Machine Learning Mid-Infrared Spectral Models for Predicting Modal Mineralogy of CI/CM Chondritic Asteroids and Bennu

L. B. Breitenfeld[1] [iD], A. D. Rogers[1] [iD], T. D. Glotch[1] [iD], V. E. Hamilton[2] [iD], P. R. Christensen[3], D. S. Lauretta[4], M. E. Gemma[5,6] [iD], K. T. Howard[7], D. S. Ebel[5,6], G. Kim[1], A. M. Kling[8] [iD], H. Nekvasil[1] [iD], and N. DiFrancesco[9]

[1]Department of Geosciences, Stony Brook University, Stony Brook, NY, USA, [2]Department of Space Science, Southwest Research Institute, Boulder, CO, USA, [3]School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA, [4]Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ, USA, [5]Department of Earth and Planetary Sciences, American Museum of Natural History, New York, NY, USA, [6]Department of Earth and Environmental Sciences, Lamont Doherty Earth Observatory of Columbia University, Palisades, NY, USA, [7]Kingsborough Community College of the City University of New York, Brooklyn, NY, USA, [8]Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, IN, USA, [9]Department of Atmospheric and Geological Sciences, SUNY Oswego, Oswego, NY, USA

Abstract Planetary surfaces can be complex mixtures of coarse and fine particles that exhibit linear and nonlinear mixing behaviors at mid-infrared (MIR) wavelengths. Machine learning multivariate analysis can estimate modal mineralogy of mixtures and is favorable because it does not assume linear mixing across wavelengths. We used partial least squares (PLS) and least absolute shrinkage and selection operator (lasso), two types of machine learning, to build MIR spectral models to determine the surface mineralogy of the asteroid (101955) Bennu using OSIRIS-REx Thermal Emission Spectrometer (OTES) data. We find that PLS models outperform lasso models. The cross-validated root-mean-square error of our final PLS models (consisting of 317 unique spectra of samples derived from 13 analog mineral samples and eight meteorites) range from ~4 to 13 vol% depending on the mineral group. PLS predictions in vol% of Bennu's average global composition are 78% phyllosilicate, 9% olivine, 11% carbonates, and 6% magnetite. Pyroxene is not predicted for the global average spectrum, though it has been detected in small amounts on Bennu. These mineral abundances confirm previous findings that the composition of Bennu is consistent with CI/CM chondrites with high degrees of aqueous alteration. The predicted mineralogy of two previously identified OTES spectral types vary minimally from the global average. In agreement with previous work, we interpret OTES spectral differences as primarily caused by relative abundances of fine particulates rather than major compositional variations.

Plain Language Summary The OTES instrument onboard the OSIRIS-REx spacecraft collects infrared emission spectra that can, in principle, be used to determine the mineralogy of Bennu, the target asteroid of the OSIRIS-REx mission. However, predicting mineral abundances on remote planetary bodies from infrared spectra is particularly complex when there are fine particles ($<\sim$100 μm) on the surface. To circumvent this problem, we created a training set of mineral mixture spectra acquired under asteroid (vacuum) conditions and used machine learning to create models for mineral abundance predictions on asteroids like Bennu. Our results support previous findings that Bennu has a composition consistent with carbonaceous chondrites, the most primitive meteorites.

1. Introduction

Carbonaceous chondritic meteorites are derived from undifferentiated materials rich in carbonaceous compounds. These meteorites typically contain chondrules; however, with high degrees of aqueous alteration (e.g., CI/CM chondrites), few or no chondrules are present (Browning et al., 1996; Takir et al., 2013). Compared to other carbonaceous chondritic materials, CI and CM chondritic meteorites are composed of primitive materials that are representative of the composition of the proto-planetary disk and record early Solar System processes (e.g., aqueous alteration). Bennu, the target of the OSIRIS-REx (Origins, Spectral Interpretation, Resource Identification, and Security–Regolith Explorer) mission, is a near-Earth asteroid with spectral properties consistent

with a composition analogous to CI and/or CM chondrites (Hamilton et al., 2019, 2021). Because asteroids like Bennu provide important information about the building blocks of the early Solar System, the development of methods for quantitative remote mineralogical analysis is desirable. This paper focuses on constructing mid-infrared (MIR; ~5–50 μm; 2,000–2,200 cm^{-1}) models to predict fine and coarse particulate mineral abundances of CI/CM chondritic materials for the interpretation of data collected by the OSIRIS-REx Thermal Emission Spectrometer (OTES; Christensen et al., 2018).

There are two main factors of interest that affect the MIR spectra collected by the OTES instrument: particle size and composition. This work undertakes quantifying these factors in an effort to disentangle them. Using the OTES MIR data, Hamilton et al. (2019, 2021) concluded that Bennu is dominated by hydrated phyllosilicates. Additionally, magnetite was identified, and the vol% of anhydrous silicates was determined to be less than ~10 (Section 2.2). In the MIR, coarse and fine particulates behave differently (Section 2.4). Bennu is an asteroid littered with boulders and a few ponds of fine particulates (e.g., Lauretta, DellaGiustina, et al., 2019). However, Hamilton et al. (2021) attributed the cause of subtle differences between OTES spectra across the asteroid to particle size (Section 2.3). This work aims to better understand the role of particle size and composition in the MIR on the asteroid Bennu.

Numerical models have not previously been applied to OTES spectra for the derivation of modal mineralogy. We explore such models here using machine learning multivariate analysis, an alternative approach to traditional linear spectral mixture analysis (LSMA) models that removes the assumption of linear mixing across all wavelengths (Section 3.6). Because such models do not assume linear mixing across wavelengths, both fine and coarse particulate minerals can be modeled together. This is appropriate given the potential for nonlinear mixing of the materials predicted by Hamilton et al. (2021) (Section 2.4). This work utilizes an integrated fine (<50 μm) and coarse (>125 μm) particulate MIR spectral library of mineral mixtures to create machine learning partial least squares (PLS) and least absolute shrinkage and selection operator (lasso) models. PLS and lasso models learn from a tailored suite of laboratory mineral mixture spectra through associated metadata. Applying machine learning models to MIR spectra enables mineral abundance predictions with robust error estimates for Bennu and potentially other asteroids covered in fine or coarse particulate material. Analysis of the returned samples from Bennu will provide the ultimate confirmation of the results in this work.

2. Background

The goal of this work is to improve our ability to model mineralogic compositions of CI/CM chondrites and the asteroid Bennu through MIR data. This is advantageous for the application to laboratory and remote sensing data. To perform the machine learning modeling detailed in this work, a sample suite of materials analogous to the application targets is necessary. Therefore, we have detailed the mineralogy of CI/CM chondrites and Bennu below. The history and challenges of modeling mineralogy using MIR spectroscopy is also provided.

2.1. CI and CM Chondrites

Previous workers have determined mineral abundances of numerous CI/CM chondritic meteorites through quantitative X-ray diffraction (XRD) analysis (e.g., Bland et al., 2004; Howard et al., 2011, 2009; King et al., 2015). These samples are dominated by hydrated, Mg-Fe silicates, with anhydrous silicates, oxides, sulfides, as well as organic compounds. Carbonates, ferrihydrite, sulfates, phosphates, and several other minor phases also occur within these meteorites (Fendrich & Ebel, 2021; Howard et al., 2009, 2011; King et al., 2015). CI chondrites generally consist of 68–84 vol% phyllosilicates, 2–10 vol% magnetite, 2–19 vol% pyrrhotite, and 3–11 vol% anhydrous silicates (e.g., olivine and pyroxene). CM chondrites are generally composed of 67–88 vol% phyllosilicates and 4–31 vol% anhydrous silicates. CI chondrites contain higher saponite, magnetite, and sulfide content, whereas CM chondrites are richer in serpentine, olivine, and pyroxene minerals (Bland et al., 2004; Howard et al., 2011; King et al., 2015). CI/CM chondrites are strongly affected by aqueous alteration on their parent asteroids and therefore commonly have few to no chondrules present (Browning et al., 1996; Takir et al., 2013). Any chondrules originally present within these rocks were altered by hydrothermal fluids that progressively converted anhydrous silicates into hydrated silicates called phyllosilicates (e.g., serpentines and saponite) (McAdam et al., 2015; Tomeoka & Buseck, 1985). The presence of particular mineral species and their abundances can indicate the degree of alteration the meteorite or asteroid experienced.

2.2. Bennu Spectral Properties

B-type asteroids like Bennu (DeMeo et al., 2009) exhibit blue slopes (decreasing with increasing wavelength) in their visible and near-infrared (VNIR) reflectance spectra, with minimal spectral features. A blue VNIR slope can be attributed to fine particulate magnetite and/or organics (Cloutis et al., 2011a, 2011b; Lantz et al., 2018; Trang et al., 2021). Spectra from the OSIRIS-REx Visible and InfraRed Spectrometer (OVIRS; Reuter et al., 2018; Simon et al., 2021) observations of Bennu display a small range of spectral slopes, with the average Bennu spectrum blue-sloped with several distinct features (Hamilton et al., 2019; Simon, Kaplan, Cloutis, et al., 2020; Zou et al., 2021).

OVIRS spectra exhibit a 2.74 μm hydration band that is attributed to phyllosilicates common in CI/CM chondrites (Hamilton et al., 2019, 2021; Praet et al., 2021). Features observed in the 3.4 μm region were attributed to carbon-bearing compounds (organics and carbonate) by Kaplan et al. (2020) and Simon, Kaplan, Hamilton, et al. (2020). Kaplan et al. (2020) identified carbonate veins present at the Nightingale sampling site and other regions of Bennu that were interpreted as primarily calcite, with minor amounts of magnesite, dolomite, and breunnerite.

In the MIR, phyllosilicate (440 cm^{-1}) and magnetite (555, 340 cm^{-1}) features are present in OTES data (Hamilton et al., 2019). A broad bowl-shaped feature (∼1,100–650 cm^{-1}) corresponding with the silicate stretching region is also present within OTES spectra. The lack of an emissivity peak around 523 cm^{-1} indicates that the combined olivine and pyroxene content is less than ∼10 vol% (Hamilton et al., 2021). DellaGiustina et al. (2021) detected pyroxene on the surface of Bennu using the OSIRIS-REx MapCam and OVIRS instruments. The strength and shape of the OTES 440 cm^{-1} feature indicates that phyllosilicates are dominant on Bennu, yet the bulk mineralogy of Bennu has not yet been constrained quantitatively through modeling. This investigation builds on the work of Hamilton et al. (2021) by helping to constrain mineral abundances.

2.3. Bennu Surface Particulates

Hamilton et al. (2021) found that nearly all OTES spectra can be modeled with linear mixtures of two spectral types, called T1 and T2, which are mostly distinguished by shape differences from ∼1,100 to 650 cm^{-1}. The authors attributed the spectral differences between T1 and T2 to various degrees of weak volume scattering caused by variations in particle size. T1 is interpreted as representative of primarily coarse to rocky materials, whereas T2 is interpreted as representative of fine particulates that may be thinly coating rough and dark boulders on the surface of the asteroid (Hamilton et al., 2021).

Prior to the arrival of OSIRIS-REx at Bennu, spectral comparisons and low thermal inertia values (Binzel et al., 2015; Emery et al., 2014; Yu & Ji, 2015) indicated a smooth surface of regolith where particles would average less than a centimeter in size (Lauretta et al., 2015). However, the surface of Bennu is rocky, with more than 200 boulders present having diameters of >10 m (Lauretta, DellaGiustina, et al., 2019). This apparent contradiction has prompted new hypotheses to explain the low thermal inertia values, including trapping of fines (dust cover) on rough boulder surfaces and/or high-porosity friable rock (Rozitis et al., 2020; Ryan et al., 2019).

The thermal breakdown of boulders on the surface of Bennu (Molaro, Walsh, et al., 2020) is a process that not only disaggregates rocks but also acts as a source of fines. Thermal fatigue may play a role in the ejections of fine particulates from Bennu's surface that were observed numerous times by the OSIRIS-REx navigation cameras (Hergenrother et al., 2019; Lauretta, Hergenrother, et al., 2019; Molaro, Hergenrother, et al., 2020).

2.4. Modeling Mineralogy From MIR Spectra

LSMA is a methodology commonly used for modeling mineralogy quantitatively. Lawson and Hanson (1974) and Ramsey and Christensen (1998) each developed non-negative LSMA methods (compared by Rogers & Aharonson, 2008). Both assume linear mixing properties in which each spectral end-member (not mineralogical end-member) in the mixture contributes to the spectrum proportionally to its abundance. Of the two methods, the Lawson and Hanson (1974) methodology exhibits lower root-mean-square error (RMSE; Rogers & Aharonson, 2008) and is thus the preferred method. The technique has a mineral-dependent accuracy of ∼5–10% when applied to laboratory data (Feely & Christensen, 1999; Rogers & Aharonson, 2008).
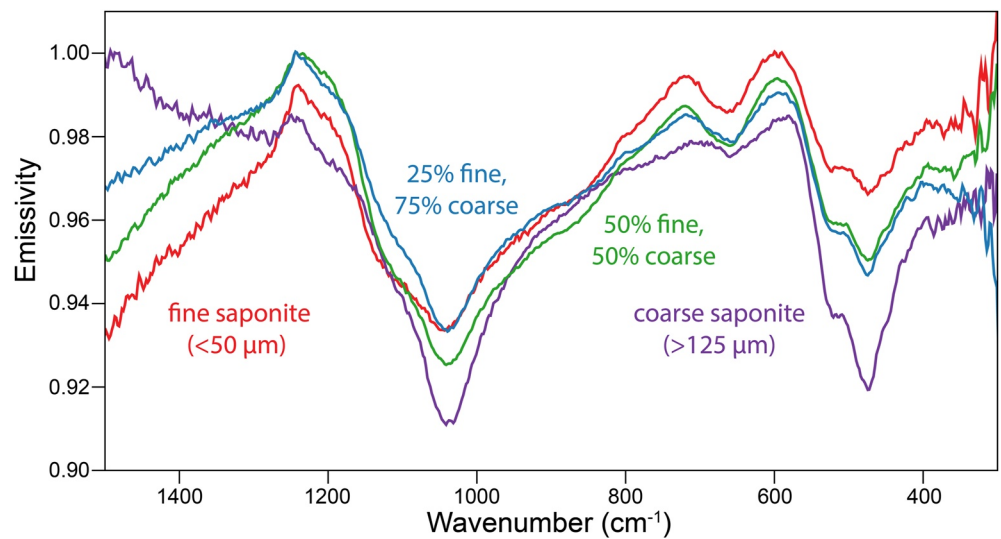
Figure 1. Mid-infrared simulated asteroid environment spectra of fine (<50 μm), coarse (>125 μm), 50% fine/50% coarse, and 25% fine/75% coarse particulates of the end-member saponite with the laboratory radiance-to-emissivity conversion method used by Ruff et al. (1997). All samples were darkened with 11 vol% carbon powder.

Many workers have studied the complexity of particle-size effects in MIR spectroscopy (Conel, 1969; Cooper et al., 2002; Logan & Hunt, 1970; Lyon, 1964; Mustard & Hays, 1997; Shirley & Glotch, 2019). Nonlinear mixing occurs in the MIR for fine particulates (Salisbury & Wald, 1992; Thomson & Salisbury, 1993) because spectra of fines include both Reststrahlen bands and transparency features, which scale with particle size (Salisbury & Wald, 1992) (Figure 1). Light can pass through smaller particles at some MIR wavelengths (i.e., where the imaginary component of the complex index of refraction, k, is low). The multiple scattering caused by fine particulate transparency at these wavelengths leads to nonlinear mixing effects (Salisbury & Wald, 1992; Thomson & Salisbury, 1993).

Traditional linear least squares models cannot reliably provide quantitative assessments of mineralogy for fine particulate materials due to nonlinear mixing. Though Ramsey and Christensen (1998) showed some success of LSMA modeling of fine particulate mixtures, those examples were limited to a few simple mixtures with two components and a two-component library, rather than more complex mixtures and a blind library. Follow-up work by Pan et al. (2015) concluded that non-negative LSMA could not reliably model fine particulates (<10 μm); in comparison, PLS retrieved values within ±10% of the absolute abundance. The nonlinear mixing behavior of fines warrants an alternative approach to LSMA in cases where fines are present. Figure S1 compares the LSMA and PLS methodologies for an identical data set of fine particulate mixtures and demonstrates the superior accuracy of the PLS method. Though it requires significant effort to develop appropriate spectral training sets, the major advantage of machine learning multivariate analysis is that it removes the requirement of linear mixing across all wavelengths (Section 3.6); therefore, we use it in this work.
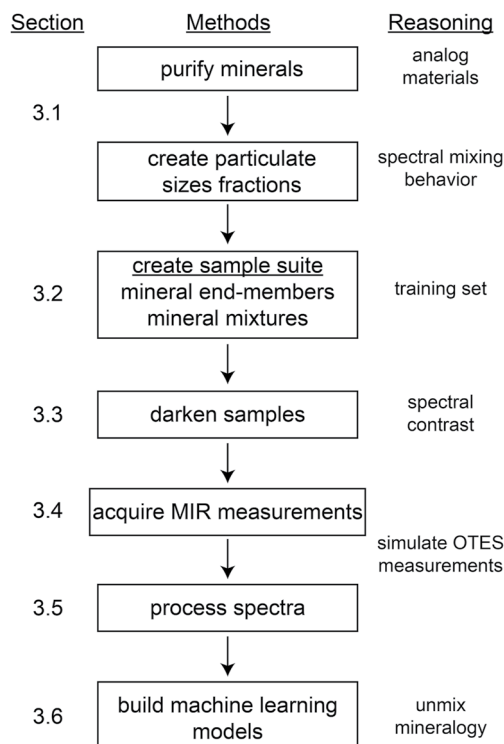
Figure 2. Conceptual diagram summarizing the methodology and reasoning for the creation of the machine learning models.

3. Methods

The machine learning models define the requirements for the suite of laboratory samples and MIR measurements (training set) for this study. The workflow for the methodology is summarized in Figure 2. The machine learning models require spectra from samples that are analogous to the materials

Table 1
Provenance of 13 Pure Mineral Samples Used in This Work as Analogs to Minerals Found on Bennu

Dana class	Mineral species	Provenance
Phyllosilicate	Antigorite (Mg-rich serpentine proxy)	Natural, locality unknown
	Cronstedtite	Natural, Czech Republic, Romania, and Germany
	Saponite	Natural, Bump Creek, Arizona
Nesosilicate	Olivine (Fo40)	Synthetic
	Olivine (Fo80)	Synthetic
	Olivine (Fo90)	Natural, San Carlos, Arizona
Inosilicate	Enstatite	Synthetic, 62%–65% clinoenstatite and 33%–37% protoenstatite
Anhydrous carbonate	Calcite	Natural, locality unknown
	Dolomite	Natural, locality unknown
Multiple oxide	Magnetite	Natural, locality unknown
Simple oxide	Ferrihydrite	Natural, Iceland
Sulfide	Pyrrhotite	Natural, locality unknown
Sulphate	Gypsum	Natural, locality unknown

Note. Natural samples with unknown localities were obtained from museum collections and dealers. Synthetic samples were produced in Stony Brook University's Experimental Petrology Laboratory using the procedures detailed in Turnock et al. (1973).

whose mineralogy will be predicted (Sections 3.1–3.3) as well as MIR measurements that closely match OTES measurements (Sections 3.4 and 3.5). After establishing the training set, parameters that affect the machine learning models must be selected (Section 3.6). The approaches for quantifying model error and applying the models to unseen data are detailed in Sections 3.7 and 3.8. Lastly, the methodology for quantifying mineral abundances within meteorite samples is discussed in Section 3.9.

3.1. Sample Provenance and Characterization

The mineral species used in this work are commonly present within CI and CM chondrites (Bland et al., 2004; Howard et al., 2011; King et al., 2015). The 13 natural and synthetic individual mineral samples are summarized in Table 1. To have sufficient sample for this investigation, synthetic samples were necessary to include, although Hamilton et al. (2020) noted spectral differences in the MIR spectra of natural versus synthetic olivine.

There are differences between terrestrial and meteoritic phyllosilicates. Cronstedtite ($Fe^{2+}_2Fe^{3+}(SiFe^{3+})O_5(OH)_4$) in terrestrial settings does not contain Mg, whereas meteoritic cronstedtite has been defined as an Fe-rich serpentine that includes Mg (e.g., Howard et al., 2009). However, the important aspect of both terrestrial and meteoritic cronstedtite is the presence of Fe^{3+} at both octahedral and tetrahedral sites (Howard et al., 2009). Given these similarities, we used terrestrial cronstedtite as an analog for meteoritic cronstedtite. Additionally, we used antigorite, a Mg-rich terrestrial serpentine, as an analog for meteoritic Mg-rich serpentine, an unnamed polymorph. Antigorite is not found within CI/CM chondrites, but owing to its availability in large quantities, it is a suitable Mg-rich serpentine proxy.

We evaluated the purity of the 13 samples through MIR spectroscopy and/or XRD (Table S1). The olivine, enstatite, ferrihydrite, gypsum, calcite, dolomite, and antigorite (Mg-rich serpentine) samples did not require any purification measures. We hand-picked for purity and/or magnetically separated the cronstedtite, saponite, and magnetite samples to remove unwanted species. The cronstedtite is derived from four separate samples; we purified and characterized each prior to combining. XRD measurements of the pyrrhotite sample indicated that it was pure, but MIR spectra revealed a phyllosilicate impurity when the sample was pressed into a pellet. This discovery was made after pyrrhotite was already included in many of our mineral mixtures. Due to this impurity and the lack of spectral features of pyrrhotite in the MIR, we did not create pyrrhotite machine learning models (Section 4.2).
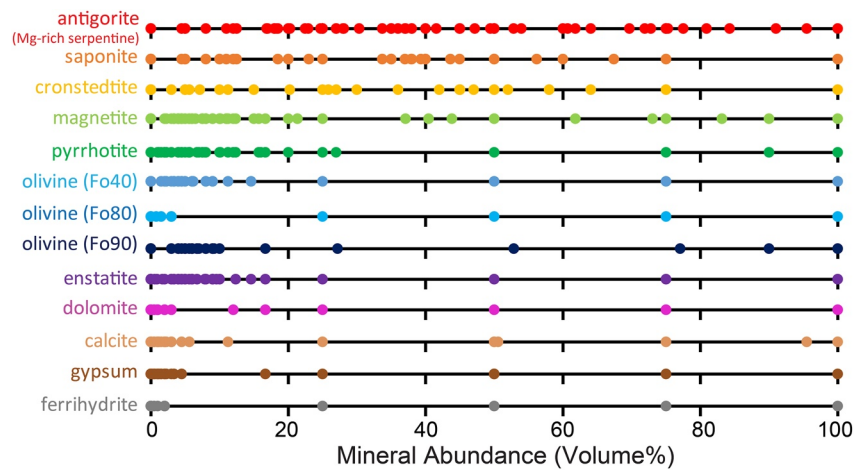
Figure 3. Normalized abundances of the 13 mineral species in the training set of 276 mineral mixtures (112 fine, 112 coarse, and 52 mixed particle sizes) and 26 pure minerals (13 fine and 13 coarse; abundances of 100%). The mineral abundances are identical between the 112 fine and 112 coarse linear mineral mixtures. The mineral abundances of the 52 mixtures that included coarse and fine particulates together overlap the plotted values.

Lastly, we crushed and sieved each sample to <50 μm (fine) and >125 μm (coarse) fractions. These two particulate size fractions were chosen to account for linear and nonlinear MIR mixing behaviors of the minerals. When coarse particulate sample was unavailable (e.g., mixtures that included synthesized minerals), we used intact whole pellets to simulate coarse particulates—specifically for olivines, enstatite, cronstedtite, ferrihydrite, and pyrrhotite. Pellets were pressed using fine particulates in a Carver hydraulic press at 10,000 PSI for ~1 min using a Pike press and anvil set. All samples were poured/placed into sample cups avoiding compaction with a consistent sample volume.

3.2. Suite of Individual Minerals and Mineral Mixtures

We used the 13 mineral samples to produce 112 fine, 112 coarse, and 52 mixed particle size mineral mixtures (276 spectra). In addition, we included pure spectra of each of the 13 individual minerals at both fine and coarse particle sizes (26 spectra), bringing the training set to a total of 302 spectra that inform the multivariate analysis models.

Pellets result in close packing of fines that mimic the spectral properties of coarse particulates in the MIR because of the reduction of multiple scattering reflections (Salisbury & Wald, 1992). With pellet samples, volume scattering is still present because the mean optical path length is greater than the diameter of the grains that compose the pellet (Clark & Roush, 1984; Hunt & Logan, 1972), resulting in a combination of linear and nonlinear mixing behaviors. Even though pellets are not a perfect spectral analog to coarse mineral grains, compaction of fine particulates (pellets) of terrestrial analog samples may better simulate textural and particulate size characteristics of meteoritic or asteroid materials because CM chondrite matrices are dominated by fine particulate phyllosilicates (Tomeoka et al., 1989).

The 112 fine mineral mixtures were physical mixtures in which each species was measured on a scale in grams to the 10 thousandths place. The 112 coarse mineral mixture spectra were synthetic spectra created through linear combinations of the coarse and pellet samples. The production of coarse synthetic spectra instead of measurements of physical mixtures was a time-saving measure. The 52 mixtures of mixed particle sizes were made by pressing a portion of a fine mineral mixture into a pellet and combining it with the remainder of the fine mineral mixture. This enabled the models to learn how fine and coarse material behave together spectroscopically. It also provided an analog of a "rock" with a coating of fines (e.g., 90 vol% coarse, 10 vol% fine).

The mineral abundances of the 112 fine or coarse mineral mixture training set are summarized in Figure 3 and Table S2. Figure 3 is a visual representation of the distribution of abundances within the sample suite by mineral, and Table S2 provides exact mineral abundances for each mixture. Some of the mixtures act as analogs to CI and CM chondrites and match meteorite literature values (Bland et al., 2004; Howard et al., 2011; King et al., 2015).
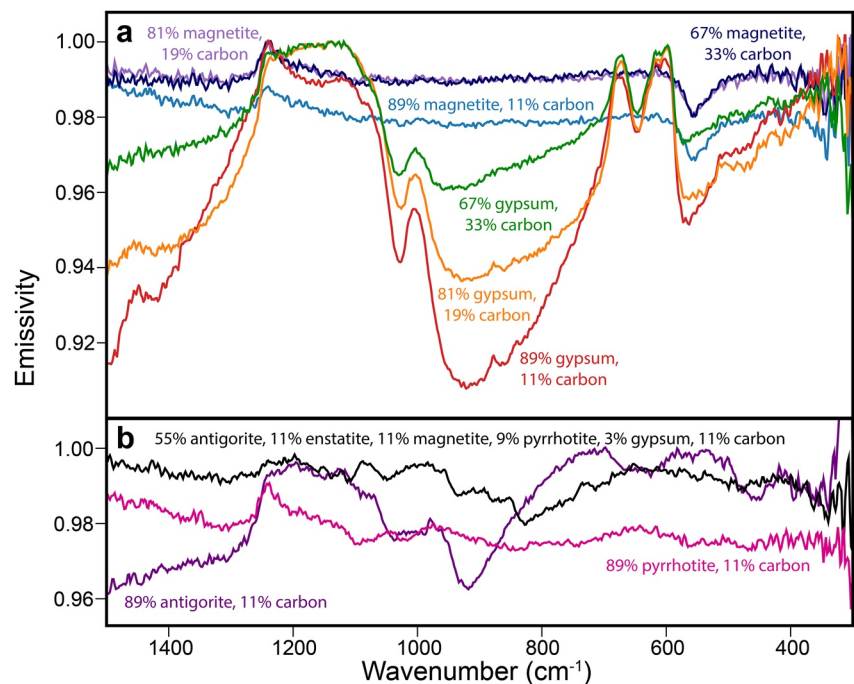
Figure 4. Simulated asteroid environment mid-infrared spectra with the standard laboratory radiance-to-emissivity conversion of (a) darkened magnetite and gypsum with 11, 19, and 33 vol% carbon powder and (b) darkened antigorite (Mg-rich serpentine), pyrrhotite, and mineral mixture #5 (black), each with 11 vol% carbon. The positive feature at 1,240 cm^{-1} in every spectrum is due to the presence of carbon powder.

Additionally, the training set includes simpler binary or ternary mineral mixtures. These mixtures are useful in broadening the mineral abundance range from 0 to 100 vol% for each mineral. This is necessary for the machine learning models because it allows for the prediction of unexpectedly high or low abundances by the models.

3.3. Constraining Sample Albedo

Sample albedo ($\sim$0.3–3.0 µm) contributes to MIR spectral differences (Donaldson Hanna et al., 2017; Logan et al., 1973; Lucey et al., 2017; Shirley, 2018). Most notably, reducing the sample albedo reduces the spectral contrast in the MIR and shifts the Christiansen Feature (CF) to longer wavelengths in a simulated asteroid environment (SAE). Accounting for spectral contrast is imperative for machine learning multivariate analysis because feature depths are directly connected to the predictive weights of particular mineral species (see Section 4.1).

Due to these MIR spectral changes, it was necessary to darken all individual minerals and mineral mixtures to simulate the albedo of Bennu's surface. There is a tradeoff between darkening samples enough to properly simulate the low albedo of Bennu while avoiding darkening to a degree that obscures spectral features. We added 2, 11, 33, and 55 vol% carbon to a naturally bright sample (gypsum) and a dark sample (magnetite) and collected VNIR and MIR spectral data to test the effect of the degree of darkening. Through this investigation and testing additional individual minerals and mineral mixtures (Figure 4), we chose to darken all samples with 11 vol% carbon to reduce spectral contrast instead of trying to match expected carbon abundances on the surface of Bennu. All samples were darkened by combining and mixing weighed lampblack carbon powder with a given mineral specie prior to combining individual minerals into mixtures to ensure homogeneity.

The lampblack carbon powder can easily change sample albedo to simulate the darkening effects of organics in the sample suite. However, carbon also introduces a positive feature in the MIR at $\sim$1,240 cm^{-1}. Unfortunately, this feature can overlap the CF of some of the individual mineral and mineral mixture spectra in the spectral library. This feature spans from 1,260 to 1,200 cm^{-1}, affecting only seven OTES spectral channels. The band was removed from the models by excluding these channels.

Throughout this study, mineral abundances were normalized by removing the 11 vol% carbon from each individual mineral and mineral mixture within the metadata of the models. Because carbon acts as a proxy for organics, the removal of carbon in the metadata abundances allows for more accurate mineral predictions. Mathematical "darkening" of the spectra (e.g., using a neutral synthetic spectrum) cannot be used to account for spectral contrast instead of physically darkening the samples. This is because synthetic additions would have no real physical abundance associated with them, which is required for the metadata of the models.

3.4. MIR Measurements

There are two main factors that the MIR emissivity measurements of this work require: measurement parameters (e.g., spectral resolution) in agreement with OTES data and temperature conditions of an airless body. The spectra were collected at a higher spectral resolution than the OTES instrument, later downgraded to match OTES resolution, and over an energy range (1,500–310 cm^{-1}) encompassing most of the OTES spectral range.

We collected all MIR spectra with a Nicolet 6700 Fourier Transform Infrared (FTIR) spectrometer for emissivity measurements. The spectrometer is equipped with a CsI beamsplitter and a DTGS detector with a CsI window, allowing for measurements ranging from 4,000 to 225 cm^{-1} at a spectral sampling of ~1.93 cm^{-1}/channel. For each spectrum, we acquired a total of 256 sample scans over ~17 min. The air within the spectrometer is continuously purged of CO_2 and water vapor. The FTIR spectrometer settings are controlled through Thermo Fisher Scientific OMNIC 9 software.

The spectrometer is coupled with the Planetary and Asteroid Regolith Spectroscopy Environmental Chamber, a custom-built planetary environmental spectroscopy chamber at the Center for Planetary Exploration (CPEx) at the Stony Brook University. This chamber allows for measurement of spectra under SAE conditions. For these measurements, the chamber is pumped down to <10^{-4} mbar over several hours. Next, the chamber's dewar is filled with liquid nitrogen, cooling it to <−125°C. The characteristics and operation of the chamber are described in detail by Shirley and Glotch (2019).

We acquired spectral measurements of a black body target at 70°C and 100°C each time we collected sample spectra. Samples were placed in aluminum sample cups that heated samples from below to 80°C while an external light with variable wattage heated the samples from above. Each sample was heated until a temperature-gradient stabilized, prior to the spectral measurement (>45 min).

A thermal gradient was induced to simulate the temperature effects of MIR measurements on an airless body like the asteroid Bennu. The dual heating system from above and below the sample induces the gradient (Henderson and Jakosky, 1994, 1997; Thomas et al., 2012). The lack of air between grains on an airless body reduces conductivity, resulting in anisothermality in the upper hundreds of microns of the regolith. The upper layer insulates the regolith below, allowing the shallow subsurface to be warmer than the outermost layer of the planetary surface. Due to the induced thermal gradient in SAE measurements and pressure/temperature conditions, SAE (equivalent to a simulated lunar environment) and ambient MIR spectral measurements have several distinctions (Donaldson Hanna et al., 2019, 2021; Logan & Hunt, 1970; Shirley & Glotch, 2019; Thomas et al., 2012). Most notably, the shift in position of the CF and an increase in spectral contrast makes SAE measurements necessary for this work. The low albedo of the samples reduces the thermal gradient. This results in smaller differences between ambient and SAE spectra than is typically seen for lunar or lunar analog samples. Previous SAE measurements of the low albedo Tagish Lake meteorite did note some differences between the ambient and SAE spectra, especially for fine particulates (Glotch et al., 2018). Donaldson Hanna et al. (2019) examined mineral mixtures and chondritic meteorites (CM, CI, CV, CR, and L5) under ambient and SAE conditions. Although there are similarities between these two types of measurements there are also differences in CF positions, feature depths, and shapes for the low albedo analog samples (Donaldson Hanna et al., 2019).

3.5. Spectral Processing

We use two blackbody measurements to produce calibrated radiance spectra (Ruff et al., 1997). We performed two methods of radiance-to-emissivity conversion: the laboratory method used by Ruff et al. (1997) which assumes an emissivity at a single brightness temperature, and the OTES method which accounts for variable temperatures in the field of view (Hamilton et al., 2019). Spectra processed using the OTES method were used for

the models in this paper to allow for their application to OTES spectra of Bennu. Spectra from the conventional laboratory conversion are available in Table S3 for future workers to build their own spectral models with unique applications due to their higher spectral resolution.

The laboratory conversion method determines the maximum brightness temperature over a user defined wave-number range (in our case, between 500 and 1,700 cm^{-1}), and then divides the radiance by a Planck function of that same temperature. This method assumes that the maximum brightness temperature over the user-defined wavenumber range is the kinetic temperature of the sample, and that the sample emissivity is unity at the frequency of the maximum brightness temperature (e.g., Realmuto, 1990). The "emcal2" function from the davinci software suite (available at http://davinci.asu.edu/) was used to produce these emissivity spectra.

We also converted our radiance spectra to emissivity through a procedure that simulates the OTES L3 "mt_emissivity" product processing. Instead of using a single Planck function, each radiance spectrum was fit using a range of Planck functions associated with different temperatures, in a linear least squares model. The measured radiance was then divided by the modeled radiance, divided by 0.97 (Hamilton et al., 2019). Our wavenumber fitting range was set to 1,500–310 cm^{-1} rather than 1,500–260 cm^{-1} to exclude the high noise of the laboratory data at lower wavenumbers. All laboratory and OTES spectra were manually normalized to a common emissivity maximum of 0.97 over the wavenumber range of 1,500–310 cm^{-1}. The equivalent normalization treatment of the model and OTES spectra allows for the most accurate application.

The majority of OTES spectra have spectral contrast of ∼2%, whereas our analog mixtures range from ∼2 to 6% with reduced spectral contrast for mixtures with multiple mineral species and especially fine particulate mixtures. Our mineral mixtures suite covers a wider compositional range than is expected for Bennu, and therefore our mineral mixtures likely simulate the spectral contrast properly. Accounting for spectral contrast is not a unique problem for machine learning and must be accounted for with other types of spectral modeling (e.g., LSMA).

Lastly, we reduced the spectral sampling to ∼8.66 cm^{-1}/channel to match the spectral sampling of the OTES data using a cubic spline interpolation. It is important for the resolution of the models to exactly match the application data given that PLS modeling relies on specific channel positions for its predictions. All spectral processing occurred prior to narrowing the spectral wavenumber range (e.g., removal of the carbon feature at ∼1,240 cm^{-1}). The spectral data that followed this pre-processing procedure to simulate the OTES radiance-to-emissivity conversion are available in Table S4.

3.6. Machine Learning Models

Machine learning models require a training set (i.e., spectra) and corresponding metadata (i.e., quantitative mineralogy). In the case of this work, the training set must include spectra representing as many possible known combinations of composition and particle size. Applying machine learning models to spectral data is relatively new to the planetary spectroscopy community and therefore methods for model optimization are still being developed. Effects on model performance of parameters such as data set size were investigated by Dyar and Ytsma (2021) for laser-induced breakdown spectroscopy (LIBS); however, a comprehensive study has not been performed for MIR spectroscopy. Because machine learning models learn from the training set and metadata, generally increasing the number of observations will improve model performance (Dyar & Ytsma, 2021); however, the cost and effort of expanding the training set at some point outweighs minor improvements.

PLS is a common machine learning technique (Geladi & Kowalski, 1986; Wold et al., 1983, 2001) that has been used in spectroscopic analysis of geologic materials. Clegg et al. (2009), Dyar et al. (2016), Breitenfeld et al. (2018) and others utilized PLS for the prediction of chemical and mineral abundances from LIBS, XANES, and Raman spectral data. In the MIR, Hecker et al. (2012) and Pan et al. (2015) have utilized PLS as a quantitative modeling technique.

PLS predictions utilize all channels of the spectral range, assigning a coefficient to every channel associated with each metadata category. This is accomplished by regressing one or multiple response variables (e.g., saponite vol%) against multiple explanatory variables (p) (Wold et al., 2001). Here, the emissivity values associated with specific channels of the MIR spectra represent the explanatory variables. While finding covariance, PLS shrinks the p-dimensional matrix to q dimensions (Wold et al., 2001) where q represents the model components. Because of this reduction, an advantage of PLS is that it can model one or several response variables simultaneously,

even if the number of observations (spectra) is more or less than the number of explanatory variables (Wold et al., 2001). Additionally, multiple response variables can be modeled even if the explanatory variables are correlated (Wold et al., 2001).

Given that PLS models are controlled by the number of components chosen, for each model, coefficients were produced for 2–20 components. Next, we chose the number of components of a selected model based on the lowest mean square error. Although choosing a model with more components can lead to higher prediction accuracy, it can also lead to overfitting of the data and make the model less applicable to outside data (Hastie et al., 2009). Our procedure optimizes a given model while keeping its applicability high.

PLS can be utilized for predictions of a single response variable (PLS1) or two or more response variables simultaneously (PLS2; Rosipal & Krämer, 2005). We use a combination of PLS1 and PLS2 in this work. We organized the PLS models by mineralogical group. For groups with multiple mineral species available for prediction (e.g., phyllosilicates), we used PLS2, whereas we used PLS1 for single mineral species (e.g., magnetite).

Least absolute shrinkage and selection operator (lasso) (Tibshirani, 1996) is an alternative machine learning technique to PLS. The lasso is a continuous shrinkage model, reducing the coefficients to as low as zero in a model (Tibshirani, 1996) rather than using every channel available like the PLS models. The lasso algorithm reduces the residual sum of squares while introducing a bound for the sum of the absolute value of the lasso coefficients (Tibshirani, 1996). In other words, lasso determines a set of predictors with the greatest effect on the response variable(s). In this study, alpha values (parameter that affects model sensitivity) were chosen for the lasso models at which the mean square error was smallest.

An advantage of lasso is that the reduction of coefficients isolates which spectral channels most strongly mark the effect of the response variable. Although PLS provides a stable model, lasso is useful when fewer spectral channels are available. For example, using lasso on laboratory hyperspectral data sets could reveal which channels would be most useful for multispectral data sets of planetary bodies. With lasso, only one response variable could be evaluated at a time, and therefore abundances were investigated by group mineralogy rather than by individual species. All machine learning models were constructing using tools available at http://nemo.mtholyoke.edu/.

3.7. Quantifying Model Errors

There are three types of RMSE that are reported in this work and match the nomenclature reported by Dyar and Ytsma (2021). RMSE-C is the internal error of the calibrated model. RMSE-CV is the cross-validated RMSE, where the number of folds (subsets of the spectral data set) equals the square root of the number of spectra. Lastly, RMSE-P is the RMSE of predicted data that are not included within the model. RMSE-C values indicate the accuracy of predicting samples within the model but do not inform the precision of predictions of unseen data (spectra that are not present in the model, such as OTES spectra). RMSE-CV values are reported as a metric to better simulate the accuracy of the prediction of unseen data. RMSE-CV is determined by breaking up the data set into subsets (equal to the square root of the total number of spectra in the model). Next, a subset (fold) is held out and the spectra within the held-out fold are predicted using a model based on the remaining observations. This process is repeated for all spectra in the model prior to calculating the RMSE. This means that a prediction is made for a spectrum by a model that does not include that spectrum. Therefore, this is a more robust error metric for the prediction of unseen data compared to RMSE-C. Lastly, to calculate a RMSE-P value, a RMSE calculation occurs after making predictions of data outside of the model and comparing to known abundance values.

3.8. Applying Models to Unseen Data

For this study, "test" PLS and lasso models that used the 302 individual mineral and mineral mixture spectra but excluded the 15 meteorite spectra were constructed. We made modal mineralogy predictions for spectra that are not present within the test models to evaluate the ability to predict unseen data by recording RMSE-P values. We applied the PLS and lasso models to meteorite samples including Essebi (C2-ung), Murchison (CM2), ALH 81002 (CM2), ALH 83100 (CM1/2), MET 01070 (CM1), QUE 93005 (CM2), QUE 97990 (CM2), and SCO 06043 (CM1). Coarse (>125 μm) and fine (<45 microns) particulate samples were investigated for each of the meteorites except for Essebi, for which only a fine particulate sample was available. The goal of this investigation was to assess the accuracy of the models by comparing predictions to quantitative XRD data. This allowed us to

Table 2
Number of Spectra Within the Training Set for the Different Test and Optimized Compositional Machine Learning Models

Compositional model name	Test models	Optimized models
Phyllosilicate	232	247
Olivine	143	158
Carbonate	136	148
Pyroxene	117	130
Magnetite	173	188

Note. The total number of spectra available is 317 (302 from minerals and mineral mixtures and 15 from meteorites).

evaluate the accuracy of the test models on unseen data representing a composition similar to Bennu's. RMSE-C, RMSE-P, and RMSE-CV values were calculated for the test models; however, RMSE-CV values were trusted as the most accurate error metric (see Section 4.3).

After investigating the robustness of the test models, the 15 meteorite spectra were encompassed in the "optimized" machine learning models to further improve the prediction accuracy of the models. These models are optimized in comparison to the test models because the addition of meteorite spectra increases the number of observations (Dyar & Ytsma, 2021) and provided better analogs than the terrestrial samples. We employed the optimized PLS models for the prediction of OTES spectra to make preliminary quantitative assessments of Bennu's mineralogy. The OTES spectra under investigation included the global average, T1, and T2 spectra (Hamilton et al., 2021).

### 3.9. Quantitative XRD of Murchison

Quantitative XRD was previously performed for all eight meteorites by Howard et al. (2009, 2011). Other than Essebi, the piece of the meteorites used for quantitative XRD differs from the aliquot we utilized for the MIR measurements. For both types of analysis, the sample was in powder form, and >100 mg was used. The average modal mineralogy that results from these analyses obscures compositional variations of the meteorite on small scales. To verify the consistency of the modal mineralogy across aliquots of the same meteorite, we made new quantitative modal mineralogy estimates of Murchison from quantitative XRD measurements. These measurements were made on a Thermo Scientific ARL Equinox 100 X-Ray Diffractometer at the American Museum of Natural History, and we utilized the same methodology as Howard et al. (2009, 2011). This investigation confirmed compositional consistency across aliquots with <3 vol% difference for each mineral group between the Murchison mineral abundances reported in Howard et al. (2011) and our sample (Table S5). This compositional consistency is in agreement with mineral abundance predictions by quantitative XRD of two to three aliquots of a given meteorite (e.g., Mighei, QUE 93005, and MET 01070) (Howard et al., 2009, 2015). We used the modal mineralogy values from Howard et al. (2011) as our ground truth quantitative mineralogy of all eight meteorite samples. There are examples of meteorites where this trend may not hold (e.g., Tagish Lake); however, for our eight meteorites, multiple lithologies are not documented, reducing the likelihood of mineralogical variability across aliquots. The error of the quantitative XRD measurements was originally estimated by Howard et al. (2011) conservatively as ranging from 2 to 5 vol% depending on the mineral. King et al. (2015) update the associated error to 1–3 vol%.

4. Results

The MIR spectral data set (Section 4.1) consists of individual minerals, mineral mixtures, and meteorites. Machine learning models can be used for the prediction of any metadata category; in this work, we made separate models for the prediction of composition and particle size. The compositional models include both coarse and fine particulates to allow the models to learn how the same composition can present differently spectroscopically based on particle size. However, the compositional models do not predict particle size, and separate models were constructed for this purpose. For the compositional models, we constructed test and optimized models (Sections 4.2 and 4.3). The number of spectra that contributed to the test and optimized compositional machine learning models are summarized in Table 2.

4.1. MIR Spectral Data

The SAE MIR data set consisting of 302 spectra is shown in Figure 5 (spectra available in Table S4). This data set consists of fine, coarse, and mixed particulate sizes and compositionally matches the mineral abundances given in Figure 3 and Table S2. Individual minerals and mineral mixtures with few components have spectra that exhibit expected absorptions. Yet, as the number of components increases, absorptions can overlap, obscuring specific minerals. Additionally, mixed-size particulate spectra exhibit characteristics of both their coarse and fine counterparts.
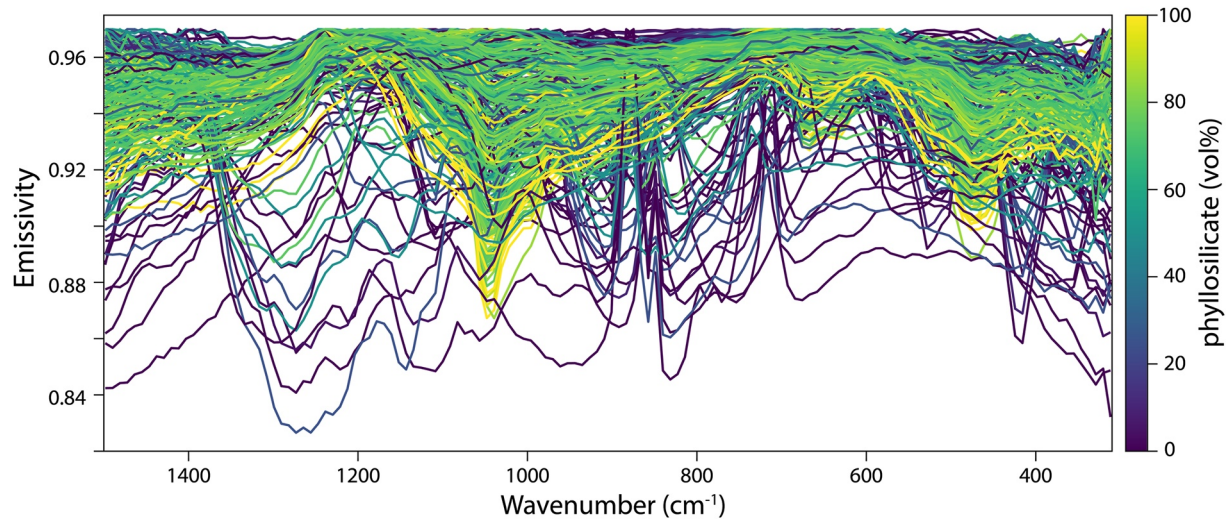
Figure 5. Simulated asteroid environment mid-infrared training set of 302 spectra processed using the OSIRIS-REx Thermal Emission Spectrometer radiance-to-emissivity conversion and normalized to a maximum emissivity of 0.97 from 1,500 to 310 cm^{-1}.

In addition to the spectral training set, 15 MIR spectra of eight carbonaceous chondritic meteorites as fine and coarse particulates are shown in Figure 6. Clear spectral differences exist between the coarse and fine particulates of the same meteorite samples, demonstrating how substantial an effect particulate size can have in the MIR. Compared to the fine particulate meteorite spectra, the coarse particulate meteorite spectra have prominent features at ~450 and ~640 cm^{-1} attributable to phyllosilicates. This trend of band depth holds for the broad phyllosilicate feature and its position shifts from ~1,050 to 700 cm^{-1} for fines to ~1,100 to 800 cm^{-1} for coarse
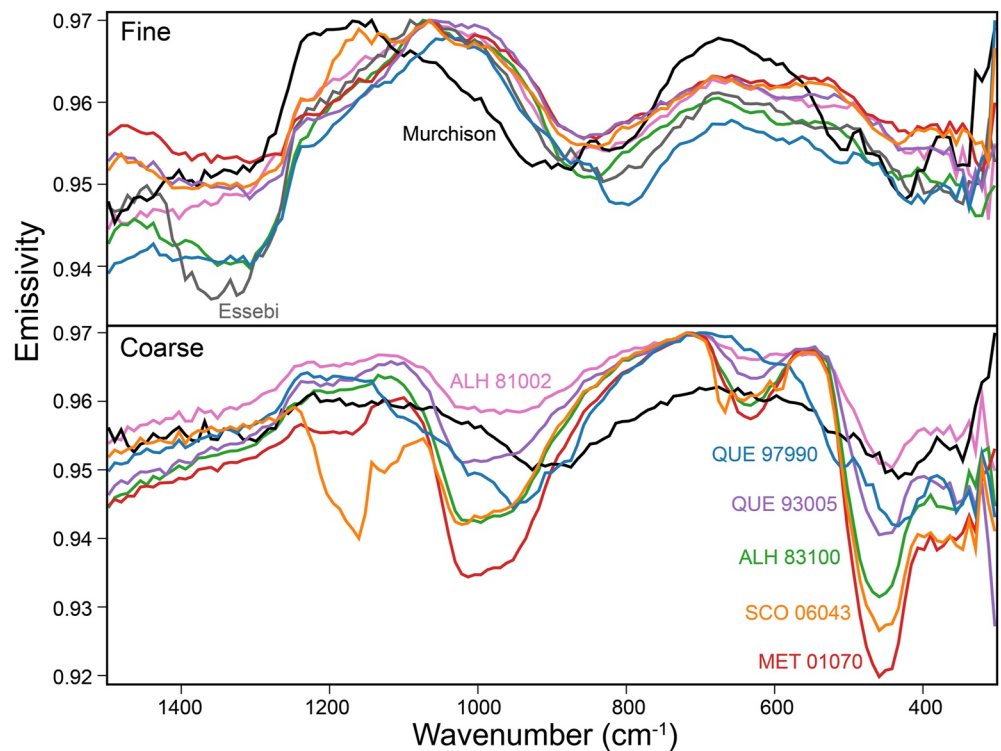
Figure 6. Simulated asteroid environment mid-infrared spectra of fine (top) and coarse (bottom) particulate meteorites ALH 81002, ALH 83100, Essebi, MET 01070, Murchison, QUE 93005, QUE 97990, and SCO 06043, processed using the OSIRIS-REx Thermal Emission Spectrometer radiance-to-emissivity conversion.

particulates with a change in feature shape. Fine particulate meteorite spectra display the expected "roll-off" (reduction in emissivity) shortward of the CF. The MIR spectral trends identified for these meteorites are consistent with those found in previous studies (e.g., Bates et al., 2020; Donaldson Hanna et al., 2021, 2019; Hamilton et al., 2019, 2021; Hanna et al., 2020).

4.2. Machine Learning Test Models

Numerous machine learning models were built to test model performance and to assist in selecting the best models for the application to OTES data. The effect of model type (PLS1, PLS2, and lasso) and spectral wavenumber range were evaluated through the test models.

A pyrrhotite model was not finalized due to the impurity within the sample and its spectral blandness in the MIR. Final models for gypsum and ferrihydrite were also not constructed because of their absence within the meteorite collection, and fewer samples with these mineral species were made, resulting in fewer observations for modeling. These exclusions leave five remaining models: phyllosilicate, olivine, carbonate, pyroxene, and magnetite. These five models include all spectra for which the mineral group of interest has an abundance >0 vol%. Although models were not finalized for pyrrhotite, gypsum, or ferrihydrite, samples including these mineral species are still present within the training set. We do not expect that their presence would inhibit the accuracy of the other prediction variables, and instead these species simply act as mixing agents. However, the pyrrhotite content was limited to <10 vol% for the phyllosilicate model because the pyrrhotite sample is contaminated with a phyllosilicate (Section 3.2).

For the test models, we evaluated the effect of model type (lasso, PLS1 and PLS2) as well as three spectral wavenumber ranges. The "whole range" extends from 1,500 to 1,260 and 1,200 to 310 cm^{-1}. The "reduced range" covers 1,200–310 cm^{-1}. The "selective ranges" for the models are 1,200–800 and 700–400 cm^{-1} (phyllosilicate), 1,100–330 cm^{-1} (olivine), 1,000–600 cm^{-1} (carbonate), 1,200–600 cm^{-1} (pyroxene), and 800–310 cm^{-1} (magnetite). The selective range models were designed to isolate spectral features characteristic of a given mineral group. All spectral ranges avoid 1,260–1,200 cm^{-1} to eliminate the region where carbon powder contributes a positive feature to the spectra (Section 3.3).

We calculated RMSE-CV and RMSE-P values for each model to compare model accuracies (Figure 7). For a given mineral group, PLS2 models yield lower RMSE-CV values compared to the PLS1 and lasso techniques. Additionally, for most mineral groups, the broadest spectral range outperformed restricted ranges with lower RMSE-CV values. To evaluate the accuracy of predicting compositional abundances similar to Bennu, we applied the machine learning test models (that do not include any meteorite spectra) to the 15 MIR spectra of carbonaceous chondritic meteorites. Next, we calculated the associated RMSE-P values by comparing predictions to quantitative XRD abundances (Howard et al., 2011) (Figure 7).

In addition to models tested in Figure 7, other PLS2 models were tested. Modeling by mineral group considering fine and coarse particulate materials separately and simultaneously (e.g., fine and coarse magnetite separately instead of a single response variable of magnetite content) resulted in models that generally overestimated mineral abundances. Additionally, PLS2 models where all mineral groups or all mineral species were predicted together also showed poor performance. When building machine learning models, testing the effect of model parameters is encouraged to determine an optimized model. Optimized model performance should be quantified through RMSE-CV or RMSE-P values.

4.3. Selection of Optimized Machine Learning Models

We isolated factors influencing model accuracy to test their relative effects (summarized in Figure 7). Generally, RMSE-P values are the best way to determine model robustness. However, because only 15 meteorite spectra were available for prediction, an outlier can skew the value. Additionally, because the eight meteorite samples from which these 15 spectra were derived have similar modal mineralogies, only a narrow compositional range was tested, and models that favor that range can appear artificially favorable. Because of these limitations in RMSE-P values, we used the RMSE-CV value as the error metric for selecting our optimized models.

Based on the RMSE-CV values from the test models, we determined that the optimized models would be constructed using PLS2 or PLS1 with the whole spectral range (1,500–1,260, 1,200–310 cm^{-1}). PLS2 was used for
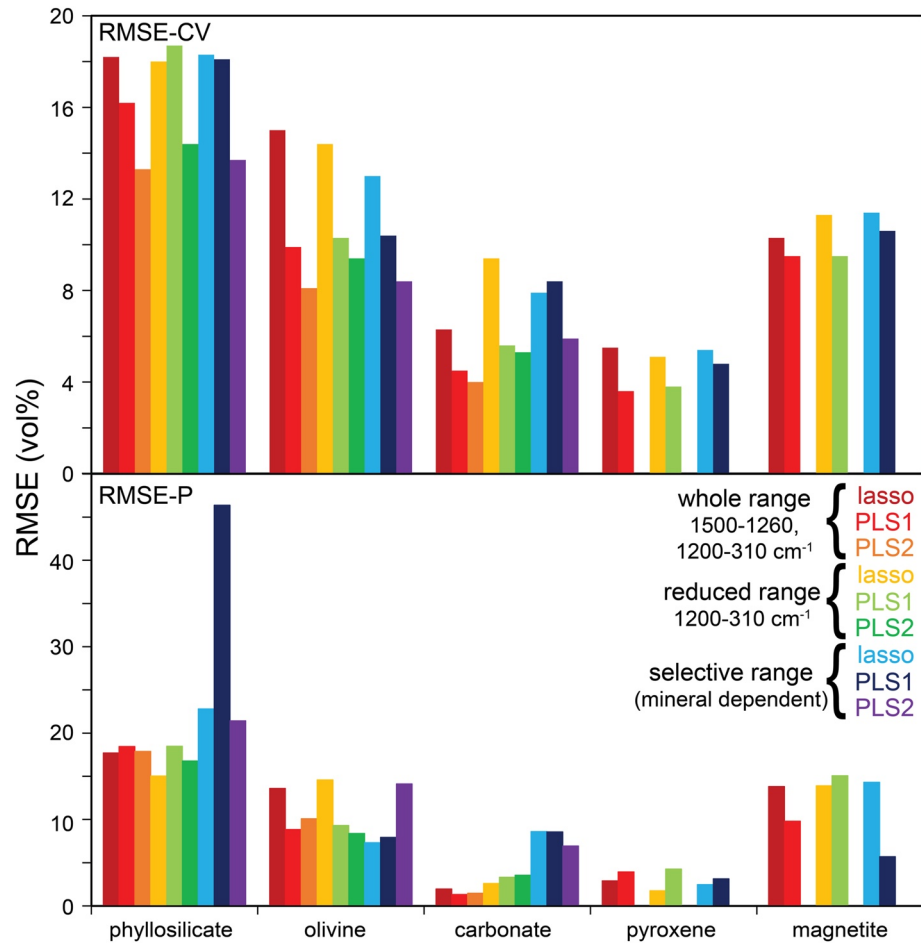
Figure 7. Root-mean-square error (RMSE)-CV (top) for lasso, PLS1, and PLS2 models of the 302-spectra training set at various spectral ranges. The selective spectral ranges for the models are 1,200–800, 700–400 cm^{-1} (phyllosilicate), 1,100–330 cm^{-1} (olivine), 1,000–600 cm^{-1} (carbonate), 1,200–600 cm^{-1} (pyroxene), and 800–310 cm^{-1} (magnetite). RMSE-P values (bottom) are generated by applying the training set models to the 15 CI/CM chondritic meteorite spectra and utilizing the predictions for the RMSE calculations.

models encompassing more than one mineral species (e.g., the phyllosilicate model including antigorite, cronstedtite, and saponite), and PLS1 was used for the models with only one mineral species (e.g., magnetite) (Table 1). After the parameters for the optimized models were selected, the 15 meteorite spectra were combined with the data set of 302 individual mineral and mineral mixture spectra to create optimized models. For each mineral group model, the predicted versus actual abundances (measured by weight or by quantitative XRD) are plotted for each mineral species in Figure 8 with associated RMSE-C and R^2 values. Note that Figure 8 summarizes internal model results. These predictions, RMSE-C, and R^2 values are less effective in evaluating the accuracy of the optimized models than RMSE-CV values (see Section 4.4).

We selected the whole spectral range (1,500–1,260, 1,200–310 cm^{-1}) owing to its optimized performance. The application of the models using the whole spectral range is admissible for averaged OTES spectra composed of hundreds of individual spectra such as those used in this work (Section 4.4). However, when applying machine learning models to individual OTES spectra that have different wavelength-dependent noise, an alternative type of model such as the reduced spectral range (1,200–310 cm^{-1}) should be considered.
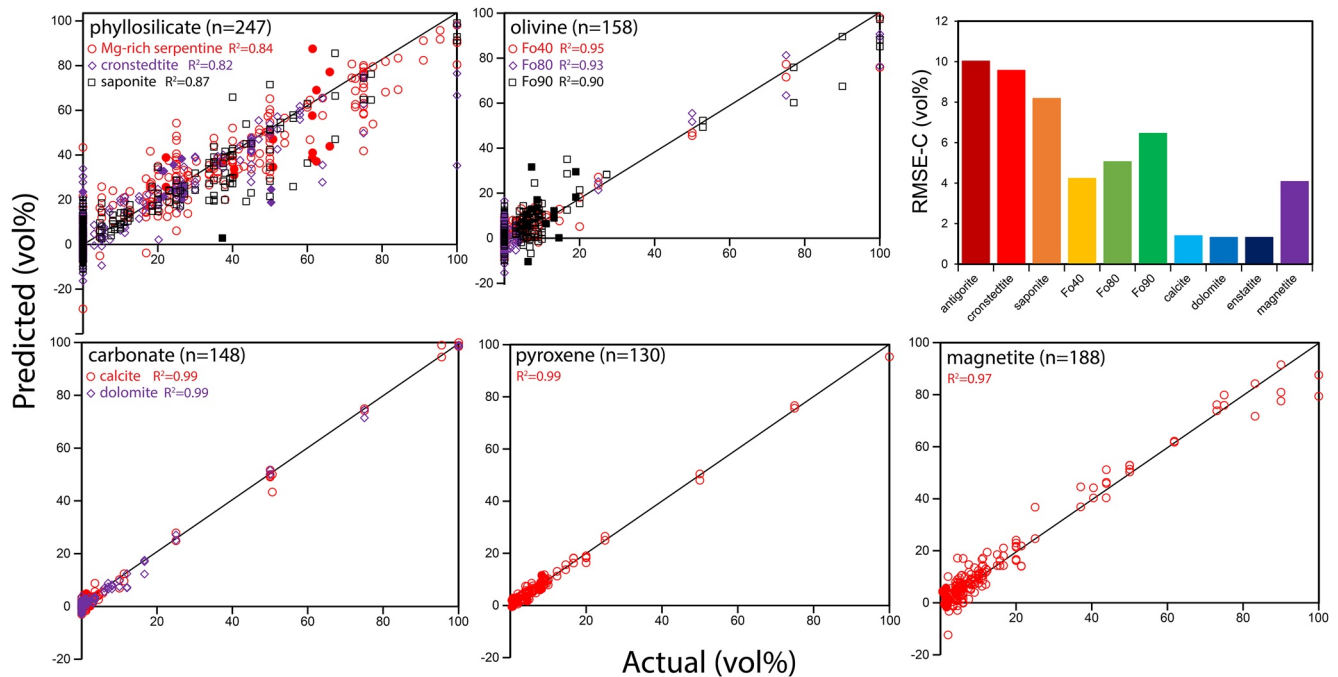
Figure 8. Predicted versus actual (weighed or quantitative X-ray diffraction) mineral abundances (vol%) from the phyllosilicate (PLS2), olivine (PLS2), carbonate (PLS2), pyroxene (PLS1), and magnetite (PLS1) models using the whole spectral range (1,500–1,260, 1,200–310 cm^{-1}). Root-mean-square error-C and R^2 values are reported. The data sets include end-members, mineral mixtures, and meteorites (data points with solid fill). Samples must have >0 vol% of the mineral group of interest to be included in the model. Due to the phyllosilicate contamination in pyrrhotite, samples with >10 vol% pyrrhotite were excluded from the phyllosilicate model.

4.4. Prediction of Bennu's Mineralogy

We applied the optimized machine learning models to three average OTES spectra: the global average, T1, and T2 (Hamilton et al., 2021) (Figure 9). Visually, the T1 and T2 spectra appear as end-members with the global average as a mixture of the T1 and T2 spectra.

The PLS predictions of the OTES spectra are summarized in Table 3. The prediction range for the mineral species between the three OTES spectra is narrow and predominantly within the error of the models. The global average
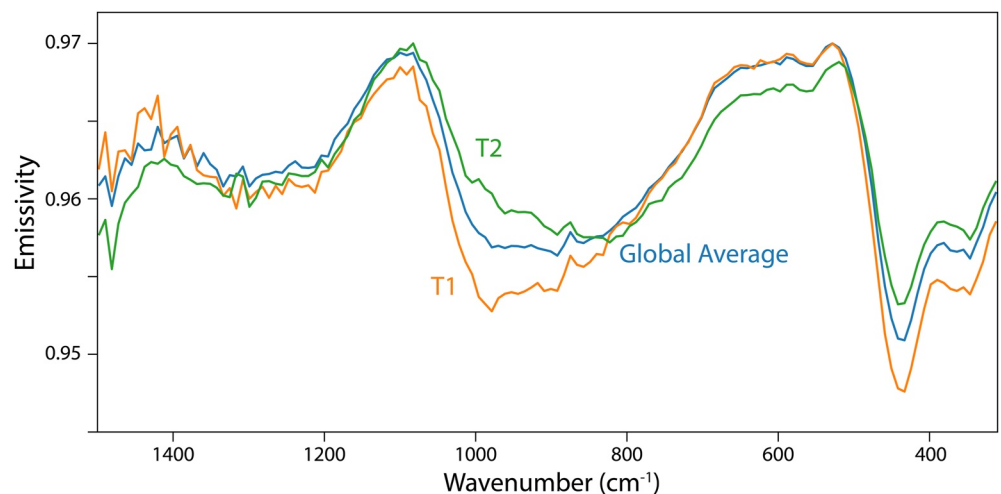
Figure 9. OSIRIS-REx Thermal Emission Spectrometer global average, T1, and T2 spectra of Bennu (Hamilton et al., 2021). Spectra are derived from the Equatorial Stations 3 data set from the OSIRIS-REx Detailed Survey. Spectra were normalized to an emissivity of 0.97 from 1,500 to 310 cm^{-1} to match the training set.

Table 3
Prediction of the Mineralogy of the Asteroid Bennu in Areal Percent Using the OTES Global Average, T1, and T2 Spectra

Prediction variable	RMSE-CV	Global average (areal %)	T1 (areal %)	T2 (areal %)
Mg-rich serpentine	13.3	56	61	51
Cronstedtite		16	15	16
Saponite		6	6	8
Olivine (Fo40)	8.1	2	2	3
Olivine (Fo80)		Not predicted	Not predicted	Not predicted
Olivine (Fo90)		7	3	13
Calcite	4	5	7	4
Dolomite		6	6	5
Pyroxene	3.6	Not predicted	Not predicted	<1
Magnetite	9.5	6	4	6
Total predicted abundance	n.a.	104	104	106

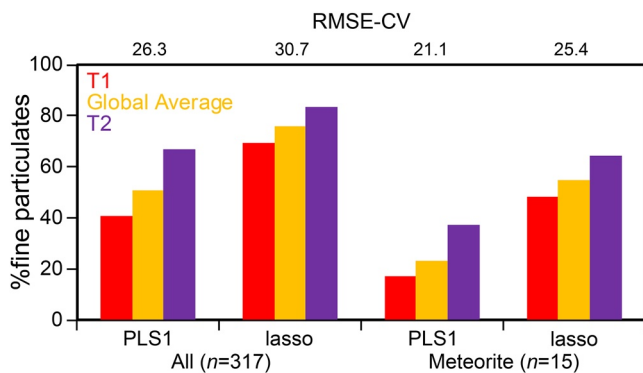*Note.* PLS models from Figure 8 were used in the predictions, and RMSE-CV values are reported.

predictions fall between the mineral abundance predictions of the T1 and T2 spectra. The optimized phyllosilicate model predicts the presence of both Mg-rich and Fe-rich phyllosilicates. For the olivine group, Fo40 and Fo90 are predicted, whereas Fo80 is not. Both calcite and dolomite are predicted at levels higher than observed in most CI/CM chondrites (Fendrich & Ebel, 2021; Howard et al., 2011; King et al., 2015). For the global average and T1 spectra, pyroxene is not predicted. The predicted magnetite content is consistent with CI chondrites and high for CM chondrites (Howard et al., 2011; King et al., 2015).

We trust model predictions within the range of RMSE-CV values for each model. Prediction values by mineral group (e.g., olivine) are more robust than by mineral species (e.g., Fo40, Fo80, and Fo90), given that similarities exist between mineral species that could offset predictions to favor one species over another. Given that the mineral abundance predictions were made using five separate mineral-group models, the total prediction abundance can exceed 100 vol% while falling within the model errors.

Figure 10. Prediction of percent fine particulates in vol% for the T1, global average, and T2 OSIRIS-REx Thermal Emission Spectrometer spectra using four different machine learning models. Both PLS1 and lasso models were constructed for data sets that included either all spectra (n = 317) or only meteorite spectra (n = 15).

4.5. Prediction of Bennu's Surface Particulate Size

In addition to the compositional models, we built four additional models with the aim of evaluating the effect of the proportion of coarse and fine particulates that contribute to OTES spectra. Two data sets ((i) all 317 spectra, (ii) 15 meteorite spectra only) and two machine learning model types ((i) PLS1, (ii) lasso) were tested, for a total of four models. A wavenumber range of 1,500–1,260, 1,200–310 cm^{-1} was used for all.

All four models predict the T1 spectrum as having a lower percentage of fine particulates compared to the T2 spectrum, with the global average as an in-between value (Figure 10). This study was not designed to estimate particulate sizes in the MIR, and therefore gaps are present within the range of available values of the response variable of interest (percent fine particles), reducing model accuracy. If additional mixtures existed between 0% and 100% fines, a more accurate model could likely be produced. The poor model accuracies are reflected with high RMSE-CV values for the four models (Figure 10). Due to the high RMSE-CV values, we are not confident in the absolute values to these predictions; however, the trend between the three OTES spectra is likely reliable. The prediction trends of all four models are consistent with the interpretation of Hamilton et al. (2021) that a greater proportion of fine particulates contribute to the T2 spectrum in comparison to the T1 spectrum.

5. Discussion

5.1. Factors Influencing the Machine Learning Models

There are many factors that impact machine learning multivariate analysis, including but not limited to spectral pre-processing, training set size, type of spectroscopy, and application (Breitenfeld et al., 2018; Dyar et al., 2012; Dyar & Ytsma, 2021). For example, with larger training sets, the accuracy of machine learning models improves (Dyar & Ytsma, 2021), incentivizing the inclusion of all available spectra. This study is the first attempt to use PLS to derive the mineralogy of fine and coarse particulates simultaneously in the MIR under SAE conditions, and therefore testing of model parameters is needed.

The application to FTIR/MIR spectroscopy is complex due to overlapping spectral features in the MIR compared to narrow features in LIBS and Raman spectra, where PLS and lasso models were previously tested on geological materials (Breitenfeld et al., 2018; Dyar et al., 2012). We investigated the effect of wavenumber range on the performance of the machine learning models. Previous workers have noted that isolating spectral features (narrowing spectral ranges) can improve the accuracy of the prediction model (Breitenfeld et al., 2018; Dyar et al., 2012). Based on our PLS test models, we determined that broadening the spectral wavenumber range improved model performance with lower RMSE-CV values (Figure 7). This testing is not in agreement with previous studies (Breitenfeld et al., 2018; Dyar et al., 2012). Hecker et al. (2012) and Pan et al. (2015) did not investigate the effects of the MIR spectral range on machine learning performance. This difference in performance with respect to spectral range may be due to the type of spectroscopy and its breadth of spectral features. PLS and lasso models may be optimized when isolating narrow features (e.g., Raman and LIBS), whereas this work suggests models of data sets with broad, overlapping spectral features (e.g., MIR) improve with broad spectral ranges. It is unclear whether MIR machine learning models always favor broader wavenumber ranges or whether this trend is specific to the training set of this study. A comparison study of a set of samples across spectroscopic techniques could address potential reasons for this discrepancy.

Lasso, PLS1, and PLS2 were tested to determine the best model type for this study (Figure 7). PLS2 outperformed the lasso and PLS1 techniques with lower RMSE-CV values. Pyroxene and magnetite PLS2 models do not exist because only a single variable is under investigation. For these models, PLS1 outperformed lasso.

For the optimized compositional models (Table 3), the phyllosilicate model's RMSE-CV value is highest. This model estimates abundances of three mineral species (Mg-rich serpentine, cronstedtite, and saponite) simultaneously. The RMSE-CV values become sequentially smaller from olivine to carbonate to pyroxene, matching the decrease in the number of prediction variables within each model. The magnetite model predicts a single mineral species, yet the associated RMSE-CV value is greater than the RMSE-CV values for the carbonate, pyroxene, and olivine models (Table 3). For this reason, we conclude that the optimized magnetite model is the poorest-performing optimized machine learning model in this study.

5.2. Analog Mineral Mixture Compared to OTES Global Average

Unlike LSMA methods, PLS cannot produce modeled spectra to directly compare to OTES spectra and to visually evaluate spectral model fit. There is no mineral mixture within the sample suite with precise mineralogy that matches the predicted mineralogy of the OTES spectra of Bennu. However, mixture #104 (Figure 11) has mineral abundances similar to the predicted composition of the global average, though it has lower Mg-rich serpentine, Fo40, and calcite content and higher pyroxene. Neither the coarse nor fine spectra of mixture #104 matches the OTES global average. Yet, a mixture of the coarse and fine particulate material with the same mineralogical proportions may share more spectral characteristics with the OTES global average spectrum. The fine spectrum of mixture #104 possesses the broader feature from $\sim$1,100–650 cm^{-1} that is similar to the global average, whereas the coarse spectrum has the diagnostic phyllosilicate feature at $\sim$460 cm^{-1}. It should be noted that a theoretical spectrum consisting of a mixture of coarse and fine particulates could have attributes of the both types of materials that mix nonlinearly across wavelengths, resulting in certain sections of the spectrum that resemble the fine and coarse spectra. The 52 spectra of mixtures consisting of coarse and fine particulates together that we included in our training set improve the likelihood that the machine learning models could properly account for the mixing of fine and coarse particulates.
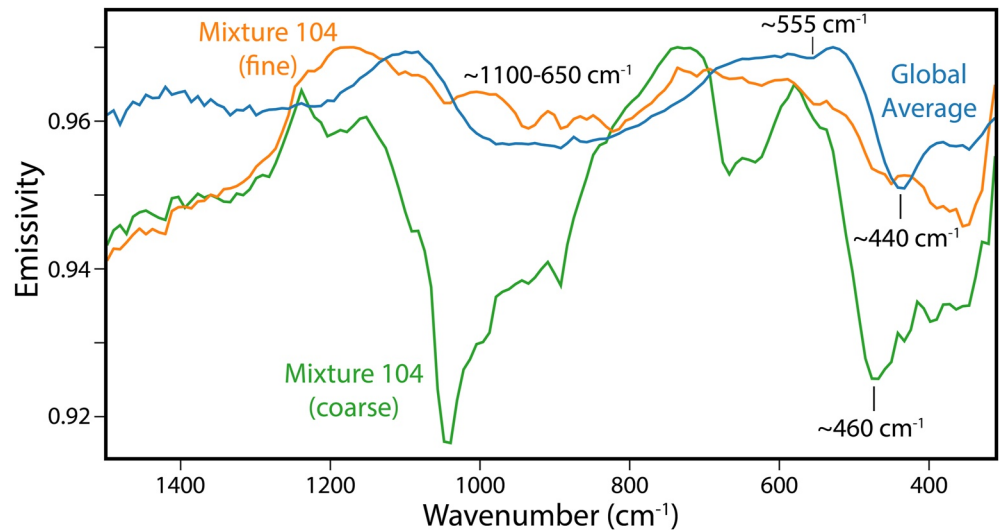
Figure 11. OSIRIS-REx Thermal Emission Spectrometer global average spectrum of Bennu with simulated asteroid environment mid-infrared coarse and fine analog laboratory spectra (mixture #104). Mixture #104 consists of 45% antigorite (Mg-rich serpentine), 17.5% cronstedtite, 8.5% saponite, 5% Fo40%, 5% Fo90%, 8% dolomite, 5% enstatite, and 6% magnetite by volume. Abundances are normalized to remove 11 vol% carbon powder that is present in mixture #104.

A limitation of this work can be observed in Figure 11. The ~460 cm^{-1} phyllosilicate feature in the mixture #104 spectra is offset from its equivalent feature in the OTES spectra at ~440 cm^{-1}. Reasons for this may be due to differences in phyllosilicate end-member mineralogy (Bishop et al., 2008). We unsuccessfully attempted to find a suitable terrestrial phyllosilicate analog with the ~440 cm^{-1} feature position. Despite the differences between the OTES global average and the mixture #104 spectrum, it should be noted that both the coarse and fine versions of mixture #104 have 6 vol% magnetite and exhibit the 555 cm^{-1} feature diagnostic of magnetite. This magnetite feature is present within the OTES spectra. This indicates that in materials with compositions similar to CI/CM chondrites, MIR detection of magnetite can be made at abundances as low as 6 vol%.

5.3. Interpretation of Bennu's Mineralogy and Surface Particulate Sizes

Carbonaceous chondrites, particularly CI chondrites, are primitive and have chemical compositions that are representative of the proto-planetary disk. Early Solar System conditions and processes like aqueous alteration influenced these materials. Therefore, these types of meteorites (and their parent asteroids) hold valuable evidence of the early Solar System environment. Determining whether the composition of Bennu falls into the category of a CI or a CM chondrite would help constrain conditions that Bennu's parent body experienced, including temperature and water/rock ratios. Zolensky et al. (1993) estimated that CM chondrites experienced temperatures <50°C with low water/rock ratios, whereas CI chondrite temperatures ranged from 50°C to 150°C with high water/rock ratios.

Our modal mineralogy predictions of the OTES spectra of Bennu (Table 3) are consistent with the mineralogy of carbonaceous chondritic meteorites (Figure 12). In comparison to the meteorite samples, phyllosilicate, olivine, and pyroxene contents fall within expected ranges.

Pyroxene was not predicted for the T1 and global average spectra. For the T2 spectrum, pyroxene was estimated to be <1 vol%. Although this prediction falls within the model error, a future investigation linking areas of Bennu where DellaGiustina et al. (2021) have detected pyroxene to OTES spectral shape may shed light on the T2 pyroxene prediction.

The prediction of carbonate content is high in comparison to typical CI/CM chondrites (Howard et al., 2011; King et al., 2015), though high carbonate content is a characteristic of one of the lithologies of the unclassified chondrite Tagish Lake (Bland et al., 2004). A possible explanation for the discrepancy between the predictions
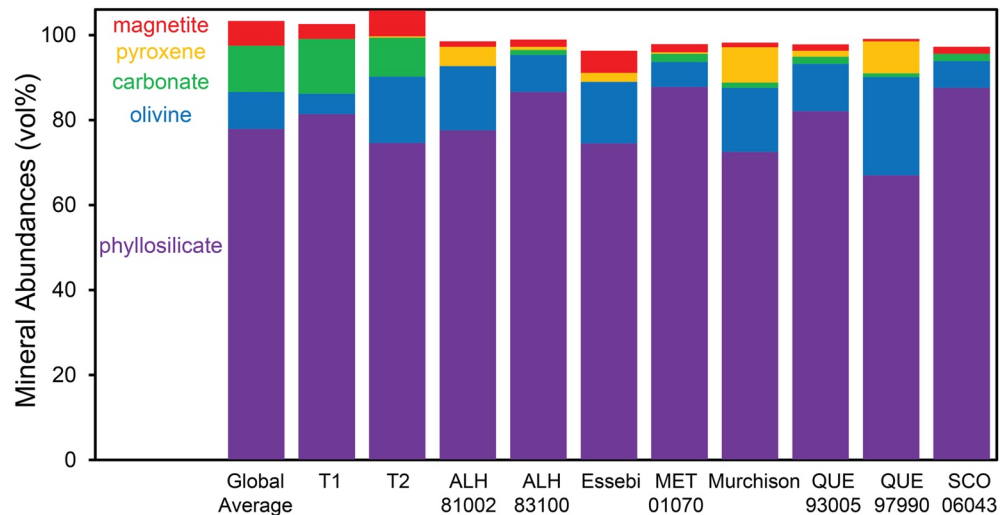
Figure 12. Mineral abundances (vol%) of the global average, T1, and T2 OSIRIS-REx Thermal Emission Spectrometer spectra from machine learning predictions in comparison to carbonaceous chondrite abundances from quantitative X-ray diffraction data (Howard et al., 2011).

and the meteorite collection is that portions of asteroids like Bennu with higher abundances of carbonates (e.g., veins; Kaplan et al., 2020) could be weaker, preventing their preservation in the meteorite record. This carbonate content discrepancy could also potentially be reinforced by low-albedo materials (e.g., carbonate potentially present within dark matrix materials) having smaller MIR spectral contributions than brighter materials (e.g., carbonate veins).

The magnetite contents of the Bennu predictions are high in comparison to CM chondrites and the meteorites included in our models. Higher magnetite content of >5 vol% has been identified in unclassified carbonaceous chondrites (e.g., Essebi) and CI chondrites (e.g., Orgueil) (Bland et al., 2004; King et al., 2015). The identification of magnetite by Hamilton et al. (2019), reinforces this interpretation of high magnetite in comparison to many carbonaceous chondrites.

The phyllosilicate model predicts the presence of Mg-rich serpentine, cronstedtite, and saponite. Typically, CM chondrites have Mg-serpentine and cronstedtite present, whereas CI chondrites have a mixture of serpentine and saponite (Howard et al., 2011; King et al., 2015). However, there are atypical CM and C2-ungrouped meteorites (e.g., Essebi, Bells) that lack well-crystalline cronstedtite in favor of disordered fine particulate Mg-serpentine and saponite (Howard et al., 2011). The low saponite abundance associated with the prediction of high serpentine content may indicate consistency with CM chondrites (Howard et al., 2011).

The Bennu olivine predictions are high in comparison to CI chondrites (King et al., 2015), yet the values fall within the observed range of some CM chondrites (Howard et al., 2011). With the exception of the T2 spectrum, estimates of anhydrous silicate content of ≤ 10 vol% are consistent with Hamilton et al.'s (2021) interpretations. The higher prediction of olivine for the T2 spectrum may be due to the longer wavelength band minimum of the broad bowl-shaped feature ($\sim 1,100$–650 cm^{-1}), given that for the silicate stretching region, olivine has the longest wavelength bands.

Hamilton et al. (2021) attributed the longer wavelength band minimum of the T2 broad bowl-shaped feature to particle size, and therefore further investigation is necessary to determine whether this variation is due to composition and/or particle size. Overall, Hamilton et al. (2021) interpreted the T1 OTES spectrum as representative of primarily coarse to rocky materials, whereas T2 includes a fine particulate coating on coarser materials. This interpretation holds for the PLS1 and lasso machine learning models designed for the prediction of the volume percentage of fine and coarse particulates sensitive to MIR spectral measurements (Figure 10). The narrow quantitative mineralogic variations between the type spectra and the particulate size abundance trend are consistent with Hamilton et al. (2021), indicating only minor compositional variations across Bennu.

5.4. Implications for Bennu's Aqueous Alteration and Space Weathering

Based on the quantitative mineralogy predictions, Bennu has properties of both CI, CM, and ungrouped carbonaceous chondrites. The distinction between CM and CI chondrites is the degree of aqueous alteration the asteroid experienced.

The PLS predictions of high phyllosilicate content, low anhydrous silicate abundances, and a greater proportion of Mg to Fe in phyllosilicates indicate high degrees of aqueous alteration (Browning et al., 1996; Howard et al., 2009). Howard et al. (2009) argued that the total phyllosilicate abundance is an unambiguous metric to indicate the degree of aqueous alteration. The PLS prediction of Fe-rich and Mg-rich olivine on Bennu does not clearly indicate high or low degrees of aqueous alteration. There are other methods for determining the degree of aqueous alteration, such as Fe/Si ratios (McSween, 1979a), oxidation of metallic Fe-Ni, and phyllosilicate/sulfide ratios of intergrown serpentine and tochilinite (Rubin et al., 2007). However, these indicators cannot be utilized for understanding the degree of aqueous alteration of Bennu until a sample is returned from the asteroid. Rubin et al. (2007) highlighted carbonate composition as another metric for aqueous alteration. Using this metric from Rubin et al. (2007), the prediction of dolomite indicates higher degrees of aqueous alteration. Although the relative abundances of the carbonate species on Bennu could vary based on error, the detection of dolomite exceeds the model error, indicating its presence is likely. Rubin et al. (2007) noted that dolomite is rare or absent in CM chondrites with low degrees of alteration. For meteoritic studies, Rubin et al.'s (2007) classification has been favored due to the difficulty in quantifying phyllosilicates. However, because the PLS phyllosilicate model allows us to estimate total phyllosilicate abundances and phyllosilicate species abundances, Howard et al.'s (2009) methodology is preferred here indicating a high degree of aqueous alteration.

In addition to aqueous alteration, the modal mineralogy of bodies like Bennu holds information about processes that have altered the asteroid over long time scales, such as space weathering. Using Hapke radiative transfer modeling, Trang et al. (2021) determined that the space weathering–byproduct magnetite may cause Bennu's blue VNIR spectral slopes. Not only do the PLS models predict magnetite on the surface of Bennu they also predict cronstedtite (Table 3), an Fe^{3+}-phyllosilicate that is needed for the formation of magnetite through space weathering. Alternatively, the formation of magnetite may be due to aqueous alteration (Kerridge et al., 1979; Takir et al., 2013) with increased magnetite content linked to higher degrees of alteration (McSween, 1979b). Deposition of magnetite into the matrix likely occurs when Fe, S, and Ni are removed from serpentine or tochilinite (Tomeoka & Buseck, 1985).

6. Conclusions and Future Work

The relative effectiveness of different machine learning parameters for the analysis of geologic materials using spectroscopic data is technique- and/or application-dependent, as optimized models and model performance vary across studies (Breitenfeld et al., 2018; Clegg et al., 2009; Dyar & Ytsma, 2021; Dyar et al., 2016; Pan et al., 2015). For the training set presented in this work, PLS consistently outperforms lasso. This is the first study to use machine learning in the MIR to investigate geologic materials that tested model parameters such as spectral range. In the MIR, features are broad and overlapping, unlike in previous applications of machine learning such as to Raman and LIBS. PLS (Wold et al., 1983) utilizes all channels for its prediction, capturing entire features, whereas lasso (Tibshirani, 1996) reduces the number of channels utilized in the prediction. The relative efficacy of PLS and lasso for different spectroscopic techniques needs to be explored further.

Quantification of mineralogy for fines in the MIR is complex due to nonlinear mixing behaviors (Salisbury & Wald, 1992; Thomson & Salisbury, 1993). The multivariate machine learning models used in this study have comparable accuracy to LSMA models typically used to quantify mineralogy from MIR spectra of coarse particulate materials; however, LSMA was not tested on our spectra of physical mixtures (Ramsey & Christensen, 1998; Rogers & Aharonson, 2008). A major success of this study is the demonstration of the ability to quantify fine and coarse particulates simultaneously. A benefit of applying machine learning techniques to spectral data includes the ability to predict materials mixing nonlinearly with high accuracy (low RMSE-CV values). A negative aspect related to utilizing machine learning techniques on spectral data sets of geologic samples includes time-consuming sample suite and training set preparation. This study demonstrates the effectiveness of utilizing machine learning techniques for mineral quantification in the MIR, providing potential promise for the application of this

methodology to MIR spectral data sets of other planetary bodies (e.g., Moon and Mars). Our MIR spectral training set acts as a base to be built upon for understanding similar materials within our Solar System.

Future studies can employ the machine learning models created in this study for non-destructive laboratory investigations of CI/CM chondrites or returned sample from Bennu. These models will be particularly useful for relating laboratory measurements to the OTES remote sensing data. Additionally, for asteroids that have similar compositions to Bennu, the machine learning models could be applied to MIR telescopic observations. Factors such as signal-to-noise ratios and different energy ranges may complicate application and require model refinement; however, this methodology may yield the ability to quantify mineralogy of asteroids with coexisting fine and coarse materials.

Modal mineralogy predictions from the global average, T1, and T2 OTES spectra (Hamilton et al., 2021) of the asteroid Bennu are consistent with those for carbonaceous chondritic materials (CI, CM, ungrouped). A noteworthy conclusion of this study includes the prediction of carbonates at abundances greater than the model error. This prediction using OTES data reinforces the detection of carbonates made by Kaplan et al. (2020) through OVIRS data. Overall, the predicted modal mineralogy of Bennu indicates that the asteroid experienced high degrees of aqueous alteration. Small variations exist in the mineralogical predictions for the global average, T1, and T2 spectra of Bennu. However, the compositional prediction variations fall within model error, and therefore differences between the OTES spectra could be caused by minor compositional variations or by a non-compositional property such as particle size. The predictions of the percentage of fine particulates for the Bennu spectra follow the trend interpreted by Hamilton et al. (2021), with T2 having a higher proportion of fines than T1. Further investigation of the role of composition and particle size will be attempted in the future through mapping the surface of Bennu.

The asteroid's compositional variability will be further investigated by applying the optimized PLS models from this study to individual OTES spectra to search for unique OTES spectra and create mineralogy maps of Bennu. Creating mineral maps of the surface of Bennu will allow for decoupling the effects of particle size and composition on the surface by connecting the maps to the OTES T1 and T2 spectra. Compositional maps of the asteroid Bennu will improve our understanding of the spatial nature of processes the rubble pile has experienced such as aqueous alteration.

Data Availability Statement

The OTES global average, T1, and T2 spectra are described by Hamilton et al. (2021) and derived from data available in the Planetary Data System at https://arcnav.psi.edu/urn:nasa:pds:context:instrument:otes.orex (Christensen et al., 2019). All supplementary tables and figures (laboratory spectra and associated mineral abundances) are archived in an external data repository at https://doi.org/10.5281/zenodo.5712028.

References

Bates, H. C., King, A. J., Donaldson Hanna, K. L., Bowles, N. E., & Russell, S. S. (2020). Linking mineralogy and spectroscopy of highly aqueously altered CM and CI carbonaceous chondrites in preparation for primitive asteroid sample return. *Meteoritics & Planetary Sciences*, *55*(1), 77–101. https://doi.org/10.1111/maps.13411

Binzel, R. P., DeMeo, F. E., Burt, B. J., Cloutis, E. A., Rozitis, B., Burbine, T. H., et al. (2015). Spectral slope variations for OSIRIS-REx target asteroid (101955) Bennu: Possible evidence for a fine-grained regolith equatorial ridge. *Icarus*, *256*, 22–29. https://doi.org/10.1016/j.icarus.2015.04.011

Bishop, J. L., Lane, M. D., Dyar, M. D., & Brown, A. J. (2008). Reflectance and emission spectroscopy study of four groups of phyllosilicates: Smectites, kaolinite-serpentines, chlorites and micas. *Clay Minerals*, *43*(1), 35–54. https://doi.org/10.1180/claymin.2008.043.1.03

Bland, P. A., Cressey, G., & Menzies, O. N. (2004). Modal mineralogy of carbonaceous chondrites by X-ray diffraction and Mössbauer spectroscopy. *Meteoritics & Planetary Sciences*, *39*(1), 3–16. https://doi.org/10.1111/j.1945-5100.2004.tb00046.x

Breitenfeld, L. B., Dyar, M. D., Carey, C. J., Tague, T. J., Wang, P., Mullen, T., & Parente, M. (2018). Predicting olivine composition using Raman spectroscopy through band shift and multivariate analyses. *American Mineralogist*, *103*(11), 1827–1836. https://doi.org/10.2138/am-2018-6291

Browning, L. B., McSween, H. Y., Jr, & Zolensky, M. E. (1996). Correlated alteration effects in CM carbonaceous chondrites. *Geochimica et Cosmochimica Acta*, *60*(14), 2621–2633. https://doi.org/10.1016/0016-7037(96)00121-4

Christensen, P. R., Hamilton, V. E., Anwar, S., Mehall, G., & Lauretta, D. S. (2019). Origins, spectral interpretation, resource identification, security, regolith explorer (OSIRIS-REx): OSIRIS-REx thermal emission spectrometer bundle. *NASA Planetary Data System*, urn:nasa:pds:orex.otes.

Christensen, P. R., Hamilton, V. E., Mehall, G. L., Pelham, D., O'Donnell, W., Anwar, S., et al. (2018). The OSIRIS-REx thermal emission spectrometer (OTES) instrument. *Space Science Reviews*, *214*(5), 1–39. https://doi.org/10.1007/s11214-018-0513-6

Clark, R. N., & Roush, T. L. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research*, *89*(B7), 6329–6340. https://doi.org/10.1029/jb089ib07p06329

Clegg, S. M., Sklute, E., Dyar, M. D., Barefield, J. E., & Wiens, R. C. (2009). Multivariate analysis of remote laser-induced breakdown spectroscopy spectra using partial least squares, principal component analysis, and related techniques. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *64*(1), 79–88. https://doi.org/10.1016/j.sab.2008.10.045

Cloutis, E. A., Hiroi, T., Gaffey, M. J., Alexander, C. O. D., & Mann, P. (2011). Spectral reflectance properties of carbonaceous chondrites: 1. CI chondrites. *Icarus*, *212*(1), 180–209. https://doi.org/10.1016/j.icarus.2010.12.009

Cloutis, E. A., Hudon, P., Hiroi, T., Gaffey, M. J., & Mann, P. (2011). Spectral reflectance properties of carbonaceous chondrites: 2. CM chondrites. *Icarus*, *216*(1), 309–346. https://doi.org/10.1016/j.icarus.2011.09.009

Conel, J. E. (1969). Infrared emissivities of silicates: Experimental results and a cloudy atmosphere model of spectral emission from condensed particulate mediums. *Journal of Geophysical Research*, *74*(6), 1614–1634. https://doi.org/10.1029/jb074i006p01614

Cooper, B. L., Salisbury, J. W., Killen, R. M., & Potter, A. E. (2002). Midinfrared spectral features of rocks and their powders. *Journal of Geophysical Research*, *107*(E4), 1-1–1-17. https://doi.org/10.1029/2000je001462

DellaGiustina, D. N., Kaplan, H. H., Simon, A. A., Bottke, W. F., Avdellidou, C., Delbo, M., et al. (2021). Exogenic basalt on asteroid (101955) Bennu. *Nature Astronomy*, *5*(1), 31–38. https://doi.org/10.1038/s41550-020-1195-z

DeMeo, F. E., Binzel, R. P., Slivan, S. M., & Bus, S. J. (2009). An extension of the Bus asteroid taxonomy into the near-infrared. *Icarus*, *202*(1), 160–180. https://doi.org/10.1016/j.icarus.2009.02.005

Donaldson Hanna, K. L., Bowles, N. E., Warren, T. J., Hamilton, V. E., Schrader, D. L., McCoy, T. J., et al. (2021). Spectral characterization of Bennu analogs using PASCALE: A new experimental set-up for simulating the near-surface conditions of airless bodies. *Journal of Geophysical Research: Planets*, *126*(2). https://doi.org/10.1029/2020je006624

Donaldson Hanna, K. L., Greenhagen, B. T., Patterson, W. R., Pieters, C. M., Mustard, J. F., Bowles, N. E., et al. (2017). Effects of varying environmental conditions on emissivity spectra of bulk lunar soils: Application to Diviner thermal infrared observations of the Moon. *Icarus*, *283*, 326–342. https://doi.org/10.1016/j.icarus.2016.05.034

Donaldson Hanna, K. L., Schrader, D. L., Cloutis, E. A., Cody, G. D., King, A. J., McCoy, T. J., et al. (2019). Spectral characterization of analog samples in anticipation of OSIRIS-REx's arrival at Bennu: A blind test study. *Icarus*, *319*, 701–723. https://doi.org/10.1016/j.icarus.2018.10.018

Dyar, M. D., Breves, E. A., Gunter, M. E., Lanzirotti, A., Tucker, J. M., Carey, C. J., et al. (2016). Use of multivariate analysis for synchrotron micro-XANES analysis of iron valence state in amphiboles. *American Mineralogist*, *101*(5), 1171–1189. https://doi.org/10.2138/am-2016-5556

Dyar, M. D., Carmosino, M. L., Breves, E. A., Ozanne, M. V., Clegg, S. M., & Wiens, R. C. (2012). Comparison of partial least squares and lasso regression techniques as applied to laser-induced breakdown spectroscopy of geological samples. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *70*, 51–67. https://doi.org/10.1016/j.sab.2012.04.011

Dyar, M. D., & Ytsma, C. R. (2021). Effect of data set size on geochemical quantification accuracy with laser-induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *177*, 106073. https://doi.org/10.1016/j.sab.2021.106073

Emery, J. P., Fernández, Y. R., Kelley, M. S. P., Warden, K. T., Hergenrother, C., Lauretta, D. S., et al. (2014). Thermal infrared observations and thermophysical characterization of OSIRIS-REx target asteroid (101955) Bennu. *Icarus*, *234*, 17–35. https://doi.org/10.1016/j.icarus.2014.02.005

Feely, K. C., & Christensen, P. R. (1999). Quantitative compositional analysis using thermal emission spectroscopy: Application to igneous and metamorphic rocks. *Journal of Geophysical Research*, *104*(E10), 24195–24210. https://doi.org/10.1029/1999je001034

Fendrich, K. V., & Ebel, D. S. (2021). Comparison of the Murchison CM2 and Allende CV3 chondrites. *Meteoritics & Planetary Sciences*, *56*(1), 77–95. https://doi.org/10.1111/maps.13623

Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, *185*, 1–17. https://doi.org/10.1016/0003-2670(86)80028-9

Glotch, T. D., Edwards, C. S., Yesiltas, M., Shirley, K. A., McDougall, D. S., Kling, A. M., et al. (2018). MGS-TES spectra suggest a basaltic component in the regolith of Phobos. *Journal of Geophysical Research: Planets*, *123*(10), 2467–2484. https://doi.org/10.1029/2018je005647

Hamilton, V. E., Christensen, P. R., Kaplan, H. H., Haberle, C. W., Rogers, A. D., Glotch, T. D., et al. (2021). Evidence for limited compositional and particle size variation on asteroid (101955) Bennu from thermal infrared spectroscopy. *Astronomy & Astrophysics*. https://doi.org/10.1051/0004-6361/202039728

Hamilton, V. E., Haberle, C. W., & Mayerhöfer, T. G. (2020). Effects of small crystallite size on the thermal infrared (vibrational) spectra of minerals. *American Mineralogist: Journal of Earth and Planetary Materials*, *105*(11), 1756–1760. https://doi.org/10.2138/am-2020-7602

Hamilton, V. E., Simon, A. A., Christensen, P. R., Reuter, D. C., Clark, B. E., Barucci, M. A., et al. (2019). Evidence for widespread hydrated minerals on asteroid (101955) Bennu. *Nature Astronomy*, *3*(4), 332–340. https://doi.org/10.1038/s41550-019-0722-2

Hanna, R. D., Hamilton, V. E., Haberle, C. W., King, A. J., Abreu, N. M., & Friedrich, J. M. (2020). Distinguishing relative aqueous alteration and heating among CM chondrites with IR spectroscopy. *Icarus*, *346*, 113760. https://doi.org/10.1016/j.icarus.2020.113760

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media.

Hecker, C., Dilles, J. H., van der Meijde, M., & van der Meer, F. D. (2012). Thermal infrared spectroscopy and partial least squares regression to determine mineral modes of granitoid rocks. *Geochemistry, Geophysics, Geosystems*, *13*(3). https://doi.org/10.1029/2011gc004004

Henderson, B. G., & Jakosky, B. M. (1994). Near-surface thermal gradients and their effects on mid-infrared emission spectra of planetary surfaces. *Journal of Geophysical Research*, *99*(E9), 19063–19073. https://doi.org/10.1029/94je01861

Henderson, B. G., & Jakosky, B. M. (1997). Near-surface thermal gradients and mid-IR emission spectra: A new model including scattering and application to real data. *Journal of Geophysical Research*, *102*(E3), 6567–6580. https://doi.org/10.1029/96je03781

Hergenrother, C. W., Maleszewski, C. K., Nolan, M. C., Li, J. Y., d'Aubigny, C. D., Shelly, F. C., et al. (2019). The operational environment and rotational acceleration of asteroid (101955) Bennu from OSIRIS-REx observations. *Nature Communications*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-09213-x

Howard, K. T., Alexander, C. O. D., Schrader, D. L., & Dyl, K. A. (2015). Classification of hydrous meteorites (CR, CM and C2 ungrouped) by phyllosilicate fraction: PSD-XRD modal mineralogy and planetesimal environments. *Geochimica et Cosmochimica Acta*, *149*, 206–222. https://doi.org/10.1016/j.gca.2014.10.025

Howard, K. T., Benedix, G. K., Bland, P. A., & Cressey, G. (2009). Modal mineralogy of CM2 chondrites by X-ray diffraction (PSD-XRD). Part 1: Total phyllosilicate abundance and the degree of aqueous alteration. *Geochimica et Cosmochimica Acta*, *73*(15), 4576–4589. https://doi.org/10.1016/j.gca.2009.04.038

Howard, K. T., Benedix, G. K., Bland, P. A., & Cressey, G. (2011). Modal mineralogy of CM chondrites by X-ray diffraction (PSD-XRD): Part 2. Degree, nature and settings of aqueous alteration. *Geochimica et Cosmochimica Acta*, *75*(10), 2735–2751. https://doi.org/10.1016/j.gca.2011.02.021

Hunt, G. R., & Logan, L. M. (1972). Variation of single particle mid-infrared emission spectrum with particle size. *Applied Optics*, *11*(1), 142–147. https://doi.org/10.1364/ao.11.000142

Kaplan, H. H., Lauretta, D. S., Simon, A. A., Hamilton, V. E., DellaGiustina, D. N., Golish, D. R., et al. (2020). Bright carbonate veins on asteroid (101955) Bennu: Implications for aqueous alteration history. *Science*, *370*(6517). https://doi.org/10.1126/science.abc3557

Kerridge, J. F., Mackay, A. L., & Boynton, W. V. (1979). Magnetite in CI carbonaceous meteorites: Origin by aqueous activity on a planetesimal surface. *Science*, *205*(4404), 395–397. https://doi.org/10.1126/science.205.4404.395

King, A. J., Schofield, P. F., Howard, K. T., & Russell, S. S. (2015). Modal mineralogy of CI and CI-like chondrites by X-ray diffraction. *Geochimica et Cosmochimica Acta*, *165*, 148–160. https://doi.org/10.1016/j.gca.2015.05.038

Lantz, C., Binzel, R. P., & DeMeo, F. E. (2018). Space weathering trends on carbonaceous asteroids: A possible explanation for Bennu's blue slope? *Icarus*, *302*, 10–17. https://doi.org/10.1016/j.icarus.2017.11.010

Lauretta, D. S., Bartels, A. E., Barucci, M. A., Bierhaus, E. B., Binzel, R. P., Bottke, W. F., et al. (2015). The OSIRIS-REx target asteroid (101955) Bennu: Constraints on its physical, geological, and dynamical nature from astronomical observations. *Meteoritics & Planetary Sciences*, *50*(4), 834–849. https://doi.org/10.1111/maps.12353

Lauretta, D. S., DellaGiustina, D. N., Bennett, C. A., Golish, D. R., Becker, K. J., Balram-Knutson, S. S., et al. (2019). The unexpected surface of asteroid (101955) Bennu. *Nature*, *568*(7750), 55–60. https://doi.org/10.1038/s41586-019-1033-6

Lauretta, D. S., Hergenrother, C. W., Chesley, S. R., Leonard, J. M., Pelgrift, J. Y., Adam, C. D., et al. (2019). Episodes of particle ejection from the surface of the active asteroid (101955) Bennu. *Science*, *366*(6470).

Lawson, C. L., & Hanson, R. J. (1974). *Solving least squares problems*. Prentice-Hall.

Logan, L. M., & Hunt, G. R. (1970). Emission spectra of particulate silicates under simulated lunar conditions. *Journal of Geophysical Research*, *75*(32), 6539–6548. https://doi.org/10.1029/jb075i032p06539

Logan, L. M., Hunt, G. R., Salisbury, J. W., & Balsamo, S. R. (1973). Compositional implications of Christiansen frequency maximums for infrared remote sensing applications. *Journal of Geophysical Research*, *78*(23), 4983–5003. https://doi.org/10.1029/jb078i023p04983

Lucey, P. G., Greenhagen, B. T., Song, E., Arnold, J. A., Lemelin, M., Hanna, K. D., et al. (2017). Space weathering effects in Diviner Lunar Radiometer multispectral infrared measurements of the lunar Christiansen Feature: Characteristics and mitigation. *Icarus*, *283*, 343–351. https://doi.org/10.1016/j.icarus.2016.05.010

Lyon, R. J. P. (1964). *Evaluation of infrared spectrophotometry for compositional analysis of lunar and planetary soils. Part II Rough and powdered surfaces* (p. 264). NASA Rep. CR-100.

McAdam, M. M., Sunshine, J. M., Howard, K. T., & McCoy, T. M. (2015). Aqueous alteration on asteroids: Linking the mineralogy and spectroscopy of CM and CI chondrites. *Icarus*, *245*, 320–332. https://doi.org/10.1016/j.icarus.2014.09.041

McSween, H. Y., Jr. (1979a). Are carbonaceous chondrites primitive or processed? A review. *Reviews of Geophysics*, *17*(5), 1059–1078. https://doi.org/10.1029/rg017i005p01059

McSween, H. Y., Jr. (1979b). Alteration in CM carbonaceous chondrites inferred from modal and chemical variations in matrix. *Geochimica et Cosmochimica Acta*, *43*(11), 1761–1770. https://doi.org/10.1016/0016-7037(79)90024-3

Molaro, J. L., Hergenrother, C. W., Chesley, S. R., Walsh, K. J., Hanna, R. D., Haberle, C. W., et al. (2020a). Thermal fatigue as a driving mechanism for activity on asteroid Bennu. *Journal of Geophysical Research: Planets*, *125*(8), e2019JE006325. https://doi.org/10.1029/2019je006325

Molaro, J. L., Walsh, K. J., Jawin, E. R., Ballouz, R.-L., Bennett, C. A., DellaGiustina, D. N., et al. (2020b). In situ evidence of thermally induced rock breakdown widespread on Bennu's surface. *Nature Communications*, *11*(1), 1–11. https://doi.org/10.1038/s41467-020-16528-7

Mustard, J. F., & Hays, J. E. (1997). Effects of hyperfine particles on reflectance spectra from 0.3 to 25 µm. *Icarus*, *125*(1), 145–163. https://doi.org/10.1006/icar.1996.5583

Pan, C., Rogers, A. D., & Thorpe, M. T. (2015). Quantitative compositional analysis of sedimentary materials using thermal emission spectroscopy: 2. Application to compacted fine-grained mineral mixtures and assessment of applicability of partial least squares methods. *Journal of Geophysical Research: Planets*, *120*(11), 1984–2001. https://doi.org/10.1002/2015je004881

Praet, A., Barucci, M. A., Clark, B. E., Kaplan, H. H., Simon, A. A., Hamilton, V. E., et al. (2021). Hydrogen abundance estimation and distribution on (101955) Bennu. *Icarus*, *363*, 114427. https://doi.org/10.1016/j.icarus.2021.114427

Ramsey, M. S., & Christensen, P. R. (1998). Mineral abundance determination: Quantitative deconvolution of thermal emission spectra. *Journal of Geophysical Research*, *103*(B1), 577–596. https://doi.org/10.1029/97jb02784

Realmuto, V. J. (1990). Separating the effects of temperature and emissivity: Emissivity spectrum normalization. In *Proceedings of the second TIMS workshop* (pp. 23–27). JPL Publication.

Reuter, D. C., Simon, A. A., Hair, J., Lunsford, A., Manthripragada, S., Bly, V., et al. (2018). The OSIRIS-REx Visible and InfraRed Spectrometer (OVIRS): Spectral maps of the asteroid Bennu. *Space Science Reviews*, *214*(2), 1–22. https://doi.org/10.1007/s11214-018-0482-9

Rogers, A. D., & Aharonson, O. (2008). Mineralogical composition of sands in Meridiani Planum determined from Mars Exploration Rover data and comparison to orbital measurements. *Journal of Geophysical Research*, *113*(E6). https://doi.org/10.1029/2007je002995

Rosipal, R., & Krämer, N. (2005). Overview and recent advances in partial least squares. In *International statistical and optimization perspectives workshop "subspace, latent structure and feature selection"* (pp. 34–51). Springer.

Rozitis, B., Ryan, A. J., Emery, J. P., Christensen, P. R., Hamilton, V. E., Simon, A. A., et al. (2020). Asteroid (101955) Bennu's weak boulders and thermally anomalous equator. *Science Advances*, *6*(41), eabc3699. https://doi.org/10.1126/sciadv.abc3699

Rubin, A. E., Trigo-Rodríguez, J. M., Huber, H., & Wasson, J. T. (2007). Progressive aqueous alteration of CM carbonaceous chondrites. *Geochimica et Cosmochimica Acta*, *71*(9), 2361–2382. https://doi.org/10.1016/j.gca.2007.02.008

Ruff, S. W., Christensen, P. R., Barbera, P. W., & Anderson, D. L. (1997). Quantitative thermal emission spectroscopy of minerals: A laboratory technique for measurement and calibration. *Journal of Geophysical Research*, *102*(B7), 14899–14913. https://doi.org/10.1029/97jb00593

Ryan, A., Emery, J. P., Rozitis, B., Christensen, P. R., Hamilton, V. E., Delbo, M., & Muñoz, D. P. (2019). Physical interpretation of Bennu's thermal inertia. *EPSC-DPS Joint Meeting 13*, EPSC-DPS2019-324-1.

Salisbury, J. W., & Wald, A. (1992). The role of volume scattering in reducing spectral contrast of reststrahlen bands in spectra of powdered minerals. *Icarus*, *96*(1), 121–128. https://doi.org/10.1016/0019-1035(92)90009-v

Shirley, K. A. (2018). *The effects of particle size and albedo on mid-infrared spectroscopy for the Moon* (Doctoral dissertation). State University of New York at Stony Brook.

Shirley, K. A., & Glotch, T. D. (2019). Particle size effects on mid-infrared spectra of lunar analog minerals in a simulated lunar environment. *Journal of Geophysical Research: Planets*, *124*(4), 970–988. https://doi.org/10.1029/2018je005533

Simon, A. A., Kaplan, H. H., Cloutis, E., Hamilton, V. E., Lantz, C., Reuter, D. C., et al. (2020). Weak spectral features on (101995) Bennu from the OSIRIS-REx Visible and InfraRed Spectrometer. *Astronomy & Astrophysics*, *644*, A148. https://doi.org/10.1051/0004-6361/202039688

Simon, A. A., Kaplan, H. H., Hamilton, V. E., Lauretta, D. S., Campins, H., Emery, J. P., et al. (2020). Widespread carbon-bearing materials on near-Earth asteroid (101955) Bennu. *Science*, *370*(6517). https://doi.org/10.1126/science.abc3522

Simon, A. A., Reuter, D. C., & Lauretta, D. S. (2021). Derivation of the final OSIRIS-REx OVIRS in-flight radiometric calibration. *Journal of Astronomical Telescopes, Instruments, and Systems*, *7*(2), 020501. https://doi.org/10.1117/1.jatis.7.2.020501

Takir, D., Emery, J. P., Mcsween, H. Y., Jr, Hibbitts, C. A., Clark, R. N., Pearson, N., & Wang, A. (2013). Nature and degree of aqueous alteration in CM and CI carbonaceous chondrites. *Meteoritics & Planetary Sciences*, *48*(9), 1618–1637. https://doi.org/10.1111/maps.12171

Thomas, I. R., Greenhagen, B. T., Bowles, N. E., Donaldson Hanna, K. L., Temple, J., & Calcutt, S. B. (2012). A new experimental setup for making thermal emission measurements in a simulated lunar environment. *Review of Scientific Instruments*, *83*(12), 124502. https://doi.org/10.1063/1.4769084

Thomson, J. L., & Salisbury, J. W. (1993). The mid-infrared reflectance of mineral mixtures (7–14 μm). *Remote Sensing of Environment*, *45*(1), 1–13. https://doi.org/10.1016/0034-4257(93)90077-b

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tomeoka, K., & Buseck, P. R. (1985). Indicators of aqueous alteration in CM carbonaceous chondrites: Microtextures of a layered mineral containing Fe, S, O and Ni. *Geochimica et Cosmochimica Acta*, *49*(10), 2149–2163. https://doi.org/10.1016/0016-7037(85)90073-0

Tomeoka, K., McSween, H. Y., Jr., & Buseck, P. R. (1989). Mineralogical alteration of CM carbonaceous chondrites: A review. *Proceedings of the NIPR Symposium on Antarctic Meteorites*, *2*, 221–234.

Trang, D., Thompson, M. S., Clark, B. E., Kaplan, H. H., Zou, X. D., Li, J. Y., et al. (2021). The role of hydrated minerals and space weathering products in the bluing of carbonaceous asteroids. *The Planetary Science Journal*, *2*(2), 68. https://doi.org/10.3847/psj/abe76f

Turnock, A. C., Lindsley, D. H., & Grover, J. E. (1973). Synthesis and unit cell parameters of Ca-Mg-Fe pyroxenes. *American Mineralogist: Journal of Earth and Planetary Materials*, *58*(1–2), 50–59.

Wold, S., Martens, H., & Wold, H. (1983). The multivariate calibration problem in chemistry solved by the PLS method. *Lecture Notes in Mathematics*, *973*, 286–293. https://doi.org/10.1007/bfb0062108

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 109–130. https://doi.org/10.1016/s0169-7439(01)00155-1

Yu, L., & Ji, J. (2015). Surface thermophysical properties determination of OSIRIS-REx target asteroid (101955) Bennu. *Monthly Notices of the Royal Astronomical Society*, *452*(1), 368–375. https://doi.org/10.1093/mnras/stv1270

Zolensky, M., Barrett, R., & Browning, L. (1993). Mineralogy and composition of matrix and chondrule rims in carbonaceous chondrites. *Geochimica et Cosmochimica Acta*, *57*(13), 3123–3148. https://doi.org/10.1016/0016-7037(93)90298-b

Zou, X. D., Li, J. Y., Clark, B. E., Golish, D. R., Ferrone, S., Simon, A. A., et al. (2021). Photometry of asteroid (101955) Bennu with OVIRS on OSIRIS-REx. *Icarus*, *358*, 114183. https://doi.org/10.1016/j.icarus.2020.114183