From spectra to mineralogy: a remote sensing approach to Earth's desert dust source regions

Thesis by

Abigail May Keebler

In Partial Fulfillment of the Requirements for

the degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2026

Defended 21. May 2026

Abigail May Keebler
ORCID: 0000-0001-6358-9834

ACKNOWLEDGEMENTS

My time at Caltech has been shaped by many wonderful mentors, collaborations, and learning opportunities that together have made this thesis and my graduate experience possible. First, I would like to thank my advisors. Bethany, thank you for teaching me how to ask the right questions and to pursue them rigorously. I am deeply grateful for your patience and guidance throughout this journey, and for always supporting my periodic field excursions. I am also grateful to my committee members—Dr. David Thompson, Dr. Katherine de Kleer, and Dr. Christian Frankenberg—for your feedback, support, and many thoughtful conversations over the years. I would also like to acknowledge Dr. George Rossman, whose playful spirit and passion for minerals has been deeply inspiring. He is dearly missed.

Beyond my advisors at Caltech, I have had the privilege of working with many outstanding scientists in the remote sensing community. The EMIT science team has been an especially rewarding collaboration. I am continually inspired by the passion and expertise of this group. I would especially like to thank Dr. Phil Brodrick for his kindness and patience in onboarding me to the EMIT team, and for helping me learn to work with the high-performance computing resources required for large-scale hyperspectral data analysis.

I am also extremely grateful to the mentors who supported me earlier in my academic journey: Dr. LeAnn Srogi, Dr. Martin Helmke, Dr. Marc Gagne, Dr. Mark Everett, Dr. Peyton Smith, and Dr. Tom Watters. In particular, I would like to thank Dr. Ed Simpson, who first introduced me to the desert. Fieldwork, beginning with the trips in southeastern Utah he invited me to join, had a lasting impact on me and shaped my trajectory in planetary science. I am incredibly grateful for his continued mentorship and support.

Fieldwork has been one of the most rewarding aspects of my training—not only for the opportunity to engage directly with geologic materials, but also for the friendships and memories formed in the process. To my field partners Rebecca, Eva, Jess, and Josh: thank you for the laughter, good spirits, and shared adventures.

Beyond research, Caltech has provided many opportunities to grow as a geologist. I am grateful for the field program here, including Ge 136 and associated enrichment trips, which have provided invaluable opportunities to learn field geology firsthand. As the saying goes, “the best geologist is the

one who has seen the most rocks." These experiences have deepened my understanding of the geology of the Southwestern United States and California, and have also provided a wonderful setting for building friendships. I am especially grateful to Dr. Joe Kirschvink for leading many of these trips.

The Ehlmann lab has been an incredibly intellectually stimulating and supportive environment. I am grateful to Eva, Oak, Samantha, Yu Yu, Jess, Mma, Valeria, Will, Erin, Jordan, Elin, Rebecca, Oscar, Amanda, Josh, and others for thoughtful feedback, discussions, and support throughout this work. I appreciate your willingness to occasionally follow me from Moon- and Mars-focused discussions back to Earth. Samantha, it has been a privilege to go through this program alongside you. Thank you for the many hours spent sharing ideas, troubleshooting, studying, and, of course, ice cream!

My time at Caltech has also been shaped by an extraordinary community of friends. I am grateful to Tina, Richard, Willow, Korbi, Amanda, Jocelyn, Ruth, James, Noel, Becky, Ryleigh, Ren, Simon, Elizabeth, Maria, Jarek, Josh, Alessandra, Jess, Noam, Eran, Holly, and many others whose companionship made this time one of the best of my life. I am so grateful for this incredible community and for each of you.

I am deeply grateful to my partner, John, whose patience and care have been a constant support as I navigated the stresses of graduate school. Thank you for keeping me grounded and for all the cups of hot tea in the evenings.

Finally, I would like to thank my family: Mom, Dad, Dawn, Tim, Robbie, and Evan. Thank you for your unconditional love and support. I am deeply grateful not only for your encouragement throughout graduate school, but also for instilling in me a lasting curiosity about the natural world, which ultimately led me to this path.

ABSTRACT

Mineral dust plays a critical role in Earth's climate and biogeochemical systems, influencing radiative forcing, cloud microphysics, and nutrient fertilization of terrestrial and marine ecosystems. These impacts are strongly dependent on dust mineralogy, particularly the abundance and speciation of iron-bearing phases, clays, and carbonates, which control both the optical properties and chemical reactivity of dust aerosols. Despite this importance, mineralogical properties of dust source regions remain poorly constrained at regional to global scales, limiting the representation of dust processes in Earth system models. Hyperspectral visible to shortwave infrared (VSWIR) remote sensing offers a promising pathway for addressing this gap by enabling spectrally resolved characterization of surface mineralogy globally. However, translating reflectance spectra into quantitative mineral abundances remains challenging due to the nonlinear nature of VSWIR spectra of intimate mineral mixtures.

This dissertation combines hyperspectral remote sensing, field spectroscopy, and laboratory analyses to improve quantitative interpretation of mineral dust source regions and to evaluate the extent to which mineralogical properties can be retrieved from VSWIR observations. The work is structured around three complementary studies that collectively link global satellite observations, field measurements, and empirical modeling approaches.

First, a global analysis of arid dust source regions is conducted using hyperspectral observations from the Earth Surface Mineral Dust Source Investigation (EMIT) mission. A systematic sampling and filtering framework is developed to extract a representative dataset of bare soil reflectance spectra from more than one billion observations. The resulting dataset characterizes global variability in surface albedo and mineralogical absorption features across major dust source regions, revealing distinct regional spectral endmembers associated with differences in mineralogy, including bright iron oxide- and kaolinite-rich Saharan surfaces and darker clay- and carbonate-dominated Asian surfaces. These results demonstrate substantial compositional diversity in dust source regions and provide constraints on surface radiative properties relevant to Earth system modeling.

Second, we investigate the physical and compositional controls on spectral variability using a novel dataset of co-located in situ VNIR reflectance spectra and laboratory measurements of mineralogy, grain size, and iron speciation. We compare the spectral variability captured in this dataset to that of the EMIT global dust source dataset to evaluate the range of global variability represented. We assess compositional controls on spectral variation, including absorption feature presence, position, and strength, as well as overall reflectance and continuum shape. Clay and carbonate absorption features show systematic but non-linear relationships with mineral presence and abundance, reflecting overlapping absorptions and mixed-phase effects. In contrast, iron oxide abundance exhibits strong, approximately linear relationships with diagnostic absorption features in fine-grained clay-rich sediments. We also identify a distinct population of hematite-bearing sands that display strong absorption features despite low hematite abundance. Radiative transfer modeling shows that grain size and bright mineral matrices significantly modulate iron

oxide spectral expression. Overall, these results highlight that composition and sediment physical context in control spectral variability.

Third, the dissertation evaluates empirical approaches for predicting quantitative mineral abundances from VSWIR spectra. Using the coupled spectral-mineralogical dataset, a partial least squares regression model is trained to estimate mineral and iron species abundances. Results show that mineral components which control continuum shape and albedo over the full spectral range, including quartz, feldspar, and iron oxides, can be predicted accurately, while phases that express diagnostic feature in a narrow spectral range, such as clays and carbonates, are less-well predicted. Application of the calibrated models to EMIT reflectance data demonstrates that PLSR models successfully identify the major minerals present in the ground-truth data. Compared with EMIT mineralogy products, the empirical approach provides several advantages. In particular, the models improve quantitative retrievals of carbonate abundance, and more frequently resolve complex multi-mineral assemblages containing combinations of clays, carbonates, and evaporite minerals within individual spectra. The models produce reasonable values when applied to spectra from sediment types outside the training dataset, suggesting promising transferability across heterogeneous dust source environments. Together, these findings demonstrate both the potential and current limitations of empirical inversion approaches for hyperspectral mineral retrieval and highlight their utility as a complement to existing feature-based remote sensing frameworks.

Together, these results provide a framework for improving quantitative interpretation of hyperspectral observations of Earth's bare sediment surface. By linking global-scale satellite data with field-based measurements and empirical modeling, this work advances the ability to retrieve physically meaningful mineralogical information from VSWIR remote sensing. These improvements are essential for better constraining the radiative and biogeochemical impacts of mineral dust in Earth system models and for extending hyperspectral approaches to future Earth and planetary remote sensing missions.

PUBLISHED CONTENT AND CONTRIBUTIONS

Keebler, A.M., Ehlmann, B.L., Thompson, D.R., Ginoux, P., Brodrick, P.G., and Green, R.O. (in review). Spectral Variability of Earth's Desert Dust Source Regions Sampled from EMIT Reflectance Data. JGR: Earth Surface, EMIT special issue.

A.M.K participated in the conception of the project, designed the sampling scheme, developed and tested the data cleaning pipeline, prepared and analyzed the data, and participated in the writing of the manuscript.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

1.1 Observing a Complex Planet: Remote Sensing and Earth System Science

Remote sensing has been humorously described as "astronomy looking the wrong way," acknowledging the methodological overlaps and Herculean effort required to establish space-based observations, whether one looks out into the universe or back towards home. Earth is unique among the terrestrial planets in our solar system, hosting extraordinarily dynamic surface processes that operate across interconnected geologic, atmospheric, hydrologic, and ecologic processes. As the only planet currently known to support life, understanding these systems is particularly important for monitoring Earth's changing climate and its impact on humans and ecosystems.

Over the past several decades, remote sensing has transformed our ability to observe Earth on global scales. Beginning with multispectral satellite missions such as the Landsat program, remote sensing provided the means to systematically monitor Earth surface processes through time. Early Landsat studies demonstrated the utility of satellite reflectance spectra observations for geological mapping, vegetation classification, agricultural monitoring, and land cover change analysis, establishing the foundation for modern Earth observation science. Foundational remote sensing studies showed the capability of multispectral observations to discriminate geologic units and alteration minerals (e.g., Rowan et al., 1974; Goetz and Rowan, 1981), monitor vegetation and ecosystem change (Tucker, 1979) and quantify large-scale land use and land cover dynamics (Townshend et al., 1991). These observations have enabled decades-long studies of land cover change, vegetation dynamics, hydrology, and surface geology across spatial scales that would be impossible to achieve through field observations alone. Advances in detector technology, computing power, and

satellite instrumentation have since dramatically expanded both the spatial and spectral resolution available for Earth system science.

More recently, the development of imaging spectroscopy, or hyperspectral remote sensing, has enabled the collection of near-continuous spectral measurements across hundreds of narrow wavelength bands. Unlike traditional multispectral systems, hyperspectral instruments can resolve narrow absorption features associated with specific minerals, vegetation pigments, moisture content, and other surface properties. Early airborne imaging spectroscopy instruments such as Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) demonstrated the potential of hyperspectral data for quantitative surface characterization and mineral identification (Green et al., 1998), while later orbital missions such as Earth Observing-1 Hyperion expanded these capabilities to global observations (Pearlman et al., 2003). More recently, the Earth surface Mineral dust source InvesTigation (EMIT) imaging spectrometer has provided high-quality global visible to shortwave infrared (VSWIR; 400-2500 nm) observations of Earth's arid regions from the International Space Station, with the specific goal of improving constraints on the composition and radiative impacts of mineral dust aerosols (Green et al., 2020; Thompson et al., 2024). Together, these instruments and missions have produced an explosion in the volume and complexity of spectral data available for Earth observation, allowing observation of Earth surface properties and processes in unprecedented spatial coverage and temporal and spectral resolutions.

Here, we use remote sensing to investigate one of Earth's most extensive and climatically sensitive biomes: the drylands. Drylands cover approximately 40% of the global land surface (Wang et al., 2012) and play a major role in global biogeochemical cycling and the planetary radiation budget. In particular, arid regions are the primary global source of mineral dust aerosols, whose climate impacts are strongly dependent on their mineralogical composition. For example, mineral dust exerts both direct and indirect radiative effects in the atmosphere, with iron oxide–

bearing phases generally enhancing absorption while many silicate and clay minerals contribute more strongly to scattering and reflectance (Shao et al., 2011; Mahowald et al., 2014). Even relatively small variations in iron oxide abundance can have a disproportionate impact on dust optical properties, as iron-bearing minerals strongly control absorption in the visible and near-infrared, meaning that changes of only a few weight percent can substantially alter the radiative forcing efficiency of dust aerosols (Scanza et al., 2015). In addition, dust deposition can provide a key source of bioavailable iron to ocean and terrestrial ecosystems (e.g., Moskowitz et al., 2016; Rizzolo et al., 2017), and the mineralogy and shape of suspended mineral grains can influence cloud microphysical processes such as nucleation (Sullivan et al., 2009), further linking dust composition to Earth system function. Accurately constraining the mineralogical composition and spatial distribution of dust source regions is therefore critical for improving representations of mineral dust in Earth system and climate models.

However, retrieving quantitative mineral abundances and grain size information from hyperspectral observations of intimate mineral mixtures remains challenging. Many existing approaches yield uncertainties on the order of tens of weight percent (Lapotre et al., 2017), reflecting the complexity of natural sedimentary systems. In particular, spectral responses are governed not only by mineral abundance but also by grain size, mixing geometry, and nonlinear interactions between absorbing and non-absorbing phases. For instance, increases in the grain size of spectrally neutral minerals such as quartz can enhance absorption band depths of co-occurring phases by increasing path length effects and multiple scattering, complicating straightforward interpretation of band strength as a proxy for abundance (Hapke, 2012; Mustard & Pieters, 1987). Improving these spectral interpretations is therefore essential, given the role of mineral dust in Earth's biogeochemical and radiative systems and their influence on climate processes across regional to global scales.

This dissertation presents an assessment of the mineralogical spectral diversity present within Earth's arid dust source regions with connections to underlying compositional drivers of variability and advances methods for interpreting VSWIR data to retrieve quantitative mineralogical information from remotely sensed observations.

1.2 Overview of Dissertation Chapters

1.2.1 Chapter 2: Earth's deserts from orbit

Chapter 2 presents a global hyperspectral characterization of Earth's arid dust source regions using observations from the EMIT imaging spectrometer. EMIT has collected over one billion visible to shortwave infrared (VSWIR) spectra from onboard the International Space Station, providing an extensive observational resource for studying Earth's arid surfaces. To make this dataset analytically tractable while preserving global representativeness, a systematic sampling framework is developed to select spectra that capture the full diversity of major dust source regions. A multi-stage quality screening pipeline is also implemented to remove contaminated or non-representative observations, including those affected by clouds, snow, surface water, and vegetation cover. The resulting dataset enables an assessment of global variability in desert surface spectral properties, including broadband albedo, which has implications for both surface radiative forcing and the optical properties of entrained mineral dust aerosols. In addition, diagnostic mineral absorption features are used to characterize broad compositional differences across regions, providing a spectral parameter space for major dust source environments. Together, these analyses establish a baseline framework for interpreting EMIT observations of Earth's arid landscapes at planetary scale.

1.2.2 Chapter 3: Connecting sediment spectra to quantitative mineralogy data

Chapter 3 investigates the physical and compositional drivers of spectral variability in dust source region surface sediments using a novel dataset of co-located in situ

reflectance spectra and paired laboratory measurements of bulk mineralogy, grain size, and iron speciation. By integrating field spectroscopy with detailed mineralogical and geochemical characterization, this dataset enables a direct assessment of how variations in mineral abundance and iron speciation are expressed in VSWIR reflectance spectra. We first quantify the spectral variability present within the in situ dataset and then link these variations to their underlying mineralogical controls. Across the dataset, we evaluate how diagnostic absorption features and continuum properties vary as a function of mineral abundance. Clay and carbonate minerals exhibit systematic but non-linear relationships between abundance and SWIR feature presence and strength. Evaporite minerals, particularly gypsum, show more direct relationships with their diagnostic SWIR absorption features, consistent with their relatively isolated spectral signatures. We identify systematic, approximately linear relationships between iron oxide abundance and diagnostic VNIR spectral absorption features for fine-grained sediments and show how quartz abundance increases iron oxide absorption strength for larger, sand-sized sediments even when iron abundance remains low. In doing so, this work moves beyond qualitative spectral interpretation toward a more quantitative framework for relating specific compositional and physical properties to observed reflectance behavior. This approach leverages the unique advantage of Earth observation studies: the ability to directly ground orbital hyperspectral measurements in field-based spectral observations and laboratory analyses, thereby enabling physically interpretable links between remotely-sensed spectra and surface composition.

1.2.3 Chapter 4: Empirical modeling of mineral abundances from hyperspectral data.

Chapter 4 leverages the coupled spectral-mineralogical dataset developed in Chapter 3 to evaluate the extent to which quantitative mineral abundances can be predicted directly from hyperspectral reflectance data. Using a partial least squares empirical modeling approach, this chapter predicts measured mineralogical composition from VSWIR spectra with a particular focus on major mineral phases relevant to dust

generation and radiative forcing. The preliminary analysis assesses both the strengths and limitations of statistical and machine learning-based approaches for inferring mineral abundances from remotely sensed spectra. In doing so, it provides insight into which mineral phases are spectrally predictable under natural conditions, and which remain challenging due to spectral mixing, compositional covariance, or weak or narrow diagnostic absorption features. It shows that empirical methods may offer a path to better surface mineralogy characterizations than the existing state-of-the-art for higher level EMIT data products of mineral abundance. This chapter establishes a foundation for future development of improved quantitative retrieval methods and highlights the potential for hyperspectral data to support physically meaningful estimates of surface mineralogy in Earth system applications.

References

Goetz, A. F., & Rowan, L. C. (1981). Geologic remote sensing. Science, 211(4484), 781-791.

Green, R. O., Eastwood, M. L., Sarture, C. M., Chrien, T. G., Aronsson, M., Chippendale, B. J., ... & Williams, O. (1998). Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). Remote sensing of environment, 65(3), 227-248.

Green, R. O., Mahowald, N., Ung, C., Thompson, D. R., Bator, L., Bennet, M., ... & Zan, J. (2020, March). The Earth surface mineral dust source investigation: An Earth science imaging spectroscopy mission. In 2020 IEEE aerospace conference (pp. 1-15). IEEE.

Hapke, B. (2012). Theory of reflectance and emittance spectroscopy. Cambridge university press.

Lapotre, M. G., Ehlmann, B. L., & Minson, S. E. (2017). A probabilistic approach to remote compositional analysis of planetary surfaces. Journal of Geophysical Research: Planets, 122(5), 983-1009.

Mahowald, N., Albani, S., Kok, J. F., Engelstaeder, S., Scanza, R., Ward, D. S., & Flanner, M. G. (2014). The size distribution of desert dust aerosols and its impact on the Earth system. Aeolian Research, 15, 53-71.

Moskowitz, B. M., Reynolds, R. L., Goldstein, H. L., Berquó, T. S., Kokaly, R. F., & Bristow, C. S. (2016). Iron oxide minerals in dust-source sediments from the Bodélé Depression, Chad: Implications for radiative properties and Fe bioavailability of dust plumes from the Sahara. *Aeolian Research*, *22*, 93-106.

Mustard, J. F., & Pieters, C. M. (1987). Quantitative abundance estimates from bidirectional reflectance measurements. Journal of Geophysical Research: Solid Earth, 92(B4), E617-E626.

Pearlman, J. S., Barry, P. S., Segal, C. C., Shepanski, J., Beiso, D., & Carman, S. L. (2003). Hyperion, a space-based imaging spectrometer. IEEE Transactions on Geoscience and Remote Sensing, 41(6), 1160-1173.

Rizzolo, J. A., Barbosa, C. G., Borillo, G. C., Godoi, A. F., Souza, R. A., Andreoli, R. V., ... & Godoi, R. H. (2017). Soluble iron nutrients in Saharan dust over the central Amazon rainforest. *Atmospheric Chemistry and Physics*, *17*(4), 2673-2687.

Rowan, L. C., Wetlaufer, P. H., Goetz, A. F., Billingsley, F. C., & Stewart, J. H. (1974). Discrimination of rock types and detection of hydrothermally altered areas in south-central Nevada by the use of computer-enhanced ERTS images (No. 883). US Goverment Printing Office.

Scanza, R. A., Mahowald, N., Ghan, S., Zender, C. S., Kok, J. F., Liu, X., ... & Albani, S. (2015). Modeling dust as component minerals in the Community Atmosphere Model: development of framework and impact on radiative forcing. Atmospheric Chemistry and Physics, 15(1), 537-561.

Shao, Y., Wyrwoll, K. H., Chappell, A., Huang, J., Lin, Z., McTainsh, G. H., ... & Yoon, S. (2011). Dust cycle: An emerging core theme in Earth system science. Aeolian Research, 2(4), 181-204.

Sullivan, R. C., Moore, M. J. K., Petters, M. D., Kreidenweis, S. M., Roberts, G. C., & Prather, K. A. (2009). Effect of chemical mixing state on the hygroscopicity and cloud nucleation properties of calcium mineral dust particles. *Atmospheric Chemistry and Physics*, *9*(10), 3303-3316.

Thompson, D. R., Green, R. O., Bradley, C., Brodrick, P. G., Mahowald, N., Dor, E. B., ... & Zandbergen, S. (2024). On-orbit calibration and performance of the EMIT imaging spectrometer. Remote Sensing of Environment, 303, 113986.

Townshend, J. R. G., Justice, C. O., Li, W., Gurney, C., & McManus, J. (1991). Global vegetation classification by reduced simple ratio analysis of NOAA-AVHRR data. International Journal of Remote Sensing, 12, 147–164.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, *8*(2), 127-150.

Wang, L., d'Odorico, P., Evans, J. P., Eldridge, D. J., McCabe, M. F., Caylor, K. K., & King, E. G. (2012). Dryland ecohydrology and climate change: critical issues and technical advances. Hydrology and Earth System Sciences, 16(8), 2585-2603.

Chapter 2

SPECTRAL VARIABILITY OF EARTH'S DESERT DUST SOURCE REGIONS SAMPLED FROM EMIT REFLECTANCE DATA

Abigail M. Keebler[1], Bethany L. Ehlmann[1,2,3,4], David R. Thompson[5], Paul Ginoux[6], Philip Brodrick[5], Robert O. Green[5]

1 California Institute of Technology, Pasadena, CA, United States

2 Laboratory for Atmospheric and Space Physics, Boulder, CO, United States

3 Department of Geological Sciences, University of Colorado, Boulder

4 Department of Astrophysical & Planetary Sciences, University of Colorado, Boulder

5 Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States

6 NOAA/OAR Geophysical Fluid Dynamics Laboratory, Princeton, NJ, United States

Key Points:

- Global dust source regions exhibit diverse mineral signatures and bare soil albedo variation 0.10-0.65% VIS, 0.20-0.85% NIR in EMIT data
- Two dominant bare soil spectral classes are brighter kaolinite-iron oxide (Sahara) and darker illite/smectite and carbonate (central Asia)
- Bare soil albedo from EMIT is 6–10% brighter than unfiltered MODIS albedo and ~40% darker than lab-measured suspended dust-source sediments

Abstract

Mineral dust plays a critical role in Earth's climate and biogeochemical cycles, influencing radiative forcing, cloud formation, and nutrient fertilization. These effects depend strongly on dust mineralogy. The Earth Mineral dust source InvesTigation (EMIT) has collected a global dataset of billions of hyperspectral observations of arid dust source regions. We sample EMIT observations from observed dust-aerosol-producing regions to build a representative hyperspectral dataset (from 0.4-2.45 μm; 7.39 nm resolution) of bare ground soil albedos in dust source regions, filtered to remove the influence of rock, vegetation, surface water,

and clouds. Principal component analyses of spectra from 60 regions yield two dominant spectral types: (1) bright Saharan surfaces with iron oxide and kaolinite signatures and albedos, and (2) darker central and east Asian surfaces with montmorillonite or illite and potentially carbonate. Across dust sources, bare ground soil albedos range from 0.10-0.65% VIS and 0.20-0.85% NIR. Region-specific endmember and outlier spectra occur on each continent, indicating dust source mineralogical diversity relevant to regional biogeochemical processes and local differences in bare soil radiative forcing. On average, EMIT filtered, bare-sediment VIS albedos are 6-10% brighter than unfiltered MODIS albedo products. Compared with laboratory measurements of suspended dust, EMIT dust source spectra show ~40% lower VIS and 7% lower NIR single-scattering albedo. Incorporation of these compositionally diverse dust source regions into Earth system models and future work focusing on refined hyperspectral retrievals of grain size and mineralogy from these source regions, informed by in situ mineralogy, will help better constrain dust's climate and biogeochemical impacts.

Plain Language Summary

Dust from Earth's deserts travels through the atmosphere and influences heating and cooling in the atmosphere, cloud formation, and ecosystems by supplying nutrients to the soil. These effects depend on the minerals contained in the dust, which vary from one site to another. We use observations from the Earth Mineral dust source InvesTigation (EMIT) satellite to analyze a large global dataset of land surface reflected light properties at hundreds of visible and near-infrared wavelengths and determine the bare ground soil properties from major dust-producing locations. The data reveal two main types of dust source surfaces: bright Saharan regions rich in iron oxides and the clay mineral kaolinite and darker regions in central and east Asia dominated by clay minerals such as montmorillonite or illite. We also identify unique regional compositions that may influence local environmental processes. Compared with previous satellite products and laboratory

measurements of dust source sediment samples, our results show brighter bare sediment surfaces than average albedos from the MODIS satellite and darker bare ground surfaces than those of suspended dust measured in lab. Using these regional differences in dust properties will improve our understanding of how to represent and model dust's effects on Earth's climate system.

1. Introduction

Mineral dust emitted from Earth's drylands plays a critical role in Earth system processes with multiple impacts on the climate. Mineral dust suspended in the atmosphere can influence radiative forcing (Sokolik & Toon, 1999) and cloud nucleation (Perlwitz and Miller, 2010). Once deposited, it impacts snow albedo (Warren and Wiscombe, 1980), ocean and soil chemistry, and biogeochemical cycles once deposited (Shao et al, 2011; Jickels et al, 2005). In all of these processes, the nature and magnitude of dust impact depends strongly on dust's mineral composition and the relative abundance of its constituent minerals (Scanza et al, 2015; Li et al, 2020). Mineralogy affects direct absorption - for example, dust containing more iron oxides absorbs more radiation in the visible spectral range due to the strong imaginary component of their refractive indices (e.g., Zhang 2015). Mineralogy also impacts cloud nucleation potential. While the clay minerals of dust have similar radiative properties in the near-infrared, their differing water uptake and hygroscopicity lead to variable cloud condensation nucleation activity (Herich et al. 2009). Laboratory studies with illite, montmorillonite, and kaolinite demonstrate that different clay types nucleate ice at different characteristic temperatures and efficiencies (Zielke et al., 2015). Mineralogy also governs dust's role in biogeochemical cycles, for example in the essential nutrients that it supplies. Iron is a limiting nutrient in the ocean, and its delivery via dust controls phytoplankton growth. The bioavailability of iron depends not only on its total abundance in the dust but also on the specific mineral phase in which it occurs, as iron is more reactive in clay minerals and poorly crystalline iron oxides like

ferrihydrite than it is in crystalline iron oxides like hematite and goethite (Shi et al., 2012). Once airborne, atmospheric chemistry processes alter the speciation of iron minerals in dust, so iron bioavailability depends both on source mineralogy and transport processes (Shi et al., 2012). Finally, dryland surfaces themselves influence the planetary radiative budget through their wavelength-dependent reflectance and absorption of incoming solar radiation (e.g., Braghiere et al, 2023). Most climate models rely on MODIS broadband reflectance products, which aggregate reflectance across wide wavelength ranges (e.g., UV–VIS or VIS–IR), but Braghiere et al. (2023) demonstrate albedo differences of up to 20% and radiative forcing biases of up to 30 W m^{-2} when comparing broadband and hyperspectral estimates that measure surface compositional variability.

Accurately representing these processes in climate models requires detailed information on dust source region characteristics under present and future climate scenarios. Historically, climate models have treated mineral dust as compositionally homogeneous, prescribing a single global mixture with fixed radiative properties such as refractive indices, single-scattering albedo (SSA), and grain-size distributions (see review in Li et al., 2021). This assumption is reasonable for simplifying model initialization but does not reflect the substantial regional-scale spatial variability in dust source mineralogy. Observed single-scattering albedo of dust sources varies regionally by more than 10% depending on source composition, with values in the near-UV and visible particularly sensitive to iron oxide content, while SSA generally increases toward the near-infrared where absorption is weaker (e.g., Di Biagio et al., 2019; Obiso et al., 2024).

Recent studies have begun incorporating spatially resolved mineralogical information to improve the realism of dust representations. For example, Song et al. (2024) compared simulations using uniform dust properties to those initialized with region-specific mineral compositions and found that including mineralogical variability significantly improved the models' ability to reproduce observed dust

optical and radiative behavior. As a second example, first results from the Earth Surface Mineral Dust Source Investigation (EMIT) imaging spectrometer aboard the International Space Station show varied iron oxide content regionally. Accounting for this mineralogy improves climate models of radiative forcing, reducing uncertainty in estimated shortwave direct radiative effect of mineral dust by over 50% (Li et al., 2026). These results highlight the importance of accurately characterizing dust mineralogy; however, determining the mineral composition of dust emitted from all major source regions worldwide remains a substantial challenge.

A variety of observational, laboratory, and remote-sensing approaches have been used to inform mineral dust composition to date. Global soil classification atlases such as those produced by the Food and Agricultural Organization provide categorical soil types, many times based on color, but these products offer only extrapolated mineral composition based on sparse ground truth mineralogy that was generalized from 239 samples to create a global database (Claquin 1999, Journet 2014). Point sources and specific regions have been studied at AERONET collection sites to directly sample airborne dust close to its source (i.e., Kandler et al, 2007; Müller et al., 2010). These samples were analyzed in the laboratory to obtain detailed mineral species, abundances, and grain-size distributions. While highly precise, such measurements are inherently point-based and cannot capture the full spatial variability of global dust sources. Similarly, numerous studies have characterized soil or sediment samples directly from dust source regions (e.g., Moskowitz et al, 2016; Gonzalez-Romero et al, 2023; Gonzalez-Romero et al, 2024a; Gonzalez-Romero et al, 2023b; Panta et al, 2022). Although these analyses yield rich mineralogical and grain size information, they remain geographically limited and do not provide a global view of dust composition. MODIS observations of mineral dust plumes over dark ocean backgrounds provide direct observations of airborne mineral dust's basic radiative properties, but multispectral data cannot resolve specific mineral assemblages or relative abundances.

Addressing variable airborne dust mineralogy globally requires detailed observations that encompass all key dust-emitting regions. Remote sensing provides an effective means of characterizing these properties globally, both for airborne dust and for dust source region surfaces. New high spectral resolution observations of dust source regions in visible to shortwave-infrared are now available that measure electronic and vibrational absorption features, which can be used to retrieve many aspects of mineral composition and improve estimates of bare surface albedo for radiative forcing.

Here, we advance the state of the art in characterizing global dust source regions by providing a comprehensive, spatially explicit assessment of their spectral and mineralogical properties and albedos globally while examining regional variability. We utilize EMIT imaging spectrometer VSWIR reflectance spectra (380–2500 nm) at ~7 nm spectral resolution and 60-m/pixel spatial resolution (Green et al., 2020; Thompson et al., 2024a). EMIT hyperspectral data has global coverage of arid land surfaces in latitudes from ~52 N to ~52 S, enabling the mapping of spectral features that can indicate surface mineralogy. We first develop a method to sample and screen all available EMIT data to build a representative dataset of Earth's dryland surface reflectance properties that minimizes the influence of clouds, snow, water, or vegetation and identifies non-bedrock regions capable of emitting dust. We compile and release this dataset of hyperspectral reflectance for all key dust-emitting regions on Earth (Keebler et al., 2026a) as well as software code for the filtering process (Keebler et al., 2026b). We quantify broadband albedos and mineral absorption features and characterize the variability of these properties worldwide, across dust source regions. We also calculate the single-scattering albedo of dust source surfaces from EMIT and compare to literature values of point source samples and MODIS. This work provides new insights into the spatial heterogeneity of dust source surface mineralogy and albedo and its implications for climate modeling.

2. Methods

To assess the spectral variability of dust source regions as observed by EMIT, we first identify chronically active sources of mineral dust which are important for regional or global earth system processes. We then build a representative dataset of reflectance spectra of these dust-producing surfaces and remove spectra containing signatures of non-sediment surfaces, including clouds, surface water, snow, and vegetation. The resulting dataset is then analyzed to assess spectral variability and quantify albedo, mineral absorption features, and single-scattering albedo to characterize regional differences in mineral dust source signatures (Figure 1).

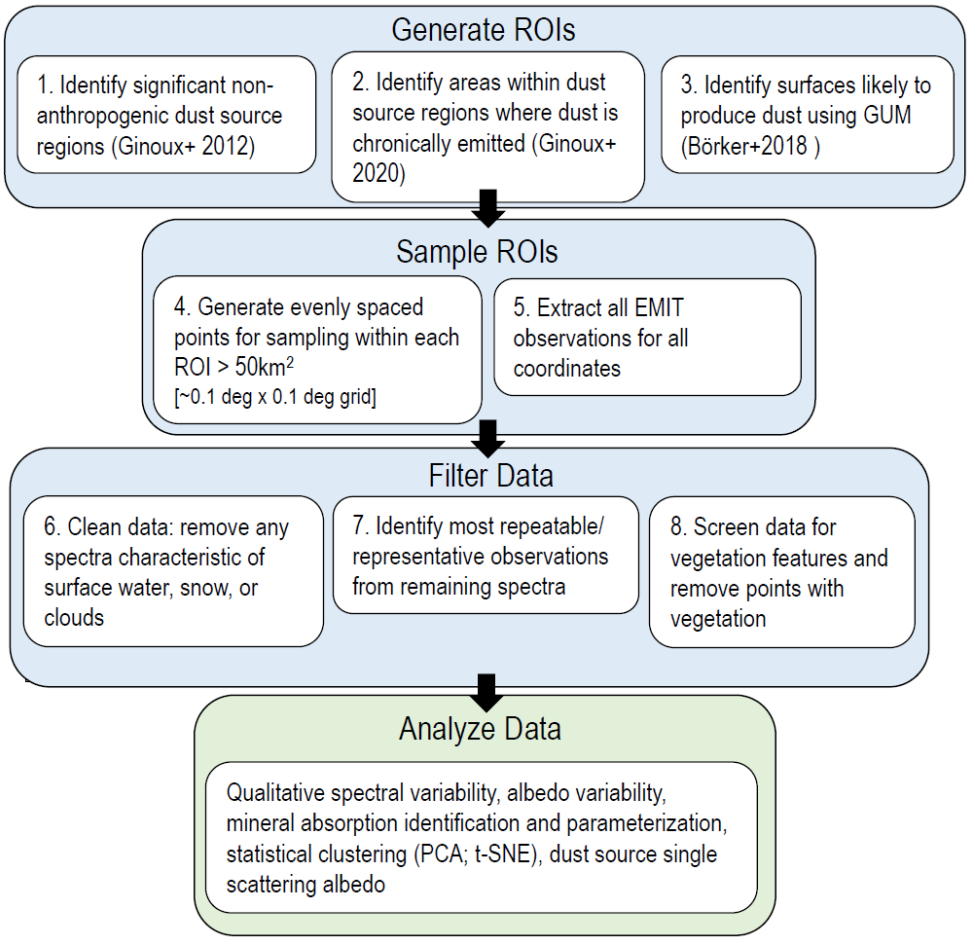

Figure 1. Workflow for data extraction, processing, and analysis. Steps include region of interest (ROI) generation, extraction of EMIT reflectance spectra, automated filtering to remove spectra of clouds, water, flooding, and vegetated surfaces; and spectral property analyses.

2.1 Study Area Selection

We focus on key mineral dust-producing regions from around the world (Figure 2). Ginoux et al. (2012) identified mineral dust-producing regions by calculating dust optical depth from MODIS data, attributing elevated dust optical depth with regional dust sources and surface type at 0.1 x 0.1 degree (~10 x 10 km). Surface types identified by Ginoux et al. (2012) differentiate between sources associated with ephemeral bodies of water ("hydro"), natural sources not associated with ephemeral bodies of water ("natural non-hydro"), and sources associated with greater than 30% anthropogenic land use ("anthropogenic non-hydro"). From this dataset, we select dust-producing regions dominantly classified as hydro and natural non-hydro for our study, excluding those that are anthropogenic in origin. Second, from these regions, we identify surfaces which are "chronic" dust emitters by filtering the Pu et al. 2020 dataset of dust optical depth (DOD) >0.75 and frequency of occurrence (FoO) >1 maps for each year from 2003-2014. We identify locations that had observed dust events for over half of the available years of data (any 7 out of 13 years of data). This ensures study regions regularly emit mineral dust and do not include anomalous areas which were aerosol emitters during, for example, a wildfire. Third, we utilize the Global Unconsolidated Materials database (Borker et al. 2018) to identify quaternary alluvium land surfaces and exclude bedrock, because, although bedrock often underlies high-DOD areas, only sediments can be entrained and transported as mineral dust. We identify all areas within EMIT's data collection mask where all three of the above criteria are met, creating 60 multipolygon regions of interest (ROIs) for this study (Figure 2) (Keebler et al., 2026a).

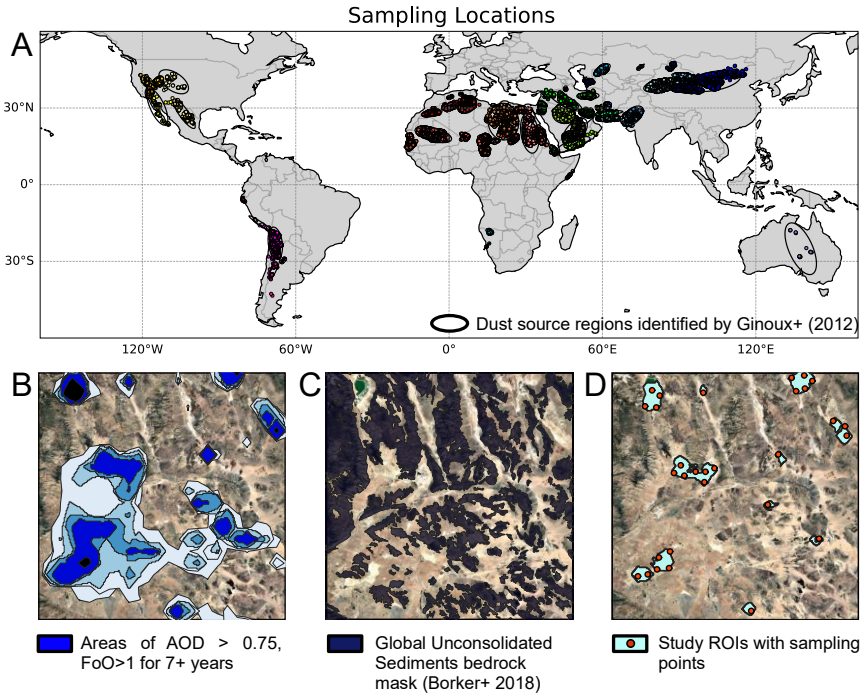

Figure 2. Dust source sampling points selected for this study. (a) Dust source regions of interest and sampling points for this study, color coded by continent. (b-d) map subset of Mojave Desert, USA showing different map layers used to generate ROIs: (b) Areas of AOD > 0.75, FoO > 1 occurring for different numbers of years of dataset; (c) the Global Unconsolidated Materials (GUM) database (Börker+ 2018) overlain on RGB satellite imagery. The GUM identifies areas of unconsolidated sediments and excludes rock outcrops; (d) ROIs produced from overlap of b (any 7 years; occurring >50% of all available observations) and c, with sampling points plotted.

2.2 Data filtering

2.2.1 Spatial subsampling

To sample the dust source ROIs, we generate a field of evenly spaced coordinates within each polygon using a Poisson disk sampling distribution. This produces a set of sampling coordinates proportional to dust source surface area with uniform spatial distribution and an enforced minimum distance between points. We chose a

minimum spacing of 11 km, corresponding to an approximate 0.1 by 0.1 degree grid, to enable analysis of spectral variability at fine spatial scale to assess intra-site heterogeneity. The 60 regions of interest (ROIs) ranged in size from 219 to 567,202 km^2, sampled with 2 to 3319 points, respectively. For each point, we extract spectra from all EMIT L2a reflectance observations of the target coordinates collected between January 2023 and June 2025. These spectra represent Hemispherical Directional Reflectance Factors (HDRFs) calculated for a virtual surface that is level with respect to the geoid (Schaepmann-Strub et al., 2006; Thompson et al., 2020). Consequently, the reflectance spectra will be darker or brighter for slopes facing away from or toward the sun, respectively. The number of spectra extracted per point depends on the number of cloud-, water-, snow-, and vegetation-free spectra, determined uniquely for each point.

2.2.2 Removal of clouds, water, snow, and vegetation

We find that using the version of EMIT-released masks available at time of analysis (Green et al, 2022; Ochoa et al, 2025) results in our dataset including some non-sediment spectra. To ensure the spectra are representative of the reflectance of sediment surfaces, we first perform a series of screening steps to exclude the unwanted signatures of water: clouds, surface water, and snow. Surface water is characterized by consistently low reflectance in the SWIR; we identify water spectra by applying thresholds for low reflectance and flatness between 1500 and 1700 nm (reflectance <0.06; flatness: -0.005<mean slope<0.005). Snow spectra are characterized by absorption features due to water ice with minima at 1021 nm and 1218 nm. These features were identified by fitting Gaussian functions to these specific wavelength minima and evaluating fitted curve position and amplitude against threshold criteria (depth >0.05, position between 1000 and 1050 nm for absorption at 1021 nm; depth >0.15, position between 1200 and 1250 nm for absorption at 1218 nm). To ensure that hydrated mineral phases of interest in dust source regions (i.e., evaporites in playas) with absorption minima at nearby

wavelengths were not confused with snow and systematically excluded from the dataset, we ran the algorithm on hydrated phase spectra from spectral libraries (e.g., Viviano-Beck et al., 2015) to ensure that the approach used to identify snow did not result in removal of hydrated phases.

To detect clouds, we first perform a simple screening for clouds that exhibit inverted water features at 980 and 1126 nm. These features are artifacts of the atmospheric correction algorithm and do not occur in any mineral spectra. We identify these clouds by detecting a local maximum at 1126 nm with an amplitude greater than or equal to 0.02 above a baseline. Clouds have a wide variety of possible spectral signatures which can make them challenging to identify with spectral parameters alone (Figure 3). To detect any remaining cloud-contaminated spectra or wet sediment, we apply pairwise spectral angle outlier detection under the assumption that the properties of the dust-source sediment are relatively constant with time. We find spectral angles between all pairs of spectra in a sampling point, and then calculate the mean and standard deviation of the set of angles. Observations with a mean spectral angle greater than one standard deviation larger than the mean of the group are flagged as outliers. Clouds are spectrally distinct from surface sediments, enabling automated identification and removal of anomalous observations while retaining valid sediment spectra. Additionally, this approach limits the chance that outliers due to flooding or abnormally wet soil are included in the dataset.

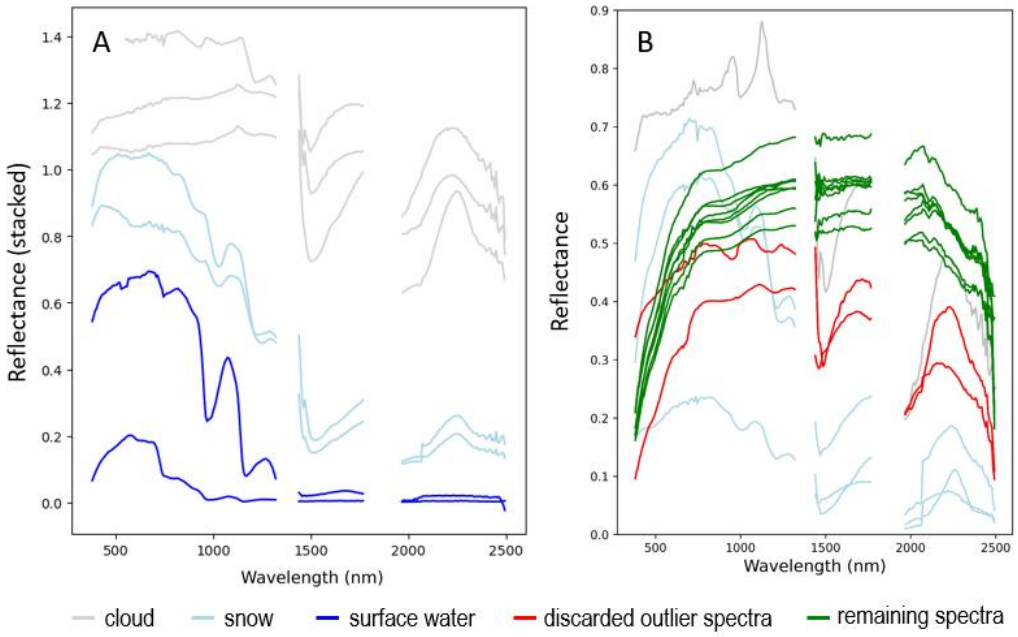

Figure 3. Screening for clouds, snow, and surface water. (a) Examples of easily screened cloud spectra (grey), snow spectra (light blue), and surface water spectra (dark blue). (b) Results of filtering process, color-coded by which spectra are removed at each step. Light blue spectra are identified as snow, grey spectra are identified as clouds with inverted water features, red spectra are clouds identified using pairwise spectral angle calculations for outlier detection, and green spectra pass onto the next screening step.

Relative to bare sediment, vegetation cover lowers visible wavelength albedo, and mineral absorptions can be obscured when superimposed with pigment and plant structure absorptions. We therefore identify and exclude spectra with features indicative of vegetation in order to minimize vegetation's influence and more accurately isolate the spectral signature of bare ground. Photosynthetic vegetation is parameterized by an absorption feature (depth > 0.001) centered at 678nm due to chlorophyll and a positive slope from 680-800 nm at the "red edge." Nonphotosynthetic vegetation is parameterized by absorptions due to cellulose at 1720 nm and 2100 nm. Accurate detection of the 2100-nm cellulose feature in nonphotosynthetic vegetation is complicated by its proximity to ~2200-nm metal-

hydroxyl absorptions features characteristic of clay minerals common in mineral dust-producing sediments.

Chlorophyll and cellulose absorptions are fit with Gaussian curves and the amplitude of the fit curve is calculated as the absorption strength. To account for the overlapping features, we fit a triple Gaussian model with one broad Gaussian for the cellulose absorption at 2100 nm, a second broad Gaussian to accommodate an adjacent ~2200-nm clay absorption if present, and a third, narrow Gaussian for a small artifact sometimes present in EMIT data, fixed at 2108 nm. An empirical threshold for absorption strength (depth > 0.001) was determined through iterative testing on a series of linear mixtures of endmember photosynthetic and non-photosynthetic vegetation spectra mixed with a variety of pure sediment spectra. We empirically evaluated a range of absorption strengths to optimize sensitivity while minimizing false positives and negatives. The empirically derived thresholds were then validated on EMIT scenes with known vegetation extent, confirming that the thresholds reliably identified both types of vegetation (Figure 4). Spectra exhibiting the target feature with absorption strength exceeding the selected threshold were excluded from subsequent analyses.

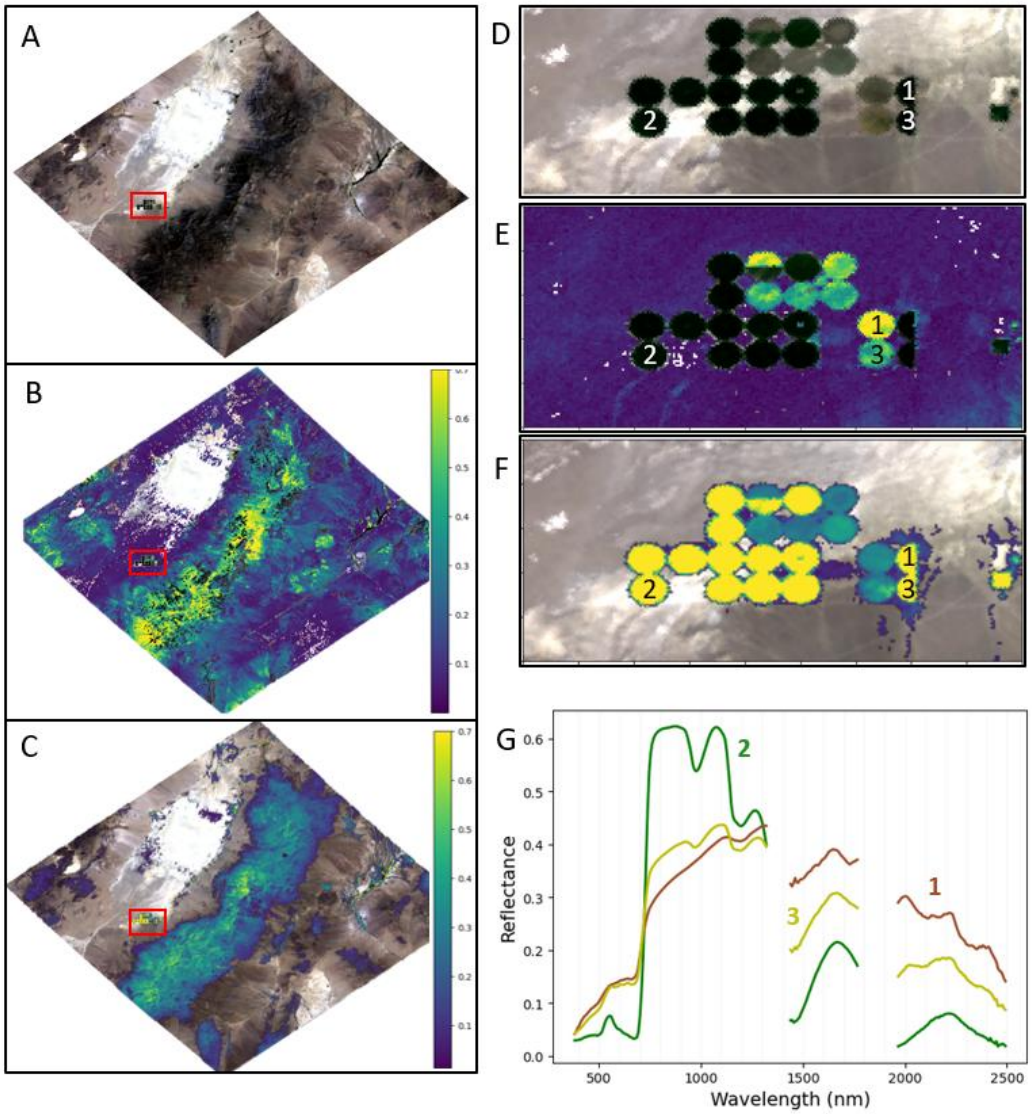

Figure 4. Vegetation feature detection and mapping. (a) RGB of EMIT image EMIT_L2A_RFL_001_20230808T174016_2322012_013 over Railroad Valley, NV. (b) RGB image overlain by detection of NPV via 1725nm and 2100nm features. Colorbar indicates 2100-nm absorption feature depth. (c) RGB image overlain by detection of GV via 678nm chlorophyll absorption. Colorbar indicates depth of absorption feature. Panels D-F: Zoom-in on (d) RGB image; (e) NPV detection map (2100nm cellulose absorption feature strength); and (f) GV detection of crop circles south of the playa, indicated by red box in panels a-c. (g) EMIT spectra from crop circles indicated in panels D-F: (1) strongest NPV signatures with weaker GV signature; (2) strong GV signature with no detected NPV signature; (3) similar strength NPV and GV signature detected.

Once all observations are filtered for clouds, snow, outlier moisture content, and vegetation signatures, the remaining accepted spectra at each sampling point are averaged. We make available this mean spectrum for each sampling point in the final set of filtered dust source region surface spectra (Keebler et al., 2026a).

2.3 Variability Analysis

Once the dataset of quality-filtered spectra is compiled for each sampling point across all dust source regions of interest, we calculate a series of parameters to characterize spectral properties. These include calculation of broadband and wavelength-specific albedos, continuum removal to isolate characteristic mineral absorption features, band depths, and single-scattering albedo. We also perform statistical dimensionality reduction techniques like principal component analysis to cluster the data, identify representative and outlier spectra, and identify primary sources of variability across the spectra, which like most Earth imaging spectroscopy data are highly correlated (Thompson et al. 2017).

2.3.1 Albedo

Broadband visible (VIS) and near-infrared (NIR) and single-wavelength albedos were calculated from EMIT reflectance spectra. Single-wavelength albedo was taken as the reflectance at a specific wavelength, and broadband values are the mean reflectance of all channels across the VIS (450-700) and NIR (2000-2500 nm). For comparison to laboratory data, reflectances were converted to single-scattering albedo using equation 8.49 in Hapke (2012), assuming isotropic scattering.

2.3.2 Mineralogical Absorptions

To assess the mineralogical diversity conveyed by the spectral dataset, we analyze spectral absorption features whose shapes and minima positions are characteristic of specific mineral phases. We performed continuum removal, in which a convex

hull is fit to the spectrum and then divided into it, normalizing the spectrum and making it easy to characterize and compare the shape, position, and depth of absorption features. To quantify mineral absorption strength, we calculate a band depth relative to the continuum according to Clark and Roush (1984).

2.3.3 Dimensionality Reduction

To characterize sources of variability across the highly correlated global spectral dataset, we employ statistical techniques to reduce dimensionality. We applied principal component analysis (PCA), which linearly transforms spectra into a smaller set of orthogonal components which capture the majority of the variance of the dataset. We then used k-means to cluster the data in principal component space, then identified representative spectra (at the center of each cluster) and outlier endmember spectra (spanning the vertices of a convex hull fit to the principal component data clusters). These representative and outlier endmember spectra facilitate characterization of typical spectral signatures for each dust source as well as identification of any local, mineralogically distinct dust source sediments within them. We apply this process to the full global dataset, to each geographic region (i.e., Northern Africa, Middle East, Asia, North America, Australia, and remaining Southern Hemisphere), and to each individual dust source region of interest.

3. Results

A total of 34,210 points were sampled from ROIs encompassing 5,979,150 km^2 of dust-producing surfaces in arid environments in northern Africa, the Middle East, Asia, North America, South America, southern Africa, and Australia. Of these sites, 23.4% were excluded due to vegetation and 1.5% were excluded because all available observations were identified as surface water, had cloud cover, or otherwise failed quality-control filtering, leaving mean spectra from 25,672 points from the 60 dust source ROIs for analysis of global spectral properties.

3.1 Statistical Dimensionality Reduction and Data Visualization

We performed PCA using nine components with the first three components explaining 99% of the variability. The first eigenvector is consistent with characteristic Saharan dust source spectra, including hematite/goethite absorptions in the visible and a kaolinite doublet near 2210 nm. The second eigenvector captures variability associated with liquid water, indicating hydrated salts and damp sediments. Higher-order components indicate additional absorptions in the visible and infrared (Fig. 5a). Representative and outlier spectra were selected from PCA K-means to span the full diversity of our global reflectance dataset (Fig. 5b).

The data are largely within two distinctive clusters: one containing most of the Saharan dust sources and the other containing most of the Asian dust sources. The centroid of the Saharan-dominated cluster is located in the Nile River basin, and the centroid of the Asian-dominated cluster is located in the Turpan Pendi depression in northwestern China. The spectra of these two centroids (Fig. 5c) represent two typical dust endmembers, reflecting globally distinct mineralogical regimes of bright, quartz, kaolinite, and iron oxide-bearing sediments and darker, illite clay-dominated sediments.

The outlier endmember spectra (Fig. 5d) span the range of spectral variability observed in principal component space and include: (1) high- and low-albedo spectra representing the brightest and darkest sediments in dust source regions; (2) hydrated sediments, particularly gypsum-bearing sediments found in evaporate deposits and damp sediments like those found in the Great Salt Lake, Utah, USA; and (3) spectrally distinct mineralogies, including extremely hematite-rich sediments from the Sahara and shallow, continuous visible–NIR slopes likely related to variable iron content and oxidation state in sources like the Etosha Pan, Namibia.

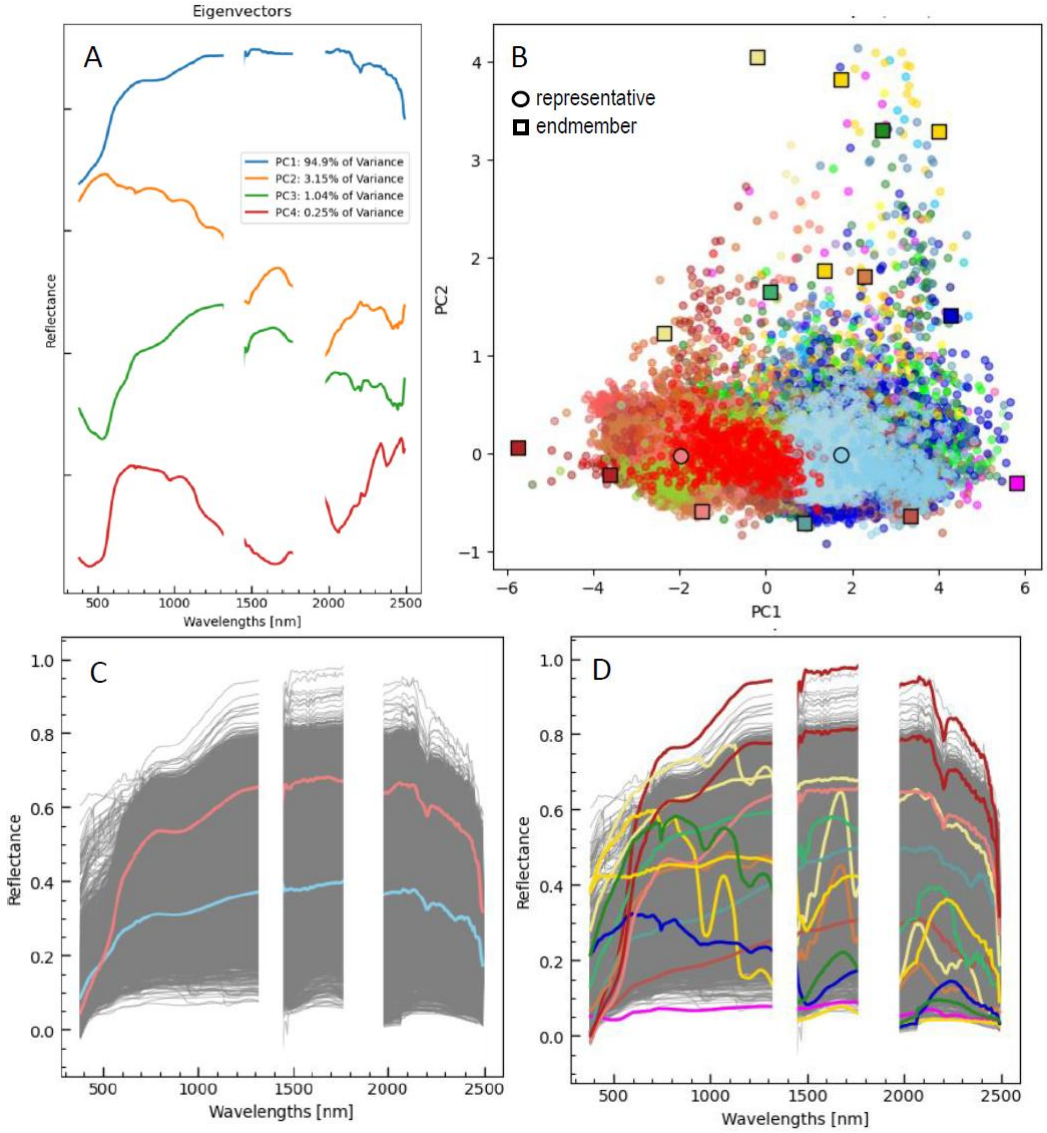

Figure 5. Principal component analysis on global spectral dataset. (a) First four eigenvectors for PCA with 9 components. (b) Scatterplot of global spectra dataset transformed into PC space (PC2 vs PC1), with cluster representatives (chosen via K-means) and endmembers/outliers (chosen by fitting convex hull to data cloud in PC space) indicated. (c) Representative spectra plotted over all spectra in the dataset; color corresponds to dust source region. (d) Endmember/outlier spectra plotted over all spectra in the dataset.

3.2 Spectral Variability by Region

Though spectral properties of Saharan and Asian dust dominate global variability, dust source regions with smaller spatial distributions play important roles in regional to local atmospheric and biogeochemical processes. To capture representative dust signatures from each of these regions (North America, Australia, and other Southern Hemisphere sources in southern Africa and South America), as well as to characterize diversity within the geographic regions which dominate mineral dust emissions (Northern Africa, the Middle East, and Asia), we applied the PCA, k-means clustering, and convex hull fitting workflow separately within each geographic region to illustrate characteristic spectral properties of mineral dust sources in different parts of the world (Figure 6).

There are geographic trends in overall radiative properties such as albedo (Figure 6). Dust source regions in Northern Africa, the Middle East, and Asia have the largest and most globally significant dust emissions. Representative spectra from dust source ROIs in Northern Africa are brightest, with strong absorptions and steep spectral slopes in the VIS and a sharp shoulder near 750 nm transitioning to flatter, high reflectance in the NIR (VIS-NIR albedo ~0.3-0.75). In contrast, Asian dust source regions have shallower VIS slopes and a more gradual VIS-NIR transition, with overall darker reflectance (VIS-NIR albedo ~0.2-0.5). Spectra from Middle Eastern dust sources display intermediate properties. Spectra from regional sources in North America, Australia, and the Southern Hemisphere tend to be darker than Saharan sources, although some sources associated with ephemeral water and evaporite minerals in North America also have very high reflectance (albedos up to ~0.6).

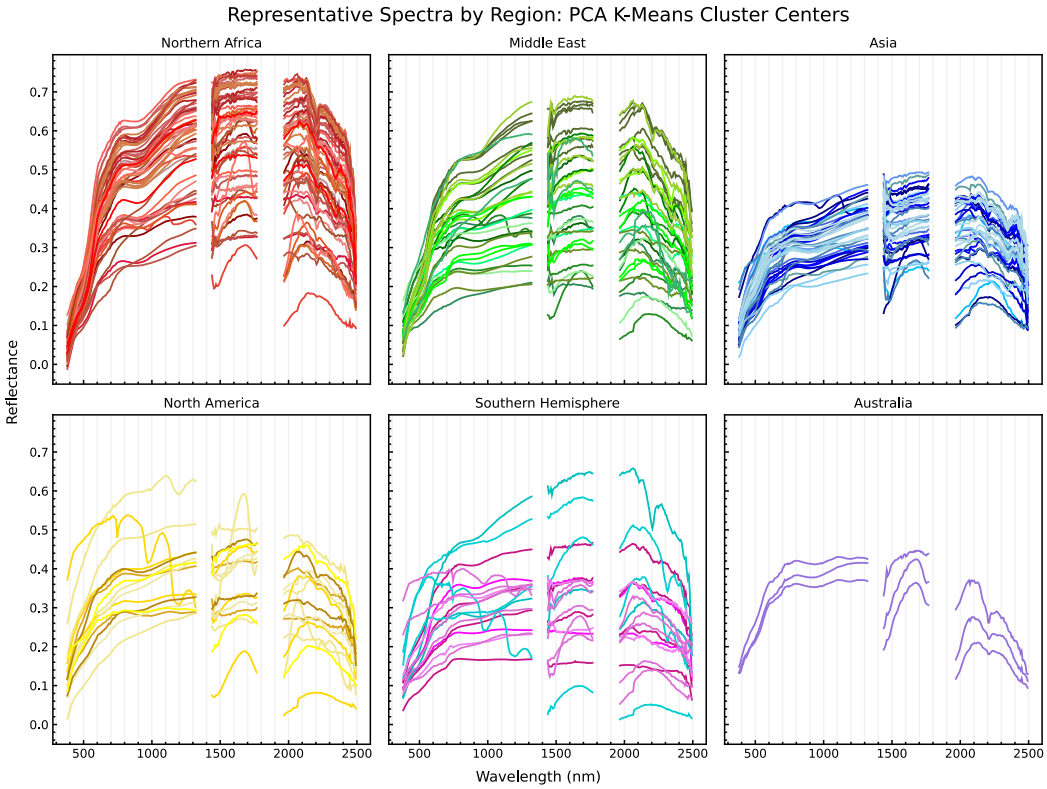

Figure 6. Representative typical spectra for each dust source ROI, grouped by geographic region. Spectra selected for plotting are closest to the center of the data cluster in PC space for each ROI.

Figure 7 shows the representative spectra from Figure 6, continuum-removed, to identify mineral absorption features and characterize mineral assemblages of each dust source region. Spectral features in the wavelength range from 400-1200 nm indicate the presence of iron oxides and other iron-bearing phases across dust source regions, as well as water features in hydrated minerals. From 2000-2500 nm, absorptions indicate clay minerals, carbonates, and sulfates. Spectra from Saharan sources consistently exhibit a strong, broad absorption with distinct minima at 500 and 535 nm, diagnostic of hematite and/or goethite. Most Saharan spectra exhibit a diagnostic doublet with minima at 2160 and 2210 nm, consistent with kaolinite. By contrast, spectra from Asian dust sources show weaker iron oxide absorptions shifted to slightly shorter wavelengths, along with a broader iron-related feature

centered near 1000 nm. Asian dust source spectra typically display a single absorption near 2210 nm, suggesting the predominance of Al clays like illite and montmorillonite. The iron-related feature could be an iron oxide or Fe substitution in a clay or carbonate mineral. Additionally, a subset of spectra from both Saharan and Asian regions exhibits absorptions near 2345 nm, indicative of illite (which has 2210 nm and 2340 nm absorptions) or carbonate phases (which has a 2320-2340 nm feature). Some dust sources in Asia, North America, and the southern hemisphere have sharp absorption features at ~960 nm and ~1150 nm, indicating the presence of hydrated salts and/or wet sediments, associated with ephemeral water.

Dust sources in North America, Australia, and the southern hemisphere also exhibit iron absorptions at ~500 and ~950 nm, but like Asian and Middle Eastern dust source regions, the absorptions are much weaker than in the Sahara (band depth ~0.05 versus up to 0.25 in the Sahara). Some spectra from these regions also have a small, broad absorption at ~1200 nm, indicating additional iron-bearing phases like maghemite. NIR absorptions >2000 nm indicate that sources in North and South America and southern Africa appear to be dominated by illite/montmorillonite, whereas sources in Australia exhibit a doublet absorption at 2210 nm attributable to kaolinite.

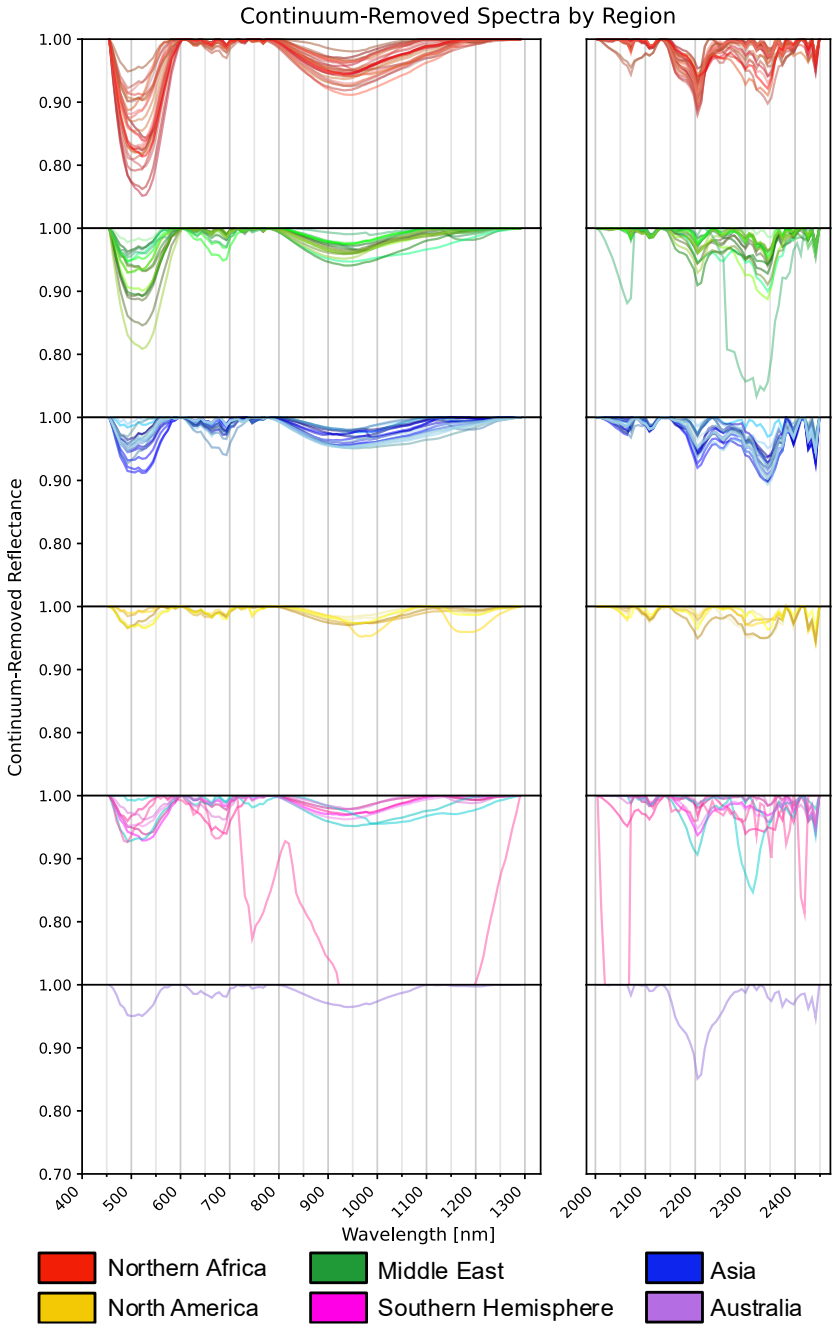

Figure 7. Continuum-removed reflectance spectra of mean of all reflectance data for each location ROI, grouped by global region. (a) VSWIR (400-1300nm) continuum-removed spectra with iron oxide absorption features (~500nm, ~900nm) evident. (b) nIR(2000-2450nm) continuum-removed spectra with clay mineral (~2200nm, 2350nm) and carbonate (2350nm) absorption features evident.

3.3 Local Spatiotemporal Spectral Variability

In addition to assessing spectral variability across dust source regions, the fine sampling scale enables study of dust source spectral variability spatially and temporally within individual point sources. As one example, we examine the Bodélé Depression in Chad, one of the world's most prominent dust source regions (Figure 8). Spatial patterns in albedo and band depths are consistent with visible image surface properties: computed visible albedo was higher within the depression's sediments and lower in the surrounding sands (Fig. 8a); Fe-oxide absorptions at 530 nm were stronger where the sands surrounding the depression are redder (on the eastern side) (Fig. 8b); and metal-OH absorptions at 2210 nm, indicative of kaolinite and other clays, were stronger in the sediments in the depression and less prominent in the surrounding sands (Fig. 8c).

For each of the mapped parameters, temporal variability is quantified by calculating the standard deviation of parameter values for all individual EMIT observations taken for each sampling coordinate that passed through the filtering procedure, then taking the mean of those calculated standard deviations from all the sampling coordinates. Spatial variability is quantified by calculating the standard deviation of parameter values for the mean spectrum from each sample coordinate. Variability in spectral shape, overall reflectance, and absorption features is far greater across sites than it is within spectra collected at a given site. In the Bodélé Depression, spatial variability (Fig. 8d) is greater than temporal variability (Fig. 8e) by a factor of 10 for band depth parameters (standard deviation of ~0.01 vs ~0.001, respectively) and a factor of 3 for VIS albedo (~0.03 vs ~0.01). These results demonstrate that observed variability in spectral parameters reflects true differences in surface properties rather than discrepancies introduced by combining multiple satellite acquisitions taken at different times.

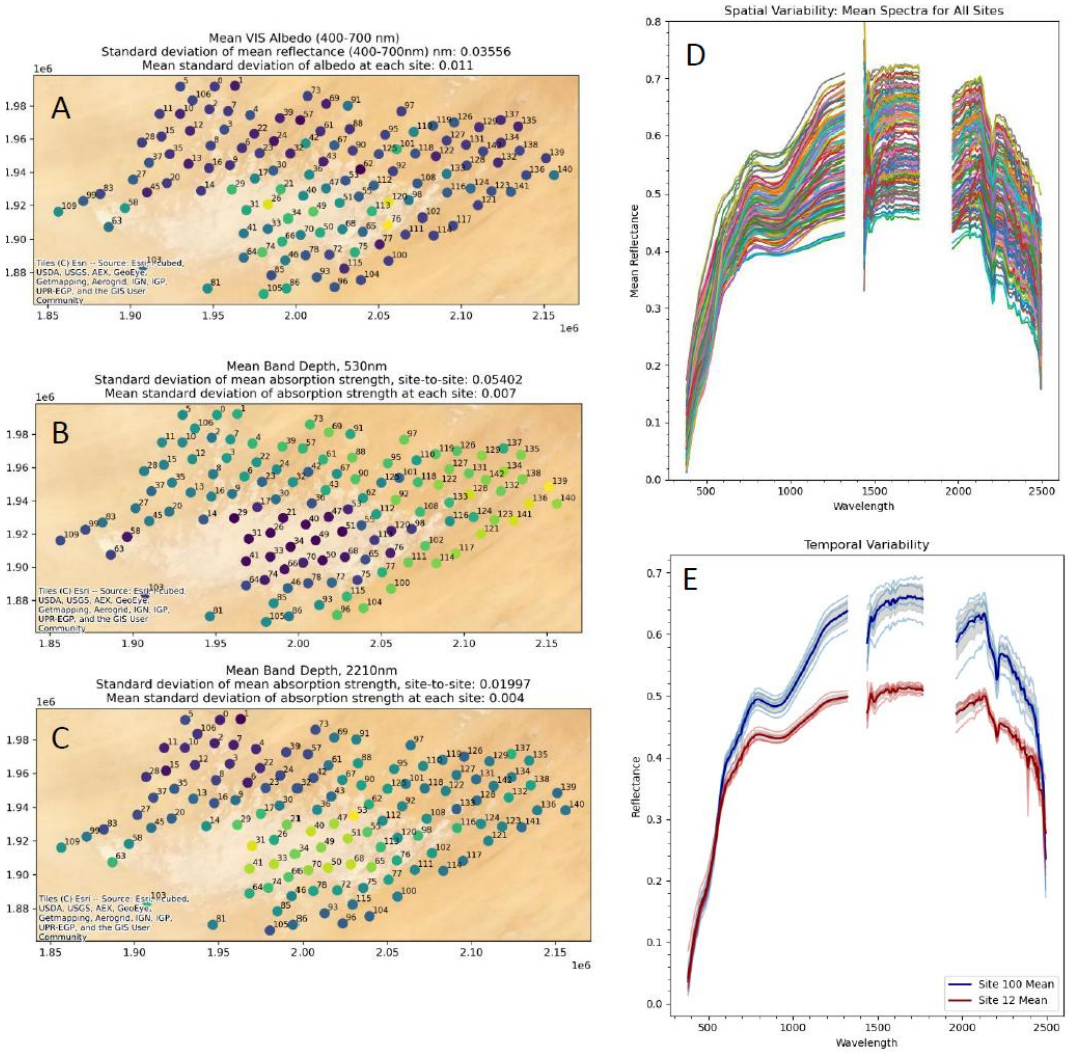

Figure 8. Spectral variability of Bodele Depression dust source, Chad. (a) map of mean VIS (400-700nm) broadband albedo from all observations per sampling point. Panels B-C: band depth maps for (b) 530-nm iron oxide absorption and (c) 2210-nm clay mineral absorption. (d) Mean spectra from each sampling point, highlighting spatial variability. (e) All observations for two sampling point, highlighting relatively small temporal variability across repeat observations of a given sample coordinate.

3.4 Global Variability Parameterization

Mineral dust's role in Earth system processes depends on regional compositional variability of dust, and dust source surface spectra are a proxy for this variance.

Maps of broadband albedo in the VIS and NIR are highest in northern Africa and for point sources corresponding to playas with bright salts, as in the southwestern USA (Figure 9). Mineral maps show the global distribution of mineral phases detected (Figure 10). The strongest iron oxide absorptions are associated with dust sources in the Sahara and the Middle East. Al-phyllosilicate absorption strengths at 2210 nm are also highest in the Sahara (maximum depth ~0.15), while absorption strengths at 2350 nm due to clays and carbonates are more evenly distributed across the globe.

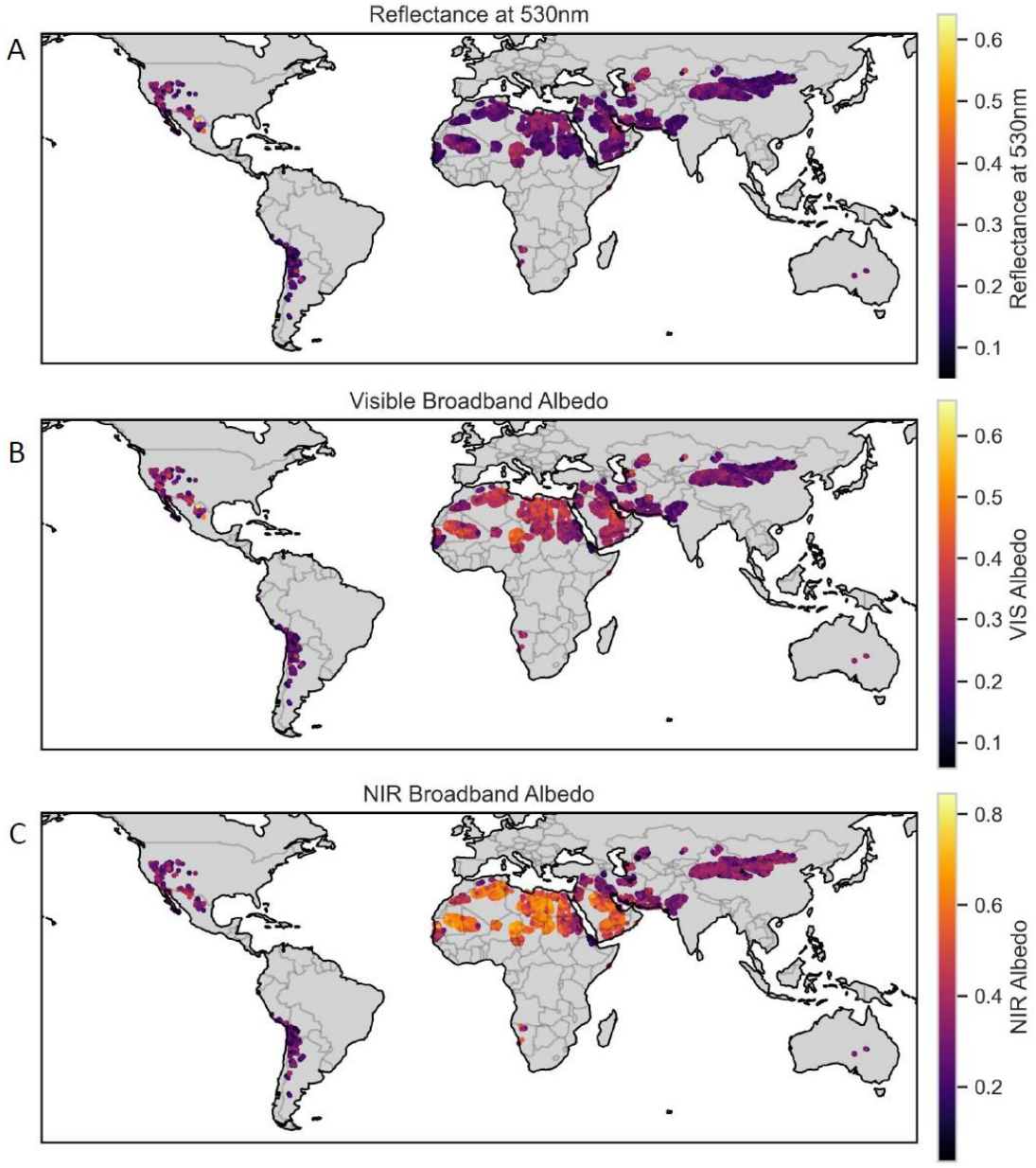

Figure 9. Albedo calculated from reflectance dataset for (a) 530nm, (b) visible broadband (400-700nm), and (c) IR broadband (1300-2450nm).

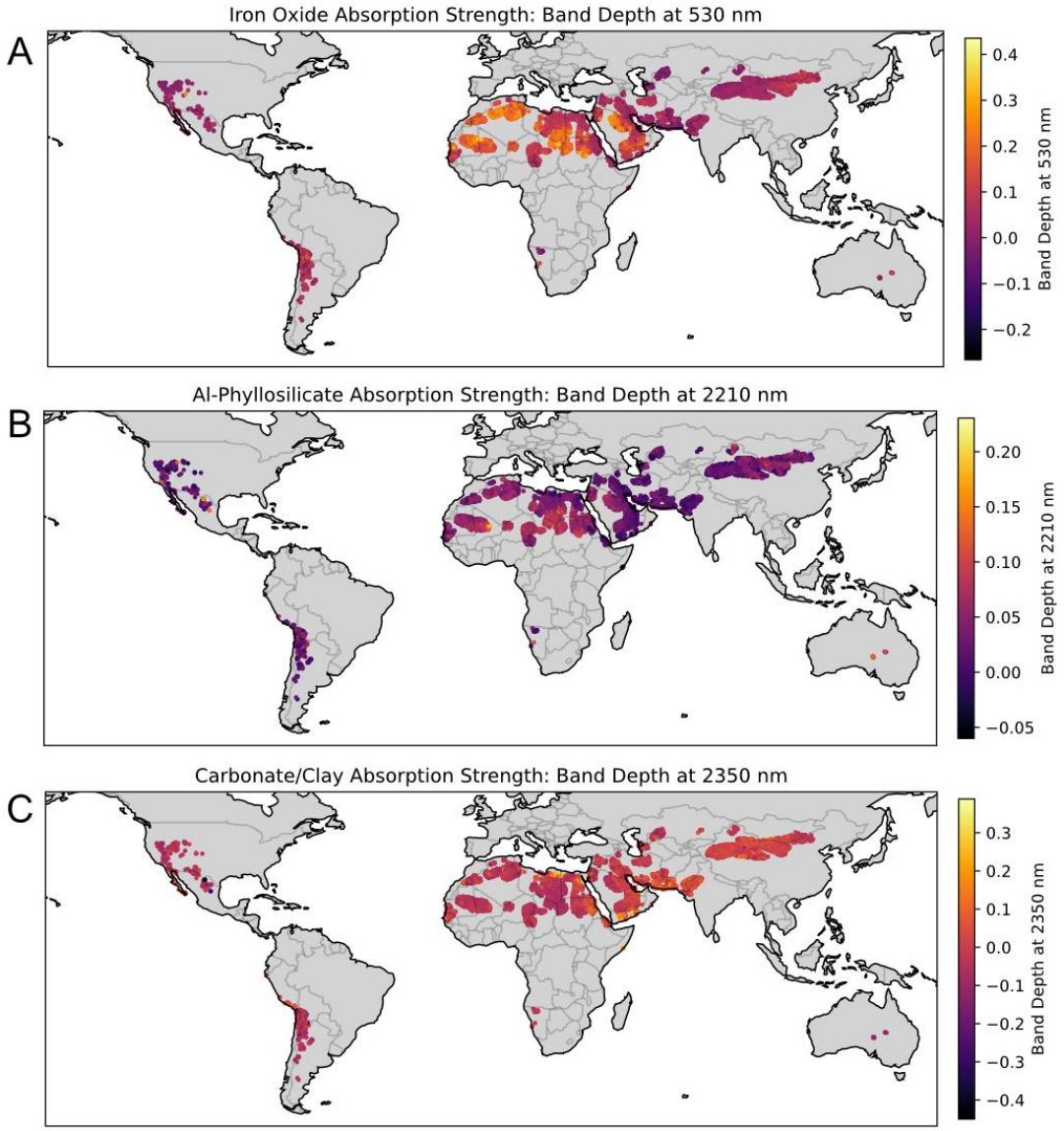

Figure 10. Absorption strength maps for minerals common in dust-producing sediments: (a) iron oxides (hematite/goethite), (b) aluminum phyllosilicates, (c) carbonates/clay minerals.

Histograms colored by geographic region (Figure 11) illustrate the spatial distribution of albedo and mineral absorption strength. VIS and NIR broadband albedos exhibit a distinctly bimodal distribution, with Saharan dust characterized by higher reflectance values, Asian sources tending to be darker by a factor of ~2x, and Middle Eastern dust sources spanning the range of values. By contrast, the

distribution of albedo at 530 nm does not have two distinct populations globally, though the Saharan dust has bimodal distribution regionally (Fig. 11c). Iron oxide absorptions at 530 nm are markedly stronger in some Saharan and Middle Eastern dust source regions than nearly any other regional ROI, except for a few in North America (Fig 11d). Histograms for clay and/or carbonate mineral absorption (Figs. 11e and 11f) make clear that Saharan dust has systematically stronger absorptions at 2210 nm (Al phyllosilicate) and weaker absorptions at 2350 nm (illite or carbonate) with Asian dust having the strongest 2350 nm absorption features.

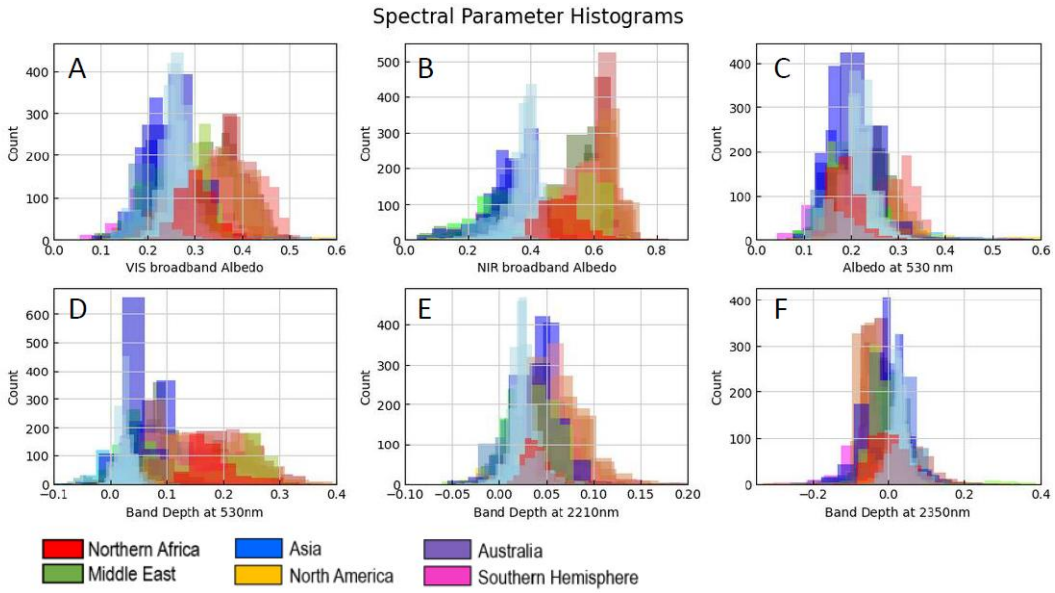

Figure 11. Histograms of spectral parameters, color-coded by region: (a) VIS broadband albedo, (b) nIRbroadband albedo, (c) reflectance at 530nm, (d) band depth at 530nm (hematite/goethite absorption strength), (e) band depth at 2210nm (kaolinite/Al-phyllosilicate absorption strength), (f) illite/carbonate absorption strength.

3.5 Comparison to MODIS Albedo Products

To evaluate how EMIT-derived dry land albedos compare with values commonly used in Earth system models, we compare our results to MODIS global albedo products and single channel albedo products. Note that EMIT albedos have been

filtered to represent bare surface, dry, dust-emitting source region sediments at 0.1 deg scale whereas MODIS products capture surface albedo at 0.5 deg scale without such filtering. Each has different utility in Earth surface models.

3.5.1 Comparison to MODIS broadband

The MODIS MCD43C3 Daily L3 Global CMG V061 albedo product provides albedo estimates for the first seven MODIS narrow bands as well as three broadband albedos, including a visible broadband spanning 300–700 nm. We compute most closely analogous visible-band albedos from the EMIT dataset from 380 nm–700 nm, which lack the UV part of the range. Because the MODIS product has high temporal resolution, we sample MODIS albedo once per month over the study period and examine the relationship between MODIS and EMIT visible albedo at each sampling location (Figure 12a). The relationship between EMIT and MODIS albedos depends on season (Figure 12b). During northern hemisphere winter months, MODIS visible albedo is often substantially higher than EMIT in some regions, likely due to the presence of snow-covered surfaces in the MODIS product. In contrast, snow-contaminated pixels are removed from the EMIT dataset using our filtering algorithm, leading to systematically lower wintertime albedos in EMIT relative to MODIS (Figure 12b).

During the dry northern hemisphere summer months, the relationship between EMIT and MODIS is less influenced by changes in snow cover and surface water, allowing a more direct comparison of how the two products represent land surface properties. Overall, EMIT and MODIS visible albedos are broadly consistent across sites, although some locations exhibit larger discrepancies. These differences are most pronounced at playa-dominated sites, where variable surface conditions such as flooding or salt crust development contribute to increased temporal and spatial heterogeneity. Across the full dataset, EMIT visible albedo is on average 16% brighter than MODIS.

We also compare MODIS narrow-band albedos from 620-670 nm and 1230-1250 nm to EMIT data at the same wavelength ranges (Figure 12c) for the best one-to-one comparison of the sensors. From 620-670 nm, the relationship exhibits a shallower slope (0.80) and a positive offset of 0.10. Although EMIT albedos are generally higher than those from MODIS, darker sites show a larger positive bias in EMIT relative to MODIS, whereas brighter sites cluster more closely around the 1:1 line. This wavelength range is influenced by both iron oxide absorptions and vegetation absorptions. Because MODIS pixels are substantially larger than EMIT pixels, a given location may contain vegetation within the MODIS footprint even when a vegetation-free EMIT pixel is sampled. In contrast, in the near-infrared (1230-1250 nm), where no diagnostic hyperspectral absorption features from minerals or vegetation are present in the EMIT data, the relationship between MODIS and EMIT reflectance has a slope of 0.96 and an offset of 0.06. This near-unity slope indicates strong consistency between the datasets in NIR spectral continuum wavelengths, while the positive offset indicates that EMIT is systematically brighter overall.

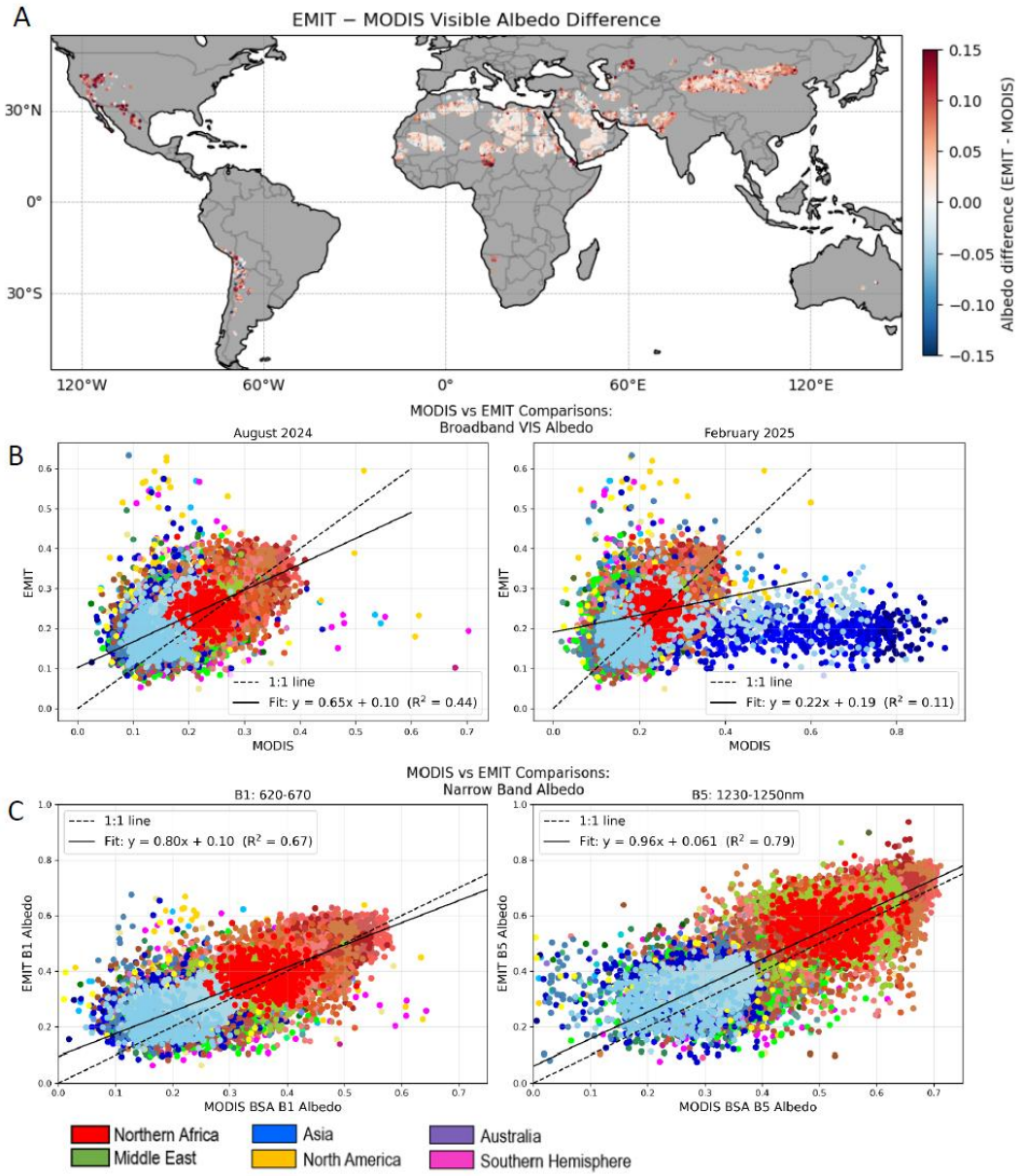

Figure 12. Comparison of albedos calculated from EMIT with MODIS albedo product. (A) map showing difference in EMIT and MODIS broadband albedos for study dust source regions for August. Areas where EMIT is brighter are red; areas where MODIS is brighter are blue. (B) point by point comparison of EMIT vs MODIS broadband albedo retrievals for August and February 2024, showing changes in MODIS surface reflectance due to changes in ground cover (snow). (C) point by point comparison of MODIS narrow bands vs EMIT mean reflectance over the equivalent wavelength range.

3.6 Single-Scattering Albedo versus Dust Region Samples

To assess how satellite-derived albedo estimates compare with laboratory measurements of single-scattering albedo (SSA) of dust source sediments, we compare remotely sensed SSA of the land surface to the SSA of dust source sediments measured in laboratory experiments by Di Biagio et al. (2019). In the lab, sediments were suspended in a chamber to allow size fractionation analogous to natural dust emission, and their spectral shortwave scattering, absorption, and size distribution were measured to retrieve SSA. We compare to SSA computed from EMIT reflectance data at the same sediment sampling locations.

At visible and shortwave near-infrared wavelengths, SSA values derived from EMIT are systematically lower than those measured in the laboratory, which is not unexpected, given that materials that are suspended in the air are finer-grained and therefore expected to be brighter. What is noteworthy is that the difference in brightness is especially pronounced at VIS wavelengths (<~600 nm) where VIS SSA of suspended dust lab samples is ~30% darker than EMIT data whereas NIR SSA is only ~3% darker than EMIT (Figure 13). SSA values derived from EMIT also have steeper slopes in the VIS from 380 nm to 660 nm than in suspended dust (~0.002 EMIT vs ~0.0002 Lab). Slopes from 660 nm to 950 nm are relatively flat in both EMIT and lab SSA (~ 0.0001 EMIT vs ~0.00004 Lab). If we assume the samples from the same area had the same initial composition, the change in shape suggests that mineral fractionation is happening during lofting in addition to grain size change, i.e., suspension leaves behind on the surface materials with more pronounced iron oxide absorptions (see section 4.2 for further discussion). The magnitude of the difference (spectral shape) is consistent in ~2/3rds of sites.

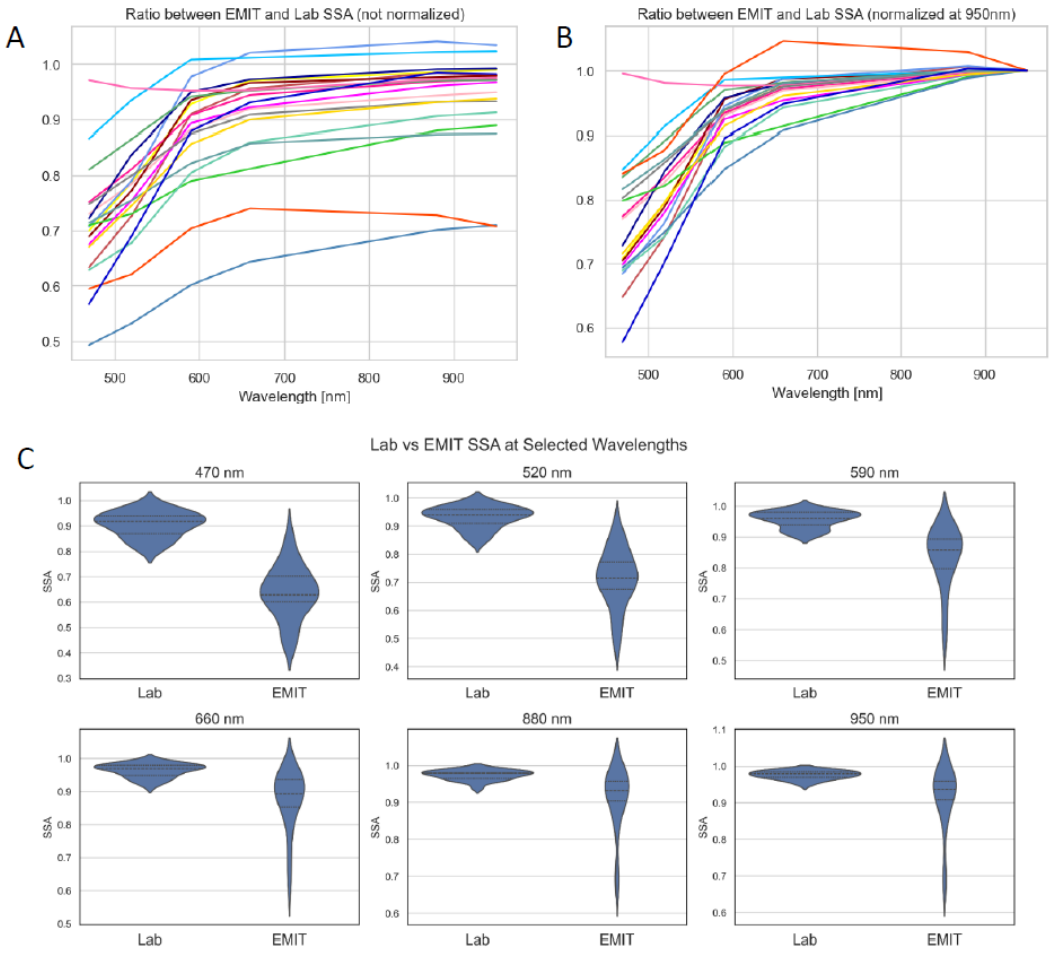

Figure 13. (a) Ratio between SSA calculated from EMIT data and measured in the lab (Di Biagio+ 2019). (b) ratio between SSA calculated from EMIT data and measured in the lab, normalized at 950 nm. (c) Violin plots comparing range of lab-measured SSA values (Di Biagio+ 2019) vs calculated SSA values from EMIT data for different wavelengths.

Additionally, across all wavelengths, spread in SSA is greater in EMIT data of regional dust sources than in lab-measured SSA of Di Biagio et al (2019) samples (0.57 average spread in EMIT data vs 0.30 average spread in lab data). This illustrates shows the diversity of global dust source mineralogy—relative to existing laboratory samples—and how EMIT observations can extend our

understanding of global dust SSA variability beyond the limited number of laboratory samples of suspended dust presently in the literature.

4. Discussion

4.1 Global variability in dust composition

Global analyses of dust source spectral properties, including PCA and k-means clustering, as well as maps and histograms of albedos and absorption features, reveal substantial heterogeneity in dust source region spectral properties globally and two dominant compositional types of mineral dust. Saharan dust sources are generally brighter (90%), with strong hematite and goethite absorptions in the visible and diagnostic kaolinite features at 2210 nm, whereas Asian dust sources are darker, exhibit weaker iron oxide absorptions, and have clay absorption features diagnostic of illite and/or montmorillonite, perhaps with carbonate. Although the Saharan sediments absorb more strongly at 530 nm due to strong presence of highly absorbing iron oxide minerals like hematite and goethite, Saharan sediments are overall brighter than Asian dust source sediments. This contrast likely reflects compositional and grain size differences in the bulk sediments: Asian dust sources contain additional iron-bearing phases that contribute to overall absorption, whereas Saharan sediments are dominated by quartz-rich sand which has high reflectance across most wavelengths and elevates albedo despite the presence of strongly absorbing iron oxides. These observations are consistent with previous lab studies that characterize mineral dust composition from major dust sources. For example, Shao et al. (2008) analyzed samples of mineral dust collected from a dust storm in northern China and found the clay fraction to be dominated by illite/smectite (83%). Compositional studies from Saharan dust source regions such as the Bodélé depression in Chad report kaolinite in addition to illite/montmorillonite, quartz, and iron oxides (hematite and goethite) (Moskowitz et al., 2016; Moreno et al., 2006).

Dust sources in North America, Australia, and elsewhere in the southern hemisphere exhibit a variety of iron and clay absorption features, indicating controls from local geological bedrock, weathering, or sediment transport differences which are important in regional climate processes. For example, mineral dust from Australia is a key source of iron for oceans in the southern hemisphere (Radhi et al., 2010). Although its iron oxide absorption features are weaker than dust sources in the Sahara, it has strong absorptions indicating clay minerals that are likely iron-bearing, corroborated by laboratory studies showing that Australian dust sources are high in iron (Radhi et al., 2010). Since iron in clay minerals is more bioavailable than iron in crystalline hematite and goethite (Shi et al., 2012), these clay mineral detections identify dust sources with important contributions to biogeochemical cycles.

Mineralogical differences are primary drivers of spectral variability and govern dust's radiative properties and role in biogeochemical processes. Importantly, Saharan dust is distinctive and its properties should not be used globally for these processes. The full global dust source region reflectance dataset (Keebler et al., 2026a) is publicly available, enabling further examination of the mineralogy and relative strength of diagnostic features for any dust source, including those relevant at hyper-local scales. Future work will prioritize advancing beyond mineral detections to quantitative mineralogy retrievals to understand source sediments' precise mineralogical contributions to Earth's mineral dust budget (see also Section 4.3).

4.2 Comparison to prior estimates of dust albedo properties

Estimates of land surface albedo used in Earth system models (ESMs) primarily rely on broadband albedo products derived from MODIS observations with specific wavelength ranges used vary depending on the specific ESM and application. Ecosystem models with a focus on vegetation processes typically employ 400-700 nm for photosynthetically active radiation and 700-2500 nm for the near-infrared

(Braghiere et al., 2023), while models such as NOAA GFDL's climate model use broader ranges (175-800 nm for UV-VIS and 800-5000 nm for NIR) which extend beyond EMIT's spectral coverage.

Consistent with our results from EMIT, MODIS albedo products show that the Sahara and Arabian Peninsula exhibit the highest broadband VIS and NIR albedos, while other dust source regions are generally darker. However, EMIT hyperspectral observations show nuances. For example, although broadband albedo in MODIS and EMIT indicates the Sahara is highly reflective relative to other dust source regions, our 530 nm albedo map (Fig. 9a) shows reduced variability between the Sahara and the rest of the world because of increased strength of iron-oxide absorptions in this wavelength range. Highly VIS-absorbing hematite and goethite likely occur at greater abundance in these sediments despite their high overall reflectance. This highlights the importance of mineral-specific features in controlling wavelength-dependent radiative behavior.

While broadly consistent, point-to-point comparisons between EMIT filtered dry, bare ground surface albedos and MODIS broadband VIS albedo show that EMIT is 16% brighter than MODIS on average. Several factors may contribute to this offset. First, the MODIS visible broadband includes wavelengths extending into the near-UV (~300–380 nm), which are not covered by EMIT and are generally more absorbing, potentially lowering MODIS broadband albedo. To mitigate this effect, we also compare EMIT to MODIS narrow-band albedos provided for wavelengths where EMIT has full coverage (Figure 12c). EMIT dust source region albedos are still 6-10% brighter than MODIS on average in MODIS B1 (620-670 nm) and MODIS B5 (1230-1250 nm). Our EMIT-based analysis explicitly filters out vegetated surfaces, whereas the coarser spatial resolution of MODIS increases the likelihood of mixed pixels containing darker vegetation. As a result, EMIT may more effectively isolate the reflectance of bare soil surfaces, leading to higher inferred albedos. A third possible difference is the measured quantity: EMIT

measures directional reflectance of a surface for the solar-observer geometry at the time of the overflight (Thompson et al., 2024b), while the MODIS product computes a total bihemispherical albedo using an assumed form for the bidirectional reflectance distribution function (Strahler et al., 1999). The two are equal only in the theoretical case of a true Lambertian surfaces, so systematic discrepancies are expected. Although MODIS effectively captures day-to-day variations in land surface albedo driven by snow and vegetation, EMIT is better suited for studies focused on intrinsic bare surface surface properties and capturing bare ground sediments for Earth system models, as its spectral resolution enables filtering and isolation of bare ground reflectance from mixed and transient surface cover.

An active area of research is the mineral fractionation during entrainment of dust from the surface into the atmosphere and effect on its radiative and biogeochemical properties, and our work highlights the continued importance of these questions. Single-scattering albedo values calculated from EMIT reflectance have lower albedo (~0.5 vs ~0.8) and steeper slope toward blue wavelengths relative to the direct lab measurements of sediment by Di Biagio et al. (2019). One cause of lowered albedo may be the spatial scales of measurement: EMIT observations capture average reflectance of a 60 x 60 m pixel, which may include darkening effects of surface roughness, shadowing, and vegetation. More fundamentally, smaller grains produce brighter reflectances (Hunt and Salisbury 1970). EMIT reflectance records the coarser-grained bulk surface sediment composition, whereas laboratory measurements record fine-grained suspended, fractionated samples that may be more representative of dust in Earth atmosphere.

However, differences in spectral shape between EMIT bare surface and lab data of suspended sediments are less straightforward to explain. Even when spectra are normalized at 950 nm to isolate shape (Figure 13), EMIT-derived SSA exhibits a steeper visible-wavelength slope and lower SSA at blue wavelengths. This result is

opposite to expectations based on prior laboratory and modeling studies, which indicate that fine, fractionated dust aerosols are often enriched in highly absorbing iron oxides such as hematite and goethite, which is expected to lead to enhanced shortwave absorption and lower SSA at blue wavelengths (Di Biagio et al., 2019; Perlwitz et al., 2015; Obiso et al., 2024). However, grain-size effects complicate this interpretation: very fine particles can exhibit reduced relative absorption strengths (band depths) due to shorter effective optical path lengths within absorbing grains and increased multiple scattering at grain boundaries (Hunt and Salisbury, 1970; Hapke, 1981). Thus, one possible explanation is that bulk surface sediments dominated by coarser grains could appear more absorbing at short wavelengths than the extremely fine, fractionated dust aerosols derived from them. A second potential explanation is residual vegetation is some EMIT pixels. Although we apply vegetation filtering, small fractions of vegetation cover (<∼20%) may remain undetected and could contribute to EMIT data's steeper slope due to vegetation's visible-wavelength absorptions. A third explanation may relate to radiometric assumptions in atmospheric removal and SSA inversion. EMIT measures radiance from combined atmosphere and a surface in hemispherical–directional reflectance (HDRF); in addition to direct sunlight, there is additional downwelling radiation from hemispherically integrated scattering. Hapke inversion to SSA assumes bidirectional reflectance factor (BRF) (§2.3.1). These derived quantities are typically 1:1 (e.g., Figure 4 in Schaepman-Strub et al., 2006). However, if scattering due to atmospheric aerosols is overestimated in some cases, blue wavelengths might be lower in EMIT-derived reflectances that are converted to SSA. Our report of SSA magnitude and shape differences between dust and surface spectra highlights the need for future work to better quantify differences between bulk surface reflectance and the optical properties of the liberated dust fraction via in situ comparisons of field and remote sensing-derived spectra and comparisons of surface and dust spectra (Section 4.3).

4.3 Needed future work

Wavelength-dependent surface reflectance measured from orbit provides a robust first-order constraint on the radiative properties of mineral dust, identifies spatially variable dust source surface optical properties, and indicates the spectrally dominant mineral phases present. Current EMIT mineral products report surface mineralogy based on a spectral feature-matching algorithm called Tetracorder that identifies diagnostic absorption features and maps the presence of spectrally distinct mineral classes, using absorption strength to approximate mineral abundance (e.g., EMIT L2B mineral products; Clark et al., 2024). Future work is needed to improve our ability to directly infer precise mineralogical abundances from hyperspectral data. Comparison of hyperspectral data of sediment samples with corresponding quantitative mineralogical, chemical, and grain size data of arid dust source regions (e.g., Gonzalez-Romero et al, 2024a; Gonzalez-Romero et al., 2023b; Panta et al., 2022) will enable the development of physical and empirical models for interpreting quantitative mineralogy from remote sensing data.

A second challenge lies in bridging the gap between surface sediment properties and the optical properties of dust actually entrained into the atmosphere. EMIT SSA values for dust source regions span a far greater range than SSA from existing laboratory samples, pointing to the importance of remote sensing data in accurately capturing dust and dust source local and regional variability for improvement of Earth system models. EMIT data capture real spatial heterogeneity in surface sediment composition of dust source regions, underscoring the importance of developing the additional steps to account for fractionation so as to incorporate airborne dust of diverse mineralogy into Earth system models. EMIT measures bulk surface reflectance, whereas atmospheric dust results from size- and mineralogy-dependent fractionation during emission and transport. Consequently, surface reflectance cannot be directly equated with atmospheric dust SSA without accounting for these fractionation mechanisms. Existing literature suggests these

processes selectively enrich fine, iron-bearing clay minerals relative to coarser, weakly absorbing components such as quartz (García-Pando et al., 2016; Di Biagio et al., 2019), fundamentally altering dust single-scattering albedo and spectral shape, though our data indicate the scenario may be more complex as our source regions have stronger Fe oxide absorptions than found in the laboratory data of suspended dust. EMIT-based surface characterization in combination with observational constraints such as regional aerosol optical property retrievals from satellite instruments (e.g., Kalashnikova et al., 2006) offers a pathway to quantify how source-specific surface properties translate into atmospheric dust radiative behavior. Given that surface reflectance already serves as a strong proxy for radiative properties of mineral dust, this study identifies the dust source locations that both represent "typical" mineral dust optical properties and span the range of variability. Targeted laboratory and field measurements from the source regions identified in this study would further constrain these links, enabling more physically consistent representations of dust optical properties in Earth system models.

5. Conclusions

EMIT provides hyperspectral observations of dust source regions worldwide, enabling characterization of dust source region composition, radiative properties, and spatial and temporal heterogeneity. The semi-automated workflow developed and released here provides a means of extracting high spectral resolution bare surface reflectance spectra. Our filtering routine effectively removes spectra impacted by clouds, surface water, snow, and vegetation, resulting in a high-quality reflectance dataset. We provide a dataset of dust source region surface hyperspectral properties that integrates multiple temporal EMIT observations at each sampling coordinate, yielding a consistent representation of typical bare land surface properties different from existing MODIS albedo products. This approach

enables characterization of surface reflectance across key dust source regions and forms the foundation for global analysis of mineralogical and radiative variability.

Except for flooded regions, we found that dust source region spectral properties were largely stable over time; within an example 24,878 km^2 region in the Bodélé depression, spatial variability was greater than temporal variability by 3-10x. We identify two spatially dominant mineral dust source region spectral types: (1) Saharan dust regions are bright (VIS albedo 0.40; NIR albedo 0.60) and iron oxide- and kaolinite-bearing and (2) central/east Asia dust regions are darker (VIS albedo 0.25; NIR albedo 0.35) with montmorillonite or illite as the spectrally dominant mineral, possibly with intermixed carbonate. We also identify region-specific endmember and outlier spectra on each continent, indicating unique mineralogies important in dust that contributes to regional to local biogeochemical processes. On average, our dry, bare sediment surface EMIT VIS albedo is 6-10% brighter than MODIS VIS albedo data products. Relative to the single scattering albedo of suspended dust samples in lab, EMIT dust source spectra have 40% lower VIS albedo and 7% lower NIR single-scattering albedo than laboratory data. While the darkening is expected due to coarser grain size and lack of roughness/shadowing in laboratory data, the difference in spectral slope is the opposite to that intuitively expected by previously reported fractionation during transport that enhances iron oxide in the airborne dust relative to source sediments. Direct measurements of in situ surface spectra and suspended dust spectra are needed with independent characterization of mineralogy and grain size of each material. Future work refining single-scattering albedo estimates, grain size, and quantitative mineralogical mapping from hyperspectral data will further improve our ability to represent dust composition and its climatic and biogeochemical impacts on the Earth system.

Acknowledgements

We are grateful for support for this work from the NASA FINESST Graduate Research Fellowship Program award to A.M.K. and B. L.E. (award number

80NSSC22K1456) and an EMIT Phase E co-investigator award to B.L.E. Thank you to the EMIT science and operations teams for collection and processing of a superb dataset. A portion of this research was conducted at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (NASA) 80NM0018D0004.

Open Research

All EMIT data used in this paper are publicly available for download through the NASA EarthData system. Reference mineral spectra from the CRISM spectral library used in this work are maintained by NASA's Planetary Data System Geosciences Node at Washington University in St. Louis and are available at https://crismtypespectra.rsl.wustl.edu/. The Python code created for this project for semi-automated data filtering will be available through Github (Keebler et al. 2026a). The spectral and albedo datasets generated from this project will be available the Caltech Research Data Repository (Keebler et al., 2026b). The data and code will be released at the sources above with doi after review and revisions.

Conflicts of Interest

The authors declare there are no conflicts of interest for this manuscript.

References

Börker, J., Hartmann, J., Amann, T., & Romero-Mujalli, G. (2018). Global Unconsolidated Sediments Map Database v1. 0 (shapefile and gridded to 0.5 spatial resolution). Supplement to: Börker, J et al.(2018): Terrestrial Sediments of the Earth: Development of a Global Unconsolidated Sediments Map Database (GUM). Geochemistry, Geophysics, Geosystems, 19 (4), 997, 1024.

Braghiere, R. K., Quaife, T., Black, E., He, L., & Chen, J. M. (2019). Underestimation of global photosynthesis in Earth system models due to

representation of vegetation structure. Global Biogeochemical Cycles, 33(11), 1358-1369.

Braghiere, R. K., Wang, Y., Gagné-Landmann, A., Brodrick, P. G., Bloom, A. A., Norton, A. J., ... & Schneider, T. (2023). The importance of hyperspectral soil albedo information for improving Earth system model projections. Agu Advances, 4(4), e2023AV000910.

Clark, R. N., & Roush, T. L. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. Journal of Geophysical Research: Solid Earth, 89(B7), 6329-6340.

Clark, R. N., Swayze, G. A., Livo, K. E., Brodrick, P. G., Dobrea, E. N., Vijayarangan, S., ... & Querol, X. (2024). Imaging Spectroscopy: Earth and Planetary Remote Sensing with the PSI Tetracorder and Expert Systems from Rovers to EMIT and Beyond. The Planetary Science Journal, 5(12), 276.

Claquin, T., Schulz, M., & Balkanski, Y. J. (1999). Modeling the mineralogy of atmospheric dust sources. Journal of Geophysical Research: Atmospheres, 104(D18), 22243-22256.

Di Biagio, C., Formenti, P., Balkanski, Y., Caponi, L., Cazaunau, M., Pangui, E., ... & Doussin, J. F. (2019). Complex refractive indices and single-scattering albedo of global dust aerosols in the shortwave spectrum and relationship to size and iron content. Atmospheric Chemistry and Physics, 19(24), 15503-15531.

Pérez García-Pando, C., Miller, R. L., Perlwitz, J. P., Rodríguez, S., & Prospero, J. M. (2016). Predicting the mineral composition of dust aerosols: Insights from elemental composition measured at the Izaña Observatory. Geophysical Research Letters, 43(19), 10-520.

Ginoux, P., Prospero, J. M., Gill, T. E., Hsu, N. C., & Zhao, M. (2012). Global-scale attribution of anthropogenic and natural dust sources and their emission rates based on MODIS Deep Blue aerosol products. Reviews of Geophysics, 50(3).

González-Romero, A., González-Flórez, C., Panta, A., Yus-Díez, J., Reche, C., Córdoba, P., ... & Pérez García-Pando, C. (2023). Variability in sediment particle size, mineralogy, and Fe mode of occurrence across dust-source inland drainage basins: the case of the lower Drâa Valley, Morocco. Atmospheric chemistry and physics, 23(24), 15815-15834.

González-Romero, A., González-Flórez, C., Panta, A., Yus-Díez, J., Córdoba, P., Alastuey, A., ... & Pérez García-Pando, C. (2024). Probing Iceland's dust-emitting sediments: particle size distribution, mineralogy, cohesion, Fe mode of occurrence, and reflectance spectra signatures. Atmospheric chemistry and physics, 24(11), 6883-6910.

González-Romero, A., González-Flórez, C., Panta, A., Yus-Díez, J., Córdoba, P., Alastuey, A., ... & Pérez García-Pando, C. (2024). Characterization of the particle size distribution, mineralogy, and Fe mode of occurrence of dust-emitting sediments from the Mojave Desert, California, USA. Atmospheric chemistry and physics, 24(16), 9155-9176.

Green, R. O., & Thompson, D. R. (2020, September). An Earth science imaging spectroscopy mission: The Earth surface mineral dust source investigation (EMIT). In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium* (pp. 6262-6265). IEEE.

Green, R. (2022). *EMIT L2A Estimated Surface Reflectance and Uncertainty and Masks 60 m V001* [Data set]. NASA Land Processes Distributed Active Archive

Center. https://doi.org/10.5067/EMIT/EMITL2ARFL.001 Date Accessed: 2025-12-22

Hapke, B. (1981). Bidirectional reflectance spectroscopy: 1. Theory. Journal of Geophysical Research: Solid Earth, 86(B4), 3039-3054.

Herich, H., Tritscher, T., Wiacek, A., Gysel, M., Weingartner, E., Lohmann, U., ... & Cziczo, D. J. (2009). Water uptake of clay and desert dust aerosol particles at sub-and supersaturated water vapor conditions. Physical Chemistry Chemical Physics, 11(36), 7804-7809.

Hunt, G.R. & Salisbury, J.W. (1970). Visible and Near-Infrared Spectra of Minerals and Rocks — I. Silicate Minerals. Modern Geology, 1, 283–300.

Jickells, T. D., An, Z. S., Andersen, K. K., Baker, A. R., Bergametti, G., Brooks, N., ... & Torres, R. (2005). Global iron connections between desert dust, ocean biogeochemistry, and climate. science, 308(5718), 67-71.

Journet, E., Balkanski, Y., & Harrison, S. P. (2014). A new data set of soil mineralogy for dust-cycle modeling. Atmospheric Chemistry and Physics, 14(8), 3801-3816.

Kalashnikova, O. V., & Kahn, R. (2006). Ability of multiangle remote sensing observations to identify and distinguish mineral dust types: 2. Sensitivity over dark water. Journal of Geophysical Research: Atmospheres, 111(D11).

Kandler, K., Benker, N., Bundke, U., Cuevas, E., Ebert, M., Knippertz, P., ... & Weinbruch, S. (2007). Chemical composition and complex refractive index of Saharan Mineral Dust at Izaña, Tenerife (Spain) derived by electron microscopy. Atmospheric Environment, 41(37), 8058-8074.

Keebler, A.M., Ehlmann, B.L., Thoompson, D.R., Ginoux, P., Brodrick, P.G., and Green, R.O. (2026a). Spectral Variability of Earth's Desert Dust Source Regions Sampled from EMIT Reflectance Data [Dataset].

Keebler, A.M., Ehlmann, B.L., Thoompson, D.R., Ginoux, P., Brodrick, P.G., and Green, R.O. (2026b). emit_data_bareground_filtering.py [Code].

Kok, J. F., Adebiyi, A. A., Albani, S., Balkanski, Y., Checa-Garcia, R., Chin, M., ... & Wan, J. S. (2021). Contribution of the world's main dust source regions to the global cycle of desert dust. *Atmospheric Chemistry and Physics*, *21*(10), 8169-8193.

Li, L., Mahowald, N., Balkanski, Y., Connelly, D., Ginoux, P., Ageitos, M. G., Hamilton, D., Kalashnikova, O., Klose, M., Miller, R. L., Obiso, V., Paynter, D. and Perez Garcia-Pando, C.: Large contribution of hematite and geothite to uncertainty in dust direct radiative forcing, Atmos. Chem. Phys., 2020.

Li, L., Mahowald, N. M., Miller, R. L., Pérez García-Pando, C., Klose, M., Hamilton, D. S., ... & Thompson, D. R. (2021). Quantifying the range of the dust direct radiative effect due to source mineralogy uncertainty. *Atmospheric chemistry and physics*, *21*(5), 3973-4005.

Li, L., Mahowald, N. M., Miller, R. L., Pérez García-Pando, C., Goncalves Ageitos, M., Ginoux, P., Obiso, V., Song, Qianqian, Brodrick, P.G., Thompson, D. R., Clark, R.N., Okin, G.S., Ehlmann, B.L., Zhoue, B., Kalashnikova, O., & Green, R.O. (2026). Spaceborne mineral mapping reduces dust's shortwave radiative impact uncertainty. *Nature Geoscience.*

Moskowitz, B. M., Reynolds, R. L., Goldstein, H. L., Berquó, T. S., Kokaly, R. F., & Bristow, C. S. (2016). Iron oxide minerals in dust-source sediments from the Bodélé Depression, Chad: Implications for radiative properties and Fe bioavailability of dust plumes from the Sahara. *Aeolian Research*, *22*, 93-106.

Moreno, T., Querol, X., Castillo, S., Alastuey, A., Cuevas, E., Herrmann, L., ... & Gibbons, W. (2006). Geochemical variations in aeolian mineral particles from the Sahara–Sahel Dust Corridor. Chemosphere, 65(2), 261-270.

Müller, D., Weinzierl, B., Petzold, A., Kandler, K., Ansmann, A., Müller, T., ... & Knippertz, P. (2010). Mineral dust observed with AERONET Sun photometer, Raman lidar, and in situ instruments during SAMUM 2006: Shape-independent particle properties. Journal of Geophysical Research: Atmospheres, 115(D7).

Obiso, V., Gonçalves Ageitos, M., Pérez García-Pando, C., Perlwitz, J. P., Schuster, G. L., Bauer, S. E., ... & Miller, R. L. (2024). Observationally constrained regional variations of shortwave absorption by iron oxides emphasize the cooling effect of dust. *Atmospheric Chemistry and Physics*, *24*(9), 5337-5367.

Ochoa, F., Brodrick, P. G., Okin, G. S., Ben-Dor, E., Meyer, T., Thompson, D. R., & Green, R. O. (2025). Soil and vegetation cover estimation for global imaging spectroscopy using spectral mixture analysis. *Remote Sensing of Environment*, *324*, 114746.

Panta, A., Kandler, K., Alastuey, A., González-Flórez, C., González-Romero, A., Klose, M., ... & Pérez García-Pando, C. (2022). Insights into the single particle composition, size, mixing state and aspect ratio of freshly emitted mineral dust from field measurements in the Moroccan Sahara using electron microscopy. *Atmospheric Chemistry and Physics Discussions*, *2022*, 1-40.

Perlwitz, J., & Miller, R. L. (2010). Cloud cover increase with increasing aerosol absorptivity: A counterexample to the conventional semidirect aerosol effect. *Journal of Geophysical Research: Atmospheres*, *115*(D8).

Perlwitz, J. P., Pérez García-Pando, C., & Miller, R. L. (2015). Predicting the mineral composition of dust aerosols–Part 1: Representing key processes. *Atmospheric Chemistry and Physics*, *15*(20), 11593-11627.

Pu, B., Ginoux, P., Guo, H., Hsu, N. C., Kimball, J., Marticorena, B., ... & Zhao, M. (2020). Retrieving the global distribution of the threshold of wind erosion from satellite data and implementing it into the Geophysical Fluid Dynamics Laboratory land–atmosphere model (GFDL AM4. 0/LM4. 0). Atmospheric Chemistry and Physics, 20(1), 55-81.

Radhi, M., Box, M. A., Box, G. P., Mitchell, R. M., Cohen, D. D., Stelcer, E., & Keywood, M. D. (2010). Size-resolved mass and chemical properties of dust aerosols from Australia's Lake Eyre Basin. Atmospheric Environment, 44(29), 3519-3528.

Scanza, R. A., Mahowald, N., Ghan, S., Zender, C. S., Kok, J. F., Liu, X., ... & Albani, S. (2015). Modeling dust as component minerals in the Community Atmosphere Model: development of framework and impact on radiative forcing. Atmospheric Chemistry and Physics, 15(1), 537-561.

Schaepman-Strub, G., Schaepman, M. E., Painter, T. H., Dangel, S., & Martonchik, J. V. (2006). Reflectance quantities in optical remote sensing—Definitions and case studies. *Remote sensing of environment*, *103*(1), 27-42.

Shao, L., Li, W., Xiao, Z., & Sun, Z. (2008). The mineralogy and possible sources of spring dust particles over Beijing. Advances in Atmospheric Sciences, 25(3), 395-403.

Shao, Y., Wyrwoll, K. H., Chappell, A., Huang, J., Lin, Z., McTainsh, G. H., ... & Yoon, S. (2011). Dust cycle: An emerging core theme in Earth system science. *Aeolian Research*, *2*(4), 181-204.

Shi, Z., Krom, M. D., Jickells, T. D., Bonneville, S., Carslaw, K. S., Mihalopoulos, N., ... & Benning, L. G. (2012). Impacts on iron solubility in the mineral dust by processes in the source region and the atmosphere: A review. Aeolian Research, 5, 21-42.

Song, Q., Ginoux, P., Gonçalves Ageitos, M., Miller, R. L., Obiso, V., & Pérez García-Pando, C. (2024). Modeling impacts of dust mineralogy on fast climate response. *Atmospheric Chemistry and Physics*, *24*(12), 7421-7446.

Strahler, A. H., Muller, J., Lucht, W., Schaaf, C., Tsang, T., Gao, F., ... & Barnsley, M. J. (1999). MODIS BRDF/albedo product: algorithm theoretical basis document version 5.0. MODIS documentation, 23(4), 42-47.

Thompson, D. R., Boardman, J. W., Eastwood, M. L., & Green, R. O. (2017). A large airborne survey of Earth's visible-infrared spectral dimensionality. Optics express, 25(8), 9186-9195.

Thompson, D. R., Green, R. O., Bradley, C., Brodrick, P. G., Mahowald, N., Dor, E. B., ... & Zandbergen, S. (2024a). On-orbit calibration and performance of the EMIT imaging spectrometer. *Remote Sensing of Environment*, *303*, 113986.

Thompson, D. R., Brodrick, P. G., Green, R. O., Kalashnikova, O., Lundeen, S., Okin, G., ... & Painter, T. (2024b). EMIT L2A Algorithm: Surface Reflectance and Scene Content Masks Theoretical Basis. Jet Propulsion Laboratory, California Institute of Technology: Pasadena, CA, USA, 23.

Viviano-Beck, C. E. (et al.), MRO CRISM Type Spectra Library, NASA Planetary Data System, 2015. https://crismtypespectra.rsl.wustl.edu

Wiscombe, W. J., & Warren, S. G. (1980). A model for the spectral albedo of snow. II: Snow containing atmospheric aerosols. Journal of Atmospheric Sciences, 37(12), 2734-2745.

Yus-Díez, J., Escribano, J., Pandolfi, M., Alastuey, A., González-Flórez, C., Gonzalez-Romero, A., Gonçalves Ageitos, M., Ivančič, M., Klose, M., Kandler, K., Obiso, V., Panta, A., Reche, C., Rigler, M., Querol, X., and Perez Garćia-Pando, C. (2025): Optical Properties and Shape-Dependent Complex

Refractive Index Retrievals of Freshly Emitted Saharan Dust, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-2571

Zielke, S. A., Bertram, A. K., & Patey, G. N. (2016). Simulations of ice nucleation by kaolinite (001) with rigid and flexible surfaces. The journal of physical chemistry B, 120(8), 1726-1734.

Chapter 3

IN SITU SPECTROSCOPY OF MINERAL DUST-PRODUCING SEDIMENTS AND CONNECTIONS TO LAB-DERIVED QUANTITATIVE MINERALOGY AND IRON CHEMISTRY DATA

Abigail M. Keebler[1], Bethany L. Ehlmann[1,2,3,4], Rebecca N. Greenberger[1], Adolfo Gonzalez-Romero[5,6,7], Xavier Querol[6], Melani Hernandez-Chiriboga[5], Carlos Perez Garcia-Pando[5,7]

1 California Institute of Technology, Pasadena, CA, United States
2 Laboratory for Atmospheric and Space Physics, Boulder, CO, United States
3 Department of Geological Sciences, University of Colorado, Boulder
4 Department of Astrophysical & Planetary Sciences, University of Colorado, Boulder
5 Barcelona Supercomputing Center (BSC), Barcelona, Spain
6 Spanish Research Council, Institute of Environmental Assessment and Water Research (IDAEA-CSIC), Barcelona, Spain
7 Catalan Institute for Research and Advanced Studies (ICREA), Barcelona, Spain

Abstract

Mineral dust composition exerts a strong control on Earth's radiative balance and biogeochemical cycles, yet remains poorly constrained at regional to global scales. Space-based hyperspectral visible to near-infrared (VNIR) observations are a tool for mapping important minerals such as iron oxides, clays, and carbonates, but quantitative relationships between surface spectral properties and composition are only partially constrained. Here, we present and evaluate a new dataset of in situ VNIR spectra collected from mineral dust sources at select field sites, paired with laboratory measurements of grain size distributions, quantitative mineralogy, and iron speciation for each sample to support global spectral mapping. Our samples capture a 60% of the range of spectral variability observed for global dust sources, greatly expanding soil samples available for comparison with remote sensing data, but still pointing to the need for representative endmembers of dark sediments from

central and east Asian sources and Saharan sands. We compare mineral absorption feature strength to XRD-derived mineral abundances to assess the extent to which compositional variability can explain observed spectral properties for clay minerals, carbonates, salts, and iron oxides. Iron oxide absorption strength correlates most strongly with abundance, exhibiting a linear relationship with the fraction of total iron present as crystalline iron oxides across all sediment types as well as with bulk iron oxide weight percent in fine-grained, clay-rich samples. In contrast, sandy sediments exhibit strong iron oxide absorptions even at relatively low iron oxide abundances, indicating a distinct compositional-spectral relationship. Forward Hapke radiative transfer modeling reproduces these trends and demonstrates the role of bright mineral matrices in enhancing iron oxide spectral signatures for large grains. These results provide a framework for estimating iron oxide abundance in dust source sediments from VNIR spectroscopy, improving constraints on dust radiative effects in the climate system.

1. Introduction

Mineral dust is a critical component of the Earth system, influencing radiative forcing, cloud microphysics, atmospheric chemistry, cryosphere processes, ocean biogeochemistry, ecosystem nutrient supply, and human health (e.g., Mahowald et al., 2014; Kok et al., 2023; Perez Garcia-Pando et al., 2014). The effects of dust on climate are strongly dependent on its mineralogical composition, because individual mineral phases exhibit distinct optical, chemical, and ice-nucleating properties (Mahowald et al., 2014; Kok et al., 2023). In particular, iron oxide minerals such as hematite and goethite exert a disproportionate influence on the shortwave radiative properties of dust aerosols owing to their strong absorption of solar radiation (Sokolik and Toon, 1999; Lafon et al., 2006; Di Biagio et al., 2019; Li et al., 2021). Iron mineralogy additionally governs the bioavailability of micronutrients delivered to marine and terrestrial ecosystems (Jickells et al., 2005; Shoenfelt et al., 2017), while variations in iron speciation and mineral associations modulate atmospheric reactivity and photochemistry (Formenti et al., 2014). Consequently, constraining the

mineralogical and geochemical variability of dust source regions remains essential for improving understanding of dust-climate interactions and reducing uncertainty in Earth system models.

Global imaging spectroscopy datasets enable the characterization of dust source mineralogy at regional to global scales. NASA's Earth surface Mineral dust source InvesTigation (EMIT) mission, launched aboard the International Space Station in 2022, is acquiring global visible-to-shortwave infrared (VSWIR) hyperspectral observations intended to map the surface mineralogy of arid dust source regions (Green et al., 2020; Thompson et al., 2024). EMIT measurements have a spectral range of 400-2500 nm and resolution of 7.5 nm, ideal for identifying absorption features associated with iron oxides, clay minerals, carbonates, sulfates, and other mineral phases (Thompson et al., 2024). However, retrieving quantitative mineralogy from hyperspectral reflectance data remains challenging because natural soils and sediments are compositionally heterogeneous mixtures whose spectra are influenced not only by mineral abundance, but also by grain size, particle coatings, crystallinity, moisture, viewing geometry, and nonlinear spectral mixing effects (Clark, 1990; Okin and Painter, 2004). Prior approaches have included spectral feature fitting and continuum-removal techniques (Clark and Roush, 1984; Swayze et al., 2014), nonlinear spectral unmixing and radiative transfer modeling (Hapke, 2012; Mustard and Pieters, 1989), and increasingly, empirical and machine-learning methods trained on laboratory or field spectral datasets (e.g., Feely and Christensen, 1999; Verpoorter et al., 2014; Mancoridis et al., 2025). Nevertheless, robust quantitative retrievals remain difficult because diagnostic absorption features often overlap, many spectrally important constituents occur in low abundance, and spectral responses vary systematically with physical sediment properties independent of composition.

To address these limitations, the FRontiers in dust minerAloGical composition and its Effects upon climaTe (FRAGMENT) project was established to develop an integrated understanding of the spectral, mineralogical, and geochemical variability of global dust source sediments and emitted dust sediment fraction through

coordinated field sampling and laboratory characterization across diverse arid environments. The project combines in situ spectroscopy, laboratory hyperspectral measurements, quantitative mineralogical analyses, grain-size characterization, and iron speciation measurements from multiple dust-producing regions spanning contrasting geologic settings and climates (Perez Garcia-Pando, 2018). In this study, we advance the current state-of-the-art by leveraging the FRAGMENT dataset to investigate quantitative relationships between hyperspectral reflectance, mineralogical composition, and iron chemistry in natural dust source sediments. We combine datasets collected during four field campaigns conducted across arid dust source regions in different parts of the world, thereby capturing a broad range of mineralogical and spectral variability. We first assess the spectral variability of the combined dataset relative to that observed in prior global surveys of dust source spectral diversity. We then evaluate how variations in mineral abundance, iron mineralogy, and sediment properties influence reflectance characteristics and diagnostic spectral features across the visible-to-shortwave infrared range. Finally, we discuss the implications of these relationships for interpreting EMIT observations and for improving quantitative mineralogical retrievals from orbital hyperspectral remote sensing data.

2. Methods

To build a dataset of diverse dust source sediment spectra and corresponding grain size distribution, quantitative mineralogy, and iron chemistry data, we conducted a series of field campaigns in four regions to acquire in situ spectra and collect corresponding samples (Figure 1). Laboratory analysis of samples to quantify grain size distribution and composition were carried out at the Institute of Environmental Assessment and Water Research at the Spanish National Research Council (IDAEA-CSIC) and visible-infrared spectra were acquired by Caltech. The resulting dataset was subsequently compared with previously produced databases of global spectral diversity of dust-producing regions (Keebler et al., submitted).

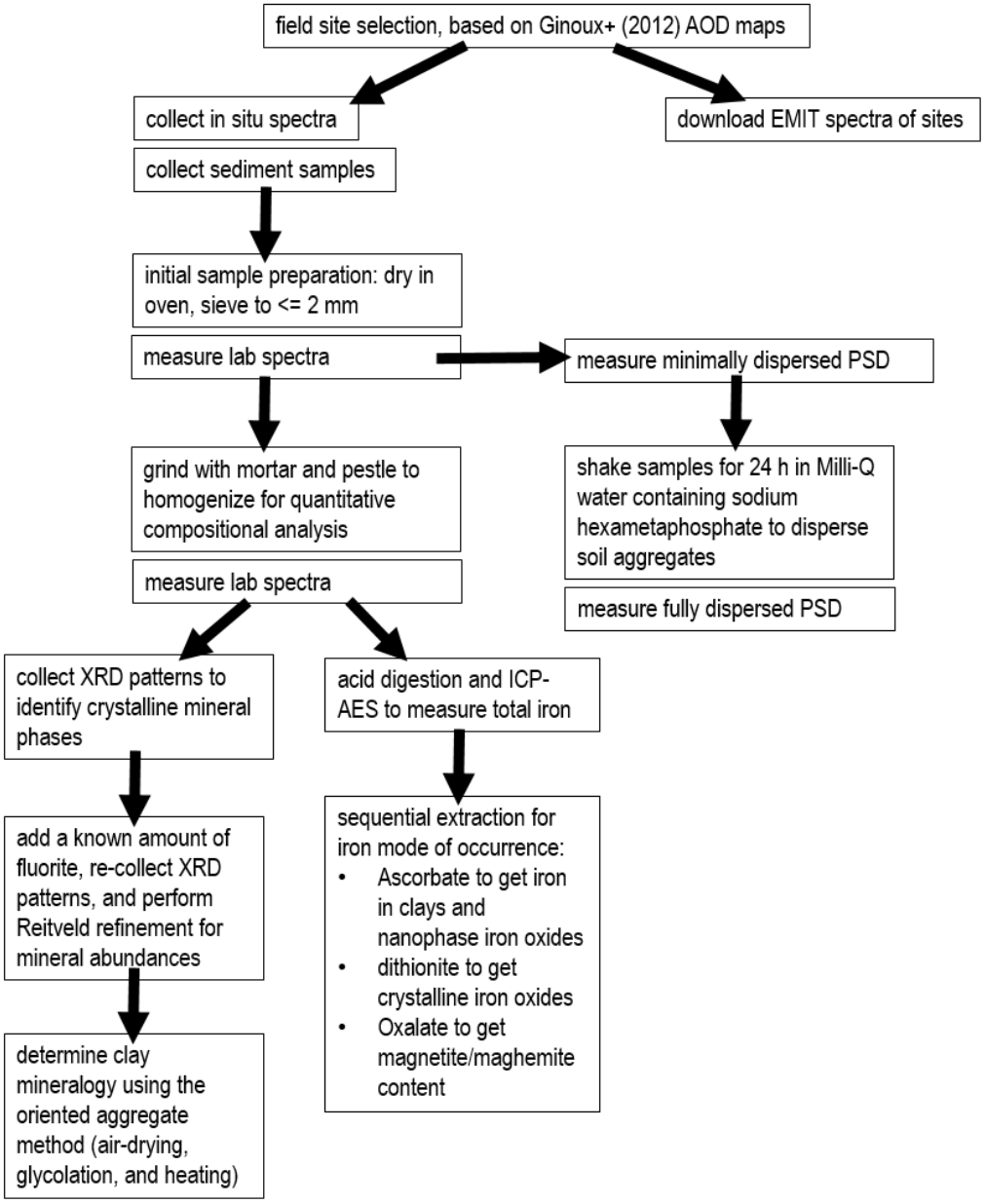

Figure 1. Overview of methods for dataset development, including field campaign planning, sample collection, and laboratory sample prep and analyses.

2.1 Field site selection

Field spectra and sample collection were conducted as a part of and in collaboration with the FRontiers in dust minerAloGical composition and its Effects upon climaTe (FRAGMENT) project (Perez Garcia-Pando, 2018). Field campaigns were

designed to capture the mineralogical and particle size diversity of sediments contributing to dust emission across major source regions with diverse properties in the Sahara, the Middle East, and North America. Campaigns were conducted in eastern Morocco (Summer 2019), the Mojave Desert, California, USA (Spring 2022), and Jordan (Fall 2022), with an additional survey across the southwestern United States (Summer 2023) to expand mineralogical endmember diversity within the dataset.

For each region, sampling targeted sediments representative of active and potential dust-emitting surfaces. Candidate locations were identified using aerosol optical depth imagery to highlight chronically active dust source regions. Within these areas, efforts were made to systematically capture with in-field spectral measurement and sediment sampling the range of surface types present, including fine-grained playas, mud-cracked surfaces, and aeolian sand deposits. Although coarser sediments such as dunes are not direct dust sources, they were included due to their role in saltation and dust liberation processes.

Sampling emphasized both representativeness and diversity. For each region, multiple sediment types were collected to account for variability in surface texture, grain size, and mineralogical composition. Additionally, atypical or visually distinct surfaces (e.g., crusts, coatings, or evaporite-rich deposits) were selectively sampled to ensure broad coverage of mineralogical endmembers. This approach aimed to produce a dataset that captures the range of sediment properties relevant to dust production within each study area (Figure 2).

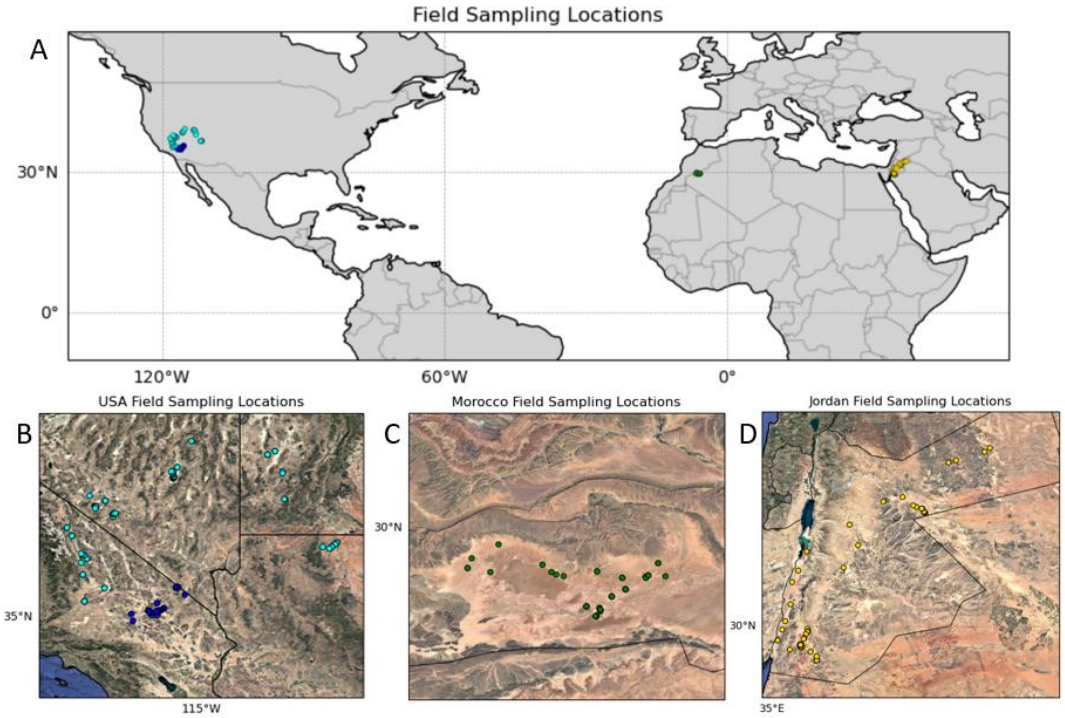

Figure 2. Sampling locations included in this study. (A) global context map; (B) North American field sites from 2022 Mojave Desert and 2023 broader Southwestern USA campaigns; (C) Moroccan field sites from 2019 campaign; (D) Jordan field sites from 2022 campaign.

2.2 Field in-situ spectra and sample collection

In situ surface spectra of dust-emitting surface sediments were collected prior to sampling using an ASD Fieldspec Pro 4 spectrometer with a spectral range of 350-2500nm and a spectral resolution of 1 nm. Measurements were acquired with a contact probe equipped with an internal light source to ensure consistent illumination and viewing geometry of quasi-hemispherical incidence angles from a halogen lamp and 18-degree emission angle across all samples. Use of a contact probe with its internal illumination source meant that there was no influence of the atmosphere's absorptions in attenuating solar radiation. At each site, four measurements were collected within an approximately 15x15 cm area to capture small-scale spatial heterogeneity associated with variations in sediment composition and surface texture.

Following spectra collection, sediments from the same area were sampled using a flat stainless steel shovel to collect the upper ~1cm of material (Figure 3). Collected sediments include crusts, underlying unconsolidated sediments, and sands. In cases where a distinct surface crust differed in texture or appearance from the underlying sediments, additional in situ spectra were collected after removal of the crust to characterize the exposed subsurface (Fig 3, Panel B). For homogeneous surfaces (e.g., sands), repeat measurements following sampling were not performed.

Figure 3. Field sampling. (A) collection of in situ spectra (B) example of sediment post-sample collection. (C) visual diversity of sediment types included in study.

2.3 Laboratory analyses

Sections 2.3.1 and 2.3.3-2.3.5 describe laboratory protocols developed and conducted by the FRAGMENT team at the CSIC laboratories to produce the XRD mineralogy, iron speciation, and grain size distribution datasets used in this work. They are described in brief here; for more detail, refer to Gonzalez-Romero et al. (2023, 2024a, 2024b).

2.3.1 Sample preparation

Once transported back to the lab, samples were dried in an oven at 40-50 °C for 48 hours to remove any residual moisture and sieved to below 2 mm to remove any large gravel and organic material. Subsamples were then allocated for the analyses described below.

2.3.2 Lab spectroscopy

Additional laboratory spectra were collected for all samples with the ASD contract probe, just as in the field. Measurements were conducted on dried and sieved material (with >2mm coarse fractions removed) and on samples that were ground with a mortar and pestle to produce dry, homogeneous powders with minimized textural and grain-size effects.

2.3.3 Grain size distribution

Grain size distributions were measured under both dry (minimally dispersed) and wet (fully dispersed) conditions using a Malvern Mastersizer 2000, following Gonzalez-Romero et al. (2023). The Malvern Mastersizer is a laser diffractometer which measures particle size distribution by analyzing the pattern of light scattering produced when a laser beam passes through a dispersed sample of particles. Minimally dispersed grain size distributions were obtained by direct measurement of dried, sieved sediment samples, preserving natural aggregate structures. Fully dispersed grain size distributions were measured after samples were agitated for 24

hours in Milli-Q water with sodium hexametaphosphate to disaggregate particles and quantify individual grain sizes.

2.3.4 XRD and Rietveld refinement for clays

Quantitative mineralogical composition was determined using X-ray diffraction (XRD) to identify mineral phases, followed by Rietveld refinement to quantify phase abundances (Rietveld, 1969; Gonzalez-Romero et al., 2023). Clay mineralogy was characterized using the oriented aggregate method, in which samples were air-dried, glycolated, and heated to induce diagnostic changes in clay structures and interlayer spacing for phase identification.

2.3.5 Wet chemistry for iron speciation

X-ray diffraction (XRD) is not well suited for quantifying iron oxides due to their typically low weight fractions and the presence of poorly crystalline nanophases; consequently, additional constraints on iron speciation were obtained using wet chemical analyses. Total iron content was measured by two-step acid digestion following Querol et al. (1993). The distribution of iron among mineral phases was then determined using a sequential extraction procedure to quantify iron associated with (1) crystalline FeIII oxides, (2) readily exchangeable iron adsorbed into or incorporated within clay minerals and nanophase Fe oxides, and (3) magnetite. Then, (4) structural iron in silicates was calculated by difference, defined as the total iron content minus the sum of these extracted fractions.

2.3.6 Optical microscopy

Select samples were imaged with an Olympus SZX10 optical microscope to examine grain morphology, aggregate textures, and mineralogical associations as a function of grain size.

2.3.7 Identification of sediment types and variability

Principal component analysis (PCA) was applied to the combined quantitative XRD mineralogical and Fe-speciation dataset to identify the dominant sources of variability and explore relationships among samples, including compositional similarities and differences between sediments from the different regions. PCA transforms the original correlated variables into a set of orthogonal principal components that successively explain decreasing proportions of the total variance.

2.4 EMIT spectra collection

For each field sampling coordinate, we retrieved all EMIT L2a reflectance observations acquired between January 2023 and January 2026. Spectra are quality-filtered following the approach of Keebler et al. (submitted) to remove observations affected by clouds, snow, surface water (e.g., flooding), vegetation, and artifacts and noise introduced by the atmospheric correction algorithm. The remaining high-quality spectra were then averaged to produce a single representative mean spectrum for each sampled coordinate.

2.5 Spectral data analysis

Spectral comparisons are conducted between field, laboratory, and satellite observations using both absolute and normalized reflectance, as well as spectral ratio techniques, to evaluate systematic differences across measurement scales. In addition to direct spectral comparisons, derived parameters such as absorption band depths and albedos were analyzed to further assess discrepancies between datasets.

To assess spectral variability and identify representative endmembers and outliers, principal component analysis (PCA) was applied to the full set of in situ spectra. The resulting principal component structure was compared to that of the global EMIT spectral dataset (Keebler et al., submitted) to evaluate how the variability captured in this study relates to observed global spectral diversity of dust source regions. We identify which sediment spectral types observed across Earth's deserts

are well-represented by this study and which sediment spectral types are not represented in this dataset.

To further quantify mineralogical and compositional controls on spectral behavior, continuum removal was applied to isolate absorption features, and spectral parameters like broadband albedo and mineral absorption strengths are calculated for key diagnostic features associated with identified minerals. Absorption strengths are quantified as band depth relative to the continuum according to Clark and Roush (1984). Both single-wavelength albedos and broadband albedos (taken as average reflectance over a specified wavelength range) were calculated, as well as visible-range spectral slopes. Relationships between spectral properties and independently derived mineralogical and geochemical data were explored using correlation matrices and scatterplots, enabling semi-quantitative assessment of the links between composition and spectral properties.

2.6 Hapke Radiative Transfer Modeling

To interpret the physical basis of the observed spectral behavior, we employ a forward radiative transfer model based on the Hapke radiative transfer model to simulate the reflectance of intimate mineral mixtures representative of our samples. Model calculations follow the formulation of Hapke (1981), following the implementation described in equations 1b-12 in Laporte et al. (2017). Optical constants were compiled from published datasets (Longtin et al., 1988) for minerals commonly present in our samples, including quartz, hematite, and kaolinite.

The primary objective of these simulations is to isolate the influence of grain size in weakly absorbing phases (e.g., quartz) on the apparent strength of diagnostic absorption features in strongly absorbing phases, particularly iron oxides such as hematite. Reflectance spectra were modeled for mixtures with mineralogical compositions and grain sizes analogous to those observed in the imaged samples described in Section 2.3.6.

3. Results

3.1 In situ mineralogy of dust source regions

The sampled sediments consist predominantly of mixtures of quartz, feldspars, clay minerals, carbonate minerals, and iron oxide phases (Figure 4). Clear regional differences in bulk mineralogy are observed. The dataset includes both fine-grained surface materials (e.g., crusts, pavements, and dried muds) and sandy sediments, which are expected to exhibit distinct mineralogical signatures. Sandy samples are typically enriched in quartz and depleted in all other mineral groups relative to crusts, consistent with textural sorting effects, except in the Mojave samples (Figure 5).

Samples from Morocco and Jordan are strongly quartz-dominated and exhibit comparatively low feldspar abundances. In contrast, North American samples (Mojave Desert and broader southwestern United States) display relatively higher feldspar contributions, particularly albite, alongside quartz (Figure 5; Figure 6). Regional variations are also evident in clay mineral assemblages (Figure 4). Kaolinite is most abundant in Jordan and Morocco, as well as in select sites in the southwestern United States, but is rare in Mojave samples. Illite/muscovite is prevalent in the United States and Morocco, but is largely absent in Jordan. Smectite is present in samples from the Mojave and broader southwestern United States but is not a significant component of the Jordanian or Moroccan samples. Carbonate occurs in minor proportions in all regions; carbonate mineralogy is dominated by calcite across all regions, with dolomite occurring in minor proportions. Evaporite minerals, including halite, are present sporadically across all field areas. Samples from the United States additionally contain more diverse evaporite assemblages, including trona, thenardite, and burkeite.

Samples exhibit organization in principal component space according to geographic source region (colors) (Figure 6). A total of 20 principal components

were required to explain 99.9% of the variance in the dataset. Along PC1, samples from the Mojave Desert and other U.S. sites are separated from samples collected in Jordan and Morocco. The U.S. samples form a continuum extending in the direction of increasing illite/montmorillonite and pargasite/feldspar loadings. In contrast, a subset of the Jordanian and Moroccan samples, composed primarily of sand samples, projects strongly toward the quartz loading vector, consistent with their quartz-rich composition. The remaining Jordanian and Moroccan samples cluster nearer the kaolinite, calcite, and FeD loading vectors, indicating greater contributions from clay minerals, carbonates, and crystalline iron oxides. Variation along PC2 is primarily associated with a contrast between samples characterized by elevated FeD values and those enriched in evaporite minerals. Loading vectors for other iron species (FeS: structural iron; FeA: iron adsorbed into clays and possible amorphous iron oxides; FeM: iron in magnetite and maghemite) align with illite- and montmorillonite-rich samples from North America. Collectively, the PCA reveals distinct mineralogical assemblage types and highlights the primary compositional gradients that differentiate the study sites.

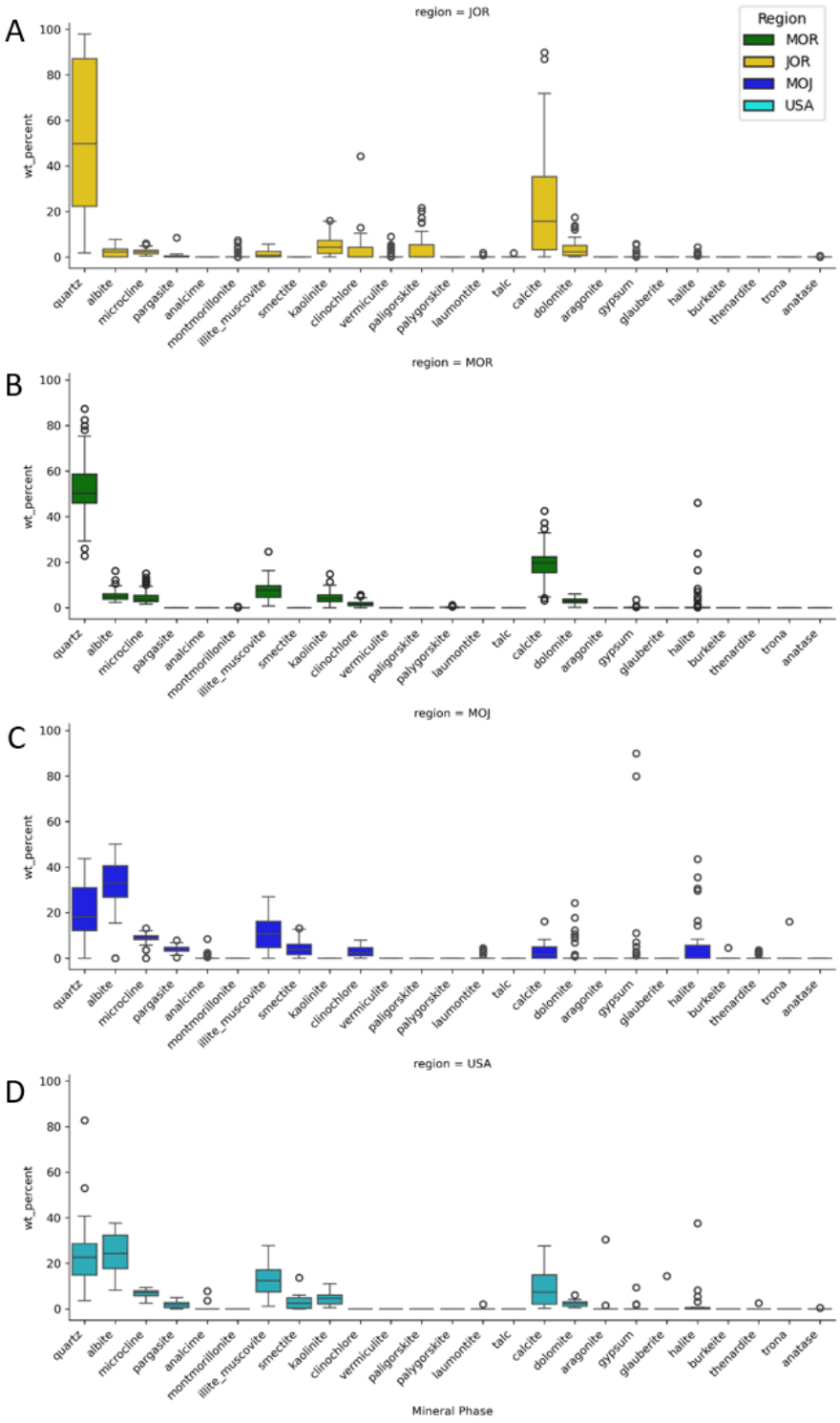

Figure 4. Box-and-whisker plots showing mineralogical distributions of sediments in study, separated by source region: (A) Jordan, (B) Morocco, (C) Mojave Desert, CA, and (D) greater southwestern USA.

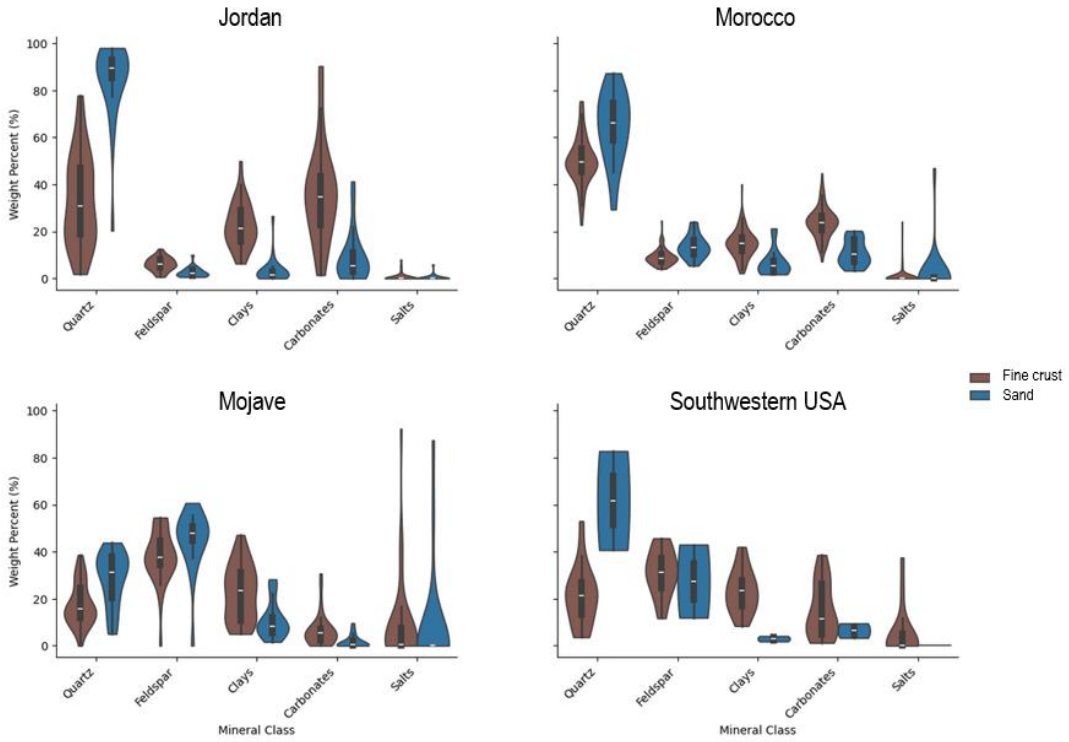

Figure 5. Violin plots showing mineral family abundance distributions, categorized by sediment type (fine-grained sediment crusts vs. sandy samples) and separated by source region.

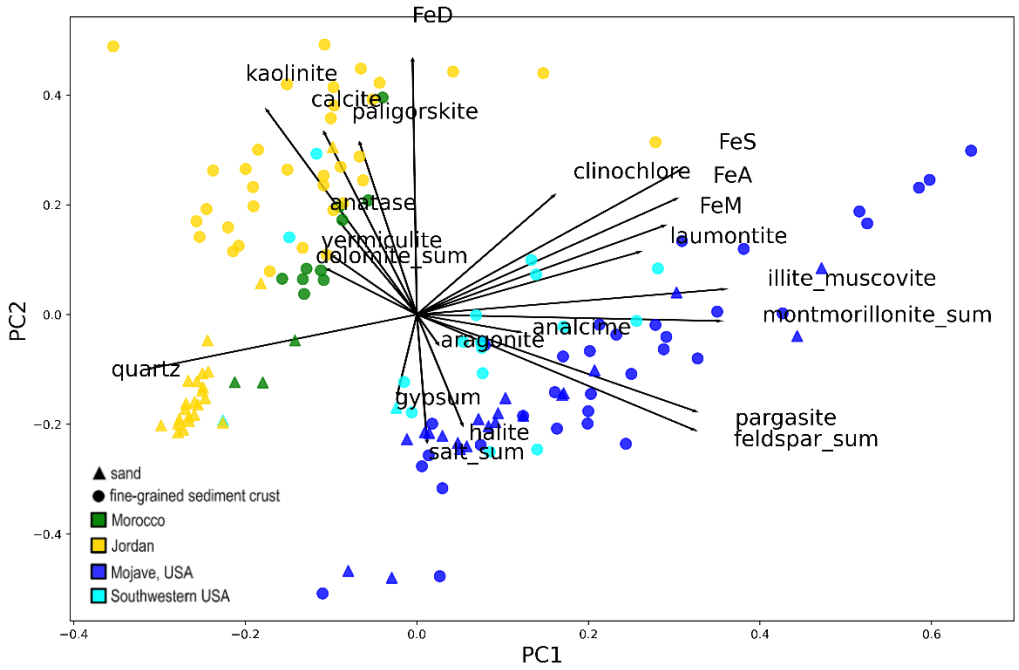

Figure 6. Principal component analysis of XRD-derived mineral abundances and iron speciation results. Samples are projected onto the first two principal components, together with loading vectors indicating the variables that contribute most strongly to each component. Symbols are colored by field site marker shape indicating fine-grained sediment crusts (circle) vs. sandy samples (triangle).

3.2 Spectral consistency across measurement scales: orbital to in situ

In situ visible/shortwave-infrared observations show broadly consistent spectral shapes and absorption features when compared with orbital EMIT observations of the region but also systematic differences that vary with wavelength region, surface texture, and scale of observation. EMIT data exhibit several known artifacts, including a small feature at 2108 nm and spikes at the edges of the major water absorption bands at 1400 and 1900nm, which are not considered further here (Keebler et al., in review).

Compositionally-driven differences between in situ and EMIT spectra are most apparent in both the visible (400-600 nm) and shortwave infrared (2300-2450 nm), where electronic absorptions due to iron-bearing minerals and vibrational absorptions from hydration and hydroxylation are expressed. Additionally, EMIT spectra often have steeper VIS and shortwave IR slope than in situ spectra (Figure 7). As expected, discrepancies in overall reflectance and band strength are observed for samples that were collected from sediment surfaces which are heterogeneous at the 60 m/pixel scale of EMIT observations. In these cases, EMIT spectra reflect a compositional and textural average of multiple surface types, whereas field spectra capture endmember-like conditions.

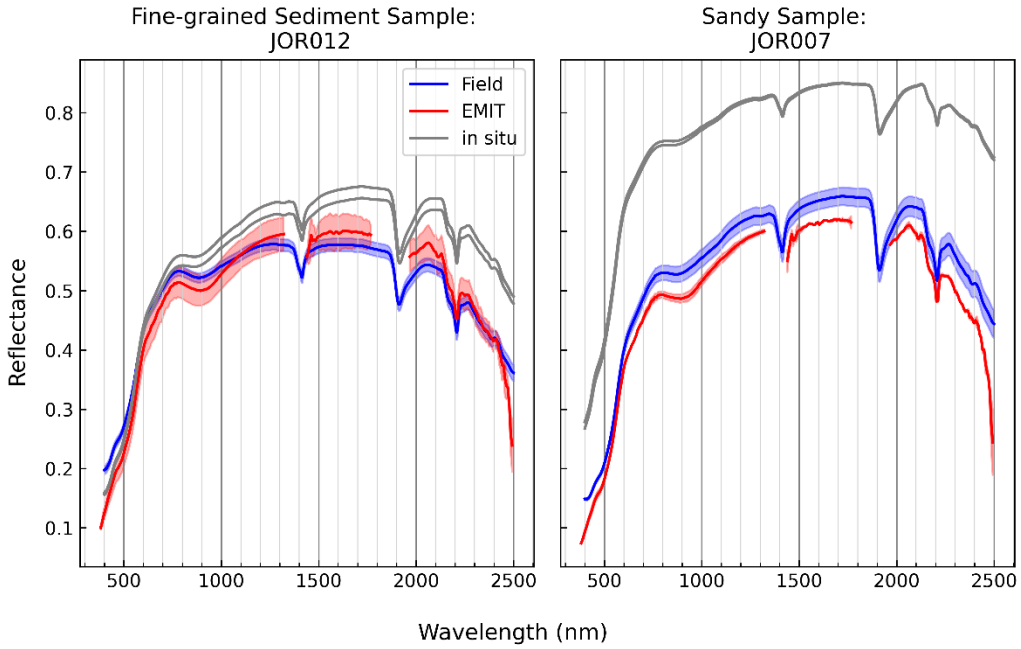

Figure 7. Comparison of EMIT vs in situ spectra and laboratory spectra for two typical surface types. JOR012 is a fine-grained sediment surface exhibiting mudcracks, typical of playa dust emission surfaces (location: 29.7112771N ,35.3820280E). JOR007 is a homogeneous sand-covered surface (location: 29.7374040N, 35.3861166E).

In situ spectra of undisturbed sediment surfaces and laboratory spectra of ground, homogenized samples exhibit broadly similar first-order spectral shapes and consistent absorption types and positions; however, laboratory measurements

generally show higher overall reflectance. This is consistent with reduced grain size and more uniform sample surface under laboratory conditions. Additional differences may arise where samples are spatially heterogeneous at fine scales or contain surface coatings or layered textures.

Grinding and homogenization for XRD and geochemical analyses reduces much of the grain-size- and texture-driven spectral variability. As a result, inter-sample spectral differences in the lab spectra dataset more directly reflect variations in mineralogy and mineral abundance. This effect is illustrated in Figure 8, where a single sediment sample was mechanically separated into grain-size fractions (1-2 mm, 500-1000 μm, 250-500 μm, 80-250 μm, 63-80 μm, 40-63 μm, 20-40 μm, and <20 μm). Each fraction was spectrally measured prior to further homogenization (Figure 8A), and subsequently re-measured following mortar-and-pestle grinding (Figure 8B). In panel A, coarse and fine fractions exhibit darker and brighter reflectances, respectively. The lack of differences in absorptions or spectral slope suggests that spectral differences among grain-size fractions is predominantly controlled by grain size, surface texture and aggregate structure rather than compositional differences, except potentially in the finest fractions where mineralogical sorting by grain size fraction extracted may be more significant. Following homogenization by grinding, the albedo differences and subtle slope differences are largely absent across original grain-size classes, reinforcing little size sorting of mineralogy between the fractions (Figure 8B). Overall, these results demonstrate that albedo and VIS continuum slope are strongly governed by physical surface properties and measurement scale. Careful consideration of grain size, mixing state, and spatial context is therefore essential when comparing spectral parameters derived from field, laboratory, and orbital datasets. On the other hand, absorption presence and position do not change, making at least semi-quantitative compositional comparisons possible across scales.

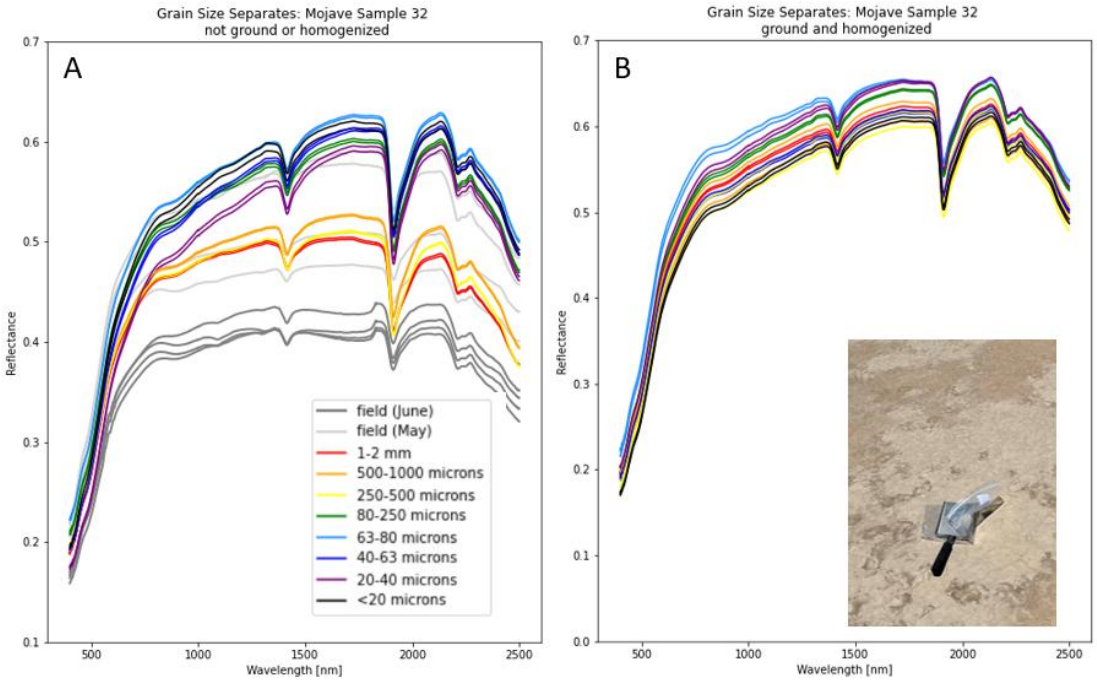

Figure 8. Effects of sediment texture on spectra. (A) grain size separates show differences between coarser and finer grain size separates of sieved sample. In situ spectra also shown in grey for comparison. (B) same grain size separates as in (A), but after having all been ground with mortar and pestle. Differences in spectral shape and overall reflectance are diminished.

3.3 Spectral variability and connections to laboratory-derived mineral abundances

Spectra of mineral dust source sediments sampled in this study span a wide range of spectral shapes and overall reflectance values with VIS albedo ranging from 0.11-0.82 and IR albedo ranging from 0.15-0.93. Continuum-removed spectra exhibit diagnostic absorptions, including iron oxide absorptions near ~530 nm and ~860 nm, clay mineral absorptions near ~2200 and ~2350 nm, and carbonate absorptions near ~2350 nm (Figure 9). Precise absorption positions vary among samples as a function of the specific mineral phases present within each mineral group. Select samples additionally display absorptions associated with hydrated evaporite minerals, including gypsum, particularly in playa environments that undergo repeated wetting and drying cycles.

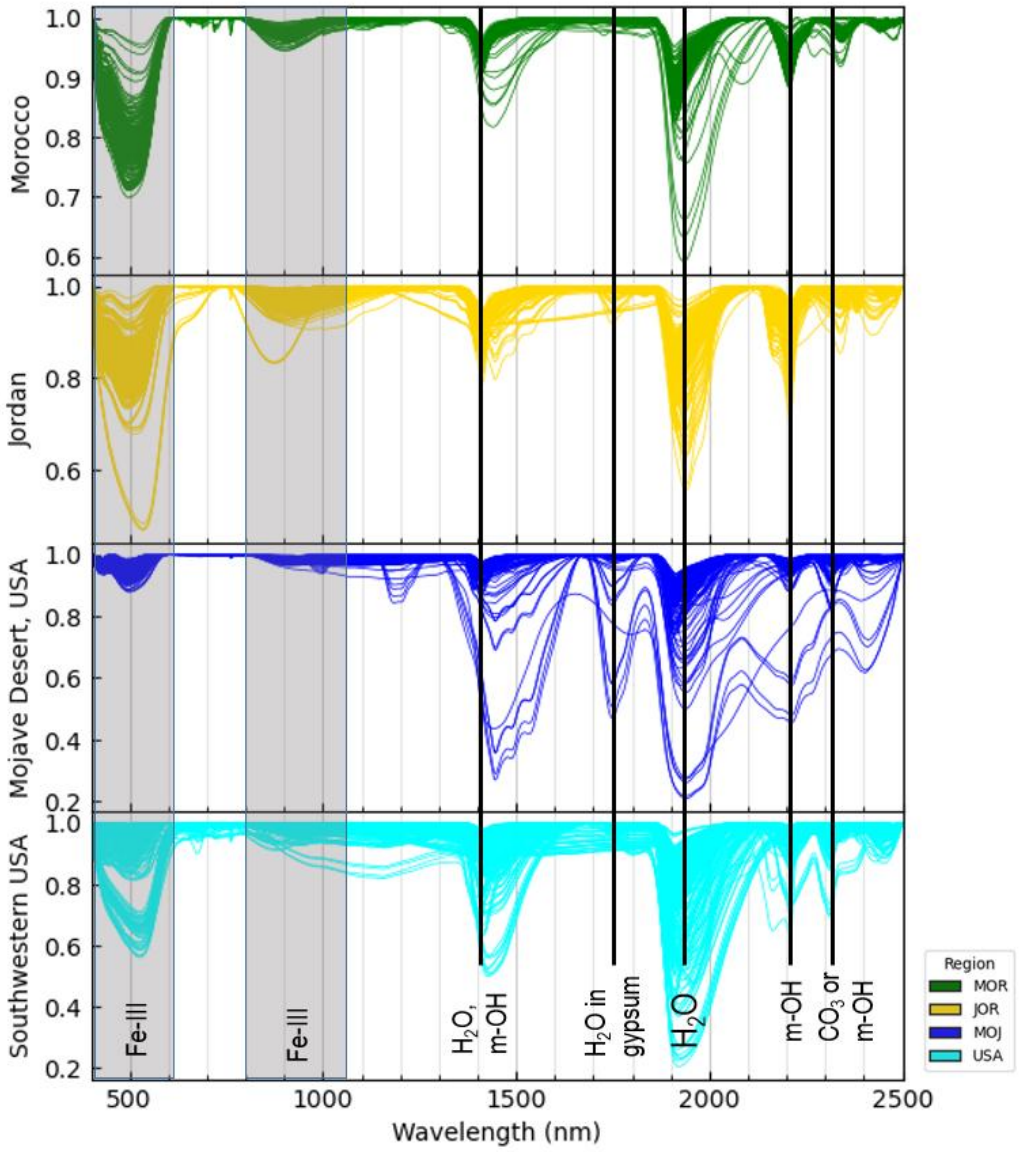

Figure 9. Continuum-removed field spectra from each of the four field campaigns: Morocco, Jordan, Mojave Desert, and selected Southwestern USA. Major absorption feature classes are annotated. Absorption features due to electronic transitions in iron-bearing minerals are shaded; vibrational absorption minima are identified with lines.

The highest-albedo samples are associated with evaporite-rich environments containing abundant carbonate and sulfate minerals, including the White Desert region of Jordan and gypsum-bearing playa systems. Sites in the southwestern United States exhibit a particularly diverse suite of evaporite minerals, including halite,

trona, thenardite, and burkeite, formed from sequential evaporative processes within interconnected playa-groundwater systems (Earman et al., 2005; Haines, 1959).

Principal component analysis (PCA) reveals orthogonal axes of variability within the highly correlated spectral dataset. The first several principal components are physically interpretable: the first principal component primarily follows overall reflectance (albedo), whereas the second principal component is strongly associated with hydrous salt-bearing sediments with overall negative VSWIR continuum slope. Subsequent principal components capture iron oxide absorption characteristics and visible-wavelength spectral slope, as well as more specific mineral absorption features, as illustrated by the eigenvectors shown in Figure 10A (for example, the kaolinite doublet absorption feature at 2210nm can be observed in eigenvectors for principal components 1, 3, and 5). K-means clustering and convex hull fitting within principal component space identify representative and endmember spectra spanning the observed range of albedos, spectral shapes, and mineral absorption characteristics (Figure 10, panels B-C).

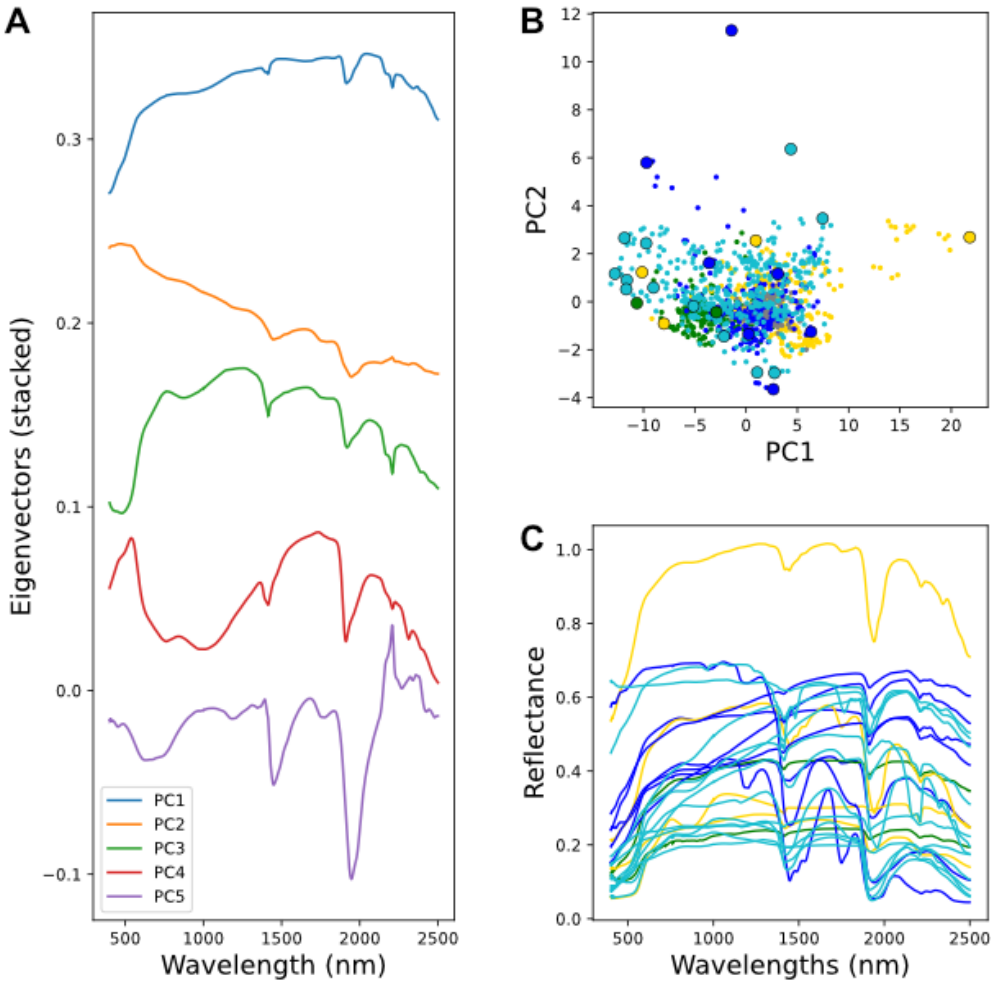

Figure 10. Spectral Variability of in situ spectra collected for this study, expressed by principal component analysis. (A) Eigenvectors indicating spectral features driving variability. (B) Scatterplot of spectra transformed into PC space (PC2 vs PC1), showing spread and clustering of data across field sites (C) Representative spectra spanning the diversity of field samples collected in this study. Spectra exhibit variability in VIS albedo (~0.11-0.82), as well as diverse iron oxide, clay, and evaporite mineral absorptions.

To evaluate how the spectral variability of the in situ dataset compares to global orbital observations, the field spectra were also projected into the principal

component space derived from the global EMIT dataset (Keebler et al., 2026) (Figure 11). PCA results indicate strong consistency between the in situ and EMIT spectral datasets. The dominant axes of variability correspond to similar spectral properties in both datasets, and the overall distribution of spectra in PC space is broadly comparable. Field spectra encompass nearly the full range of variability observed in the EMIT orbital data. This difference is expected given the substantially greater geographic and environmental diversity sampled at orbital scale. Spectral types not represented within the field dataset include the highest-albedo sandy sediments observed in parts of the Sahara (Figure 11, Region 2), although these materials may be less relevant for dust emission due to their limited fine-grained fraction. Additional underrepresented spectral regimes include hydrated sediments and uncommon evaporites (Figure 11, Regions 3, 4, & 5), though a handful of endmember examples of these sediments are present within the in situ dataset. The global EMIT dataset also contains spectra characterized by a distinctive low-angle positive slope between ~400-1300 nm, associated with iron-bearing clay-rich sediments common in parts of Asia (Figure 11, Region 1), and carbonate-bearing sediments in the middle east and southern Africa (Figure 11, Region 6); no clear analogs to this spectral type are present in our field measurements. These differences highlight important regions of spectral variability that remain undersampled by field study of dust properties and should be considered when extrapolating relationships derived from the field dataset to global orbital observations. These results highlight an important consideration for model development: empirical and machine-learning frameworks trained exclusively on field spectra may not fully capture the broader range of spectral variability present in global orbital datasets.

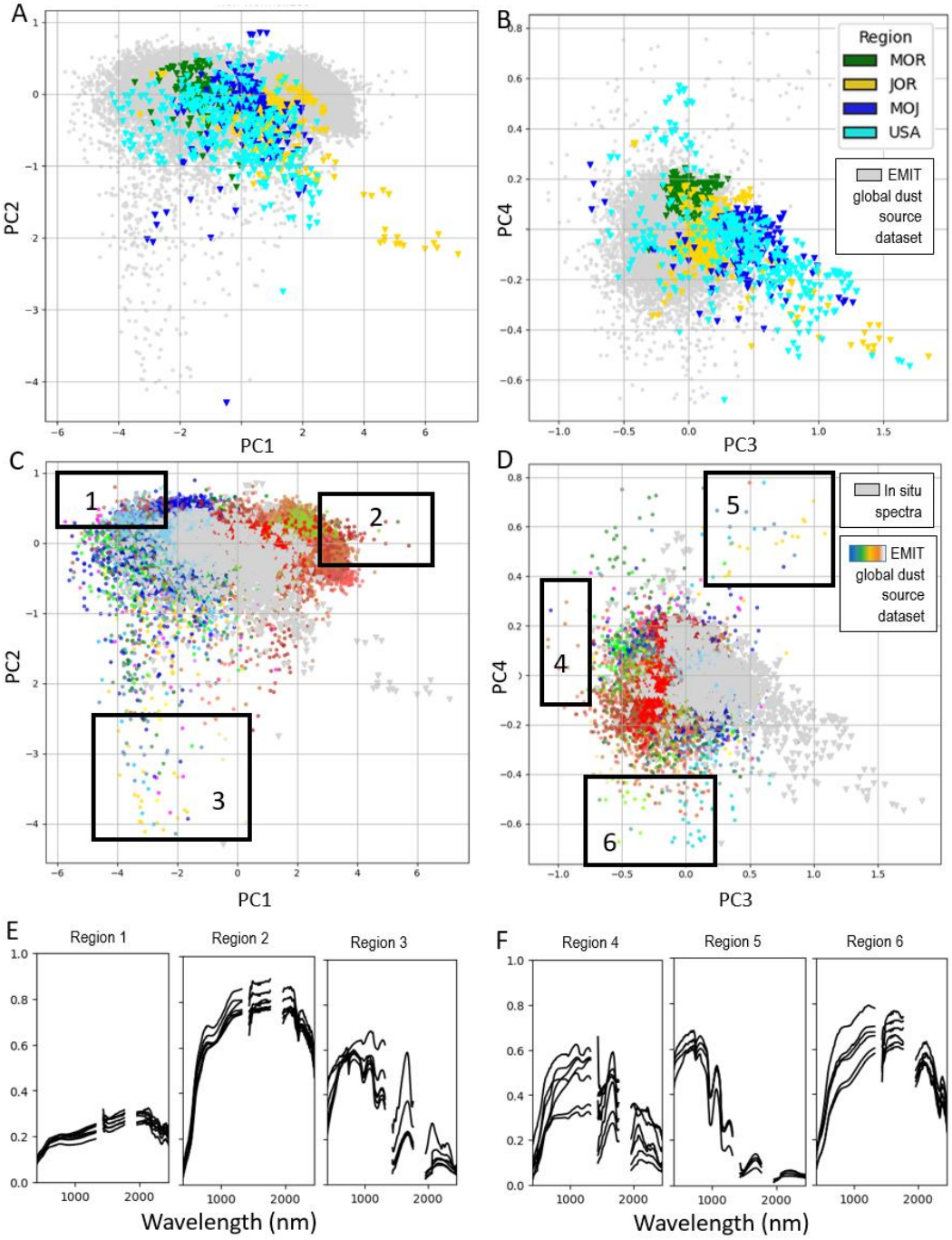

Figure 11. Spectral variability of our field sites as compared to that of all dust source regions on Earth as sampled from global EMIT dataset (Keebler et al., 2026a). (A-B) PC 1vs2 (A) and PC 3vs4 (B) plots for in situ spectra collected in this study transformed into the principal component space of the EMIT global desert dust source dataset. In situ spectra are colored by source region and EMIT global dataset is in grey. (C-D) reversed color scheme of (A-B), illustrating span of different global

regions beyond EMIT and allowing for identification of dust source regions whose variability is not spanned by our in situ dataset. (E) Spectral endmember types in PC 1vs2 space which are not represented by the in situ dataset. (F) Spectral endmember types in PC 3vs4 space which are not represented by the in situ dataset.

Correlation analyses between band depths, spectral slopes, albedo, and X-ray diffraction-derived mineral abundances and iron chemistry data reveal systematic relationships between spectral properties and composition (Figure 12). To evaluate these relationships, we compute both Pearson correlations to assess linear associations and Spearman correlations to capture monotonic (potentially nonlinear) relationships, reflecting the nonlinear nature of VSWIR spectral mixing processes. In general, linear relationships (Figure 12A) are weaker than monotonic relationships (Figure 12B), indicating that many spectral-mineral relationships are not well described by simple linear trends. One notable exception is gypsum, which exhibits a strong linear correlation with band depth at ~1760 nm, a feature attributed to structural water in gypsum. This relationship likely reflects both the uniqueness of this absorption feature among the measured mineral assemblages and the influence of a small number of high-gypsum samples with strongly expressed spectral features, which disproportionately strengthen the linear trend. Outside of this case, evaporite minerals lacking strong diagnostic absorption features show little correlation with spectral parameters. In contrast, quartz, despite lacking prominent absorption features in the VSWIR range, exhibits measurable correlations with spectral properties, likely reflecting its covariation with broader compositional trends such as iron oxide abundance. The strongest and most consistent relationships are observed between mineralogy and iron oxide spectral metrics, particularly visible spectral slope and iron oxide absorption strength. These results underscore the dominant influence of iron-bearing phases on spectral variability across the dataset, consistent with their strong effects on both continuum shape and diagnostic absorption features.

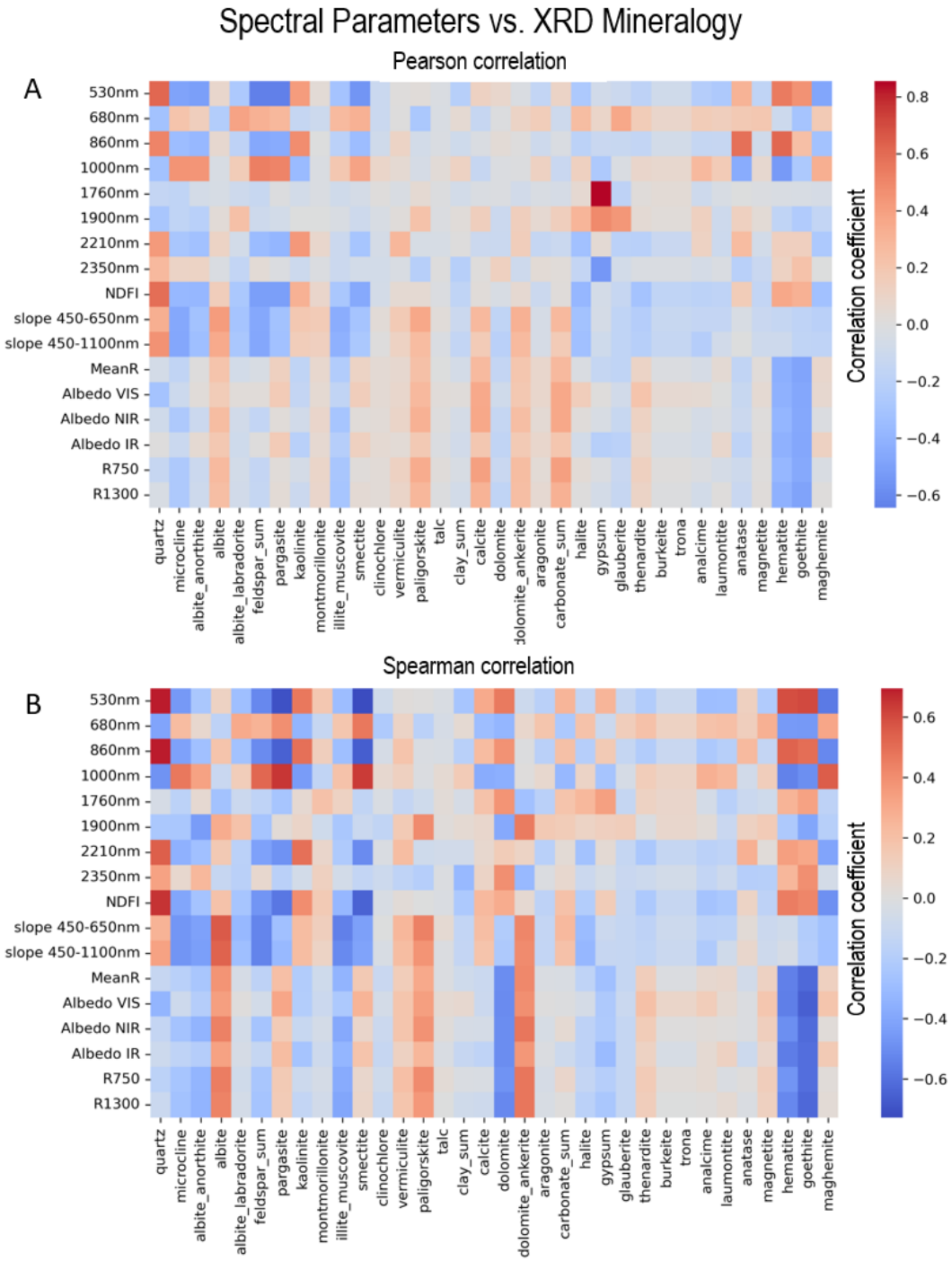

Figure 12. Correlation matrices for spectral properties vs. quantitative mineralogy from XRD. (A) Pearson correlation (linear) (B) Spearman correlation (monotonic).

3.4 Connections between spectral properties and mineralogy

Comparing diagnostic absorption band depths with quantitative mineral abundances provides insight into how mineralogical variability influences the expression of spectral absorption features. Clay minerals exhibit diagnostic Al-OH absorption features near 2210 nm, which in this dataset are primarily associated with kaolinite and illite/muscovite. Kaolinite abundance shows a weak positive relationship with 2210 nm band depth ($R^2 = 0.188$), with substantial scatter and a shallow slope (Figure 13A). Despite the weak correlation, the trend indicates that stronger 2210 nm absorptions are generally associated with increased kaolinite abundance. In contrast, illite/muscovite abundance exhibits a more complex relationship with 2210 nm band depth (Figure 13B). Samples from the southwestern United States show a modest positive association between illite abundance and band depth, whereas Mojave samples display a nonlinear distribution and samples from Jordan and Morocco exhibit little apparent relationship. This behavior likely reflects the non-unique nature of the 2210 nm absorption feature, which is shared among multiple Al-bearing clay minerals. In particular, several kaolinite-rich samples from Jordan exhibit strong 2210 nm absorptions despite containing little to no illite, resulting in samples with high band depth but low illite abundance. The relationship between 2210 nm band depth and the combined abundance of kaolinite and illite/muscovite is negligible ($R^2 = 0.016$; Figure 13C), indicating that summing the abundances of these minerals does not improve prediction of absorption strength. These results demonstrate that while the 2210 nm feature is effective for identifying the presence of Al-OH-bearing clays, band depth alone is not a reliable predictor of clay mineral abundance in mixed assemblages.

Surprisingly, carbonate abundance has no relationship with the 2350 nm band depth metric (Figure 13D). The distribution of samples occupies a triangular region in abundance-band depth space, with the lower boundary defined by a positive relationship between carbonate abundance and absorption strength. However, this positive relationship is in negative band depth space, since gypsum-bearing samples

exhibit strongly negative 2350 nm band depths as a consequence of overlapping absorptions. More broadly, the carbonate absorption feature is comparatively weak (<5% for samples with highest wt.% calcite) relative to the diagnostic absorptions of other mineral groups examined here, limiting the utility of simple band-depth metrics as quantitative indicators of carbonate abundance.

In contrast, gypsum abundance exhibits a strong relationship with the 1760 nm band depth metric (Figure 13E). This absorption feature is unique to gypsum within the studied mineral assemblages and is not significantly affected by overlapping absorptions from other common phases. The distribution of gypsum abundances is sparse in the dataset, which includes numerous samples containing little to no gypsum as well as a few gypsum-dominated samples exceeding 80 wt.%. At high gypsum abundance, the 1760 nm absorption deepens to ~30%, providing a robust spectral indicator of elevated gypsum concentrations.

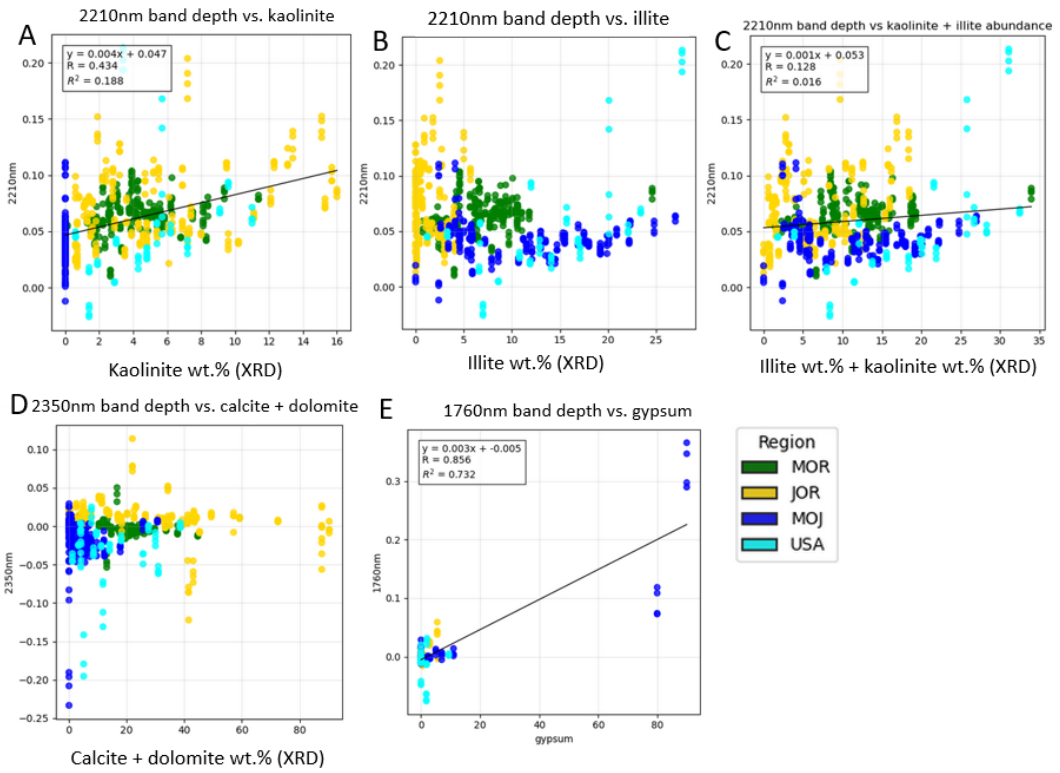

Figure 13. Correlation plots for relationships between diagnostic spectral absorption band depths and mineral abundances determined by quantitative XRD, colored by sampling location. Scatterplots compare band depth measurements to the abundance of (A) kaolinite versus the 2210 nm absorption feature, (B) illite/muscovite versus the 2210 nm absorption feature, (C) calcite + dolomite versus the 2350 nm absorption feature, and (D) gypsum versus the 1760 nm absorption feature. Solid lines indicate linear regression fits where statistically significant relationships were observed.

3.5 Connections between spectral properties and iron speciation

Hematite and goethite were identified by XRD in a subset of samples; however, diagnostic spectral absorption features indicate their presence in a substantially larger number of samples, giving confidence that these iron oxides are more widely present than XRD detections alone suggest. This discrepancy likely reflects the combined effects of low mass abundance and poor crystallinity, which can limit XRD detectability while still producing measurable spectral absorptions. In this study, iron abundance was quantified by species using wet-chemical analyses, allowing direct

evaluation of relationships between spectral parameters and crystalline iron oxide abundance.

Sequential extraction via wet chemistry can reliably quantify abundance of iron in iron oxides, but unlike with XRD, hematite and goethite are unable to be distinguished by the wet chemistry method. Thus, we compare iron oxide abundance from wet chemistry to the maximum absorption depth due to the combined, superimposed absorption features of hematite at 530 nm and goethite at 500 nm. To evaluate spectral sensitivity to iron oxides, we compare the combined hematite-goethite absorption feature centered near ~530 nm with iron oxide abundance derived from wet chemistry measurements. Most samples exhibiting a pronounced ~530 nm absorption feature show a positive relationship with measured iron oxide abundance (Figure 14), supporting the use of this feature as a proxy for iron-bearing phase abundance. However, the relationship is not universal across all sediment types. Some samples display strong visible-wavelength iron oxide absorptions despite relatively low measured iron oxide abundance (Figure 14, panel A), indicating that factors beyond total iron oxide abundance strongly influence spectral expression. But that Fe oxides are driving this absorption, as expected, rather than Fe in other phases is indicated by the linear fit when related to the proportion of iron as crystalline iron oxides. Relationships also emerge between spectral parameters and nominally non-absorbing or weakly absorbing mineral phases. For example, ~530 nm band depth due to hematite and goethite exhibits positive correlations with quartz abundance (Figure 14, panel C), and bulk albedo likewise varies systematically with quartz content. These observations suggest that spectrally neutral phases indirectly influence diagnostic absorption features by modifying scattering behavior and altering the spectral contrast of absorbing minerals. Such effects demonstrate that spectral expression is controlled not only by absorbing phases themselves, but also by the physical and mineralogical context of the surrounding sediment mixture.

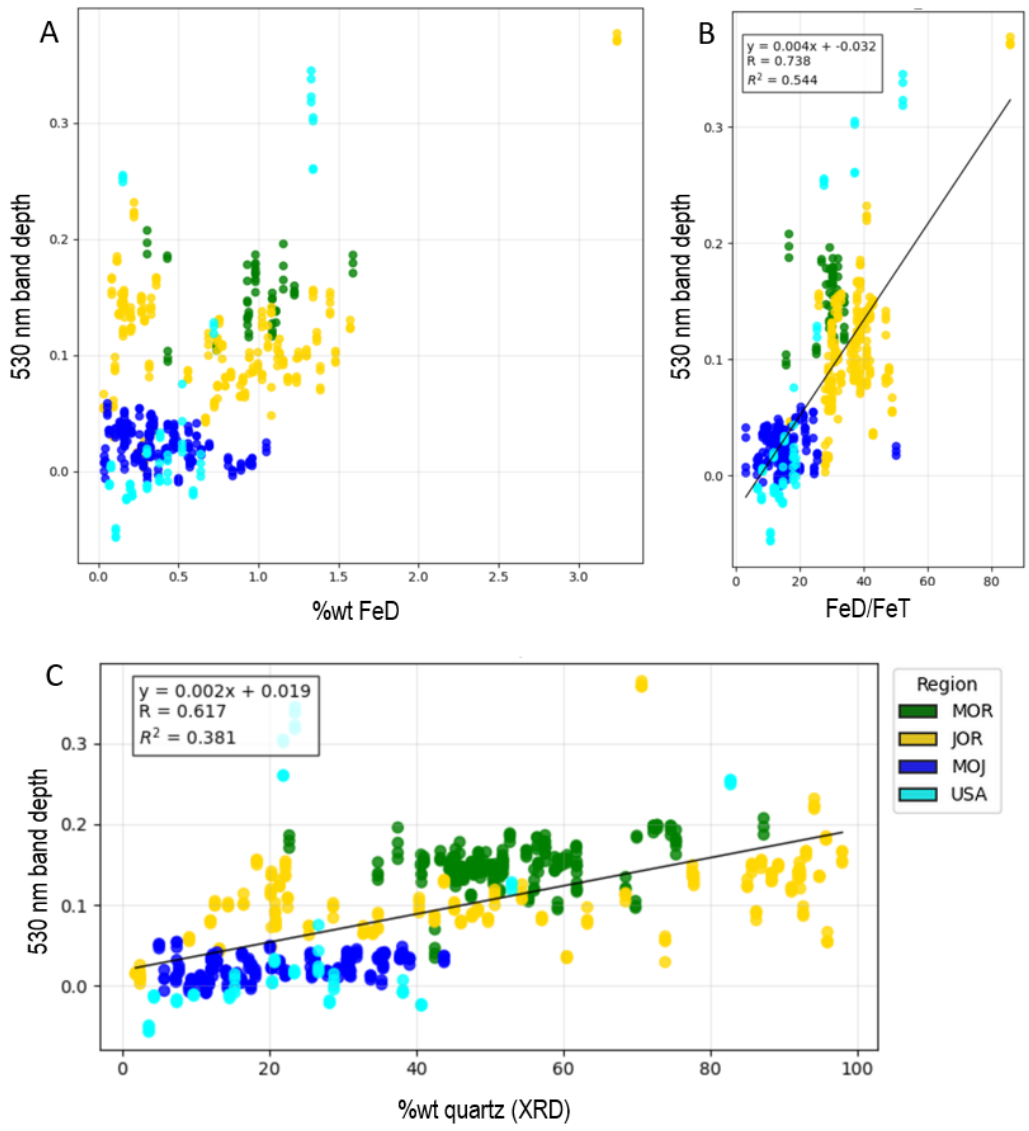

Figure 14. Band depth due to hematite and goethite vs: (A) FeD, iron oxide abundance from iron speciation chemistry (B) FeD/FeT, fraction of total iron from iron speciation chemistry that is in crystalline iron oxides (C) quartz abundance from XRD.

A simple forward radiative transfer modeling using Hapke theory further supports these interpretations. Figure 16 illustrates the expected hematite absorption strength produced under varying hematite abundances and quartz-clay mixing ratios, assuming coarse sand-sized quartz grains mixed with finer kaolinite and hematite particles. Grain size choices for the quartz, clay, and hematite in the forward models

were partially informed by optical microscope images (Figure 15) of select sediment samples with similar band depth due to hematite at 530 nm, but different %wt hematite. Spectra were modeled for up to 5 wt.% hematite, corresponding to the maximum abundance present in the study dataset. The modeling results indicate that samples containing high proportions of coarse quartz can exhibit substantially amplified hematite absorption features relative to mixtures dominated by finer-grained matrices (Figure 14, panel A). In these mixtures, coarse quartz grains decrease multiple scattering and increase the effective optical path length through absorbing phases, thereby lowering albedo and strengthening visible-wavelength absorptions even at relatively low hematite abundances.

These effects are also evident in the forward-modeled spectra (Fig. 16B). Panel B shows spectra containing a constant hematite abundance of 5 wt.% but varying proportions of quartz and kaolinite, as well as different quartz grain sizes. For the kaolinite-rich mixtures (blue curves), which contain only 25 wt.% quartz and are representative of fine-grained dust-producing sediments, changes in quartz grain size produce only minor differences in continuum shape, overall reflectance, and hematite absorption strength. In contrast, the quartz-rich mixtures (tan curves), containing approximately 85 wt.% quartz and representative of sandy materials, exhibit a pronounced grain-size effect. When quartz is fine-grained, the spectra are of similar albedo and continuum shape to the kaolinite-rich mixtures; however, increasing the quartz grain size to sand-sized fractions substantially decreases overall reflectance and enhances the apparent strength of hematite absorption features despite the hematite abundance remaining unchanged.

Similar behavior is observed in the in situ spectra (Figure 14C). Spectra of samples B and D from Figure 15 represent clay-rich and coarse quartz-rich materials, respectively. Although both samples exhibit comparably strong Fe-oxide absorption features, the quartz-rich sample contains only 0.4 wt.% FeD and is strongly absorbing in the VNIR, whereas the clay-rich sample contains 1.6 wt.% FeD. This contrast is consistent with the modeling results and demonstrates that coarse quartz-rich

matrices can produce strong visible Fe-oxide absorptions at substantially lower iron oxide abundances than finer-grained, clay-rich materials.

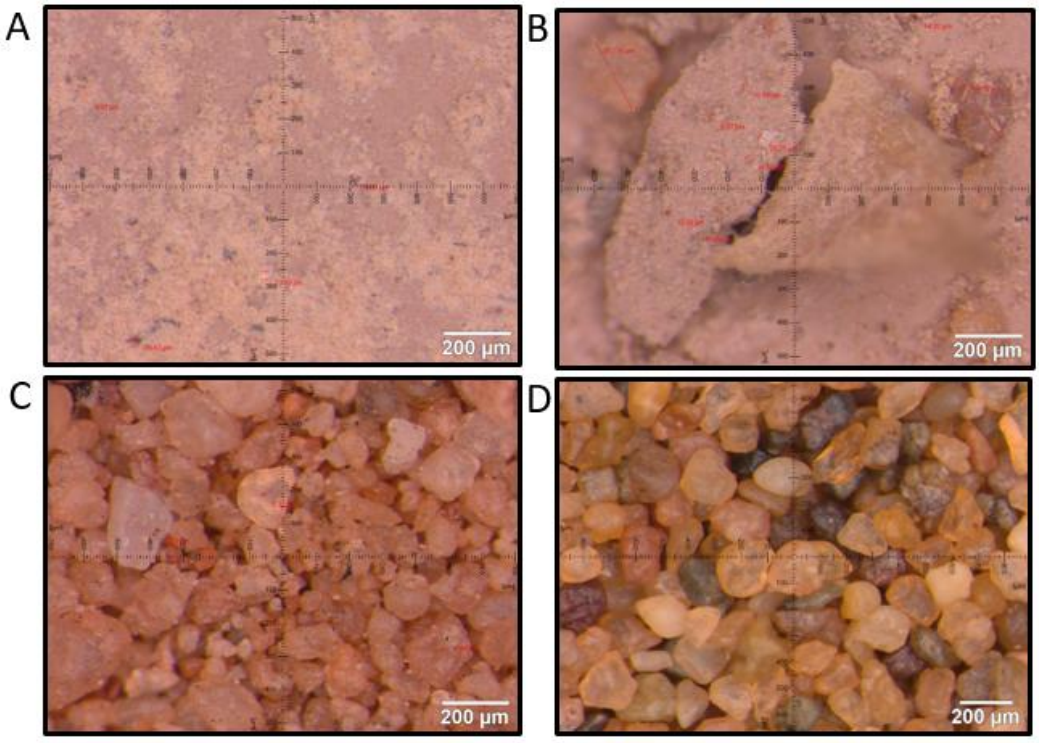

Figure 15. Optical microscope images of sediment samples showing coarse quartz grains and fine (sub-10 micrometer) clay-bearing aggregates and iron oxides. (A-B) examples of sediment crusts with high clay abundance and low to moderate quartz abundance. Both samples exhibit strong strong iron oxide absorptions features at 530 nm and elevated %wt FeD measured via wet chemistry. (C-D) examples of sandy samples with high quartz abundance, strong iron oxide absorptions features at 530 nm, and low %wt FeD measured via wet chemistry.

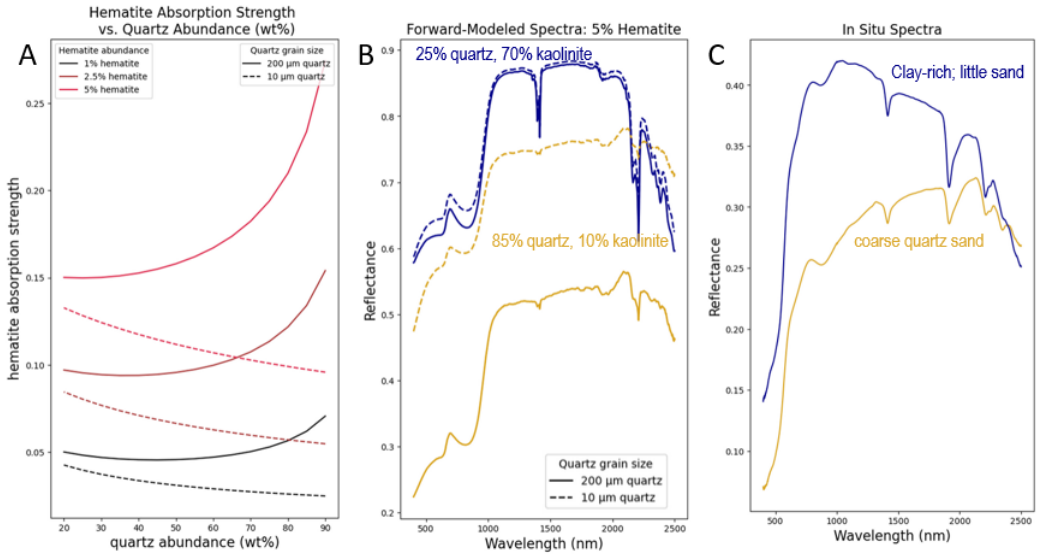

Figure 16. (A) Expected hematite absorption band depth as a function of abundance for different fixed hematite abundances. For constant hematite abundance, models predict hematite absorption strength increasing with quartz abundance for coarse quartz grains and decreasing with quartz abundance for fine quartz grains. (B) Select forward-modeled spectra for constant hematite abundance, varying fractions of quartz vs kaolinite, and varying quartz grain size. (C) in-situ spectra for two samples (MOR033 and MOR040) with similar band depths due to hematite but with different sediment textures.

These results demonstrate that absorption strengths in dust source regions are strongly modulated by the coupled effects of grain size, texture, and the abundance of both absorbing and non-absorbing phases. Consequently, mineralogy-spectra relationships cannot be interpreted solely in terms of direct abundance-band depth relationships. Instead, robust interpretation of mineralogical composition from reflectance spectra requires multivariate and physically informed approaches that explicitly account for sediment texture, mixture effects, and radiative transfer processes.

4. Discussion

4.1 Summary of the in situ measurements of dust source regions

Regionally, Morocco and Jordan—examples of dust source areas in the Sahara and Middle East—exhibit the strongest and most spectrally distinct iron oxide signatures in the dataset. These regions are consistently characterized by pronounced iron oxide absorptions and higher FeD values, in agreement with recent findings on iron oxide absorption band strength (Keebler et al., submitted), global mineralogical interpretations from EMIT (e.g., Li et al., 2026), and prior studies identifying these regions as enriched in hematite- and goethite-bearing dust source materials (e.g., Moskowitz et al., 2016). In addition, both regions contain measurable kaolinite, consistent with previous observations of clay mineral assemblages in North African and Middle Eastern dust sources.

In contrast, samples from the United States, which represent dust sources that are more regionally important rather than globally dominant (Kok et al., 2021), display a distinct mineralogical signature. These sites are characterized by comparatively weaker iron oxide absorptions and lower FeD values, with iron phases identified via XRD as magnetite and maghemite, particularly in the Mojave Desert and Great Basin regions. Relative to Morocco and Jordan, these U.S. samples also show higher feldspar abundances relative to quartz and a broader representation of clay mineral variability. Notably, Mojave samples exhibit a greater contribution of smectite, which is reflected spectrally by a single absorption minimum centered near ~2200 nm. This contrasts with the doublet ~2210 nm feature observed in our Jordanian and some Moroccan samples.

Overall, the dataset captures a wide range of quartz-feldspar ratios together with three dominant clay mineral groups (illite, smectite, and kaolinite), expressed in varying proportions across regions. Iron-bearing phases further contribute a key axis of variability, with implications for dust radiative properties and climate

forcing given the strong optical effects of iron oxides. Taken together, these results demonstrate that the dataset spans a large fraction of the compositional and spectral parameter space relevant to dust source regions globally. Importantly, the combination of paired in situ mineralogical constraints and hyperspectral reflectance measurements provides a more complete framework for interpreting spectral variability in terms of underlying mineralogy than has been available in previous dust source databases. This represents a substantive advance in the state of the art, enabling more physically grounded interpretation of hyperspectral observations and improved attribution of spectral signatures to specific mineralogical assemblages in dust-emitting regions.

4.2 Implications for interpreting mineralogy from remotely sensed spectra

Our results demonstrate key considerations for interpreting mineralogy from remotely sensed spectral data, particularly when comparing laboratory, in situ, and orbital observations. For example, across scales, we find that spectral shape and overall reflectance are strongly influenced not only by mineralogical composition but also by physical properties such as surface texture and the degree of aggregate clumping. These effects are especially evident when comparing laboratory spectra of sieved grain size separates, where each size fraction is spectrally distinct in overall reflectance and continuum shape, but appear uniform once crushed and homogenized. Any mineralogical interpretations derived from albedo or satellite-based spectra must explicitly account for these textural controls, which can significantly modulate both continuum level and band depth.

In situ measurements are able to resolve fine-scale variability, including localized coatings, grain-size sorting, and small-scale textural differences that are effectively averaged out at EMIT spatial resolution. These occurrences are spatially restricted and are likely of limited importance for broader, regional-scale biogeochemical interpretations. Overall, our results suggest that while fine-scale mineralogical heterogeneity is partially obscured at orbital resolution, EMIT effectively captures

the dominant patterns of spectral variability, supporting its utility for regional-scale mineralogical and surface process studies.

Importantly, while overall albedo, spectral shape, and band depth differ between in situ and EMIT data, the absorption occurrences and positions are broadly comparable, and thus orbital measurements preserve much of the essential spectral diversity present at the surface. EMIT data also spans dust source types with compositional and spectral types that are not well represented in our in situ dataset (Fig. 10). This suggests that while our field measurements capture much of the dominant spectral variability observed from orbit, certain dust source sediment compositional classes are absent in our ground-based observations. For a comprehensive spectral-mineralogical dataset from dust source regions, efforts should be made to collect data from so-far undersampled regions. As a result, interpretations or models derived solely from the in situ dataset may not be fully applicable to those spectral types present in the orbital data.

4.3 Challenges in going from spectra to mineral abundance

As expected from radiative transfer theory and previous studies on spectra of sediment mixtures, spectral parameters do not scale linearly with mineral abundance. Diagnostic absorption features often show only partial or context-dependent agreement with independently measured abundances. For example, wet chemistry-derived iron oxide abundance can correlate with iron absorption band depth in some sediment suites, but these relationships are dependent on the grain sizes and mineral assemblages, underscoring the non-unique nature of spectral-abundance relationships.

This complexity is further compounded by the overlap of absorption features among common mineral groups. Many clays exhibit broad, superimposed absorptions in similar wavelength regions, and key diagnostic features can overlap with those of other phases; for instance, illite and carbonate minerals share absorption features at 2350 nm, making it difficult to uniquely attribute band depth variations to a single

mineral without additional constraints. As a result, deconvolving mixed spectral signatures remains a major limitation for direct mineral identification and abundance estimation, particularly in natural systems where multiple phases co-occur at fine scales.

At the same time, there are encouraging results that demonstrate the potential of spectral approaches. We observe meaningful correlations between specific band depths and quantitative XRD-derived mineral abundances, and similarly strong relationships between percent weight iron from chemical analyses and iron-related spectral features. These findings suggest that, despite the inherent complexity, certain mineral groups, including iron oxides, can be robustly estimated from spectral data. However, these successes also highlight the broader challenge: signals from different minerals are highly mixed and context-dependent, meaning that simple univariate metrics such as band depth alone are insufficient for reconstructing full mineral assemblages from orbital spectra.

Multivariate approaches provide a promising path forward. In particular, PCA analyses effectively isolate dominant axes of spectral variability and identify spectral endmembers that correspond well with variations in laboratory-derived mineral abundances. This suggests that empirical relationships are well suited to capturing the coupled effects of composition and physical properties that govern spectral variability. Overall, these results underscore the importance of using this dataset not only to evaluate existing band-depth-based approaches, but also to develop more comprehensive empirical methods for translating spectra into mineralogy.

4.4 Needed future work

Future work will leverage this dataset to validate and improve existing approaches for deriving mineralogical information from spectral data. In particular, systematic comparisons between predicted and observed mineral abundances could help assess the robustness of current mineralogy modeling methods across a wider range of sediment types and surface conditions. Building on this foundation, the dataset also

provides an opportunity to develop and test improved empirical approaches for estimating mineral abundances from reflectance spectra, including PSLR and other regression or machine learning models.

In addition, more comprehensive laboratory analyses are needed to refine constraints on iron oxide content, especially iron oxide phase-specific abundance in sediments, as these are critical for interpreting dust direct radiative effect and biogeochemical cycle contributions. Further work is also necessary to better understand how sediment fractionation during dust emission influences radiative properties and underlying mineralogy in comparison to bulk parent sediments that we observe with satellite and field observations. Developing datasets that explicitly compare spectra of fractionated dust with their parent sediment sources would improve our ability to better estimate fractionated, lofted dust composition based on surface spectral observations of source sediments.

5. Conclusions

We compile a novel dataset of in situ sediment spectra from dust source regions, linked with grain-size distributions, quantitative XRD mineralogy, and iron abundance partitioned by species. We evaluate the spectral variability represented within this dataset and compare it to the global diversity of desert dust source spectra, finding that our measurements capture many of the major spectral types observed worldwide, despite the limited number of field sites and the incomplete representation of global compositional variability.

Dust source sediment compositions range from the iron-oxide and kaolinite bearing, quartz-rich sediments to smectite, maghemite, and feldspar-rich sediments of the Mojave. Broadly, sediments in the dataset are ~20-90 %wt quartz, ~10-50% clay minerals, ~0-40% carbonate minerals, and have 0-1.5 %wt Fe in crystalline Fe-III oxides. Dust source sediments exhibit a wide range of reflectance values (0.11-0.82 VIS albedo) and include diagnostic absorption features associated with iron oxides,

clay minerals, carbonates, and evaporite salts. Principal component analysis further parameterizes the dominant modes of spectral variability. Directly calculated spectral parameters and statistically derived components exhibit correlations with quantitative mineralogical measurements. For example, quartz abundance exhibits linear relationships with band depth calculated at 530 nm, and Figure 10, panel B indicates monotonic correlations between clay mineral and carbonate abundances and spectral features. However, these relationships are influenced by multiple interacting factors, including grain size, sediment texture, aggregation state, and the superposition of overlapping absorption features, suggesting nonlinear relationships which may benefit from an empirical approach.

For iron oxides specifically, we identify that hematite- and goethite-bearing sediment spectra are directly and nearly linearly correlated to measured iron oxide abundance (FeD). The nature of these relationships varies among sediment types, particularly between sand-rich and clay-rich materials. Across all sediment classes, however, iron oxide absorption strength exhibits a consistent relationship with the fraction of total iron that is hosted within iron oxide phases (FeD/FeT). These results highlight the importance of iron-bearing minerals as first-order controls on spectral variability in desert dust source sediments. Future work will focus on developing empirical and physically informed models capable of predicting spectral behavior from mineralogical composition, with the goal of improving interpretation of field and orbital hyperspectral observations of global dust source regions.

Acknowledgements

We are grateful for support for this work from the NASA FINESST Graduate Research Fellowship Program award to A.M.K. and B. L.E. (award number 80NSSC22K1456) and an EMIT Phase E co-investigator award to B.L.E. We also thank the Resnick Sustainability Institute for support. Many thanks to Eva Scheller, Rebecca Greenberger, Jessica Mueller, and Joshua Anadu for assistance during field spectra and sample collection. Thank you to Melani Hernandez-Chiriboga for providing thorough training in lab methods at IDAEA-CSIC.

References

Clark, R. N., & Roush, T. L. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. Journal of Geophysical Research: Solid Earth, 89(B7), 6329-6340.

Clark, R. N., King, T. V., Klejwa, M., Swayze, G. A., & Vergo, N. (1990). High spectral resolution reflectance spectroscopy of minerals. Journal of Geophysical Research: Solid Earth, 95(B8), 12653-12680.

Di Biagio, C., Formenti, P., Balkanski, Y., Caponi, L., Cazaunau, M., Pangui, E., ... & Doussin, J. F. (2019). Complex refractive indices and single-scattering albedo of global dust aerosols in the shortwave spectrum and relationship to size and iron content. Atmospheric Chemistry and Physics, 19(24), 15503-15531.

Earman, S., Phillips, F. M., & McPherson, B. J. (2005). The role of "excess" CO2 in the formation of trona deposits. Applied Geochemistry, 20(12), 2217-2232

Feely, K. C., & Christensen, P. R. (1999). Quantitative compositional analysis using thermal emission spectroscopy: Application to igneous and metamorphic rocks. Journal of Geophysical Research: Planets, 104(E10), 24195-24210.

Formenti, P., Caquineau, S., Desboeufs, K., Klaver, A., Chevaillier, S., Journet, E., & Rajot, J. L. (2014). Mapping the physico-chemical properties of mineral dust in western Africa: mineralogical composition. Atmospheric Chemistry and Physics, 14(19), 10663-10686.

García-Pando, C. P., Stanton, M. C., Diggle, P. J., Trzaska, S., Miller, R. L., Perlwitz, J. P., ... & Thomson, M. C. (2014). Soil dust aerosols and wind as predictors of seasonal meningitis incidence in Niger. Environmental health perspectives, 122(7), 679.

Green, R. O., Mahowald, N., Ung, C., Thompson, D. R., Bator, L., Bennet, M., ... & Zan, J. (2020, March). The Earth surface mineral dust source investigation: An Earth science imaging spectroscopy mission. In 2020 IEEE aerospace conference (pp. 1-15). IEEE.

González-Romero, A., González-Flórez, C., Panta, A., Yus-Díez, J., Reche, C., Córdoba, P., ... & Pérez García-Pando, C. (2023). Variability in sediment particle size, mineralogy, and Fe mode of occurrence across dust-source inland drainage basins: the case of the lower Drâa Valley, Morocco. Atmospheric chemistry and physics, 23(24), 15815-15834.

González-Romero, A., González-Flórez, C., Panta, A., Yus-Díez, J., Córdoba, P., Alastuey, A., ... & Pérez García-Pando, C. (2024a). Characterization of the particle size distribution, mineralogy, and Fe mode of occurrence of dust-emitting sediments from the Mojave Desert, California, USA. Atmospheric chemistry and physics, 24(16), 9155-9176.

González-Romero, A., González-Flórez, C., Panta, A., Yus-Díez, J., Córdoba, P., Alastuey, A., ... & Pérez García-Pando, C. (2024b). Probing Iceland's dust-emitting sediments: particle size distribution, mineralogy, cohesion, Fe mode of occurrence, and reflectance spectra signatures. Atmospheric chemistry and physics, 24(11), 6883-6910.

Hapke, B. (2012). Theory of reflectance and emittance spectroscopy. Cambridge university press.

Haines, David V. Core logs from Searles Lake, San Bernardino County, California. No. 1045-E. US Government Printing Office, 1959.

Jickells, T. D., An, Z. S., Andersen, K. K., Baker, A. R., Bergametti, G., Brooks, N., ... & Torres, R. (2005). Global iron connections between desert dust, ocean biogeochemistry, and climate. science, 308(5718), 67-71.

Keebler, A.M., Ehlmann, B.L., Thompson, D.R., Ginoux, P., Brodrick, P.G., and Green, R.O. (submitted). Spectral Variability of Earth's Desert Dust Source Regions Sampled from EMIT Reflectance Data. JGR: Earth Surface, EMIT special issue.

Kok, J. F., Adebiyi, A. A., Albani, S., Balkanski, Y., Checa-Garcia, R., Chin, M., ... & Wan, J. S. (2021). Contribution of the world's main dust source regions to the global cycle of desert dust. Atmospheric Chemistry and Physics, 21(10), 8169-8193.

Kok, J. F., Storelvmo, T., Karydis, V. A., Adebiyi, A. A., Mahowald, N. M., Evan, A. T., ... & Leung, D. M. (2023). Mineral dust aerosol impacts on global climate and climate change. Nature Reviews Earth & Environment, 4(2), 71-86.

Lafon, S., Sokolik, I. N., Rajot, J. L., Caquineau, S., & Gaudichet, A. (2006). Characterization of iron oxides in mineral dust aerosols: Implications for light absorption. Journal of Geophysical Research: Atmospheres, 111(D21).

Lapotre, M. G., Ehlmann, B. L., & Minson, S. E. (2017). A probabilistic approach to remote compositional analysis of planetary surfaces. Journal of Geophysical Research: Planets, 122(5), 983-1009.

Li, L., Mahowald, N. M., Miller, R. L., Pérez García-Pando, C., Klose, M., Hamilton, D. S., ... & Thompson, D. R. (2021). Quantifying the range of the dust direct radiative effect due to source mineralogy uncertainty. Atmospheric chemistry and physics, 21(5), 3973-4005.

Longtin, D. R., Shettle, E. P., Hummel, J. R., & Pryce, J. D. (1988). A wind dependent desert aerosol model: Radiative properties (Vol. 88, No. 112). Air Force Geophysics Laboratory, Air Force Systems Command, United States Air Force.

Mahowald, N., Albani, S., Kok, J. F., Engelstaeder, S., Scanza, R., Ward, D. S., & Flanner, M. G. (2014). The size distribution of desert dust aerosols and its impact on the Earth system. Aeolian Research, 15, 53-71.

Mancoridis, V., Bue, B., Lee, J. H., Thorpe, A. K., Cusworth, D., Ayasse, A., ... & Duren, R. (2025). Multi-Platform Methane Plume Detection via Model and Domain Adaptation. IEEE Transactions on Geoscience and Remote Sensing.

Moskowitz, Bruce M., Richard L. Reynolds, Harland L. Goldstein, Thelma S. Berquó, Raymond F. Kokaly, and Charlie S. Bristow. "Iron oxide minerals in dust-source sediments from the Bodélé Depression, Chad: Implications for radiative properties and Fe bioavailability of dust plumes from the Sahara." Aeolian Research 22 (2016): 93-106.

Mustard, J. F., & Pieters, C. M. (1989). Photometric phase functions of common geologic minerals and applications to quantitative analysis of mineral mixture reflectance spectra. Journal of Geophysical Research: Solid Earth, 94(B10), 13619-13634.

Okin, G. S., & Painter, T. H. (2004). Effect of grain size on remotely sensed spectral reflectance of sandy desert surfaces. Remote Sensing of Environment, 89(3), 272-280.

Pérez García-Pando, C. (PI). FRAGMENT: Frontiers in dust mineralogical composition and its effects upon climate. ERC Consolidator Grant, Barcelona Supercomputing Center (2018–2024).

Querol, X., Alastuey, A., Chinchon, J. S., Turiel, J. F., & Soler, A. L. (1993). Determination of pyritic sulphur and organic matter contents in Spanish subbituminous coals by X-ray power diffraction. International Journal of Coal Geology, 22(3-4), 279-293.

Rietveld, H. M. (1969). A profile refinement method for nuclear and magnetic structures. Applied Crystallography, 2(2), 65-71.

Shoenfelt, E. M., Sun, J., Winckler, G., Kaplan, M. R., Borunda, A. L., Farrell, K. R., ... & Bostick, B. C. (2017). High particulate iron (II) content in glacially sourced dusts enhances productivity of a model diatom. Science advances, 3(6), e1700314.

Sokolik, I. N., & Toon, O. B. (1999). Incorporation of mineralogical composition into models of the radiative properties of mineral aerosol from UV to IR wavelengths. Journal of Geophysical Research: Atmospheres, 104(D8), 9423-9444.

Swayze, G. A., Clark, R. N., Goetz, A. F., Livo, K. E., Breit, G. N., Kruse, F. A., ... & Ashley, R. P. (2014). Mapping advanced argillic alteration at Cuprite, Nevada, using imaging spectroscopy. Economic geology, 109(5), 1179-1221.

Thompson, D. R., Green, R. O., Bradley, C., Brodrick, P. G., Mahowald, N., Dor, E. B., ... & Zandbergen, S. (2024). On-orbit calibration and performance of the EMIT imaging spectrometer. Remote Sensing of Environment, 303, 113986.

Verpoorter, C., Kutser, T., Seekell, D. A., & Tranvik, L. J. (2014). A global inventory of lakes based on high-resolution satellite imagery. Geophysical Research Letters, 41(18), 6396-6402.

Chapter 4

EMPIRICAL ESTIMATION OF QUANTITATIVE MINERALOGY FROM HYPERSPECTRAL REFLECTANCE DATA OF DESERT DUST SOURCE REGIONS

Abigail M. Keebler[1], Bethany L. Ehlmann[1,2,3,4]

1 California Institute of Technology, Pasadena, CA, United States

2 Laboratory for Atmospheric and Space Physics, Boulder, CO, United States

3 Department of Geological Sciences, University of Colorado, Boulder

4 Department of Astrophysical & Planetary Sciences, University of Colorado, Boulder

Abstract

Quantitative estimation of surface mineralogy from visible-shortwave infrared (VSWIR) reflectance spectra is essential for characterizing dust-source regions and their compositional variability. However, existing physics-based and spectral library approaches, including current EMIT retrieval frameworks, are challenged by the coupled effects of mineral mixing, grain size variability, particle shape, porosity, and multiple scattering, which complicate robust quantitative inversion. Here, we explore an empirical alternative that bypasses explicit physical modeling by learning direct relationships between spectra and mineral composition. We develop Partial Least Squares Regression (PLSR) models trained on a novel dataset of dust-source sediments that links field VSWIR reflectance spectra with laboratory-measured mineralogy, iron chemistry, and grain size. Models were optimized using five-fold cross-validation and evaluated across multiple preprocessing and hyperparameter configurations. The resulting model achieves fair predictive accuracy for quartz, feldspar, and iron oxides ($R2 \sim 0.6$) and is able to identify the presence of mixed mineral assemblages, including clays, carbonates, and evaporites. We evaluate model performance using EMIT reflectance spectra co-located with ground-truth mineralogy and compare predictions against EMIT Level 2B and Level 3 products. We further assess extrapolation behavior using spectra outside the training domain.

The PLSR model reproduces major mineral constituents identified by EMIT while improving quantitative estimates for calcite across all test cases and achieving comparable or improved performance for quartz, feldspar, clay minerals, and iron oxides. These results demonstrate the potential of empirical spectral-mineral models as a complementary approach for improving quantitative mineralogy retrieval from VSWIR observations.

1. Introduction

Imaging spectrometers such as the Earth surface Mineral dust source InvesTigation (EMIT) instrument provide visible to shortwave infrared (VSWIR) reflectance spectra of Earth's surface at global scales. A major goal of this investigation is to provide quantitative mineralogy of bare surfaces in Earth's deserts to determine the mineralogy of dust emitted into Earth's climate system (Green et al., 2020, Thompson et al., 2024). VSWIR reflectance characteristics depend on the sediments' mineralogy, iron chemistry, and grain sizes. In particular, diagnostic absorption features associated with iron oxides, phyllosilicates, carbonates, and salts, as well as scattering controlled by silicates like quartz and feldspars, provide the basis for quantitative mineralogical interpretation of mineral intimate mixtures (i.e., Hapke, 2012; Clark et al., 1990; Clark et al., 2024). These compositional properties are of central importance for understanding mineral dust radiative forcing, nutrient transport, and land-atmosphere interactions in Earth system models (i.e., Mahowald et al., 2014; Shao et al., 2011).

However, extracting quantitative mineralogical information from hyperspectral data is not a solved problem and is inherently challenging due to complex, nonlinear interactions of light with densely packed particulate matter and the role of both composition and physical properties (Hapke, 2008). Spectral mixing is influenced not only by mineral abundance and chemical solid solution of constituent materials, but also by grain size, shape, porosity and multiple scattering effects, which can strongly modulate absorption band depth and continuum shape. For example,

increases in grain size of spectrally non-absorbing phases such as quartz can enhance the apparent depth of absorption features associated with co-occurring absorptive minerals such as hematite through nonlinear path length effects (Chapter 3). These effects complicate the direct inversion of mineral abundance from spectral features in mixed sediments.

These challenges manifest in current EMIT mineralogical data products, which are derived using the Tetracorder algorithm. Tetracorder is an expert decision-making algorithm which identifies constituent minerals by comparing observed absorption features with entries in the USGS spectral library and identifying best-fit endmembers (Clark et al., 2024). While this approach provides physically interpretable mineral maps at global scale, it exhibits reduced sensitivity to some mineral phases within compositionally complex mixtures. Because the algorithm identifies minerals by matching observed spectra to library spectral endmembers, spectrally dominant phases can mask weaker or overlapping absorption features from less dominant constituents. For example, clay minerals often exhibit strong and distinctive shortwave infrared absorption features that dominate the spectral fit, which can lead to carbonate minerals being underestimated or omitted when their broader and less diagnostic absorptions occur within similar wavelength regions. Iron oxide abundances may also be inaccurately estimated depending on concentration, grain size effects, and the surrounding spectral mixing context. Although recent advances in EMIT processing workflows have incorporated expanded spectral libraries and mixtures of mineral endmembers to better represent natural variability (Clark et al., 2024), such approaches remain challenged by the complexity of natural mineral assemblages. In practice, it is not feasible to build a spectral library which fully explicitly represents all possible mixtures and proportions of mineral components encountered in natural sediments, particularly in cases where multiple clay species, carbonate phases, and iron oxides co-occur, producing overlapping and non-unique absorption features within the same spectral regions and where grain size is also variable.

An alternative approach for estimating mineral abundances from reflectance spectra is the use of physics-based radiative transfer models, such as Hapke modeling frameworks that invert for mineralogical composition using laboratory-derived optical constants of constituent endmembers (e.g., Mustard & Pieters, 1989; Lapotre et al., 2017). These methods are physically grounded and explicitly model light interaction with intimate mineral mixtures, offering the potential for quantitative mineral abundance retrievals that are directly tied to first-principles scattering physics. However, their rigorous application to natural dust source sediment mixtures is not possible at this point. The most important limitation is that radiative transfer model performance depends strongly on the accuracy and completeness of the optical constants used to represent pure mineral phases. Existing optical constant datasets (e.g., Longtin et al., 1988; Querry et al., 1985) often do not span the full spectral range required for visible-to-shortwave infrared remote sensing applications. Crucially, the values in the literature vary substantially between source materials for a given mineral, introducing uncertainty into modeled spectral behavior (Zhang et al., 2015) For example, Zhang et al. (2015) demonstrate that climate models employing hematite optical constants from different datasets will produce markedly different radiative properties. In particular, optical constants from Longtin et al. (1988) lead to an underestimation of hematite's absorption, rendering it comparable to that of clay minerals such as illite, whereas using those from Querry et al. (1985) can substantially overestimate its absorptive capacity. For several mineral phases relevant to natural dust sediments, appropriate optical constants are either partially constrained or entirely unavailable (e.g., illite). Natural sediments are complex substances with variable solid solution chemistries of minerals (which affects optical constants), particle size distributions, and sediment grain aggregate structures, which can alter spectral behavior in ways that are difficult to explicitly parameterize. Additional complexity arises because dust-source sediments commonly contain grain sizes (<10 μm) that are comparable to VSWIR wavelengths, leading to a transition from geometric to Mie scattering regimes. In this regime, the assumptions underlying Hapke radiative transfer

models—particularly, the treatment of particles as optically large scatterers with well-defined single-scattering behavior—are increasingly violated, making it difficult to accurately parameterize reflectance for fine-grained mineral aggregates (Hapke 2012). As a result, although physics-based inversion approaches are powerful, they are not yet ready to robustly apply across compositionally diverse natural dust source sediments.

In contrast to physically constrained spectral unmixing approaches, empirical and machine learning-based methods offer an alternative strategy that utilizes statistical relationships between observed spectra and measured mineral composition and grain size. These methods treat hyperspectral reflectance as a high-dimensional predictor space and infer mineral abundances through regression or nonlinear mapping functions without assumptions about physics of light interactions or need to recognize endmembers. Empirical and machine learning methods can provide complementary or alternative pathways for estimating mineral composition directly from reflectance spectra, particularly in complex or mixed mineral assemblages (e.g., Jahoda et al., 2021; Hajaj et al., 2024). Empirical models can implicitly capture nonlinear interactions between spectral features and compositional variables, including effects arising from grain size, intimate mixing, and secondary scattering processes, which are presently difficult to represent explicitly in physics-based inversion frameworks. A challenge to date has been the unavailability of empirical datasets of known spectra and mineralogy to train these models.

Here, we develop and evaluate an empirical framework for jointly estimating iron speciation mass fraction, quartz abundance, and additional major mineral phases from visible to shortwave infrared (VSWIR) reflectance spectra. We leverage a new coupled dataset of VSWIR spectra, iron chemistry, and quantitative mineralogy described in Chapter 3 to build the models. We assess the potential for quantitative retrieval of sediment composition using an empirical linear regression method. We (1) develop and evaluate a partial least squares empirical model to estimate

quantitative mineralogy, grain size distributions, and iron chemistry from paired in situ reflectance observations, (2) relate model performance and feature importance to spectral properties to identify which spectral properties are most strongly controlled by variations in mineralogy, and (3) apply the trained models to EMIT reflectance data and compare resulting predictions to the existing state-of-the-art EMIT mineral detection and abundance estimate data products.

2. Methods

2.1 Training Dataset

The dataset used to develop these empirical models is described in greater detail in Chapter 3. It consists of in situ visible-to-shortwave infrared (VSWIR) reflectance spectra collected from dust source sediment surfaces, providing a close analog to satellite-based hyperspectral observations such as those acquired by EMIT. These spectral measurements, shown in in Figure 1, are paired with laboratory-derived measurements of grain size distribution (wet and dry sieving/analysis), iron speciation chemistry, and quantitative X-ray diffraction (XRD) mineralogy (Gonzales-Romero 2023 & 2024). Samples were collected from multiple arid and semi-arid dust source regions, including the southwestern United States, Morocco, and Jordan. These sites span a range of geologic and geomorphic settings and include unconsolidated sands and sediment crusts.

All samples were analyzed using laboratory procedures described in Chapter 3 (Section 2: Methods). Full VSWIR reflectance spectra spanning 400–2500 nm are available for all samples. Quantitative XRD analyses and grain size distribution measurements were also performed for the complete dataset. Laboratory iron speciation measurements were only available for a subset of samples from Morocco and the southwestern United States due to the labor-intensive nature of the analytical procedures. To ensure consistency across all modeled outputs, accuracy in iron mineral content that is more accurately measured with the wet chemistry

approach, and avoid introducing missing-data bias, the final modeling dataset was restricted to 140 spectral-mineralogical sample pairs for which the full suite of laboratory measurements (spectral, mineralogical, grain size, and iron chemistry) was available.

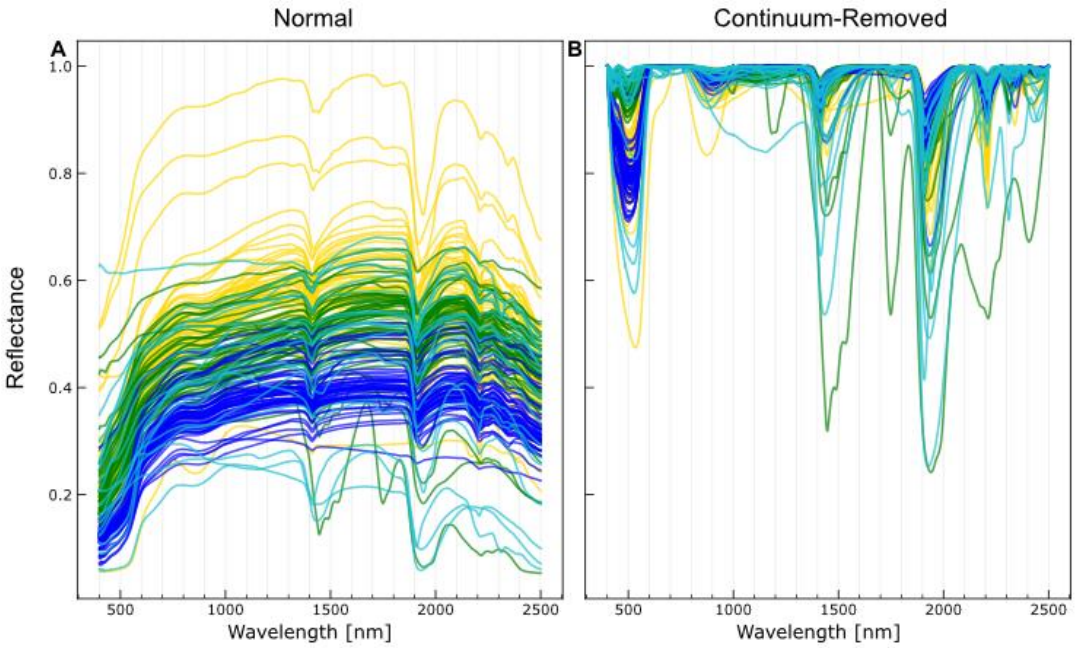

Figure 1. Spectral dataset used as training and testing input for empirical models. (A) normal and (B) continuum-removed to show characteristic mineralogical absorption features.

The sampled sediments are compositionally diverse and primarily consist of mixtures of quartz, feldspars, clay minerals, carbonate minerals, and iron oxide phases; mineral family abundance distributions for the samples are shown in Figure 2. Clay mineral assemblages are primarily composed of kaolinite, illite, and smectite-group minerals, with smaller contributions from chlorite, montmorillonite, palygorskite, and vermiculite in a subset of samples. Carbonate-bearing samples are dominated by calcite, with some dolomite. Iron oxide and oxyhydroxide phases include hematite, goethite, magnetite, and maghemite, which vary in abundance across arid source regions and are of particular interest due to their strong spectral

influence in the visible and near-infrared regions. A subset of samples collected from evaporite-rich playa environments additionally contains salt mineral phases, including gypsum, halite, trona, burkeite, and thenardite. These phases introduce additional spectral complexity due to hydration state variability and strong absorption features in the shortwave infrared.

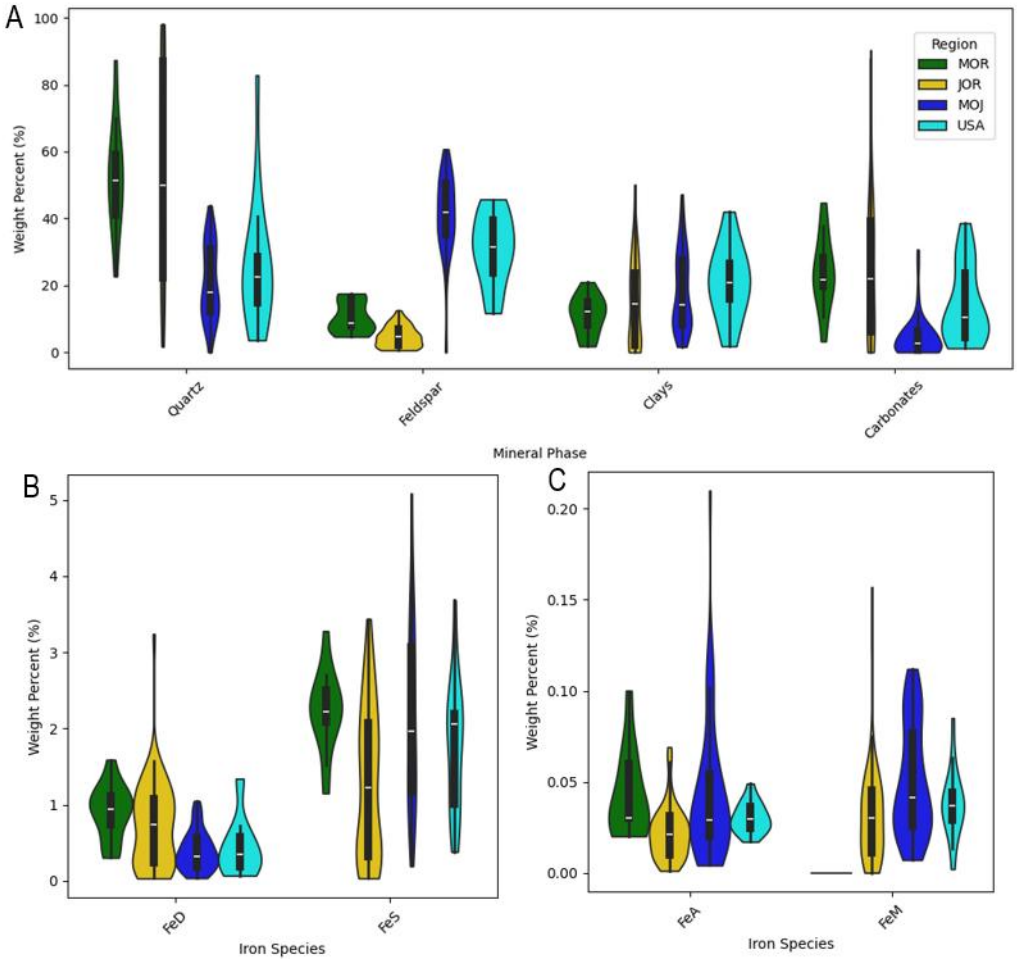

Figure 2: Distributions of laboratory-derived compositional data for samples, separated by region of origin. (A) major mineral groups, detected and quantified with XRD. (B-C) %wt Fe by species. (B) FeD (iron in insoluble, crystalline FeIII-iron oxides) and FeS (iron in structure of silicates). (C) FeA (iron adsorbed in clay minerals) and FeM (iron in magnetite and maghemite). Note that FeM was not measured for samples from Morocco.

Across the sampled regions, distinct mineral assemblages are observed that reflect differences in sediment provenance and weathering environment. Figure 3 presents a principal component analysis (PCA) of the mineralogical dataset, with mineral loading vectors plotted in principal component space to illustrate the dominant axes of compositional variability and mineral associations within the dataset.

The primary axis of variability (PC1) separates samples collected from Morocco and Jordan from those collected in the Mojave Desert and other southwestern United States sites. Moroccan and Jordanian samples are generally characterized by higher abundances of quartz, calcite, and kaolinite, whereas the Mojave and additional U.S. samples are associated with greater feldspar, illite, and montmorillonite abundances. Minor phases including pargasite and zeolite minerals also cluster with the U.S. samples. The secondary axis of variability (PC2) primarily distinguishes evaporite-bearing sediments containing minerals such as gypsum and halite. These samples were collected from playa environments and exhibit distinct compositional signatures relative to silicate-dominated sediments.

The PCA results demonstrate that the dataset represents some mineralogical compositions representative of key diverse global dust source environments as described in Chapters 2 and 3. Sediments from major dust sources in the Sahara and Middle East are characterized by relatively bright reflectance spectra associated with abundant quartz, hematite and goethite, and kaolinite-rich clay assemblages (Chapter 2, Section 4.1). In the training dataset, these sources are represented by samples from Morocco and Jordan. Dust source sediments containing less hematite and goethite, and whose primary clay mineral absorptions are due to illite and/or montmorillonite, are represented in the training dataset by samples from the southwestern USA. Samples from the USA also represent some of the evaporite minerals (i.e., gypsum) observed in the global spectral dataset. However, this training dataset does not provide comprehensive representation of the full diversity of all dust source sediments. Extremely reflective Saharan sands, carbonate-rich

Saharan sediments, and low-reflectance, iron-bearing-clay-rich Asian dust source sediments are among the key sediment types not represented in the training dataset (Chapter 3, Figure 11).

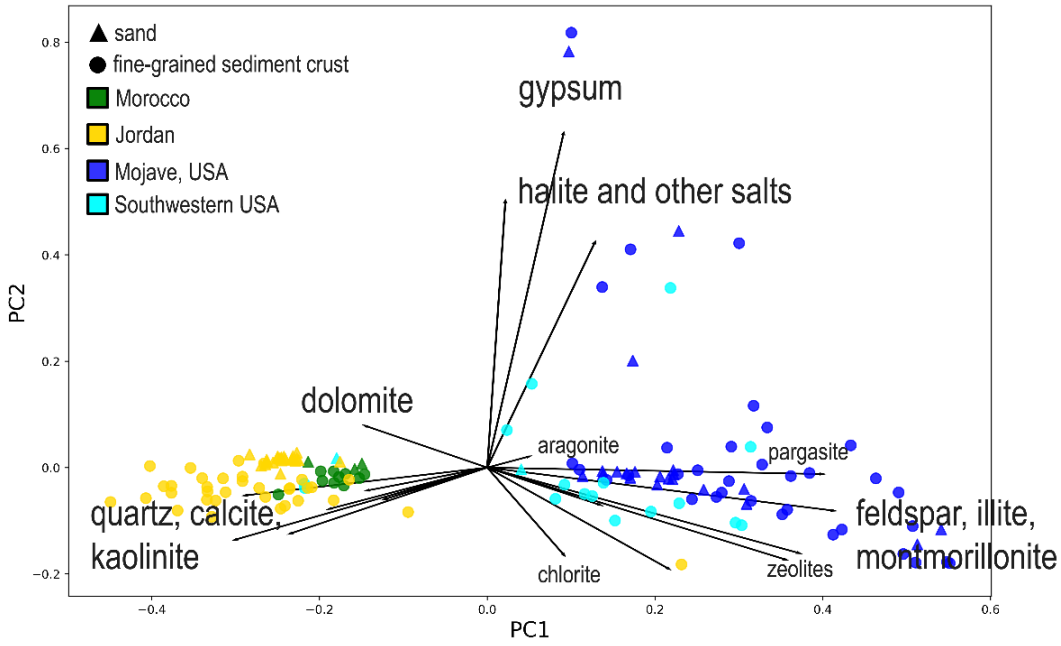

Figure 3: Principal Component Analysis of mineralogical data for samples in the training dataset with mineral loading vectors plotted in principal component space to illustrate the dominant axes of compositional variability. Data color indicates sampling source region and data marker shape indicates sediment type.

A detailed discussion of spectral variability across samples is provided in Chapter 3. That section also describes the primary physical and compositional controls on spectral reflectance behavior in this dataset, which form the basis for the empirical modeling framework developed in this study.

2.2 Modeling Approach

Here, we employ a supervised machine learning regression approach to evaluate the extent to which hyperspectral reflectance spectra can be used to predict quantitative sediment composition. Hyperspectral reflectance spectra serve as the predictor variables, representing the signal that would be observed by an orbital

imaging spectrometer. Target variables include mineral abundances, iron speciation expressed as weight percent iron by phase, and grain size distribution parameters. Reflectance spectra were treated as high-dimensional feature vectors, with each wavelength band representing an individual predictor, and target compositional variables were modeled simultaneously within a multi-output regression framework. To evaluate the sensitivity of model performance to spectral representation, multiple preprocessing strategies were tested, including spectral sub-setting and continuum removal.

Models were trained using the full raw VSWIR reflectance spectra to capture integrated spectral variability associated with overall mineral assemblage, including albedo and brightness effects related to grain size, quartz abundance, and iron-bearing phases. To focus on retrievals of clay, carbonate, and evaporite mineral abundances, models were also trained on the subset of the spectral range (SWIR; 1500-2450nm) to isolate diagnostic absorption features. For both the full spectral range and the SWIR subset, separate models were trained on both the original spectra and on continuum-removed spectra to reduce the influence of broad spectral slope and albedo effects. Across models, performance was compared via error metrics to identify the pre-processing that produced the highest predictive accuracy.

Partial Least Squares Regression (PLSR) was used to evaluate linear relationships between hyperspectral reflectance spectra and sediment composition. PLSR is a latent-variable regression technique that projects high-dimensional, collinear predictor variables into a reduced set of orthogonal components that maximize covariance between predictors (spectra) and response variables (mineralogy, iron speciation, and grain size) (Wold et al., 2001). This makes it particularly well-suited for hyperspectral datasets, where adjacent spectral bands are highly correlated.

PLSR models were implemented using the PLSRegression class in the scikit-learn library in Python. Model performance was evaluated using five-fold cross-validation

(GroupKFold in scikit-learn) using an 80-20 train-test split. That is, the training dataset of ~140 unique spectra with associated laboratory mineralogical data was divided into five groups. A PLSR model was trained on four of the five groups, then predictions were made for the fifth group. This process was then repeated four times so that all five groups were each used as the testing set once. Then, test-set predictions were assembled post-training to assess model performance with error metrics like coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE), computed separately for each target variable. Model complexity was controlled through the number of latent components (n_components), which defines the dimensionality of the reduced spectral space. This parameter was systematically varied from 2-20 components, and optimal model configuration was selected based on cross-validated predictive performance.

Variable Importance in Projection (VIP) scores derived from the fitted PLSR models were analyzed to identify spectral regions contributing most strongly to model performance. VIP scores summarize the relative contribution of each spectral band to the latent-variable regression framework and provide a means of assessing which wavelength regions are most strongly associated with variability in the modeled compositional parameters.

2.3 Application to EMIT data and comparison to EMIT Mineralogy Models

2.3.1 Final Model and Datasets

We then apply the model to corresponding hyperspectral reflectance data from EMIT which are co-located with our in situ spectral sampling sites (Chapter 3) and compare with corresponding EMIT mineral identification data products for these spectra. The final PLSR model, using the optimal number of latent components determined during model tuning, was trained using the full in situ spectral dataset and corresponding laboratory mineralogical abundances. Prior to training the model, the in situ spectra were resampled from 1-nm collection spectral resolution

to the EMIT wavelength spacing and spectral resolution (~7.5 nm) so the model could be applied. Spectral channels affected by detector-edge artifacts and atmospheric water absorption bands and adjacent low-quality regions were removed from the training spectra to match EMIT spectral masking procedures. Model prediction error metrics and VIP scores from the new model trained on the coarser wavelength spacing were compared to the model trained on the full in situ spectral resolution as described in Section 2.2 to confirm that the model performance was not impacted by resampling.

The final PLSR model was applied to select EMIT spectra co-located with the field sampling sites described in Chapter 3. To assess model performance, we compare PLSR-predicted mineralogy and iron oxide abundance with state-of the art EMIT mineralogy products. EMIT currently has two mineralogy products: the Level 2B (L2B) Estimated Mineral Identification product and the Level 3 (L3) Aggregated Spectral Abundance product.

The EMIT L2B Estimated Mineral Identification product provides per-spectrum mineral identifications for 60 m x 60 m pixels based on the Tetracorder algorithm (see Introduction). The L2B product reports two mineral identifications for each spectrum corresponding to distinct spectral absorption regions. Group 1 identifications are primarily derived from visible to near-infrared (VIS–NIR) wavelengths (~400-1300 nm), where electronic absorptions associated with iron-bearing minerals are located, including iron oxides and ferrous iron in silicates. Group 2 identifications are derived primarily from shortwave infrared (SWIR) wavelengths (~2000-2450 nm), which contain diagnostic vibrational absorption features associated with clay minerals, carbonates, and sulfates. This enables evaluation of individual spectra but does not yield comprehensive mineral abundance estimates across all mineral groups.

The EMIT L3 Aggregated Spectral Abundance product provides quantitative estimates of the abundances of ten climatically relevant minerals (kaolinite, illite,

montmorillonite, vermiculite, chlorite, hematite, goethite, calcite, dolomite, and gypsum) plus quartz and feldspar. Mineral mass fractional abundances are reported for a 0.5 x 0.5 degree spatial grid (approximately 55 x 55 km) by aggregating mineral identifications and estimated abundances for all EMIT bareground spectra within each grid cell. The EMIT L3 algorithm applies masking to remove pixels containing bedrock, fractional vegetation cover, and anthropogenic structures (EMIT L3 ATBD), such that per-grid-cell mineralogy is intended to represent primarily alluvium and unconsolidated sediments.

To evaluate the performance of the PLSR model relative to state-of-the-art EMIT mineral retrieval products, we apply the model to selected EMIT spectra and compare outputs across methods. Where available, results are compared to ground-truthed mineralogy derived from in situ sampling and laboratory analyses to assess relative performance across mineral types and spectral characteristics.

2.3.2 Comparison site selection

Site selection was guided by three criteria. First, we select approximately four sites that capture key compositional endmembers within both the training dataset and the broader global dataset, enabling evaluation across a range of mineral assemblages. For these sites, PLSR and EMIT-derived mineral detections and abundance estimates are directly compared to ground-truth observations. Second, to assess model generalizability beyond the training distribution, we include three sites whose spectral characteristics fall outside the variability represented in the in situ training data. Because no ground-truth mineralogy is available for these locations, we compare PLSR and EMIT outputs and evaluate them in the context of expert spectral interpretation. Finally, all sites are chosen to be as spatially homogeneous as possible at the scale of the EMIT L2B and L3 products to ensure meaningful comparison across methods. The EMIT L2B product has a spatial resolution of 60 × 60 m, which allows relatively fine-scale site selection, whereas the EMIT L3 product aggregates observations to a ~55 × 55 km grid, which can encompass

substantial sub-pixel heterogeneity; this is particularly challenging in regions with localized sources (e.g., North American dust sources).

In total, seven dust source sites were selected for model comparison and evaluation: four within the compositional span of the training dataset and three representing dust sources outside this range. Site spectra are shown in Fig. 4. Sampling location maps at the spatial scales of the EMIT L2B and L3 products are provided in Fig. 6 to assess spatial homogeneity. For the L2B scale, an approximate 60 x 60 m EMIT pixel footprint is shown around each sampling location (Fig. 6, panels F-I and M-O). For the L3 scale, the EMIT L3 grid is overlaid to illustrate the size of each dust source relative to the grid cell and to contextualize the degree of compositional similarity among sediments contributing to each retrieval. The global unconsolidated materials mask is also displayed to indicate bedrock areas excluded from the retrieval. Although vegetation-contaminated pixels are additionally removed during aggregation, these are not explicitly mapped here.

For the four sites with available ground-truth mineralogy, locations were selected to capture the compositional variability represented within the training dataset. Corresponding EMIT spectra are shown in Fig. 4A alongside the in situ spectra of sediments from which laboratory mineralogical measurements were obtained.

To assess how the selected samples relate to the training dataset and to global dust source spectral variability, we project the into the principal component space derived from the global EMIT dust-source dataset (Chapter 2). Figure 5 shows the EMIT global dust source dataset in PC space. The in situ data are overlaid in the same PC space to illustrate the extent of variability captured by the training dataset, shown colored by sample region. The spectra used in this study are plotted with large circle markers to indicate site selection both within and beyond the span of the training set.

To represent quartz-, hematite-, kaolinite-, and calcite-bearing sediments characteristic of Saharan and Middle Eastern dust sources (Chapter 2), sites were selected from the Draa River Valley, Morocco, and Wadi Rum, Jordan. The Draa River Valley is a ~2,900 km² basin, from which sample MOR025 (Chapter 3) was selected as representative of the dominant surface materials. Wadi Rum is an extensive dust source region (~800 km²), from which sample JOR021 (Chapter 3) was selected. Both sites were chosen because their spectra are representative of the broader source regions; in global principal component space, they plot near cluster centers of the in-situ dataset and of the global dataset (Figure 5). Additionally, they occupy relatively homogeneous areas at both the EMIT L2B (Figure 6, Panels H-I) and L3 (Figure 6, Panels D-E) spatial scales, facilitating meaningful comparison between PLSR-derived mineralogy and EMIT mineral abundance products. Although some heterogeneity is present within the L3 grid cells, much of the non-source terrain is excluded through application of the EMIT masking procedures (Figure 6, Panels D-E).

To represent feldspar-rich, clay-rich, and evaporite-bearing sediments characteristic of many North American dust sources and other arid environments, two sites were selected from western North America. Railroad Valley Playa, Nevada (sample USA053), is a large playa system (~400 km²) composed of fine-grained sediments typical of playa dust sources and was selected as a representative example of this depositional environment. While the site does not fully occupy an L3 grid cell (Figure 6B), substantial portions of the surrounding mountainous terrain are excluded by the unconsolidated-materials mask, and additional vegetated areas are expected to be removed through the EMIT fractional-cover filtering procedure. Mesquite Lake Playa, California (sample MOJ013), was selected as a relatively homogeneous evaporite-rich playa exhibiting strong diagnostic evaporite absorption features. Although the playa is spatially homogeneous at the scale of individual EMIT L2B pixels (Figure 6G), its spatial extent (~80 km²) is considerably smaller than the EMIT L3 grid cell (Fig 6C).

Mesquite Lake Playa was included to evaluate L2B mineral identification performance for sediments containing both clay minerals and evaporites, and to assess the extent to which these mineralogical signatures are retained in the aggregated L3 product despite the source occupying only a small fraction of the L3 grid cell.

The trained PLSR models were also applied to three select EMIT spectra that fall outside the spectral variability represented in the training dataset (Figure 5) to assess model performance when extrapolating. These spectra are shown in Fig. 4B and correspond to dust source regions in Turpan Pendi, China; Grand Erg Oriental, Algeria; and the Libyan Desert. The selected sites represent distinct spectral endmembers not captured within the training data. The sample from Turpan Pendi of dark, iron- and clay-rich sediments from Central Asia, with absorption features characteristic of iron-rich chlorite and/or illite content. The bright Saharan sample is characteristic of quartz-rich dune sands and exhibits absorption features consistent with goethite at 500nm and kaolinite at 2210 nm. The carbonate-rich sample from Libya shows strong absorption features at 2350 nm and towards 2500 nm, consistent with calcite. Figure 5 illustrates how spectra outside the range of the training dataset project into global principal component space. The Asian dust and bright Saharan sands fall outside the in situ training dataset envelope in PC2–PC1 space, while the carbonate-rich sample lies outside the training span in PC4–PC3 space. All three sites occur in hyper-arid environments with minimal vegetation cover and are characterized by large, relatively homogeneous source regions at both the EMIT L2B and L3 spatial scales (Fig. 6, panels J-O). Ground-truth mineralogical measurements are not available for these locations. Consequently, evaluation is based on comparison between PLSR-derived mineralogical estimates and EMIT mineral abundance products relative to expert spectral interpretation, providing a qualitative assessment of model behavior when applied beyond the spectral domain represented in the in situ training dataset.

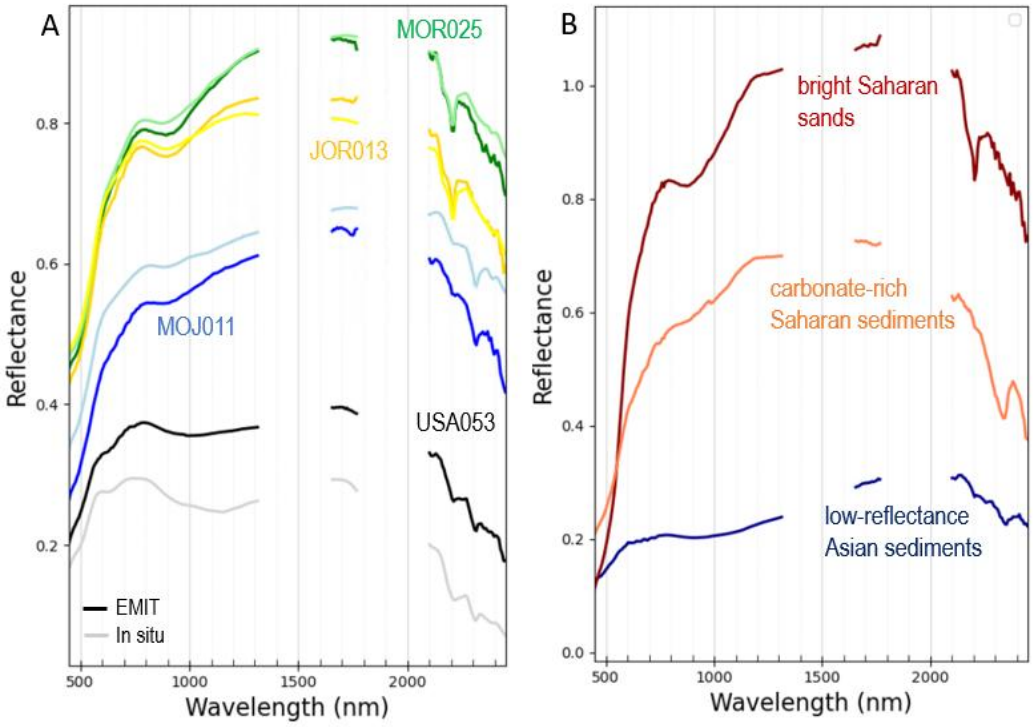

Figure 4. Selected EMIT spectra used to compare PLSR-derived mineral predictions with EMIT operational mineralogy products. (A) EMIT spectra from sites included in the in situ training dataset, shown with analogous in situ spectra used to train the model. (B) EMIT spectra representing spectral types not captured within the in situ training dataset.

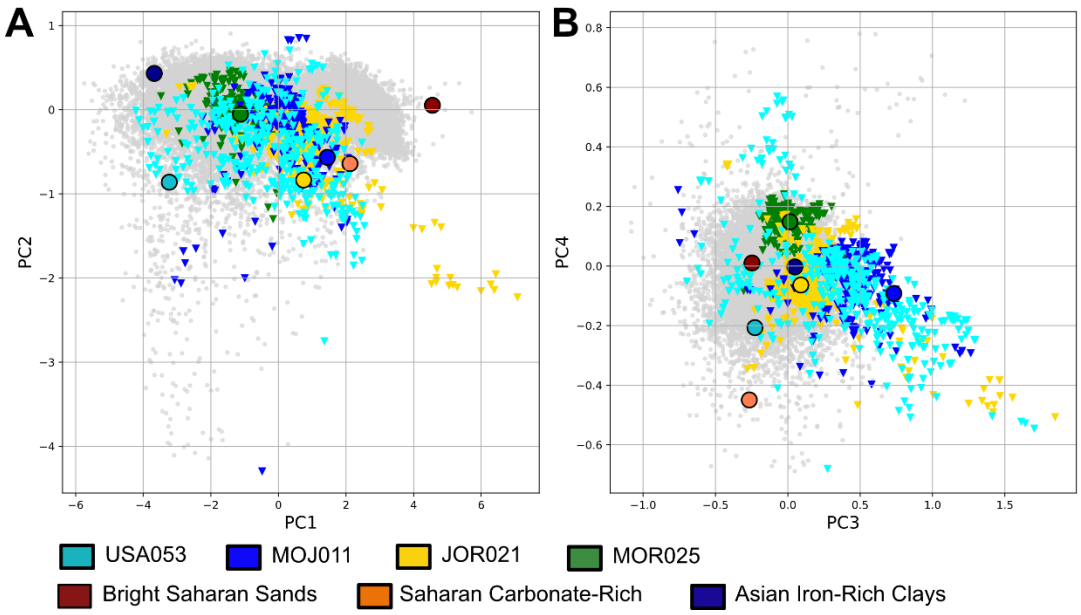

Figure 5. Principal component analysis of the full global EMIT dust-source dataset (grey points; Chapter 2) with in situ spectra dataset projected into the same PC space (colors; Chapter 3). Selected sites for this study are plotted with large circles. (A) PC2 vs PC1. (B) PC4 vs PC3.

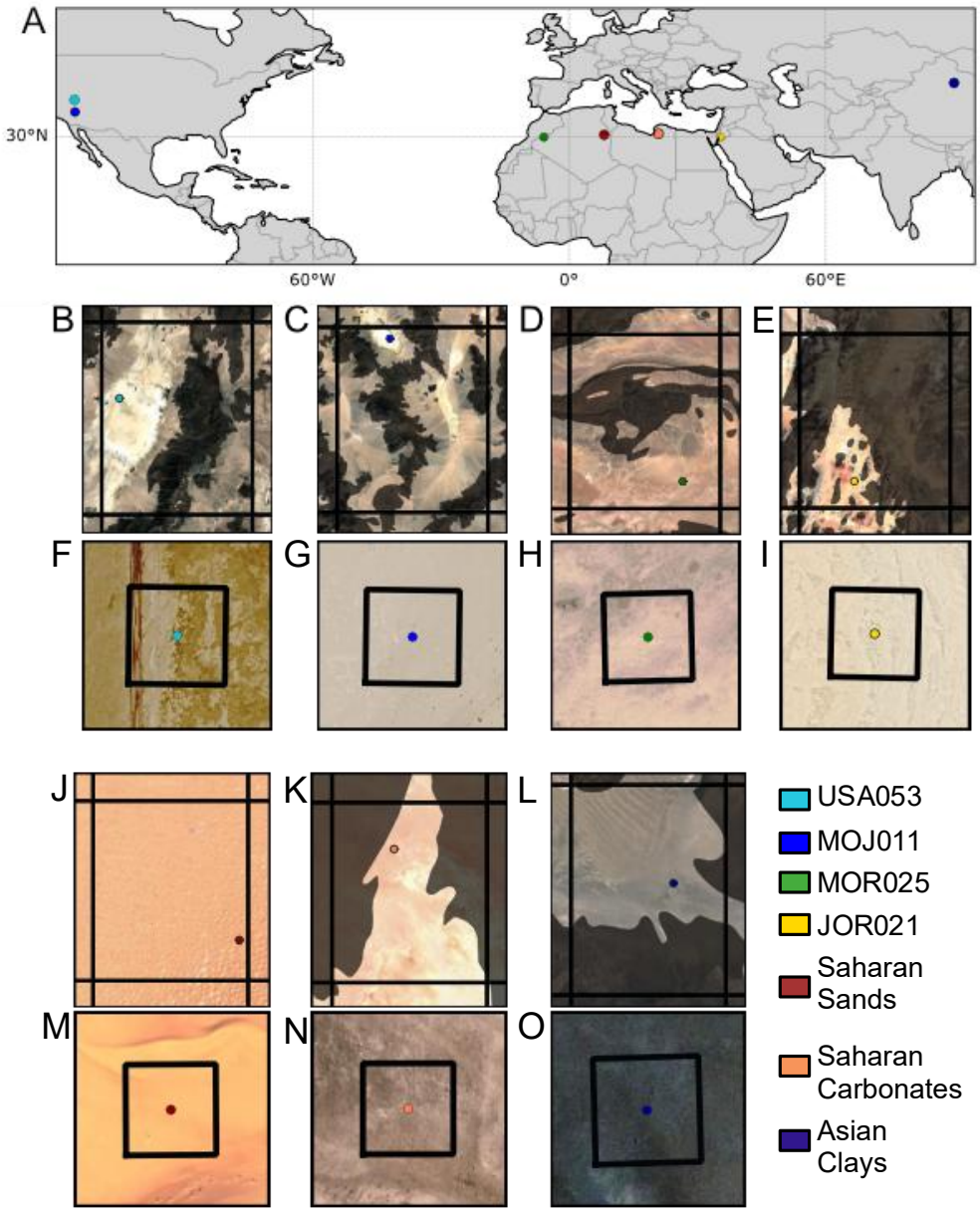

Figure 6. Sampling locations for EMIT observations used to evaluate PLSR results against EMIT mineralogy products in this study, color-coded by source. (A) Global context map. (B-E) EMIT L3 0.5-degree grid cells (black lines) containing observations that are within the span of the training dataset. (F-I) close-up of in-span sampling sites with simulated 60-m EMIT pixel footprint (black box) to show spatial heterogeneity at L2B scale. (J-K) EMIT L3 0.5-degree grid cells (black lines) containing sampling locations that are not co-located with in situ sampling sites and fall outside the spectral variability spanned by the training dataset. (M-O) close-up of out-of-span sampling points with simulated 60-m EMIT pixel footprint (black box).

3. Results

3.1 Partial Least Squares Regression

We trained Partial Least Squares Regression (PLSR) models to predict mineralogical and geochemical parameters from hyperspectral reflectance spectra and report results from the best-performing configuration. Model performance was optimized over a range of latent component numbers (2-25), spectral preprocessing choices, and target variable groupings. Model performance was assessed using validation error metrics and coefficient of determination (R^2) values, with performance improving up to approximately eight latent components before additional components produced minimal improvement and eventually increased prediction error The optimal model configuration for estimating mineral mass fractions used 8 components and the full (400-2450nm) non-continuum-removed spectra.

Figure 7 presents observed versus predicted scatterplots for all modeled targets and Table 1 summarizes performance metrics (R^2, RMSE, and MAE) for each variable. The strongest predictive accuracy was achieved for quartz, feldspar, and fraction of total iron that comes from crystalline Fe-III oxides (FeD fraction), each with R^2 values greater than 0.6. Moderate predictive performance ($R^2 > 0.3$) was obtained for calcite, % wt FeD, % wt iron in mineral structures (FeS), fraction of total iron in FeS, kaolinite, illite, and smectite.

In the case of some mineral targets, negative mineral abundance predictions arise from the unconstrained nature of the Partial Least Squares Regression model, which does not enforce physical bounds on predicted values. These negative values are most commonly observed in phyllosilicate minerals, where overlapping SWIR absorption features (e.g., Al-OH vibrational features at ~2200nm due to different clay phases) result in strong collinearity among clay endmembers. In cases where clay abundance is low or poorly constrained by the observed spectral signal, the

regression solution may extend below zero as part of a multivariate error-minimization process in latent space, reflecting statistical extrapolation rather than physically meaningful mineralogy. Error in carbonate, clay, and sulfate mineral abundance retrieval is likely higher overall than error for quartz, feldspar, and iron oxides because they exert control over absorption features, continuum shape, and albedo over a much smaller spectral range (~1700-2500 nm compared to the full spectral range).

To interpret model behavior, Figure 8 shows variable importance in projection (VIP) scores as a function of wavelength. The VIP spectra highlight multiple diagnostic absorption features that align with known mineralogical absorptions. Prominent contributions occur in wavelength regions consistent with iron oxide electronic transitions, as well as absorption features associated with Al-OH-bearing phyllosilicates such as kaolinite and illite. Collectively, these patterns indicate that the model is not relying on arbitrary spectral correlations but is instead exploiting physically meaningful absorption features consistent with established mineral spectral behavior.

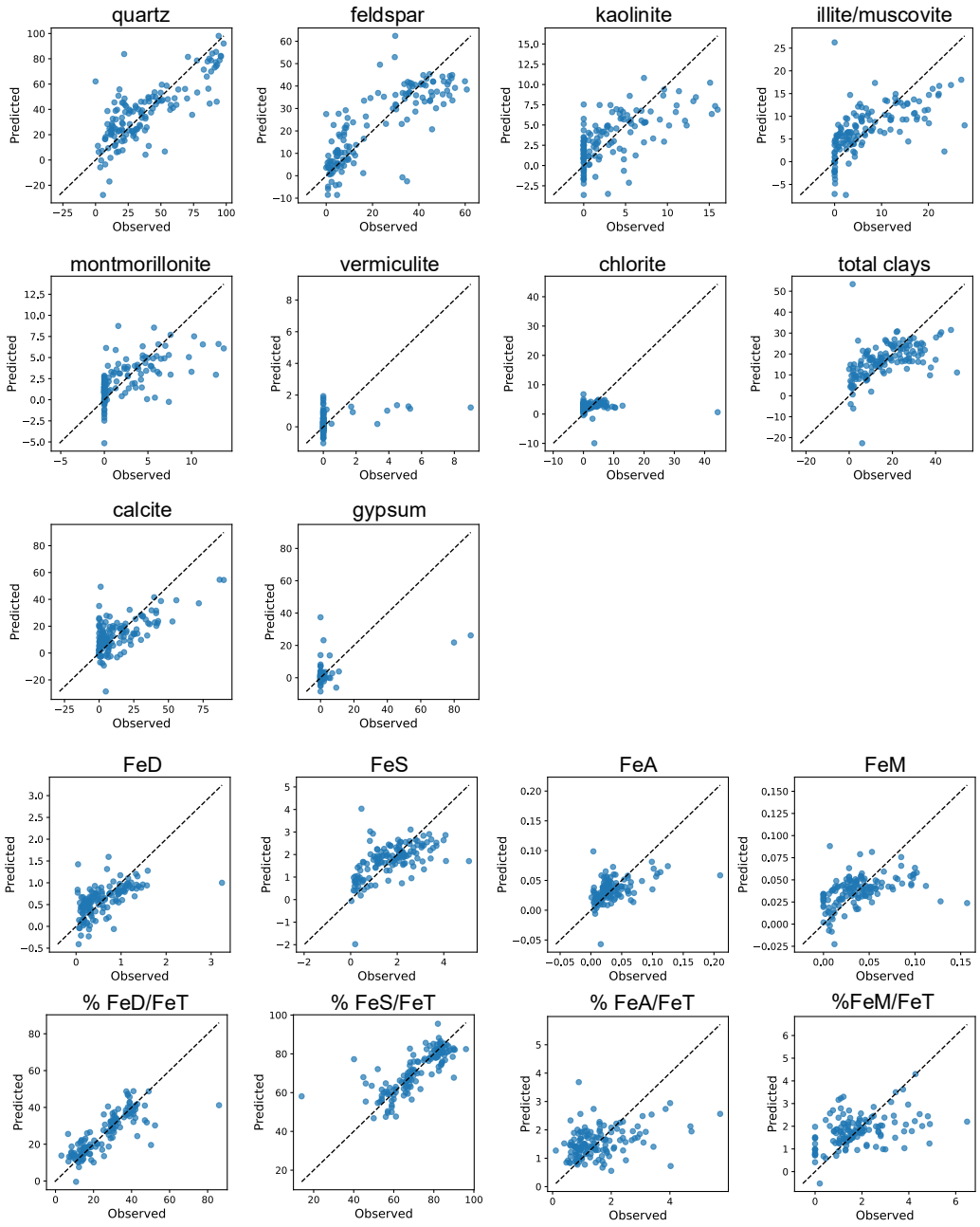

Figure 7. PLSR model performance. Observed versus predicted values for each modeled target using the optimal PLSR configuration (8 components, non-continuum-removed spectra) with the 1:1 line plotted for reference.

Target	R^2	MAE	RMSE
quartz	0.60911	12.9194	17.1179
feldspar	0.66617	7.82615	10.8192
kaolinite	0.41981	2.38434	3.06834
illite/muscovite	0.36545	4.08937	5.57883
montmorillonite	0.38808	1.75248	2.45341
vermiculite	0.107	0.55109	1.07136
clinochlore	-0.0927	2.32926	4.64669
total clays	0.22112	7.47686	10.3437
calcite + dolomite	0.40912	9.67472	12.9225
gypsum	0.27236	3.12516	8.66784
FeD	0.32168	0.278	0.38374
FeS	0.31076	0.65116	0.87588
FeA	0.15289	0.0173	0.02568
FeM	0.17847	0.02014	0.02701
% FeD/FeT	0.66427	4.7031	7.44324
% FeS/FeT	0.60006	5.26696	8.11593
% FeA/FeT	0.11206	0.63972	0.87967
%FeM/FeT	0.16248	0.79108	1.11748

Table 1. Error metrics for each predicted target plotted in Figure 7.

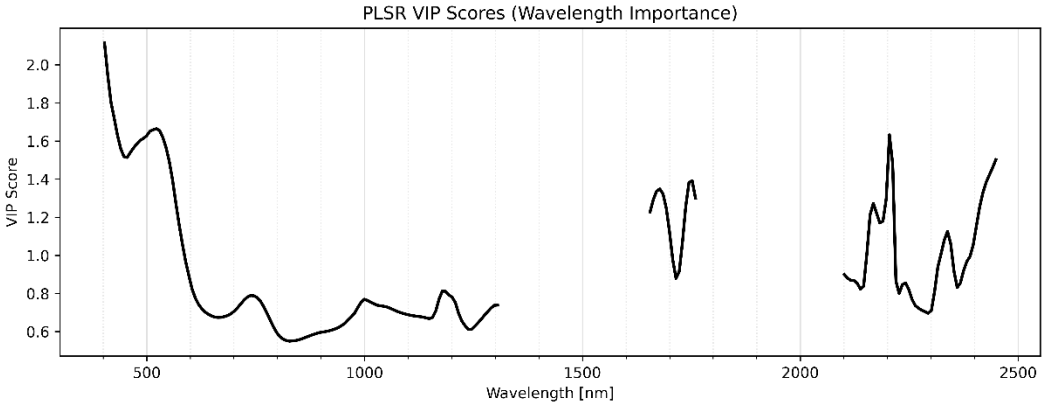

Figure 8. Variable Importance in Projection (VIP) scores as a function of wavelength for the final PLSR model, highlighting spectral regions contributing most strongly to model predictions. Elevated VIP values align with diagnostically meaningful absorption features associated with iron oxides, gypsum, Al-OH–bearing phyllosilicates (e.g., kaolinite and illite), and carbonate minerals.

3.2 Performance on EMIT data

3.2.1 Comparison with EMIT Retrievals for Training-Consistent Spectra

Table 2 presents a comparison of mineral phase detection for the seven selected EMIT spectra described in Section 2.3.2. We compare L2B mineral detections and PLSR mineral detections with the actual lab-detected mineralogy (when available) or expert analysis of spectral features to identify likely constituent minerals.

When evaluating the four samples with ground-truth quantitative XRD measurements, the EMIT L2B product generally captures the dominant mineral constituents but does not consistently resolve all coexisting mineral phases. For example, in sample MOJ011, gypsum is predicted alongside clay minerals by the PLSR model and is present in the laboratory mineralogy, yet it is not identified in the L2B output. This behavior is expected in complex mineral assemblages containing clays, carbonates, and sulfates, where overlapping diagnostic absorption features can complicate endmember-based spectral matching. Because the L2B algorithm identifies minerals through comparison to a spectral library, its performance may be limited when spectra represent mixtures that are not well represented by available library endmembers.

A similar pattern is observed for iron-bearing minerals. Laboratory analyses indicate that several samples contain multiple iron-bearing phases, including combinations of hematite, goethite, magnetite, and maghemite (e.g., JOR021), as well as mixtures of iron oxides like hematite and other iron-bearing minerals such as the amphibole pargasite (e.g., MOJ011 and USA053). In these cases, the EMIT L2B product typically reports a single iron-bearing mineral detection within its Group 1 classifications, whereas the laboratory mineralogy and PLSR results indicate the presence of multiple iron-bearing phases. These results suggest that the PLSR approach may be better suited for identifying constituents of complex mineral mixtures.

Sample	Method	Quartz/Feldspar	Iron-bearing minerals	Clay minerals	Carbonate Minerals	Evaporate Minerals
MOR025	Actual (Lab)	quartz, feldspar	hematite	kaolinite, illite/muscovite	calcite	--
	Detected (EMIT L2B)	--	nanohematite	illite/muscovite	--	--
	Detected (PLSR)	quartz, feldspar	hematite/goethite, magnetite	kaolinite, illite/muscovite	calcite	--
JOR021	Actual (Lab)	quartz	hematite/goethite, magnetite	kaolinite, illite/muscovite	calcite + dolo	--
	Detected (EMIT L2B)	--	hematite + goethite	kaolinite, illite/muscovite	--	--
	Detected (PLSR)	quartz, minor feldspar	hematite/goethite, magnetite	kaolinite, illite/muscovite	calcite	--
MOJ011	Actual (Lab)	quartz, feldspar	pargasite, maghemite	illite, montmorillonite	dolomite	gypsum, halite
	Detected (EMIT L2B)	--	hematite + goethite	montmorillonite	dolomite	--
	Detected (PLSR)	quartz, feldspar	hematite/goethite, magnetite	illite	dolomite	gypsum
USA053	Actual (Lab)	feldspar	hematite, magnetite/maghemite, pargasite	illite/muscovite	calcite	--
	Detected (EMIT L2B)	--	cummingtonite	montmorillonite	--	--
	Detected (PLSR)	feldspar	hematite/goethite, magnetite	illite, minor kaolinite, montmorillonite	calcite	--
Asian Iron-rich clays	Expert spectral analysis	likely both	chlorite	chlorite, illite/muscovite	--	--
	Detected (EMIT L2B)	--	chlorite (thuringite)	illite/muscovite		
	Detected (PLSR)	quartz + feldspar	hematite/goethite, magnetite	illite	--	--
Bright Saharan Sands	Expert spectral analysis	quartz	hematite/goethite	kaolinite	--	--
	Detected (EMIT L2B)	--	goethite	kaolinite, illite/muscovite	--	--
	Detected (PLSR)	quartz	hematite/goethite	kaolinite	calcite	
Saharan Carbonate-Rich	Expert spectral analysis	likely both	hematite/goethite	--	calcite	--
	Detected (EMIT L2B)	--	goethite	--	calcite	--
	Detected (PLSR)	quartz	hematite/goethite, magnetite	illite	calcite	gypsum

Table 2. Comparison of mineral phase detection for EMIT L2B and PLSR for each sample. Minerals detected with XRD are listed for samples when available. For samples without available ground-truth laboratory mineral identification, minerals thought to be present based on expert spectral analysis are listed.

To evaluate quantitative mineralogy performance, Table 3 compares laboratory-derived mineral abundances with predictions from the EMIT L3 mineralogy product and the PLSR model for the target mineral phases.

Overall, the results show a mixed performance pattern that depends on mineralogical composition. In quartz-rich samples from the Sahara and Jordan, the EMIT L3 product systematically overestimates quartz and feldspar abundances, whereas the PLSR model provides lower absolute errors (approximately 2–9 wt% for PLSR compared to ~20 wt% for EMIT L3). In contrast, for feldspar-dominated samples from the Mojave and USA sites, EMIT L3 performs somewhat better, with lower absolute errors (~10 wt%) compared to PLSR (~15 wt%).

For clay minerals, both approaches show broadly comparable performance, with absolute errors typically ranging from <1 to 10 wt% across samples. In contrast, carbonate minerals (calcite and dolomite) are more accurately captured for all four samples by the PLSR model, which yields a mean absolute error of approximately 4.5 wt% compared to ~24.5 wt% for EMIT L3. Finally, both methods perform similarly for crystalline iron oxide phases, with consistent absolute errors on the order of ~0.25 wt%.

Sample ID	Method	Quartz + Feldspar	Kaolinite	Illite + Muscovite	Montmorillonite	Vermiculite	Chlorite	Calcite + Dolomite	Gypsum	Hematite	Goethite	FeD	FeM
MOR025	Lab	63.454	5.590	8.179	0.000	0.000	0.615	21.646	0.054	--	--	0.931	0.0417
	EMIT L2B*	--	no	yes	no	no	no	no	no	yes	no		
	EMIT L3	81.779	0.147	7.599	0.586	0.000	0.002	8.347	0.000	0.868	0.879	1.160	--
	PLSR	61.867	5.064	7.289	1.388	0.362	3.066	15.226	0.000	--	--	0.939	0.042
JOR021	Lab	61.600	5.500	5.700	0.000	9.000	0.000	18.100	0.000	--	--	0.828	0.028
	EMIT L2B*	--	yes	yes	no	no	no	no	no	yes	yes		
	EMIT L3	86.833	6.218	5.494	0.024	0.000	0.007	0.395	0.000	0.669	0.577	0.831	--
	PLSR	52.389	8.274	8.538	3.339	1.540	3.481	18.648	0.000	--	--	1.098	0.056
MOJ011	Lab	47.100	0.000	13.900	7.600	0.000	2.900	6.900	2.300	--	--	0.323	0.086
	EMIT L2B*	--	no	no	yes	no	no	yes	no	yes	yes		
	EMIT L3	57.965	0.775	9.176	0.177	0.001	0.741	30.583	0.003	0.248	0.623	0.565	--
	PLSR	31.695	2.553	10.997	3.403	0.270	3.886	21.291	7.315	--	--	0.537	0.065
USA053	Lab	46.300	3.400	27.700	0.000	0.000	0.000	14.100	0.000	--	--	0.305	0.085
	EMIT L2B*	--	no	no	yes	no	no	no	no	no	no		
	EMIT L3	40.068	0.034	1.885	0.707	0.019	0.001	57.466	0.029	0.137	0.121	0.172	--
	PLSR	32.903	5.304	17.027	5.405	0.904	4.255	15.081	0.000	--	--	0.843	0.081

* detections only

Table 3. Quantitative mineralogical results for selected sampling locations. For each sample, ground-truth laboratory-derived mineral abundances from in situ analyses

are shown in grey (“Lab”), alongside corresponding EMIT L2B mineral detections, EMIT L3 aggregated mineralogy and PLSR-predicted mineral abundances for each mineral phase. For quantitative abundances, predicted values within 5% relative error of the lab value are highlighted in green and values within 15% relative error of the lab value are highlighted in pale green.

Figure 9 compares modeled and measured mineral abundances for the four samples with ground-truth mineralogy, with points colored by mineral group. Open symbols represent abundances derived from the EMIT L3 product, whereas filled symbols represent abundances predicted by the PLSR model. The plot provides a mineral-specific assessment of model performance relative to laboratory measurements.

For the dominant mineral phases, particularly quartz, calcite, and kaolinite, PLSR-derived abundances generally plot closer to the 1:1 line than the corresponding EMIT L3 estimates, indicating improved agreement with the ground-truth mineralogy. For lower-abundance clay minerals, including illite and montmorillonite, both approaches exhibit larger deviations from the 1:1 relationship and comparable overall performance. Iron oxide abundances show similar scatter for both methods, with neither approach demonstrating a clear advantage relative to the laboratory measurements. Overall, the largest differences between the two methods are observed for the major mineral constituents, where the PLSR model more closely reproduces measured abundances.

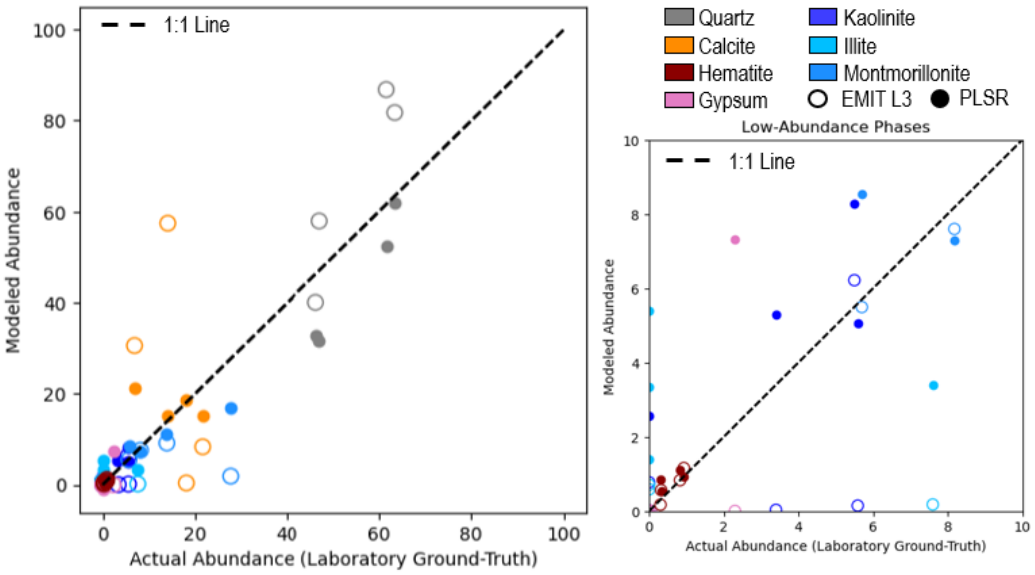

Figure 9. Modeled vs. actual mineral abundances, colored by mineral family/type, for each of the 4 samples with ground truth mineralogy data. Open circles represent abundance results from EMIT L3 and filled circles are abundances derived from PLSR.

3.2.2 Extrapolation to Non-Training Spectral Types

For these three EMIT observations, laboratory ground truth mineralogy is not available; however, the spectral signatures can be directly evaluated in terms of consistency between observed reflectance features and model-derived mineral identifications. This allows a qualitative assessment of how the PLSR predictions compare with the EMIT mineralogy retrievals in spectral regimes not represented in the training dataset.

Table 4 presents quantitative results for the three spectra, comparing PLSR-derived mineral abundances with the EMIT L3 aggregated mineralogy product. In the absence of ground truth, the comparison focuses on whether the PLSR model yields materially different mineralogical interpretations relative to the EMIT retrievals in these out-of-training-set cases and how these each relate to expert spectral analysis of phases present.

Several patterns of agreement emerge. For the Asian dust sample, both models suggest relatively high quartz and feldspar abundances, likely indicating higher feldspar and a composition more consistent with iron-bearing clay dominance, potentially reflecting the contribution of chlorite and other Fe-bearing phyllosilicates (rather than coarse-grained quartz-rich sand). In contrast, the carbonate-rich sample shows substantially lower quartz and feldspar abundances in the PLSR results compared to EMIT L3, differences in prediction of silicate phases combined with bright, carbonate-dominated spectral signatures. Clay mineral abundances are broadly consistent between the two approaches. Notably, PLSR predicts calcite and dolomite in the Saharan sand sample, whereas these phases are not represented in the corresponding EMIT L3 output. Iron oxide abundances (FeD) are of similar order of magnitude across methods for most samples, although PLSR yields lower values for the Asian dust sample, likely reflecting attribution of low VNIR reflectance and absorption features to iron-bearing clays rather than discrete crystalline iron oxide phases such as hematite or goethite.Overall, while absolute agreement between methods is not expected in these extrapolation cases, the comparison indicates some differences in how each approach partitions spectrally overlapping features and distributes mineral abundances across compositionally complex mixtures.

Sample	Result Type	Method	Quartz + Feldspar	Kaolinite	Illite + Muscovite	Montmorillonite	Vermiculite	Chlorite	Calcite + Dolomite	Gypsum	Hematite	Goethite	FeD	FeM
Asian Iron-rich Clays	Detections only	Expert	yes	--	yes	--	--	yes	--	--	--	--	--	--
		EMIT L2B	--	--	yes	--	--	yes	--	--	--	--	--	--
	Detections + quantifications	EMIT L3	80.331	0.002	10.203	0.001	0.000	4.082	0.447	0.000	0.228	1.836	1.313	--
		PLSR	70.409	1.497	11.752	1.958	0.000	3.175	0.000	0.000	--	--	0.394	0.064
Saharan Carbonate-Rich	Detections only	Expert	yes	--	yes	--	--	--	yes	--	yes	yes	yes	--
		EMIT L2B	--	--	--	--	--	--	yes	--	--	yes	--	--
	Detections + quantifications	EMIT L3	35.364	0.154	1.260	5.460	0.000	0.001	49.774	6.176	0.669	0.577	0.831	--
		PLSR	12.044	1.176	5.022	0.269	0.000	3.811	38.002	16.720	--	--	0.389	0.051
Bright Saharan Sands	Detections only	Expert	yes	yes	--	--	--	--	--	--	yes	yes	yes	--
		EMIT L2B	--	yes	yes	--	--	--	--	--	--	yes	--	--
	Detections + quantifications	EMIT L3	79.713	7.457	11.122	0.000	0.001	0.000	0.116	0.003	0.650	1.244	1.236	--
		PLSR	80.665	9.026	0.000	0.000	1.796	2.994	26.223	0.000	--	--	1.113	0.027

Table 4. Mineral abundance model results for EMIT spectra outside the training data domain, comparing EMIT Level-3 aggregated mineral abundances with PLSR-derived predictions.

4. Discussion

4.1 Model performance

For this dataset, PLSR predictive performance is strongest for variables that are known (from Chapter 3 and established spectral mineralogy literature) to be closely linked to iron oxide absorptions. This includes both iron-bearing parameters such as FeD fraction and FeS fraction, and, notably, non-iron-oxide-bearing silicate minerals such as quartz and feldspar. The latter result is particularly significant, as quartz and feldspar do not exhibit strong diagnostic absorptions in the visible to shortwave infrared (VSWIR) and are therefore not directly observable in a traditional feature-based sense. Instead, their predictability from VSWIR reflectance spectra suggests that they are being inferred from overall albedo, or indirectly through covariation with iron oxide spectral features or their silicate weathering products. This is particularly notable because it implies that VSWIR spectra carry robust predictive information on bulk mineralogical composition of minerals that are not easily accessible through traditional feature-based approaches or optical constants-based approaches.

Clay and carbonate mineral estimation showed consistently poorer performance relative to iron oxide-linked parameters. While clay minerals such as kaolinite, illite, and smectite exhibit diagnostic Al-OH absorptions in the shortwave infrared, PLSR model performance was less accurate and more variable than both iron-bearing phases and non-absorbing silicates such as quartz and feldspar. Notably, performance for aggregate measures such as total clay and carbonate abundance was also comparatively poor, suggesting that the challenge is not limited to inter-species discrimination but extends to bulk group estimation. This suggests that broad continuum-scale spectral characteristics across many wavelengths dominate the retrieval behavior, resulting in comparatively robust prediction of minerals such as quartz, feldspar, and iron oxides that strongly influence overall spectral shape and albedo over many channels. In contrast, minerals characterized by narrower

absorption features spanning only a limited number of wavelength channels are less reliably predicted, likely because these localized spectral signatures are underweighted relative to broader continuum variability within the modeling framework. Future work should focus on disentangling clay-carbonate spectral overlap using hybrid approaches that explicitly account for mineralogical co-occurrence in these spectrally complex sediment systems.

4.2 Implications for application to the global EMIT dataset

Application of the trained PLSR model to EMIT reflectance data indicates that the approach can provide more complete mineral identification than the EMIT Level 2B product while yielding mineral abundance estimates that are comparably or more accurate than the EMIT Level 3 product for several key dust-source minerals. Note that this comparison involves point-source ground-truth measurements, but EMIT L3 products are derived from aggregated spectra from the 0.5° grid, such that the L3 estimates reflect contributions from additional materials within the scene, diluting the strength of target mineral signals. In particular, the PLSR retrievals appear to provide improved sensitivity to carbonate abundance than the EMIT L3 data, while slightly improving quartz/feldspar prediction precision and maintaining comparable performance in clay and iron oxide estimation. Although these minor phases contribute less to the bulk mineralogy, their identification is important for reconstructing sediment provenance, weathering history, evaporitic processes, and dust-climate interactions. The ability to resolve more compositionally complex assemblages therefore offers a more complete representation of dust source mineralogy and environmental conditions within arid landscapes.

Importantly, model behavior remained consistent when applied to spectra that extend beyond the compositional variability represented in the training dataset, including Asian dust samples, bright quartz-rich Saharan sands, and carbonate-rich Saharan sediments. While direct ground-truth validation is not currently possible for these independent datasets, the similarity between PLSR-derived mineralogy and EMIT

Level 3 retrievals and to expert-analyzed spectra suggests that the empirical model retains reasonable predictive capability outside the training domain. Notably, the agreement observed for these previously unseen sediment types is comparable to that obtained for EMIT spectra associated with the original training dataset, indicating that the models are not simply reproducing known examples through overfitting, though future work is needed to ground truth the composition of all the dust source region endmembers. The relatively consistent performance across distinct sedimentological and mineralogical environments further supports the robustness and transferability of the spectral-mineralogical relationships captured by the PLSR framework.

These results suggest that PLSR-based quantitative mineral retrievals may provide a valuable complementary tool for current and future EMIT mineral dust applications. The enhanced sensitivity to carbonate abundance is particularly relevant given the importance of carbonate-rich dust for atmospheric alkalinity, cloud interactions, and biogeochemical cycling. In addition, the approach enables quantitative mineralogical characterization at the spatial resolution of individual EMIT observations, avoiding the need to aggregate mineralogical information over coarse regional grids that may obscure localized heterogeneity in dust source regions. This is especially advantageous for geographically discrete sources such as playas in the southwestern United States, where substantial mineralogical variability can occur over spatial scales much smaller than typical 0.5° climatological grids. Continued expansion of the training dataset to encompass a broader range of sediment compositions and surface conditions should further improve model generalizability and strengthen the utility of this approach for remote sensing investigations of global dust source mineralogy.

4.3 Future work

Several avenues remain for improving predictive performance and physical interpretability for empirical models. First, future work should focus on more

explicit resolution of clay mineral speciation. The relatively poor model performance for spectra of clay-bearing samples here suggests that current spectral information and modeling approaches are insufficient to fully separate phyllosilicate subtypes in mixed natural sediments, motivating the need for more physically informed or spectrally constrained approaches. More sophisticated modeling architectures should be explored, including multi-stage or hierarchical frameworks in which mineral classes and phases are first detected before attempting quantitative abundance estimation.

Although this study focuses on a Partial Least Squares Regression framework, nonlinear ensemble methods (e.g., Random Forest) remain a potentially powerful alternative for modeling complex, non-linear relationships between hyperspectral reflectance and sediment mineralogy. In preliminary testing, such methods can capture higher-order interactions and non-linear spectral responses that are not explicitly represented in linear latent-variable models. However, these approaches require sufficiently large and diverse training datasets to avoid overfitting and ensure robust generalization.

Model performance is ultimately constrained by the representativeness of the training dataset. Although the current dataset (~140 samples) provides strong compositional coverage of major dust-source mineral assemblages, it does not fully capture the spectral variability observed across global dust source environments. As highlighted in earlier chapters, several spectral endmembers remain underrepresented in the available in situ and laboratory measurements, potentially limiting model generalizability when applied to spectra that fall outside the training domain. Future work should therefore prioritize targeted expansion of the training dataset to include these underrepresented mineralogical and spectral endmembers, either through additional field sampling or through physically realistic data augmentation strategies. Increasing the size and diversity of the training dataset would support development of a more robust empirical framework for

mineralogical prediction from global hyperspectral observations and enable more rigorous evaluation of nonlinear modeling approaches, including ensemble methods.

5. Conclusions

We leverage a paired dataset of hyperspectral reflectance spectra and laboratory-derived measurements of mineral abundance, iron chemistry, and grain size to develop empirical models that predict quantitative mineral composition directly from spectra. For this work, we focus on Partial Least Squares Regression. The highest-performing model achieves fair ($R^2 \sim 0.6$) predictive skill for quartz, feldspar, and iron oxide weight percent, while moderate performance ($R^2 \sim 0.3$) is achieved for clay and carbonate minerals. In general, models are effective at identifying the presence of these phases but tend to underestimate higher clay abundances, indicating limitations in utilizing the metal-OH and metal-CO_3 absorptions in conjunction with other spectral ranges for mineral identification and quantification in clay-rich systems.

Application of the trained PLSR model to EMIT reflectance data shows overall consistency with the EMIT Level 2B (L2B) mineral identification product in terms of detecting major mineral constituents, while providing a more complete representation of mineral assemblages. In particular, the PLSR model more consistently identifies multi-mineral mixtures affecting the same spectral region (i.e., amphiboles and iron oxides; clay minerals, carbonates, and gypsum). The PLSR model also provides estimates of silicate phases such as quartz and feldspar that are not explicitly reported in the L2B product due to the absence of diagnostic VSWIR absorption features. Relative to the EMIT Level 3 (L3) aggregated mineral abundance product, PLSR-derived estimates show comparable to improved quantitative performance across multiple mineral groups when evaluated against ground-truth mineralogy. Notably, improved agreement is observed for calcite and

dolomite, where PLSR yields more accurate abundance estimates than L3 in all samples. Comparable performance is observed for clay minerals and iron oxides.

For applications requiring finer spatial resolution than the EMIT L3 product's 0.5 x 0.5 degree grid, as well as for analysis of individual spectra or localized dust source regions, the empirical PLSR framework provides a useful complement to both EMIT standard products. Overall, these results demonstrate that empirical spectral-mineral models can enhance both the completeness of mineral identification (relative to L2B) and the quantitative accuracy of abundance estimates (relative to L3), particularly in compositionally complex or mixed-mineral regimes. Future work should focus on expanding training dataset diversity, especially for underrepresented high-abundance clay- and carbonate-bearing compositions, and on exploring more flexible modeling frameworks to further improve performance in globally heterogeneous spectral environments.

References

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Clark, R. N., King, T. V., Klejwa, M., Swayze, G. A., & Vergo, N. (1990). High spectral resolution reflectance spectroscopy of minerals. Journal of Geophysical Research: Solid Earth, 95(B8), 12653-12680.

Clark, R. N., Swayze, G. A., Livo, K. E., Brodrick, P. G., Noe Dobrea, E., Vijayarangan, S., ... & EMIT and TREX teams. (2024). Imaging spectroscopy: Earth and planetary remote sensing with the PSI Tetracorder and expert systems from rovers to EMIT and beyond. The Planetary Science Journal, 5(12), 276.

Green, R. O., Mahowald, N., Ung, C., Thompson, D. R., Bator, L., Bennet, M., ... & Zan, J. (2020, March). The Earth surface mineral dust source investigation: An Earth science imaging spectroscopy mission. In 2020 IEEE aerospace conference (pp. 1-15). IEEE.

González-Romero, A., González-Flórez, C., Panta, A., Yus-Díez, J., Reche, C., Córdoba, P., ... & Pérez García-Pando, C. (2023). Variability in sediment particle size, mineralogy, and Fe mode of occurrence across dust-source inland drainage basins: the case of the lower Drâa Valley, Morocco. Atmospheric chemistry and physics, 23(24), 15815-15834.

González-Romero, A., González-Flórez, C., Panta, A., Yus-Díez, J., Córdoba, P., Alastuey, A., ... & Pérez García-Pando, C. (2024). Characterization of the particle size distribution, mineralogy, and Fe mode of occurrence of dust-emitting sediments from the Mojave Desert, California, USA. Atmospheric chemistry and physics, 24(16), 9155-9176.

Hajaj, S., El Harti, A., Pour, A. B., Jellouli, A., Adiri, Z., & Hashim, M. (2024). A review on hyperspectral imagery application for lithological mapping and mineral prospecting: Machine learning techniques and future prospects. Remote Sensing Applications: Society and Environment, 35, 101218.

Hapke, B. (2008). Bidirectional reflectance spectroscopy: 6. Effects of porosity. *Icarus*, *195*(2), 918-926.

Hapke, B. (2012). Theory of reflectance and emittance spectroscopy. Cambridge university press.

Jahoda, P., Drozdovskiy, I., Payler, S. J., Turchi, L., Bessone, L., & Sauro, F. (2021). Machine learning for recognizing minerals from multispectral data. Analyst, 146(1), 184-195.

Lapotre, M. G., Ehlmann, B. L., & Minson, S. E. (2017). A probabilistic approach to remote compositional analysis of planetary surfaces. *Journal of Geophysical Research: Planets*, *122*(5), 983-1009.

Longtin, D. R., Shettle, E. P., Hummel, J. R., & Pryce, J. D. (1988). *A wind dependent desert aerosol model: Radiative properties* (Vol. 88, No. 112). Air Force

Geophysics Laboratory, Air Force Systems Command, United States Air Force.

Mustard, J. F., & Pieters, C. M. (1987). Quantitative abundance estimates from bidirectional reflectance measurements. Journal of Geophysical Research: Solid Earth, 92(B4), E617-E626.

Querry, M. R. (1985). *Optical constants* (No. CRDCCR85034).

Thompson, D. R., Green, R. O., Bradley, C., Brodrick, P. G., Mahowald, N., Dor, E. B., ... & Zandbergen, S. (2024). On-orbit calibration and performance of the EMIT imaging spectrometer. Remote Sensing of Environment, 303, 113986.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. Chemometrics and intelligent laboratory systems, 58(2), 109-130.

Zhang, X. L., Wu, G. J., Zhang, C. L., Xu, T. L., & Zhou, Q. Q. (2015). What is the real role of iron oxides in the optical properties of dust aerosols?. *Atmospheric Chemistry and Physics*, *15*(21), 12159-12177.

Chapter 5

SUMMARY, IMPLICATIONS, AND OUTSTANDING QUESTIONS

This dissertation addresses a central challenge in imaging spectroscopy and Earth system science: translating increasingly abundant hyperspectral observations into quantitative information about surface mineralogy relevant to radiative forcing and biogeochemical cycling processes. This work advances the interpretation of global dust mineralogy from imaging spectroscopy observations through the development and analysis of coupled spectral, mineralogical, and grain size datasets focused on arid mineral dust source regions.

5.1 Summary of contributions and scientific implications

This work developed and analyzed a globally representative dataset of hyperspectral reflectance observations from the EMIT mission to characterize the spectral diversity present within Earth's arid dust source regions. By examining distributions of mineral absorption presence and strength, as well as broadband albedo variability across geographically distinct dust-producing regions, this work supports future efforts to map surface mineralogy and dust source properties at regional and global scales. Importantly, we show that dust source regions exhibit highly heterogeneous spectral and mineralogical properties. The Sahara and Asian deserts dominate global dust emissions (Kok et al., 2021); these regions differ substantially in both surface albedo and mineralogical composition. Saharan dust sources are bright (VIS albedo ~0.4) and characterized by hematite, goethite, and kaolinite absorptions, while Asian desert dust sources are lower-reflectance (VIS albedo ~0.25) and exhibit signatures of illite, montmorillonite, and carbonates. Outside of these dominant dust source region spectral types, dust source region endmembers, including in the Middle East, North America, and the southern hemisphere exhibit evaporite and hydrous mineral signatures, absorptions due to less-common iron-bearing clays like chlorite, and strong carbonate features. Dust source composition and radiative properties are not

globally uniform. As a result, Earth system models cannot assume uniform dust source composition, as we demonstrate substantial spatial variability in mineralogy and spectral properties that depends on source region and sediment physical characteristics. In addition, the bareground hyperspectral dataset developed in this study provides a direct observational constraint on Earth surface mineralogy and radiative properties across dust source regions. Previous work by Braghiere et al. (2023) demonstrates that replacing broadband albedo parameterizations with hyperspectral information significantly improves estimates of surface radiative forcing. Building on this, the dataset presented here offers an expanded basis to further constrain land surface radiative properties and reduce uncertainty in modeled surface–atmosphere radiative exchange.

In addition to the satellite-based analysis, this dissertation developed a novel in situ spectral database containing coupled measurements of VSWIR reflectance spectra, quantitative mineralogy, and grain size distributions of dust source sediment samples from four dust source regions. Here, we demonstrate how compositional variability and particle size effects influence reflectance spectra of natural sediments of intimately mixed mineral constituents. Collectively, these results demonstrate that the spectral detectability and apparent abundance of iron oxides is strongly conditional on sediment physical properties, particularly composition and grain size of the matrix. Iron oxides are most effectively detected and their abundance most accurately reflected in fine-grained, clay-rich sediments, where diagnostic absorptions scale approximately linearly with bulk abundance. In contrast, in coarse-grained, sandy sediments, iron oxides can produce disproportionately strong absorption features even at low bulk abundances due to enhanced multiple scattering and increased optical path length within bright quartz-rich matrices. Under these conditions, spectral signatures may therefore overestimate iron oxide abundance if matrix grain size effects are not explicitly accounted for. These findings have important implications for remote sensing-based dust source mineralogy. Iron oxides can be reliably inferred from diagnostic absorption band strengths, but only under

specific sedimentological conditions, particularly in fine-grained, clay-rich materials rather than sandy materials. For example, Chapter 2 shows that Saharan sands exhibit some of the strongest iron oxide absorption features and are key contributors to dust emissions, yet they are unlikely to be the most iron-rich dust source sediments globally in terms of bulk abundance. This highlights the need for joint constraints on both composition and physical properties, such as quartz abundance and grain size, in order to produce robust mineralogical estimates from reflectance spectra.

This work further demonstrated the potential utility for empirical approaches like partial least squares regression in estimating the abundance of several mineral phases from reflectance spectra, particularly for those influencing the whole spectral wavelength range and albedo properties, e.g. quartz, feldspars, and iron oxides. The observed relationships between spectral properties and mineral abundance indicate that quantitative mineralogical retrieval from VSWIR observations is feasible with continued refinement of methods, and may be superior to feature fitting or forward modeling approaches. At the same time, the results highlight important limitations and remaining challenges associated with more compositionally and spectrally complex phases, particularly clay minerals and mixed mineral assemblages where vibrational absorptions overlap and occur over a relatively small wavelength range. Additionally, the spectra should be well-spanned by the training set, which is not yet complete for Earth's dust source regions. These results underscore the inherent complexity of natural soil and sediment spectra and emphasize the need for continued development of physically informed and empirically robust interpretation methods.

The implications of this work extend beyond mineralogical mapping alone. Mineral dust composition strongly influences the radiative properties, atmospheric transport behavior, and biogeochemical impacts of dust aerosols, particularly through the abundance and speciation of iron-bearing minerals. Improving the interpretation of mineralogy from remotely sensed observations therefore contributes directly to broader efforts to better constrain dust-climate interactions within Earth system models. More broadly, the methods and datasets developed in this dissertation

provide a foundation for future imaging spectroscopy studies focused on dryland ecology, soil surface characterization, land degradation, and planetary surface composition. As hyperspectral remote sensing missions continue to expand in spatial coverage and spectral capability, the ability to robustly interpret spectra of natural mixtures will become increasingly important across both Earth and planetary sciences.

5.2 Remaining questions and future work

5.2.1 Expanding Empirical Mineralogical Modeling

Although this work demonstrates the promise of empirical approaches for retrieving quantitative mineralogical information from reflectance spectra, substantial opportunities remain for improving both the accuracy and generalizability of these methods. One avenue for improvement is the continued expansion of joint spectral and mineralogical training datasets spanning a broader diversity of environmental settings, sediment compositions, and grain size distributions. Natural soils and sediments are compositionally heterogeneous and often exhibit strong covariance among mineral phases, weathering states, and physical properties. Expanding the geographic and compositional diversity represented within spectral databases will therefore be essential for developing models that are robust across the full range of conditions observed within global drylands.

Future work would particularly benefit from improved characterization of iron mineralogy and iron speciation within dust source sediments as iron-bearing minerals exert a disproportionately strong influence on visible and near-infrared reflectance properties and play an important role in determining the radiative impacts of mineral dust aerosols. However, iron is commonly present in multiple mineral phases and oxidation states, including hematite, goethite, ferrihydrite, structural iron within clays, and nanophase or coating materials, each of which influences spectral behavior differently. Our collaborators in this work exerted

particular efforts to constrain the phases containing iron, but this is not commonly reported. Better constraints on iron speciation in future sample sets would improve both mineralogical interpretation and the representation of dust optical properties within Earth system models. Future studies integrating spectroscopy with complementary laboratory techniques such as Mössbauer spectroscopy or X-ray Absorption Near-Edge Structure (XANES) may provide valuable pathways toward disentangling these effects and should be compared with wet chemistry speciation determinations used here.

This work also highlights the persistent challenge of interpreting clay mineral mixtures from VSWIR spectra. Compared to quartz, feldspars, and iron oxides, predictive relationships for clay and carbonate minerals were generally weaker and more variable. However, this complexity likely reflects the inherently difficult nature of clay and carbonate spectral behavior within natural sediments as well as limitations unique to the modeling approaches applied here. Clay minerals frequently exhibit overlapping absorption features, variable hydration states, and interstratification. But characterization is critical because clay and carbonate minerals commonly host iron and other chemically important elements relevant to dust biogeochemistry and nutrient delivery. Improving the spectral characterization and quantitative retrieval of clay and carbonate mineral assemblages therefore remains an important area for future work in both Earth system and planetary remote sensing applications.

5.2.2 From surface sediment to airborne dust

An important limitation of this work in constraining quantitative mineralogy as it pertains to mineral dust aerosols is that the analyzed samples represent bulk surface sediments, rather than the size-resolved fraction of particles that ultimately become entrained into the atmosphere as airborne dust. Although surface sediments provide critical information about source region composition, emitted dust aerosols are likely compositionally distinct from their parent materials due to preferential entrainment

of fine particles, aerodynamic sorting, aggregate breakdown, and other transport-related fractionation processes.

Understanding these fractionation processes represents an important next step for improving the interpretation of remotely sensed dust source properties and their incorporation into climate models. This was not addressed by our work but is a key next step from dust source region surface mineralogy retrievals and has been treated by others (Perez Garcia-Pando et al., 2016; Perlwitz et al., 2015; Di Biagio et al., 2019). In particular, future work should investigate how mineral abundances vary as a function of particle size within both source sediments and emitted dust fractions. Iron oxides, clays, and other fine-grained phases may become preferentially enriched within airborne dust relative to coarser parent sediments, potentially altering the optical, radiative, and biogeochemical behavior of transported aerosols. Field sampling, laboratory dust generation experiments, wind tunnel studies, and size-resolved mineralogical analyses would help improve understanding of how surface composition translates into atmospheric dust composition. Acquiring spectra of the liberated dust fraction and the bulk source sediment whence it came would improve our ability to characterize systematic differences in the radiative processes of bulk vs fractionated aerosol sediments. Integrating satellite observations of active dust plumes and airborne sampling campaigns could help directly connect source region mineralogy with transported aerosol properties. Such efforts would improve constraints on dust radiative forcing, atmospheric transport modeling, and nutrient deposition processes while helping bridge the gap between surface spectroscopy and atmospheric dust characterization.

5.2.3 Future Remote Sensing Missions and Planetary Applications

The continued growth of the imaging spectroscopy field presents substantial opportunities for continuing and extending the work presented in this dissertation. Upcoming Earth orbital missions (e.g., Surface Biology and Geology, SBG), along with future airborne and orbital imaging spectrometers, will provide increasingly

frequent, spatially extensive, and spectrally detailed observations of Earth's surface. These datasets will significantly expand the ability to investigate dryland processes, mineral dust source regions, soil composition, vegetation dynamics, and surface-atmosphere interactions at regional to global scales. At the same time, the growing volume and complexity of hyperspectral observations will continue to increase the need for robust methods capable of quantitatively interpreting natural spectral variability.

Although this dissertation focuses primarily on mineral dust source regions, the approaches developed here have applications extending far beyond dust mineralogy alone. Improved understanding of soil and sediment spectral behavior may support future studies of dryland ecology, vegetation and soil interactions, biological soil crusts, land degradation, and surface hydrology. Imaging spectroscopy also provides important opportunities for monitoring ecosystem change and human impacts within vulnerable arid and semiarid environments undergoing rapid environmental change.

Furthermore, many of the challenges explored throughout this dissertation are equally relevant to planetary remote sensing beyond Earth. Spectral mixing, particle size effects, compositional ambiguity, and quantitative mineralogical retrieval remain central challenges in the interpretation of remotely sensed observations of planetary surfaces throughout the solar system. As imaging spectroscopy capabilities continue to expand for both Earth and planetary missions, improved understanding of how mineralogical and physical properties influence reflectance spectra will remain essential for interpreting the surfaces of planets, moons, asteroids, and other planetary bodies. In this sense, the methods and questions explored in this dissertation contribute not only to understanding Earth's drylands and dust systems, but also to the broader scientific goal of interpreting planetary surfaces through remote observations.

References

Braghiere, R. K., Wang, Y., Gagné-Landmann, A., Brodrick, P. G., Bloom, A. A., Norton, A. J., ... & Schneider, T. (2023). The importance of hyperspectral soil albedo information for improving Earth system model projections. *Agu Advances*, *4*(4), e2023AV000910.

Di Biagio, C., Formenti, P., Balkanski, Y., Caponi, L., Cazaunau, M., Pangui, E., ... & Doussin, J. F. (2019). Complex refractive indices and single-scattering albedo of global dust aerosols in the shortwave spectrum and relationship to size and iron content. *Atmospheric Chemistry and Physics*, *19*(24), 15503-15531.

Kok, J. F., Adebiyi, A. A., Albani, S., Balkanski, Y., Checa-Garcia, R., Chin, M., ... & Wan, J. S. (2021). Contribution of the world's main dust source regions to the global cycle of desert dust. *Atmospheric Chemistry and Physics*, *21*(10), 8169-8193.

Pérez García-Pando, C., Miller, R. L., Perlwitz, J. P., Rodríguez, S., & Prospero, J. M. (2016). Predicting the mineral composition of dust aerosols: Insights from elemental composition measured at the Izaña Observatory. *Geophysical Research Letters*, *43*(19), 10-520.

Perlwitz, J. P., Pérez García-Pando, C., & Miller, R. L. (2015). Predicting the mineral composition of dust aerosols–Part 1: Representing key processes. Atmospheric Chemistry and Physics, 15(20), 11593-11627.

.