PlanetGSD 1.0: a cross-planetary grain-size distribution dataset from the Earth, the Moon, and the Mars

Jun Zhang[1,2], Yong Li[1*]

[1] Key Laboratory of Mountain Hazards and Engineering Resilience /Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu 610213, China

[2] State Key Laboratory of Hydroscience and Engineering; Department of Hydraulic Engineering, Tsinghua University, Beijing 100084, China

*Correspondence to: Yong Li, ylie@imde.ac.cn

Abstract

Comparative studies of surface processes across planetary bodies are hindered by the lack of consistently parameterized, openly accessible soil data, especially the grain-size distributions (GSDs) data. Here we present PlanetGSD 1.0, the first standardized and unified cross-planetary GSD database. It comprises 6,527 measurements from Earth (4,419 samples, 20 geomorphic settings), the Moon (379 samples, 8 missions), and Mars (1,729 rover-derived estimates, 4 landing areas), covering seven orders of magnitude in grain size (0.0001–600 mm). The textural fractions have been transferred into a unique parameter set (μ, D_c, n) derived from the unified GSD (UGSD) function, accompanied by quality metrics, georeferenced metadata, site-level statistics, and open-source analysis codes. Technical validation confirms high fitting quality across all samples (97.8% with $R^2 > 0.95$) and robust inter-operator reproducibility for Martian image-derived measurements (coefficient of variation < 8.3% for key parameters). The complete dataset is openly available on Figshare (Zhang, 2026) under the CC BY 4.0 license. PlanetGSD 1.0 enables robust cross-planetary comparison of regolith properties, benchmarking of simulants, and data-driven landing site assessment, establishing a foundational resource for planetary science.

1. Introduction

The GSD of soil governs the mechanical and hydraulic properties; GSD data under a unified

framework are crucial for comparative planetology, geotechnical engineering, and the planning of future exploration missions (Carrier, 2003; Fedo et al., 2015; Iverson & Vallance, 2001). On Earth, GSD classification underpins soil taxonomy and geotechnical site characterization (Carrier, 2003; Das, 2019). On the Moon, GSD controls regolith trafficability, dust mobilization, and the design of in-situ resource utilization (ISRU) systems (Colwell et al., 2007). On Mars, GSD provides important constraints on the transition from volcanic to sedimentary surface processes and on past hydrological activity (Grotzinger et al., 2015; Rivera-Hernández et al., 2020). Importantly, primary volcanic deposits tend to be poorly sorted with unimodal or bimodal distributions, whereas reworked aeolian or fluvial deposits exhibit better sorting and distinct Weibull shape parameters. The Medusae Fossae Formation — a >5000 km-long deposit of friable, fine-grained materials — exemplifies this volcanic-to-sedimentary transition, interpreted as a pyroclastic deposit from explosive volcanism (Ojha & Lewis, 2018; Wilson & Head, 2007), subsequently shaped by aeolian erosion into yardangs and other wind-carved landforms.

However, existing GSD datasets remain fragmented across disciplines and planetary targets, with no unified analytical framework for cross-body comparison. Terrestrial soil databases (e.g., ISRIC-WISE, HWSD, POLARIS; Batjes, 2009; Chaney et al., 2016; Shangguan et al., 2014) provide broad geographic coverage but report only texture-class percentages and use Earth-centric schemes (USDA, FAO) that are not transferable to extraterrestrial contexts. Complementary large-scale resources such as the global SoilGrids product and the Webb et al. (2000) soil texture database offer additional spatial coverage but report only texture classes rather than complete GSD curves. Lunar GSD data are scattered across mission-specific catalogs (Graf, 1993; Heiken et al., 1991), with varying sieve protocols and no unified parameterization; recent Chang'e-5/-6 measurements (Li et al., 2022, 2024; Zhang et al., 2022) remain isolated from the Apollo/Luna corpus. Martian GSD estimates derived from rover imaging at four landing sites (Gale, Jezero, Gusev, Meridiani) employ site-specific protocols and resolution limits (Fedo et al., 2015; Stack et al., 2020; Weitz et al., 2018), and no cross-site compilation exists. Compounding this fragmentation, conventional GSD descriptors (percentile indices, moment statistics) reduce continuous distributions to summary values without direct links to physical processes, while standard unimodal fitting functions (lognormal, Rosin – Rammler, Fredlund) cannot represent the multimodal,

polygenetic distributions common in natural regolith (Bittelli et al., 1999). The universal GSD (UGSD) function (Yong et al., 2013, 2017) addresses these limitations through a flexible four-parameter formulation capable of capturing multimodal curves, but has not previously been applied as a unified cross-planetary framework.

Here we present PlanetGSD 1.0, a standardized database of 6,527 GSD measurements compiled from Earth (4,419 samples, 20 geomorphic settings, 18 countries), the Moon (379 samples from Apollo, Luna, and Chang'e missions), and Mars (1,729 rover-derived estimates from Gale, Jezero, Gusev, and Meridiani), covering seven orders of magnitude in grain size (0.0001–600 mm). The GSD measurements derive from distinct techniques — laboratory sieving / laser diffraction for Earth, dry sieving for lunar returned samples, and image-based segmentation for Mars — introducing method-specific biases (Martian right-censoring below 40–150 μm; lunar loss of fines due to clogging; terrestrial laser diffraction sensitivity to grain shape). All samples have been harmonized using the UGSD function (Zhang et al., 2023), achieving a median fitting R^2 of 0.988. Compared with existing resources, PlanetGSD 1.0 provides four key advances: (i) the first unified parameterization across three planetary bodies; (ii) complete raw GSD cumulative curves (mass percentage passing) retained alongside fitted parameters; (iii) site-level statistics (see Section 2.4 for full definition, formula, and application to grain size analysis) enabling stochastic field generation (see Section 2.5 for detailed methodology); and (iv) fully open data, codes, and metadata under the CC BY 4.0 license. The UGSD harmonization framework provides a common parametric language that retains method-specific biases as explicitly documented limitations. Future work (PlanetGSD 2.0) will integrate thermal inertia and photometric datasets to enable regional-to-global extrapolation.

PlanetGSD 1.0 is designed to serve three primary objectives: (i) to enable cross-body comparative analysis of GSDs under a unified parametric framework; (ii) to support planetary mapping and process identification by linking site-level GSD parameters to geologic units, depositional environments (aeolian, fluvial, impact, volcanic), and surface processes; and (iii) to provide a quantitative benchmark for regolith simulant development for future lunar and Martian missions. Recent analysis of Chang'e-6 farside samples (Qi et al., 2025) reveals finer-grained, more poorly sorted, and more cohesive regolith than nearside sites, underscoring the need for a systematic,

cross-mission GSD database to enable robust cross-planetary comparison of regolith properties and data-driven landing-site assessment.

This paper is organized as follows. Section 2 describes the data acquisition methods for each planetary body. Section 3 presents the data records, including dataset structure, content summaries, and key fields. Section 4 reports the technical validation results. Section 5 provides illustrative applications and identifies open questions. Section 6 and Section 7 are usage notes and conclusions, respectively.

2. Methods

2.1 Overview of the PlanetGSD dataset

The PlanetGSD dataset comprises 6,527 soil samples from three planetary bodies—Earth, Moon, and Mars—integrated within a unified analytical framework based on the UGSD function. This compilation enables systematic comparison of granular characteristics across terrestrial and extraterrestrial environments with contrasting surface processes, climatic regimes, and geological histories. The database is organized with multi-level metadata, including geographic coordinates, geomorphic context, mission information, and original data sources, all standardized to facilitate cross-planetary analysis.

Earth component. The terrestrial samples represent the full spectrum of Earth's surface environments, spanning five continents and encompassing 20 distinct geomorphic settings. These include gravity-driven deposits (debris flows, landslides, moraines), fluvial systems (river valleys, floodplains), aeolian environments (deserts, loess), pedogenic landscapes (vegetated slopes, grasslands), and coastal zones. This diversity captures grain-size signatures associated with major process regimes, providing a comprehensive terrestrial reference for comparison with extraterrestrial materials. The Earth data integrate field-collected samples and literature-derived measurements, all processed using standardized laboratory protocols (sieving and laser diffraction).

Lunar component. The lunar samples originate from eight sample-return missions conducted between 1969 and 2024, including Apollo (USA), Luna (USSR), and Chang'e (China) programs. These missions provide geographic coverage of major lunar geological provinces, including mare basalts, highland terrains, and the farside South Pole-Aitken basin. All samples are regolith materials processed by impact gardening and space weathering, with grain-size distributions originally

measured by dry sieving. The lunar data are harmonized from multiple mission catalogs, enabling consistent cross-mission comparison.

Martian component. The Martian samples are derived from high-resolution rover imagery at four NASA landing sites: Gale Crater (Curiosity), Jezero Crater (Perseverance), Gusev Crater (Spirit), and Meridiani Planum (Opportunity). These sites represent diverse Martian geological environments, including lacustrine mudstones and deltaic deposits (Gale, Jezero), weakly altered olivine basalt plains (Gusev), and sulfate-rich evaporite sequences (Meridiani), spanning from Noachian to Amazonian periods. Grain-size distributions were obtained through semi-automated image segmentation, capturing grain sizes from silt to gravel.

Database structure. All samples in PlanetGSD 1.0 are standardized to a consistent framework: raw cumulative GSD curves are preserved in their original bin resolutions, and each sample is fitted to the UGSD function to extract interpretable parameters (μ, D_c, n). Site-level statistics, including three-parameter Weibull distributions for μ populations, are provided to enable stochastic field generation and comparative analysis. Detailed sample-level metadata, including geographic coordinates, geomorphic setting, mission information, and original data citations, are provided in Supplementary Tables S1–S7.

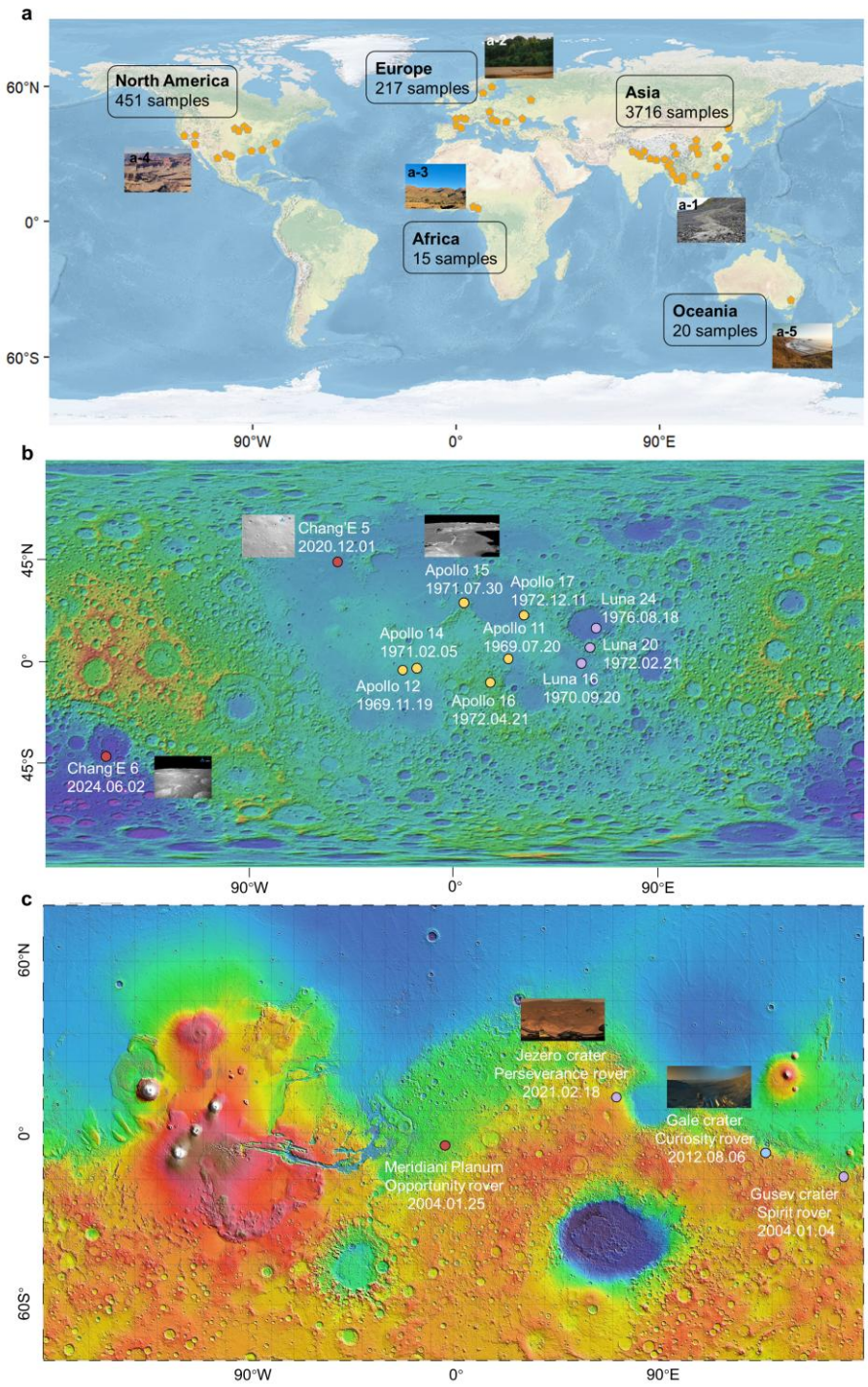

Figure 1. Global distribution of three terrestrial planetary soil sampling sites and representative photographic images. (a) Sampling locations across 18 countries on five continents: Asia (China, Singapore, South Korea, Thailand, Iran), Europe (Finland, Sweden, France, Bulgaria, Romania, Poland, Germany, Russia), Africa (Nigeria), North America (USA, Canada), and Australia. Representative soil environments are shown: (a-1) wide-graded debris-flow deposits (Jiangjia Gully, Southwest China); (a-2) arable land (Sweden); (a-3) gully terrain (Nigeria); (a-4) canyon (USA); and (a-5) coastal beach (Australia). (b) Sampling locations of lunar soils, predominantly collected near spacecraft landing sites. The insets show surface images from Apollo 15, CE-5, and CE-6, with

mission dates annotated. The topographic basemap is the Lunar Reconnaissance Orbiter Laser Altimeter (LRO LOLA) color shaded relief (388 m resolution). Images courtesy of the China National Space Administration (CNSA). (c) Sampling locations of Martian soils across four distinct regions, with insets showing overviews of Gale and Jezero Craters. The basemap is the Mars Orbiter Laser Altimeter (MOLA) shaded relief topography. Images courtesy of NASA.

Table 1 summarizes the key characteristics of the three planetary datasets, including sample size, geographic coverage, analytical methods, grain-size ranges, and data sources. Table 2 provides detailed information for each sampling site, including sample counts, dominant soil types, and associated surface processes.

Table 1. Summary of the PlanetGSD dataset by planetary body

Component	Sample size	Geographic / landing-site coverage	Analytical method	Grain-size range (mm)	Primary data sources
Earth	4,419	18 countries on 5 continents; 20 geomorphic settings	Combined sieving ($\geqslant$0.25 mm) + laser diffraction (<0.25 mm)	0.0001–20	Field campaigns (2015–2023); 43 peer-reviewed publications
Moon	379	8 sample-return missions (Apollo 11,12,14,15,16,17; Luna 16,20,24; Chang'e-5,-6)	Dry sieving	0.001–10	Graf (1993); Heiken et al. (1991); Zhang et al. (2022); Li et al. (2024)
Mars	1,729	4 rover landing sites (Gale, Jezero, Gusev, Meridiani)	Image-based grain segmentation (MAHLI, WATSON, MI, Mastcam, Mastcam-Z, RMI)	0.04–25	New analysis of PDS imagery (this study)

Table 2. Sampling sites, sample counts, dominant soil types, and associated surface processes

Planet	Site / Region	Sample count	Dominant soil types	Associated surface processes
Earth	Jiangjia Gully (China)	438	Debris-flow deposits	Gravity-driven mass movements
	Minjiang River (China)	825	Fluvial deposits	Fluvial transport and deposition
	Xichang (China)	960	Vegetated soils	Pedogenesis, root-soil interaction
	Horqin (China)	65	Grassland soils	Aeolian erosion, biotic processes
	Tibet (China)	107	Alpine meadow soils	Cryogenic processes
	Kunlun Mountains (China)	21	Permafrost and desert soils	Cryogenic, aeolian
	Malipo (China)	570	Red soils	Chemical weathering, pedogenesis
	Nevada (USA)	120	Desert soils	Aeolian erosion, arid weathering
	Michigan (USA)	31	Sandy soils	Glacial, coastal processes
	California (USA)	38	Coastal plain soils	Fluvial, coastal
	Canada	41	Fine loamy soils	Glacial, pedogenic
	Finland	108	Forest soils	Pedogenic, cryogenic

	Sweden	35	Arable soils	Agricultural, pedogenic
	Germany	40	Brown soils	Pedogenic
	France	45	Various soils	Fluvial, pedogenic
	Bulgaria	21	Agricultural soils	Pedogenic
	Poland	14	Forest soils	Pedogenic
	Romania	30	River plain soils	Fluvial
	Russia	32	Forest and plain soils	Pedogenic, fluvial
	Nigeria	15	Agricultural soils	Pedogenic, tropical weathering
	Australia	20	Coastal soils	Coastal, aeolian
	Singapore	32	Granite-derived soils	Tropical weathering
	South Korea	12	Red-yellow soils	Pedogenic
	Japan	42	Volcanic, forest soils	Volcanic, pedogenic
	Thailand	28	Siliceous soils	Tropical weathering
	Iran	35	Aeolian and marine sediments	Aeolian, coastal
Moon	Apollo 11	21	Mare basaltic regolith	Impact comminution, space weathering
	Apollo 12	20	Mare basaltic regolith	Impact comminution, space weathering
	Apollo 14	26	Highland ejecta	Impact comminution, mixing
	Apollo 15	63	Mare-highland boundary	Impact comminution
	Apollo 16	5	Highland regolith	Impact comminution, multi-source mixing
	Apollo 17	43	Mare-highland contact	Impact comminution
	Luna 16	-	Mare basaltic regolith	Impact comminution
	Luna 20	-	Highland regolith	Impact comminution
	Luna 24	-	Mare basaltic regolith	Impact comminution
	Chang'e-5	1	Mare basaltic regolith	Impact comminution, space weathering
	Chang'e-6	1	Farside SPA basin	Impact comminution
Mars	Gale Crater (11 sub-sites)	384	Lacustrine, aeolian, fluvial	Lake sedimentation, aeolian transport
	Jezero Crater (8 sub-sites)	722	Deltaic, lacustrine, volcanic	Delta deposition, lake environment
	Gusev Crater (10 sub-sites)	366	Volcanic, aeolian, hydrothermal	Flooding, volcanism, alteration
	Meridiani Planum (6 sub-sites)	257	Sulfate-rich, aeolian	Evaporite deposition, aeolian transport

2.2 Data acquisition methods

2.2.1 Field sampling (Earth)

Terrestrial field samples (2,847) were collected during campaigns conducted between 2015 and 2023 across 18 countries on five continents (Fig. 1a; Supplementary Tables S1–S5). Sampling targeted the upper 10 cm of the surface layer to ensure consistency with the actively weathering zone and to facilitate comparison with orbital remote sensing data. At vegetated sites, surface organic litter was removed prior to sampling.

The sampling strategy was tailored to geomorphic context: systematic grid-based transects for homogeneous terrain (plains, grasslands, deserts) and targeted sampling of distinct geomorphic units for heterogeneous environments (debris-flow deposits, landslides, moraines). Individual sample mass ranged from 200 g to 2 kg (median 500 g) to ensure sufficient material for grain-size analysis

and representative sampling of heterogeneous deposits. Samples were sealed in polyethylene bags with GPS coordinates recorded, transported to the laboratory, and stored at 4°C prior to analysis.

All field-collected samples were processed following standardized protocols (BS 1377-2, ASTM D422-63, ASTM D6913-04, ISO 11277:2009) using combined sieving and laser diffraction methods, as detailed in Section 2.3.1. Complete instrument specifications and quality control procedures are provided in Text S1. No additional replicate analyses or certified reference material measurements were performed specifically for this study.

2.2.2 Literature compilation (Earth and Moon)

Terrestrial literature data. In addition to field-collected samples, we compiled grain-size distribution data for 1,572 terrestrial samples from 43 peer-reviewed publications. These samples expand the geographic and environmental coverage of the terrestrial component, including soils from Europe (Finland, Sweden, France, Germany, Poland, Romania, Russia), North America (USA, Canada), Asia (China, Japan, South Korea, Thailand, Iran), Australia, and Africa (Nigeria). The compiled data represent diverse surface environments such as forested hillslopes, agricultural lands, grassland soils, desert regions, coastal zones, and loess-mantled terrains, complementing the field-collected samples and enhancing the representation of major process regimes (Text S2).

Lunar mission data. Lunar grain-size distributions were compiled from eight sample-return missions (Apollo, Luna, Chang'e) spanning 1969–2024. Data were digitized from the NASA Lunar Soils Grain Size Catalog (Graf, 1993) for Apollo and Luna samples, and from mission-specific publications for Chang'e-5 (Zhang et al., 2022) and Chang'e-6 (Li et al., 2024). All lunar samples were originally analyzed using standardized dry sieving protocols. Detailed mission information, sample counts, and metadata are provided in Table 1 and Table S6.

2.2.3 Image-based analysis (Mars)

Martian GSD data were derived from high-resolution rover images for 1,729 sampling targets at four NASA landing sites: Gale Crater (Curiosity), Jezero Crater (Perseverance), Gusev Crater (Spirit), and Meridiani Planum (Opportunity). All images were obtained from NASA's Planetary Data System (PDS) Imaging Node (https://pds-imaging.jpl.nasa.gov/). Source image identifiers (PDS Product IDs) for all 1,729 samples are provided in Dataset S3.

Imaging systems. Grain-size measurements utilized both arm-mounted close-range imagers

and mast-mounted cameras. Close-range systems include the Mars Hand Lens Imager (MAHLI) on Curiosity (Edgett et al., 2012), the Wide Angle Topographic Sensor for Operations and Engineering (WATSON) on Perseverance, and the Microscopic Imager (MI) on Spirit and Opportunity (Herkenhoff et al., 2004). These instruments achieve pixel scales of 13–32 μm at working distances of 2–5 cm, enabling resolution of sand-sized grains down to approximately 40 μm. Mast-mounted systems—Mastcam on Curiosity (Bell et al., 2017), Mastcam-Z on Perseverance (Bell et al., 2021), and SuperCam Remote Micro-Imager (RMI) on Perseverance (Maurice et al., 2021; Wiens et al., 2021)—provide pixel scales of 74–450 μm depending on focal length and target distance, limiting grain-size resolution to medium sand and coarser fractions (>150 μm). Table 3 summarizes the key characteristics, resolution limits, and detection thresholds for all imaging systems used in this study.

Image selection and analysis. A total of 2,160 images meeting quality criteria (known target distance, viewing angle <30° from vertical, solar elevation >20°) were selected from the PDS archive. Grain-size distributions were measured using semi-automated image segmentation software developed by Shi et al. (2021) and Zhao et al. (2023), implemented in Mathematica. The software's grain identification capability has been validated using three approaches (Karunatillake et al., 2014; Zhao et al., 2021). First, it was qualitatively compared with BASEGRAIN, ImageJ WEKA, and ENVI tools on 57 MAHLI and MI images from Mars, showing superior accuracy and speed. Second, using terrestrial basaltic sand (0.1–1.0 mm) as a known analog with manually placed pebbles (5–10 mm), the algorithm successfully segmented foreground pebbles while excluding background sand (Karunatillake et al., 2014, Part 2, Section 3.4). Third, the same study demonstrated that the algorithm provides consistent results across repeated runs, whereas manual segmentation yields substantial internal inconsistency (areal mismatch of 35–50 %). The software identifies individual grains through edge detection and watershed segmentation, calculates equivalent circular diameter (ECD) for each grain, and constructs cumulative size distributions after binning into logarithmic size classes. For each image, 150–500 grains (median 280) were segmented. The geographic distribution of all four landing sites is shown in Figure 2. Complete image selection criteria, analysis protocols, resolution limits, and uncertainty estimates are provided in Text S3.

Table 3. Martian rover imaging systems and grain-size detection limits

Instrument	Pixel scale (μm/pixel)	Working distance	Min. detectable grain size (mm)	Min. detectable grain size (conservative/5-pixel)	Notes	Reference
MI (Microscopic Imager)	31	6.8 cm	0.09 mm	0.16 mm	Fixed scale	Herkenhoff et al., 2004
MAHLI (Mars Hand Lens Imager)	14-22	2–10 cm	0.04 – 0.07 mm	0.07 – 0.11 mm	5-pixel conservative threshold used in this study	Edgett et al., 2012
WATSON	~14 (best) / 14-77 (range)	2–25 cm	0.04 mm	0.07 mm	Similar to MAHLI	Beegle et al., 2015
Mastcam (M-100)	74-220	1–10 m	0.22 – 0.66 mm	0.37 – 1.10 mm	Coarse fraction only	Bell et al., 2017
Mastcam-Z (left)	148 (at 2 m)	0.5–∞ m	0.44 mm	0.74 mm	—	Bell et al., 2021
RMI (SuperCam)	40 (at 2 m)	1.6–∞ m	0.12 mm	0.20 mm	Fisheye distortion; central 40 % analyzed	Maurice et al., 2021

Note: Minimum detectable grain size is estimated as 3× pixel scale (optimistic) and 5× pixel scale (conservative). Following Karunatillake et al. (2014) and Shi et al. (2024), we adopt the conservative 5-pixel threshold due to the point spread function (PSF) of the optics and the ubiquitous presence of Martian dust, which reduces effective contrast. For MAHLI at 14 μm/pixel, this yields a practical lower bound of ~70 μm rather than the theoretical 40 μm.

2.2.4 Instrument-specific processing protocols

Different imaging instruments require tailored preprocessing before grain segmentation (Kozakiewicz et al., 2018):

MAHLI and WATSON (color, arm-mounted): These instruments acquire images at variable working distances (2-25 cm) with corresponding pixel scales of 14-77 μm/pixel (Aileen et al., 2014). We exclusively use the highest-resolution images available for each target (working distance ≤10 cm, pixel scale ≤32 μm/pixel). The focus motor count recorded in image metadata enables scale calculation using the empirical relationship $p = 6.90 + 3.52 \times w$ (where p = μm/pixel, w = working distance in cm; Minitti et al., 2013). When available, images acquired after dust removal tool (DRT) brushing are prioritized to minimize dust obscuration (Sacks et al., 2016).

RMI (grayscale, mast-mounted, fisheye): RMI images exhibit barrel distortion, particularly toward the field edges. To minimize geometric bias, we analyze only the central 40 % of the field of view where distortion is < 2% based on calibration target measurements. Grayscale images require contrast enhancement (adaptive histogram equalization) prior to segmentation. Resolution

246 at typical working distances (2-5 m) yields pixel scales of 40-100 μm/pixel, limiting grain detection
247 to medium sand and coarser.
248 Cross-instrument consistency: For 30 targets imaged by both MAHLI and RMI at the same
249 location, we find median D_{50} agreement within ± 12% (see Section 4.4), confirming that instrument-
250 specific processing yields comparable results.
251 **2.2.5 Discrimination of composite grains**
252 A critical step in image-based granulometry is distinguishing discrete detrital grains from
253 composite features (e.g., small grains sitting on larger grains, grain aggregates, or surface coatings).
254 Our protocol follows a two-step approach:
255 **Algorithmic segmentation:** The watershed algorithm identifies individual grains based on
256 local intensity minima, effectively separating touching grains. However, this algorithm may fail
257 when grains partially overlap or when a small grain lies atop a larger one.
258 **Manual inspection and correction:** After automated segmentation, each image is visually
259 inspected by the operator. For ambiguous cases—where a small grain appears embedded within or
260 attached to a larger grain—the following rules are applied:
261 • If $\geqslant$ 50% of the small grain's perimeter is visible and distinct from the underlying grain,
262 it is counted as a separate grain.
263 • If the feature is a surface texture, cement, alteration rind, or dust aggregate, it is not
264 counted as a discrete grain.
265 • If the visible perimeter is < 50% (e.g., a partially buried grain), it is considered part of the
266 larger grain.
267 Additionally, following Shi et al. (2024), images where foreground coverage (e.g., coarse
268 granules or pebbles overlying finer matrix) exceeds 5% are excluded from analysis, as such high-
269 contrast composite textures compromise reliable grain boundary detection.
270 This two-step protocol ensures consistent discrimination across the 1,729 Martian samples.
271 Complete operator guidelines are provided in Supplementary Text S3.

2.3 Grain-size analysis and harmonization

2.3.1 Granulometric analysis methods

The dataset integrates GSDs obtained through three granulometric methods: combined sieving-laser diffraction (terrestrial field samples and literature), dry sieving (lunar missions), and image-based segmentation (Mars rovers). Table 4 summarizes the principles, grain-size ranges, advantages, and limitations of each method.

Table 4. Comparison of GSD analytical methods for grain-size determination across the three planetary datasets

Methods	Principles	Grain-size range	Advantages	Limitations	Applied to
Combined sieving–laser diffraction	Sieving separates coarse fractions by physical passage through mesh; laser diffraction measures fine fractions by angular light scattering	0.0001–20 mm	Full size-range coverage (clay to gravel); high precision for fine fractions; standardized protocols (ASTM/ISO)	Time-intensive; requires laboratory access	Earth (field samples and literature)
Dry sieving	Mechanical shaking through nested sieve stacks	0.001–10 mm	Direct mass-based measurement; well-established for coarse sediments; large sample masses possible	Poor resolution for fines (<0.1 mm); clay fractions not resolved	Moon (all missions)
Image-based segmentation	Grain boundary detection from high-resolution images; equivalent circular diameter calculation	0.04–25 mm	In situ analysis; no sample return required; spatial context preserved	Truncated at fine end ($\geqslant 40\ \mu$m); spherical grain assumption (standard practice in 2D image analysis; see Heilbronner & Barrett, 2013; Pizzati et al., 2023); 2D projection bias	Mars (rover images)

Note: 1. Mars imaging data are right-censored below 40–150 μm (depending on the instrument), meaning that fitted μ values may systematically underestimate fine-grain content; 2. Earth data are from combined sieving and Malvern laser diffraction. The Malvern instrument is based on the principle of laser diffraction and is sensitive to grain shape; non-spherical grains introduce systematic bias; 3. Lunar sieving provides mass-based grain-size distributions down to $\sim$1 μm but captures no shape information (e.g., aspect ratios, roundness, sphericity). Fine fractions (<0.1 mm) suffer from poor resolution due to sieve clogging and electrostatic adhesion.

Although the raw data are measured by different methods, the UGSD function provides a unified flexible parametric framework to fit the measurements and transfer the textural fractions to a unique parameter set (μ, D_c, n), which enables comparative analysis while preserving method-

specific biases as quantifiable uncertainties. Critically, Martian image-based data are right-censored below 40–150 μm (Table 4), so fitted μ values may systematically underestimate fine-grain content. Lunar sieving resolves grains down to $\sim$1 μm but misses clay morphology; terrestrial laser diffraction covers the full range but is shape-sensitive.

Figure 2 shows representative GSD patterns from each planetary body, illustrating the range of shapes captured in the dataset. With the *x*-axis of all subfigures standardized to the same range (0.001–1000 mm), the GSDs are more directly comparable. The terrestrial examples (Fig. 2a-1) exhibit the widest size range (0.001–1000 mm) and display multi-peak patterns. Specifically, poorly sorted landslide deposits, bimodal debris-flow materials, and finer-grained vegetated soils are included, and the cumulative probability curves of mass movements (landslides and debris flows) samples are relatively scattered (Fig. 2a-2). Lunar samples (Fig. 2b-1) are very fine (<4 mm), showing the relatively narrow, fine-grained distributions characteristic of impact-comminuted regolith; however, the cumulative probability curves of different lunar soils are very close to each other (Fig. 2b-2). Martian samples (Fig. 2c-1) thus exhibit a very narrow range (0.3–2 mm), with all curves showing a multi-modal pattern; notably, the cumulative curves of Martian soils are almost straight lines (Fig. 2c-2). Martian samples thus exhibit intermediate characteristics: some sites (e.g., Meridiani Planum) show well-sorted distributions indicative of aeolian sorting (Kozakiewicz et al., 2025), while others (e.g., Gale Crater) preserve multi-modal or poorly sorted signatures reflecting primary sedimentary textures (Kapui et al., 2018; Milliken et al., 2014).

The diversity of grain-size curves across and within planetary bodies—ranging from unimodal to multimodal distributions, with varying degrees of sorting and skewness—highlights the complexity of natural granular materials. Traditional descriptors such as percentile-based indices

(D_{10}, D_{50}, D_{90}) or unimodal distribution models (e.g., lognormal, Weibull) are insufficient to capture this variability in a standardized manner, particularly for cross-planetary comparison where analytical methods and size ranges differ fundamentally. To address this limitation, we employ the Universal Grain-Size Distribution (UGSD) function (Section 2.3.2), which provides a flexible, four-parameter formulation capable of representing diverse curve shapes within a unified framework, enabling systematic comparison of granular characteristics across the planetary bodies.

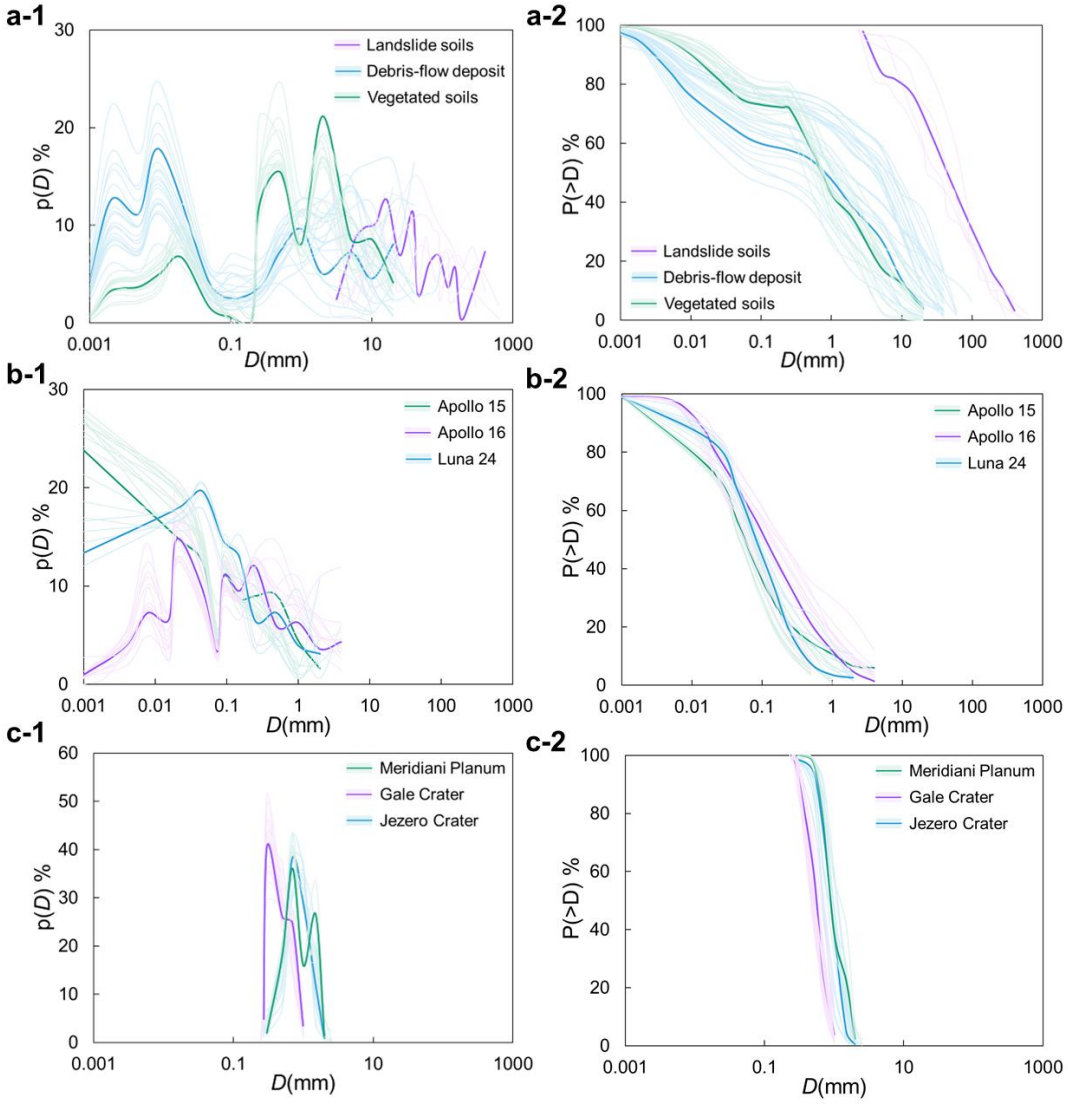

Figure 2. Representative GSD curves from each planetary body. Left column: frequency curves (Y-axis: mass percentage of grains within each bin); right column: cumulative curves (Y-axis: cumulative percentage passing). (a) Terrestrial: landslide deposits, debris-flow deposits, and

vegetated soils; (b) Lunar: Apollo 15, Apollo 16, and Luna 24; (c) Martian: Meridiani Planum, Gale Crater, and Jezero Crater.

2.3.2 Data harmonization

Cross-planetary comparison requires harmonization of data collected using different methods with distinct size ranges and bin resolutions. Rather than applying interpolation or re-binning, which could introduce artifacts, we employ the UGSD function to fit the measured GSD data.

The UGSD function, developed and validated in our previous work (Yong et al., 2013, 2017; Zhang et al., 2023, 2025), describes the cumulative GSD as:

$$P(D) = f(\mu)D^{-\mu}\exp(-(D/D_c)^n) \quad (3)$$

where μ is a power exponent positively correlated with fine-grain content, D_c is the characteristic grain size, the normalized coefficient $f(\mu) = a\exp(-b\mu)$, with empirically determined constants a and b); and n is a shape exponent controlling the coarse-tail steepness.

All 6,527 samples were fitted to the UGSD function (Eq. 3) using nonlinear least-squares optimization. Figure 3 shows the rescaled UGSD curves for Earth (a), Moon (b), and Mars (c), demonstrating the unified fitting performance across planetary bodies. Median R^2 values are 0.991 (Earth), 0.989 (Moon), and 0.983 (Mars), with 97.8% of all samples yielding $R^2 > 0.95$ (Table 9). The terrestrial and lunar populations exhibit strongly left-skewed R^2 distributions, with >85% of samples achieving $R^2 > 0.98$. The slightly lower Martian median is attributable to the truncation of the fine tail below ~0.04–0.15 mm imposed by imaging resolution limits (Table 3), which reduces the number of constraining data points available for curve fitting.

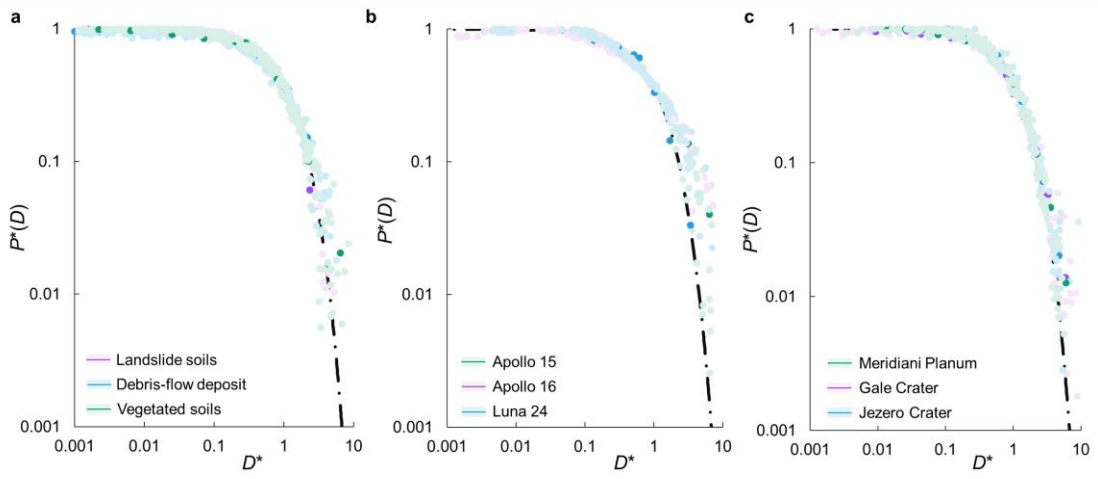

Figure 3. Rescaled UGSD fitting performance across the three planetary bodies. Each panel shows all measured cumulative GSD curves collapsed onto dimensionless axes by normalizing grain size D by the fitted characteristic size D_c and plotting the rescaled cumulative percentage $P(>D)/f(\mu)$ against $(D/D_c)^n$. (a) Earth; (b) Moon; (c) Mars.

All samples included in the dataset contain sufficient size fractions for reliable UGSD fitting (minimum five data points per sample). For curves truncated at coarse or fine ends (e.g., Martian samples missing fractions <40 μm), fitting is performed on the available portion, and truncation thresholds are documented in the metadata (Dataset S4). Complete quality control protocols and uncertainty estimates for all datasets are provided in Texts S1–S4. Parameter definitions and typical ranges for the dataset are summarized in Table 5.

For Martian image-derived data, all grain-size distributions are truncated at the instrument-specific minimum detectable grain size, defined as three times the pixel scale (Table 3). Grains smaller than this threshold are not resolvable and are therefore excluded from the fitted GSD curves. Rather than extrapolating or interpolating across the unresolved fine tail, we retain the truncation as a conservative lower bound: cumulative percentages at the smallest size bin are treated as "≥" values, and the fitted UGSD parameters are reported with a quality flag indicating right-censoring. This approach follows the reviewer's recommendation to truncate rather than extrapolate across unconstrained size ranges. All truncated fractions are documented in Dataset S3 metadata.

Comparison with alternative parameterizations. To demonstrate the advantage of the UGSD function, we fitted all 6,527 samples to three alternative models (lognormal, Rosin-Rammler, and Fredlund three-parameter) and found that UGSD achieved higher median R^2 values than all three

alternatives (Table S10).

Table 5. UGSD function parameters

Parameter	Symbol	Description	Typical range
Power exponent	μ	Correlated with fine-grain content	0.01–3.5
Characteristic size	D_c	Representative grain diameter (mm)	0.01–20
Shape exponent	n	Controls coarse-tail steepness	0.1–2.0

2.3.3 Harmonization across 2D and 3D methods

A fundamental challenge in cross-planetary comparison is the incompatibility between 2D image-based measurements (Mars) and 3D mass-based measurements (Earth dry sieving + laser diffraction; Lunar dry sieving). Rather than attempting direct conversion of raw percentages, we adopt a three-pronged harmonization strategy:

(1) Comparative framework. All Martian samples are analyzed exclusively through the same image-based method, enabling internal comparisons across landing sites. Lunar and terrestrial data remain mass-based. Cross-method comparisons are performed only at the parameter level (UGSD μ, D_c, n) rather than raw cumulative percentages, as these parameters have been shown to be more robust to methodological differences (Kozakiewicz et al., 2018).

(2) Stereological conversion where applicable. For direct comparisons between 2D ECD (equivalent circular diameter) distributions and 3D sieve data, we apply a shape correction factor of 1.3-1.5 following terrestrial studies (Heilbronner & Barrett, 2013; Pizzati et al., 2023). This converts 2D measurements to estimated 3D grain sizes. The correction factor is applied only in interpretive discussions (Section 5) and is clearly flagged as an approximation.

(3) Uncertainty propagation. The inter-operator variability documented in Section 4.4 (CV ≈ 5-8 % for key parameters) implicitly captures the combined effects of the 2D-to-3D approximation and operator-dependent grain delineation. This uncertainty is reported alongside all fitted parameters in Dataset S4.

2.4 Probability Distribution of UGSD Parameters

The UGSD provide quantitative constraints on method-induced variability. Rather than erasing method differences, the UGSD framework makes them explicit and traceable by transferring the

396 measured textural fractions to the parameter set (μ, D_c, n). Moreover, as discussed below, the
397 probability distribution of the exponent μ provides a potential method to recover the "loss" of fine
398 content in the sampling methods, e.g., as in the case of Mars.

399 The UGSD parameters (μ, D_c, n) exhibit spatial variability across sampling sites, reflecting the
400 inherent heterogeneity of granular materials under different surface processes. Among these, the
401 exponent μ—which correlates positively with fine-grain content—serves as the most sensitive
402 indicator of local textural characteristics and is therefore the primary focus of our probabilistic
403 analysis. The characteristic size D_c also varies spatially and its distribution can be analyzed similarly,
404 while the shape exponent n primarily reflects broad-scale differences between planetary bodies or
405 major process regimes (e.g., $n \approx 1$ for Earth and Moon, $n > 1$ for Mars) and is not considered for
406 site-scale probability distributions. Figure 4a displays the (μ, D_c) points of all the samples from the
407 three bodies. Soils from the Earth, Moon, and Mars. The cluster of the Earth covers soils of almost
408 all the types from continents, while clusters of the Moon and Mars are limited to samples in the
409 landing sites. Future explorations of the Moon and Mars are expected to fill the data gaps.

410 The (μ, D_c) value can be assigned to each point in principle; while for a target area, no matter
411 of the scale, they present a probability distribution. For example, at the site level, the population of
412 μ values is found to follow the three-parameter Weibull distribution:

413 $$f(\mu;\lambda,k,\mu_0)=\frac{k}{\lambda}\left(\frac{\mu-\mu_0}{\lambda}\right)^{k-1}\exp\left(-\left(\frac{\mu-\mu_0}{\lambda}\right)^{k}\right) \quad (4)$$

414 where μ_0 is the location parameter (the minimum measured μ value), λ is the scale parameter
415 (characteristic spread), and k is the shape parameter (controlling distribution tail behavior). The
416 three-parameter form was adopted because it provides greater flexibility than the two-parameter
417 version, accommodating potential shifts in the minimum μ value that may arise from site-specific
418 depositional or weathering histories. Fitting was performed using maximum likelihood estimation

(scipy.stats.weibull_min.fit in Python).

Table 6 summarizes the ranges of fitted Weibull parameters across the three planetary bodies, and Figure 4b displays the site-level (k, λ) points for all 42 locations of the samples in the dataset. One sees distinct point groups for the three planetary bodies, supporting the utility of the k-λ parameter pair as a diagnostic descriptor for cross-planetary comparison. Complete site-level Weibull parameters, including geological context and dominant process annotations, are provided in Dataset S4 (Sheet 2).

Table 6. Weibull parameter ranges by planetary body

Planet	**sites**	***k* range**	**λ range**	**μ_0 range**
Moon	6	0.053–0.077	1.83–4.48	0–0.028
Earth	20	0.005–0.354	0.49–6.38	0–0.294
Mars	16	0.367–3.180	1.01–5.64	0–3.142

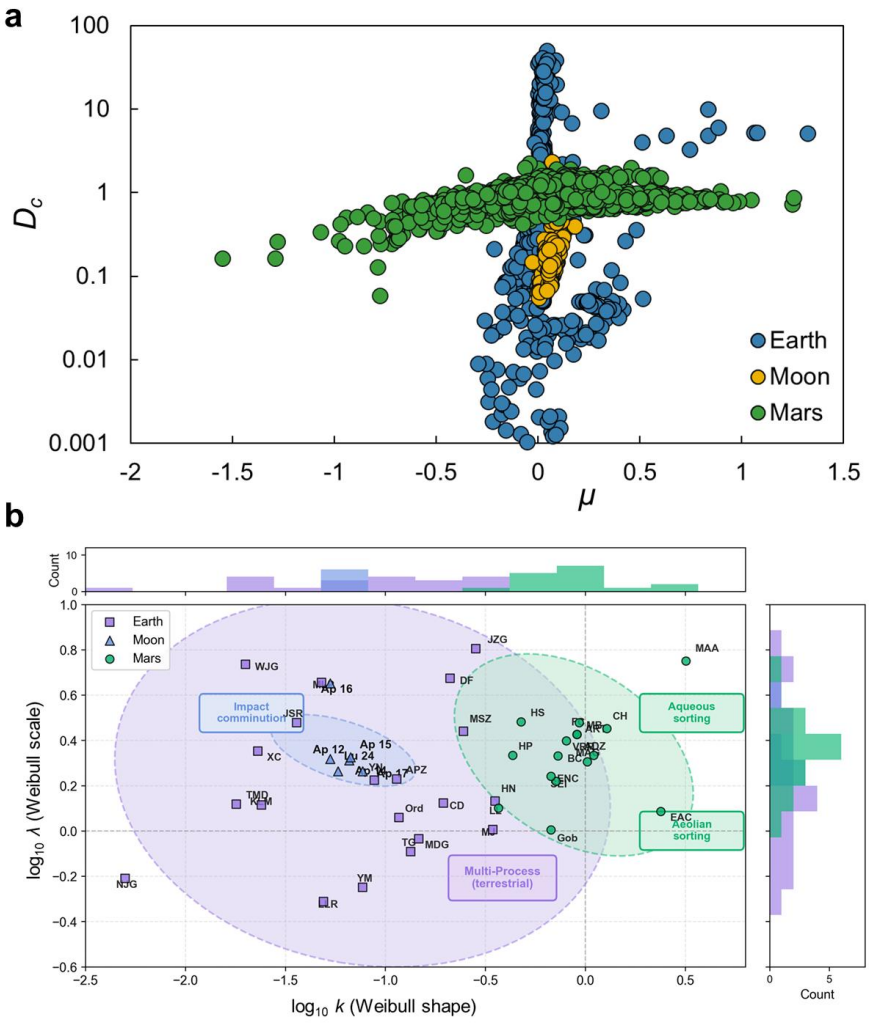

Figure 4. Clustering of UGSD parameters (μ, D_c) and Weibull distribution parameter (k, λ) for the PlanetGSD 1.0 sites. (a) (μ, D_c) clustering in the Earth, Moon, and Mars. (b) (k, λ) clustering for the three bodies. The marginal histograms show the distribution of k (top) and λ (right) for each planetary body. Notably, some terrestrial sites (e.g., mass movement deposits and periglacial colluvium) plot within the lunar field. These terrestrial samples are dominated by physical comminution with minimal chemical weathering, producing $k-\lambda$ signatures similar to lunar impact-gardened regolith. This overlap does not indicate identical surface processes, but rather demonstrates that the Weibull parameter space captures textural similarity across different genetic environments. Full site acronyms and geological descriptions are provided in Table S5 (Zhang, 2026).

2.5 From GSD measurements to random parameter fields

The preceding sections established a two-step transformation of raw grain-size measurements into a unified probabilistic framework (Fig. 5). First, the UGSD function converts the original GSD curves into a compact set of interpretable parameters (μ, D_c, n), reducing high-dimensional textural data to physically meaningful indices (Section 2.3). Second, at each sampling site, the population of μ values is characterized by a three-parameter Weibull distribution (Section 2.4), capturing the local spatial variability of fine-grain content in a probabilistic form.

This transformation fundamentally changes the nature of the dataset: from a conventional collection of discrete observations to a probabilistic GSD parameter library. In this library, each site is represented not by individual measurements but by the probability distribution of its granular characteristics (μ_0, k, λ), enabling stochastic simulation and extrapolation beyond the original sampling locations.

The probabilistic parameterization naturally leads to the concept of granular fields—spatially continuous representations of soil heterogeneity derived from the site-level Weibull parameters. Following the workflow of Zhang et al. (2022), μ-fields are generated through a three-step Monte Carlo procedure:

1) UGSD fitting of measured GSD data to obtain μ for each sample (as described in Section 2.3).
2) Weibull parameterization and Monte Carlo simulation – For each sampling site, the population of μ values is fitted to the three-parameter Weibull distribution (Section 2.4). An ensemble of N random μ values is then generated by sampling from this fitted distribution, with N specified by the user (default N = 1,000).
3) Spatial interpolation – The simulated μ values are assigned to a Digital Elevation Model (DEM)-registered grid via inverse-distance-weighted (IDW) interpolation (power = 2).

To ensure that the generated μ-fields respect realistic spatial continuity — an essential characteristic of geological and geomorphic units — the IDW interpolation is applied within spatial autocorrelation constraints defined by an exponential variogram model, with correlation length empirically determined from field observations (10 grid units at hillslope scale, 50 m at watershed

scale). Users can specify DEM resolution, interpolation radius, and the number of Monte Carlo realizations.

The resulting μ-fields provide spatially explicit estimates of grain-size characteristics that preserve local heterogeneity while enabling upscaling to regional scales. Because soil properties such as porosity, permeability, and shear strength can be expressed as functions of μ (Jiang et al., 2024), these granular fields can be further transformed into property fields for hydro-geophysical modeling, rover mobility simulation, or in-situ resource utilization assessment. The complete workflow, from raw measurements to μ-fields, is illustrated in Figure 5, and the implementation codes are provided in Code S1–S3.

Importantly, the probabilistic framework presented here—specifically, the finding that site-level μ populations follow a three-parameter Weibull distribution—offers a principled way to partially overcome the spatial sampling bias inherent in mission-specific datasets. Because the Weibull model captures the statistical variability of GSD parameters, it enables stochastic simulation of grain-size characteristics for unsampled terrains (e.g., Martian highlands or dune fields not yet visited by rovers). Thus, while PlanetGSD 1.0 does not eliminate spatial bias, it provides a probabilistic pathway to propagate observed heterogeneity to unobserved locations—a point we return to in Section 6.3.

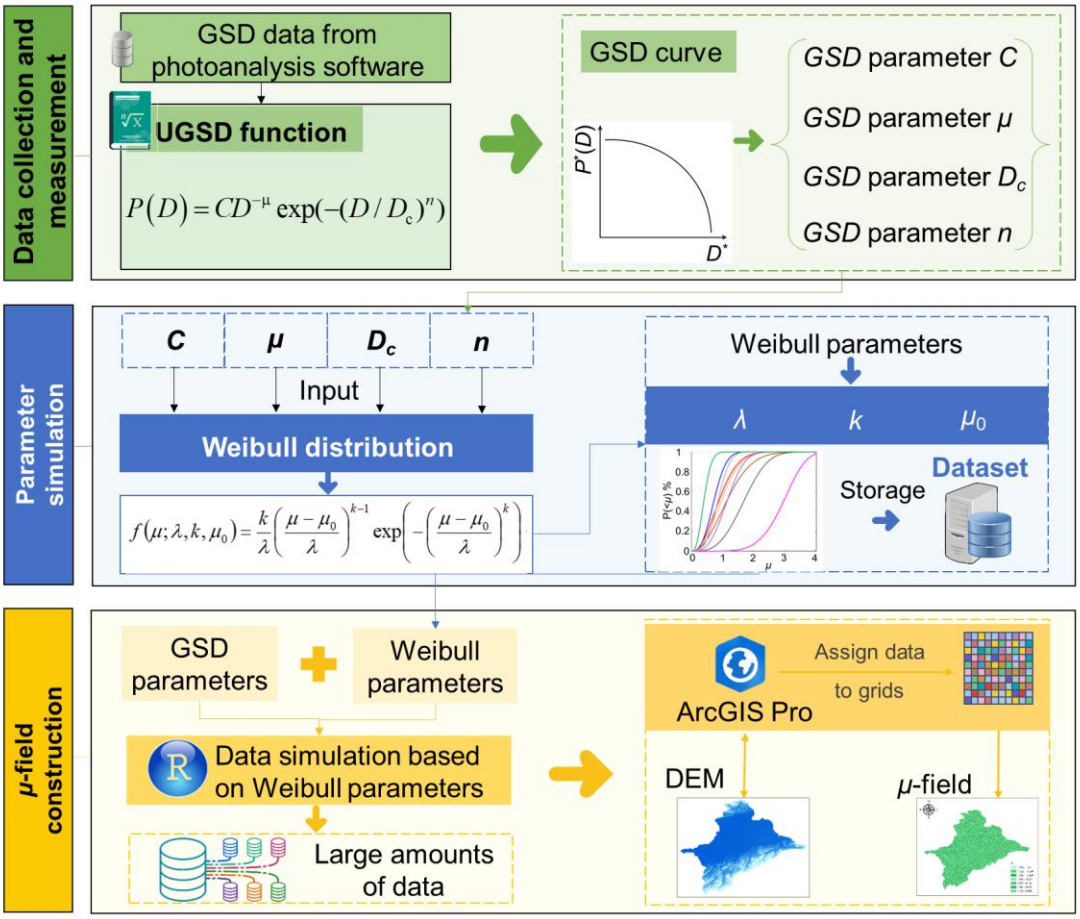

Figure 5. Workflow for generating spatially continuous μ-fields from raw GSD measurements. The process involves three stages: (1) UGSD fitting to extract μ for each sample; (2) Weibull parameterization of site-level μ populations and Monte Carlo simulation of random μ values; (3) IDW interpolation onto DEM grids to produce continuous μ-fields. This transformation converts conventional observational data into a probabilistic parameter library, enabling stochastic simulation and cross-site comparison.

3. GSD Dataset Construction

The UGSD function has harmonized the textural data to a set of analytical parameters (μ, D_c, n), while preserving the integrity of the measurements. This allows to construct a new-type dataset of soil from the traditional measured data of texture or moisture of soils.

3.1 Database structure and file inventory

3.1.1 Folder organization

All datasets are provided in Microsoft Excel format (.xlsx) to accommodate multi-sheet structures that organize GSD curves, metadata, and quality flags within single files. The complete

PlanetGSD 1.0 database is organized in a hierarchical folder structure with separate directories for data files, analysis codes, and documentation:

```
PlanetGSD_v1.0.0/
|-- data/
|-- codes/
|-- README.txt
```

3.1.2 File inventory

A complete inventory of all files in the PlanetGSD 1.0 repository is provided in Table 7, including file names, sizes, and brief descriptions.

Table 7. File inventory for PlanetGSD 1.0

File	Size	Description
S1_GSD_data_from_Earth.xlsx	145 KB	Raw cumulative GSD curves for 4,419 terrestrial samples, with geographic coordinates and metadata
S2_GSD_data_from_Moon.xlsx	36 KB	Raw cumulative GSD curves for 379 lunar samples, with mission identifiers and sample metadata
S3_GSD_data_from_Mars.xlsx	3,579 KB	Raw cumulative GSD curves for 1,729 Martian samples, with PDS image source IDs and geological context
S4_UGSD_parameters_and_Weibull.xlsx	107 KB	Two-sheet file: (Sheet 1) UGSD fitted parameters for all 6,527 samples; (Sheet 2) site-level Weibull statistics for 42 locations
S5_Mars_image_reference_numbers.xlsx	25 KB	Complete list of NASA PDS Product IDs for all 1,729 rover images used
S1_mu_field_generation.m	-	MATLAB code for stochastic μ-field generation
S2_UGSD_fitting.R	-	R code for data pre-processing and UGSD parameterization
S3_Weibull_fitting.py	-	Python code for Weibull distribution fitting and diagnostic plots
README.txt	-	Detailed dataset documentation, version history, and contact information

3.2 Database structure overview

The PlanetGSD 1.0 database is organized into five core datasets (S1–S5), each containing raw GSD measurements, fitted parameters, and associated metadata:

- Dataset S1 (Earth GSD) contains 4,419 terrestrial grain-size distributions with complete cumulative curves (17 size fractions from 0.0001 to 20 mm) and georeferenced metadata,

integrating field-collected samples (2,847) and literature-derived data (1,572).

- Dataset S2 (Lunar GSD) comprises 379 lunar grain-size distributions from eight sample-return missions (Apollo, Luna, Chang'e), with 9–15 size fractions per sample depending on mission protocol.
- Dataset S3 (Mars GSD) provides 1,729 grain-size distributions derived from semi-automated analysis of 2,160 rover images, with 8 logarithmic size bins (0.04–25 mm) and full PDS image traceability.
- Dataset S4 (UGSD and Weibull parameters) contains fitted UGSD parameters (C, μ, D_c, n) for all 6,527 samples (Sheet 1) and site-level three-parameter Weibull statistics (μ_0, k, λ) for 42 sampling locations (Sheet 2), with quality flags and geological context annotations.
- Dataset S5 (Site descriptions, acronyms, and dominant surface processes) provides a comprehensive list of all 42 sampling locations across Earth, the Moon, and Mars. For each site, the dataset includes: planet, acronym, full site name, geomorphic setting, and dominant surface process(es).

All datasets are provided in Microsoft Excel format (.xlsx) with multi-sheet structures to organize raw data, metadata, and quality flags within single files. Users are referred to the Figshare repository Zhang (2026) for direct data access.

3.3 Data dictionary

To facilitate correct interpretation and usage of the PlanetGSD 1.0 datasets, Table 8 provides complete field definitions for all variables appearing in Datasets S1–S5. The table includes field names, descriptions, data types, units, and example values, serving as a detailed reference to complement the database structure overview in Section 3.2.

Table 8. Data dictionary for PlanetGSD 1.0

Field	Description	Type	Unit	Example
sample_id	Unique sample identifier	String	—	Earth_DF_001
planet	Planetary body	String	—	Earth / Moon / Mars
site_name	Sampling site or mission name	String	—	Jiangjia Gully
latitude	Sampling latitude	Float (Decimal Degrees)	degrees	26.253
longitude	Sampling longitude	Float	degrees	103.117

elevation	Sampling elevation (Earth only)	Float	m	1,450
geomorphic_setting	Depositional or surface environment	String (Free text, see Table S9 for controlled vocabulary)	—	Debris flow
$D_{0.0001}$ to D_{20}	Cumulative mass percentage passing each size	Float (%)	%	2.3, 5.1, ...
$n_{fractions}$	Number of reported size fractions	Integer	—	17
data_source	Original data source	String	—	Field / Graf (1993)
μ	UGSD location parameter	Float	—	0.35
D_c	UGSD characteristic grain size	Float	mm	0.82
n	UGSD shape exponent	Float	—	0.48
R^2	Coefficient of determination for UGSD fit	Float	—	0.994
quality_flag	Data quality indicator	Integer (1 = good, 0 = flagged, 2 = excluded)	—	1 (good) / 0 (flagged)

4. Technical Validation

The validation of PlanetGSD 1.0 was conducted through a four-step procedure: (1) assessing UGSD fitting quality across all 6,527 samples; (2) evaluating Weibull parameterization quality at the site level; (3) verifying compiled data against original source publications; and (4) assessing inter-operator reproducibility for Martian image-derived measurements.

4.1 UGSD fitting quality

All 6,527 samples were fitted to the UGSD function (Eq. 3) using nonlinear least-squares optimization. Goodness-of-fit was quantified using the coefficient of determination (R^2). Figure 3 shows the rescaled UGSD curves for Earth (a), Moon (b), and Mars (c), demonstrating the unified fitting performance across planetary bodies.

Median R^2 values are 0.991 (Earth), 0.989 (Moon), and 0.983 (Mars), with 97.8% of all samples yielding $R^2 > 0.95$ (Table 9). The terrestrial and lunar populations exhibit strongly left-skewed R^2 distributions, with > 85% of samples achieving $R^2 > 0.98$. The slightly lower Martian median is attributable to the truncation of the fine tail below ~0.04–0.15 mm imposed by imaging resolution limits, which reduces the number of constraining data points available for curve fitting. Although a small number of lunar data points in Figure 3b deviate slightly from the reference curve, these points correspond predominantly to coarse grains ($D > 1$). Because soil physical properties are governed primarily by fine grains (Zhang et al., 2025), these minor deviations have negligible influence on the overall validity of our approach.

Following the outlier screening protocol (Section 2.3.2), 23 samples (0.4% of total) with $R^2 < 0.94$ were excluded after manual review due to data-entry errors or highly irregular curve shapes inconsistent with natural GSD. An additional 24 samples (0.4%) with $0.90 < R^2 < 0.94$ were retained but flagged (quality_flag = 0) in Dataset S4 for user awareness. The remaining 6,480 samples (99.3%) carry quality_flag = 1 and are recommended for analysis without restriction.

Comparison with alternative parameterizations. To demonstrate the advantage of the UGSD function, we fitted all 6,527 samples to three alternative models (lognormal, Rosin-Rammler, and Fredlund three-parameter) and found that UGSD achieved higher median R^2 values than all three alternatives (Table S10).

Table 9. UGSD fitting quality statistics

Planet	samples	Median R^2	Mean R^2	% with $R^2 > 0.95$	% with $R^2 < 0.90$	K-S test pass rate
Earth	4,419	0.991	0.987	98.5	0.3	98.2%
Moon	379	0.989	0.985	97.9	0.5	97.1%
Mars	1,729	0.983	0.978	96.2	1.1	94.8%
Total	6,527	0.988	0.984	97.8	0.5	97.3%

Kolmogorov-Smirnov test at significance level $\alpha = 0.05$; pass rate indicates percentage of samples where the UGSD fit is not significantly different from the measured distribution ($p > 0.05$).

The physical interpretation of UGSD parameters is supported by controlled experiments and field observations (Yong et al., 2017; Zhang et al., 2023): μ increases with fine-grain content (with $\mu > 1$ typical of aeolian settings, $\mu \approx 0.5$ for fluvial-aeolian transitions, and $\mu < 0.5$ for lag or impact-mixed deposits), D_c marks the grain-size break point related to transport energy, and n reflects coarse-tail sorting efficiency. All four UGSD parameters are physically constrained ($C > 0$, $D_c > 0$, $n > 0$); μ has no sign constraint because grain sizes below 1 μm yield negative ln(diameter) values, which is common in fine-grained samples. Statistically, the model is well-constrained: each GSD curve is fitted from > 10 independent size fractions (Earth/Moon) or > 150 grain measurements (Mars), and subsampling cross-validation confirms that the fitted UGSD parameters are stable, with standard deviations substantially smaller than the site-level variability across samples.

4.2 Weibull parameterization quality

Three-parameter Weibull distributions were fitted to site-level μ populations for all 42 sites (20 Earth, 6 Moon, 16 Mars). Fit quality was assessed using two complementary approaches: the

Kolmogorov-Smirnov (K-S) goodness-of-fit test at significance level $\alpha = 0.05$, and visual inspection of Q-Q (quantile-quantile) plots comparing observed μ quantiles against theoretical Weibull quantiles.

Of the 42 sites, 39 (93%) passed the K–S test, indicating that the three-parameter Weibull model adequately describes the μ population structure at most sampling locations. The three sites that did not pass are all terrestrial (MJ, DF, JZG) and exhibit bimodal μ distributions attributable to the mixing of geomorphically distinct lithological units within a single sampling area. Their Weibull parameters are retained in Dataset S4 (Sheet 2) with explanatory notes in the geological_description field; users may choose to exclude these sites or to fit separate Weibull distributions to the identified sub-populations.

4.3 Verification against published values

To verify the accuracy of compiled and processed data, we cross-checked subsets of samples from each planetary body against original source publications. This procedure tests for systematic errors introduced during data transcription, digitization, and processing.

(1) Lunar samples. Median grain sizes (D_{50}) for 50 Apollo samples were compared with values reported in the Lunar Sourcebook (Heiken et al., 1991). The mean absolute deviation between our compiled values and the Lunar Sourcebook was 4.2 μm (range: 0.3–11.8 μm), within the reported measurement uncertainty of ±5 μm for the original sieving analyses (Graf, 1993). No systematic bias was observed (mean signed deviation: +0.8 μm), confirming that digitization from the NASA Lunar Soils Grain Size Catalog did not introduce directional errors.

(2) Martian samples. D_{50} values for 30 Gale Crater samples were compared with those reported by Fedo et al. (2015) and Weitz et al. (2018) for overlapping locations. The mean absolute deviation was 0.08 mm (range: 0.01–0.22 mm), consistent with the inter-operator variability of ±15% documented for image-based grain-size measurements (see Section 4.4). No systematic offset was detected between our measurements and published values (mean signed deviation: −0.02 mm).

(3) Terrestrial samples. For 100 randomly selected literature-derived samples with data available in the original publications, we verified cumulative grain-size percentages at the D_{10}, D_{50}, and D_{90} percentile levels. All values matched within rounding error (mean absolute difference: 0.3%;

maximum: 0.8%). For the subset of 127 samples where original studies used hydrometer rather than laser diffraction methods, clay content (<2 μm) agreed within ±2.1% (mean absolute deviation) when cross-validated against laser diffraction measurements in overlapping datasets, consistent with known inter-method variability (Konert & Vandenberghe, 1997).

These checks confirm that data transcription and processing introduced no systematic errors into PlanetGSD 1.0.

4.4 Inter-operator reproducibility of Martian measurements

Because all Martian GSD data were derived from semi-automated image analysis, inter-operator variability represents an important component of measurement uncertainty. To quantify this variability, three independent analysts processed identical image sets (n = 30 images spanning all four landing sites) using the same grain-segmentation software (Shi et al., 2021; Zhao et al., 2023).

Results are summarized in Table 10. Inter-operator agreement was assessed using three metrics: (i) the coefficient of variation (CV) of D_{50} across the three analysts; (ii) the mean absolute percentage difference in cumulative grain-size fractions; and (iii) the resulting variation in fitted UGSD parameters.

Table 10. Inter-operator reproducibility for Martian image-based grain-size measurements (n = 30 images, 3 analysts)

Metric	Mean	Range	Notes
CV of D_{50}	8.30%	2.1–18.7%	Higher CV for images with more grain overlap
Mean absolute % difference in cumulative fractions	6.20%	1.5–14.3%	Largest differences in finest resolvable fraction
CV of UGSD parameter μ	5.70%	1.2–12.4%	Propagated from grain boundary delineation differences
CV of UGSD parameter D_c	7.10%	2.3–15.8%	—
R^2 variation	±0.008	0.001–0.021	Fitting quality robust to operator differences

The primary source of inter-operator variability is grain boundary delineation for overlapping or partially obscured grains. Images with >30% grain overlap showed systematically higher CV (mean: 14.2%) compared to images with <10% overlap (mean: 3.8%). This operator-dependent uncertainty is propagated into the UGSD parameter confidence intervals reported in Dataset S4.

These inter-operator variability estimates (CV ≈ 5–8% for key parameters) are comparable to

those reported in previous Mars grain-size studies using similar methods (Fedo et al., 2015; Weitz et al., 2018) and are substantially smaller than the inter-site variability observed across the four Martian landing sites, confirming that measurement uncertainty does not obscure geological signal in the dataset. However, caution is warranted in attributing the observed inter-site variability solely to geological signals. Alternative factors may also contribute to the relatively small variability observed among the four Martian landing sites.

First, the inherent resolution limits of image-based measurements (Table 3) truncate the fine tail of all Martian GSD curves below ~0.04–0.15 mm, depending on the instrument. This right-censoring reduces the number of constraining size fractions available for UGSD fitting, which may artificially reduce apparent inter-site variability in the derived parameters—meaning that some of the observed similarity across sites could reflect what the instruments cannot resolve, rather than a true geological signal.

Second, the four landing sites—Gale, Jezero, Gusev, and Meridiani—are all located in low-latitude regions with broadly similar aeolian regimes. This geographic concentration does not capture the full range of Martian surface textural diversity (e.g., high-latitude mantling terrains, dust-dominated regions, or polar layered deposits).

We note that while Martian dust storms are known to mobilize and transport fine grains globally (Kahre et al., 2017; Senel et al., 2021), the grain-size fractions most affected by this process (silt and clay, <62 μm) are below the detection limits of the rover imagers used in this study (Table 3). Therefore, the potential homogenizing effect of dust storms cannot be directly evaluated with our data and is not invoked as an explanation for the observed inter-site variability.

The inter-operator variability reported above (CV ≈ 5–8 %) inherently captures the combined effects of operator-dependent grain boundary delineation and the spherical shape approximation, as both contribute to the observed dispersion in fitted parameters.

5. Illustrative applications and open questions

This section demonstrates three representative applications of PlanetGSD 1.0 and identifies open questions for future research.

5.1 Cross-planetary comparison of GSD signatures and μ-field generation

The UGSD parameterization enables not only direct comparison of GSD characteristics but

also continuous representations of spatial heterogeneity. Using the site-level Weibull statistics derived in Section 2.4 and the Monte Carlo workflow described in Section 2.5, we generated μ-fields—spatially explicit maps of the UGSD parameter μ—for three representative sites, one from each planetary body (Fig. 6).

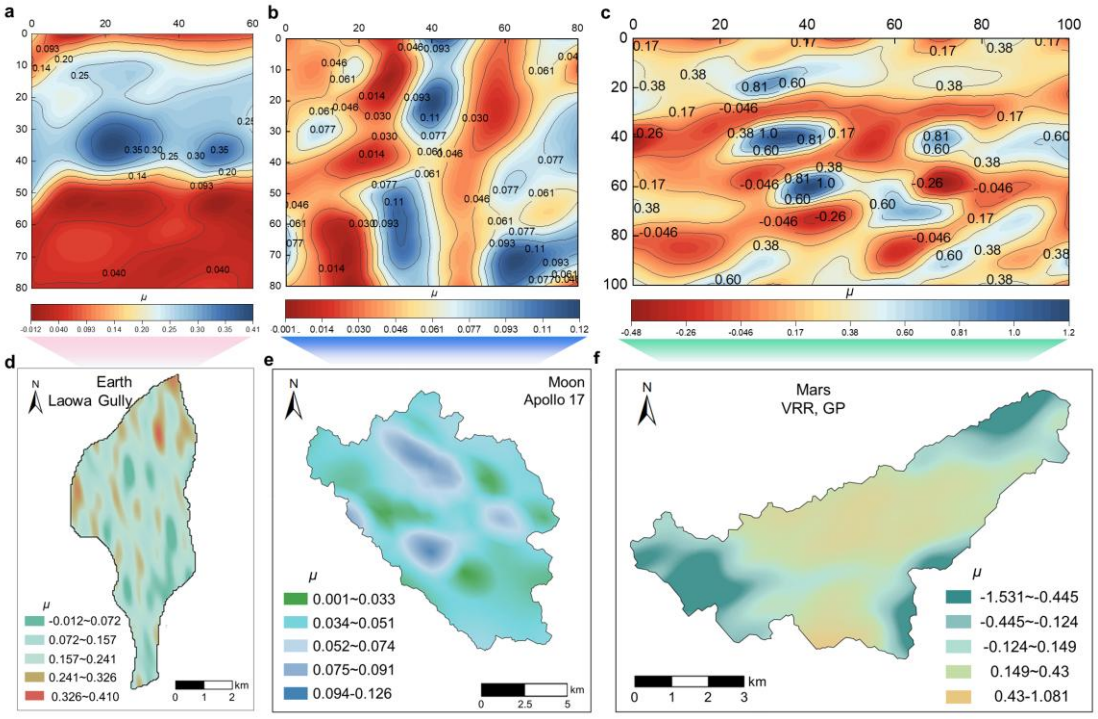

Figure 6. Spatial distribution of the UGSD parameter μ across three planetary bodies. The top row (a–c) shows μ-field maps for representative sites: (a) Laowa Gully, Earth; (b) Apollo 17 Taurus-Littrow valley, Moon; (c) Vera Rubin Ridge–Greenheugh Pediment area, Gale Crater, Mars. The bottom row (d–f) displays enlarged views of the corresponding areas at higher spatial resolution: (d) Laowa Gully; (e) Apollo 17 site; (f) Vera Rubin Ridge–Greenheugh Pediment area.

The Laowa Gully (Earth) simulation produces a spatially homogeneous μ-field (range = 0.422), representing the null hypothesis of uniform grain-size distribution under uniform pedogenic processes (Zhang et al., 2022). This pattern can serve as a baseline for detecting anthropogenically-induced or naturally-occurring spatial heterogeneity in regions where limited field data suggest uniformity. The Apollo 17 Taurus-Littrow valley (Moon) simulation exhibits fine-scale random heterogeneity (~100 m correlation length) without persistent spatial gradients (range = 0.125, Moran's I = 0.08). This pattern represents the stochastic mixing expected from impact gardening

693 — dominant regolith — forming process on airless bodies—where meteoroid bombardment
694 produces spatially uncorrelated grain-size distributions (McKay et al., 1991; Szalay et al., 2019).
695 The Vera Rubin Ridge–Greenheugh Pediment area in Gale Crater (Mars) simulation shows a
696 pronounced spatial gradient (range = 1.612, Moran's I = 0.76). This pattern demonstrates how
697 directional geomorphic features (e.g., valley networks, crater rims, or lithological contacts extracted
698 from DEMs) can induce systematic spatial organization in modeled grain-size distributions.
699 These contrasting patterns illustrate the flexibility of our spatial parameterization approach
700 (Section 2.4) in generating diverse heterogeneity patterns under varying geomorphic constraints.
701 While direct field validation of these simulations remains challenging for planetary surfaces due to
702 the scarcity of in-situ grain-size transects at comparable scales, the demonstrated patterns provide
703 testable hypotheses for future mission targets. For terrestrial applications (e.g., Laowa Gully), the
704 same methodology can generate spatially explicit heterogeneity fields when user-provided DEM
705 data are available, with field validation deferred to future work. The provided code (Code S1)
706 enables users to generate analogous μ-fields for any study site using their own DEM data.
707 The overlap between the lunar field and a subset of terrestrial sites (e.g., Jiangjia Gully, Kunlun
708 Mountains) in Figure 4 warrants explanation. Lunar regolith is dominated by impact comminution
709 (Devine et al., 1982; King, 1977), producing extremely fine grained, poorly sorted distributions with
710 low μ values (Table 5). The overlapping terrestrial samples are dominated by physical fragmentation
711 under cold or arid conditions (e.g., glacial till, periglacial colluvium, debris flow deposits), where
712 chemical weathering is minimal (Attal et al., 2015; Weinman et al., 2011). Their Weibull parameters
713 (low k, low μ) quantitatively mirror the textural signature of impact dominated fragmentation. This
714 overlap thus highlights a fundamental insight: different geological processes that share a dominance
715 of physical comminution over chemical alteration can produce similar grain size distribution
716 fingerprints (Ovalle & Dano, 2011; Toneva & Peukert, 2007). The parameter space therefore
717 discriminates process regimes (physical vs. chemical weathering) rather than planetary provenance
718 per se. We have elaborated this interpretation in the revised Figure 4 caption and Section 6.3.
719 For a complementary view of the parameter space itself, the distribution of all 6,527 samples
720 in μ-D_c space is provided in Figure S2, where the three planetary populations occupy largely distinct
721 domains: Earth spans the μ range from −0.012 to +0.41, Moon occupies a narrow near-zero range

(0.001 to 0.126), and Mars displays the broadest μ range (from −1.531 to +1.081). The non-overlapping Weibull shape parameter k ranges (Table 6) further confirm systematic differences among the three bodies, supporting the utility of the k–λ parameter pair as a diagnostic descriptor for cross-planetary discrimination.

5.2 Regolith simulant benchmarking and spatial visualization

PlanetGSD 1.0 provides reference GSD parameters against which lunar and Martian simulant materials can be quantitatively evaluated. Users can compute the UGSD and Weibull parameters of candidate simulants using the provided analysis codes (Codes S1–S3) and compare them with the site-level parameter ranges documented in Dataset S4. For example, the Weibull parameters for Apollo mare sites (k = 0.053–0.077, λ = 1.83–4.48) define a target envelope for lunar simulants intended to replicate mare regolith; simulants falling outside this envelope can be identified as texturally non-representative. Table 11 provides example analog matches linking target terrains to PlanetGSD reference sites to facilitate such comparisons.

It is important to note that Figure 6 presents an interpolated μ field generated from the Monte Carlo workflow described in Section 2.5, which already accounts for spatial autocorrelation via an exponential variogram model to ensure realistic spatial continuity.

However, even with spatial autocorrelation, the current interpolation does not explicitly incorporate mapped geologic or geomorphic boundaries (e.g., lithological contacts, fault lines, or geomorphic units). Purely statistical spatial correlation without such contextual constraints is of limited value for predictive mapping, especially with sparse data points.

Therefore, the μ field in Figure 6 is presented as a continuous visual summary of sparse point measurements, not as a final predictive soil map. It serves two modest but necessary purposes: (i) it provides an intuitive, spatially continuous visualization of the general trend of grain size, helping to generate hypotheses or guide future sampling; and (ii) it offers a baseline continuous covariate for future geostatistical models (e.g., kriging with external drift) or machine-learning approaches that will explicitly integrate geologic units, geomorphic boundaries, or remote-sensing data (planned for PlanetGSD 2.0).

Future work (PlanetGSD 2.0) will explicitly incorporate such spatial constraints to generate

more robust, context-aware predictions.

5.3 Open questions and future directions

PlanetGSD 1.0 opens several avenues for future research, but it also has limitations that should guide its use and development:

(1) Gaps in geologic and environmental representation. The terrestrial component is dominated by gravity-driven and fluvial deposits (debris flows, landslides, alluvial fans), while other geomorphic settings remain underrepresented. These include playa lakes, aeolian dune fields, glacial outwash plains, coastal sediments, tropical weathering profiles, and cold-desert soils (e.g., Antarctic dry valleys). The current geographic bias toward Asia (73% of samples) is less critical than the underrepresention of these process-based endmembers. Lunar samples are predominantly from nearside mare regions, with limited representation of highland, farside, and polar materials. Martian samples are limited to four low-latitude landing sites, lacking high-latitude and dust-dominated terrains. Priority targets for future data collection include underrepresented geomorphic settings on Earth (as listed above), lunar farside and polar regions, and new Martian landing sites (e.g., Zhurong at Utopia Planitia, and high-latitudes sites such as Arcadia Planitia). A future version of PlanetGSD (e.g., PlanetGSD 2.0) should further expand farside and polar lunar sampling beyond the current CE-6 sample and include Martian terrains that have not yet been visited by rovers (e.g., highlands, polar regions, dune fields).

(2) Analytical resolution limits. Martian GSD curves are right-censored below 40–150 μm depending on imaging instrument, truncating the fine tail of distributions. Validation against in-situ measurements from future sample-return missions will be essential to quantify systematic biases introduced by this truncation.

(3) Process attribution. While μ shows strong empirical correlations with depositional environments (e.g., $\mu > 1.0$ for aeolian-dominated settings; $\mu < 0.5$ for lag deposits), the mechanistic links between UGSD parameters and specific surface processes remain incompletely understood. The overlapping parameter fields observed in Figure 4 (e.g., lunar samples falling within the terrestrial envelope) indicate that the k–λ space captures textural similarity across different genetic environments rather than providing unambiguous process discrimination. Such overlaps should not be overinterpreted as evidence of identical formative processes. The primary value of the PlanetGSD

779 parameterization lies in (i) compact representation of GSDs, (ii) quantitative cross-comparison
780 across planetary bodies and analytical methods, and (iii) stochastic field generation (Section 2.5).
781 Robust process attribution requires integration with independent geological, mineralogical, and
782 geochemical context.

783 **(4) Integration with remote sensing.** Orbital datasets (thermal inertia, hyperspectral
784 mineralogy, radar backscatter) provide indirect constraints on surface texture at global scales.
785 Calibrating these proxies against UGSD parameters could enable planetary-scale mapping of grain-
786 size characteristics beyond the spatial reach of rover-based measurements.

787 **(5) Climate change effects.** Long-term terrestrial GSD records may capture the influence of
788 changing precipitation patterns and temperature regimes on weathering and sediment transport.
789 Coupling PlanetGSD with climate datasets could provide insights into how surface processes
790 respond to environmental forcing—a question relevant to both Earth system science and interpreting
791 Martian paleoclimate records.

792 **(6) Coupling with orbital remote sensing.** Future versions of PlanetGSD should integrate
793 thermal inertia and photometric datasets (see limitation 7 in Section 6.3). Key open questions
794 include: Can thermal inertia predict the UGSD parameter μ at rover landing sites? Do photometric
795 phase curves correlate with Weibull k or λ? Answering these questions would enable extrapolation
796 of GSD characteristics from point-scale measurements to regional and global scales.

797 **6. Usage Notes**

798 **6.1 Recommended applications**

799 PlanetGSD 1.0 is designed to support the following research applications:

800 **(1) Comparative planetology.** The unified UGSD parameterization enables direct comparison
801 of grain-size signatures across Earth, Moon, and Mars. The $k - \lambda$ parameter domains (Figure 4)
802 provide a quantitative framework for classifying samples by planetary provenance.

803 **(2) Regolith simulant benchmarking.** Lunar and Martian simulants can be quantitatively
804 compared with natural planetary materials by computing their UGSD and Weibull parameters using
805 Codes S2 and S3. Dataset S4 provides reference parameter values for all sampled sites to facilitate
806 such comparisons.

807 **(3) Landing site assessment.** While geologic mapping and remote sensing analyses (e.g.,

spectroscopy, thermal inertia) remain the primary methods for determining geological context, pre-landing GSD estimates can be generated as ancillary data by identifying PlanetGSD entries with geological context similar to candidate landing sites. Such estimates provide additional constraints on surface mechanical properties (e.g., trafficability, drilling resistance) that are not directly accessible from orbital data alone. Table 11 provides example analog matches to illustrate this approach.

Table 11. Example analog matches for landing-site GSD estimation

Target terrain	PlanetGSD analog	Analog k range	Analog λ range
Phyllosilicate-bearing plains	Gale Crater Murray formation (MAG, VRR)	1.8–2.5	2.0–2.5
Olivine-rich volcanic terrain	Gusev Columbia Hills (LL, HP)	0.4–0.6	1.3–2.2
Sulfate-rich evaporites	Meridiani Planum	0.5–0.9	1.5–3.0
Deltaic deposits	Jezero Crater delta	1.0–1.5	2.0–4.0

(4) Stochastic granular field generation. Using Code S1 and the fitted Weibull parameters (Dataset S4), users can generate spatially continuous μ-fields for geotechnical modeling, rover mobility simulation, or landscape evolution studies. The workflow accepts user-supplied DEM data and outputs μ-fields as GeoTIFF files compatible with standard Geographic Information System (GIS) software. Users can specify DEM resolution, interpolation radius, and the number of Monte Carlo realizations.

(5) Integration with remote sensing. The μ-field framework provides sub-pixel grain-size information that complements orbital datasets such as thermal inertia and radar backscatter. Coupling μ-fields with geophysical transfer functions enables the derivation of spatially continuous porosity and cohesion estimates for engineering applications at future landing sites.

6.2 Methodological guidance for users

To ensure robust application of PlanetGSD 1.0, users should consider the following guidance:

Parameter interpretation. While the UGSD parameter μ correlates positively with fine-grain content and the Weibull shape parameter k reflects the width of the μ distribution, these are empirical descriptors derived from curve fitting. Direct translation of μ to physical quantities such as porosity, permeability, or shear strength should be undertaken with caution. Such applications are most robust

when μ is coupled with complementary site-specific data or established geotechnical transfer functions (e.g., Zhang et al., 2022; Jiang et al., 2024). The parameters are primarily intended for comparative purposes—e.g., ranking sites by relative fineness or sorting—rather than for absolute grain-size inference.

Martian data handling. All Martian GSD curves are right-censored below instrument-specific resolution limits (Table 2). Users conducting analyses sensitive to the fine fraction (<0.1 mm) should treat cumulative percentages in the smallest size bin as lower bounds rather than absolute values. For statistical applications, we recommend incorporating the truncation thresholds (provided in Dataset S3 metadata) into uncertainty estimates. Sample-return measurements will be essential for future validation of these image-derived GSDs.

Lunar sample mass. Lunar samples with mass <5 g are flagged (quality_flag = 0) in Dataset S2 metadata due to potential sampling bias toward coarser fractions. Users requiring unbiased fine-fraction statistics should apply mass-based filtering to retain only quality_flag = 1 samples.

6.3 Limitations and caveats

Users should be aware of three principal limitations when using PlanetGSD 1.0, organized around data completeness, data integrity, and appropriate application scale.

6.3.1 Data completeness: sampling bias and future refinements

PlanetGSD 1.0 reflects the historical focus of planetary exploration. Lunar samples are predominantly from nearside, low-latitude mare regions, with highland and farside materials underrepresented. Martian grain-size estimates are restricted to rover-accessible terrains (often flat, low-risk, and rock-free surfaces). The terrestrial component is biased toward poorly sorted deposits (debris flows, colluvium, moraines), with well-sorted aeolian dunes and littoral sands underrepresented.

Implication for users. This sampling is not spatially random and does not represent the full geological diversity of either body. However, rather than ignoring this bias, PlanetGSD 1.0 makes it explicit: each sample retains its geographic metadata and landing-site context; site-level Weibull parameters describe local μ variability without extrapolating to unsampled landscapes. The unified UGSD framework renders this bias visible and quantifiable—a prerequisite for addressing it in

future versions.

Future refinements. As more samples become available—particularly from the lunar highlands, lunar farside (e.g., Chang'e-6), and diverse Martian sedimentary environments (e.g., Jezero delta, Medusae Fossae Formation)—the (μ, D_c) and (k, λ) points for each planetary body are expected to enlarge the converge (cf. Figure 4a and 4b), enabling more robust probabilistic comparisons. Two important remote-sensing techniques not yet incorporated—thermal inertia (Fergason et al., 2006; Mellon et al., 2000) and photometric Hapke modeling (Hayne et al., 2017)—offer complementary constraints on effective grain size at regional to global scales. Coupling PlanetGSD 1.0 with these datasets is a high-priority direction for PlanetGSD 2.0.

6.3.2 Data integrity: missing and censored grain-size fractions

GSDs in PlanetGSD 1.0 are incomplete in two distinct ways: physical absence of certain size fractions due to surface processes, and censoring due to instrument resolution limits.

Physical missing fractions. In some environments, specific grain-size classes may be genuinely absent from the sediment (e.g., the 2–4 mm gap in Jiangjia Gully debris-flow deposits). Such physical gaps are not measurement errors but real textural signatures. Users should be aware that these physical gaps may affect the shape of the grain-size distribution and, consequently, the fitted UGSD parameters (μ, D_c, n). When such gaps are present, the three-parameter Weibull distribution may provide a poorer fit, and users are encouraged to inspect the per-site grain-size curves directly rather than relying solely on summary parameters.

Instrument-induced censoring. All Martian GSD curves are right-censored below instrument-specific detection limits (Table 2). No current Martian imager can resolve clay-sized grains (<2 μm) or silt-sized grains (<40–150 μm depending on the instrument). Applying a conservative 5-pixel threshold to MAHLI at 14 μm/pixel yields a practical lower bound of ~70 μm, meaning very fine sand (62.5–125 μm) is only marginally resolvable. The presence of ubiquitous Martian dust further degrades effective resolution: dust aggregates (<100 μm) on grain surfaces obscure true boundaries, and uniform dust mantling reduces image contrast, making segmentation thresholds less reliable. To mitigate these effects, we use images acquired after dust removal tool (DRT) brushing when available; for unbrushed targets, we manually mask regions with obvious dust

aggregates. The fine tail (< minimum detectable size) is truncated, not extrapolated (Section 2.3.2). Users should treat fine-fraction statistics as conservative lower bounds.

Implication for UGSD and Weibull parameters. Censoring affects μ (which correlates with fine-grain content) more severely than D_c or n. The three-parameter Weibull fit partially accommodates this through the location parameter μ_0—a shifted minimum μ value that reflects the effective lower bound of the measurement. Anomalously high μ_0 values for Martian sites (Table 6) directly encode instrument-induced censoring, making the bias explicit rather than hidden.

6.3.3 Application scale: distribution parameters vs. site-specific transfer functions

PlanetGSD 1.0 supports two distinct modes of use, corresponding to different spatial scales and scientific questions. Regional-scale (cross-planetary) comparison.

Regional-scale (cross-planetary) comparisons. Users should rely on Weibull distribution parameters (k, λ) aggregated at the site or sample level. These parameters describe the shape and spread of μ distributions across multiple samples within a geomorphic or geological unit. Their strength is that they capture textural similarity across different genetic environments (e.g., terrestrial physical comminution sites overlapping with lunar impact-gardened regolith in k–λ space), enabling process-independent comparative planetology. Users should not interpret individual μ or D_c values at this scale, as these are sensitive to local sampling and measurement bias.

Local-scale (process) analysis. At the scale of individual sampling sites or sub-meter transects, users should use raw UGSD parameters (μ, D_c, n) or the full grain-size distribution curves. These can serve as input to transfer functions that relate grain-size statistics to specific surface processes. For example, the ratio μ/D_c may discriminate between debris-flow and fluvial deposits; the shape exponent n distinguishes impact-dominated regolith ($n \approx 1$) from aeolian-reworked sediments ($n > 1$). However, as noted in Section 5.3, mechanistic attribution of specific surface processes from UGSD or Weibull parameters alone is not yet reliable without site-specific validation (e.g., independent geomorphic or geochemical evidence).

What not to do. Spatial extrapolation beyond sampled locations (e.g., using the μ-fields generated by Code S3) is provided for visualization and hypothesis generation only. In areas with

sparse sampling or complex lithological boundaries, interpolated values may not accurately represent local grain-size characteristics. Users requiring spatially continuous predictions should couple PlanetGSD 1.0 with orbital remote-sensing data (thermal inertia, photometry) as planned for PlanetGSD 2.0.

6.4 Data updates and community contributions

PlanetGSD is conceived as a living database that will evolve with new missions and community input. Future versions (e.g., PlanetGSD 2.0) will incorporate:

- Additional terrestrial environments (aeolian dunes, glacial outwash, coastal sediments, loess sequences);
- Samples from ongoing and future missions (Mars Sample Return, Artemis lunar samples, OSIRIS-REx and Hayabusa2 asteroid samples);
- Extended grain-size ranges from improved analytical techniques;
- User-contributed datasets meeting quality standards.

Researchers wishing to contribute data should follow the standardized format:

- Complete cumulative grain-size percentages at standard size classes (or at original measured sizes with clear documentation);
- Geographic coordinates (WGS84 for Earth, Moon 2000/Mars 2000 for extraterrestrial bodies);
- Site metadata (geomorphic setting, sampling depth, lithology, depositional environment);
- Clear documentation of analytical methods and quality control procedures.

All contributed datasets will undergo the same validation and quality control checks applied to the original PlanetGSD compilation (see Technical Validation) before inclusion. Version history and a detailed changelog will be maintained in the repository README file.

7. Conclusions

The PlanetGSD 1.0 dataset establishes a unified framework for characterizing grain-size distributions across Earth, the Moon, and Mars, revealing that soils from these disparate planetary bodies conform to a single functional form—the Universal Grain-Size Distribution (UGSD). Key

conclusions include:

1. **UGSD universality.** The UGSD function fits 6,527 samples from three planets with a median R^2 = 0.988, demonstrating that granular configuration transcends genetic and environmental differences. The exponent n serves as a scaling factor that classifies soils into categories that reflect textural organization inherited from formation processes.
2. ***μ* as process indicator.** The UGSD parameter μ effectively discriminates among depositional environments: μ > 1.0 (aeolian-dominated), $\mu \approx 0.5$ (fluvial-aeolian transition), $\mu < 0.5$ (lag deposits, impact-influenced mixtures). Negative μ values record fine-fraction deficiencies resulting from winnowing or analytical truncation.
3. **Probabilistic behavior.** Site-level μ populations follow three-parameter Weibull distributions, enabling the stochastic generation of spatially continuous μ-fields that preserve local heterogeneity while allowing upscaling to regional scales.
4. **Cross-planetary comparison.** The μ-D_c parameter space clusters soils by planetary body and process regime: Earth spans the widest range (diverse processes); the Moon occupies a narrow, near-zero μ range (impact comminution); Mars displays the broadest μ range (complex aqueous-aeolian-volcanic interactions).
5. **Applications.** Granular fields support hydro-geophysical modeling, ISRU targeting, and mission planning by providing spatially explicit estimates of soil properties derived from limited point measurements.

The UGSD framework thus provides a quantitative, mechanistically grounded approach to soil characterization that unifies terrestrial and extraterrestrial regolith studies. By reducing high-dimensional texture data to interpretable indices and enabling probabilistic field generation, it bridges the gap between limited point measurements and spatially explicit process modeling, offering a new paradigm for planetary surface science.

Data Availability

The complete PlanetGSD 1.0 database is openly available at Figshare (https://doi.org/10.6084/m9.figshare.32362083, Zhang, 2026) under the Creative Commons Attribution 4.0 International License (CC BY 4.0). During peer review, the data can be accessed

anonymously via the same URL. The repository contains five Excel files:

- 'S1_GSD_data_from_Earth.xlsx': 4,419 terrestrial samples with cumulative curves and metadata

- 'S2_GSD_data_from_Moon.xlsx': 379 lunar samples from eight sample-return missions

- 'S3_GSD_data_from_Mars.xlsx': 1,729 Martian samples derived from rover image analysis

- 'S4_UGSD_parameters_and_Weibull.xlsx': UGSD parameters for all samples and site-level Weibull statistics

- 'S5_Mars_image_reference_numbers.xlsx': PDS image Product IDs for all 1,729 Mars samples

Code Availability

All analysis codes are openly available in the PlanetGSD 1.0 repository at Figshare (https://doi.org/10.6084/m9.figshare.31569616). The repository contains three Python codes:

- Code S1 (MATLAB): Stochastic μ-field generation (requires MATLAB R2020a or later)

- Code S2 (R): Data pre-processing and UGSD parameterization (requires R $\geq$ 4.0)

- Code S3 (Python): Weibull distribution fitting and diagnostic plots (requires Python $\geq$ 3.9 with NumPy, SciPy, Matplotlib, and Pandas)

The codes require Python $\geq$ 3.9 with NumPy, SciPy, Matplotlib, and Pandas. Version history is maintained in the Figshare repository.

Acknowledgements

This research is supported by the National Youth Science Foundation (No. 42501091); the NSFC, China (42322703, 42271092); the Open Funds of the Key Laboratory of Mountain Hazards and Engineering Resilience, CAS (No. KLMHER-K22); the China National Postdoctoral Program for Innovation Talent (No. GZC20231347); the China Postdoctoral Science Foundation (No. 2023M741997).

Competing Interests

The authors declare no competing interests.

Author Contributions

J. Z.: Conceptualization, Data curation, Formal analysis, Software, Visualization, Writing – Original Draft. **Y. L.:** Data curation, Validation, Writing – Review & Editing.

Reference

Aileen Yingst, R., Edgett, K., Williams, R., Hamilton, V., Mangold, N., & Bridges, N. (2014, May). The Texture of Mars: Observations of Rock and Outcrop Targets Over 360 Martian Sols at the Gale Field Site. In EGU General Assembly Conference Abstracts (p. 7263).

Attal, M., Mudd, S. M., Hurst, M. D., Weinman, B., Yoo, K., & Naylor, M. (2015). Impact of change in erosion rate and landscape steepness on hillslope and fluvial sediments grain size in the Feather River basin (Sierra Nevada, California). Earth Surface Dynamics, 3(1), 201-222. https://doi.org/10.5194/esurf-3-201-2015, 2015.

Batjes, N.H. (2009). Harmonized soil profile data for applications at global and continental scales: updates to the WISE database. Soil Use and Management 5:124 – 127. http://dx.doi.org/10.1111/j.1475-2743.2009.00202.x.

Beegle, L., Bhartia, R., White, M., DeFlores, L., Abbey, W., Wu, Y. H., ... & Sobron, P. (2015, March). SHERLOC: Scanning habitable environments with Raman & luminescence for organics & chemicals. In 2015 IEEE aerospace conference (pp. 1-11). IEEE. http://dx.doi.org/10.1109/AERO.2015.7119105.

Bell III, J. F., Maki, J. N., Mehall, G. L., Ravine, M. A., Caplinger, M. A., Bailey, Z. J., ... & Wolff, M. J. (2021). The Mars 2020 perseverance rover mast camera zoom (Mastcam-Z) multispectral, stereoscopic imaging investigation. Space science reviews, 217(1), 24. https://doi.org/10.1007/s11214-020-00755-x.

Bell III, J. F., Godber, A., McNair, S., Caplinger, M. A., Maki, J. N., Lemmon, M. T., ... & Deen, R. G. (2017). The Mars Science Laboratory Curiosity rover Mastcam instruments: Preflight and in-flight calibration, validation, and data archiving. Earth and Space Science, 4(7), 396-452. https://doi.org/10.1002/2016ea000219.

Bittelli, M., Campbell, G. S., & Flury, M. (1999). Characterization of grain-size distribution in soils with a fragmentation model. Soil Science Society of America Journal, 63(4), 782-788. https://doi.org/10.2136/sssaj1999.634782x.

Blott, S. J., & Pye, K. (2001). GRADISTAT: a grain size distribution and statistics package for the analysis of unconsolidated sediments. Earth surface processes and Landforms, 26(11), 1237-1248. https://doi.org/10.1002/esp.261.

Bockheim, J. G., Hartemink, A. E. (2013). Classification and distribution of soils with lamellae in the USA. Geoderma, 206, 92-100. https://doi.org/10.1016/j.geoderma.2013.04.014.

Cabrol, N. A., Grin, E. A., Carr, M. H., Sutter, B., Moore, J. M., Farmer, J. D., ... & de Goursac, O. (2003). Exploring Gusev Crater with Spirit: Review of science objectives and testable hypotheses. Journal of Geophysical Research: Planets, 108(E12). https://doi.org/10.1029/2002je002026.

Carrier III, W. D. (2003). Grain size distribution of lunar soil. Journal of Geotechnical and Geoenvironmental Engineering, 129(10), 956-959.

Chaney, N. W., Wood, E. F., McBratney, A. B., Hempel, J. W., Nauman, T. W., Brungard, C. W., & Odgers, N. P. (2016). POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. Geoderma, 274, 54-67. https://doi.org/10.1016/j.geoderma.2016.03.025.

Christensen, P. R., Morris, R. V., Lane, M. D., Bandfield, J. L., & Malin, M. C. (2001). Global mapping of Martian hematite mineral deposits: Remnants of water-driven processes on early Mars. Journal of Geophysical Research: Planets, 106(E10), 23873-23885. https://doi.org/10.1029/2000je001415.

Colwell, J. E., Batiste, S., Horányi, M., Robertson, S., & Sture, S. (2007). Lunar surface: Dust dynamics and regolith mechanics. Reviews of Geophysics, 45(2). https://doi.org/10.1029/2005rg000184.

Das, B. M. (2019). Advanced soil mechanics. CRC press.

Demas, S. Y., Hall, A. M., Fanning, D. S., Rabenhorst, M. C., Dzantor, E. K. (2004). Acid sulfate soils in dredged materials from tidal Pocomoke Sound in Somerset County, MD, USA. Soil Research, 42(6), 537-545. https://doi.org/10.1071/SR03089.

Devine, J. M., McKay, D. S., & Papike, J. J. (1982). Lunar regolith: Petrology of the< 10 μm fraction. Journal of Geophysical Research: Solid Earth, 87(S01), A260-A268. https://doi.org/10.1029/JB087iS01p0A260.

Edgett, K. S., Yingst, R. A., Ravine, M. A., Caplinger, M. A., Maki, J. N., Ghaemi, F. T., Goetz, W. (2012). Curiosity's Mars hand lens imager (MAHLI) investigation. Space science reviews, 170, 259-317. https://doi.org/10.1007/s11214-012-9910-4.

Eshel, G., Levy, G. J., Mingelgrin, U., & Singer, M. J. (2004). Critical evaluation of the use of laser diffraction for grain-size distribution analysis. Soil Science Society of America Journal, 68(3), 736-743. https://doi.org/10.2136/sssaj2004.7360.

Fedo, C. M., McGlynn, I. O., & McSween Jr, H. Y. (2015). Grain size and hydrodynamic sorting controls on the composition of basaltic sediments: Implications for interpreting Martian soils. Earth and Planetary Science Letters, 423, 67-77. https://doi.org/10.1016/j.epsl.2015.03.052.

Fergason, R. L., Christensen, P. R., Bell III, J. F., Golombek, M. P., Herkenhoff, K. E., & Kieffer, H. H. (2006). Physical properties of the Mars Exploration Rover landing sites as inferred from Mini - TES – derived thermal inertia. Journal of Geophysical Research: Planets, 111(E2). https://doi.org/10.1029/2005je002583.

Folk, R. L., & Ward, W. C. (1957). Brazos River bar [Texas]; a study in the significance of grain size parameters. Journal of sedimentary research, 27(1), 3-26. https://doi.org/10.1306/74d70646-2b21-11d7-8648000102c1865d.

Fredlund, M. D., Fredlund, D. G., & Wilson, G. W. (2000). An equation to represent grain-size distribution. Canadian Geotechnical Journal, 37(4), 817-827. https://doi.org/10.1139/cgj-37-4-817.

Graf, J. C. (1993). Lunar Soils Grain Size Catalog (No. NASA-RP-1265).

Grotzinger, J. P., Gupta, S., Malin, M. C., Rubin, D. M., Schieber, J., Siebach, K., ... & Wilson, S. A. (2015). Deposition, exhumation, and paleoclimate of an ancient lake deposit, Gale crater, Mars. Science, 350(6257), aac7575. https://doi.org/10.1126/science.aac7575.

Hayne, P. O., Bandfield, J. L., Siegler, M. A., Vasavada, A. R., Ghent, R. R., Williams, J. P., ... & Paige, D. A. (2017). Global regolith thermophysical properties of the Moon from the Diviner

Lunar Radiometer Experiment. Journal of Geophysical Research: Planets, 122(12), 2371-2400. https://doi.org/10.1002/2017je005387.

Heiken, G., Vaniman, D., French, B.M., and Schmitt, J., 1991. Lunar Sourcebook: A User's Guide to the Moon. Cambridge, UK.

Heilbronner, R., & Barrett, S. (2013). Image analysis in earth sciences: Microstructures and textures of earth materials (Vol. 129). Springer Science & Business Media.

Herkenhoff, K. E. (2004, December). High-Resolution Textures of Soils and Rocks at Gusev Crater and Meridiani Planum From the Mars Exploration Rover Microscopic Imagers. In AGU Fall Meeting Abstracts (Vol. 2004, pp. P13B-02).

Hwang, S. I., Lee, K. P., Lee, D. S., & Powers, S. E. (2002). Models for estimating soil grain-size distributions. Soil Science Society of America Journal, 66(4), 1143-1150. https://doi.org/10.2136/sssaj2002.1143.

Hynek, B. M., & Di Achille, G. (2017). Geologic map of Meridiani Planum, Mars. US Geological Survey Scientific Investigations Map, 3356. https://doi.org/10.3133/SIM3356.

Isham, V., Cox, D. R., Rodriguez-Iturbe, I., Porporato, A., & Manfreda, S. (2005). Representation of space–time variability of soil moisture. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 461(2064), 4035-4055. https://doi.org/10.1098/rspa.2005.1568.

Iverson, R. M., & Vallance, J. W. (2001). New views of granular mass flows. Geology, 29(2), 115-118. https://doi.org/10.1130/0091-7613(2001)029<0115:nvogmf>2.0.co;2.

Jiang, H., Li, Y., Zou, Q., Zhang, J., Cui, J., Cheng, J., ... & Yao, H. (2024). A GSD-driven approach to deriving stochastic soil strength parameters under hybrid machine learning models. European Journal of Soil Science, 75(6), e70009. https://doi.org/10.1111/ejss.70009.

Kahre, M. A., Murphy, J. R., Newman, C. E., Wilson, R. J., Cantor, B. A., Lemmon, M. T., & Wolff, M. J. (2017). The Mars dust cycle. The atmosphere and climate of Mars, 18, 295.

Kapui, Z., Kereszturi, A., Kiss, K., Szalai, Z., Újvári, G., Hickman-Lewis, K., ... & Westall, F. (2018). Fluvial or aeolian grains? Separation of transport agents on Mars using earth analogue observations. Planetary and Space Science, 163, 56-76.

Karunatillake, S., Wray, J. J., Gasnault, O., McLennan, S. M., Rogers, A. D., Squyres, S. W., ... & Olsen, N. (2014). Sulfates hydrating bulk soil in the Martian low and middle latitudes. Geophysical Research Letters, 41(22), 7987-7996. https://doi.org/10.1002/2014GL061136.

King, E. A. (1977). The lunar regolith: physical characteristics and dynamics. Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, 285(1327), 273-278. https://doi.org/10.1098/rsta.1977.0065.

Konert, M., & Vandenberghe, J. E. F. (1997). Comparison of laser grain size analysis with pipette and sieve analysis: a solution for the underestimation of the clay fraction. Sedimentology, 44(3), 523-535. https://doi.org/10.1046/j.1365-3091.1997.d01-38.x.

Kozakiewicz, J., Maj, D., Mol, S., Sobucki, M., Michaels, T., & Frodyma, N. (2025). Seasonality of aeolian landforms on Meridiani Planum, Mars. Icarus, 425, 116325. https://doi.org/10.1016/j.icarus.2024.116325.

Kozakiewicz, J. Image analysis algorithm for detection and measurement of Martian sand grains. Earth Sci Inform 11, 257–272 (2018). https://doi.org/10.1007/s12145-018-0333-y.

Lai, C. D., Murthy, D. N. P., & Xie, M. (2011). Weibull distributions. Wiley Interdisciplinary Reviews: Computational Statistics, 3(3), 282-287. https://doi.org/10.1002/wics.157.

Li, C., Hu, H., Yang, M. F., Liu, J., Zhou, Q., Ren, X., Ouyang, Z. (2024). Nature of the lunar far-side samples returned by the Chang'E-6 mission. National Science Review, 11(11), nwae328. https://doi.org/10.1093/nsr/nwae328.

Li, C., Hu, H., Yang, M. F., Pei, Z. Y., Zhou, Q., Ren, X., ... & Ouyang, Z. (2022). Characteristics of the lunar samples returned by the Chang'E-5 mission. National science review, 9(2), nwab188. https://doi.org/10.1093/nsr/nwab188.

Liu, T., Michael, G., Zuschneid, W., Wünnemann, K., & Oberst, J. (2021). Lunar megaregolith mixing by impacts: Evaluation of the non-mare component of mare soils. Icarus, 358, 114206. https://doi.org/10.5194/epsc2020-186.

Mangold, N., Gupta, S., Gasnault, O., Dromart, G., Tarnas, J. D., Sholes, S. F., ... & Williford, K. H. (2021). Perseverance rover reveals an ancient delta-lake system and flood deposits at Jezero crater, Mars. Science, 374(6568), 711-717. https://doi.org/10.1126/science.abl4051.

Maurice, S., Wiens, R. C., Bernardi, P., Cais, P., Robinson, S., Nelson, T., ... & Sobron, P. (2021).

The SuperCam instrument suite on the Mars 2020 rover: Science objectives and mast-unit description. Space Science Reviews, 217(3), 47. https://doi.org/10.1007/s11214-021-00807-w.

McKay, D. S., Heiken, G., Basu, A., Blanford, G., Simon, S., Reedy, R., ... & Papike, J. (1991). The lunar regolith. Lunar sourcebook, 567, 285-356.

McLennan, S. M., Bell Iii, J. F., Calvin, W. M., Christensen, P. R., Clark, B. D., De Souza, P. A., ... & Yen, A. (2005). Provenance and diagenesis of the evaporite-bearing Burns formation, Meridiani Planum, Mars. Earth and Planetary Science Letters, 240(1), 95-121. https://doi.org/10.1016/j.epsl.2005.09.041.

Mellon, M. T., Jakosky, B. M., Kieffer, H. H., & Christensen, P. R. (2000). High-resolution thermal inertia mapping from the Mars global surveyor thermal emission spectrometer. Icarus, 148(2), 437-455. https://doi.org/10.1006/icar.2000.6503.

Milliken, R. E., Ewing, R. C., Fischer, W. W., & Hurowitz, J. (2014). Wind-blown sandstones cemented by sulfate and clay minerals in Gale Crater, Mars. Geophysical Research Letters, 41(4), 1149-1154. https://doi.org/10.1002/2013GL059097.

Minitti, M. E., Kah, L. C., Yingst, R. A., Edgett, K. S., Anderson, R. C., Beegle, L. W., ... & Van Beek, T. (2013). MAHLI at the Rocknest sand shadow: Science and science-enabling activities. Journal of Geophysical Research: Planets, 118(11), 2338-2360. https://doi.org/10.1002/2013JE004426.

Mitchell, J. K., Houston, W. N., Carrier III, W. D., & Costes, N. C. (1974). Apollo soil mechanics experiment S-200 (No. NASA-CR-134306).

Ojha, L., Lewis, K., Karunatillake, S., & Schmidt, M. (2018). The Medusae Fossae Formation as the single largest source of dust on Mars. Nature Communications, 9(1), 2867. https://doi.org/10.1038/s41467-018-05291-5.

Ovalle, C., & Dano, C. (2011). A model of confined comminution for granular materials. In 2nd International Symposium on Computational Geomechanics.

Pizzati, M., Mantovani, L., Lisotti, A., Storti, F., & Balsamo, F. (2023). Grain size distributions in Earth Sciences: a review of techniques and a new procedure to match 2D and 3D analyses. EGUsphere, 2023, 1-71. https://doi.org/10.5194/egusphere-2023-2636, 2023.

Qi, S., Li, L., Hou, X., Qiao, S., Ma, X., Lu, X., ... & Wu, F. Y. (2026). Strongly cohesive lunar soil

identified at the Chang'e-6 landing site. Nature Astronomy, 10(2), 214-223. https://doi.org/10.1038/s41550-025-02715-3.

Rivera-Hernández, F., Sumner, D. Y., Mangold, N., Banham, S. G., Edgett, K. S., Fedo, C. M., ... & Wiens, R. C. (2020). Grain size variations in the Murray formation: Stratigraphic evidence for changing depositional environments in Gale crater, Mars. Journal of Geophysical Research: Planets, 125(2), e2019JE006230. https://doi.org/10.1029/2019JE006230.

Rohatgi, A. (2020). Webplotdigitizer: Version 4.5. URL https://automeris. io/WebPlotDigitizer, 411, 860.

Ruff, S. W., Niles, P. B., Alfano, F., & Clarke, A. B. (2014). Evidence for a Noachian-aged ephemeral lake in Gusev crater, Mars. Geology, 42(4), 359-362. https://doi.org/10.1130/g35508.1.

Sacks, L. E., Edgar, L. A., Edwards, C. S., & Anderson, R. B. (2016, December). Grain-scale analyses of Curiosity data at Marias Pass, Gale crater, Mars: Methods comparison and depositional interpretation. In AGU Fall Meeting Abstracts (pp. P23B-2166).

Senel, C. B., Temel, O., Lee, C., Newman, C. E., Mischna, M. A., Muñoz-Esparza, D., ... & Karatekin, Ö. (2021). Interannual, seasonal and regional variations in the Martian convective boundary layer derived from GCM simulations with a semi-interactive dust transport model. Journal of Geophysical Research: Planets, 126(10), e2021JE006965. https://doi.org/10.1029/2021JE006965.

Shangguan, W., Gong, P., Liang, L., Dai, Y., & Zhang, K. (2014). Soil diversity as affected by land use in China: consequences for soil protection. The Scientific World Journal, 2014(1), 913852. https://doi.org/10.1155/2014/913852.

Shi, Y., Zhao, S., Karunatillake, S., Cousin, A., Zhao, J., & Xiao, L. (2024). Sorting and eathering trends of soil at Gale Crater, Mars: Implications for regional pedological processes. Journal of Geophysical Research: Planets, 129(12), e2024JE008598. https://doi.org/10.1029/2024JE008598.

Shi, Y., Zhao, S., Karunatillake, S., & Xiao, L. (2021, April). Semi-automated Image Segmenting Software for Martian Soil Granulometry. In EGU General Assembly Conference Abstracts (pp.

EGU21-7872).

Squyres, S. W., Grotzinger, J. P., Arvidson, R. E., Bell III, J. F., Calvin, W., Christensen, P. R., ... & Soderblom, L. A. (2004). In situ evidence for an ancient aqueous environment at Meridiani Planum, Mars. science, 306(5702), 1709-1714. https://doi.org/10.1126/science.1104559.

Stack, K. M., Williams, N. R., Calef III, F., Sun, V. Z., Williford, K. H., Farley, K. A., ... & Aileen Yingst, R. (2020). Photogeologic map of the perseverance rover field site in Jezero Crater constructed by the Mars 2020 Science Team. Space Science Reviews, 216(8), 127. https://doi.org/10.1007/s11214-020-00739-x.

Szalay, J. R., Pokorný, P., Sternovsky, Z., Kupihar, Z., Poppe, A. R., & Horanyi, M. (2019). Impact ejecta and gardening in the lunar polar regions. Journal of Geophysical Research: Planets, 124(1), 143-154. https://doi.org/10.1029/2018JE005756.

Toneva, P., & Peukert, W. (2007). A general approach for the characterization of fragmentation problems. Advanced Powder Technology, 18(1), 39-51. https://doi.org/10.1163/156855207779768160.

Webb, R., Rosenzweig, C. E., & Levine, E. R. (2000). Global soil texture and derived water-holding capacities (Webb et al.). ORNL Distributed Active Archive Center (DAAC) dataset 10.3334/ORNLDAAC/548 (2000, 548. https://doi.org/10.3334/ORNLDAAC/548.

Weinman, B. A., Yoo, K., Mudd, S. M., Attal, M., Hurst, M. D., Maher, K., ... & Singhvi, A. K. (2011, December). Chemical vs. Physical Contributions to Grainsize Distributions in Hillslope Soils along a Denudation Gradient in the Sierra Nevada, California. In AGU Fall Meeting Abstracts (Vol. 2011, pp. EP53A-02).

Weitz, C. M., Sullivan, R. J., Lapotre, M. G., Rowland, S. K., Grant, J. A., Baker, M., & Yingst, R. A. (2018). Sand grain sizes and shapes in eolian bedforms at Gale Crater, Mars. Geophysical Research Letters, 45(18), 9471-9479. https://doi.org/10.1029/2018GL078972.

Weitz, C. M., Anderson, R. C., Bell III, J. F., Farrand, W. H., Herkenhoff, K. E., Johnson, J. R., ... & Sullivan, R. J. (2006). Soil grain analyses at Meridiani Planum, Mars. Journal of Geophysical Research: Planets, 111(E12). https://doi.org/10.1029/2005je002541.

Wiens, R. C., Maurice, S., Robinson, S. H., Nelson, A. E., Cais, P., Bernardi, P., ... & Willis, P. (2021). The SuperCam instrument suite on the NASA Mars 2020 rover: Body unit and

combined system tests. Space Science Reviews, 217(1), 4. https://doi.org/10.1007/s11214-020-00777-5.

Wilson, L., & Head, J. W. (2007). Explosive volcanic eruptions on Mars: Tephra and accretionary lapilli formation, dispersal and recognition in the geologic record. Journal of volcanology and geothermal research, 163(1-4), 83-97. https://doi.org/10.1016/j.jvolgeores.2007.03.007.

Yong, L., Chengmin, H., Baoliang, W., Xiafei, T., & Jingjing, L. (2017). A unified expression for grain size distribution of soils. Geoderma, 288, 105-119. https://doi.org/10.1016/j.geoderma.2016.11.011.

Yong, L., Xiaojun, Z., Pengcheng, S., Yingde, K., & Jingjing, L. (2013). A scaling distribution for grain composition of debris flow. Geomorphology, 192, 30-42. https://doi.org/10.1016/j.geomorph.2013.03.015.

Zhang, H., Zhang, X., Zhang, G., Dong, K., Deng, X., Gao, X., ... & Yang, M. (2022). Size, morphology, and composition of lunar samples returned by Chang'E-5 mission. Science China Physics, Mechanics & Astronomy, 65, 1-8. https://doi.org/10.1007/s11433-021-1818-1.

Zhang, J., Li, Y., Cui, Y., Wu, Z., Xue, Y., Cheng, J., ... & Luo, A. (2025). Unity of terrestrial and extraterrestrial soils in granular configuration. Earth and Planetary Science Letters, 654, 119239. https://doi.org/10.1016/j.epsl.2025.119239

Zhang, J., Li, Y., Yang, T., Liu, J., Guo, X., Yao, Y., 2023. A universal grain-size distribution of soil with scaling invariance. European Journal of Soil Science, 74(2), e13354. https://doi.org/10.1111/ejss.13354.

Zhang, Jun (2026). Dataset for "PlanetGSD 1.0: a cross-planetary grain-size distribution dataset from Earth, the Moon, and Mars". figshare. Dataset. https://doi.org/10.6084/m9.figshare.32362083.

Zhang, J., Li, Y., Liu, D., Jiang, N., Yang, T., Guo, X., & Yao, Y. (2022). The probability-based granular field of vegetated soils. Earth Surface Processes and Landforms, 47(13), 3100-3116. https://doi.org/10.1002/esp.5446.

Zhao, S., Karunatillake, S., & Shi, Y. (2023). Semi-automated granulometry software for Martian soil analysis (beta) [Computational Notebook]. Zenodo. https://doi.org/10.5281/zenodo.7998487.

Zhao, S., Shi, Y., Karunatillake, S., & Xiao, L. (2021, March). Computational photoanalysis software for Martian soil Granulometry. In 52nd Lunar and Planetary Science Conference (No. 2548, p. 1959).

Zobeck, T. M., Gill, T. E., Popham, T. W. (1999). A two-parameter Weibull function to describe airborne dust grain size distributions. Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group, 24(10), 943-955. https://doi.org/10.1002/(sici)1096-9837(199909)24:10<943::aid-esp30>3.3.co;2-0.

Supplementary Information

Supplementary Table S1 provides detailed site metadata for terrestrial samples organized by geographic region, Table S2 for lunar samples, and Table S3 for Martian samples. Table S4 compiles the unimodal grain size distribution (UGSD) parameters and associated Weibull parameters for soils from all three planetary bodies. Table S5 is site descriptions, acronyms, and dominant surface processes for all sampling locations. The geomorphological contexts of the terrestrial, lunar, and Martian sampling sites are described in Supplementary Texts S1, S2, and S3, respectively, with Text S3 additionally detailing the sub-site distribution and geological setting for Mars. Supplementary Codes S1, S2, and S3 are provided for simulating grain size distribution (GSD) data, calculating Weibull parameters, and generating grain fields, respectively.