Exploratory comparison of Texas soil ion profiles with Phoenix Mars measurements

Lissandra Kruse Fuganti

Abstract

Archived soil measurements were screened for records that could guide further Mars-regolith simulant research. The analysis examines 14,805 laboratory layers from 2,165 Texas pedons in a frozen USDA National Cooperative Soil Survey dataset. After auditing the source data, relative ion profiles are compared numerically with individual Phoenix Wet Chemistry Laboratory samples. Positive calcium, magnesium, sodium and potassium concentrations are converted from charge equivalents to molar units and compared through Aitchison distances between relative four-cation profiles. Chloride and separately determined sulfate provide secondary sensitivity analyses; pH and electrical conductivity remain separate descriptive contrasts. The maximal four-cation cohort contains 4,251 layers, whereas 3,172 belong to the complete-eight-variable panel. Among horizons entirely within 0–5 cm, sample 10N03221 has the smallest four-cation point distance to each Phoenix reference. The leading samples change when other endpoints are included, and reference-error scenarios substantially alter their ranks. The selected records therefore require further investigation before their suitability as simulants can be assessed. Perchlorate and ammonium measurements are absent from the Texas panel, extraction equivalence is unvalidated, and full chemical similarity remains unidentifiable. This partial comparison identifies both candidate records and the measurements needed to test their resemblance under matched laboratory conditions.

Keywords: Mars; Phoenix; terrestrial soils; soluble ions; compositional analysis; regolith simulants; reproducibility

1 Introduction

I used public soil measurements to ask which Texas records merit closer laboratory comparison with Phoenix chemistry. Matching an extract's relative ion profile is a narrower objective than reproducing mineralogy, physical behavior or plant-growth conditions. The analysis focuses on ion profiles; the other properties require separate measurements and experiments.

The Phoenix Wet Chemistry Laboratory (WCL) measured aqueous chemistry at a specific Martian landing site. The early report established the importance of perchlorate; later publications provide detailed ion reductions and a separate sulfate determination. The NASA Planetary Data System preserves the calibrated and derived MECA archive. These observations characterize the Phoenix site and should not be treated as a universal Martian regolith composition. [1–4]

I selected Texas because the USDA archive provides identifiable soil layers, chemistry, depths and method metadata. Measurements can therefore be kept with their original samples instead of being replaced by city averages. The first stage of the study checks coverage, units, missing data and quality flags. The second calculates relative-profile distances to Phoenix.

The resulting candidate list concerns selected ion proportions under the stated assumptions. It does not establish agricultural suitability, authorize collection at the reported locations or validate a simulant. A full matched-method comparison remains outside the available evidence.

2 Materials and methods

Public inputs and reproducible joins

The Texas snapshot was retrieved on 28 October 2026 from the USDA National Cooperative Soil Survey Soil Characterization Database through the public Soil Data Access endpoint. It contains 14,805 unique laboratory sample identifiers representing 2,165 pedons. Each chemical measurement stays with its source row. Missing endpoints are not filled from another horizon, and original identifiers, units, preparation codes and method codes are retained. Field meanings and analytical procedures are interpreted using the USDA documentation. [5–7]

A supplementary taxonomy query is joined by exact pedon key, with one supplement row per pedon. Sampled-as, correlated and laboratory-assigned names are kept separately because their classification contexts differ. None is assumed to be a current soil-series determination. At least one source name is available for 14,166 layers; 639 have no reported name. Coordinate-based locality assignments use frozen 2025 Census place polygons and preserve outside-place and missing-coordinate outcomes. Historical sample-to-municipality assignments have not been independently verified. [8]

The analysis package separates frozen inputs, the Phoenix endpoint ledger, scripts, tests and derived outputs. SHA-256 checks verify input integrity. Execution uses bundled relative paths and requires no network during analysis. The notebook copies the package to a writable directory, allowing execution with read-only Kaggle inputs. Separate scripts and outputs distinguish the source-data audit from the numerical comparison with Phoenix.

Reference endpoints and analytical compatibility

Phoenix reference samples are Rosy Red, Sorceress 1 and Sorceress 2, retained separately rather than averaged. The ion references come from published experimental reductions. Raw telemetry was not reprocessed. WCL combined approximately 1 cm^3 of material with 25 mL of leaching solution containing calibrant ions, not simply pure water. Direct comparison with a terrestrial saturation extract is consequently limited. The chloride endpoints retain experiment-specific qualifications, and sulfate comes from a separate titration rather than a contemporaneous sulfate-selective electrode vector. [2,3]

Missing endpoints and measurement qualifications are recorded in the ledger. Rosy Red has no measured conductivity reference in the selected table; Sorceress 1 lacks the sulfate endpoint; Sorceress 2 has unreliable pH and an unresolved upper magnesium uncertainty. Sorceress 1 ammonium is a nondetection without a numeric censoring limit; Rosy Red and Sorceress 2 estimates are interference-qualified, and the Sorceress 2 uncertainty spans zero. No missing reference value is filled from another Phoenix sample. [2,3]

Texas saturation-extract ions are reported in millimoles of charge per litre, equivalent to meq/L. For molar profiles, calcium and magnesium are divided by two; sulfate is likewise divided by two in the conditional six-ion analysis. Sodium, potassium and chloride retain their numerical values. The conversion resolves the unit difference but leaves extraction effects unresolved. The separate Texas water-extract subset contains only 15 layers from two pedons and retains an unresolved sulfate-unit conflict; it is not used as a harmonized substitute.

Relative profile distance and separate endpoint contrasts

The primary comparison uses positive Ca, Mg, Na and K molar concentrations. For a D-part vector x, its centered log-ratio transformation is

$$\mathrm{clr}(x)_j = \ln(x_j) - (1/D)\ \Sigma_k \ln(x_k)$$

The squared distance to a declared reference r is

$$d(x,r)^2 = \Sigma_j\ [\mathrm{clr}(x)_j - \mathrm{clr}(r)_j]^2$$

Both sums run from 1 to D; reported distances are the nonnegative square roots of the squared expression. The calculation compares relative proportions independently of the extract's overall concentration scale. [9] Zero distance would mean identical ratios in the selected subcomposition, not identical soils. The four-part profile has three independent log-ratio dimensions. No percentage-similarity transformation or acceptance threshold is introduced.

Missing, zero and negative values are excluded from the relevant log-ratio calculation, with exclusion counts retained. No pseudocounts or assumed detection limits are used. The five-ion sensitivity panel adds chloride. The six-ion panel also adds the separately measured sulfate endpoint. This conditional, mixed-endpoint analysis does not reconstruct simultaneous solution chemistry. Sorceress 1 is not scored for six ions.

pH differences and conductivity ratios are calculated separately, only where the relevant reference is available. They remain descriptive contrasts affected by extraction method and are not combined into a weighted similarity score. pH is already logarithmic. Conductivity units can be reconciled, although the extraction conditions still differ.

Cohorts depth and sensitivity

Two cohorts are analyzed: the complete-eight-panel cohort, requiring pH, conductivity and all six saturation ions, and maximal profile availability, requiring only the ions used in each profile. Four strata distinguish all depths; valid-depth records without repeated pedon/depth intervals; horizons wholly contained within 0–5 cm; and horizons merely overlapping 0–5 cm. A 0–16 cm measurement is not relabeled as a measured 0–5 cm composition.

Published marginal reference errors define finite corner scenarios for Rosy Red and Sorceress 1. The four-cation panel has 16 such scenarios per eligible reference. The scenarios do not give confidence intervals or exhaustive bounds over a continuous uncertainty region. They exclude unknown Texas measurement error, error covariance and extraction bias. Sorceress 2 receives point comparisons but no corner scenarios because its upper magnesium uncertainty remains unresolved.

Distances retain full precision; competition ranks use distances rounded to 12 decimal places for deterministic ties. Distinct-pedon summaries use each pedon's closest observed horizon. No synthetic average chemistry is calculated. This choice favors more intensively sampled pedons.

3 Results

Coverage and source data structure

The initial audit identifies 3,882 complete-eight-variable layers from 869 pedons and 147 reported counties. Of these layers, 3,604 have coordinates: 109 fall inside 18 Census places and 3,495 outside mapped places; another 278 lack coordinates. These coverage counts do not establish geographic representativeness.

Across available measurements, saturation-paste pH has a median of 7.6 and a 5th–95th percentile interval of 5.0–8.3. Conductivity has a median of 1.07 dS/m and a corresponding interval of 0.30–12.2 dS/m. These are archived-layer distributions. Ten eight-field method-code signatures occur in the complete panel. The audit flags 123 rows with at least one negative ion value, 175 layers with missing or invalid thickness, and 277 rows sharing a pedon/depth interval. The flags identify records for review; they do not by themselves establish measurement errors or laboratory replication. Fourteen layers in one Ochiltree pedon report a positive longitude inconsistent with Texas; these coordinates are flagged and preserved, not silently corrected. Legacy coordinate accuracy is unquantified and some locations may be generalized or county centroids, so decimal precision does not establish positional accuracy.

Table 1 reports the layers eligible for logarithmic profiles in each cohort and depth stratum.

Table 1. Four-cation analysis cohorts by depth/QC stratum. Values are layers (pedons).

Stratum	Complete-eight panel	Maximal profile availability
All depths, positive four cations	3,172 (793)	4,251 (1,035)
Valid depth, no repeated interval	3,127 (786)	4,177 (1,020)
Entire horizon within 0–5 cm	127 (125)	137 (135)
Horizon overlaps 0–5 cm	584 (544)	741 (697)

Among the 3,882 complete panels, 710 have at least one nonpositive cation and cannot enter the four-cation logarithmic comparison. In the maximal cohort, 9,530 rows lack a required cation and another 1,024 have observed nonpositive values. The corresponding all-depth five-ion cohorts contain 3,167 and 3,689 layers; conditional six-ion cohorts contain 3,152 and 3,613 layers. The complete output contains 124,117 profile/cohort/stratum/reference records. These are repeated analytical comparisons, not 124,117 independent soil samples.

Shallow records selected for investigation

Sample 10N03221, pedon 40712, is the closest four-cation point profile among strictly 0–5 cm-contained horizons for all three references, in both cohorts. Its distances are 0.776 to Rosy Red, 0.828 to Sorceress 1 and 0.866 to Sorceress 2. The source reports Amarillo as both sampled-as and correlated name, Hockley County, and a 0–3 cm interval. The coordinates fall outside Census places; no city identity is inferred.

Table 2. Leading strict-shallow records for three Rosy Red profile definitions in the complete-eight cohort. All are exploratory point-rank leaders, not validated simulants. Distances cannot be compared directly across different-dimensional panels.

Profile	Sample; source soil name	County; depth (cm)	Distance	Scenario rank range
Four cations	10N03221; Amarillo (sampled-as and correlated)	Hockley; 0–3	0.776	1–8
Five ions, adding chloride	10N03217; Amarillo (sampled-as and correlated)	Hockley; 0–3	1.424	1–15
Six ions, conditional mixed endpoint	07N00501; Shippingpens (sampled-as); Pokorny-hypergypsic saline (correlated)	Culberson; 0–5	1.833	1–15

Reported coordinates are 33.504669, −102.321106 for 10N03221; 33.504627, −102.321007 for 10N03217; and 31.904612, −104.362221 for 07N00501. These coordinates are not independently verified collection locations. All three records carry method code 4F2 across the eight-variable panel. The names retain their source classification contexts; their current series status has not been independently confirmed.

For 10N03221, original Ca, Mg, Na and K values are 5.7, 10.4, 3.5 and 0.9 meq/L, corresponding to 2.85, 5.2, 3.5 and 0.9 mmol ions/L. Chloride is 6.6 meq/L and sulfate 10.2 meq/L; pH is 7.8 and conductivity 1.88 dS/m. The pH contrast is +0.1 relative to Rosy Red and +0.2 relative to Sorceress 1. Conductivity ratios are approximately 1.37 relative to Sorceress 1 and 0.99 relative to Sorceress 2. These descriptive contrasts remain subject to the extraction-method differences.

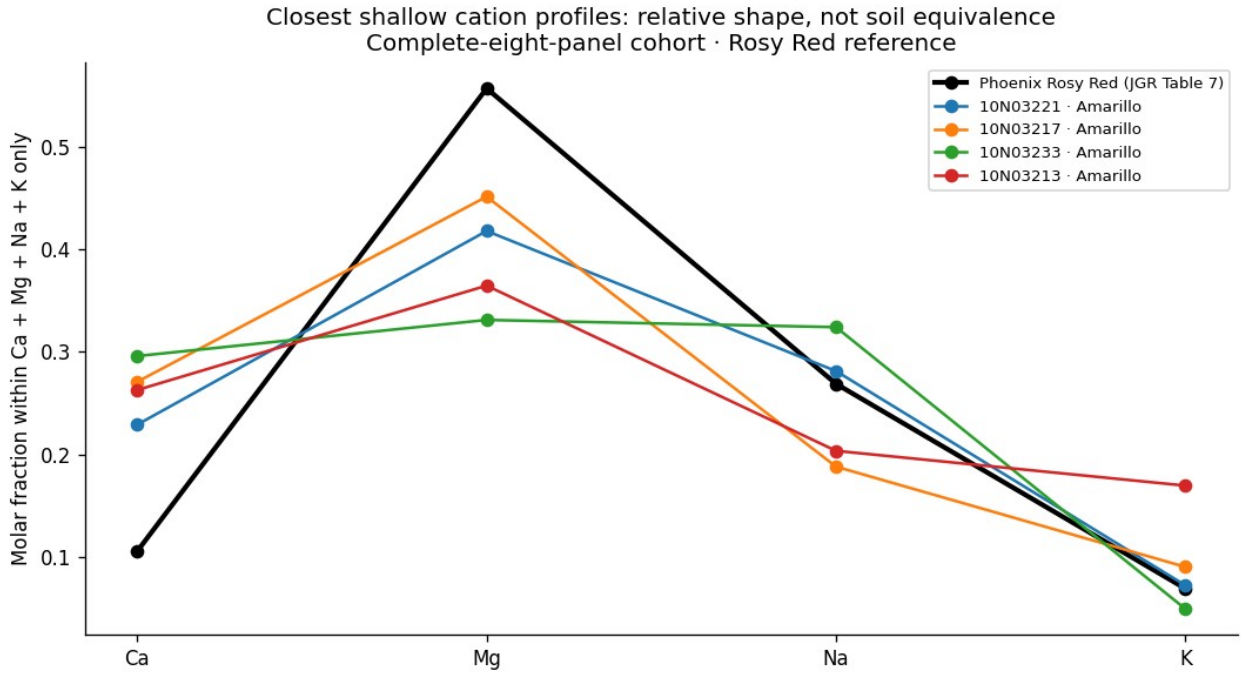


Figure 1. Relative molar fractions within Ca, Mg, Na and K for the four closest strict-shallow records and Rosy Red. Fractions describe only the selected subcomposition; they do not demonstrate absolute concentration or soil equivalence.

Endpoint reference and depth sensitivity

The closest point-reference profile is sensitive to reference uncertainty. For 10N03221, Rosy Red corner-scenario ranks range from 1 to 8. In the complete-eight-panel cohort, Sorceress 1 ranks range from 1 to 7; with maximal profile availability, the range is 1 to 8. Other shallow records become closest under some of the same scenarios. Several candidates should therefore be retained for investigation; the analysis does not establish a unique optimum.

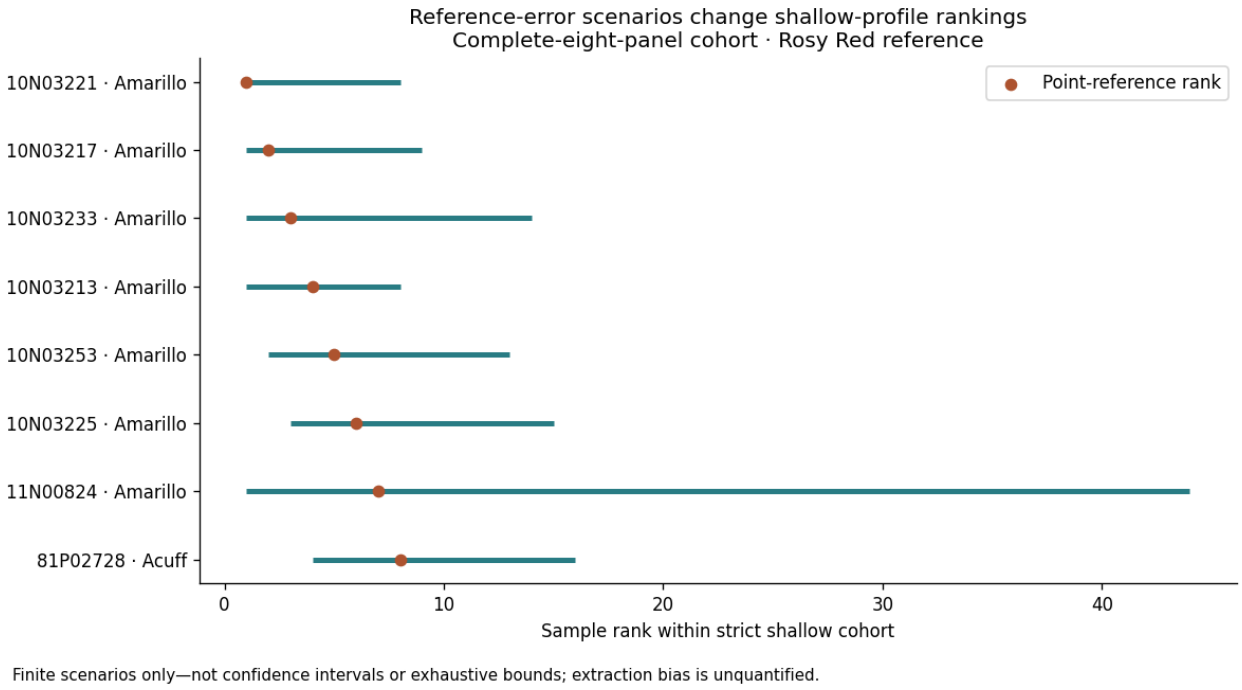


Figure 2. Point-reference ranks and ranks attained under finite marginal-error corner scenarios for selected strict-shallow records. Horizontal ranges are not confidence intervals or exhaustive uncertainty bounds; extraction bias remains unquantified.

Adding chloride changes the closest strict-shallow record to 10N03217, with distances of 1.424, 1.874 and 1.674 to Rosy Red, Sorceress 1 and Sorceress 2, respectively. The conditional six-ion panel instead selects 07N00501, a 0–5 cm Culberson County record, with distances of 1.833 to Rosy Red and 1.920 to Sorceress 2. Distances from panels with different dimensions do not share an acceptance scale. The relevant result is that including additional endpoints changes which record is selected.

The depth restriction also changes the selected records. Across all depths, 95P05500 is closest to Rosy Red and Sorceress 1 for four cations, at distances 0.469 and 0.127. However, it represents 282–323 cm in Lynn County. Sorceress 2 instead selects 95P05497 at 189–239 cm in the same pedon. Both records describe deep horizons and cannot represent measured shallow chemistry. In the overlapping shallow stratum, five-ion and six-ion leaders include 0–12 cm and 0–16 cm horizons, respectively, emphasizing why overlap must remain separate from strict containment.

4 Discussion

The analysis identifies the closest observed profiles within each specified cohort, but their interpretation depends on the ions included. Even identical four-cation ratios would leave perchlorate, ammonium, omitted anions, absolute concentration and extraction response unresolved. The Texas schema contains no measured perchlorate or ammonium-ion concentrations. References to ammonium in extraction-reagent names describe the reagent, not measured ammonium. These missing measurements prevent a full comparison and provide no evidence of either zero concentrations or chemical dissimilarity.

Selective extraction is the main methodological limitation. Ratios remove a common multiplicative concentration factor while retaining the effects of differential dissolution, adsorption, precipitation and complexation. A soil with a similar saturation-extract profile may yield a different profile under a WCL-like protocol. The sulfate analysis adds a further limitation because it combines distinct experimental endpoints. It remains a secondary, conditional calculation. None of these distances establishes mineralogical or environmental equivalence to Mars.

The archive also has an uneven sampling structure. It is not a probability sample of Texas, multiple horizons share pedons, and several candidate pedons have nearby reported coordinates. Amarillo-named records

are common among the closest shallow profiles, which warrants targeted examination without implying that the entire soil taxon is suitable. Selecting each pedon's closest observed horizon favors pedons with more measurements. Future sampling should account for this bias.

5 Next experimental steps

The candidates need paired laboratory measurements, using archived samples where available or comparable resampled material where access is permitted. A preregistered protocol should preserve exact horizon identity and include both the original terrestrial method and a documented WCL-like extraction. Soil mass or volume, liquid ratio, extraction time, temperature, headspace composition, preparation and filtration must be recorded. Paired extracts should quantify the requested ions, pH and conductivity, with blanks, analytical replicates, calibration checks and detection limits. Free-ion and total-dissolved endpoints must be distinguished.

Mineralogy, texture, organic matter and water retention must also be measured against the needs of the intended simulant application. Acceptance tolerances should be set for that purpose and tested against independently characterized controls before judging the closest database records.

Controlled plant-growth work, such as the simulant experiments of Wamelink and colleagues, could follow as a separate study with its own controls and safety assessments. [10] No growth, germination, remediation or food-safety experiment was performed here.

6 Conclusion

The Texas archive supports a reproducible numerical comparison of selected ion profiles with Phoenix measurements. It identifies shallow records for laboratory investigation, with different candidates emerging as the endpoints and reference assumptions change. The comparison remains partial: missing analytes and unvalidated extraction equivalence prevent a full Texas–Phoenix similarity assessment or validation of a simulant. Paired measurements are needed to determine whether the observed profile resemblance persists under matched methods.

Data code and computational transparency

The executable analysis is publicly available in the Texas soils and Phoenix exploratory ion profiles notebook, using version 4 of the public Texas soil chemistry and Census localities dataset. Publication and execution were verified on 29 October 2026. The notebook preserves explicit exploratory labels rather than presenting the outputs as a full-similarity score.

The public GitHub repository contains the reproducibility package. The version used for this report is pinned to commit dadbb875447536400a8852b561f1621dbcb24fc3.

The earlier Brazilian analysis is available separately as public analysis code only. Its full data-backed run remains private; that code-only release does not redistribute the mixed-license source collection. It should not be read as a validated full chemical comparison.

Kaggle version 3 (script version 363326864) completed all six code cells successfully and freshly generated the 35 analysis outputs. The Output tab contains the profile tables and figures under texas_numerical_study/outputs and the descriptive source-data reports under texas_numerical_study/eda/outputs. Key reports include the full exploratory distance table, profile run summary, Phoenix endpoint ledger, shallow-profile records with original chemistry and location, quality flags, distributions and method-stratified summaries. Archive compatibility was corrected to accept Kaggle-expanded source files with verified hashes; this adaptation did not change the scientific calculations.

The frozen Texas package includes public-source inputs, provenance and checksums, the numerical reference ledger, executable scripts and 35 outputs: 25 from the source-data analysis and 10 from the profile study. Independent verification recomputed all 124,117 comparisons across 64 computed groups, with a maximum absolute distance difference of 1.78×10^{-15}. All six notebook cells, nine source-data tests and the profile tests passed in both local input modes: original gzip files and expanded CSV files. Each mode freshly generated 35 outputs; 34 numerical and figure outputs were byte-identical across modes, while the source-hash inventory appropriately recorded different file representations. These local checks are distinct from platform execution. USDA and Census attribution is retained; newly written code is offered under Apache 2.0. No agency endorsement is implied.

AI-assisted computational preparation supported code organization, data auditing, reproducibility checks and manuscript preparation. The reported work is an analysis of public records and published reference endpoints; it does not describe newly performed laboratory or field experiments. Source citations, preserved inputs and explicit computational outputs provide the basis for independent checking.

References

1. NASA Planetary Data System. *Phoenix MECA Non-Imaging Calibrated and Derived Data Bundle*. doi:10.17189/1518770.

2. Kounaves, S. P., et al. (2010). Wet Chemistry experiments on the 2007 Phoenix Mars Scout Lander mission: Data analysis and results. *Journal of Geophysical Research: Planets*, 115, E00E10. doi:10.1029/2009JE003424.

3. Kounaves, S. P., et al. (2010). Soluble sulfate in the martian soil at the Phoenix landing site. *Geophysical Research Letters*, 37, L09201. doi:10.1029/2010GL042613.

4. Hecht, M. H., et al. (2009). Detection of perchlorate and the soluble chemistry of martian soil at the Phoenix lander site. *Science*, 325, 64–67. doi:10.1126/science.1172466.

5. USDA National Cooperative Soil Survey / Kellogg Soil Survey Laboratory. *NCSS Soil Characterization Database*. Snapshot retrieved 28 October 2026. Database query interface; public-domain distribution metadata.

6. USDA NRCS (2025). *NCSS Soil Characterization Database Data Content Standard*, December 2025. Official content standard.

7. USDA NRCS. *Kellogg Soil Survey Laboratory Methods Manual*, Soil Survey Investigations Report 42, version 6, part 1. Official methods manual.

8. U.S. Census Bureau (2025). *TIGER/Line Shapefiles: Texas Places*. Release documentation.

9. Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society, Series B*, 44, 139–160. doi:10.1111/j.2517-6161.1982.tb01195.x.

10. Wamelink, G. W. W., Frissel, J. Y., Krijnen, W. H. J., Verwoert, M. R., and Goedhart, P. W. (2014). Can plants grow on Mars and the Moon: A growth experiment on Mars and Moon soil simulants. *PLOS ONE*, 9, e103138. doi:10.1371/journal.pone.0103138.