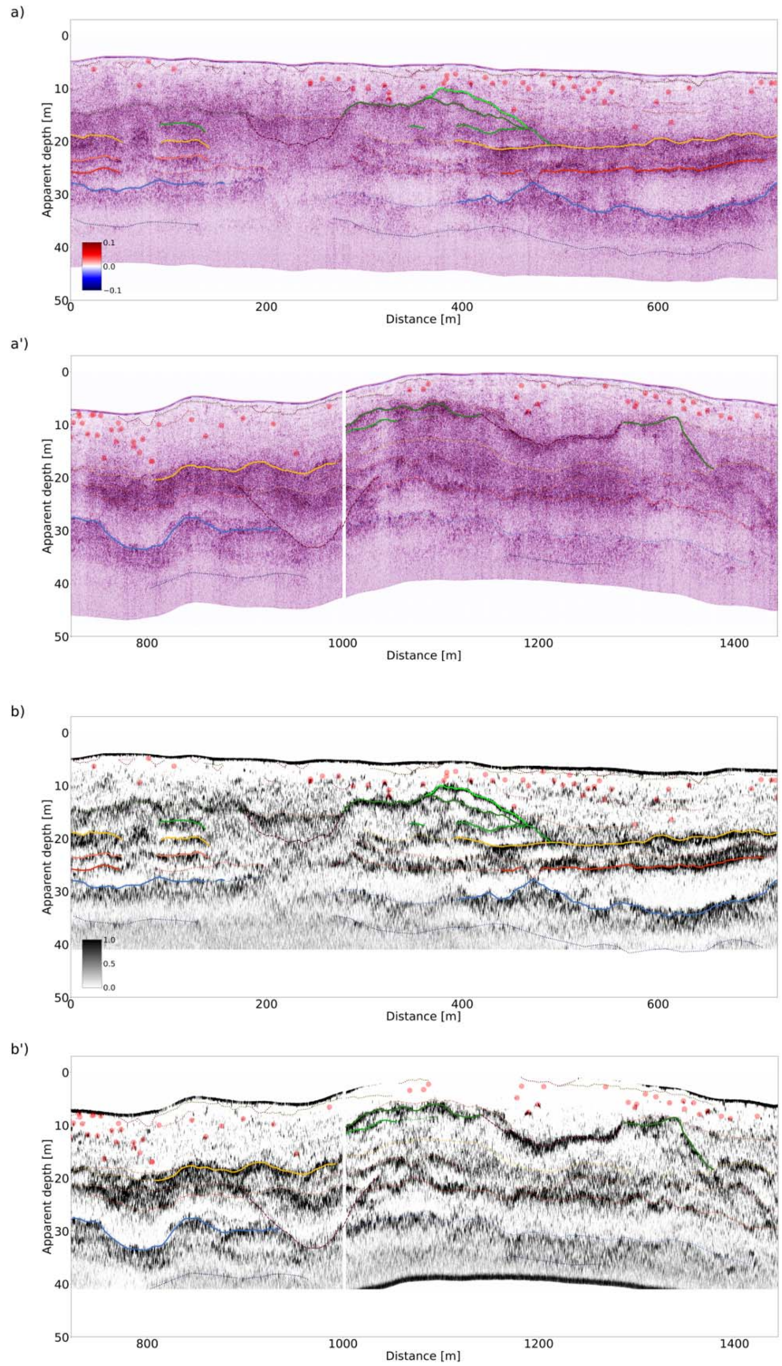

Figure 2. LPR total interpreted LPR dataset in amplitude (a, a') and automated Deep Learning horizons extraction (b, b'). Light red dots represent localized scatterers, while continuous, dashed and dotted lines follow the main recognized reflectors.

The new Deep Learning horizon extraction (Fig. 2b) interpreted some reflectors which are almost continuous along the entire profile, as well as other horizons present only in specific locations. By integrating the reflector probability (Fig. 2b) with the reflection amplitude (Fig. 2a) and integrated attribute analyses (Fig. 3; Fig. S15-S20) we can interpret single horizons and both their spatial correlation and facies: similar colours represent the same stratigraphic level along the radar profile (Fig. 2b, c; Fig. 3).

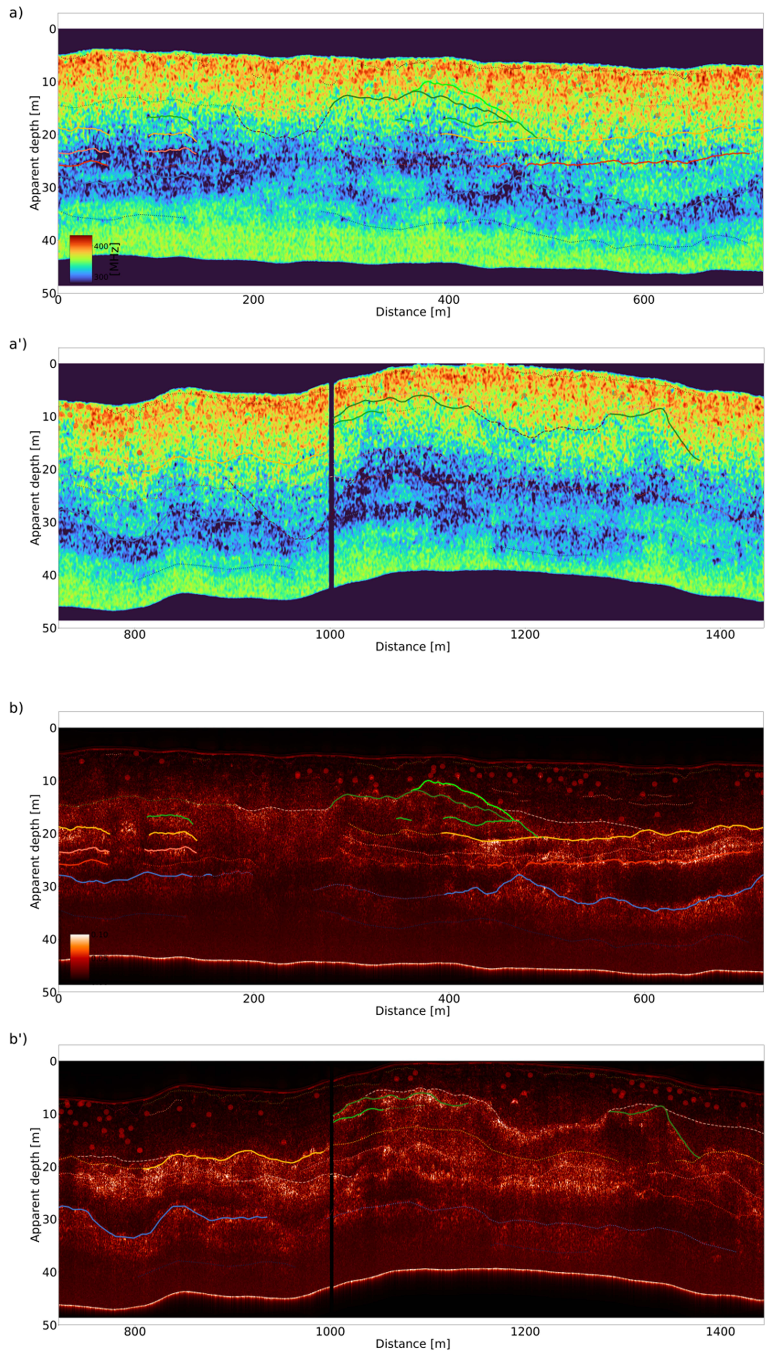

Figure 3. LPR total interpreted LPR dataset: smoothed dominant frequency (a, a'); sweetness (b, b').

Moreover, we interpreted beneath the entire rover path different electromagnetic (EM) units (U1–U4) on the basis of their EM signature, geometry and relative location, as detailed below (Fig. 4).

U1 has a low overall reflectivity without clear and coherent high amplitude reflections. This unit can be observed from the topographic surface down to an almost constant depth of ~12 m, seen from the landing site position to a distance of 275 m, decreasing down to as low as 6 m at a distance of 380 m and then approaching 15 m between 500 and 950 m (Fig. 4). Its thickness is again lower (5–10m) until 1,350 m then increases again at the end of the profile. Within U1 we imaged two new types of structures, namely: several (at least 20) with a concave shape, while several others are local sub-horizontal reflectors. None of these features have been previously reported, probably because they have an overall low signal amplitude, but appear very clear when phase or other composite attributes like *sweetness* (see Methods) are considered (Fig. 3, S17). The concave structures appear close to the surface and have a mean width equal to 16.4 m (maximum 23.2 m; minimum 8.3 m) and some of them are partially overlapping. We interpret all those structures as filled craters produced by either small meteorites or, most likely, as secondary craters that are very frequent in this area. Some of these craters appear concealed at the top with quite discontinuous but still recognizable sub-horizontal reflector. In addition to these reflectors, some other deeper and significantly longer ones have been imaged within U1 and specifically between 450 and 780 m. They show a maximum (apparent) equal to 6° between 450 and 500 m where they lie over dipping layers of U2. The maximum lateral extension of a single reflector reaches 70 m, demonstrating that the regolith is not entirely chaotic but, at least locally, layered and showing stratification that follows the former (i.e. deeper) morphology.

Stronger reflectors are beneath ~12 m depth from the beginning of the profile, including discrete horizontal and slightly dipping reflectors down to a depth of ~30 m. These layered zones can be classified into two separate stratigraphic units based on their amplitude, signature and lateral continuity (Fig. S16-S18): the top one (U2) that is ~8–10 m thick and entails two roughly equally thick layers (green in Fig. 2 and 4), and the slightly thicker beneath (~12 m, U3) that contains up to three layers (orange to red in Fig. 2, light blue in Fig. 4).

U2 is present in the first 480 m of the profile and from ~950 to 1350 m, while U3 can be observed throughout the profile. At a depth of ~30 m, a strong reflector appears (with local lower reflectivity) with significantly different characteristics from those of the facies above (see e.g. Fig. S18), and persists laterally throughout the observed profile (U4 in Fig. 4). Actual stratigraphic structure is also determined by the excavated local materials mixed with that ejecta and reworked by multiple impacts. The final stratigraphic layers rather reflect the mixture of the primary ejecta and the excavated local materials (i.e., ejecta deposits) (Xu et al., 2021) and the thicknesses of the U1–U4 layers described here are broadly in agreement with previously interpreted thicknesses. In this regard, the U1 has been interpreted as fine-grained regolith (e.g. ,Lai et al., 2020; Chen et al., 2022; Zhang et al., 2021) dominated by Finsen ejecta which was then reworked, mixed and overturned by numerous impacts but compositionally it is very similar to the ejecta itself (Dong et al., 2021; Lin et al., 2020; Guo et al., 2021). In addition to several low amplitude interfering events made clear by phase analysis (see e.g. Figs.S16), there are some localized scatterers having different amplitudes, alternatively interpreted as decimeter-sized boulders ejected during the formation of Finsen crater, including an unknown fraction of local rocks (Chen et al., 2022), or as broken pieces of glass-bearing breccia projectiles excavated from pre-existing small craters on the lunar far side (Lin et al., 2020). Some authors further divide U1 into two sub-units: the topmost is more homogeneous with weaker amplitude because the surface materials have undergone a longer weathering period, while the lower portion has a high overall reflectivity interpreted as a less weathered material (Zhang et al., 2021). This division is not apparent in our analysis, even if the shallower part of U1 seems to have higher numbers of scatterers than the deeper one. Scatterers produce diffractions hyperbolas on radar sections which have been exploited to estimate the EM velocity (and from it the dielectric permittivity) of this shallow zone (e.g. refs. (Dong et al., 2020; Dong et al., 2021; Lai et al., 2020; Chen et al., 2022) even if the rover path is not straight and its speed is not constant thus resulting in distorted hyperbolic patterns, as previously pointed out.

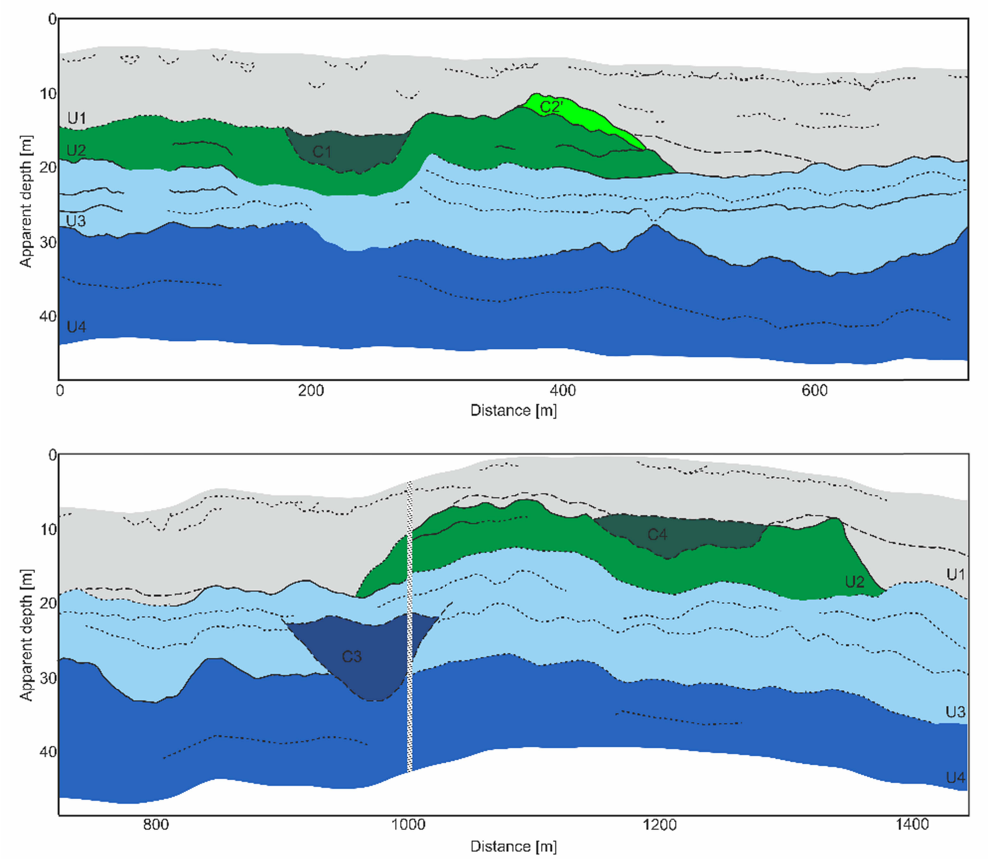

Figure 4. Subsurface units assessment from LPR data interpretation (Figs. 2, 3). U1 (grey), U2 (in green), U3 (in light blue), U4 (in blue) represent different macro units. C1, C2’, C3 and C4 are interpreted as paleo-crater filling and related materials, while dotted and dashed lines mark layering and peculiar structures within the main units (see text for description and details).

Discrete layers within U2 and U3 (~12–30 m deep) correspond to what has been previously described as different coarse ejecta deposits (i.e., the mixture zone of Finsen’s primary ejecta, pre-Finsen primary ejecta (Maksutov, Von Kármán L′, and Von Kármán L), and local basalt materials) (Xu et al., 2021).

The lowest U4 most probably represents a mare basalt layer > 30 m deep, as already pointed out by several studies (e.g. Zhang et al., 2021; Guo et al., 2021), with some discontinuous highly attenuated internal layers also due to low overall signal-to-noise ratio. This seems to be confirmed by the signal frequency behaviour (Fig. 3); however, a conclusive interpretation of such a unit is not possible just from the analysis of LPR data.

In general, the observed stratigraphy is quite in agreement with strata described earlier (Xu et al., 2021; Zhang et al., 2021; Lai et al.,2021), while notable discrepancies and new imaged structures will be discussed especially regarding U1, U2 and paleo craters and related structures C1 to C4.

4 Discussion

The landing site of Chang'E-4 shows a morphology with alternating topographic lows and highs reflecting Finsen ejecta rays, which are transected by the Yutu-2 rover (Fig. 1, 5 + supplementary). In particular, based on a fine-scale DTM map we detected four crater shapes (C1 to C4 in Fig. 1 and 5) crossed or very close to the rover path. Crater C1 is buried just below the regolith and developed within U2. It was first interpreted by Zhou et al., 2021 and then confirmed by several other authors. We estimated its maximal excavation depth (d) as a function of the crater diameter (D) using the relation d=0. 084D (Melosh, 1989; Warner et al., 2017) (Fig. 4). The obtained result of 8.8 m, being D=105 m, is in very good agreement with the interpreted paleo crater which extends to a maximum depth based on the LPR data of 8.0 m. If we consider the entire zone in which the layering is absent as the crater diameter, having an extension at its top of about 125 m, an excavation depth of 10.5 m is obtained: it matches the vertical extension in which the layers are absent i.e., from the white dashed line (top of the crater filling materials) down to the red dashed line (Fig. 2).

The same analysis was performed on C3 and C4. C3 has a diameter equal to about 142 m and an estimated excavation depth from LPR data of 12.4 m while C4 has a similar width and a depth of just 6 m. By applying the previously reported relation, we obtain a d value equal to 11.9 m which is in good agreement with C3 while it is not with C4 possible because this latter crater was modified after the main impact, as suggested also by its very irregular shape.

A similar analysis performed on the 20 shallow craters (some of them being coalescent), (Fig. 2) gives a mean diameter of 16.4 m and a consequent excavation depth of 1.4 m; also, in this case, it is quite similar to the one imaged by the LPR data, which values range from 0.9 to 2.4 m. The similar size and their close and regular spatial distribution suggest they were created as secondary craters.

A peculiar structure is labelled as C2' in Fig. 4. It lies on the top of U2 and is apparent between 360-460 m on the profile. In this portion LPR path crosses the rim of crater C2 whose center is to the south of the profile. C2 is an elliptical crater evident of the surface; it is ~145 m (North-South) by ~195 m (East-West) wide.

This smaller (max wideness in NW-SE direction equal to about 115 m) and younger crater (Fig. 6) superimposed on main crater C2 rim (estimated to be younger than 100 Ma, REF) appears quite fresh. In particular, it is apparent when considering the surface azimuth (Fig. 6c), even if the surface slope is smooth (Fig. 6d). Indeed, from the satellite imagery (Fig. 6a) this structure is not recognizable.

Very high reflectivity and reflection continuity of the LPR horizon (light green in Fig. 2) imply that it entails, at least partially, possible impact melt that was created on the rim of C2.

Deep Learning driven interpretation of the LPR data, linked with integrated attribute analysis and satellite imagery, can extract the stratigraphic horizons, correlate them spatially and group their main units, instead of obtaining by a more subjective manual line drawing. While the overall structure revealed by this method agrees reasonably well with previous observations

(Feng et al., 2022; Chen et al., 2022; Zhang et al., 2021; Lai et al., 2021) among the others), it allows for the first time the recognition of unexpected and less evident sub-surface structures. Importantly, using this method, we were able to distinguish stratigraphic units with different electromagnetic characteristics, as well as recognize their correlation with the present-day topography. For instance, U2 is not present in the central part of the LPR profile (480-950 m). Notably, based on topography and distribution of lows and highs, it can be observed that this segment of missing U2 unit corresponds to the low terrain where Finsen ejecta were originally less deposited (Fig. 5). Based on this observation, we suggest that the U2 unit, whose top is the first strong layered reflector beneath the regolith, corresponding to the top of the Finsen ejecta. At the end of the released data, U2 is no longer present since the path is approaching another low topographic area in which Finsen ejecta is not expected (Fig. 1, Fig. 5). Therefore, U2 layer can be identified as directly deposited by Finsen event.

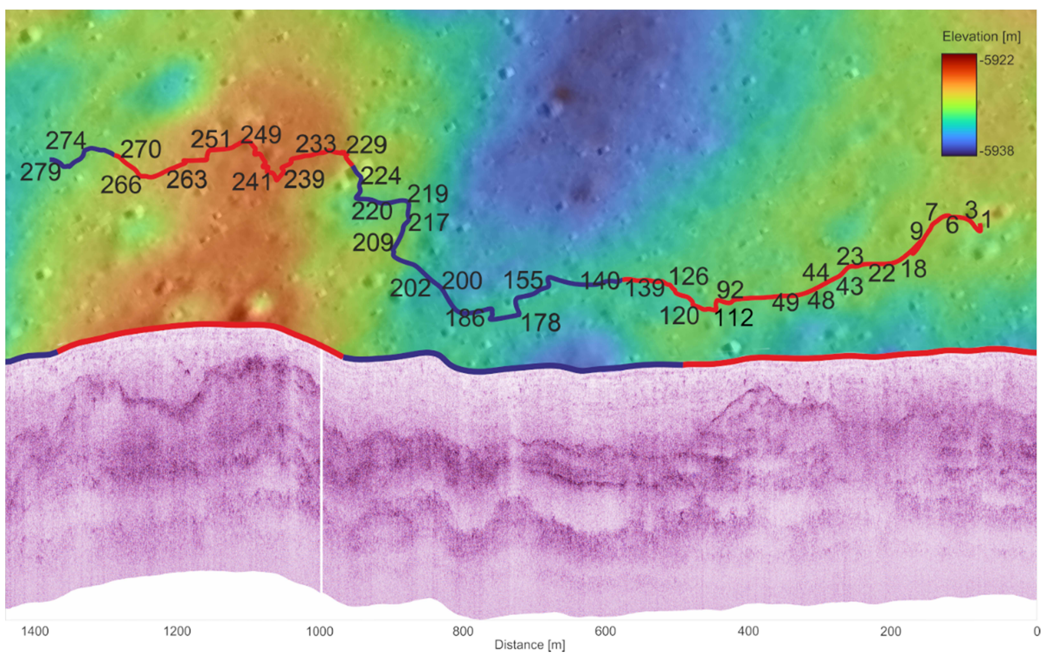

Figure 5. Correlation of surface and sub-surface structures. Plain view of the Rover path and correlation with LPR processed profile (in amplitude). Red and blue segments highlight high and low topography zones, respectively. The red dot marks the landing point. Data from: https://quickmap.lroc.asu.edu/.

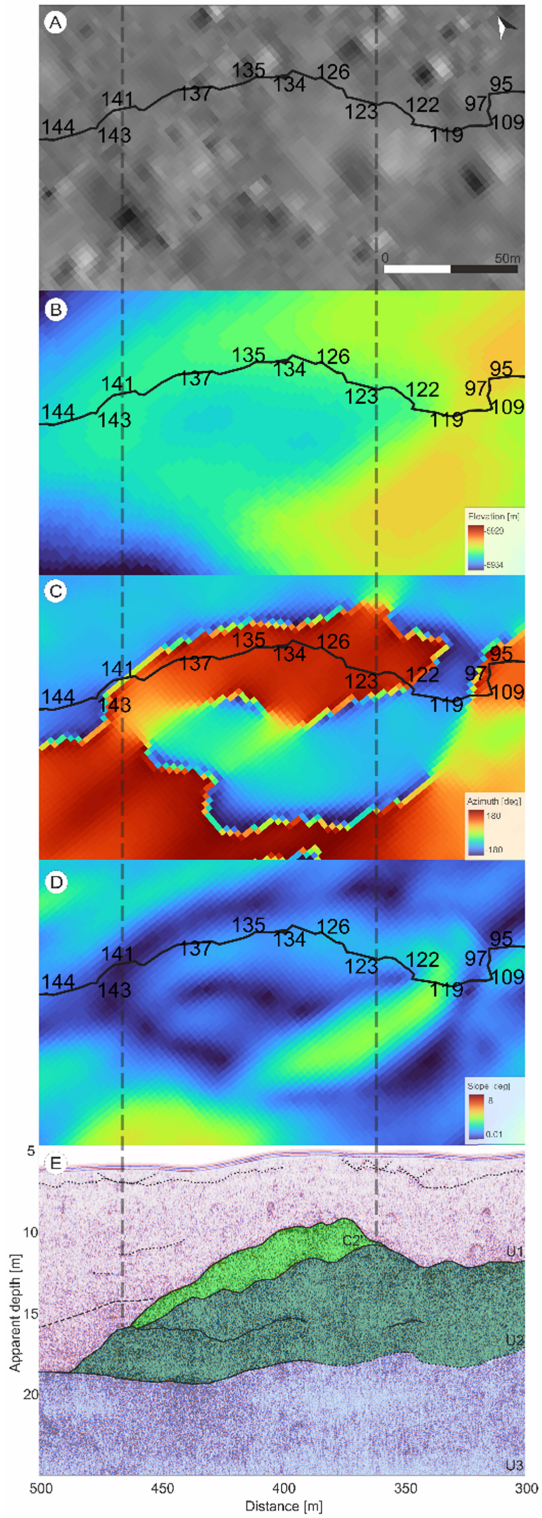

Figure 6. Analysis of the LPR portion between SOL 95 and 145 and comparison with surface morphology. a) satellite photograph; b) surface elevation; c) aspect; d) slope; e) LPR interpreted data. See discussion for details.

All available studies consider a homogeneous regolith without clear internal reflectors, only with local high amplitude scatterers with the exception of Feng et al., 2023 which did not directly recognize layering within the regolith but evidenced lateral and vertical macro electrical permittivity changes within it, on the base of an algorithm for permittivity estimation exploiting a new approach for diffraction hyperbolas fitting, limited to the shallower 150 ns (i.e. 12 m considering a constant EM velocity equal to 0.16 m/ns). In any case, their proposed model is 1-D and so, by definition, the obtained layers are perfectly parallel and horizontal. We here show not only that the layering is visible and clearly imaged on the radar profile but that it is slightly dipping and follows the deeper paleo-topography. This observation is an independent support toward the fact that the Finsen ejecta were deposited in alternating lows and highs which formed a paleo-topography of the terrain as early as ~3 Ga ago, or even earlier. The regolith layers were then deposited following this paleo-relief and were not entirely annihilated by the subsequent impacts. As layers do not show complete homogeneous mixing, it is expected that the Finsen ejecta would not completely mix with ejecta from previous craters and the layers dominated by previous ejecta may still exist; however the topmost layer is a mixture of dominantly Finsen ejecta with pre-materials, in agreement with geological mapping suggesting dominance of the Finsen ejecta in this entire portion of the VK crater, crossed by SW-NE ejecta rays from Finsen and dominated by characteristic orthopyroxene (LCP) (Huang et al., 2018), Fig. Sup. S3.

In addition to regolith internal layering, for the first time at least 20 shallow buried craters have been detected directly on the LPR dataset. They are not apparent on the base of reflection amplitude (Fig. 7) suggesting, as expected, that the filling material is very similar to the surrounding one. However, using signal attributes and in particular phase attributes (Fig. 7) the lateral limit of such a crater is evident and can be quite easily laterally recognized even when the Deep Learning horizon extraction does not clearly recognize apparent structures.

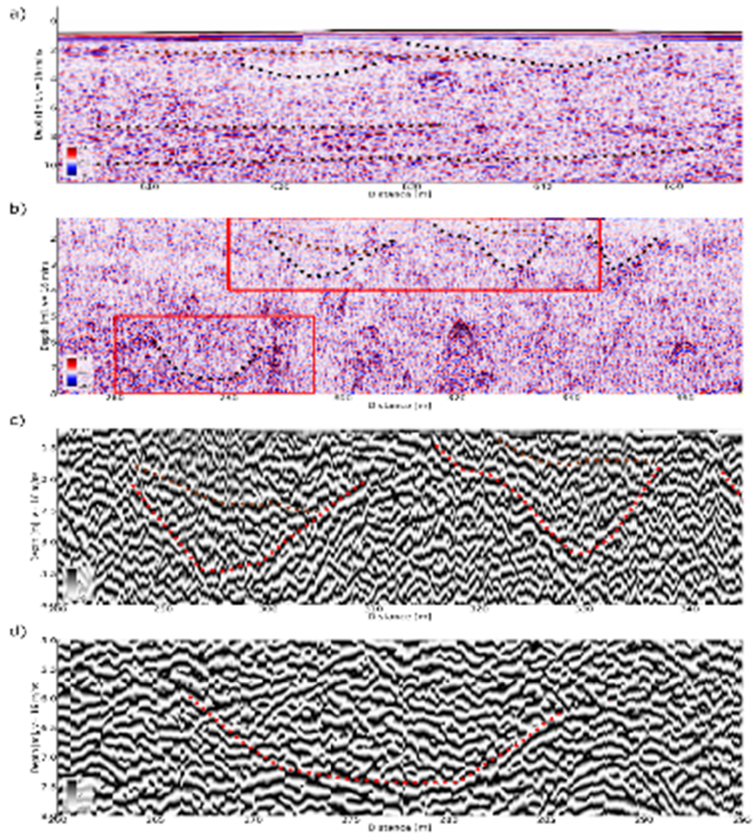

Figure 7. Details of layering and shallow crateriform structures within the regolith (U1), a) and b). c) and d) show the cosine of instantaneous phase within red boxes in b).

5 Conclusions

New LPR data and Deep Learning-based interpretation allowed to identify new and somewhat unexpected subsurface structures on the far side of the Moon. In particular, while available studies consider a homogeneous regolith with only local scatterers and no apparent reflectors, we not only show that several layers are visible on the LPR data, but also that they are not always horizontal but rather follow the deeper paleo-topography. The DL automated horizon probability procedure integrated with the analysis of combined signal attributes allowed the discovery of fine-scale features in regolith and ejecta layers which were not previously imaged, probably due to their low overall amplitude and elusive nature. In particular, we recognized at least 20 shallow buried crater-like structures within the regolith and further four developed deeper within different stratigraphic units. We made a relevant step forward in correlating layers and defining their different geological meaning. The LPR dataset was not interpreted as a stand-alone information, but it was fully integrated with satellite-derived information, and specifically to surface photographs and detailed elevation models, finding that subsurface units are well correlated with the present-day topography. This observation is an independent support toward the fact that the Finsen ejecta was deposited in alternating lows and highs which formed a paleo-topography of the terrain. Thanks to the exploited integrated data approach, we assigned specific geological and geomorphological meaning to the identified subsurface reflectors, defining four different units along the considered rover path, and describing their relationship.

The obtained results proved the importance of integrated analysis of Lunar data for subsurface assessment and structure identification, which are in turn crucial for possible resources evaluation. Further research will be addressed to the calculation of DL-based attributes.

Acknowledgments

A.Č. acknowledges Rita Levi Montalcini fellowship by the Italian Ministry of University and Research (MUR).

We thank the Chang'E-4 payload team for mission operations and China National Space Administration for providing the Chang'E-4 data that made this study possible. This work was supported by the National Natural Science Foundation of China (11773023, 11941001, U1631124) and the Civil Aerospace Pre-research Project (D020302). The Chang'E-4 data used in this work is processed and produced by "Ground Research and Application System (GRAS） of China's Lunar and Planetary Exploration Program, it can be downloaded at http://moon.bao.ac.cn/ce5web/searchOrder_dataSearchData.search

Open Research

Data can be found on 2019 DOI : https://dx.doi.org/10.12350/CLPDS.GRAS.CE4.LPR-2B-2019.vA 2020 DOI : https://dx.doi.org/10.12350/CLPDS.GRAS.CE4.LPR-2B-2020.vA. Ground Research and Application System of China's Lunar and Planetary Exploration Program. Chang'E-4 Lunar Penetrating Radar Level 2B scientific Dataset. China National Space Administration, 2020.

The codes related to the horizon extraction algorithm can be found at https://github.com/Giacomo-Roncoroni/CE4-HrEx, while the codes for data pre-processing at https://figshare.com/s/36c46ad26ab1aadcfcd7. Processed data are available at https://figshare.com/s/9ce7f1cb8ff0fb8d90c9.

References

Byrne, C. J. The South Pole-Aitken Basin and the South Polar Region. in *The Far Side of the Moon: A Photographic Guide* (ed. Byrne, C. J.) 60–93 (Springer New York, 2008). doi:10.1007/978-0-387-73206-0_8.

Cao, H. *et al.* From Schrödinger to Von Kármán: An Intriguing New Geological Structure Revealed by the Chang'e-4 Lunar Penetrating Radar. *Geophys. Res. Lett.* **50**, e2022GL101413 (2023).

Chang, Y. *et al.* An updated constraint on the local stratigraphy at the Chang'E-4 landing site. *Earth Planet. Phys.* **5**, eepp2021007 (2021).

Chen, R. *et al.* Sub-surface stratification and dielectric permittivity distribution at the Chang'E-4 landing site revealed by the lunar penetrating radar. *Astron. Astrophys.* **664**, A35 (2022).

Chopra, S. & Marfurt, K. J. *Seismic Attributes for Prospect Identification and Reservoir Characterization*. (Society of Exploration Geophysicists and European Association of Geoscientists and Engineers, 2007). doi:10.1190/1.9781560801900.

Dong, Z. *et al.* Dielectric Properties of Lunar Subsurface Materials. *Geophys. Res. Lett.* **47**, (2020).

Dong, Z. *et al.* Properties of Lunar Regolith on the Moon's Farside Unveiled by Chang'E-4 Lunar Penetrating Radar. *J. Geophys. Res. Planets* **126**, (2021).

Fang, G.-Y. *et al.* Lunar Penetrating Radar onboard the Chang'e-3 mission. *Res. Astron. Astrophys.* **14**, 1607–1622 (2014).

Feng, J., Siegler, Matthew. A. & White, M. N. Dielectric properties and stratigraphy of regolith in the lunar South Pole-Aitken basin: Observations from the Lunar Penetrating Radar. *Astron. Astrophys.* **661**, A47 (2022).

Feng, J., Siegler, M. A., Su, Y., Ding, C., & Giannakis, I. Layered structures in the upper several hundred meters of the Moon along the Chang'E-4 rover's first 1,000-m traverse. Journal of Geophysical Research: Planets, 128, e2022JE007714. https://doi.org/10.1029/2022JE007714 (2023).

Fernandes, V. A., Fritz, J., Weiss, B. P., Garrick-Bethell, I. & Shuster, D. L. The bombardment history of the Moon as recorded by 40Ar-39Ar chronology. *Meteorit. Planet. Sci.* **48**, 241–269 (2013).

Forte, E., Dossi, M., Pipan, M. & Colucci, R. R. Velocity analysis from common offset GPR data inversion: theory and application to synthetic and real data. *Geophys. J. Int.* **197**, 1471–1483 (2014).

Forte, E., Pipan, M., Casabianca, D., Di Cuia, R. & Riva, A. Imaging and characterization of a carbonate hydrocarbon reservoir analogue using GPR attributes. *J. Appl. Geophys.* **81**, 76–87 (2012).

Fortezzo, C. M., Spudis, P. D. & Harrel, S. L. Release of the digital unified global geologic map of the moon at 1:5,000,000. (2020).

Giannakis, I., Zhou, F., Warren, C. & Giannopoulos, A. *Inferring the Shallow Layered Structure at the Chang'E-4 Landing Site: A Novel Interpretation Approach Using Lunar Penetrating Radar*. https://essopenarchive.org/doi/full/10.1002/essoar.10506249.1 (2021) doi:10.1002/essoar.10506249.1.

Gou, S. *et al.* Absolute model age of lunar Finsen crater and geologic implications. *Icarus* **354**, 114046 (2021).

Guo, D., Fa, W., Zeng, X., Du, J. & Liu, J. Geochemistry of the Von Kármán crater floor and thickness of the non-mare ejecta over the Chang'e-4 landing area. *Icarus* **359**, 114327 (2021).

Hagedoorn, J. G. A process of Seismic Reflection Interpretation. *Geophys. Prospect.* **2**, 85–127 (1954).

Hochreiter, S. & Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **9**, 1735–1780 (1997).

Huang, J. *et al.* Geological Characteristics of Von Kármán Crater, Northwestern South Pole-Aitken Basin: Chang'E-4 Landing Site Region. *J. Geophys. Res. Planets* **123**, 1684–1700 (2018).

Ivanov, B. A. Size-Frequency Distribution of Small Lunar Craters: Widening with Degradation and Crater Lifetime. *Sol. Syst. Res.* **52**, 1–25 (2018).

Jia, Y. *et al.* The scientific objectives and payloads of Chang'E−4 mission. *Planet. Space Sci.* **162**, 207–215 (2018).

Jiao, Y., McMechan, G. A. & Pettinelli, E. In situ 2-D and 3-D measurements of radiation patterns of half-wave dipole GPR antennas. *J. Appl. Geophys.* **43**, 69–89 (2000).

Lai, J. *et al.* A Complex Paleo-Surface Revealed by the Yutu-2 Rover at the Lunar Farside. *Geophys. Res. Lett.* **48**, (2021).

Lai, J. *et al.* First look by the Yutu-2 rover at the deep subsurface structure at the lunar farside. *Nat. Commun.* **11**, 3426 (2020).

Li, C. *et al.* Pitfalls in GPR Data Interpretation: False Reflectors Detected in Lunar Radar Cross Sections by Chang'e-3. *IEEE Trans. Geosci. Remote Sens.* **56**, 1325–1335 (2018).

Li, C. *et al.* The Moon's farside shallow subsurface structure unveiled by Chang'E-4 Lunar Penetrating Radar. *Sci. Adv.* **6**, eaay6898.

Lin, H. *et al.* New Insight Into Lunar Regolith-Forming Processes by the Lunar Rover Yutu-2. *Geophys. Res. Lett.* **47**, e2020GL087949 (2020).

Ling, Z. *et al.* Composition, mineralogy and chronology of mare basalts and non-mare materials in Von Kármán crater: Landing site of the Chang'E−4 mission. *Planet. Space Sci.* **179**, 104741 (2019).

Lu, Y. *et al.* Chronological sequence of Chang'E-4 landing zone within Von Kármán crater. *Icarus* **354**, 114086 (2021).

Mannor, S., Peleg, D. & Rubinstein, R. The cross entropy method for classification. in *Proceedings of the 22nd international conference on Machine learning - ICML '05* 561–568 (ACM Press, 2005). doi:10.1145/1102351.1102422.

Melosh, H. J. *Impact Cratering: A Geologic Process*. (Oxford University Press, 1989).

Mendes-Moreira, J., Jorge, A. M., Soares, C. & de Sousa, J. F. Ensemble Learning: A Study on Different Variants of the Dynamic Selection Approach. in *Machine Learning and Data Mining in Pattern Recognition* (ed. Perner, P.) 191–205 (Springer Berlin Heidelberg, 2009).

Oliveros, R. B. & Radovich, B. J. Image-processing Display Techniques Applied to Seismic Instantaneous Attributes Over the Gorgon Gas Field, North West Shelf, Australia. in SEG-1997-2064 (1997).

Pettinelli, E., Lauro, S. E., Mattei, E., Cosciotti, B. & Soldovieri, F. Stratigraphy versus artefacts in the Chang'e-4 low-frequency radar. *Nat. Astron.* **5**, 890–893 (2021).

Roncoroni, G., Forte, E., Bortolussi, L. & Pipan, M. Efficient extraction of seismic reflection with Deep Learning. *Comput. Geosci.* **166**, 105190 (2022a).

Roncoroni, G., Forte, E., Bortolussi, L., Gasperini, L. & Pipan, M. Polarity assessment of reflection seismic data: a Deep Learning approach. *BGO* (2022b) doi:10.4430/bgo00409.

Sénéchal, P., Perroud, H. & Sénéchal, G. Interpretation of reflection attributes in a 3-D GPR survey at Vallée d'Ossau, western Pyrenees, France. *GEOPHYSICS* **65**, 1435–1445 (2000).

Wang, R. *et al.* A Novel Approach for Permittivity Estimation of Lunar Regolith Using the Lunar Penetrating Radar Onboard Chang'E-4 Rover. *Remote Sens.* **13**, (2021).

Warner, N. H. *et al.* Near Surface Stratigraphy and Regolith Production in Southwestern Elysium Planitia, Mars: Implications for Hesperian-Amazonian Terrains and the InSight Lander Mission. *Space Sci. Rev.* **211**, 147–190 (2017).

White, L. F. *et al.* Evidence of extensive lunar crust formation in impact melt sheets 4,330 Myr ago. *Nat. Astron.* **4**, 974–978 (2020).

Wu, W. *et al.* Lunar farside to be explored by Chang'e-4. *Nat. Geosci.* **12**, 222–223 (2019).

Xiao, Z. *et al.* Ejecta From the Orientale Basin at the Chang'E-4 Landing Site. *Geophys. Res. Lett.* **48**, e2020GL090935 (2021).

Xu, L., Zhang, X., Qiao, L. & Lai, J. Evaluating the Thickness and Stratigraphy of Ejecta Materials at the Chang'e-4 Landing Site. *Astron. J.* **162**, 29 (2021).

Yilmaz, Ö. *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data*. (Society of Exploration Geophysicists, 2001). doi:10.1190/1.9781560801580.

Zhang, J., Zhou, B. & Lin, Y. Reply to: Stratigraphy versus artefacts in the Chang'e-4 low-frequency radar. *Nat. Astron.* **5**, 894–897 (2021).

Zhang, L. *et al.* Rock abundance and evolution of the shallow stratum on Chang'e-4 landing site unveiled by lunar penetrating radar data. *Earth Planet. Sci. Lett.* **564**, 116912 (2021).

Zhang, L. *et al.* Stratigraphy of the Von Kármán Crater Based on Chang'E-4 Lunar Penetrating Radar Data. *Geophys. Res. Lett.* **47**, (2020).

Zhao, W., Forte, E., Fontana, F., Pipan, M. & Tian, G. GPR imaging and characterization of ancient Roman ruins in the Aquileia Archaeological Park, NE Italy. *Measurement* **113**, 161–171 (2018).

Zhou H. et al., Yutu-2 Radar Sounding Evidence of a Buried Crater at Chang'E-4 Landing Site, IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-19, 2022, Art no. 4600419, doi: 10.1109/TGRS.2021.3090528.

Zhou, H. *et al.* Heterogeneous Weathering Process of Lunar Regolith Revealed by Polarimetric Attributes Analysis of Chang'E-4 Lunar Penetrating Radar Data Acquired During the Yutu-2 Turnings. *Geophys. Res. Lett.* **49**, (2022).

Zhou, H., Feng, X., Dong, Z., Liu, C. & Liang, W. Application of Denoising CNN for Noise Suppression and Weak Signal Extraction of Lunar Penetrating Radar Data. *Remote Sens.* **13**, 779 (2021).