Optimizing Calibration Transfer for Laser-Induced Breakdown Spectroscopy (LIBS): A Multi-Instrument Investigation Across Atmospheric and Laser Power Variations

Advancing Post-Processing Techniques Across Experimental Setups

Authored by: Ines Belkhodja

Thesis Advisor: M. Darby Dyar

A Thesis Submitted to the Department of Physics and Astronomy, Mount Holyoke College, in Partial Fulfillment of the Requirements for the Degree of Bachelors of Arts with Honors

May 2025

Acknowledgements

I wish to express my deepest gratitude to Dr. Darby Dyar for her extraordinary mentorship and for welcoming an inexperienced freshman into her research group. Her guidance not only helped master my Python programming by starting with archiving SuperLIBS data to the Planetary Data System (PDS), but also revealed the profound connections between my research interests and planetary science. Your help and guidance has been transformative for my intellectual and professional development throughout these four years.

Special huge thanks go to Dr. Kate Lepore and Cai Ytsma for their patient guidance through every stage of this research process since the beginning. By writing a paper together to my thesis, their insightful comments and advice on methodology and results were always invaluable and helped me grow as a researcher.

To my committee members, I extend my deepest appreciation for their transformative impact on my academic journey. Dr. Thomas Burbine's introductory astronomy course, which I fought to enroll in as an eager freshman, ignited my passion for planetary science and set me on this research path. Dr. Laura Lyman's Intermediate Statistics course proved equally foundational, providing the framework that became crucial for the statistical analysis in this work. I am profoundly grateful for their generous support, insightful questions, and expert guidance throughout this process - from coursework to to final evaluation. Their constructive feedback has shaped not only this thesis but my development as a researcher!

I also owe immense thanks to my closest friends—Lakshyaa Sreekumar and Manasvi Agarwal—who have witnessed my growth since freshman year and kept me grounded (and sane) throughout this process. To Nick Vasconcellos, my biggest cheerleader, thank you for all the snacks and food during the late nights working together. To Abi Asteraye and Latika Joshi for their unwavering encouragement throughout the research process. And to my parents, whose own academic overcoming and achievements inspired me to push beyond my limits and take on this challenge.

Finally, I am deeply grateful to the Mount Holyoke Astronomy & Physics and Computer

Science Departments for accepting a determined, if occasionally confused, student. Thank you to Dr. Katie Lester for your guidance through these few years and to Dr. Heather Pon-Barry for her help in machine learning—your insights in this field and from your class were invaluable.

Abstract

Laser-Induced Breakdown Spectroscopy (LIBS) analyzes the elemental composition by measuring plasma emissions from laser-ablated samples. This study develops a universal calibration model for LIBS applications across Earth (760 Torr), Mars (7 Torr CO_2), and vacuum (<1 Torr) conditions with laser energies spanning 3–9 mJ. Using 2,157 geological standards analyzed across four LIBS systems (SuperLIBS, ChemLIBS, and pLIBS Z300/Z903), we evaluate Partial Least Squares (PLS) regression performance through four validation scenarios, incorporating data from current Mars rover LIBS instruments (ChemCam and SuperCam) and their full and onboard calibration target sets. To account for atmospheric effects on plasma formation, we implement Piecewise Direct Standardization (PDS), a calibration transfer technique that mathematically adjusts spectral responses between different environmental conditions. The comprehensive calibration set developed across different LIBS setups provides crucial insights into prediction accuracy under Mars conditions while establishing a quantitative framework of LIBS applicable to diverse planetary environments, from Venus to airless bodies.

Keywords – Laser-Induced Breakdown Spectroscopy, LIBS, Multivariate Analysis, Mars, ChemCam, SuperCam, PyHAT, Machine Learning, Partial Least Squares, PLS, Calibration Transfer, Piecewise Direct Standardization - Partial Least Squares, PDS-PLS

Contents

List of Figures

List of Tables

1 Introduction

Laser-induced breakdown spectroscopy (LIBS) is a powerful application of atomic emission spectroscopy that provides quantitative, rapid, multi-element analysis of minerals and rocks (Fabre, C. et al. 2014). It operates by a laser pulse focused onto a sample to create a small plasma from which the optical emissions are recorded spectroscopically under three key conditions (Laura Bush 2019). First, the laser properties, such as energy, duration, wavelength, and beam quality, influence plasma generation. Second, the characteristics of the sample, including thermal conductivity and melting and vaporization temperatures, play a crucial role. Third, the ambient atmosphere in which the plasma forms also affects its characteristics. The plasma, composed primarily of neutral atoms and ions in the first few ionization states of the sample's elements, emits light that is diffracted by a spectrometer and detected as an electrical signal (Cousin, A. et al. 2011). LIBS spectra analysis relies on quantized valence electron transitions that occur in the ultraviolet (UV), visible (VIS), and visible to near-infrared (VNIR) regions of the electromagnetic spectrum (Dyar, M.D. et al. 2012). These emission peaks depend on the abundance of specific elements in the sample, which forms the basis for both qualitative and quantitative chemical analysis (Laura Bush 2019). Hence, by analyzing the plasma emission spectra, one can determine the sample's composition and concentration information.

The first research on LIBS emerged in 1963. Since then, more than 13,700 papers have explored LIBS, covering its fundamentals and diverse applications (Laura Bush 2019). Over the past three decades, significant advances have enhanced LIBS performance and highlighted its advantages over other spectroscopic techniques. One major benefit of LIBS is its ability to perform in-situ or remote analyses without requiring extensive sample preparation (Fu, H. et al. 2017). Another demonstrated advantage is its effectiveness in multivariate modeling, which enables the analysis of any element in the periodic table. This versatility makes LIBS highly suitable for applications across fields like geology, materials science, biomedical research, environmental monitoring, and planetary exploration (Ytsma, C. & Dyar, M.D. 2019). For example, in space exploration, ChemCam and SuperCam aboard NASA's *Curiosity* and

Perseverance rovers, respectively, use LIBS to analyze Martian rock samples and provide valuable insights into the planet's geology and potential habitability in different locations such as in the Gale Crater and the Jezero Crater (Rammelkamp, K. et al. 2021; Cousin, A. et al. 2025).

Despite its versatility in applications like metals, alloys, and liquid slag analysis, LIBS faces significant challenges that can compromise its accuracy and reliability (Gottlieb, C. et al. 2017). One key issue is sample heterogeneity, as LIBS analyzes a small, specific portion of a solid sample. This sensitivity to uneven sample composition can lead to sensitivity in results, which decreases the accuracy and reproducibility performances compared to other spectroscopic techniques. Furthermore, LIBS is influenced by chemical matrix effects, where the signal intensity depends not only on the concentration of elements but also on the surrounding matrix (Clegg, S.M. et al. 2007). The sample's elemental and molecular composition influences the laser-to-sample coupling efficiency, which affects plasma emission intensity (Tucker, J.M. et al. 2010). Atmospheric pressure and composition also impact plasma emission, adding to the challenge of precise quantification (Clegg, S.M. et al. 2007). These factors hinder the accurate determination of elemental composition from spectral intensities. Another key issue is that understanding how to properly and effectively calibrate LIBS data. Calibration remains essential to convert raw spectral data into accurate quantitative results by accounting for instrument variability, matrix effects, and environmental conditions (Sun, C. et al. 2019). The key to interpreting Mars or any other types of LIBS data is calibrations that relate laboratory standards to unknowns examined in other settings and enable predictions of chemical composition. Traditional calibration methods, such as the standard calibration curve (SCC), often fall short of providing robust spectroscopic analytical data performance due to their reliance on univariate models (Bilge, G. et al. 2021).

To overcome these limitations, more advanced techniques have been increasingly integrated into LIBS analysis to enable reliable quantitative chemical analyses from spectral data. Among these, multivariate analysis and calibration-free laser-induced breakdown spectroscopy have shown meaningful potential (Anderson, R. et al. 2012; Ciucci, A. et al. 1999). The multivariate techniques, such as principal component analysis (PCA), partial least squares

regression (PLS), least absolute shrinkage and selection operator (LASSO) and artificial neural networks (ANN), have been particularly examined to be effective (Anderson, R. et al. 2012; Boucher, T. et al. 2015a; Ytsma, C. & Dyar, M.D 2018). These multivariate approaches fall under the broader domain of machine learning, a branch of artificial intelligence (AI) that enables computers to learn from data and make predictions (IBM 2023). Machine learning enhances the development of spectroscopic data treatment models by providing powerful and adaptive algorithms to better understand the predictive performances. Another domain where machine learning enhances LIBS is in improving the versatility of instruments. However, not all researchers have the means to test their data under another instrument's conditions required for robust model development. A powerful solution to this challenge is calibration transfer (CT), which enables models trained on one instrument or set of conditions to be effectively applied to others by using mathematical algorithms (Boucher, T. et al. 2015a).

Despite these recent advancements with calibration transfer, previous studies have underscored the need for larger, well-characterized datasets to improve model generalizability. For instance, work by Boucher, T. et al. (2015a) demonstrated that models spanning diverse laser power densities better accommodate pulse-to-pulse signal fluctuations, yet such datasets remain scarce. The extent to which conclusions can be generalized still depends on data size and composition, but this remains untested due to limited availability of comprehensive calibration suites. Another paper by Wiens, R.C. et al. (2012) suggest that "increasing the number of standards used in the training set will significantly increase the accuracy as well". Currently, the most robust datasets are restricted to specific sample types (such as experiments with only igneous/meta-igneous rocks), which mitigates matrix effects but limits scope (Boucher, T. et al. 2015a).

Addressing this gap, this thesis investigates a dataset of 2,157 geologically diverse samples by developing a supermodel trained on four subsets collected from three instruments under varying conditions (Earth, Mars, and vacuum/space) and laser powers (2–9 mJ). The primary goal is to assess whether a single, unified model can reliably predict compositions under novel conditions; for instance, looking at the performance metric when applied to data from an unfamiliar instrument or environment. The secondary goal is to examine multiple

scenarios with and without calibration transfer, and compare these model performance on unseen datasets through five statistical metrics. This work not only quantifies improvements in predictive accuracy but also offers practical insights for optimizing LIBS workflows in planetary science and other related fields. Ultimately, it lays another foundation for future studies to clarify the utility of these methods as even larger, more complex LIBS datasets emerge.

2 Background

2.1 LIBS History

Since the discovery of optical lasers in the early 1960s, spectroscopy has undergone significant advances, leading to powerful techniques such as laser-induced breakdown spectroscopy (LIBS). This technique was first used by Runge, E.F. et al. (1963), who produced atomic emission spectra through simultaneous vaporization and excitation of materials using a giant-pulse ruby laser. While this discovery laid the foundation for LIBS, its widespread adoption was delayed until the 1980s when improved CCD detectors and spectrographs enabled more precise measurements (Tucker, J.M. et al. 2010).

The technique faces particular challenges in geological applications where established microanalytical methods dominated. However, a turning point came in the early 2010s when NASA's need for remote planetary surface exploration propelled LIBS into prominence as a vital tool for stand-off geochemical analysis. This space exploration demand drove innovations that expanded LIBS applications to Venus (Clegg, S.M. et al. 2014), lunar (Lasue, J. et al. 2012), icy satellites such as Europa (Pavlov, S.G. et al. 2011), asteroids (Pavlov, S.G. et al. 2012), and Mars missions (Maurice, S. et al. 2016), while simultaneously fostering the development of portable field instruments (Kim, Y. et al. 2022).

Recent advances have further enhanced LIBS capabilities through integration with complementary techniques. For instance, fusion strategies combining LIBS with Raman spectroscopy, X-ray fluorescence (XRF), and hyperspectral imaging now provide more comprehensive material characterization (Ferreira, D.S. et al. 2024). The technique's simplicity and minimal sample preparation requirements have led to hundreds of publications documenting its versatility across diverse fields including space exploration, archaeological science, environmental monitoring, and material science (Chen, T. et al. 2020).

2.2 Principles of LIBS Operation

The fundamental principle of LIBS involves quantized valence-electron transitions. When a high-energy laser pulse is focused on the material surface, the strong heating induces surface ablation. The removed material interacts with the incident beam to generate a small, high-temperature plasma (typically 1 mm in diameter at several thousand degrees), where excited atoms and ions emit characteristic emission lines that serve as elemental fingerprints (Andor 2025; L'Hermite, D. et al. 2016).

The emitted photons are captured as spectral lines by a spectrometer through fiber optics or lenses, with each line corresponding to a unique elemental wavelength (L'Hermite, D. et al. 2016). Plasma formation depends critically on laser parameters including wavelength, pulse duration, and fluence (Sirven, J.B. et al. 2004). The plasma evolution follows distinct phases: initial breakdown, continuum emission, and characteristic line emission. The resulting spectra reflect both the plasma composition and the chemical composition of the ablated material (Steven Buckley 2014), where all the chemical elements of the periodic table can be detected by LIBS with a variable sensitivity (L'Hermite, D. et al. 2016).

A LIBS system consists of (i) a laser (such as a solid-state Nd:YAG laser), (ii) mirrors and lenses (which direct and focus the laser pulse on the sample), (iii) a sample holder, (iv) a plasma radiation collection system (and/or an off-axis parabolic mirror), (v) a spectrometer, a (vi) detector, as well as (vii) a delay generator/system that synchronizes plasma generation and spectra collection (Guedes, W.N. et al. 2023). These components are shown in Figure 2.1.

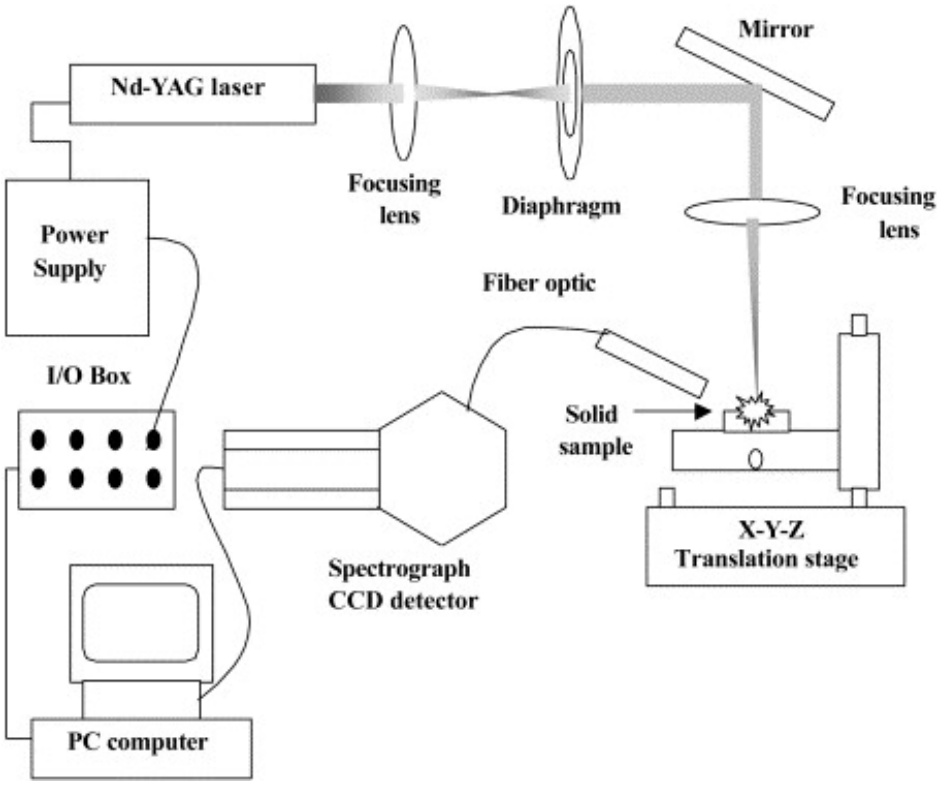

Figure 2.1: Detailed Example of LIBS Configuration

A LIBS spectrum may be simple or complex. All elements have at least one, and typically many, emission line(s) in the spectral region between 200 and 900 nm, so that the LIBS emission generated by a single laser shot will reveal the elements present in a sample, shown in an example emission spectrum plot in Figure 2.2 (Harmon, R.S. & Senesi, G.S 2021).

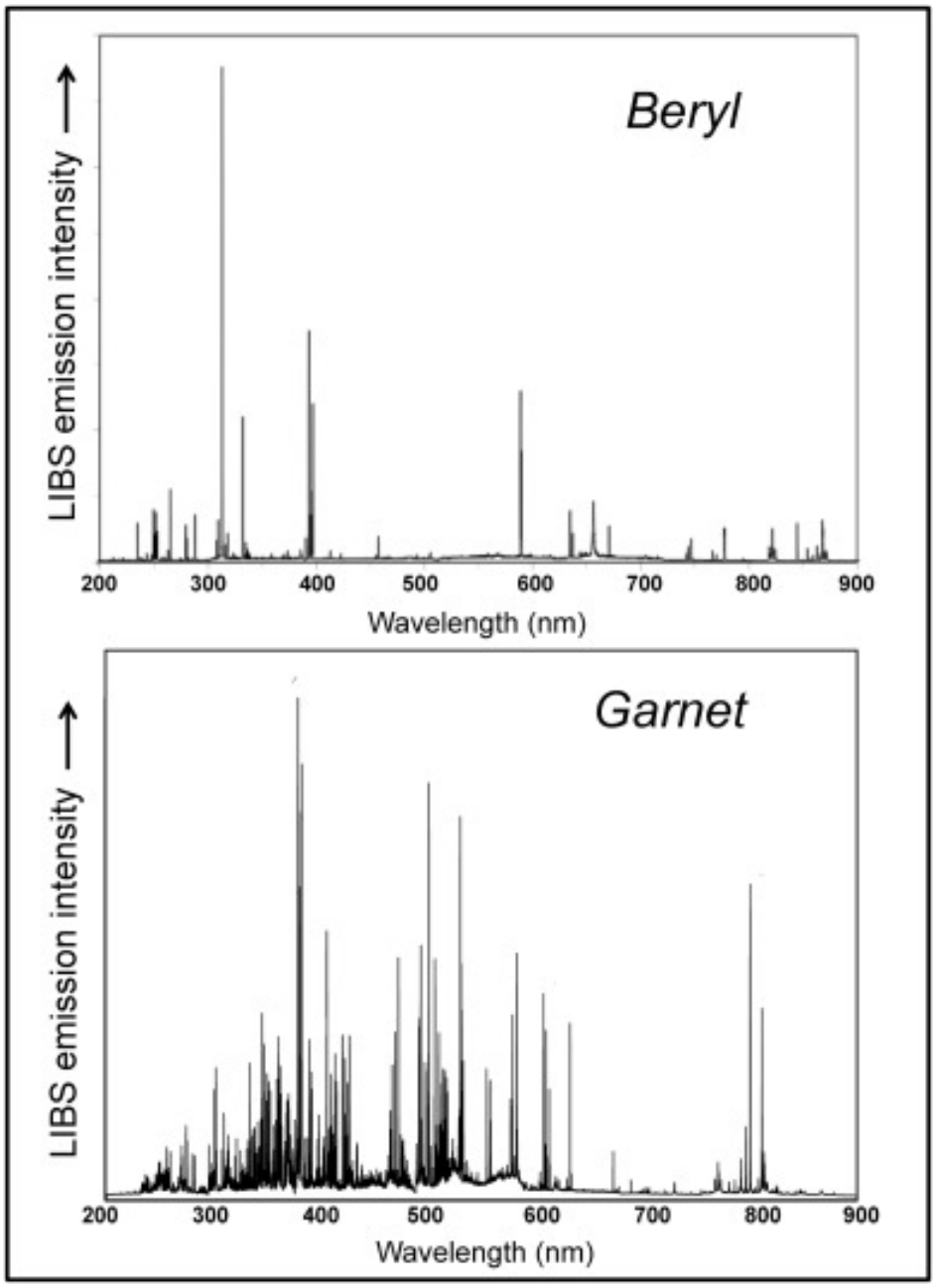

Figure 2.2: Example of a simple (above) and complex (below) LIBS emission spectrum over the wavelength range from 200 to 900 nm acquired with commercial laboratory system for a beryl and a garnet (Harmon, R.S. & Senesi, G.S 2021)

2.3 LIBS Instrumentation and Analytical Challenges

In practical LIBS analysis, qualitative analysis means material identification and/or classification, while quantitative analysis refers to the determination of chemical component

concentration (Li, L.N. et al. 2021). This technique's appeal for field applications stems from its ability to analyze nearly all elements rapidly and remotely. However, LIBS faces several analytical challenges that complicate quantitative elemental analysis.

The most significant limitation is matrix effects, where a sample's physical and chemical properties influence emission line intensities independent of elemental concentration (Grünberger, S. et al. 2023). These effects arise from the chaotic nature of LIBS plasmas, where high-temperature interactions between neutral/ionized species, collisional dynamics, and self-absorption perturb emission line intensities (Matsumura, T. et al. 2024). For example, 5% iron in basalt produces 30% stronger emission than the same concentration in sandstone due to differences in plasma formation dynamics (Lepore, K. et al. 2017). Atmospheric conditions further compound these effects, as local pressure and composition alter plasma energetics, though recent studies show atmospheric differences may not dominate error (Chen, T. et al. 2020).

The transient nature of laser-induced plasmas exacerbates these issues. During analysis, plasmas evolve rapidly in temperature, density, and opacity (Cristoforetti, G. et al. 2013), while laser-sample coupling efficiency varies with surface morphology. Unconsolidated materials like regolith present unique challenges as laser ablation creates pits that confine the plasma differently than flat surfaces (Wiens, R.C. et al. 2013). Field conditions introduce additional variability—changing distances and incidence angles alter energy deposition, while typical 5 mJ laser pulses show ±7% intensity variation for calcium lines between consecutive shots (Steven Buckley 2014).

Spectral interference constitutes another major challenge, where overlapping emission lines from different elements compromise detection limits in complex mineral assemblages (Nancy McMillan 2018). This problem is particularly acute for transition metals with rich emission spectra or when analyzing minor/trace constituents. Recent advances address these limitations through specialized instruments like ChemLIBS that simulate planetary rover conditions, comprehensive spectral databases such as the USGS Spectral Library, and advanced preprocessing techniques including baseline correction and normalization (Castro, J.P. et al. 2020). These approaches collectively improve LIBS's quantitative capabilities while

maintaining its core advantages of rapid, in-situ analysis.

2.4 Analytical Techniques in LIBS

2.4.1 From Univariate to Multivariate Approaches

The analysis of LIBS data increasingly relies on chemometric methods to address its inherent complexities, particularly chemical matrix effects (Herreyre, N. et al. 2023). Chemometrics, the application of mathematical, statistical, and logical techniques to extract maximum chemical information from analytical data, has become indispensable for LIBS applications (Harmon, R.S. & Senesi, G.S 2021). This synergy stems from LIBS's unique capacity to generate high-dimensional spectral datasets, where each measurement can be analyzed on different different spectral regions channels across the UV-VIS-NIR range (Fernández-Manteca, M.G. et al. 2023). Such multivariate data contains intricate relationships between emission lines that conventional univariate analysis cannot exploit. Modern chemometric approaches overcome these limitations through several mechanisms: noise reduction via signal averaging across correlated wavelengths, compensation for spectral interferences through multivariate regression, and automated outlier detection using pattern recognition algorithms (Harmon, R.S. & Senesi, G.S 2021). The field has developed an extensive toolkit for LIBS applications, including principal component analysis (PCA) for dimensionality reduction (Sirven, J.B. et al. 2006), partial least squares (PLS) regression for quantitative calibration (Sirven, J.B. et al. 2006), and machine learning techniques like artificial neural networks (ANNs) (Sirven, J.B. et al. 2006; Motto-Ros, V. et al. 2007) and support vector machines (SVMs) for complex nonlinear modeling (Zhu, X. et al. 2014). These methods collectively enable more accurate compositional analysis by simultaneously considering all spectral information rather than isolated emission lines, while also providing capabilities for material classification and discrimination that transcend traditional analytical approaches.

LIBS analysis has evolved from traditional linear univariate methods to sophisticated multivariate techniques to address the technique's inherent challenges. Univariate analysis, while straightforward, proves inadequate for complex geological samples due to its reliance

on single emission peaks and vulnerability to matrix effects and spectral peak overlaps (Dyar, M.D. et al. 2016). Modern implementations now predominantly employ multivariate analysis (MVA) that simultaneously examines multiple spectral variables, either through selected peaks or full spectrum utilization (Lepore, K. et al. 2017; Ytsma, C. & Dyar, M.D. 2019). These mathematical approaches establish critical relationships between spectral features and known elemental concentrations, compensating for LIBS's inherent analytical challenges (Lepore, K. et al. 2017). One of the first studies done by Sirven, J.B. et al. (2007) aplied three chemometric methods to LIBS spectra acquired remotely in the laboratory in simulated Martian conditions on a series of natural rock samples. Now, the ChemCam instrument aboard NASA's *Curiosity* rover demonstrates the vital role of calibration, where it has a comprehensive set of LIBS spectra from well-characterized geological standards both from a laboratory setting and currently onboard the rover (Clegg, S.M. et al. 2017). Furthermore, this approach of MVA offers significant advantages over traditional univariate methods, as demonstrated by Dyar, M.D. et al. (2016), who found univariate predictions based on single emission lines consistently produced the least accurate results regardless of wavelength selection.

Principal Component Analysis (PCA) and Partial Least Squares Regression (PLS) represent two foundational multivariate analysis techniques that effectively mitigate matrix effects in laser-induced breakdown spectroscopy (LIBS), particularly for major element analysis (Clegg, S.M. et al. 2009). PLS has become the predominant method for geological LIBS applications, including the analysis of Martian surface materials (Dyar, M.D. et al. 2016; Luque-García, J.L. et al. 2002). Unlike traditional regression methods such as ordinary least squares (OLS), PLS is particularly advantageous by dealing with multicollinearity or data with many correlated predictor variables (X) and a limited number of observations, and high dimensional data.

2.4.1A Partial Least Squares

Partial Least Squares (PLS) regression is a robust multivariate technique that generalizes and combines features from principal component analysis (PCA) and multiple linear regression.

The effectiveness of PLS derives from its linear model formulation (Y = Xb + E), which exists in two distinct variants: PLS-1, designed for single-response prediction, and PLS-2, which extends the framework to multi-response analysis by exploiting inherent correlations between output variables (Jan Gerretzen 2016).

The method operates by iteratively extracting latent components that simultaneously maximize the explained variance in both the predictor space (X) and the covariance between X and the response variable(s) (Y). Unlike PCA, which identifies components based solely on variance within X, PLS explicitly incorporates the relationship with Y when constructing these latent dimensions (Dante M. Pirouz 2010). This characteristic makes PLS particularly powerful for analyzing spectroscopic data, where it establishes robust correlations between variations in elemental composition (the response variable Y) and spectral features across wavelength channels (the predictor variables X) (Dyar, M.D. et al. 2011).

The algorithm proceeds through three key phases: (1) decomposition of X into orthogonal latent components, (2) regression of Y on these components, and (3) reconstruction of the relationship between original X and Y variables. During the decomposition phase of these latent variables, PLS employs either the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm or straightforward implementation of a statistically inspired modification of the PLS method (SIMPLS) (Aylin Alin 2009). Each component is computed as a weighted linear combination of the original variables, with weights determined by their contribution to both X variance and Y prediction. The predictive performance of PLS models can be optimized through careful selection of the number of latent components, typically using cross-validation techniques to avoid overfitting, which will be used in this study. The resulting models not only provide accurate concentration predictions but also offer interpretability through examination of component weights and regression coefficients, which can reveal important spectral features associated with elemental composition.

A key advantage of PLS is that it requires no prior information about emission line locations or wavelength calibration (Dyar, M.D. et al. 2011). The ChemCam team adopted the PLS method to train hundreds of standards, which are predictions compared to their quantitative elemental compositions (Wiens, R.C. et al. 2012; Liu, S. et al. 2025). However, the method

faces limitations for trace element analysis effects, where minor elements can be indirectly predicted through their systematic relationships with major elements that share similar geochemical properties and behavior in mineral structures. This fundamental challenge can only be addressed through artificially-doped standards that disrupt natural geochemical substitution patterns (Dyar, M.D. et al. 2016). The chemical analysis with MVA in LIBS follows a two-phase process that is similar to how common machine learning algorithms work: calibration (training) and prediction (testing). During calibration, models establish relationships between spectral features (X) and known concentrations (Y) using representative training datasets. The prediction phase then applies these relationships to new, unknown samples. This framework underpins all quantitative LIBS analysis, from basic univariate calibrations to advanced machine learning approaches.

2.4.2 Calibration Transfer

A persistent challenge in LIBS is the variability introduced when measurements are taken under different conditions—whether across multiple instruments, environmental settings, or temporal variations in the same device (Boucher, T. et al. 2015b). These discrepancies can distort spectral data, leading to inconsistencies that complicate comparisons and undermine the reliability of analytical results.

Calibration transfer (CT) is another form of chemometrics and addresses this challenge by establishing mathematical or computational mappings between datasets (Anderson, R. et al. 2022). The general setup and input for CT techniques is spectra of a set of samples, such as rocks or powders that are recorded on two different instruments or on the same instrument under two varying conditions. This pioneering work came from University of Massachusetts graduate students such as Thomas Boucher (Boucher, T. et al. 2015b). These transfer techniques address spectral discrepancies between instruments, where fast advancements have been achieved. Machine learning and how to encompass model prediction accuracy enhances these approaches through methods such as direct standardization (DS) and piecewise direct standardization (PDS), which align calibration models across instruments (Jerome Workman 2017). Particularly effective is the hybrid Piecewise Direct Standardization with Partial Least

Squares (PDS-PLS) method, which applies PLS regression within spectral segments to create smoother alignments while avoiding discontinuities at window boundaries (Lepore, K. et al. 2024). This combination of multivariate analysis and machine learning represents the current state-of-the-art in addressing both matrix effects and instrumental variations in LIBS spectroscopy, where this study will further dive into how we can apply them to pressing questions.

Recent work by Anderson, R. et al. (2022) has demonstrated the potential of calibration transfer techniques to improve quantitative analysis of SuperCam calibration standards. Their study showed that appropriately transformed laboratory data become spectrally closer to Mars measurements, as evidenced by reduced distances between principal component clusters of Mars and transferred laboratory spectra. However, as noted in their conclusions, significant work remains to fully characterize the effects of these transfer methods, a key gap this study aims to address.

The development of robust calibration transfer protocols is particularly crucial given the practical constraints of LIBS analysis. Building comprehensive, chemically diverse calibration datasets requires substantial analytical resources, typically hundreds to thousands of carefully characterized reference samples measured under controlled conditions. Most research teams lack the instrumentation, time, and funding to develop such calibration sets for each new application. Effective transfer methods would allow researchers to leverage existing high-quality databases while minimizing new data collection. By improving the accuracy and reliability of calibration transfer, it can both enhance prediction performance and make LIBS quantification more accessible to a broader range of scientific studies, particularly for planetary missions and field applications where traditional calibration approaches are impractical.

Furthermore, based on Lepore, K. et al. (2024), this study critically evaluates the performance of LIBS under Earth, Mars, and vacuum atmospheric conditions, providing key insights into how pressure and composition influence spectral accuracy. Under Earth-like pressures (760 Torr), plasma self-absorption broadens spectral peaks, reducing resolution, whereas Mars-like conditions (7 Torr CO_2) produce sharper, more intense emissions due to reduced

collisional effects (Lepore, K. et al. 2024). In contrast, vacuum environments (100–300 mTorr) yield weaker signals due to rapid plasma expansion but still permit reliable quantification. Normalized spectra from all three atmospheric conditions from Lepore, K. et al. (2024) is shown in the Figure 2.3. A significant finding is that combining training data from all three atmospheres into a single model achieves near-optimal accuracy (Lepore, K. et al. 2024). The study also demonstrates that calibration transfer techniques are highly effective when training and test spectra are collected under similar but non-identical conditions, which means that both environments can produce measurable plasma emissions, just with different peak shapes and intensities. Calibration transfer results show significant improvements in prediction accuracy; for example, SiO_2 errors were reduced from 21.0% to 6.14% when Earth data were algorithmically adjusted to Mars conditions (Lepore, K. et al. 2024). These findings establish a critical benchmark for this thesis, which builds upon the demonstrated success of combining multi-atmospheric LIBS data to create a comprehensive spectral library and the inner workings of calibration transfer.

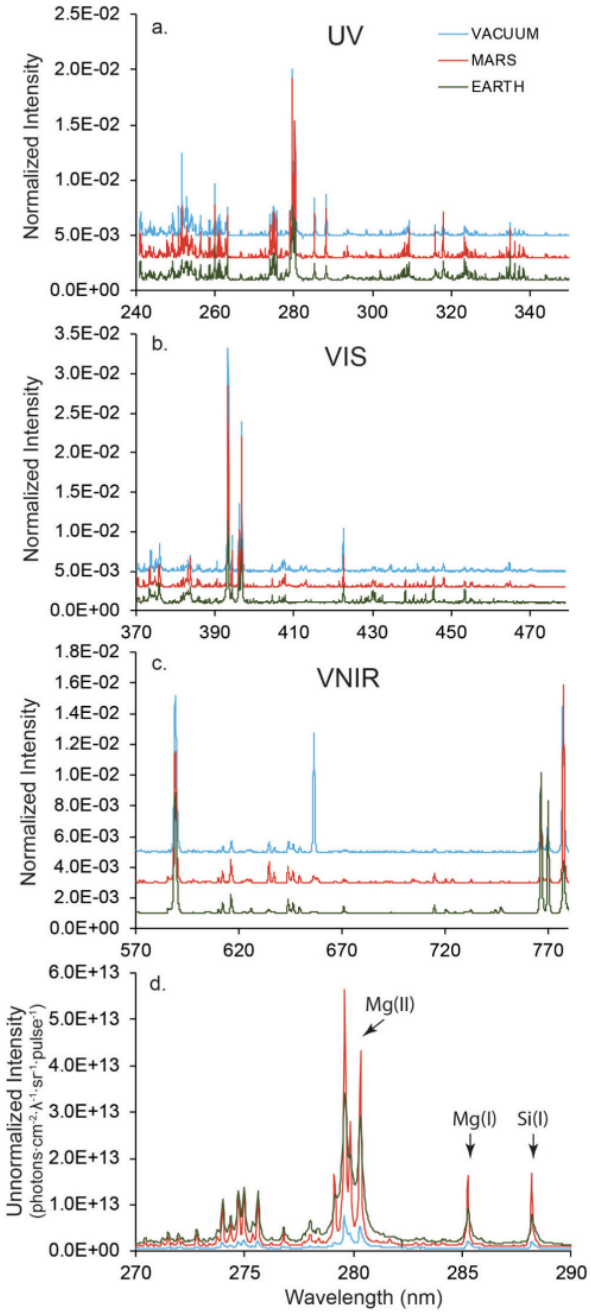

Figure 2.3: Normalized (a, b, and c) and unnormalized (d) spectra of sample GBW07105 measured under Earth, Mars, and vacuum atmospheric conditions with a laser pulse of 2.5 mJ. Earth, Mars, and vacuum spectra in panels a-c are offset for clarity on the intensity axis by 0.001, 0.003, and 0.005, respectively (Lepore, K. et al. 2024)

3 Data

3.1 Millennium Set Standards

There has been increasing demand for a high-quality spectral LIBS-derived database for use in geochemical studies on Earth and Mars and calibration support, as the community is still figuring it out (Lepore, K. et al. 2024). It takes a lot of time and effort to acquire and characterize such a large database under many different conditions and geological diversity of reference targets. As a result, the Mineral Spectroscopy Laboratory at Mount Holyoke College (MHC) over these past two decades has compiled a geochemically diverse LIBS spectral database of rock powder standards using four instruments under varying laser energies and atmospheric conditions. This data set is also in support of the Mars Science Laboratory mission for the ChemCam instrument (Boucher, T. et al. 2015b).

The sample suite at MHC includes roughly 3,000 rock powder standards with a grain size of $\ll 10\,\mu$m pressed into 1.6 cm diameter pellets for LIBS analysis. Samples represent a diverse collection of igneous (70%), sedimentary (25%), and metamorphic (5%) rocks and collected from several sources with support from the NASA Mars Fundamental Research program (Dyar, M.D. et al. 2024). All other un-doped rock standards that are used in this study use either literature values or analyses from the laboratory of Dr. J. Michael Rhodes at the University of Massachusetts using his standard protocols with X-ray Fluorescence (XRF), loss on ignition (LOI) and ferrous iron content (Dyar, M.D., Ytsma, C.R., Lepore, K. 2019; Rhodes, J.M. & Vollinger, M.J. 2004). Other available geological standards were purchased from the U.S. Geological Survey, the Centre de Recherches Pétrographiques et Géochimiques (CRPG) and Brammer Standard Company, Inc (Dyar, M.D & Ytsma, C. 2021).

3.1.1 Major Elements

The decay of atoms from an excited state back to their ground state results in the emission of photons. Spectral analysis of the emitted light makes it possible to identify elements present in the sample from their characteristic emission lines: major elements, minor and

trace elements (Wiens, R.C. et al. 2012). This study focuses only on the major elements.

The major elements detected by LIBS include Na, Mg, Al, Si, K, Ca, Ti, and Fe. These elements are primarily found in geological materials as oxides (e.g., SiO_2, Fe_2O_3) due to their affinity for oxygen (Maurice, S. et al. 2016). In LIBS analysis, they are typically quantified in weight percent (wt%) of their oxide forms, based on their most common oxidation states in rocks and minerals.

Sodium (Na) is commonly present as Na_2O and shows prominent LIBS emission lines at 589.0 nm and 589.6 nm (D-lines) (Body, D. & Chadwick, B.L. 2001). It is geologically significant in feldspars and volcanic glasses. Magnesium (Mg), quantified as MgO, displays strong LIBS lines at 279.5 nm, 280.3 nm, 285.2 nm, 383.8 nm and is a dominant component in mafic minerals like olivine and pyroxene (Knight, A.K. et al. 2000; Body, D. & Chadwick, B.L. 2001). Aluminum (Al) occurs as Al_2O_3 and has key LIBS lines at 309.28 nm nm and 396.1 nm (Body, D. & Chadwick, B.L. 2001; Prasetya, O.D. et al. 2018). It is a major constituent of feldspars and clay minerals. Silicon (Si), present as SiO_2, shows primary emission lines at 288.2 nm (Si I) and 390.6 nm (Si II), and is fundamental to all silicate minerals that dominate Martian geology (Knight, A.K. et al. 2000; Body, D. & Chadwick, B.L. 2001). Potassium (K) appears as K_2O with strong emission at 766.5 nm and 769.9 nm (Body, D. & Chadwick, B.L. 2001). Its presence indicates evolved magmas and is found in K-feldspar and micas. Calcium (Ca), quantified as CaO, has intense lines at 393.4 nm, and is key to minerals like plagioclase and carbonates (Knight, A.K. et al. 2000; Body, D. & Chadwick, B.L. 2001). Titanium (Ti) occurs as TiO_2 and exhibits hundreds of LIBS lines, with the strongest at 334.9 nm, 336.1 nm, and 498.1 nm (Body, D. & Chadwick, B.L. 2001). It is important in accessory minerals like ilmenite. Iron (Fe) appears in multiple oxide states (FeO, Fe_2O_3) and shows numerous emission lines, notably at 238.2 nm, 259.9 nm, 293 nm and 404.6 nm, being abundant in basalts and pyroxenes (Body, D. & Chadwick, B.L. 2001; Idris, N. et al. 2017).

Table 3.1 demonstrates the mean compositions of these major elements from all the standards collected over the years.

Table 3.1: Compositions of Geochemical Standards (Pressed Pellets and Powders) for Major (Dyar, M.D., Ytsma, C.R., Lepore, K. 2019)

Element	Form	# Standards	Mean Concentration (wt%)
SiO_2	wt%	3035	58.5
TiO_2	wt%	3031	1.2
Al_2O_3	wt%	3035	13.5
Fe_2O_3	wt%	3022	8.0
MgO	wt%	3027	5.6
CaO	wt%	3032	6.0
Na_2O	wt%	3031	2.7
K_2O	wt%	3027	2.2

3.1.2 Outliers

Outliers in the elemental composition data were detected using the Isolation Forest algorithm, an unsupervised machine learning method that identifies anomalies by measuring how easily data points can be isolated through random feature splits (*Isolation Forests* n.d.). This approach is particularly effective for detecting samples with sparse or unusual chemical compositions, as anomalous concentrations require fewer random splits to isolate from the dataset. This method was implemented with an outlier probability threshold set at 5%, following similar protocols used in MarSCoDe data analysis (Lepore, K., Dyar, M.D., Ytsma, C.R. 2024).

The algorithm works by recursively partitioning the data along randomly selected elemental concentration axes, with outliers identified as points that separate with significantly fewer splits than normal samples. The analysis revealed that most compositional outliers occupied extreme high or low concentration ranges of major elements. In the few cases where mid-range compositions were flagged as outliers, these were retained to ensure comprehensive analysis, as they may represent legitimate but uncommon geochemical signatures (Lepore, K., Dyar, M.D., Ytsma, C.R. 2024).

3.2 Instrumentation and spectra collection

The spectra collected on four instruments in the Mineral Spectroscopy Laboratory at Mount Holyoke College, South Hadley, Massachusetts are SuperLIBS 10K and 18K, ChemLIBS, pLIBSZ300, and pLIBSZ903. It is the largest LIBS calibration dataset of known samples compositions to date (Lepore, K., Ytsma C.R., Dyar, M.D. 2024).

3.2.1 SuperLIBS

SuperLIBS enables rapid sample analysis across various plasma temperatures and atmospheric conditions, including air and vacuum environments. It is used to support SuperCam calibration, which is the LIBS instrument on the Mars 2020 *Perseverance* rover. It is a highly sensitive instrument with 2D CCD (e2v CCD42–10 back-illuminated high performance AIMO CCD sensors) detectors capable of analyzing spectra collected at a range of laser powers and resolutions and identical to those on SuperCam on *Perseverance* (Lepore, K. et al. 2018; Lepore, K. et al. 2024). SuperLIBS was built with the capacity to analyze 100 standards at up to 10 distinct laser powers and three atmospheres (Earth, vacuum (up to 100 mTorr), Mars) in a single, automated 24-hour run (Lepore, K. et al. 2018).

The system holds the target material ablated using a Quantel Ultra 100 Nd:YAG laser operating between 1 and 10 mJ at 1064 nm, with a beam diameter of 3 mm and an 8 ns pulse width. The laser beam is focused to a spot diameter of approximately 110 μm using a 300 mm focal length lens and directed at the target surface at a 13˚angle via an Nd:YAG dielectric mirror. Light emitted from the laser-induced plasma is separated into three wavelength regions—UV (240-340 nm), VIS (375-475 nm), and VIS-NIR (490-790 nm)—by a demultiplexer and transmitted via fiber optic cables to three Princeton Instruments Iso-Plane 160 spectrometers. Each spectrometer uses gratings similar to or higher than the wavelength ranges and resolutions of SuperCam: resolutions of ≈0.08 nm in the UV and VIS region, and at both ≈0.08 nm and ≈0.40 nm in the VNIR region. Low- and high-resolution spectra are referred to as 10 K and 18 K, respectively, due to the approximate number of wavelength pixels present in each spectrum (Lepore, K. et al. 2024). For the UV and VIS regions,

spectra are collected with PIXIS cameras (Princeton Instruments), while the VIS-NIR region utilizes a PI-MAX-4 camera (Princeton Instruments) with gating capabilities. This setup synchronizes spectrometers, cameras, and the laser flash lamp using trigger pulses.

As part of a larger project funded by NASA's Planetary Data Archiving, Restoration, and Tools program, samples were analyzed by Dr. Kate Lepore at MHC under three different atmospheric conditions with both ChemLIBS and SuperLIBS equipped with vacuum chambers for target analysis: ∼760 Torr Earth atmosphere, 7 Torr CO_2 to mimic Mars, and ∼100–300 mTorr vacuum. Spectra were collected on each target at three (Earth; 2.4, 3.2, and 4.0 mJ) or four (Mars and vacuum; 2.4, 4.0, 5.6, and 7.2 mJ) laser energies at five locations (Lepore, K. et al. 2024). The lens-to-sample distance (300 mm), ablation spot size (∼110 μm diameter), and pulse width (7 ns) were constant throughout data collection. These parameters and the range in laser energies result in laser fluences between 3.6 and 11 GW cm^{-2}. The area sampled by the collection optics is approximately 2 mm in diameter. This collection area was chosen in order to sample the entire plasma under Earth, Mars-like, and vacuum atmospheres. These results were published in a series of papers highlighting use of sub-models and effects of plasma temperature, line and peak binning, and atmospheric variations on LIBS accuracy (Lepore, K., Ytsma C.R., Dyar, M.D. 2022; Lepore, K., Dyar, M.D., Ytsma, C.R. 2023; Lepore, K., Dyar, M.D., Ytsma, C.R. 2024; Lepore, K. et al. 2024).

Plasma emission was collected over a 10 ms interval to ensure that the entire lifetime of a single plasma was recorded during each integration. Thirty shots were recorded at each location; up to five shots were discarded to eliminate any contamination of the target surface. Spectra were averaged over 29 (vacuum) or 25 (Mars and Earth) shots and five locations to produce a single spectrum for each reference target at each laser energy and atmosphere (Lepore, K. et al. 2024).

Low-(10 K) and high-(18 K) resolution spectra were collected in the VNIR region with two different spectrometer gratings (400 and 1800 g mm^{-1}, respectively). The low-resolution grating was chosen to collect spectra with a spectral resolution similar to ChemCam and SuperCam. These Mars instruments assume that high resolution is not required to resolve low-density VNIR peaks. Using SuperLIBS, both high- and low-resolution VNIR spectra

were collected under identical protocols to determine if information is lost at low resolution. All other SuperLIBS protocols were chosen to match SuperCam as well as possible. Because the 1800 g mm^{-1} grating covers a relatively narrow wavelength range (approximately 50 nm), ten overlapping ranges were sampled at one location each for a total range of ∼500–860 nm. The UV, VIS, and low-resolution VNIR gratings cover a wider range, so only two ranges were sampled at 5 locations each for total ranges of ∼230–350 nm, ∼370–480 nm, and ∼510–870 nm, respectively. Spectra were concatenated to generate a single spectrum for each shot prior to analysis (Lepore, K., Dyar, M.D., Ytsma, C.R. 2024).

3.2.2 ChemLIBS

ChemLIBS is the first generation of LIBS instruments operating at MHC, starting in 2013. ChemCam-analog LIBS setup uses advanced laser technology for high-precision material analysis in controlled settings, but has a lower sample throughput and sensitivity compared to SuperLIBS. It is modeled after ChemCam, the LIBS instrument on the *Curiosity* rover, and supports the refinement of LIBS techniques for planetary exploration and laboratory studies.

The ChemLIBS instrument at Mount Holyoke College operates at 1064 nm using a Quantel Ultra 100 Nd:YAG laser with a 6 ns pulse, 10 Hz repetition rate, and 130 μm focused beam diameter. Pulse energy is recorded by a Newport 818E series pyroelectric detector and meter. Spectra are accumulated with three Ocean Optics HR2000+ spectrometers that cover the same wavelength ranges used by the ChemCam instrument (Ytsma, C.R. & Dyar, M.D. 2022). ChemLIBS operates on a fixed, continuous wavelength axis from 240.811 nm to 905.574 nm at a spectral resolution of 0.14–0.55 nm. The 'UV' camera resolution instrument covers 245 to 338 nm with 0.04–0.05 nm/pixel, the camera resolution 'VIS' instrument covers 388 to 469 nm with 0.04–0.05 nm/pixel, and the 'VNIR' camera resolution instrument covers 496 to 849 nm with 0.19–0.23 nm/pixel. The full spectrum contains 5485 wavelength channels. Spectra were acquired from 1 s integrations of 300 pulses from six locations per pellet (Ytsma, C.R. & Dyar, M.D. 2022).

These results of different atmospheres (Earth, Mars, and Vacuum) were published in a series

of papers highlighting use of sub-models and effects of plasma temperature, line and peak binning, and atmospheric variations on LIBS accuracy (Lepore, K., Ytsma C.R., Dyar, M.D. 2022; Lepore, K., Dyar, M.D., Ytsma, C.R. 2023; Lepore, K., Dyar, M.D., Ytsma, C.R. 2024; Lepore, K. et al. 2024).

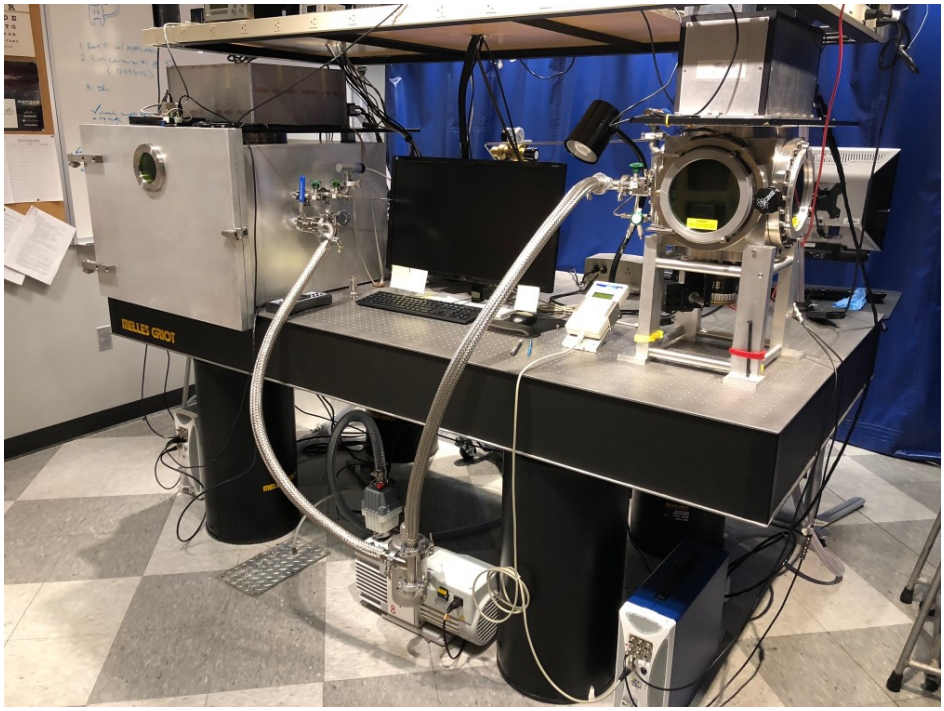

Figure 3.1: SuperLIBS (left) and ChemLIBS (right) instruments at the Mount Holyoke College Mineral Spectroscopy Lab: ChemLIBS was built to provide calibration support for ChemCam on *MSL Curiosity* (Lepore, K. et al. 2018)

3.2.3 pLIBS

The portable LIBS (pLIBS) instrument is a handheld analyzer used for fieldwork and rapid sample analysis under ambient conditions, with geological applications such as in situ Li quantification in rocks and minerals, slag analysis, geochemical imaging, and geochemical fingerprinting (Schlatter, N. & Lottermoser, B.G. 2023). It analyzes targets under ambient conditions, with a pulse of Ar gas prior to ablation to clear the sample surface and improve the signal to noise ratio (Lepore, K., Dyar, M.D., Ytsma, C.R. 2024). For this reason, when comparing spectra collected using pLIBS to those collected with SuperLIBS under Earth atmospheric conditions, the VNIR regions with visible Ar peaks were masked prior to

post-processing and modeling. Future applications may include geochemical characterization by astronauts on lunar and Martian endeavors. Two kinds of pLIBS instruments were used for the analysis: pLIBSZ300 and pLIBSZ903. The spectral resolution is between 0.2 and 0.7 nm. pLIBSZ300 was loaned from NASA Goddard Space Flight Center and pLIBSZ903 was loaned from SciAps, a company that specializes in portable analytical instruments.

3.2.3A pLIBS Z300

The pLIBS analyzer SciAps Z-300 features a pulsed Nd:YAG Class 3B laser, which emits laser light with a wavelength of 1064 nm and an energy of 5–6 mJ/pulse, with a pulse duration of 1 ns (Schlatter, N. & Lottermoser, B.G. 2023). The firing rate is user-adjustable within a range of 1 to 50 Hz. The laser focuses on a 100-μm spot size, and the emitted light is detected across a spectral range of 190 to 950 nm using three spectrometers equipped with time-gated charge-coupled diodes (Wise, M.A. et al. 2022). These spectrometers have the following resolution specifications: 190–365 nm (FWHM 0.18 nm), 365–620 nm (FWHM 0.24 nm), and 620–950 nm (FWHM 0.35 nm), covering 23,432 spectral channels (Wise, M.A. et al. 2022). By default, light collection occurs after a gate delay (t_d) of approximately 650 ns, with an integration time of 1 ms. The instrument can operate in argon atmosphere to enhance UV emission (190–300 nm). The pLIBS Z-300 operates on a fixed, continuous wavelength axis from 180 nm to 960 nm at a spectral resolution of 0.2–0.7 nm.

3.2.3B pLIBS Z903

The pLIBS instrument SciAps Z-903 hand-held LIBS analyzer operates at a fixed wavelength range and laser energy. Spectra were collected on a 4 $\times$ 3 raster on each sample, then averaged prior to downloading spectra files from the instrument. The pLIBS Z-903 operates on a fixed, continuous wavelength axis from 180 nm to 960 nm at a spectral resolution of 0.2 to 0.5 nm.

	Resolution (nm)	Atmosphere	Laser Energy (mJ)
SL 10K	0.08 (UV) 0.08 (VIS) 0.41 (VNIR)	M, E, V	2.4, 3.2, 4.0, 5.6, 7.2
SL 18K	0.08 (UV) 0.08 (VIS) 0.08 (VNIR)	M	2.4, 4.0, 5.6, 7.2
CL	0.14 (UV) 0.14 (VIS) 0.55 (VNIR)	M, E, V	1.6, 2.5, 3.5
Z300	0.03 (UV)	E	–
Z903	0.07-0.20 (UV)	E	–

Table 3.2: LIBS Reference Database Specifications (Lepore, K.H. et al. 2023)
Atmospheres: M = Mars, E = Earth, V = Vacuum

3.3 Unseen Data

3.3.1 ChemCam

The ChemCam (Chemistry & Camera) instrument aboard NASA's *Curiosity* rover represents a revolutionary advancement in planetary surface analysis, being the first instrument to perform laser-induced breakdown spectroscopy (LIBS) on another world. Since *Curiosity*'s 2012 landing in Gale Crater, this mast-mounted system has transformed our approach to remote geochemical studies by combining three key components: a precision laser capable of creating 300-550 μm diameter analysis spots, a telescope for light collection, and three spectrometers covering ultraviolet to near-infrared wavelengths (240-850 nm) (NASA 2024). Each laser pulse vaporizes a microscopic target area (0.3-0.6 mm) into a luminous plasma, with the emitted light revealing the sample's elemental composition. The accompanying Remote Micro-Imager simultaneously captures high-resolution context images, enabling direct correlation between chemical signatures and geological textures (NASA 2024).

As the pioneering LIBS instrument on Mars (Wiens, R.C. et al. 2013), ChemCam has compiled an unprecedented dataset of planetary surface compositions, acquiring nearly 200,000 LIBS spectra from approximately 1,500 individual targets during its operational lifetime (Alex Bolshakov n.d.). This extensive dataset revealed the need for improved calibration methods,

particularly when analyzing unexpected mineralogies like the alkali-rich feldspars discovered in Gale Crater (Payré, V. et al. 2017).

Clegg, S.M. et al. (2017) addresses these challenges by employing an expanded set of 408 laboratory standards analyzed under Mars-like atmospheric conditions (7 Torr CO_2). These standards span a significantly broader compositional range than those available before launch, better representing the mineralogical diversity encountered during *Curiosity*'s mission. The Dyar laboratory at Mount Holyoke College (MHC) provided 16 igneous samples, many of which were also in the original ChemCam flight model calibration (Wiens, R.C. et al. 2013). The advanced calibration approach combines two complementary multivariate techniques: a refined Partial Least Squares (PLS) regression that first classifies targets into geochemical families before applying specialized sub-models, and an Independent Component Analysis (ICA) method that identifies hidden spectral patterns (Clegg, S.M. et al. 2017). By weighted averaging of these approaches, the model achieves more accurate composition determinations across diverse samples while quantifying prediction uncertainties (Clegg, S.M. et al. 2017).

3.3.2 SuperCam

Building upon the proven architecture of ChemCam, the SuperCam instrument aboard NASA's *Perseverance* rover has been conducting groundbreaking analyses in Jezero crater since its successful landing in February 2021 (Cousin, A. et al. 2022). This next-generation instrument enhances its predecessor's capabilities through several critical upgrades, most notably the integration of laser-induced breakdown spectroscopy (LIBS) with Raman and visible-infrared spectroscopy (Anderson, R. et al. 2022). Together, these techniques enable comprehensive characterization of both chemical composition and mineral structure for targets as small as 0.5 mm at working distances up to 7 meters (Maurice, S. et al. 2021). The improved spectral resolution and expanded wavelength coverage provide better data quality while maintaining backward compatibility with ChemCam's extensive dataset, allowing for direct comparison of results between the two rover missions.

The foundation of this study relies on SuperCam's carefully curated library of 1198 laboratory spectra obtained from 334 well-characterized reference samples. As Anderson, R. et al. (2022)

documented, this comprehensive spectral collection was meticulously designed to capture the full mineralogical diversity expected in Jezero crater, incorporating insights from orbital remote sensing and previous surface missions. The reference suite spans common rock-forming minerals including silicates, sulfates, carbonates, and oxides, along with more specialized compositions like manganese ores and sodalite. Additionally, it includes laboratory analogs of the flight calibration targets (SCCTs) mounted on *Perseverance*'s deck (Anderson, R. et al. 2022).

For SuperCam, the original training/test set divisions specified in its metadata were preserved, as these groupings were carefully designed during the instrument's calibration process. ChemCam's metadata, however, lacked predefined training/test sets, necessitating implementation of a stratified sampling approach where samples were allocated to Fold 6 specifically to serve as the test set.

3.3.3 Mars Calibration Targets

Calibration targets containing geological samples/analogs with known compositions are used on the *Curiosity* and *Perseverance* rovers to test the accuracy and precision of analyses on Mars via repeated analyses (Ytsma, C.R., Dyar, M.D., Lepore, K. 2023).

3.3.3A ChemCam Mars Calibration Targets

Curiosity carried body-mounted ChemCam calibration targets consisting of three synthetic silicate glasses, a natural volcanic glass (Macusanite), four sulfate-bearing ceramics, and a certified graphite target (Fabre, C. et al. 2011). The glasses provide calibration based on representative basaltic compositions on Mars (picrite, shergottite), an analog of primitive Noachian crust (norite), and for extended calibration a more felsic igneous member (macusanite); the macusanite also contains a small amount of water (up to 0.5 wt.% H_2O) (Vaniman, D. et al. 2022). There are a total of 8 calibration targets where they were assembled at MHC, shown in Table 3.3.

Table 3.3: ChemCam Onboard Calibration Targets

Target Name	Number	Description
Macusanite	1	Felsic magmatic glass
Shergottite	4	Synthetic basaltic glass
Kga-d_Med-S	6	Medium sulfur ceramic
NAu-2_Low-S	7	Low sulfur ceramic
NAu-2_Med-S	8	Medium sulfur ceramic
NAu-2_Hi-S	9	High sulfur ceramic
Norite	2	Synthetic basaltic glass
Picrite	3	Synthetic basaltic glass

3.3.3B SuperCam Mars Calibration Targets

A key advantage of SuperCam's LIBS system is the inclusion of a much larger number of onboard calibration targets that cover a greater compositional range than ChemCam's (Anderson, R. et al. 2022). Replicates of the onboard SuperCam Calibration Targets (SCCTs) are included in the spectral library. SuperCam requires constant monitoring of the signal acquired from LIBS and Raman, and this is why a set of 36 calibration targets have been developed (Cousin, A. et al. 2022). Among those 36, there are 23 LIBS specific targets selected in order to fulfill both instrumental calibration needs and science objectives, shown in Table 3.4. These targets were derived from igneous minerals, standards, natural rocks, basaltic glasses doped with minor elements and synthesized targets. A Titanium plate, a former ChemCam shergottite target and a martian meteorite thin section are also on-board as calibration support (Cousin, A. et al. 2022).

Table 3.4: SuperCam Onboard Calibration Targets

Target	Metadata Name	Number	Index	Description
scct lanke0101	LANKE0104	20	4.3	Ankerite (Mg-carbonate)
scct lbhvo20406	LBHVO20401	18	4.1	Basalt analog (BHVO-2)
scct lca530106	LCA530101	9	2.3	Calparencite (Ca carbonate)
scct lcmb0006	LCMB0001	8	2.2	Si-rich rock (chert)
scct ljmn10106	LJMN10101	22	4.5	Mn-rich standard (JMN-1)
scct ljsc10304	LJSC10301	19	4.2	Soil analog (JSC-1)
scct lside0101	LSIDE0104	21	4.4	Siderite (Fe-carbonate)
scct nte010301	NTE010304	23	5.1	Silicate doped w/ Cu
scct nte020106	NTE020101	24	5.2	Silicate doped w/ Cr and Ba
scct nte030106	NTE030101	25	5.3	Silicate doped w/ Zn and Rb
scct nte040106	NTE040101	26	5.4	Silicate doped w/ Li and Sr
scct nte050301	NTE050304	27	5.5	Silicate doped w/ Ni
scct pmian0106	PMIAN0104	15	3.4	Andesine (plagioclase)
scct pmidn0302	PMIDN0303	13	3.2	Diopside (clinopyroxene)
scct pmien0602	PMIEN0603	16	3.5	Enstatite (pyroxene)
scct pmifa0306	PMIFA0301	14	3.3	Olivine (silicate)
scct pmifs0505	PMIFO0302	10	2.4	Ferrosilite (orthopyroxene)
scct pmior0507	PMIFS0504	12	3.1	Orthoclase (feldspar)
scct sherg02	SHERG02	28	5.6	ChemCam spare
scct tapag0206	TAPAG0204	11	2.5	Phosphate
scct titanium	-	-	-	Ti
scct tserp0102	TSERP0103	17	3.6	Serpentine
scct tsrich0404	TSRICH0402	7	2.1	Basalt-Sulfate mix

4 Methodology

4.1 Modeling Protocols

4.1.1 PyHAT

Numerous landed planetary spectroscopy instruments such as ChemCam (Wiens, R.C. et al. 2012), SuperCam (Maurice, S. et al. 2021), APXS (Rieder, R. et al. 2003), Mössbauer (Klingelhöfer, G. et al. 2003), PIXL (Allwood, A. et al. 2020), and Mini-TES (Christensen, P.R. et al. 2003) acquire spectra from individual target locations (Anderson, R.B. and Aneece, I.P. 2022). Interpretation of these point spectra is vital to the understanding of the geology of the targets analyzed and the geologic context in which they occur but point spectra can be difficult to work with. Even for members of instrument teams, it can be challenging to apply the latest processing and analysis techniques to the data, particularly owing to the preponderance of closed-source scripts. Additionally, many planetary scientists are not expert programmers and may not have funding to pay for expensive software packages and licenses, limiting their ability to interpret point spectral data (Anderson, R.B. and Aneece, I.P. 2022). Therefore, Python Hyperspectral Analysis Tool (PyHAT) Point Spectra Graphical User Interface (GUI) (PyHAT-PSG) was developed by the United States Geological Survey (USGS). This software is a versatile, free, and open-source library for the sole purpose of analyzing planetary spectral data using Python. PyHAT has a graphical user interface (GUI) that includes to work with simple CSV format files, looking up and matching up metadata to the spectra based on the values in user-specified columns, and removing, combining, splitting spectra datasets, and running multivarirate calibration methods and calibration transfer techniques.

4.1.2 ANACONDA

The ANACONDA cloud platform provides an efficient solution for data processing and analysis in this study (Anaconda Inc. 2025). This web-based environment allowed direct access to Jupyter Notebooks through a browser interface, eliminating the need for local

Python installations while maintaining full analytical capabilities (Anaconda Inc. 2025). The platform's accessibility and pre-configured tools significantly reduced setup time and technical barriers.

Key applications of ANACONDA for this study included data organization and preprocessing, particularly for standardizing CSV file headers and transposing the spectral data matrices. ANACONDA's visualization tools enabled the creation of this study's consistent figures for the results with carefully selected color palettes to enhance data interpretation, which made it easy to change any details about the sizing or location of text.

4.1.3 Cai R. Ytsma's Code

This study benefits significantly from computational contributions by Cai, a former research undergraduate student of Dr. Darby Dyar. Cai's foundational codebase establishes the critical infrastructure for this research, including: (1) systematic file organization protocols for training data preparation, (2) implementation of the core PLS regression algorithm with .asc file output capabilities, (3) generation of stratified k-fold validation sets, and (4) predictive modeling using PDS-PLS with comprehensive metric reporting. The existing code architecture shows particularly valuable for rapidly correct structure for the various scenario combinations during the training and testing phases.

4.2 Spectral Processing

Processing includes steps for pre- and post-processing. Pre-processing is a crucial step in data analysis, involving the removal of so-called data artifacts to make the preprocessed data better suited for subsequent data analysis. It directly influences the successful outcomes of all following steps in the pipeline and therefore the success of the entire experiment. With many papers and evaluations already published, the LIBS community is aware of what makes or breaks the preproccessing data analysis steps required. LIBS spectra require sophisticated pre-processing before analysis can be undertaken; the effects of these steps have been thoroughly examined. Data collected at MHC are preprocessed using a protocol similar to that used for ChemCam and SuperCam (Lepore, K. et al. 2024). This processing consists

of two steps. The first step is removing the non-laser-induced background or known as dark subtraction, wavelength alignment to Ti, and correction for instrument response, which takes spectra from "spectral intensity" or counts on the camera (non-unit) to units of spectral radiance, which can be compared among instruments (Wiens, R.C. et al. 2013). Now, the emission peak spectra in units of photons emitted per pulse in radiance units. The second step for processing is the postprocessing which involves continuum or baseline removal, and normalization (Wiens, R.C. et al. 2013).

All spectra used in this study is part of a larger LIBS reference database that will be made permanently available on the NASA Planetary Data System (PDS) Geosciences Node in 2026 (Lepore, K. et al. 2024).

4.2.1 Preprocessing: Part One

ChemLIBS and SuperLIBS follow the same preprocessing steps described below. pLIBS spectra are not treated until postprocessing because it is not possible to run an instrument response on them.

4.2.1A Dark-subtraction

To obtain an accurate measure of the signal, a dark spectrum is subtracted from the measured spectrum (*Subtracting dark spectra* n.d.). For every 25 samples analyzed, a dark spectrum is taken using the same integration time and detector temperatures. The dark spectrum typically consists of identical light collections but without the laser pulses. This subtraction removes an offset of 300–500 DNs, or counts on the camera, added during the analog-to-digital conversion of data from each spectrometer, to prevent negative values (Wiens, R.C. et al. 2013).

4.2.1B Wavelength Alignment to Ti

Titanium spectra are collected during every run for ChemLIBS and SuperLIBS. The MATLAB/Python code used is `icoshift` (Interval Correlation Optimized Shifting), an open-source and highly efficient program designed for solving signal alignment problems. It

is first run to align Ti to a baseline spectrum, and this alignment is then applied to each target spectrum.

4.2.1C Correction for Instrument Response

The Ocean Optics DH2000 calibration lamp is ensured to gather accurate wavelength calibration across the ultraviolet (UV), visible (VIS), and visible-near-infrared (VNIR) regions. The lamp emits light at well-characterized wavelengths, serving as a reference for spectral intensity. The system is calibrated to the known emission intensity of the lamp at the exact target ablation location with the detected spectra. This process to determine instrument sensitivity and response is performed separately for each camera and wavelength range in both SuperLIBS and ChemLIBS.

4.2.2 Preprocessing: Part Two

This step of pre-processing involves refining the raw spectral data to improve accuracy, by resampling onto a common axis between the files, removing the baseline, and normalizing signal interpretation for further analysis.

The same pre-processing steps are followed for each individual instruments and atmosphere before creating the common dataset.

4.2.2A Resampling on Common Axis

When working with data from two instruments that have different spectral resolutions, it is often necessary to resample one dataset to match the wavelengths of the other. PyHAT facilitates this process through its user interface. The first drop-down menu allows the user to select the dataset to be resampled from the available options, while the second drop-down menu specifies the reference dataset. The module uses the SciPy function scipy.interpolate.interp1d to linearly interpolate the selected dataset onto the wavelengths of the reference dataset. Importantly, the module does not extrapolate beyond the minimum or maximum wavelengths of the original dataset. If the dataset to be resampled or the reference dataset extends beyond the wavelength range of the other, both datasets are truncated to their overlapping range.

The universal common axis reference data spans from a minimum wavelength of 246.69 nm in the UV to a maximum wavelength of 849.03 nm in the VNIR range, with the common axis file containing approximately 6,500 pixel counts. For instance, if a dataset to be resampled has a range of 400 nm to 800 nm and the reference dataset spans 246.69 nm to 849.03 nm, the overlapping range (400 nm to 800 nm) is used for resampling. This step ensures that all resampled datasets match the spectral channel spacing of the reference dataset within the overlapping wavelengths.

4.2.2B Baseline Removal

LIBS spectra often exhibit background signals caused by bremsstrahlung and recombination processes (Sobran, P. et al. 2012). Baseline and random noise can negatively impact both qualitative and quantitative analyses if left unaddressed.

To correct for baseline drift, this study employs the adaptive iteratively reweighted Penalized Least Squares (airPLS) algorithm (Lepore, K. et al. 2024). This method iteratively adjusts the weights of sum squared errors (SSE) between the fitted baseline and the original signal. The weights are updated adaptively based on the difference between the previously fitted baseline and the original signal (Zhang, Z.M et al. 2010). It ensures robust and automated correction without requiring prior information or user intervention (Zhang, Z.M et al. 2010). The user can select smoothness, maximum number of iterations, and the convergence threshold for airPLS. The pLIBS spectra are reported in arbitrary units of intensity on a standard, fixed wavelength axis (Lepore, K., Dyar, M.D., Ytsma, C.R. 2024).

4.2.2C Normalization

Normalization recalculates peak intensities into a standardized range by dividing each measured wavelength's intensity by the sum of all intensities within that range (L1 norm) (Lepore, K. et al. 2024). This process is implemented through a custom Python function that: (1) converts each spectrum to an array, (2) divides all intensities by their sum, and (3) reconstructs the normalized dataset. The normalization is performed independently for three wavelength ranges: UV (246.7-338.4 nm), VIS (387.92-464.54 nm), and VNIR (508.3-849 nm).

Although pLIBS generates a continuous spectrum, the normalization algorithm successfully segmented it into the three discrete spectral ranges to maintain consistency with ChemLIBS and SuperLIBS data during cross-instrument comparisons.

Figures 4.1 4.2 4.3 4.4 show the final pre-proccessing of the spectra for each instrument/atmosphere/laser power variation by using the `pkey` column in the legend. These plots also show sample 9340H in its full spectrum, and separated regions: UV , VIS, VNIR, respectively.

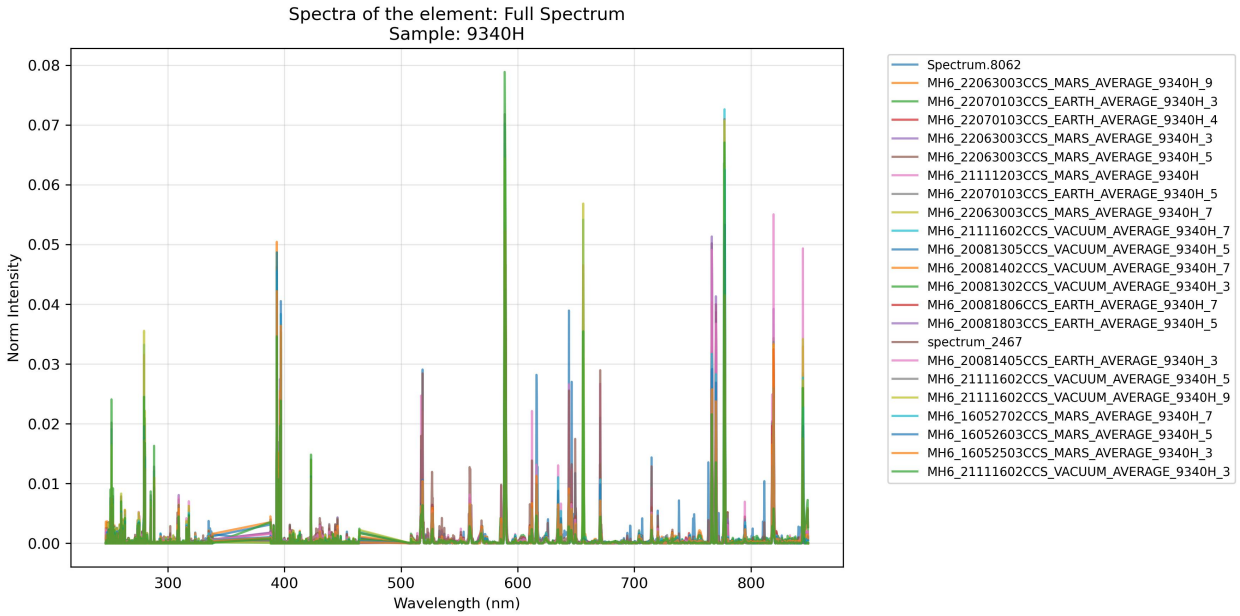

Figure 4.1: LIBS spectrum of sample 9340H showing the full wavelength range (246-870 nm). Characteristic emission lines across UV-VIS-NIR regions are visible.

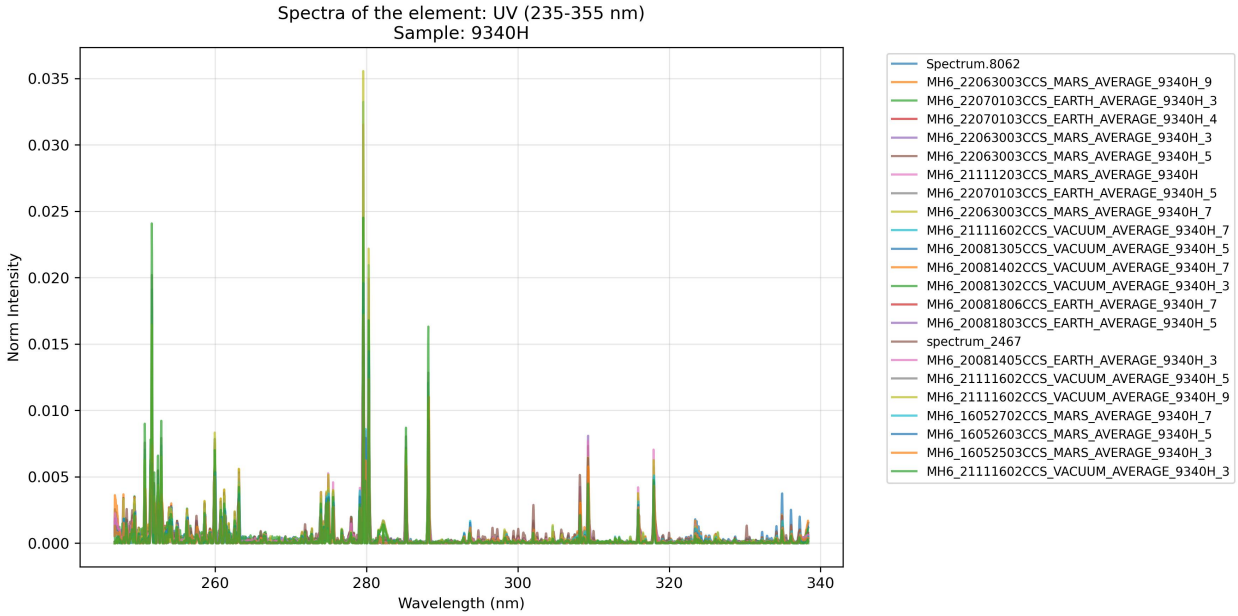

Figure 4.2: UV region (235-355 nm) of sample 9340H

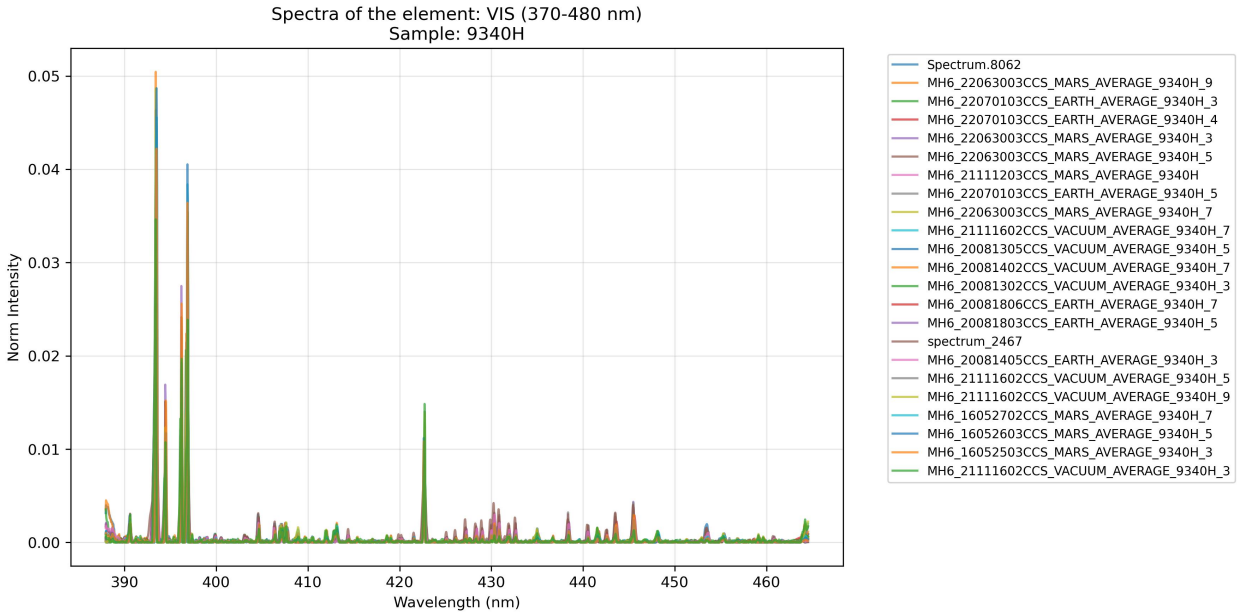

Figure 4.3: VIS spectrum (370-480 nm) of sample 9340H

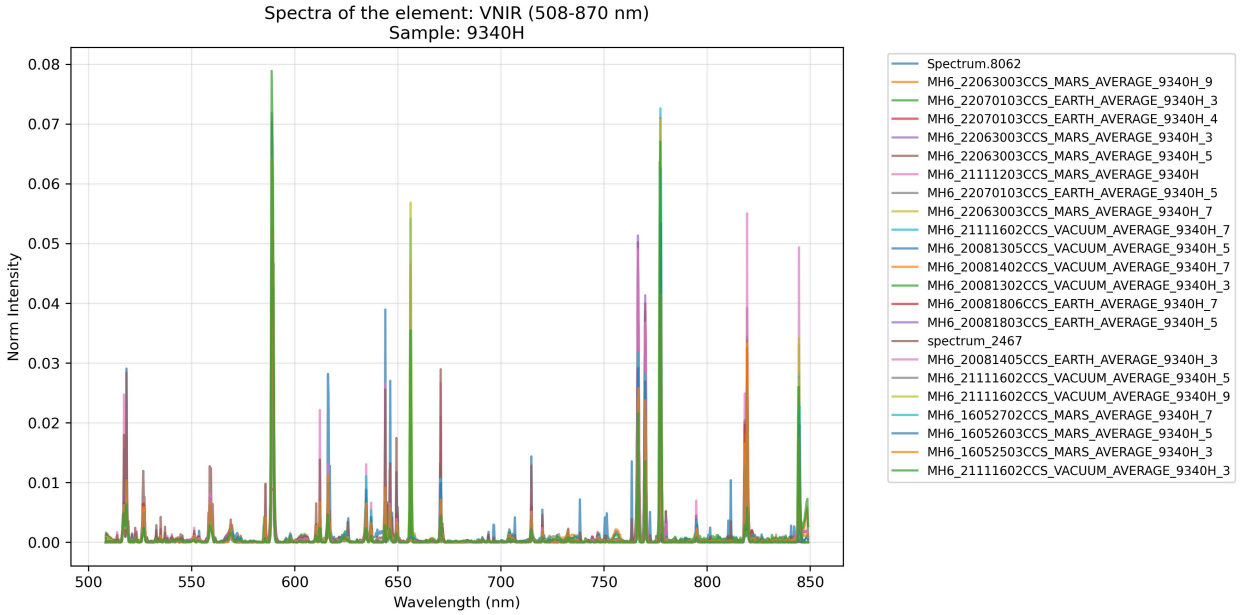

Figure 4.4: VNIR region (508-870 nm) of sample 9340H

4.3 Modeling Files

4.3.1 Metadata

The metadata files in this study share a consistent column structure across all datasets. Each metadata file contains the following standardized columns that provide essential geochemical information about the samples. The first column, labeled `pkey`, serves as the primary key identifier for each sample entry, which collects the unique reference throughout the database. Following this, the `Sample Name` column contains the specific designation or label assigned to each geological sample for identification purposes.

Subsequent columns present the same order of oxide composition measurements in weight percent (wt%) for eight major elements: silicon dioxide (SiO_2), titanium dioxide (TiO_2), aluminum oxide (Al_2O_3), iron oxide (Fe_2O_3), magnesium oxide (MgO), calcium oxide (CaO), sodium oxide (Na_2O), and potassium oxide (K_2O). These measurements constitute the fundamental geochemical characterization of each sample.

The dataset then includes a parallel set of columns with the suffix `Folds` appended to each

oxide measurement. These columns contain the cross validation fold information that is derived from the common dataset stratified folds columns, depending on the samples. The consistent fold assignments across all files and datasets make sure that samples maintain their designated training or testing roles throughout model development for key reproducible and valid results.

4.3.2 Spectra

The spectral data files maintain a consistent organizational structure while accommodating variations based on instrumentation and atmospheric conditions. Each file begins with the `pkey` column that corresponds exactly to the primary key identifiers in the associated metadata files.

The first row of each spectra file, following the `pkey` column, contains the resampled wavelength axis ranging from 246.67 nm to 849.03 nm. The remaining data matrix of 6572 spectra channels consists of normalized spectral intensity values used for examining the patterns to predict the composition of elements.

4.3.3 Common Dataset

Lepore, K. et al. (2024) demonstrated that prediction accuracies approach optimal performance when training data encompass all atmospheric conditions and laser energies. To construct a robust common dataset, the samples are all filtered within each instrument-atmosphere combination, retaining only those appearing in at least one laser power file. This approach generated nine files per instrument-atmosphere pair, resulting in a final dataset of 2,157 unique samples, which are shown in Table 4.1 with the other instruments' number of reference targets.

Instrument/Atmosphere	# Reference Targets
SL 10K MARS	2676
SL 10K EARTH	2644
SL 10K VACUUM	2858
SL 18K MARS	2872
CL MARS	2951
CL EARTH	2767
CL VACUUM	2695
pLIBS Z300	2849
pLIBS Z903	2686
Common Dataset	2157

Table 4.1: Number of reference targets sampled in each instrument configuration (Lepore, K.H. et al. 2023)

Table 4.2 shows the first ten sample names and their known elemental composition from the Common Dataset CSV file used for future methodology steps.

Sample Name	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	Na_2O	K_2O
0201H	48.49	1.31	17.47	9.94	2.32	13.64	3.81	1.29
0211H	47.8	1.86	15.98	13.43	4.09	8.58	3.43	3.45
0214H	53.02	1.31	16.86	9.01	0.88	12.55	4.72	0.77
0230H	56.39	0.45	18.97	9.57	0.54	3.08	6.12	4.35
0232H	48.52	1.11	17.03	9.85	6.5	11.09	3.67	1.54
0242H	54.38	1.3	14.9	9.49	0.8	11.54	4.78	2.13
0243H	56.5	1.44	12.21	9.96	1.15	10.72	5.08	2.21
0245H	50.88	0.93	18.04	9.4	4.86	9.23	4.49	2.04
0248H	43.06	0.93	18.38	13.37	2.47	12.91	6.05	1.94
0249H	47.38	1.26	17.75	11.36	5.25	10.52	4.03	1.85

Table 4.2: Preview of the first 10 samples and their elemental composition

4.4 Partial Least Squares

The complete common dataset for all instrument-atmosphere combinations is prepared for multivariate analysis using partial least squares (PLS) regression. To enable effective machine learning model training, the data is split into an 75% training set for PLS development and a 25% test set for model validation. The PLS analysis and cross-validation are implemented using Cai's computational framework.

4.4.1 Stratified Folds for Cross Validation

This study implements a robust validation framework combining stratified 5-fold cross-validation with an independent held-out test set, following the methodology established by Dyar, M.D & Ytsma, C. (2021). Their comprehensive analysis demonstrated that held-out testing provides superior accuracy assessment compared to cross-validation alone, as it better represents real-world prediction scenarios. The cross-validation process carefully maintains the distribution of target element concentrations across all folds through a stratification procedure where samples are first sorted by concentration values, then systematically distributed to ensure each fold captures the full compositional range. This approach is implemented independently for each major element to account for varying geochemical distributions. The choice of five folds aligns with findings from Anderson, R. et al. (2022), who showed this configuration optimally balances computational efficiency and reliable error estimation while adequately representing sample variability. Validation results reveal distinct performance characteristics between the two methods: RMSE-P values follow smooth polynomial trends across all elements, suggesting the held-out samples properly represent the source population, while RMSE-CV shows greater dispersion that may reflect irregular sampling across cross-validation folds. This observation reinforces the importance of the dual validation approach, where cross-validation fine-tunes model parameters while the held-out set provides the definitive performance assessment.

The common dataset of 2,157 samples is stratified to guarantee that the training and testing datasets are consistent across all instruments and atmospheric conditions for accurate model evaluations. By maintaining the same distribution of elements in each fold, any mix-ups are avoided in understanding the composition of these elements.

The selection of the test fold is not done randomly, but through calculating the most representative fold. The `get_most_representative_fold` method from Cai's code is used to identify the most representative fold for each element in a stratified dataset and ensures that the test data truly represents the full geochemical variability. The method begins by identifying the fold columns (1, 2, 3, 4, and 5) in the common dataset composition stratified metadata file. For each element, the histogram of the full dataset is computed using the

Freedman-Diaconis rule to determine optimal binning. The mean absolute difference between the histogram of each fold and the full dataset histogram is then calculated, and the fold with the smallest difference is identified as the most representative. The histograms of the full dataset and each fold for each element are plotted to allow them to be visualized for comparison. The TiO_2 original dataset and histogram differences plots are shown in Figure 4.5, where the smallest fold difference is fold 1, which now becomes the test fold. Once the most representative folds are determined, the metadata are updated by assigning the specific fold value (fold 6) to the identified representative folds for each element. The rest of the histograms for each element is in the Appendix Section.

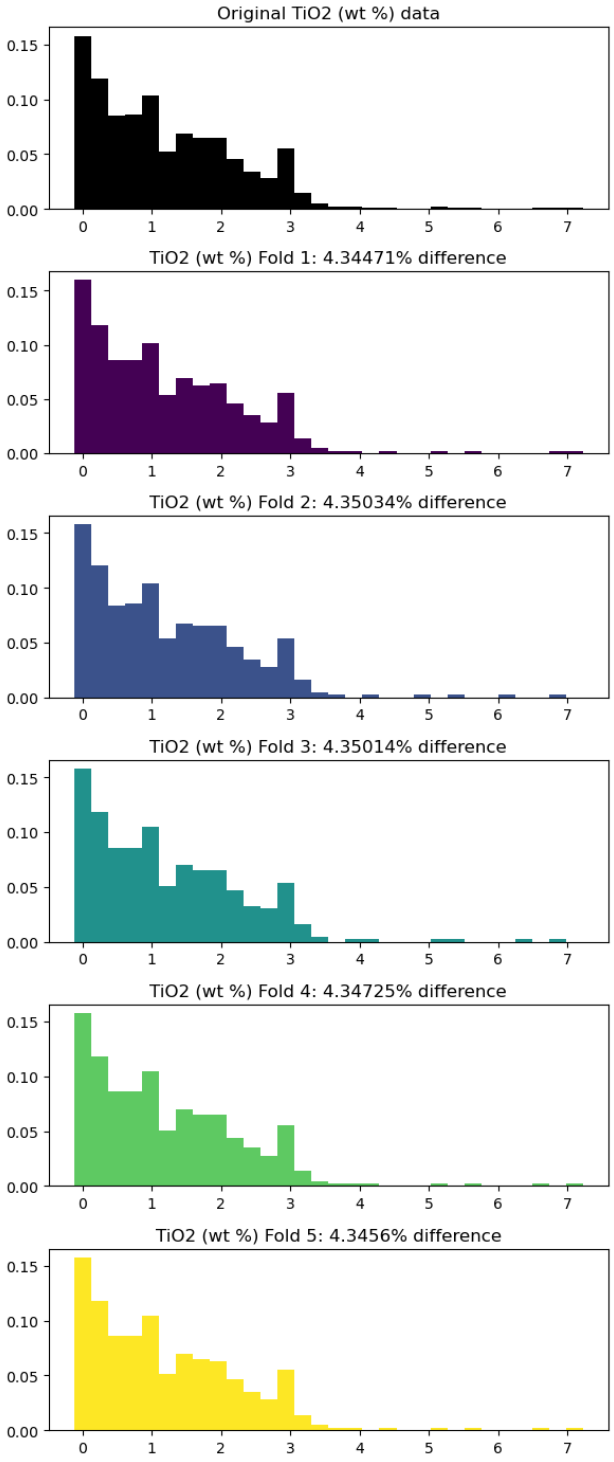

Figure 4.5: Getting the Most Representative Fold for TiO_2 Histograms

4.4.2 Training

The five-fold cross-validation approach provides robust assessment of model generalizability by systematically evaluating performance across different data partitions. This study employs partial least squares (PLS) regression, the predominant multivariate technique for LIBS analysis (Dyar, M.D & Ytsma, C. 2021). PLS operates by projecting the high-dimensional spectral data (6572 channels in this study) into a reduced latent space that maximizes covariance between the predictor matrix $\mathbf{X}$ (spectral intensities) and response matrix $\mathbf{Y}$ (elemental concentrations). Through an iterative algorithm, PLS first identifies orthogonal components that capture the maximum variance in $\mathbf{X}$ while being most predictive of $\mathbf{Y}$, then applies ordinary least squares regression in this reduced space to minimize prediction error.

The critical hyperparameter k (number of latent components) is optimized via cross-validation across the range 2–30. For each candidate k, the training data underwent five complete iterations where models trained on four folds predicted the held-out fifth fold. The optimal k is selected based on minimizing the average root mean square error of cross-validation (RMSE-CV), balancing model complexity against overfitting risks (Lepore, K. et al. 2024). This dimensionality reduction addresses the inherent collinearity in LIBS spectra while preserving the chemically meaningful variance needed for accurate concentration predictions.

Final PLS models are then trained using all available calibration data at the optimized component count, with their predictive performance rigorously evaluated using the statistical metrics detailed in subsequent sections.

4.5 Calibration Transfer: PDS-PLS

This study employs PDS-PLS (Piecewise Direct Standardization - Partial Least Squares) for several compelling reasons. First, the method has proven particularly effective for its accuracy and relative speed when instruments are similar through previous studies as PDS combines PLS in order to remove sharp edges within the transferred spectra (Lepore, K., Dyar, M.D., Ytsma, C.R. 2024; Wang, Y. et al. 1991). The choice is further supported by training the giant supermodel with PLS modeling for methodological consistency. PDS divides the

spectra into small windows and determines the number of neighboring wavelengths used to map spectra from the target instrument to the reference instrument, while PLS reduces the dimensionality of the data and models the relationship between spectral features and target compositions (Lepore, K. et al. 2024).

To optimize model performance, cross-validation works with the Calibration Transfer Cross-Validation (CV) module in PyHAT. From specific parameters settings inputted explained in the next section, the module outputs a calculated root mean squared error (RMSE) to evaluate performance, with lower RMSE indicating better alignment between transformed and target spectra. There is more information about the specific training subsets of what data is calibration transferring to/from in the Results section.

4.5.1 Parameters Selection

The selection of specific parameter ranges for both window size (5-31 channels) and PLS components (1-8) is guided by fundamental spectroscopic principles and previous studies. If the window size is too small, the differences between instruments might not be contained in the selected spectrum, resulting in poor transfer performance (Zhang, X. et al. 2022). If the window size is chosen too large, the selected spectral data will involve some irrelevant information, which not only requires too much computation, but leads to the phenomenon of overcorrection, which eventually also leads to poor performance of the model transfer (Zhang, X. et al. 2022). The upper limit of 31 channels prevents excessive broadening that could mask important spectral details, while the lower limit of 5 channels looks at sufficient neighboring information for robust transformations.

For PLS components, which is the same parameter as previously mentioned for the training stage of the model, the 1-8 range is determined through systematic cross-validation studies that revealed most chemically relevant variance in LIBS spectra and analyzed in past papers (Lepore, K. et al. 2024). Fewer components suffice for simple elemental compositions, while the upper range handles complex mineral matrices and occasional spectral interference.

4.6 Unseen Data: ChemCam, SuperCam, and Onboard Calibration Targets

To evaluate the performance of the trained models on unseen data, the same exact modeling protocols used for each instruments and atmospheric conditions are applied to the ChemCam, SuperCam and their respective calibration target datasets to ensure consistency in preprocessing and evaluation steps. The spectra data from these datasets are first resampled to match the resolution of the common data. Baseline removal and normalization are then performed using Cai's code. Predictions are generated using the same stratified fold (Fold 6) as in the baseline modeling step to maintain consistency in evaluation.

4.7 Data Organization: Experimental Scenarios

This study employs a systematic experimental framework (Tables 4.3–4.4) designed to address fundamental challenges in LIBS cross-instrument analysis through four carefully constructed scenarios. The design progresses from controlled benchmarks to complex operational simulations, each targeting specific knowledge gaps in calibration transfer methodology. Scenario 0 serves as the foundational benchmark, establishing baseline performance metrics through three configurations: the primary reference (A1/A2) quantifies optimal laboratory performance, while secondary benchmarks (B1a/B2a and C1a/C2) evaluate raw and calibration-transferred instrument data that are not incorporated in the reference model, respectively. These controls determine rigorous isolation of instrument-specific effects from methodological limitations.

The core investigation unfolds through Scenarios 1-4, each motivated by distinct research imperatives. Scenario 1 addresses the critical need for portable methodologies in planetary science by testing whether reference models (A1) can directly analyze new instruments (B2a) or their CT-processed counterparts (C2) without mission-specific recalibration. Scenario 2 evaluates practical hybrid training strategies (A1+B1a $\rightarrow$ B2b and A1+C1 $\rightarrow$ C2) to determine if limited instrument-specific data can bridge domain gaps - a solution particularly relevant for resource-constrained missions. Scenario 3 systematically examines the value of

Mars-specific data (B1b) through progressive training set augmentations where the calibration transfer target domain matches the test set, providing empirical evidence to guide future mission calibration campaigns. Finally, Scenario 4 challenges conventional assumptions about transfer symmetry by inverting the CT direction (SuperLIBS→SuperCam), with important implications for data repurposing across missions with different instrumental characteristics.

This graduated experimental design dives deep into systematic investigation of how strategic training set modifications influence model performance across instruments. The analysis specifically quantifies: (1) the relative benefits of incorporating raw versus transformed instrument data, (2) the role of spectral similarity between training and test conditions, and (3) the optimization of limited calibration resources. By evaluating these factors through carefully designed scenarios that progress from idealized controls to operational simulations, the work establishes quantitative relationships between training set composition and prediction accuracy under realistic analytical constraints.

Table 4.3: Data Variables Definitions

Variable	Description
A	Complete reference dataset containing all MHC laboratory spectra
A1	Train data
A2	Test data
B	New instrument data from ChemCam or SuperCam (LANL pre-launch) or Mars Onboard Calibration Targets or SuperLIBS 10K/18K Mars
B1a	Training data from ChemCam or SuperCam
B1b	Training data from Mars Calibration Targets or SuperLIBS 10K/18K Mars
B2a	Test data from ChemCam or SuperCam
B2b	Test data from Mars Calibration Targets or SuperLIBS 10K Mars
C	New instrument data from ChemCam or SuperCam (LANL pre-launch) or Mars Onboard Calibration Targets or SuperLIBS 10K/18K Mars with Calibration Transfer
C1a	Training data from ChemCam or SuperCam CT to SL 10K Mars
C1b	Training data from SuperLIBS 10K/18K Mars CT to SuperCam
C2	Test data from ChemCam or SuperCam CT to SL 10K Mars

Table 4.4: Experimental Scenarios

Sc.	Train Data	Test Data	Experimental Design
0	A1	A2	Reference model
0	B1a	B2a	Unseen instrument
0	C1a	C2	Unseen transformed instrument
1	A1 A1+B1a	B2a	Evaluate model augmentations on unseen data
1	A1 A1+C1	C2	Evaluate model augmentations on unseen transformed data
2	A1 B1a C1 A1+B1a A1+C1	B2b	Evaluate model augmentations on one instrument from model data
3	B1a A1+B1a B1a+B1b A1+B1a+B1b B1b+B1b+B1a*	B2b	Evaluate model augmentations on unseen instrument data
4**	B1b C1b	B2a	Evaluate one transformed instrument data on unseen instrument data

*B1b appears twice due to adding SL 10k/18K Mars and Mars Onboard Calibration Targets for a SuperCam only scenario
**SuperCam only scenario

4.8 Evaluation of model performance

4.8.1 Root Mean Squared Error of Prediction (RMSE-P)

The Root Mean Squared Error of Prediction (RMSE-P) quantifies the average magnitude of error between the predicted and actual oxide concentration from the test spectra. It is defined as:

$$\text{RMSE-P} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

where N is the number of samples, y_i is the measured oxide concentration, and $\hat{y}_i$ is the corresponding value predicted by the PLS model (Lasue, J. et al. 2012). RMSE-P is expressed in the same units as the target variable (in this case, wt%), making it a useful metric for interpreting the typical prediction error in a physical context.

Lower RMSE-P values indicate better predictive accuracy. Since this metric directly reflects the deviation of model predictions from the reference measurements, it is widely used to assess the practical performance of regression models in spectroscopic applications.

4.8.2 R^2 (Coefficient of Determination)

The coefficient of determination (R^2) quantifies how well the predicted oxide concentrations from the PLS model explain the variability in the measured values. It is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

where y_i and $\hat{y}_i$ are the measured and predicted oxide concentrations, respectively, $\bar{y}$ is the mean of the measured values, and N is the number of samples (Ejiro Onose 2023). This metric compares the model's predictive performance to a baseline model that simply predicts the mean of the measured values .

In this study, R^2 values are capped at a minimum of 0.0 to avoid misinterpretation because

negative values indicate that the model performs worse than predicting the mean and are not meaningful in this context.

4.8.3 Slope

The slope (β_1) of the ordinary least squares regression between predicted ($\hat{y}_i$) and measured (y_i) oxide concentrations reveals proportional calibration biases. Calculated as:

$$\beta_1 = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

where y_i represents reference concentrations (wt%), $\hat{y}_i$ denotes LIBS-predicted values (wt%), $\bar{y}$ and $\bar{\hat{y}}$ are their respective means, and N is the number of test samples in fold 6 (Valentina Alto 2025). The slope interpretation follows three cases: (1) $\beta_1 = 1.0$ indicates ideal proportional agreement, (2) $\beta_1 < 1.0$ reflects systematic under-prediction of high concentrations and over-prediction of low values, while (3) $\beta_1 > 1.0$ shows the opposite trend.

4.8.4 Normalized Intercept

The intercept (β_0) from partial least squares regression is normalized by the mean measured concentration ($\bar{y}_{\text{train}}$) to enable cross-element comparison of constant biases across oxides with varying concentration ranges. The normalization is calculated as:

$$\text{Normalized Intercept} = \frac{\beta_0}{\bar{y}_{\text{train}}}, \quad \text{where} \quad \bar{y}_{\text{train}} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

where y_i represents reference concentrations (wt%) in the training set, $\bar{y}_{\text{train}}$ is their arithmetic mean, and N is the number of training samples. The mean is employed rather than the median for three critical reasons: (1) it preserves the mathematical relationship with the ordinary least squares regression framework where $\beta_0 = \bar{y} - \beta_1\bar{x}$, maintaining theoretical consistency; (2) it accounts for all training samples equally, unlike the median which only considers the central value; and (3) it properly scales the intercept relative to the full concentration distribution, particularly important for elements with skewed abundance distributions. Interpretation

follows: (1) values approaching 0 indicate minimal constant bias, (2) positive values reveal systematic over-prediction at low concentrations, while (3) negative values demonstrate under-prediction at low concentrations.

Table 4.5: PLS Regression Evaluation Metrics and Their Significance

Metric	Definition	Motivation for PLS Analysis
RMSE-P	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$	Quantifies *absolute prediction error* in wt% units. Critical for assessing real-world analytical precision. Lower values indicate better absolute accuracy.
RMSE-P%	$\left(\frac{\text{RMSE-P}}{\text{Median Concentration}}\right) \times 100$	Normalizes errors to element abundance levels. Creates fair comparison across major/trace elements by showing relative error magnitude. Particularly valuable for low-concentration species.
R^2	$1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$	Measures *proportion of variance explained*. Values close to 1.0 indicate strong spectral-composition relationships. Capped at 0 to prevent misinterpretation of negative values.
Slope (β_1)	$\frac{\sum(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sum(y_i - \bar{y})^2}$	Diagnoses *proportional calibration biases*: $\beta_1 < 1.0$ suggests under-prediction of high concentrations; $\beta_1 > 1.0$ indicates the opposite. Ideal value = 1.0.
Normalized Intercept	$\beta_0 / \bar{y}_{\text{train}}$	Reveals *constant baseline biases*. Near 0 indicates no systematic offset. Positive/negative values show consistent over/under-prediction at low concentrations. Normalization creates cross-element comparison.

5 Results

5.1 Benchmark Performance

The results of training and testing the giant combined model called "All" in the subsequent tables present both RMSE-P and RMSE-P% metrics to enable comprehensive evaluation of model performance across elements with varying concentration ranges. While RMSE-P provides absolute error measurements in weight percent (wt%), RMSE-P% expresses these errors relative to each element's median concentration (Table 5.1), calculated as:

$$\text{RMSE-P\%} = \left(\frac{\text{RMSE-P}}{\text{Median Concentration}} \right) \times 100$$

This dual approach offers distinct advantages: (1) RMSE-P maintains physical interpretability in wt% units, while (2) RMSE-P% enables fair performance comparisons across major and minor elements by normalizing errors to typical abundance levels. Lower RMSE-P% values indicate better relative prediction accuracy, particularly important for trace elements where absolute errors might appear deceptively small yet represent significant proportional deviations.

Table 5.1: Median Concentrations of Major Elements from "All" model

Element	**Median Concentration (wt%)**
SiO_2 (wt%)	51.78
TiO_2 (wt%)	1.05
Al_2O_3 (wt%)	13.72
Fe_2O_3 (wt%)	9.32
MgO (wt%)	3.95
CaO (wt%)	7.19
Na_2O (wt%)	2.71
K_2O (wt%)	1.17

The results files for "All" with PLS are presented in the tables are divided into two main sections: training and testing. The training-related columns describe the performance of the PLS model during the model-building phase. Here, `n_train` indicates the number of samples

used to train the model, while `model_type` specifies the type of model employed, which in this case is PLS. The `rmsecv` (Root Mean Square Error of Cross-Validation) measures the model's prediction error during cross-validation, providing an estimate of how well the model generalizes to unseen data. The `model_params` column indicates the number of latent variables or components used in the PLS model, which are critical for capturing the underlying structure of the data. The number of latent components (`model_params`) directly controls the bias-variance tradeoff, where too few components introduce bias by failing to capture important spectral features, while too many increase variance by modeling noise. The `model_intercept` represents the constant term in the regression equation, and `rmsec` (Root Mean Square Error of Calibration) quantifies the error between the observed and predicted values in the training set. The R^2 `_train` (Coefficient of Determination for Training) and `adj_`R^2 `_train` (Adjusted R^2 for Training) describe the proportion of variance in the dependent variable explained by the model, with the latter adjusting for the number of predictors to avoid overfitting.

The testing-related columns evaluate the model's performance on an independent test set, which was not used during training. The `test_fold` indicates the cross-validation fold used for testing which will always be fold 6 from the stratification step, while `n_test` specifies the number of samples in the test set. The `RMSE-P` (Root Mean Square Error of Prediction) measures the prediction error on the test set, which provides an assessment of the model's ability to generalize to new data. The R^2 `_test` (Coefficient of Determination for Testing) and `adj_`R^2 `_test` (Adjusted R^2 for Testing) indicate how well the model explains the variance in the test set, with the adjusted R^2 accounting for the number of predictors.

Tables 5.2 and 5.3 show the results for SuperLIBS 10k Earth.

Table 5.2: SuperLIBS 10k Earth Modeling Results (Training)

Variable	n_train	rmsecv	model_params	model_intercept	rmsec
SiO_2 (wt%)	5461	4.91	13	57.47	4.36
TiO_2 (wt%)	5396	0.40	17	1.29	0.37
Al_2O_3 (wt%)	5364	1.88	17	13.38	1.71
Fe_2O_3 (wt%)	5435	1.88	19	8.51	1.70
MgO (wt%)	5128	1.26	27	4.76	1.08
CaO (wt%)	5441	1.24	22	6.28	1.04
Na_2O (wt%)	5413	0.57	18	2.70	0.52
K_2O (wt%)	5407	0.75	11	1.90	0.70

Table 5.3: SuperLIBS 10k Earth Modeling Results (Testing)

Variable	R^2 _train	adj_R^2 _train	test_fold	n_test	RMSE-P	R^2 _test
SiO_2 (wt%)	0.89	0.89	6	1374	5.40	0.83
TiO_2 (wt%)	0.86	0.86	6	1353	0.37	0.85
Al_2O_3 (wt%)	0.79	0.79	6	1336	1.91	0.75
Fe_2O_3 (wt%)	0.88	0.88	6	1370	1.77	0.87
MgO (wt%)	0.94	0.94	6	1280	1.26	0.91
CaO (wt%)	0.95	0.95	6	1361	1.13	0.94
Na_2O (wt%)	0.84	0.84	6	1363	0.63	0.76
K_2O (wt%)	0.86	0.86	6	1353	0.69	0.86

5.1.1 Individual

This table shows the gathered RMSE-Ps columns from the individual nine datasets and provides a comprehensive comparison of the model's predictive performance across different atmospheres/instruments, when they are not merged together yet. The results indicate varying levels of accuracy, with lower RMSE-P values suggesting that the model makes more accurate predictions and fits the data well (Shittu Olumide 2023).

Table 5.4: Individual RMSE-P Values (Part 1)

Variable	CL Earth	CL Mars	CL Vacuum	pLIBS Z300	pLIBS Z903
SiO_2 (wt %)	5.51	4.38	4.38	5.90	5.30
TiO_2 (wt %)	0.42	0.38	0.41	0.41	0.42
Al_2O_3 (wt %)	2.34	1.56	1.63	1.96	1.95
Fe_2O_3 (wt %)	2.02	1.69	1.92	2.19	1.98
MgO (wt %)	1.32	1.14	1.24	1.61	1.68
CaO (wt %)	1.45	1.07	1.02	1.54	1.22
Na_2O (wt %)	0.64	0.52	0.53	0.70	0.70
K_2O (wt %)	0.74	0.63	0.68	0.87	0.70
Averages	1.80	1.42	1.48	1.90	1.75

Table 5.5: Individual RMSE-P Values (Part 2)

Variable	SL 10k Earth	SL 10k Mars	SL 10k Vacuum	SL 18k Mars
SiO_2 (wt %)	5.40	4.87	4.37	4.65
TiO_2 (wt %)	0.37	0.38	0.35	0.31
Al_2O_3 (wt %)	1.91	1.61	1.61	1.34
Fe_2O_3 (wt %)	1.77	1.65	1.64	1.59
MgO (wt %)	1.26	1.08	1.07	0.92
CaO (wt %)	1.13	1.03	0.92	0.91
Na_2O (wt %)	0.63	0.58	0.48	0.62
K_2O (wt %)	0.69	0.69	0.59	0.71
Averages	1.65	1.49	1.38	1.38

Table 5.6: Individual RMSE-P% Values (Part 1)

Variable	CL Earth	CL Mars	CL Vacuum	pLIBS Z300	pLIBS Z903
SiO_2 (wt %)	9.52%	7.57%	7.57%	10.20%	9.16%
TiO_2 (wt %)	33.07%	29.92%	32.28%	32.28%	33.07%
Al_2O_3 (wt %)	17.53%	11.69%	12.21%	14.68%	14.61%
Fe_2O_3 (wt %)	24.05%	20.12%	22.86%	26.07%	23.57%
MgO (wt %)	28.51%	24.62%	26.78%	34.77%	36.29%
CaO (wt %)	23.62%	17.43%	16.61%	25.08%	19.87%
Na_2O (wt %)	23.88%	19.40%	19.78%	26.12%	26.12%
K_2O (wt %)	38.14%	32.47%	35.05%	44.85%	36.08%
Averages	24.79%	20.28%	21.65%	26.63%	24.85%

Table 5.7: Individual RMSE-P% Values (Part 2)

Variable	SL 10k Earth	SL 10k Mars	SL 10k Vacuum	SL 18k Mars
SiO_2 (wt %)	9.33%	8.42%	7.55%	8.04%
TiO_2 (wt %)	29.13%	29.92%	27.56%	24.41%
Al_2O_3 (wt %)	14.31%	12.06%	12.06%	10.04%
Fe_2O_3 (wt %)	21.07%	19.64%	19.52%	18.93%
MgO (wt %)	27.21%	23.33%	23.11%	19.87%
CaO (wt %)	18.40%	16.78%	14.98%	14.82%
Na_2O (wt %)	23.51%	21.64%	17.91%	23.13%
K_2O (wt %)	35.57%	35.57%	30.41%	36.60%
Averages	22.32%	20.80%	19.14%	19.48%

5.1.2 All

Table 5.8 presents the gathered RMSE-Ps columns from the combined common dataset of all nine datasets.

Table 5.8: All RMSE-P Values and RMSE-P%

Variable	RMSE-P	RMSE-P%
SiO_2 (wt %)	5.09	9.83%
TiO_2 (wt %)	0.41	39.05%
Al_2O_3 (wt %)	1.95	14.21%
Fe_2O_3 (wt %)	1.94	20.82%
MgO (wt %)	1.38	34.94%
CaO (wt %)	1.24	17.25%
Na_2O (wt %)	0.64	23.62%
K_2O (wt %)	0.73	62.39%
Average	1.67	27.27%

5.1.3 All - Individual

Tables 5.9, 5.10, 5.11, and 5.12 present the RMSE-P values derived from the combined common dataset, and are subsequently stratified by individual datasets to quantify the variation in prediction accuracy from .

Table 5.9: All (Individual) RMSE-P Values (Part 1)

Variable	CL Earth	CL Mars	CL Vacuum	pLIBS Z300	pLIBS Z903
SiO_2 (wt %)	5.93	4.32	4.78	5.84	5.81
TiO_2 (wt %)	0.44	0.44	0.44	0.46	0.46
Al_2O_3 (wt %)	2.43	1.76	1.96	2.26	2.11
Fe_2O_3 (wt %)	2.11	1.86	1.99	2.17	2.08
MgO (wt %)	1.40	1.26	1.41	2.01	1.73
CaO (wt %)	1.62	1.20	1.16	1.73	1.28
Na_2O (wt %)	0.70	0.59	0.63	0.81	0.76
K_2O (wt %)	0.82	0.72	0.75	0.92	0.75
Averages	1.93	1.52	1.64	2.02	1.87

Table 5.10: All (Individual) RMSE-P Values (Part 2)

Variable	SL 10k Earth	SL 10k Mars	SL 10k Vacuum	SL 18k Mars
SiO_2 (wt %)	5.43	4.94	4.77	4.53
TiO_2 (wt %)	0.41	0.37	0.37	0.34
Al_2O_3 (wt %)	2.01	1.80	1.68	1.77
Fe_2O_3 (wt %)	2.01	1.86	1.80	1.75
MgO (wt %)	1.48	1.23	1.20	1.28
CaO (wt %)	1.20	1.06	1.06	1.13
Na_2O (wt %)	0.70	0.61	0.54	0.68
K_2O (wt %)	0.78	0.67	0.62	0.71
Averages	1.75	1.57	1.50	1.52

Table 5.11: All (Individual) RMSE-P% Values (Part 1)

Variable	CL Earth	CL Mars	CL Vacuum	pLIBS Z300	pLIBS Z903
SiO_2 (wt %)	10.25%	7.47%	8.26%	10.09%	10.04%
TiO_2 (wt %)	34.65%	34.65%	34.65%	36.22%	36.22%
Al_2O_3 (wt %)	18.20%	13.18%	14.68%	16.93%	15.81%
Fe_2O_3 (wt %)	25.12%	22.14%	23.69%	25.83%	24.76%
MgO (wt %)	30.24%	27.21%	30.45%	43.41%	37.37%
CaO (wt %)	26.38%	19.54%	18.89%	28.18%	20.85%
Na_2O (wt %)	26.12%	22.01%	23.51%	30.22%	28.36%
K_2O (wt %)	42.27%	37.11%	38.66%	47.42%	38.66%
Averages	26.65%	22.79%	23.98%	29.79%	26.41%

Table 5.12: All (Individual) RMSE-P% Values (Part 2)

Variable	SL 10k Earth	SL 10k Mars	SL 10k Vacuum	SL 18k Mars
SiO_2 (wt %)	9.39%	8.54%	8.25%	7.83%
TiO_2 (wt %)	32.28%	29.13%	29.13%	26.77%
Al_2O_3 (wt %)	15.06%	13.48%	12.58%	13.26%
Fe_2O_3 (wt %)	23.93%	22.14%	21.43%	20.83%
MgO (wt %)	31.97%	26.57%	25.92%	27.65%
CaO (wt %)	19.54%	17.26%	17.26%	18.40%
Na_2O (wt %)	26.12%	22.76%	20.15%	25.37%
K_2O (wt %)	40.21%	34.54%	31.96%	36.60%
Averages	24.81%	21.68%	20.84%	21.34%

The All-Individual RMSE-P and RMSE-P over median concentrations values corroborate the trends observed in the Individual results. The SL 10k Vacuum dataset consistently demonstrates the lowest RMSE-P and RMSE-P % values for most variables, with an overall average of 1.50 and 20.84 %, respectively. This superior performance likely stems from the absence of atmospheric interference, which minimizes spectral noise and variability. In contrast, datasets such as CL Earth and pLIBS Z300 show elevated RMSE-P RMSE-P % values, suggesting reduced predictive accuracy under Earth-like atmospheric conditions or specific instrumental configurations.

These findings prompted an evaluation of optimal datasets for calibration transfer. While the SuperLIBS 10k Vacuum dataset's consistently low RMSE-P and RMSE-P % values make it a strong candidate, the SuperLIBS 10k Mars dataset provides more appropriate reference conditions given that the unseen test data (ChemCam and SuperCam) are from instruments calibrated to be in Mars-like environments. Hence, the calibration transfer is from ChemCam/SuperCam to SuperLIBS 10k Mars when running PDS-PLS.

5.2 Calibration Transfer: PDS-PLS

5.2.1 Calibration Transfer Cross-Validation

The PyHAT calibration transfer implementation imposed a key constraint: only subsets of unique samples common to both source datasets (SuperCam/ChemCam) and the target

dataset (SL 10k Mars) could be used for cross-validation. This limitation arose because although the datasets contained identical geological samples measured under different atmospheric conditions, they employed different primary keys (pkeys) for sample identification. To address this challenge, a systematic preprocessing approach was developed. First, common sample names between source and target datasets were identified through metadata matching. From these matched samples, subsets were created, consisting of approximately 500 rows of spectra (104 samples) from SuperCam and 630 rows of spectra (160 samples) from ChemCam, with each sample appearing only once in these subsets.

Following input of these subsets into PyHAT, cross-validation optimized the parameters of window size and number of components. The analysis revealed optimal performance for ChemCam at 2 components and window size 3 (average RMSE = 0.000499), while SuperCam showed different optimal parameters at 5 components and window size 9 (RMSE = 0.00058879).

5.2.2 Calibration Transfer Implementation

Using the optimized parameters from cross-validation, the calibration transfer was applied to the full datasets. The same filtered subsets of ChemCam and SuperCam (matched to the SL 10k Mars dataset) served as inputs, but with all available SuperCam and ChemCam data incorporated to transfer calibrations across the entire spectra. The resulting output file will subsequently be combined with the combined supermodel for evaluation under different scenarios.

5.3 ChemCam

The following tables and plots reveal the results based on each performance metric.

Table 5.13: Average RMSE-P (wt%) across all oxides for ChemCam scenarios

Scenario	Description	Avg RMSE-P
Scenario 0	Train on All Test on All	1.67
	Train on CC test on CC	2.02
	Train on CC CT test on CC CT	1.91
Scenario 1	Train on All test on CC	3.47
	Train on All test on CC CT	3.32
	Train on All+CC test on CC	2.28
	Train on All+CC CT test on CC CT	2.52
Scenario 2	Train on All test on SL 10k Mars	1.57
	Train on CC test on SL 10k Mars	4.64
	Train on CC CT test on SL 10k Mars	2.48
	Train on All+CC test on SL 10k Mars	1.63
	Train on All+CC CT test on SL 10k Mars	1.57
Scenario 3	Train on CC test on Mars Cal Targ	2.24
	Train on All+CC test on Mars Cal Targ	2.17
	Train on CC+Cal Targ test on Mars Cal Targ	1.16
	Train on All+CC+Cal Targ test on Mars Cal Targ	1.45

Table 5.14: Average RMSE-P over median concentration percentage (%) across all oxides for ChemCam scenarios

Scenario	Description	Avg RMSE-P Median (%)
Scenario 0	Train on All Test on All	27.27
	Train on CC test on CC	37.05
	Train on CC CT test on CC CT	36.83
Scenario 1	Train on All test on CC	45.12
	Train on All test on CC CT	51.87
	Train on All+CC test on CC	35.85
	Train on All+CC CT test on CC CT	34.36
Scenario 2	Train on All test on SL 10k Mars	25.00
	Train on CC test on SL 10k Mars	82.39
	Train on CC CT test on SL 10k Mars	57.96
	Train on All+CC test on SL 10k Mars	27.24
	Train on All+CC CT test on SL 10k Mars	25.53
Scenario 3	Train on CC test on Mars Cal Targ	49.60
	Train on All+CC test on Mars Cal Targ	34.73
	Train on CC+Cal Targ test on Mars Cal Targ	27.88
	Train on All+CC+Cal Targ test on Mars Cal Targ	18.26

Table 5.15: Average R^2 across all oxides for ChemCam scenarios

Scenario	Description	Avg R^2
Scenario 0	Train on All Test on All	0.83
	Train on CC test on CC	0.81
	Train on CC CT test on CC CT	0.82
Scenario 1	Train on All test on CC	0.54
	Train on All test on CC CT	0.65
	Train on All+CC test on CC	0.73
	Train on All+CC CT test on CC CT	0.70
Scenario 2	Train on All test on SL 10k Mars	0.86
	Train on CC test on SL 10k Mars	0.27
	Train on CC CT test on SL 10k Mars	0.59
	Train on All+CC test on SL 10k Mars	0.84
	Train on All+CC CT test on SL 10k Mars	0.85
Scenario 3	Train on CC test on Mars Cal Targ	0.05
	Train on All+CC test on Mars Cal Targ	0.07
	Train on CC+Cal Targ test on Mars Cal Targ	0.32
	Train on All+CC+Cal Targ test on Mars Cal Targ	0.23

Table 5.16: Average Slope across all oxides for ChemCam scenarios

Scenario	Description	Avg Slope
Scenario 0	Train on All Test on All	0.83
	Train on CC test on CC	0.84
	Train on CC CT test on CC CT	0.84
Scenario 1	Train on All test on CC	0.74
	Train on All test on CC CT	0.74
	Train on All+CC test on CC	0.75
	Train on All+CC CT test on CC CT	0.81
Scenario 2	Train on All test on SL 10k Mars	0.85
	Train on CC test on SL 10k Mars	0.83
	Train on CC CT test on SL 10k Mars	0.75
	Train on All+CC test on SL 10k Mars	0.84
	Train on All+CC CT test on SL 10k Mars	0.86
Scenario 3	Train on CC test on Mars Cal Targ	0.86
	Train on All+CC test on Mars Cal Targ	0.39
	Train on CC+Cal Targ test on Mars Cal Targ	1.14
	Train on All+CC+Cal Targ test on Mars Cal Targ	0.42

Table 5.17: Average Normalized Intercept across all oxides for ChemCam scenarios

Scenario	Description	Avg Norm Intercept
Scenario 0	Train on All Test on All	0.17
	Train on CC test on CC	0.17
	Train on CC CT test on CC CT	0.16
Scenario 1	Train on All test on CC	0.25
	Train on All test on CC CT	0.27
	Train on All+CC test on CC	0.27
	Train on All+CC CT test on CC CT	0.21
Scenario 2	Train on All test on SL 10k Mars	0.15
	Train on CC test on SL 10k Mars	0.45
	Train on CC CT test on SL 10k Mars	0.32
	Train on All+CC test on SL 10k Mars	0.16
	Train on All+CC CT test on SL 10k Mars	0.14
Scenario 3	Train on CC test on Mars Cal Targ	1.25
	Train on All+CC test on Mars Cal Targ	1.41
	Train on CC+Cal Targ test on Mars Cal Targ	-0.48
	Train on All+CC+Cal Targ test on Mars Cal Targ	0.88

5.3.1 ChemCam: Scenario 0

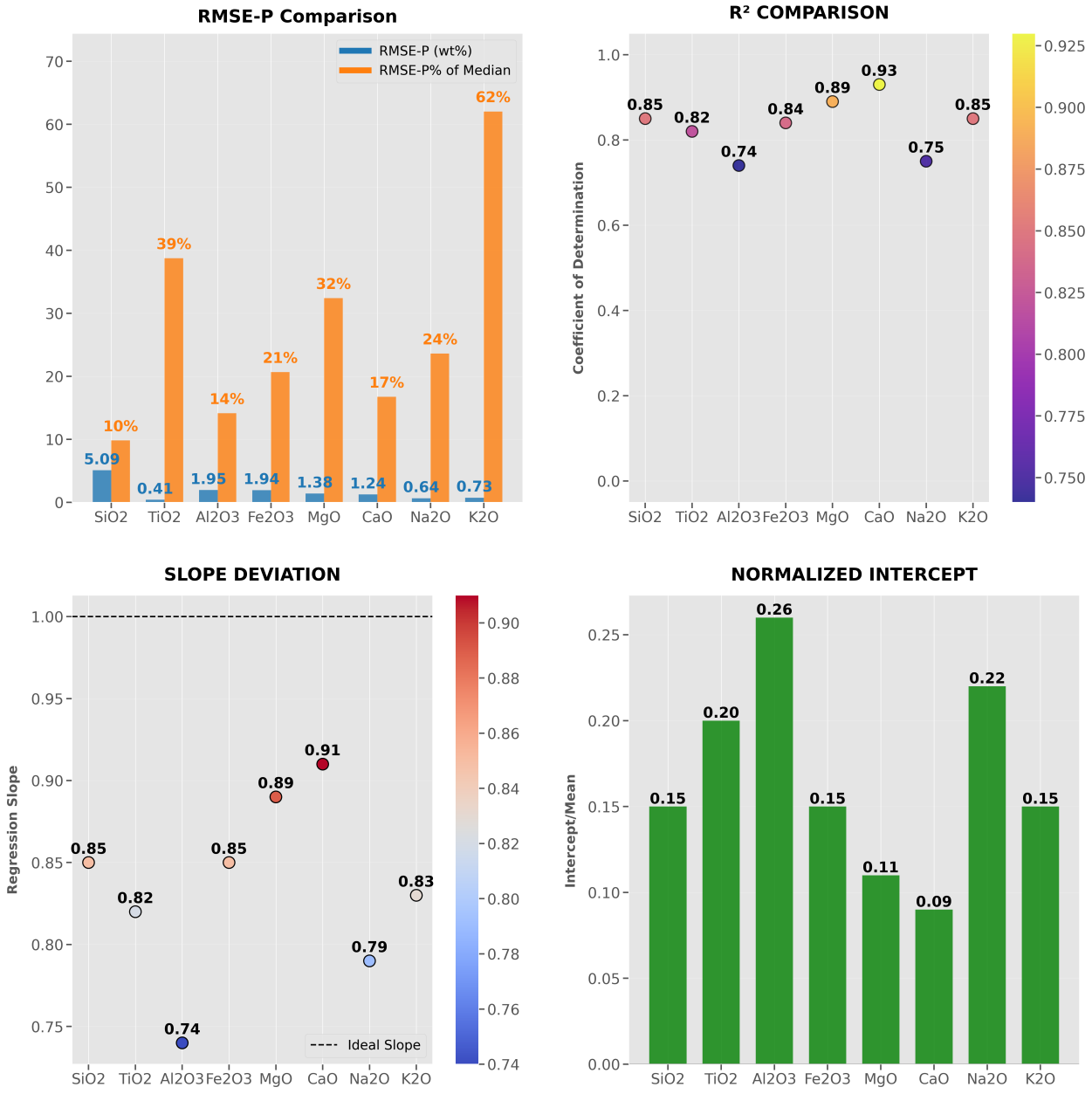

Figure 5.1: ChemCam Scenario 0 Plots: Train on All, test on All

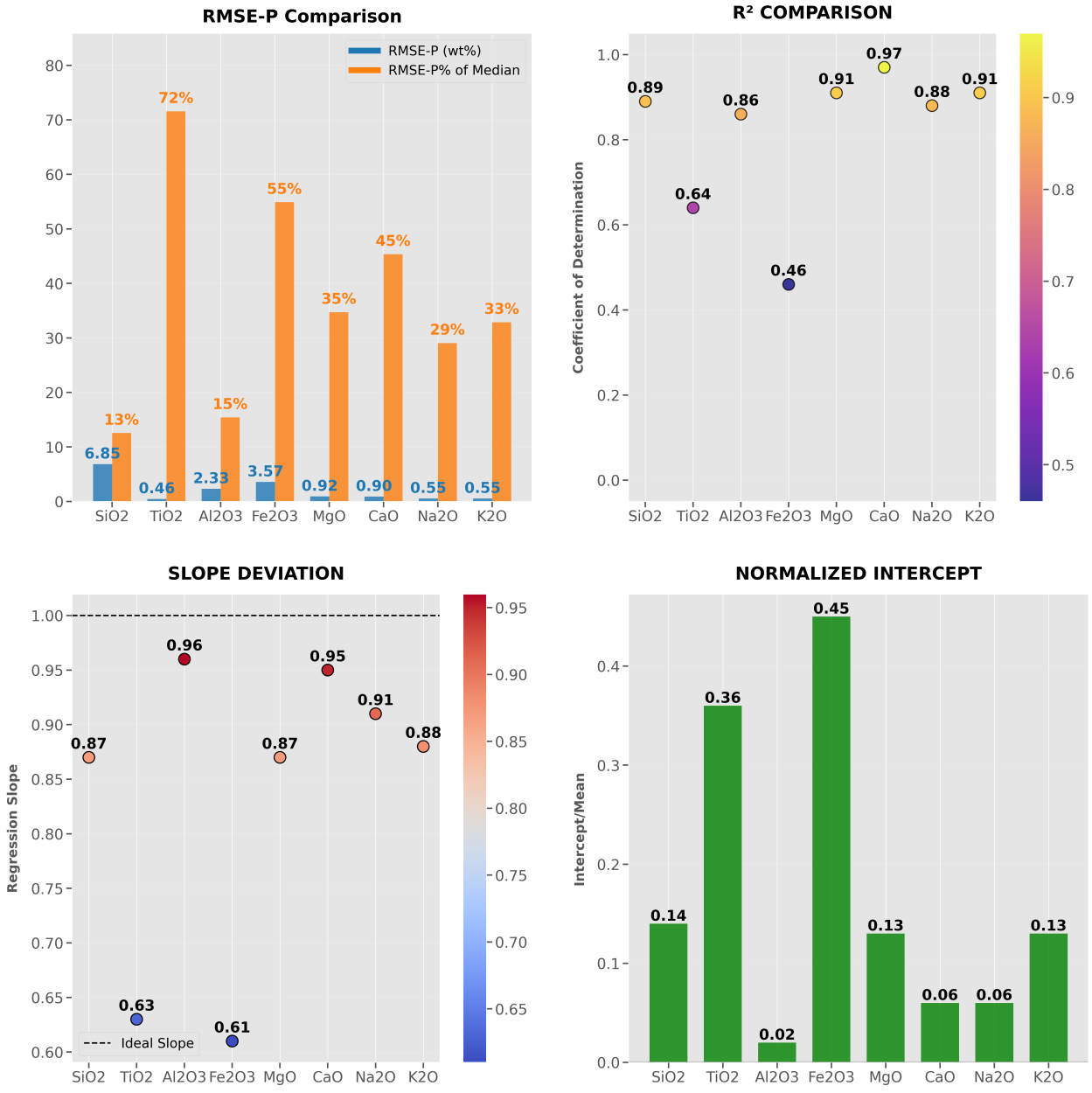

Figure 5.2: ChemCam Scenario 0 Plots: Train on CC, test on CC

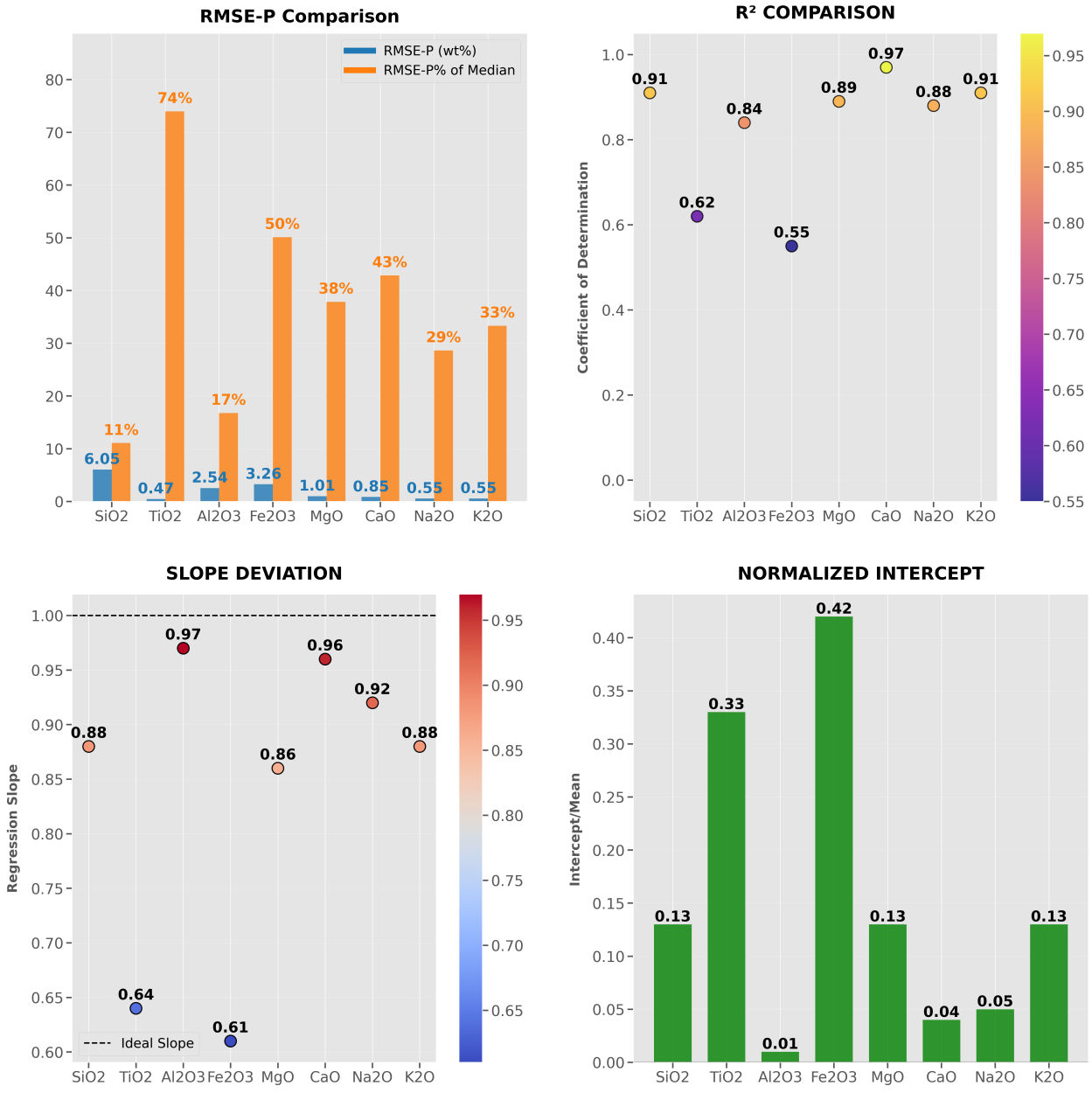

Figure 5.3: ChemCam Scenario 0 Plots: Train on CC CT, test on CC CT

5.3.2 ChemCam: Scenario 1

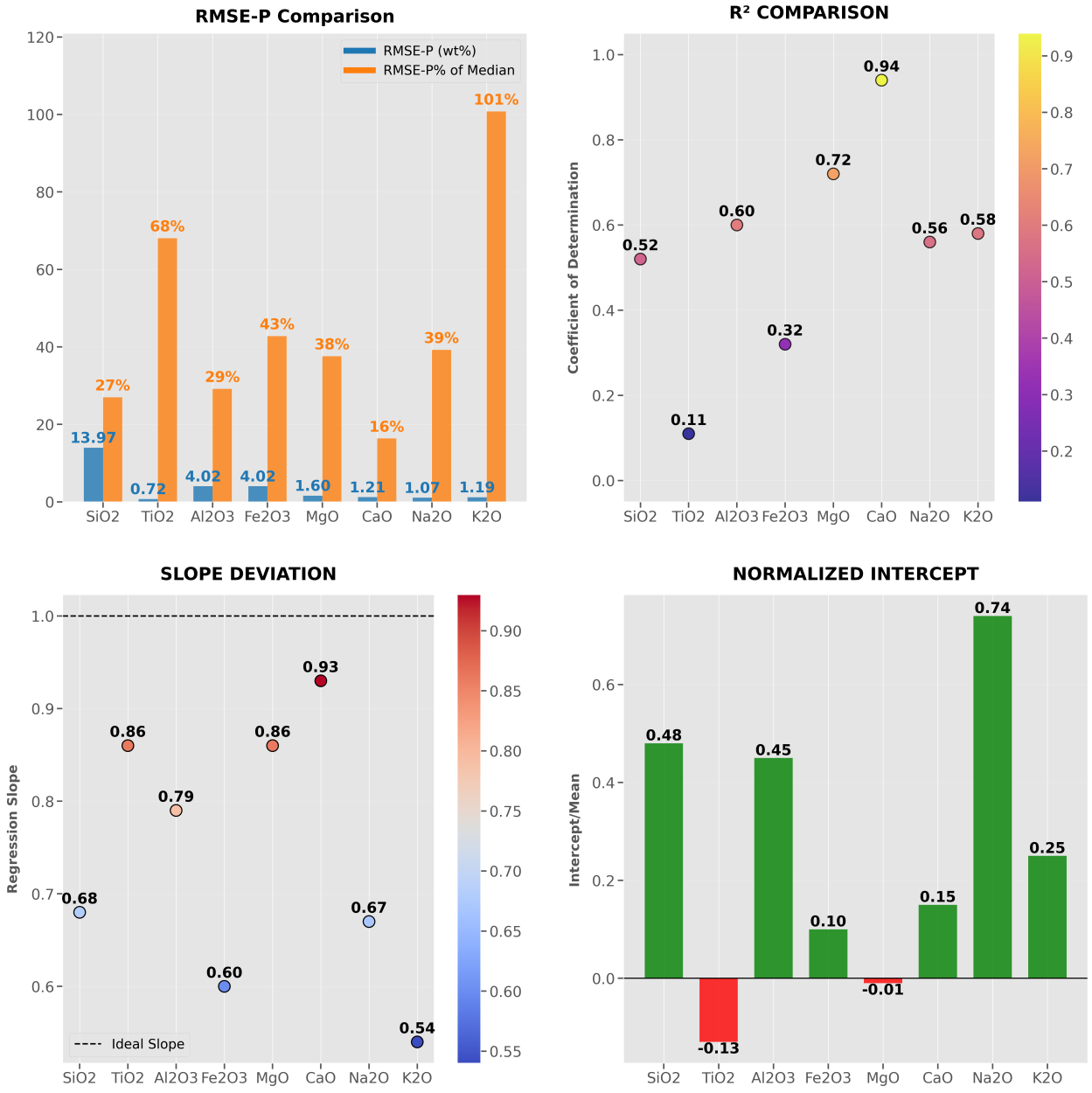

Figure 5.4: ChemCam Scenario 1 Plots: Train on All, test on CC

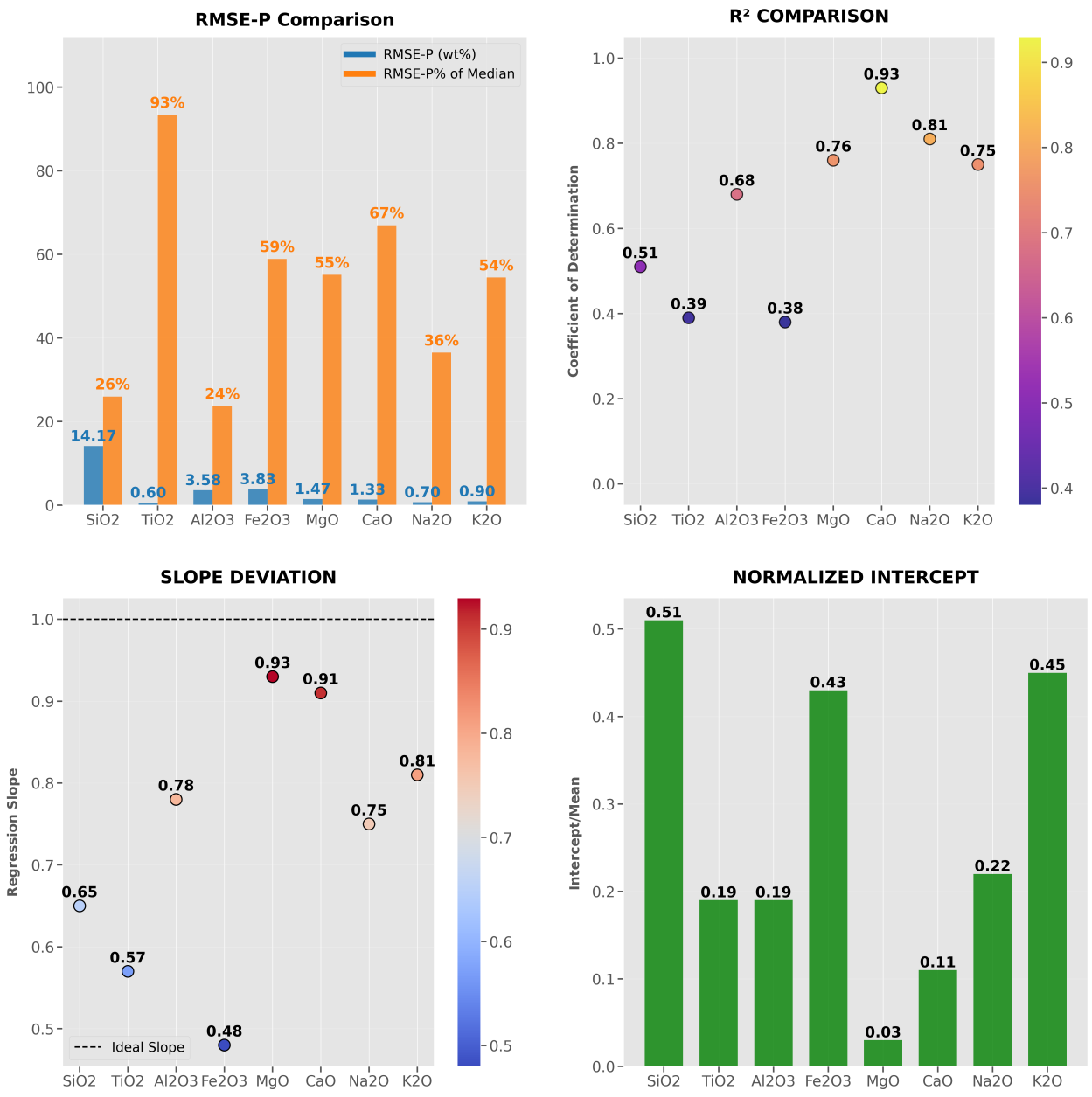

Figure 5.5: ChemCam Scenario 1 Plots: Train on All, Test on CC CT (PDS PLS)

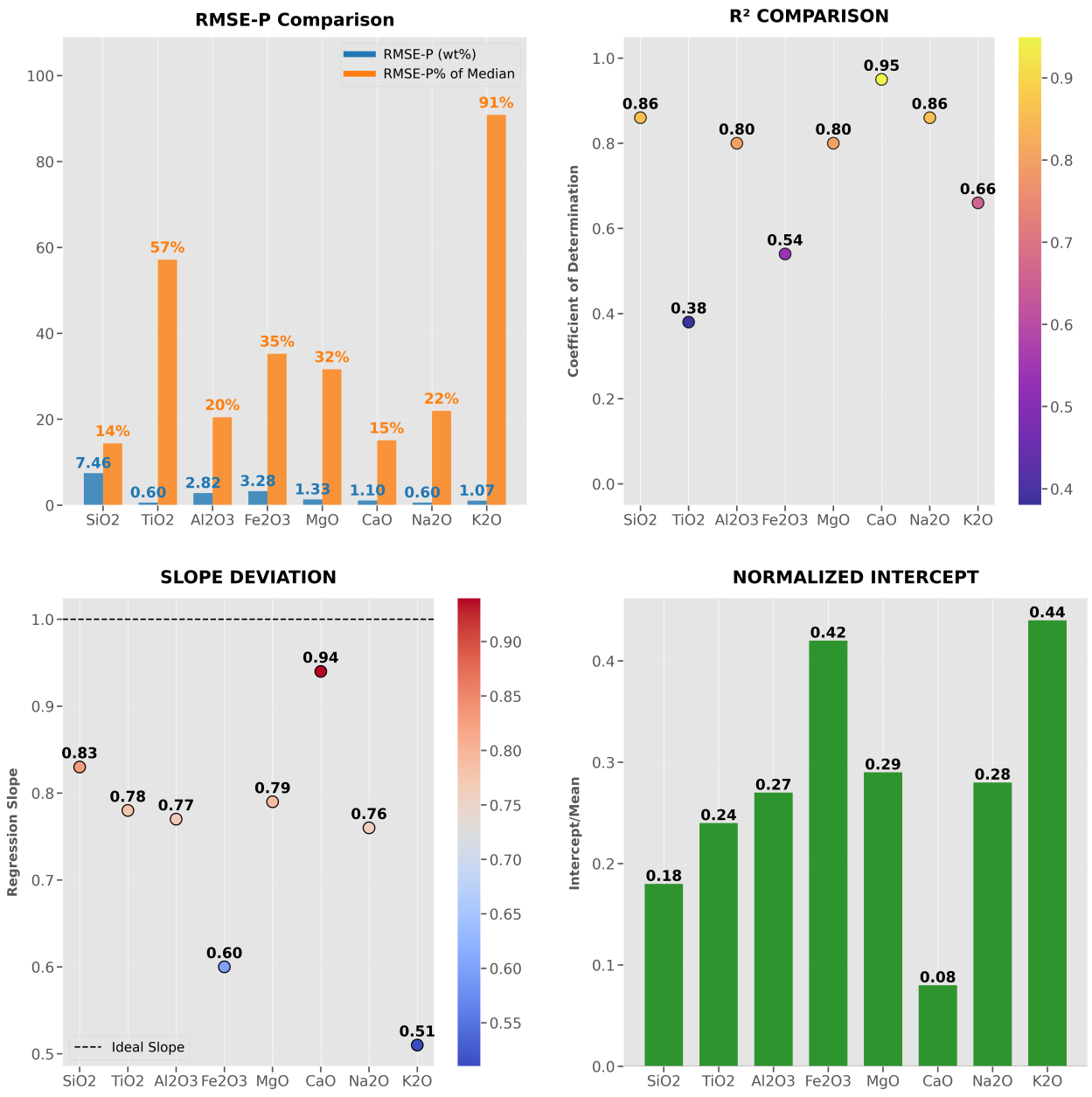

Figure 5.6: ChemCam Scenario 1 Plots: Train on All+CC, Test on CC

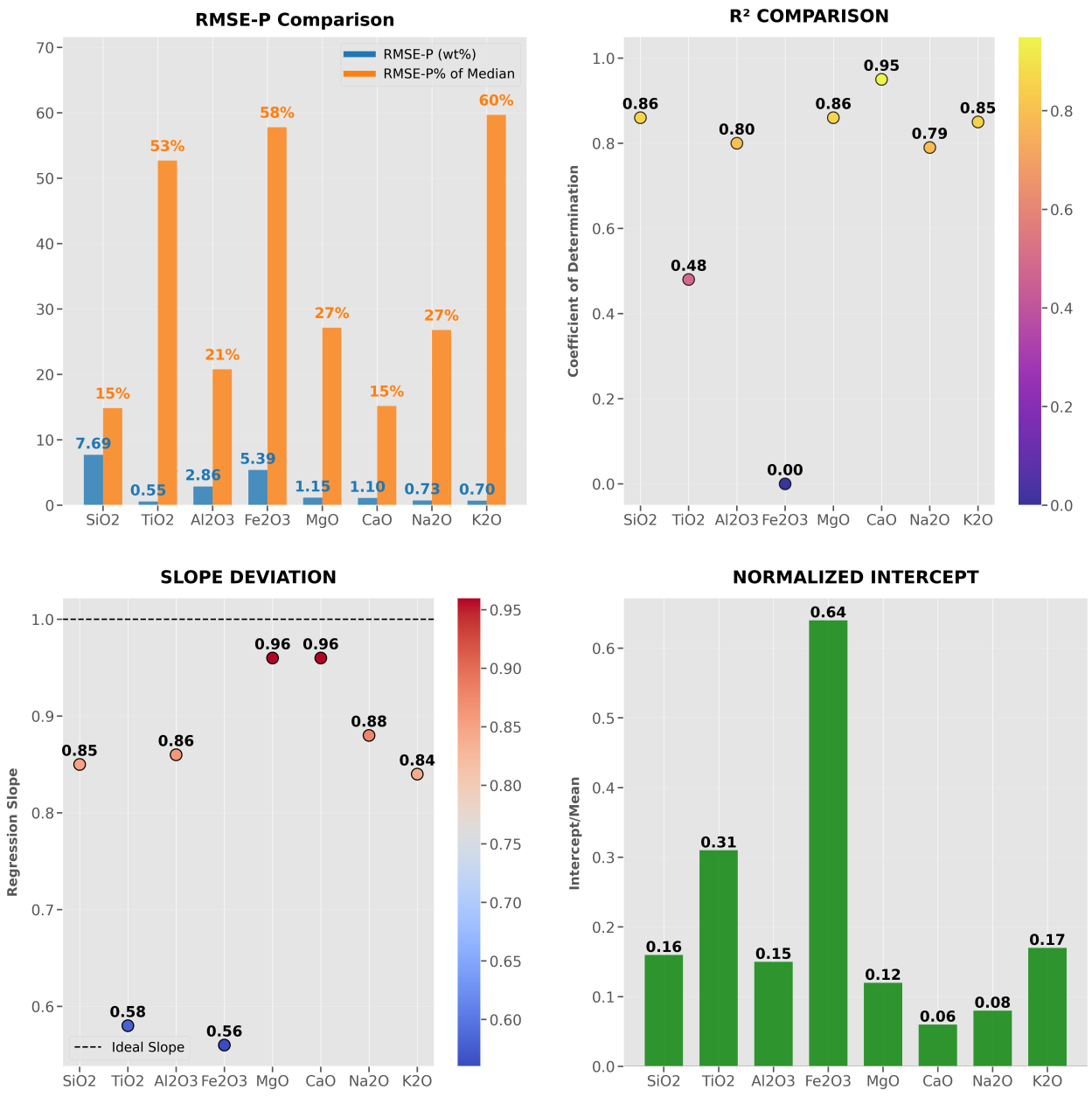

Figure 5.7: ChemCam Scenario 1 Plots: Train on All+CC CT, Test on CC CT (PDS PLS)

5.3.3 ChemCam: Scenario 2

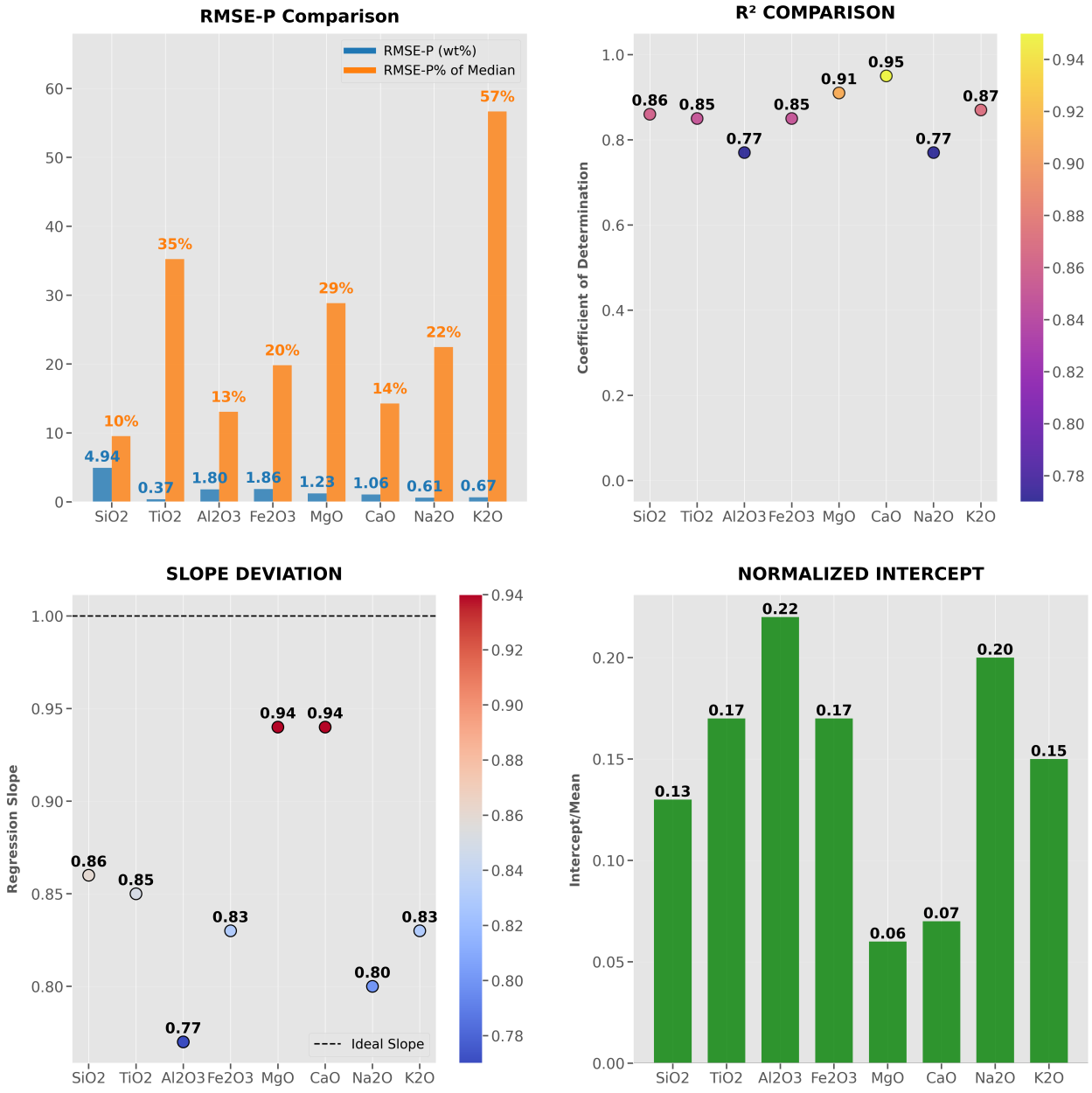

Figure 5.8: ChemCam Scenario 2 Plots: Train on All, Test on SL 10k Mars

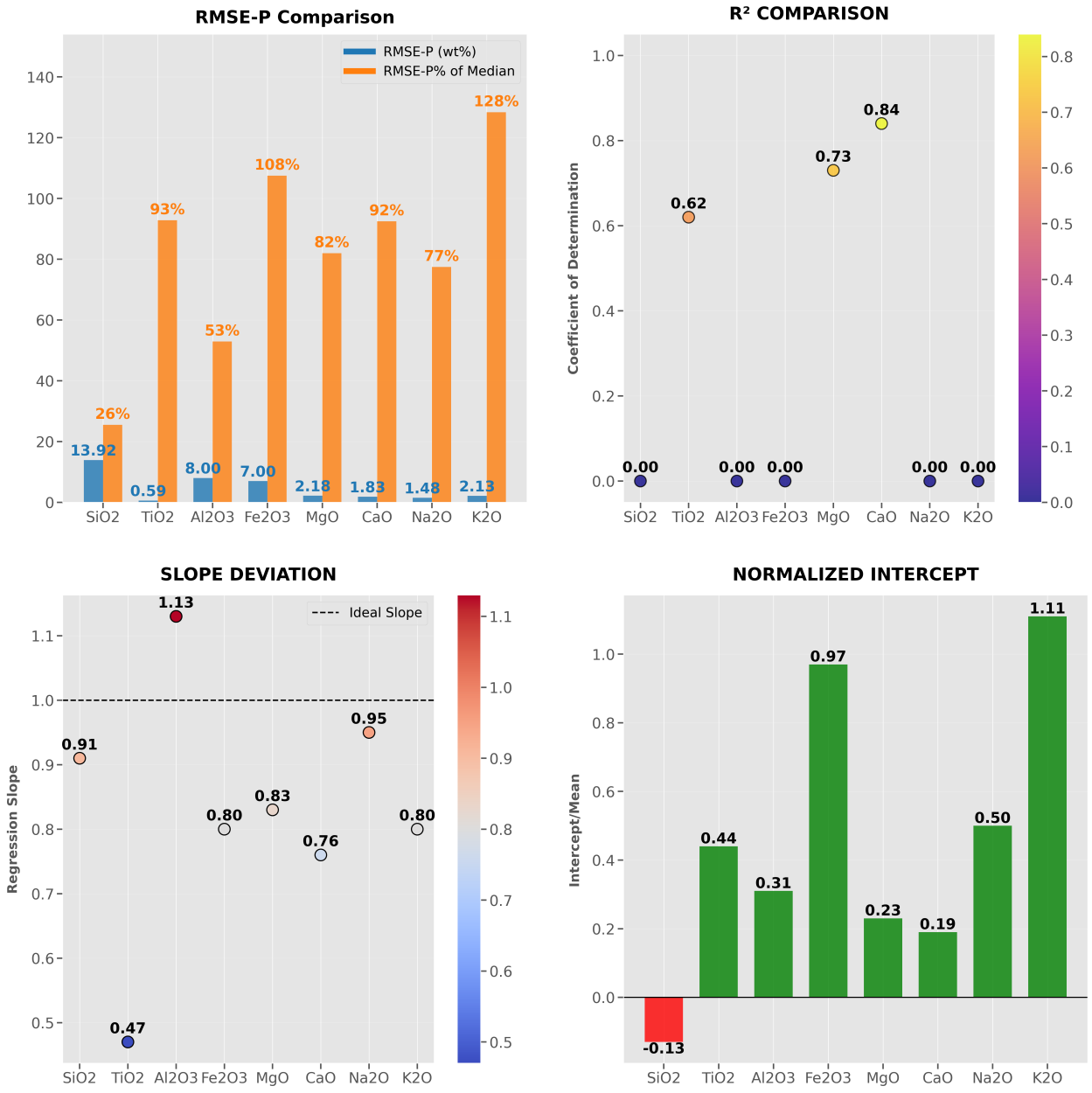

Figure 5.9: ChemCam Scenario 2 Plots: Train on CC, Test on SL 10k Mars

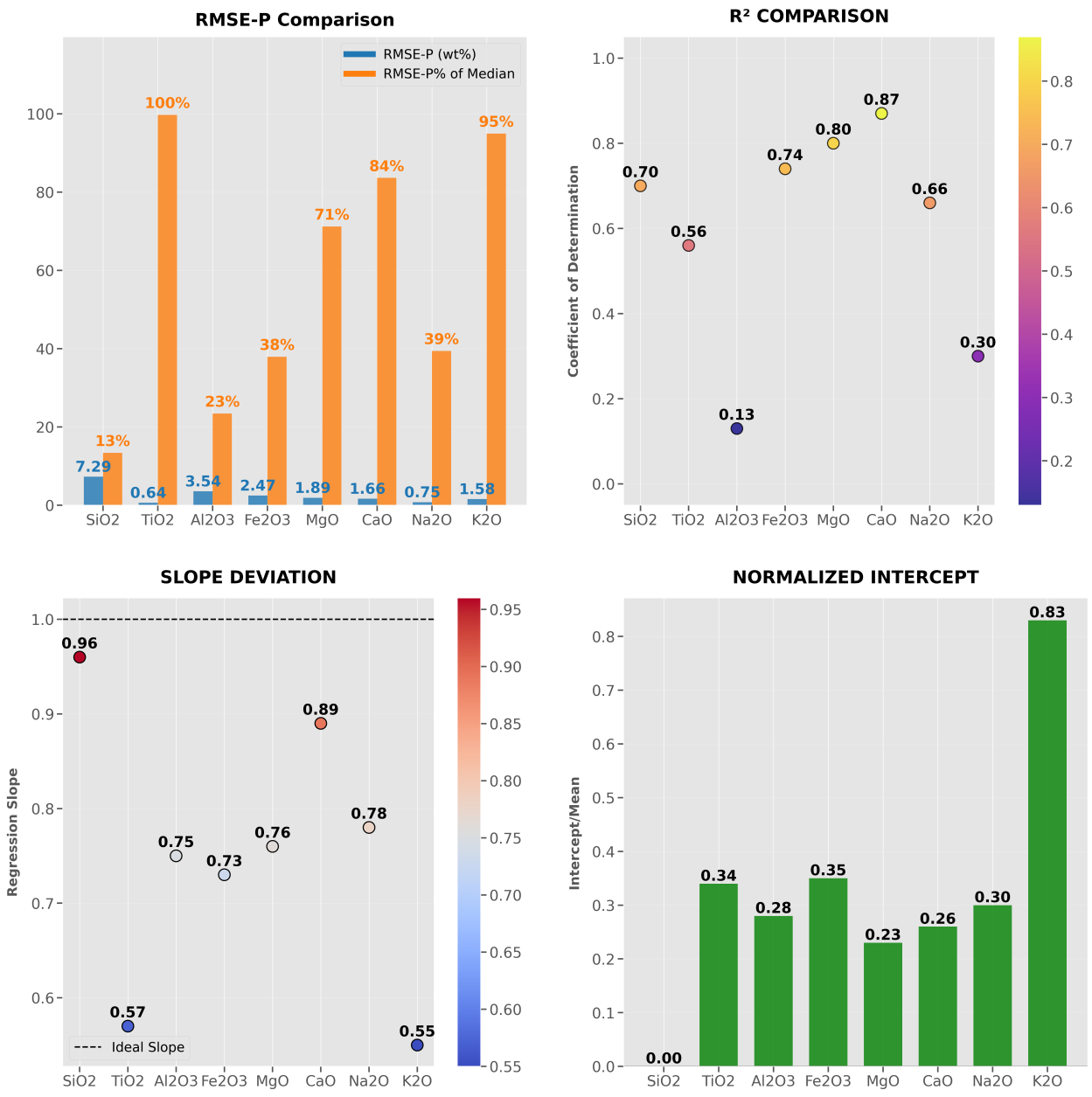

Figure 5.10: ChemCam Scenario 2 Plots: Train on CC CT, Test on SL 10k Mars

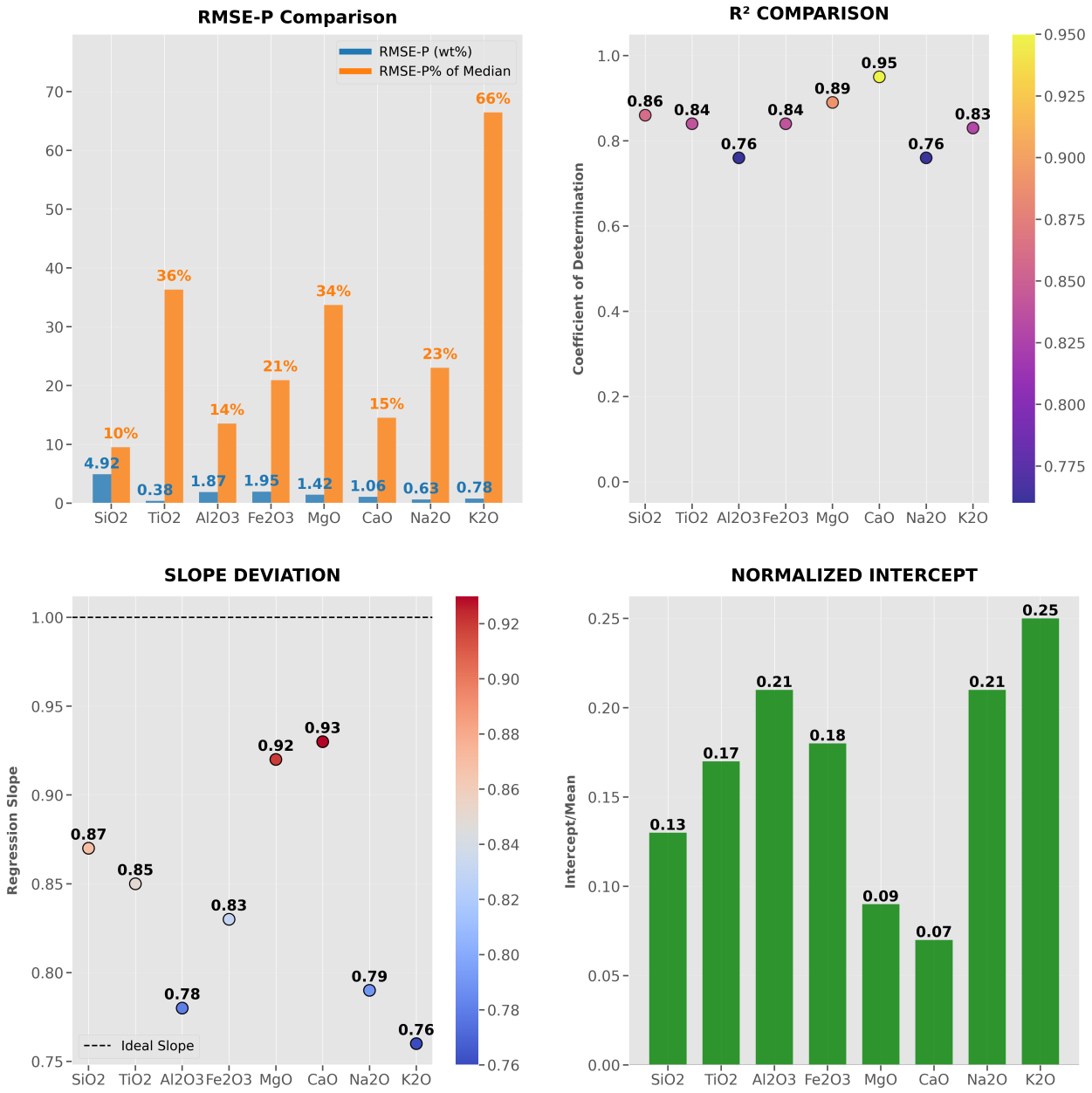

Figure 5.11: ChemCam Scenario 2 Plots: Train on All+CC, Test on SL 10k Mars

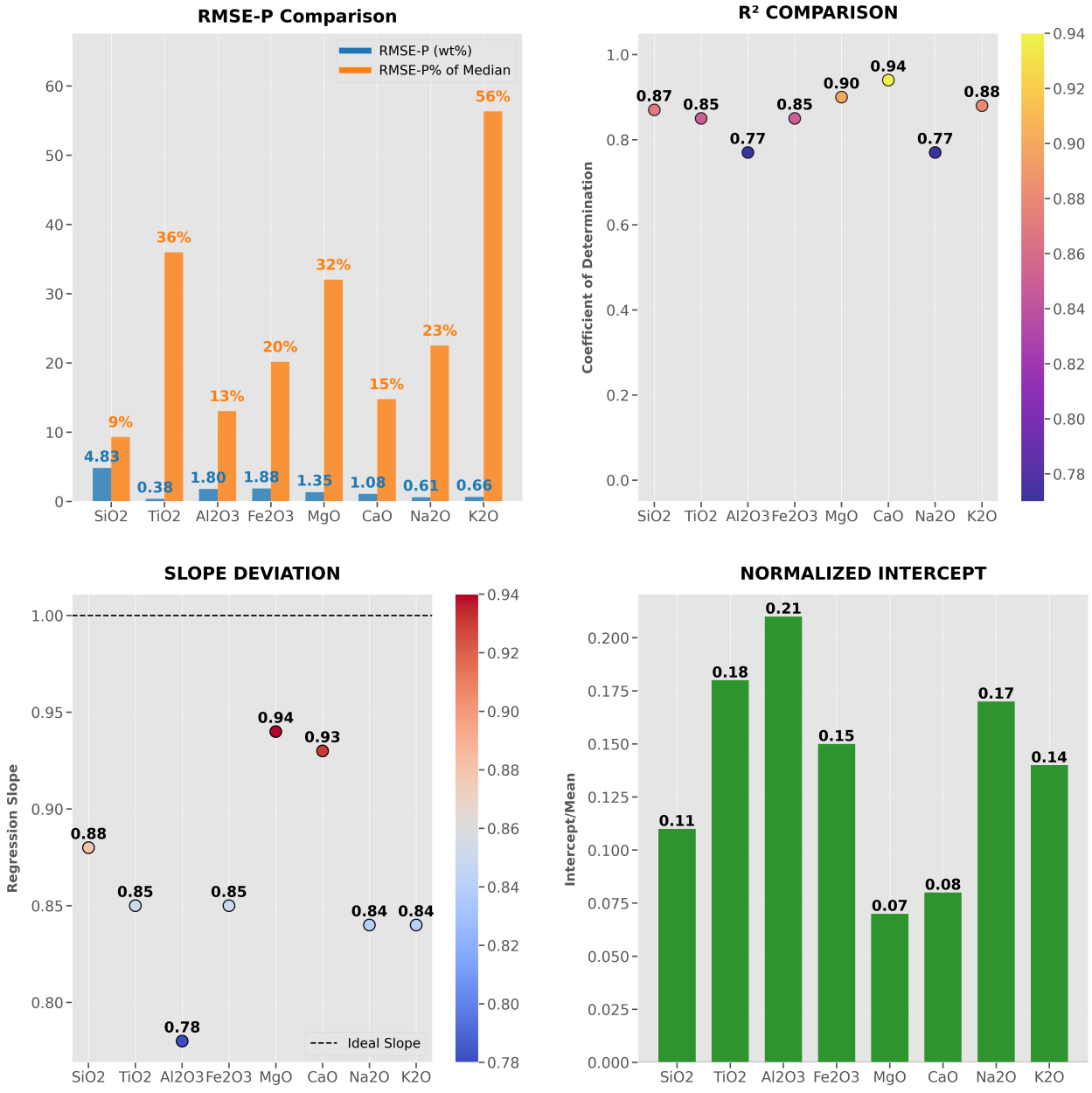

Figure 5.12: ChemCam Scenario 2 Plots: Train on All+CC CT (PDS PLS), Test on SL 10k Mars

5.3.4 ChemCam: Scenario 3

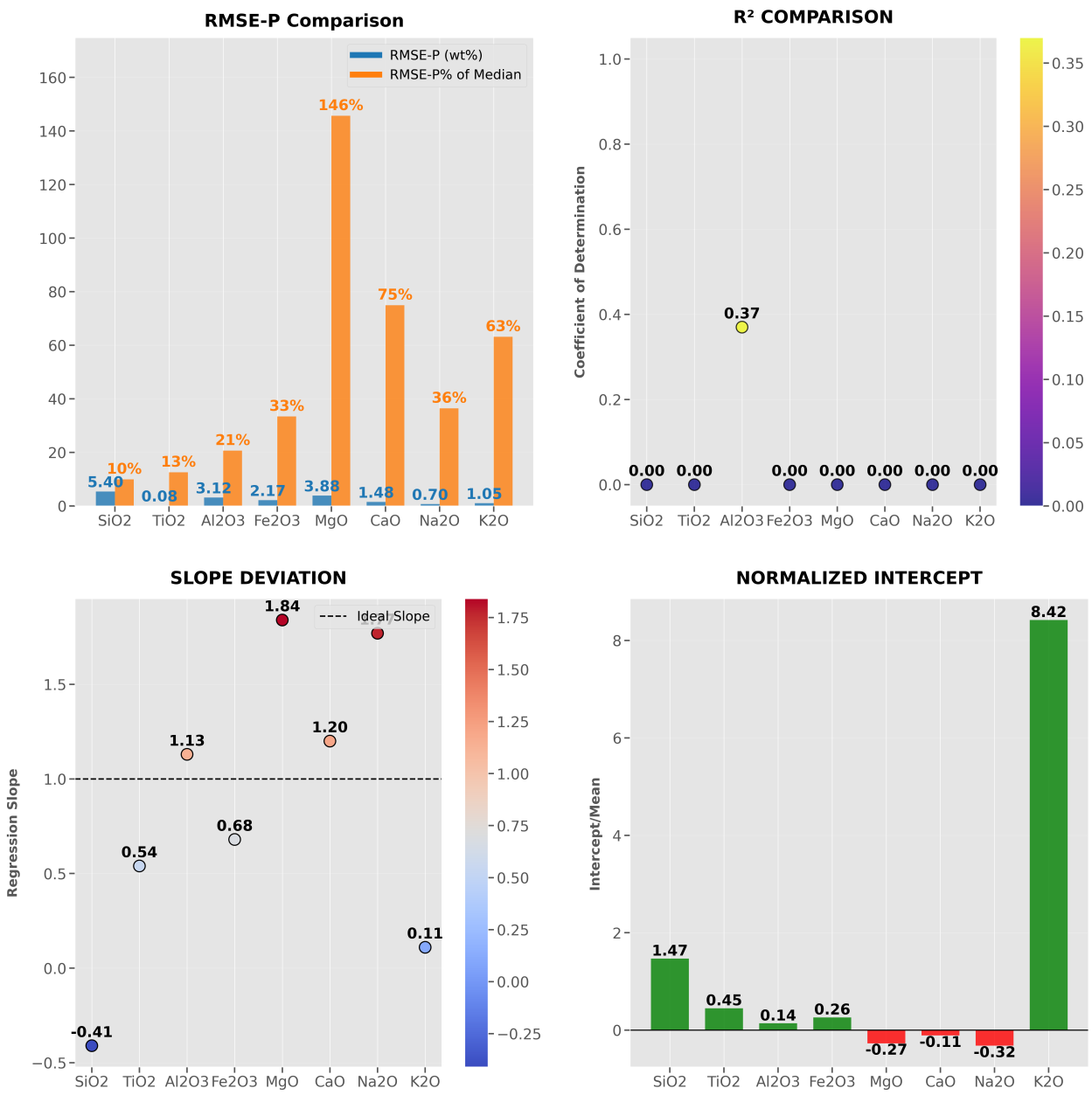

Figure 5.13: ChemCam Scenario 3 Plots: Train on CC, Test on ChemCam Mars Calibration Targets

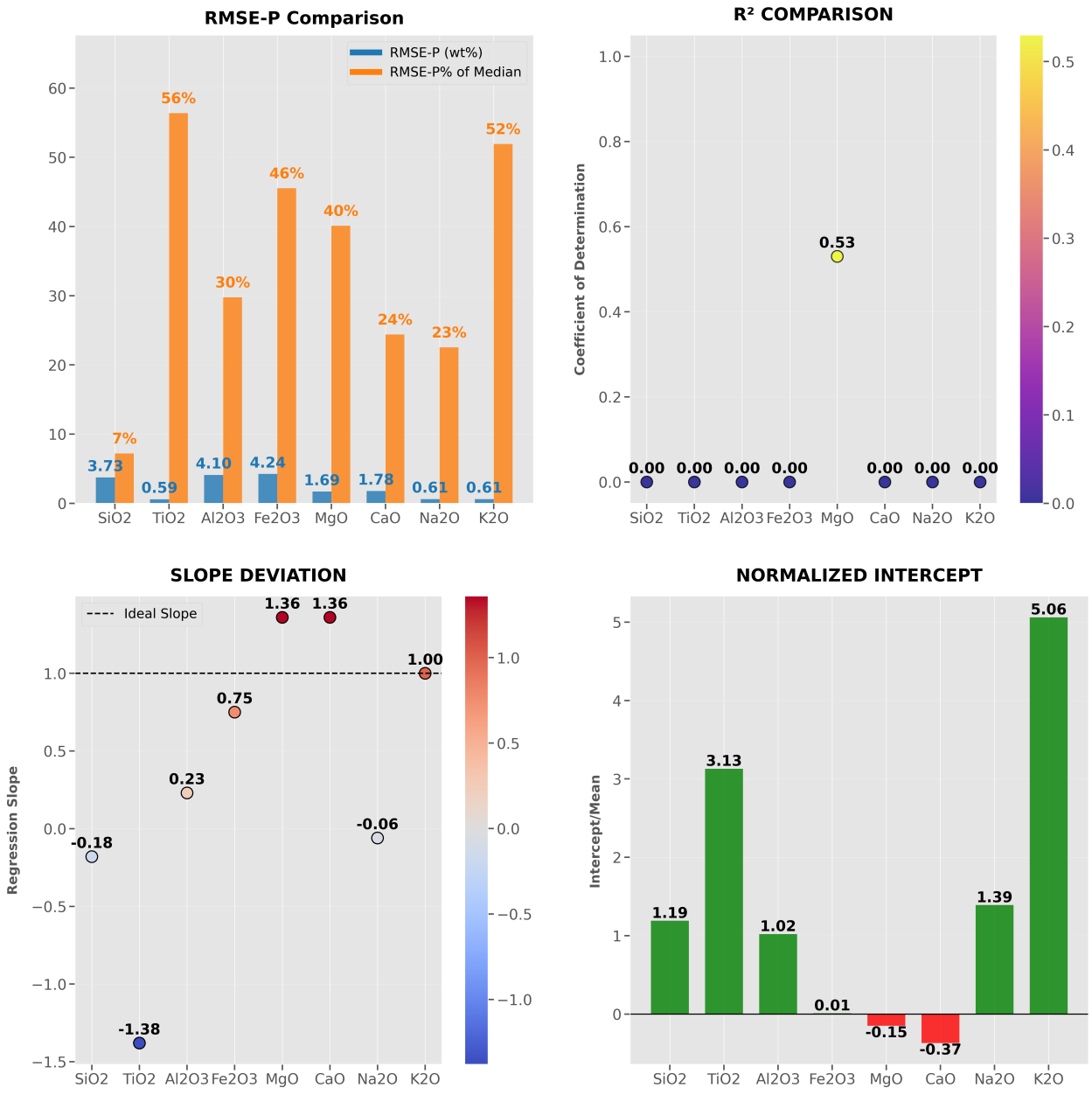

Figure 5.14: ChemCam Scenario 3 Plots: Train on All+CC, Test on ChemCam Mars Calibration Targets

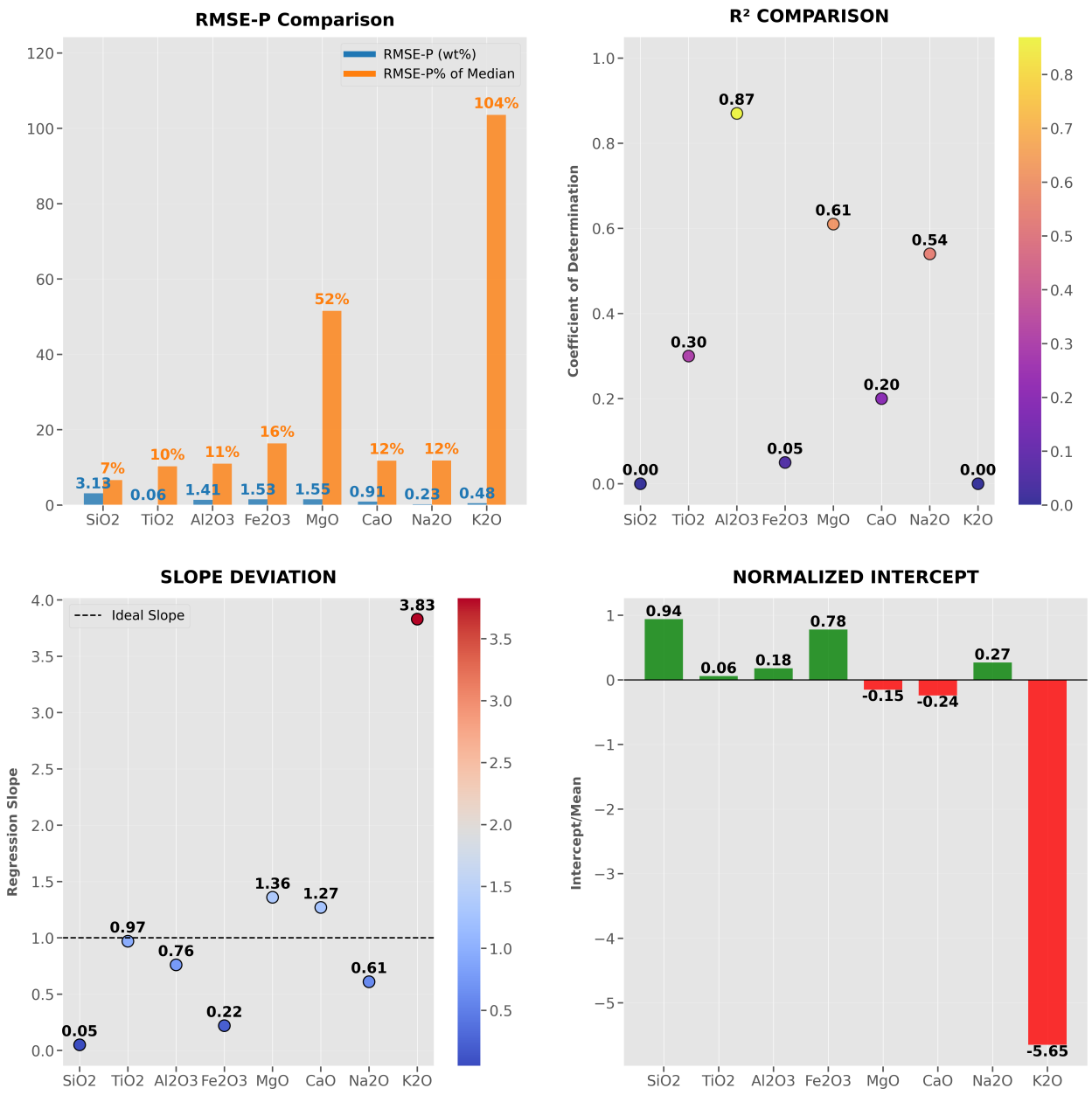

Figure 5.15: ChemCam Scenario 3 Plots: Train on CC+Calibration Targets, Test on ChemCam Mars Calibration Targets

5.4 SuperCam

The following tables and plots reveal the results based on each performance metric.

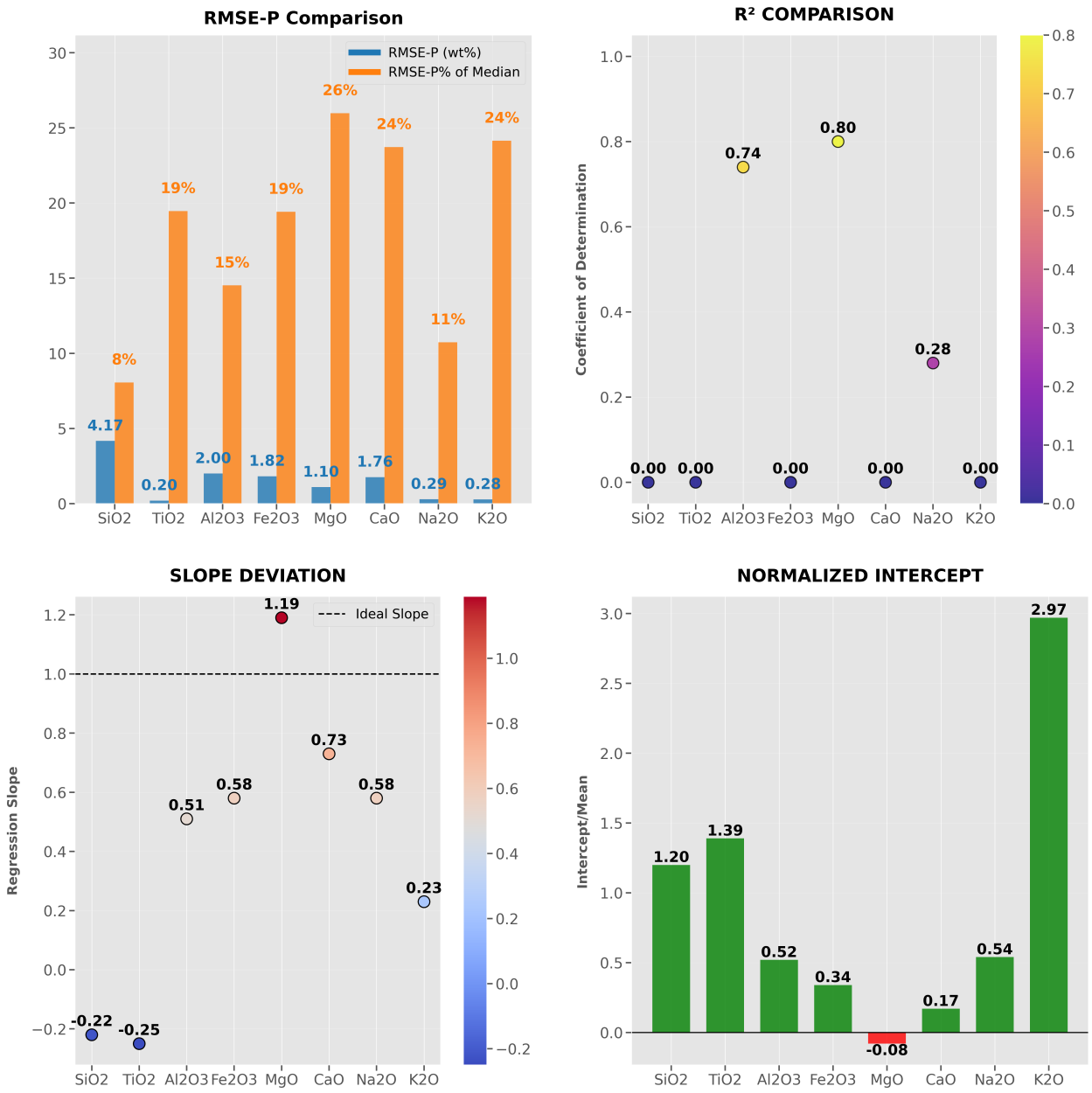

Figure 5.16: ChemCam Scenario 3 Plots: Train on All+CC+Calibration Targets, Test on ChemCam Mars Calibration Targets

Table 5.18: Average RMSE-P across all oxides for SuperCam scenarios

Scenario	Description	Avg RMSE-P
Scenario 0	Train on All Test on All	1.67
	Train on SC test on SC	2.42
	Train on SC CT test on SC CT	2.67
Scenario 1	Train on All test on SC	5.62
	Train on All test on SC CT	3.92
	Train on All+SC test on SC	2.34
	Train on All+SC CT test on SC CT	2.78
Scenario 2	Train on All test on SL 10k Mars	1.57
	Train on SC test on SL 10k Mars	7.33
	Train on SC CT test on SL 10k Mars	2.85
	Train on All+SC test on SL 10k Mars	1.64
	Train on All+SC CT test on SL 10k Mars	1.70
Scenario 3	Train on SC test on Mars Cal Targ	5.48
	Train on All+SC test on Mars Cal Targ	4.61
	Train on SC+Cal Targ test on Mars Cal Targ	6.11
	Train on All+SC+Cal Targ test on Mars Cal Targ	3.89
	Train on SL 10k/18k Mars+SC+Cal Targ test on Mars Cal Targ	3.82
Scenario 4	Train on SL 10k/18k Mars test on SC	7.53
	Train on SL 10k/18k Mars CT SC test on SC	4.68

Table 5.19: RMSE-P over median concentration percentage (%) for SuperCam scenarios

Scenario	Description	Avg RMSE-P Median (%)
Scenario 0	Train on All Test on All	27.27
	Train on SC test on SC	50.04
	Train on SC CT test on SC CT	54.24
Scenario 1	Train on All test on SC	71.12
	Train on All test on SC CT	59.10
	Train on All+SC test on SC	37.00
	Train on All+SC CT test on SC CT	42.51
Scenario 2	Train on All test on SL 10k Mars	25.00
	Train on SC test on SL 10k Mars	95.06
	Train on SC CT test on SL 10k Mars	68.07
	Train on All+SC test on SL 10k Mars	27.53
	Train on All+SC CT test on SL 10k Mars	29.47
Scenario 3	Train on SC test on Mars Cal Targ	104.58
	Train on All+SC test on Mars Cal Targ	71.37
	Train on SC+Cal Targ test on Mars Cal Targ	70.88
	Train on All+SC+Cal Targ test on Mars Cal Targ	70.06
	Train on SL 10k/18k Mars+SC+Cal Targ test on Mars Cal Targ	58.16
Scenario 4	Train on SL 10k/18k Mars test on SC	100.5
	Train on SL 10k/18k Mars CT SC test on SC	60.28

Table 5.20: Average R^2 across all oxides for SuperCam scenarios

Scenario	Description	Avg R^2
Scenario 0	Train on All Test on All	0.83
	Train on SC test on SC	0.74
	Train on SC CT test on SC CT	0.68
Scenario 1	Train on All test on SC	0.28
	Train on All test on SC CT	0.38
	Train on All+SC test on SC	0.72
	Train on All+SC CT test on SC CT	0.66
Scenario 2	Train on All test on SL 10k Mars	0.86
	Train on SC test on SL 10k Mars	0.22
	Train on SC CT test on SL 10k Mars	0.49
	Train on All+SC test on SL 10k Mars	0.84
	Train on All+SC CT test on SL 10k Mars	0.82
Scenario 3	Train on SC test on Mars Cal Targ	0.36
	Train on All+SC test on Mars Cal Targ	0.41
	Train on SC+Cal Targ Mars test on Cal Targ	0.28
	Train on All+SC+Cal Targ Mars test on Cal Targ	0.49
	Train on SL 10k/18k Mars+SC+Cal Targ test on Mars Cal Targ	0.46
Scenario 4	Train on SL 10k/18k Mars test on SC	0.09
	Train on SL 10k/18k Mars CT SC test on SC	0.40

Table 5.21: Average Slope across all oxides for SuperCam scenarios

Scenario	Description	Avg Slope
Scenario 0	Train on All Test on All	0.83
	Train on SC test on SC	0.78
	Train on SC CT test on SC CT	0.69
Scenario 1	Train on All test on SC	0.57
	Train on All test on SC CT	0.68
	Train on All+SC test on SC	0.70
	Train on All+SC CT test on SC CT	0.64
Scenario 2	Train on All test on SL 10k Mars	0.85
	Train on SC test on SL 10k Mars	0.40
	Train on SC CT test on SL 10k Mars	0.65
	Train on All+SC test on SL 10k Mars	0.82
	Train on All+SC CT test on SL 10k Mars	0.81
Scenario 3	Train on SC test on Mars Cal Targ	0.57
	Train on All+SC test on Mars Cal Targ	0.71
	Train on SC+Cal Targ test on Mars Cal Targ	0.54
	Train on All+SC+Cal Targ test on Mars Cal Targ	0.85
	Train on SL 10k/18k Mars+SC+Cal Targ test on Mars Cal Targ	0.80
Scenario 4	Train on SL 10k/18k Mars test on SC	0.80
	Train on SL 10k/18k Mars CT SC test on SC	0.89

Table 5.22: Average Normalized Intercept across all oxides for SuperCam scenarios

Scenario	Description	Avg Norm Intercept
Scenario 0	Train on All Test on All	0.17
	Train on SC test on SC	0.21
	Train on SC CT test on SC CT	0.26
Scenario 1	Train on All test on SC	0.51
	Train on All test on SC CT	0.42
	Train on All+SC test on SC	0.27
	Train on All+SC CT test on SC CT	0.31
Scenario 2	Train on All test on SL 10k Mars	0.15
	Train on SC test on SL 10k Mars	0.57
	Train on SC CT test on SL 10k Mars	0.17
	Train on All+SC test on SL 10k Mars	0.18
	Train on All+SC CT test on SL 10k Mars	0.19
Scenario 3	Train on SC test on Mars Cal Targ	0.61
	Train on All+SC test on Mars Cal Targ	0.43
	Train on SC+Cal Targ test on Mars Cal Targ	0.75
	Train on All+SC+Cal Targ test on Mars Cal Targ	0.24
	Train on SL 10k/18k Mars+SC+Cal Targ test on Mars Cal Targ	0.42
Scenario 4	Train on SL 10k/18k Mars test on SC	0.34
	Train on SL 10k/18k Mars CT SC test on SC	0.01

5.4.1 SuperCam: Scenario 0

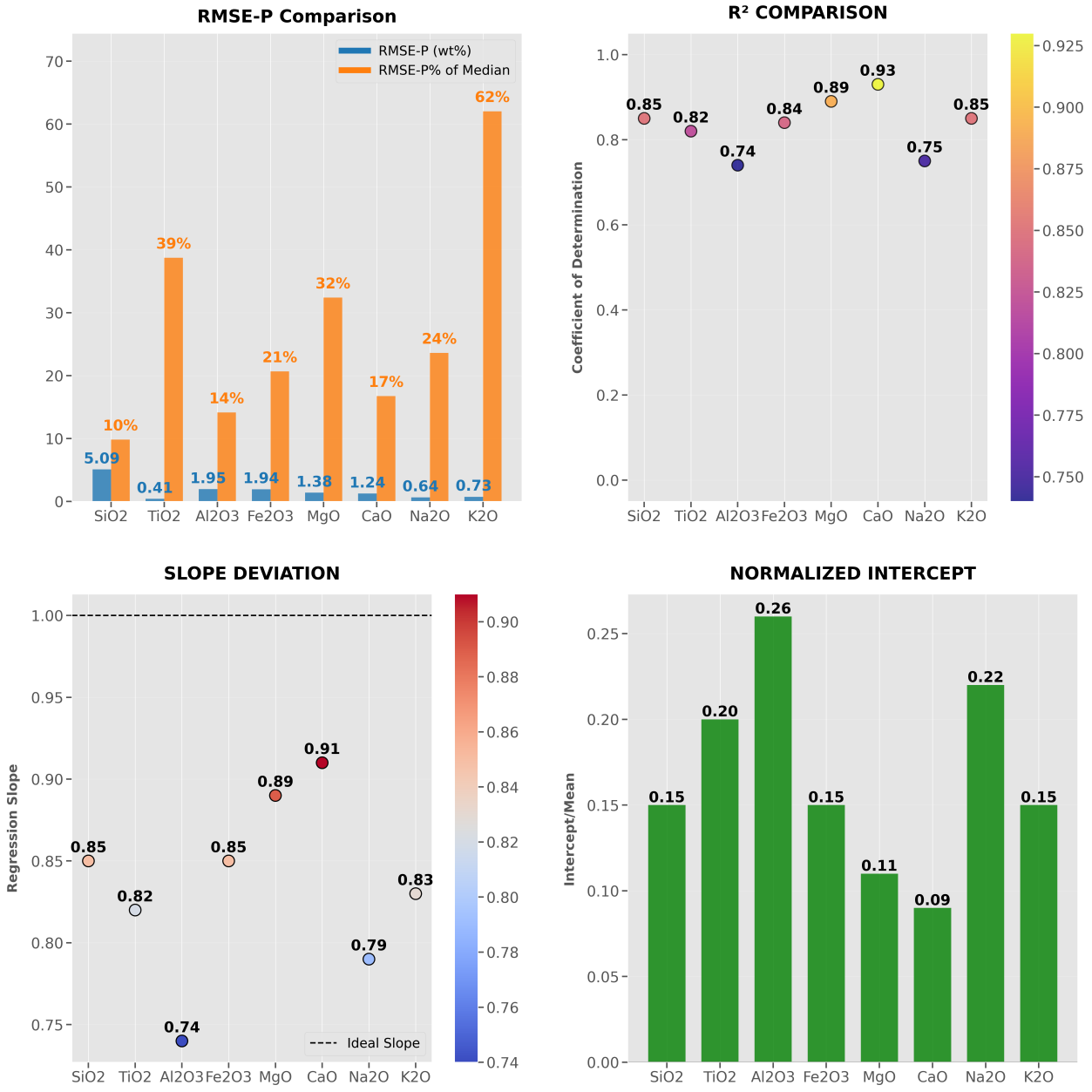

Figure 5.17: SuperCam Scenario 0 Plots: Train on All, test on All

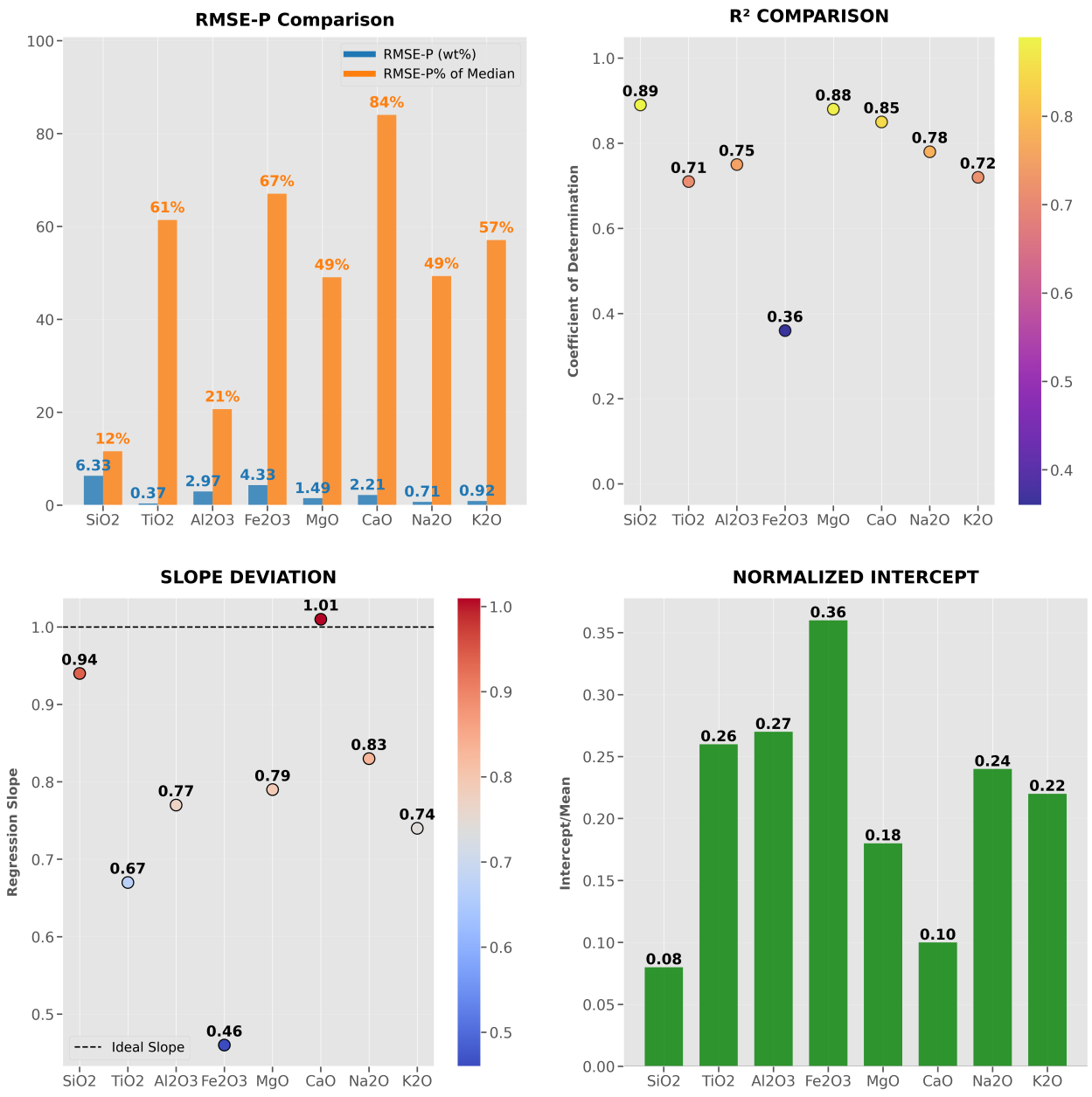

Figure 5.18: SuperCam Scenario 0 Plots: Train on SC, test on SC

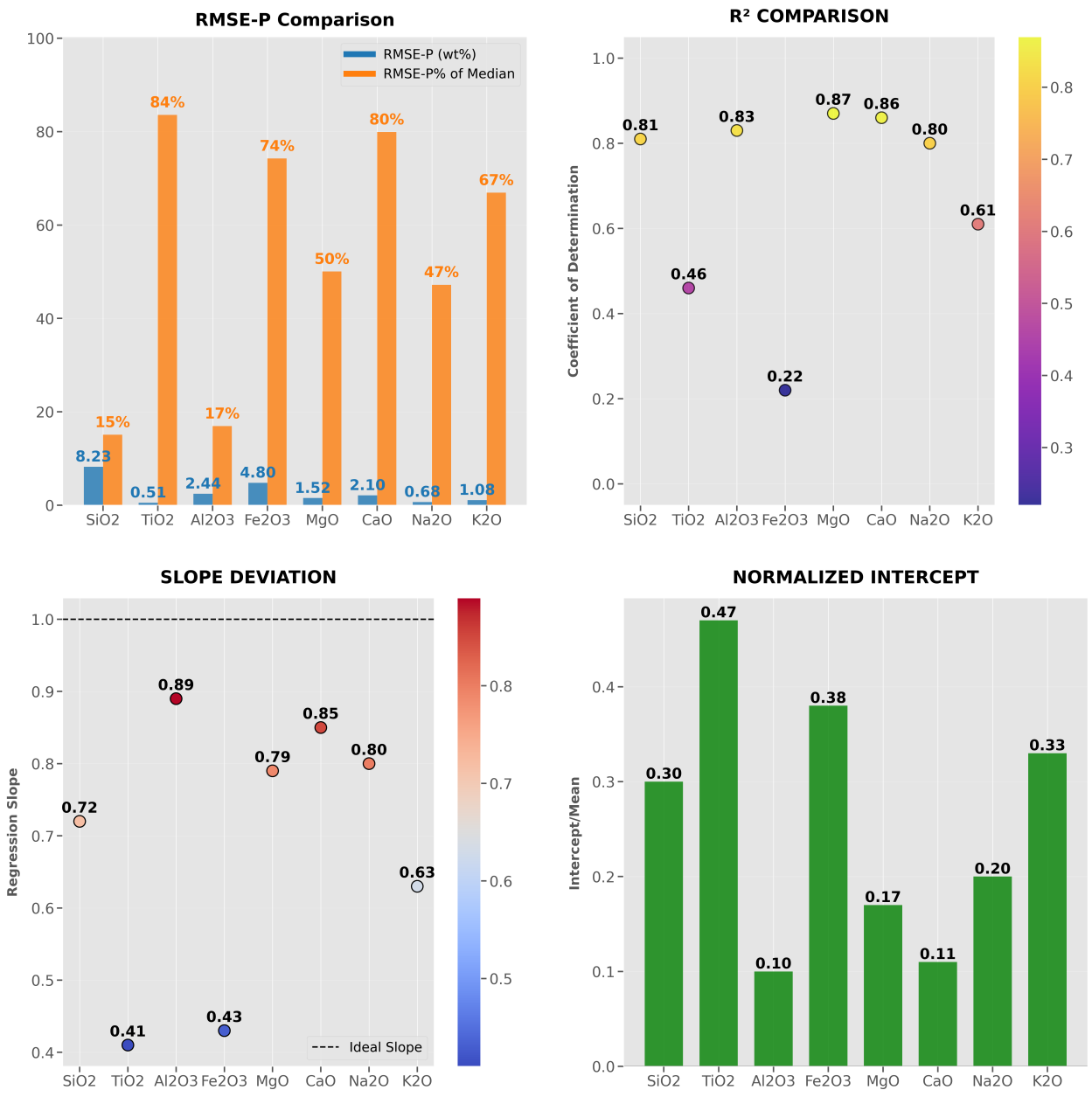

Figure 5.19: SuperCam Scenario 0 Plots: Train on SC CT, test on SC CT

5.4.2 SuperCam: Scenario 1

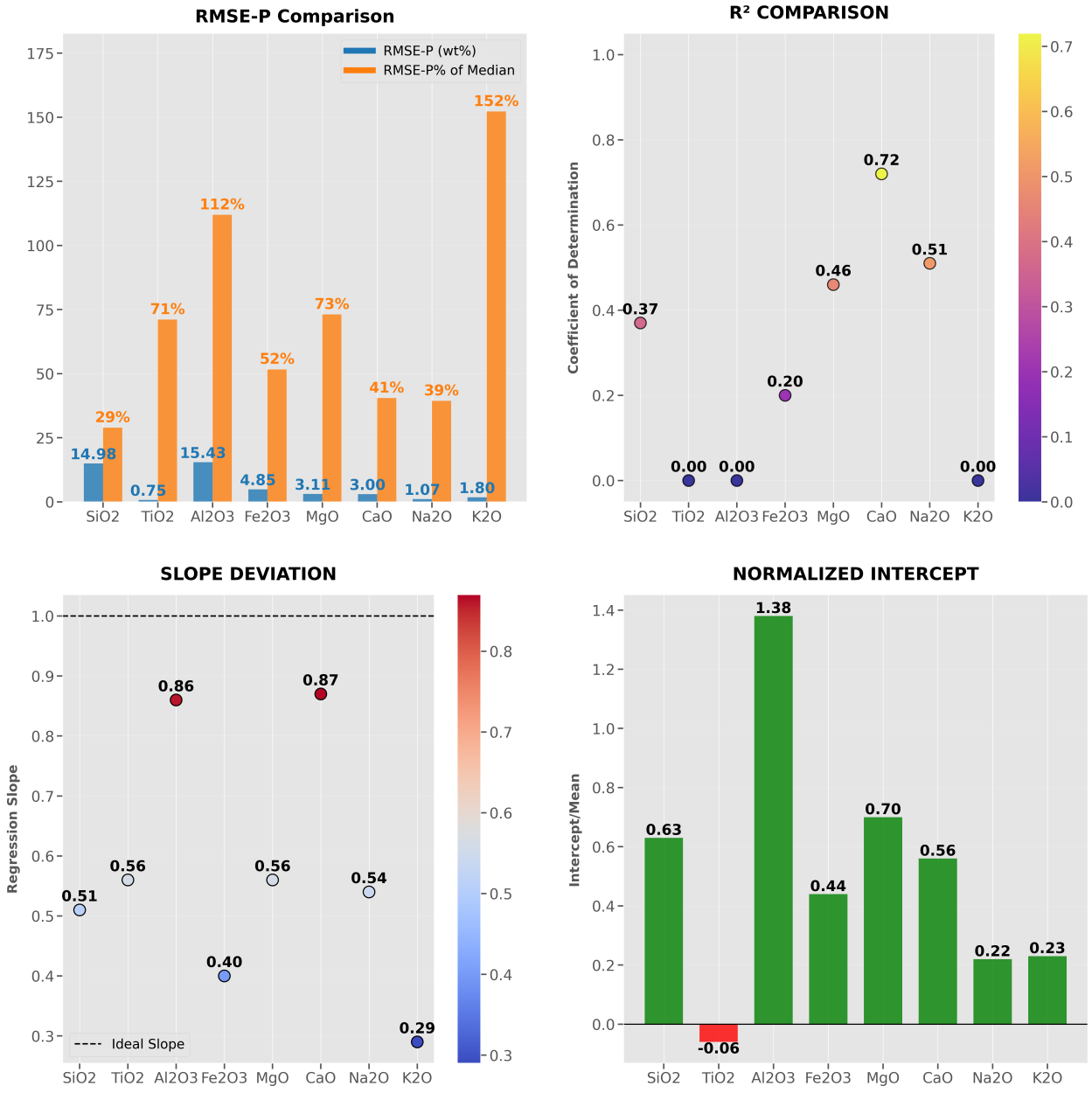

Figure 5.20: SuperCam Scenario 1 Plots: Train on All, Test on SC

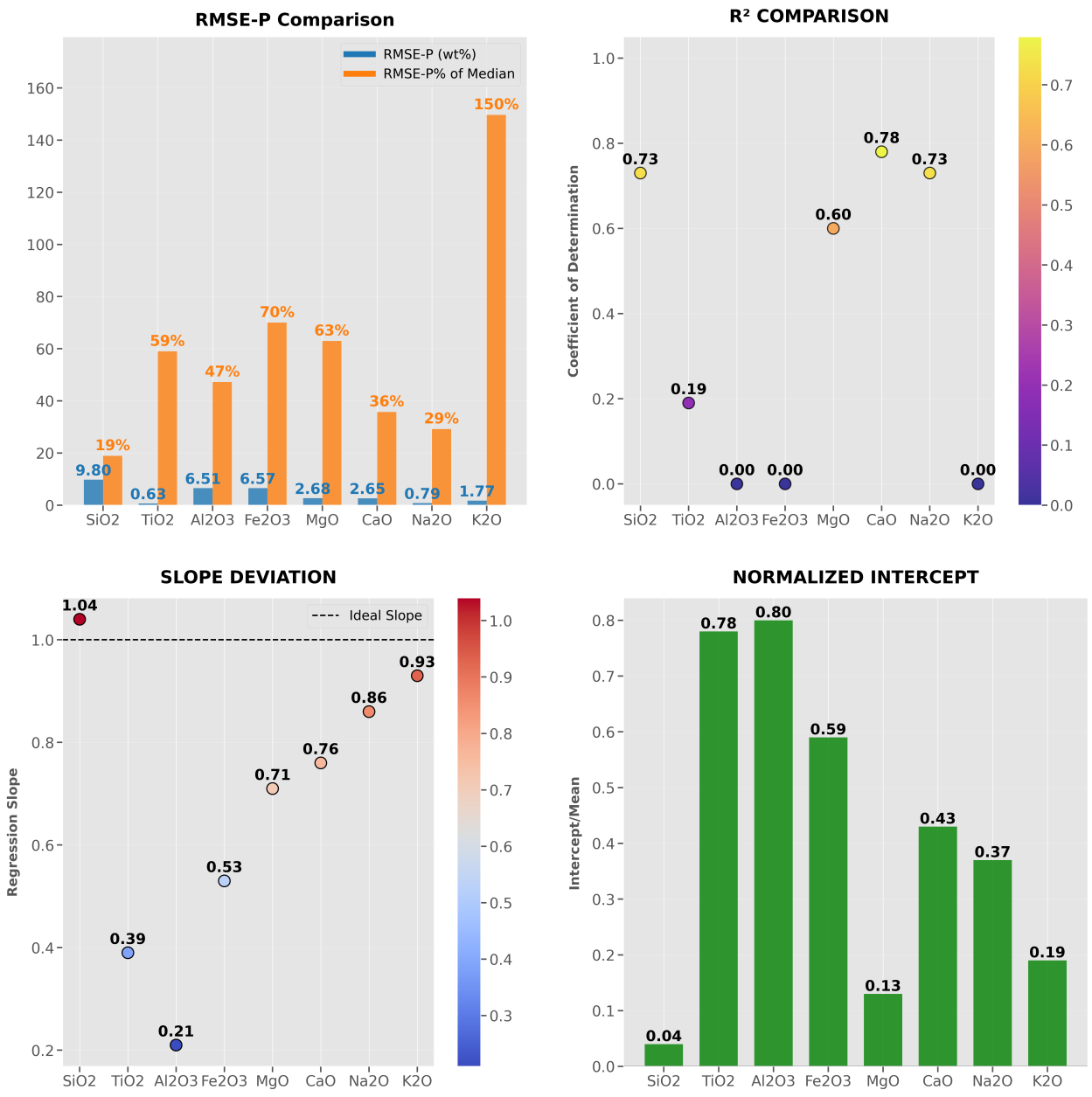

Figure 5.21: SuperCam Scenario 1 Plots: Train on All, Test on SC CT (PDS PLS)

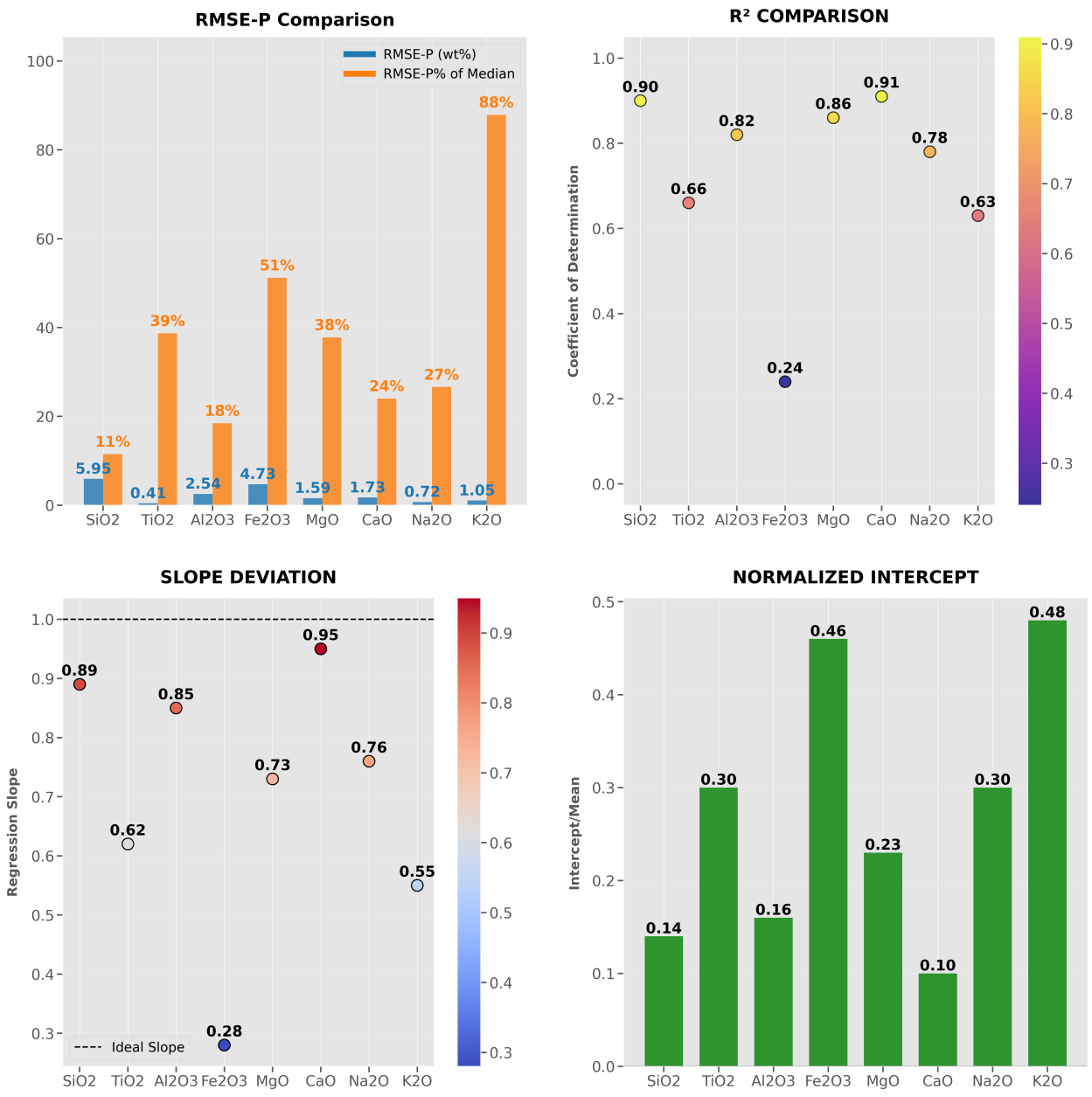

Figure 5.22: SuperCam Scenario 1 Plots: Train on All+SC, Test on SC

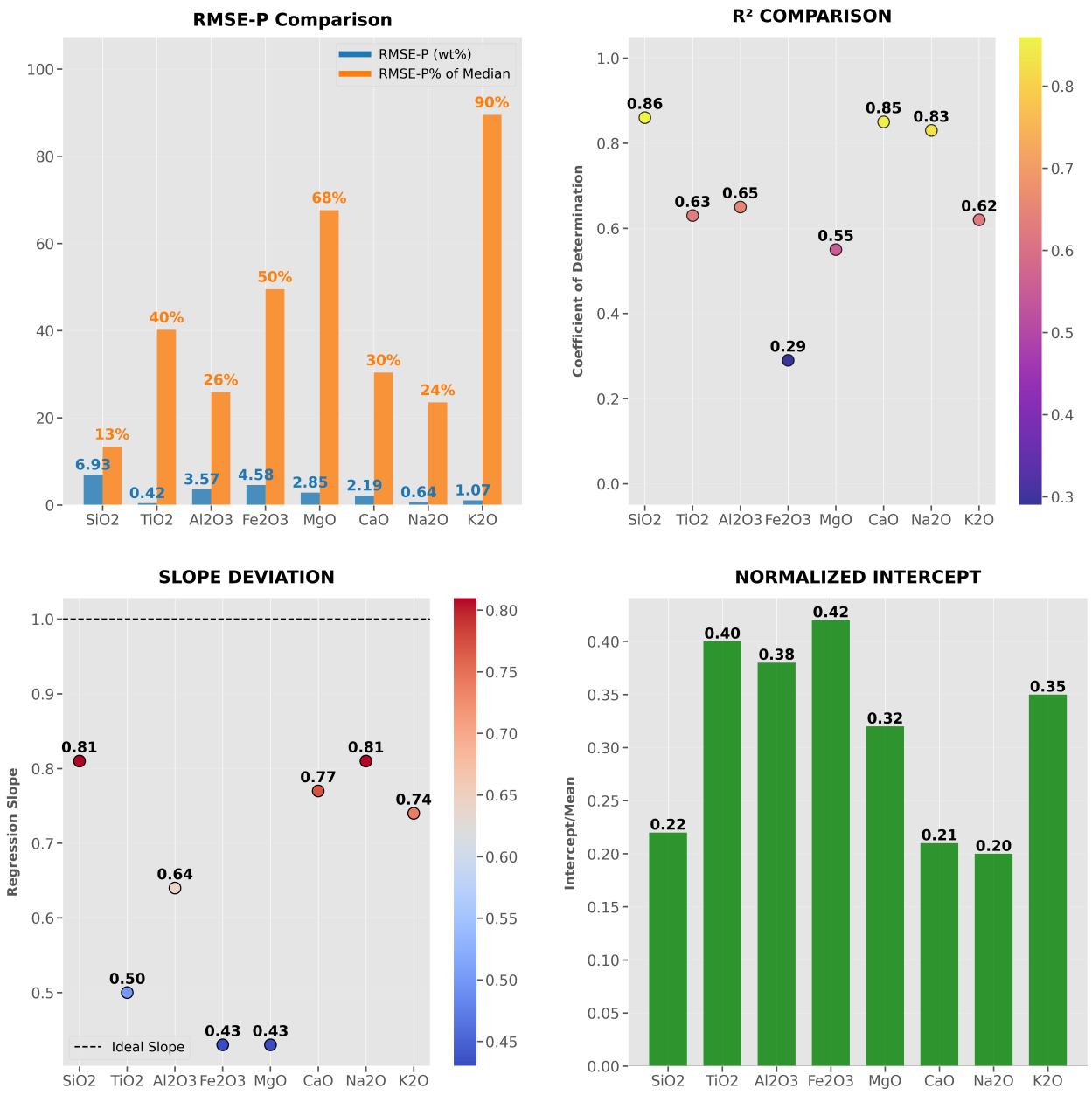

Figure 5.23: SuperCam Scenario 1 Plots: Train on All+SC CT, Test on SC CT (PDS PLS)

5.4.3 SuperCam: Scenario 2

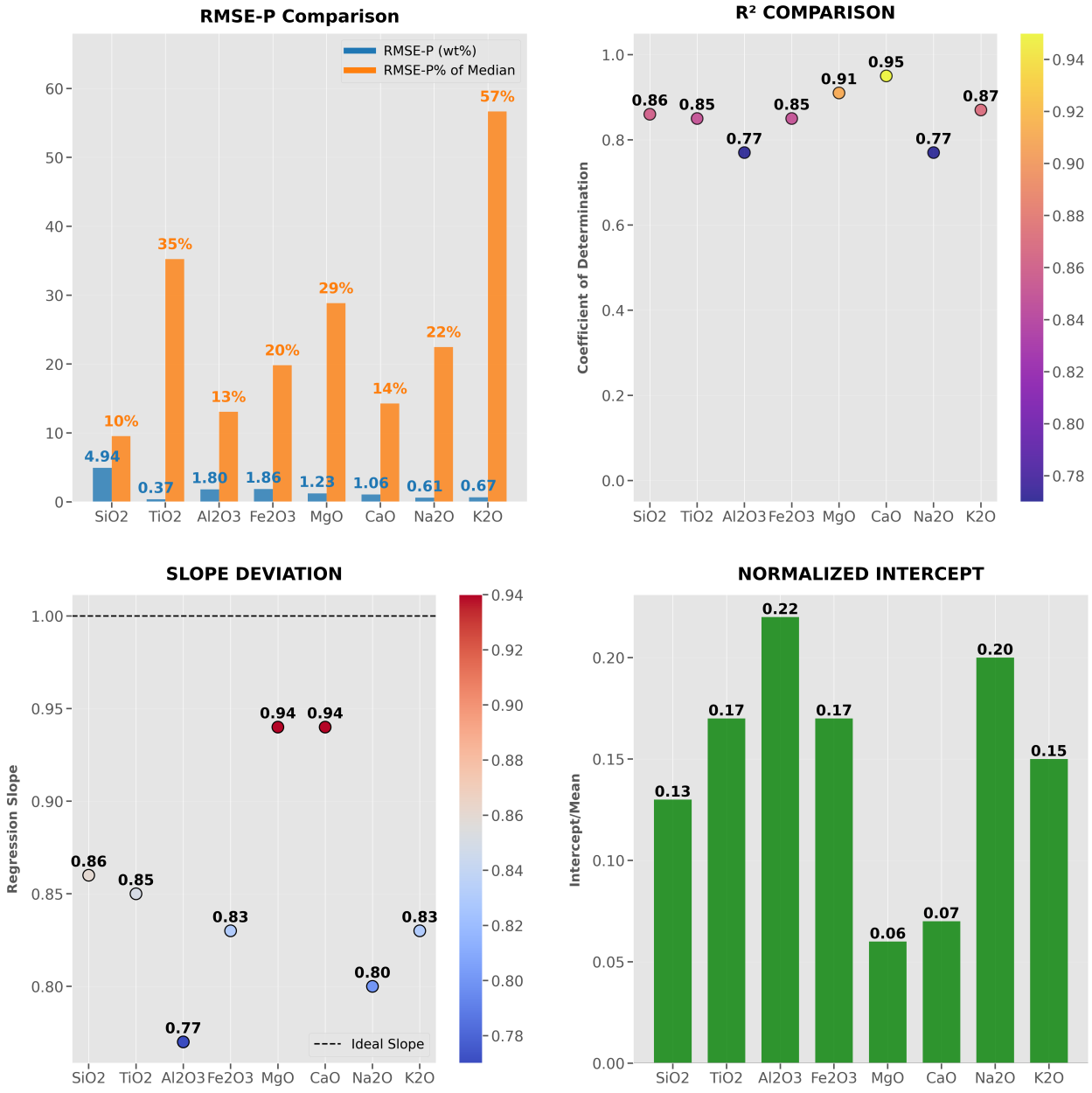

Figure 5.24: SuperCam Scenario 2 Plots: Train on All, Test on SL 10k Mars

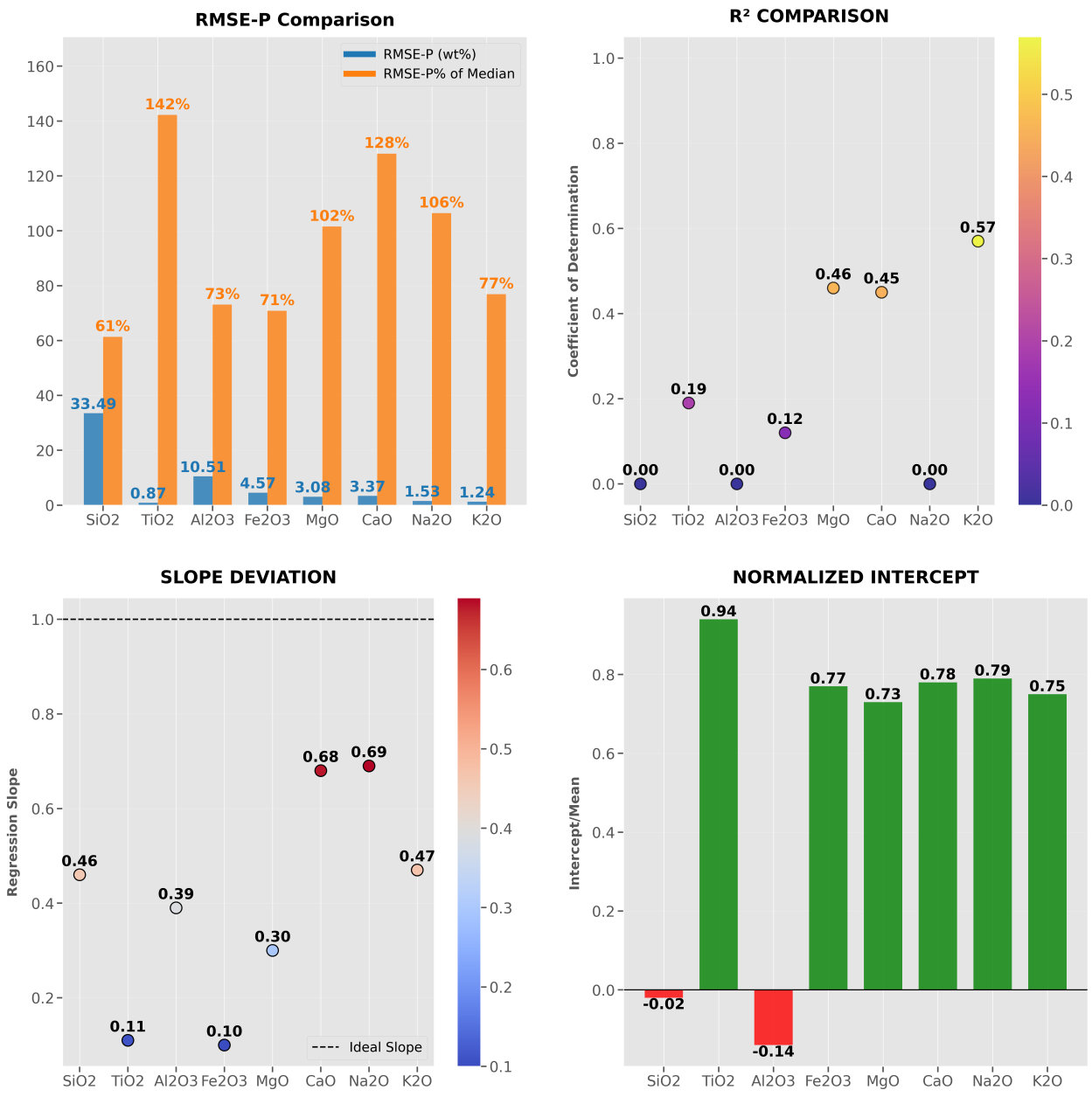

Figure 5.25: SuperCam Scenario 2 Plots: Train on SC, Test on SL 10k Mars

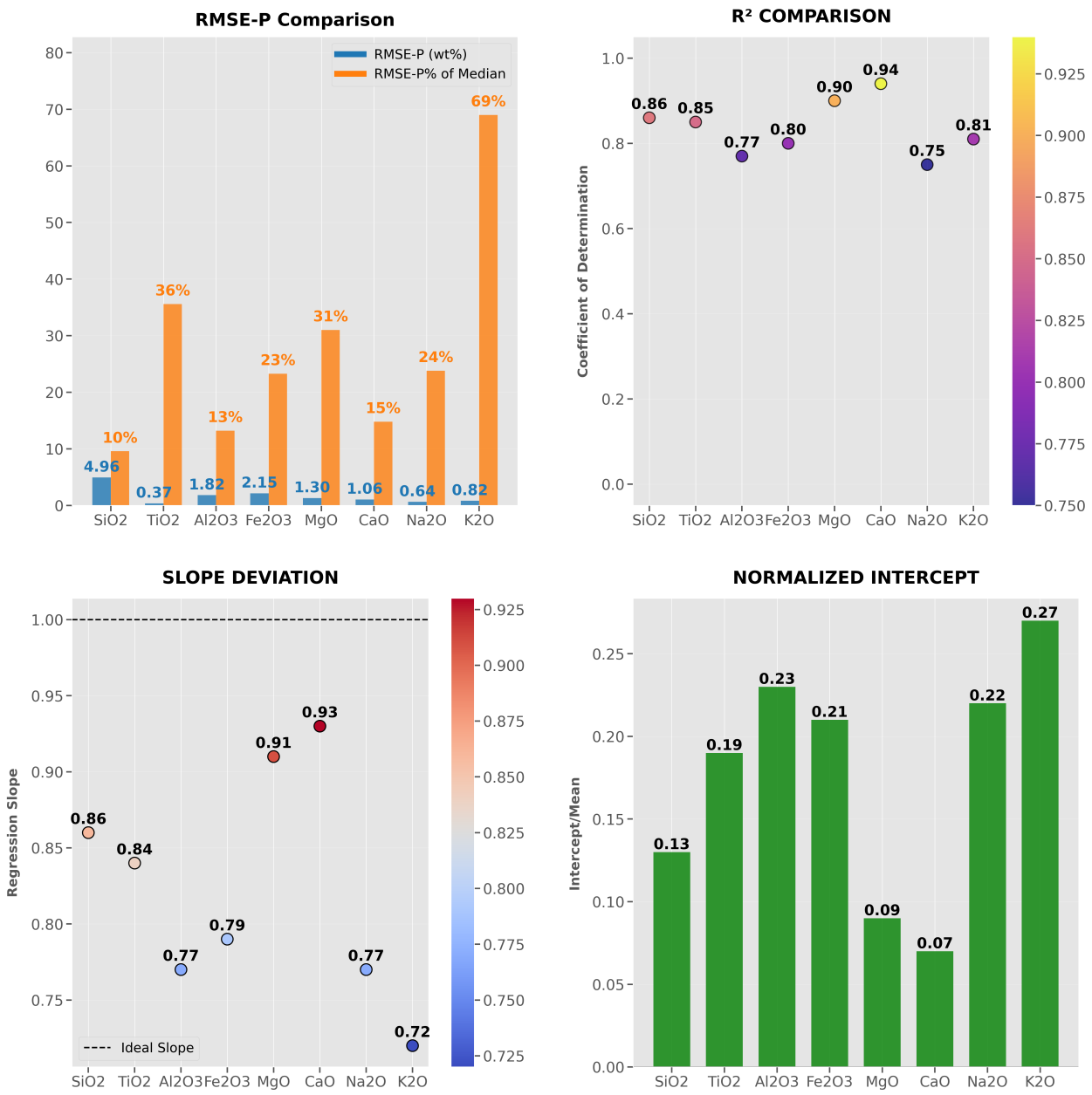

Figure 5.26: SuperCam Scenario 2 Plots: Train on All+SC, Test on SL 10k Mars

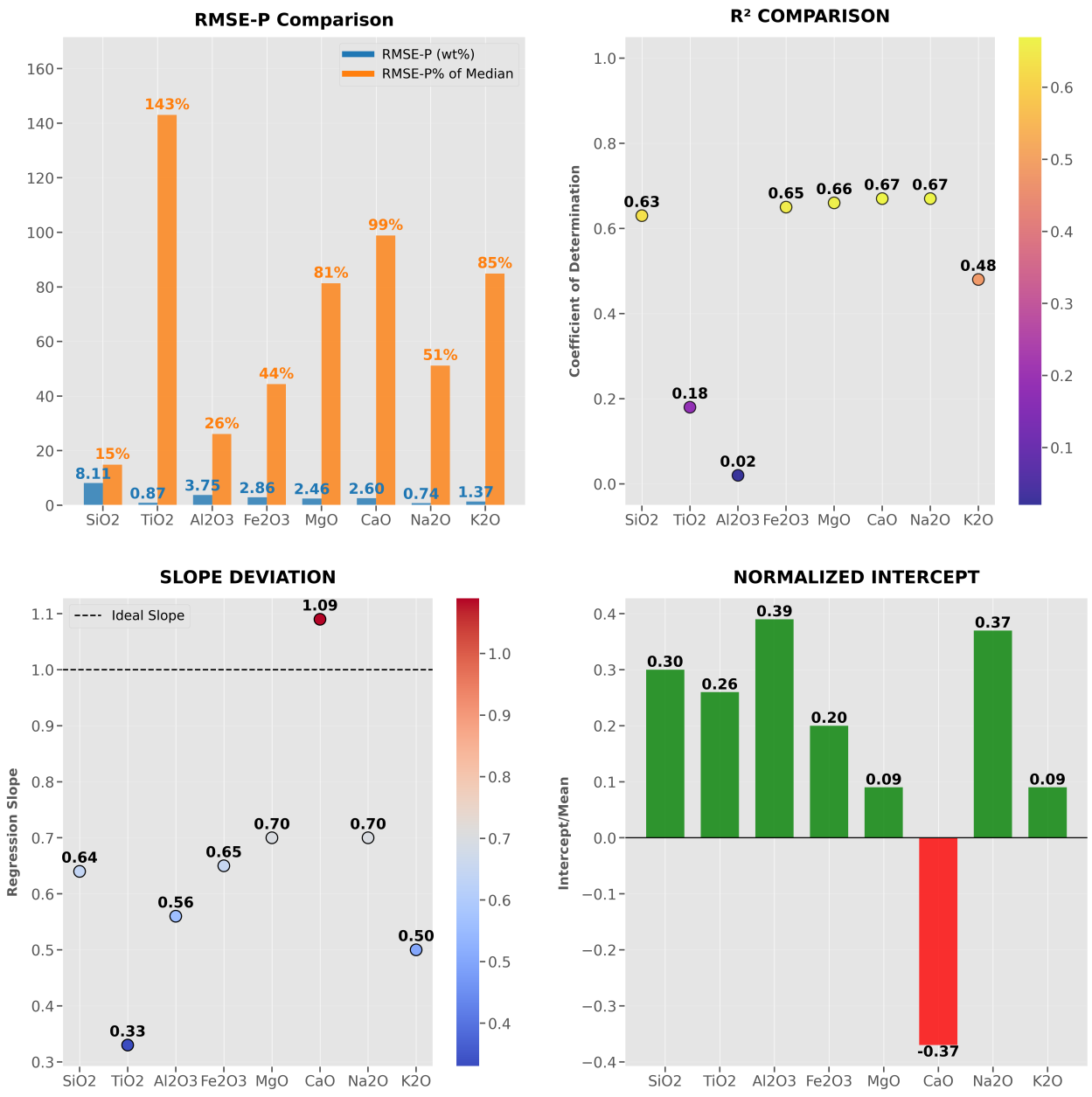

Figure 5.27: SuperCam Scenario 2 Plots: Train on SC CT, Test on SL 10k Mars

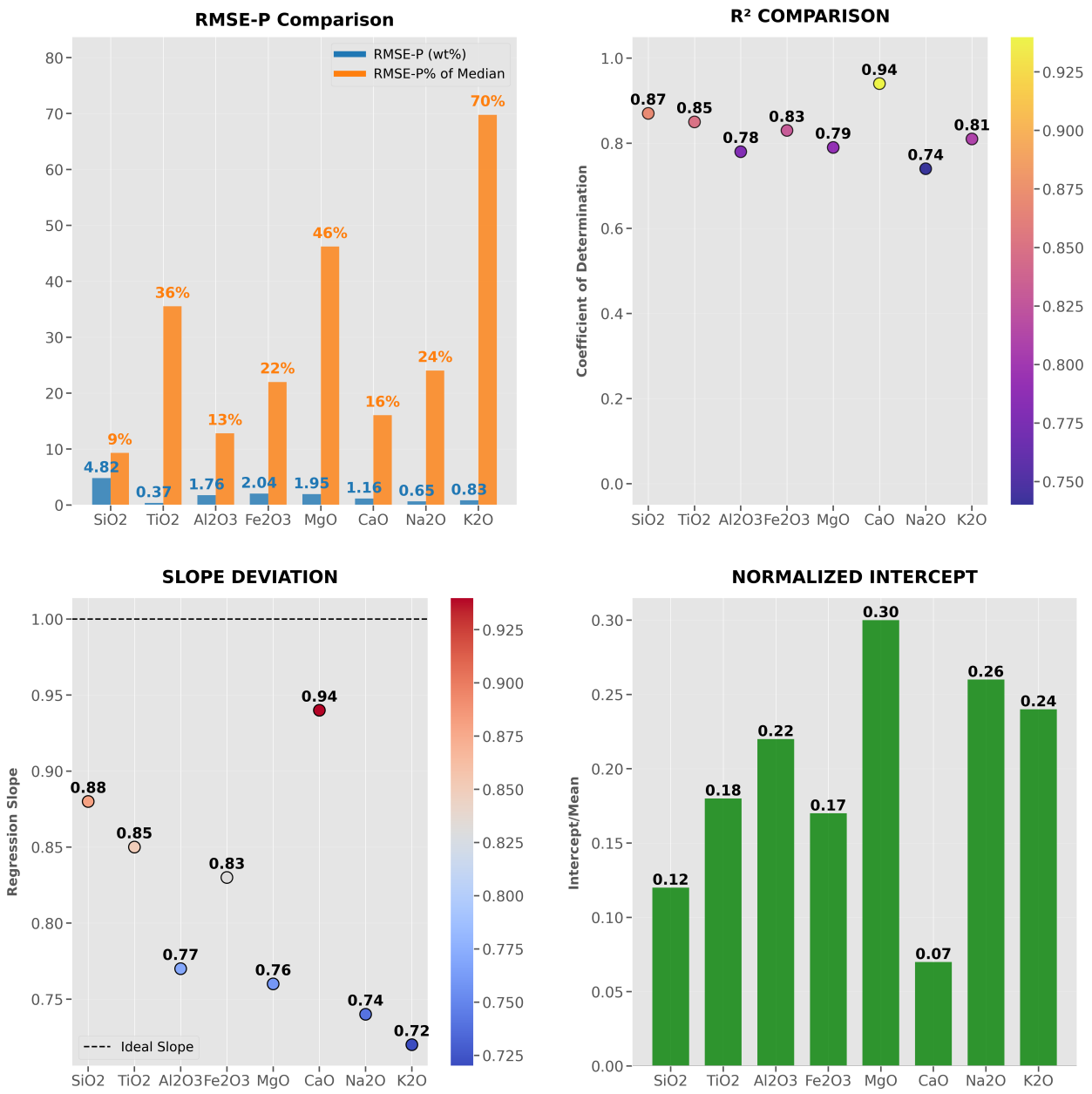

Figure 5.28: SuperCam Scenario 2 Plots: Train on All+SC CT (PDS PLS), Test on SL 10k Mars

5.4.4 SuperCam: Scenario 3

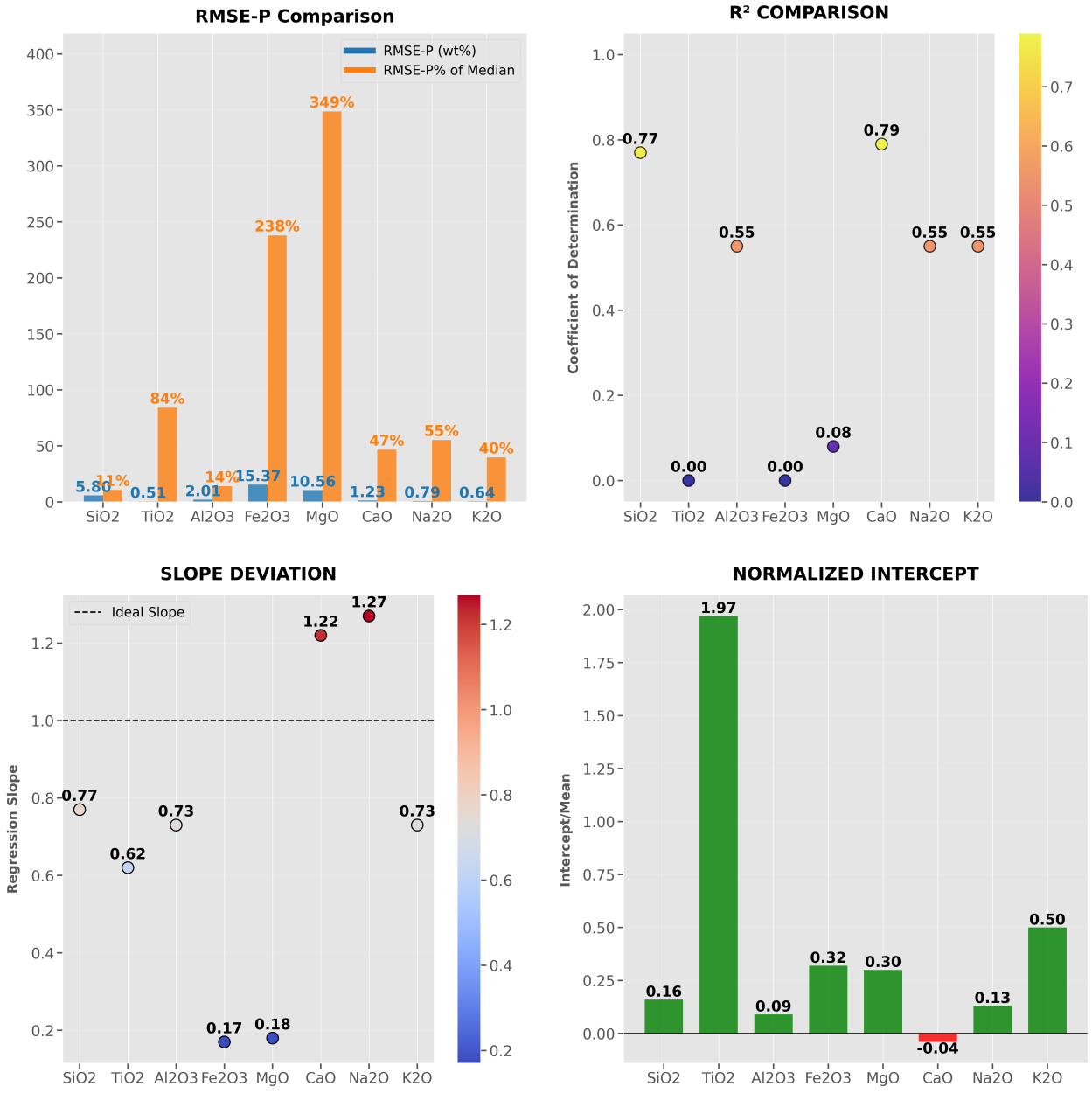

Figure 5.29: SuperCam Scenario 3 Plots: Train on SC, Test on SuperCam Mars Calibration Targets

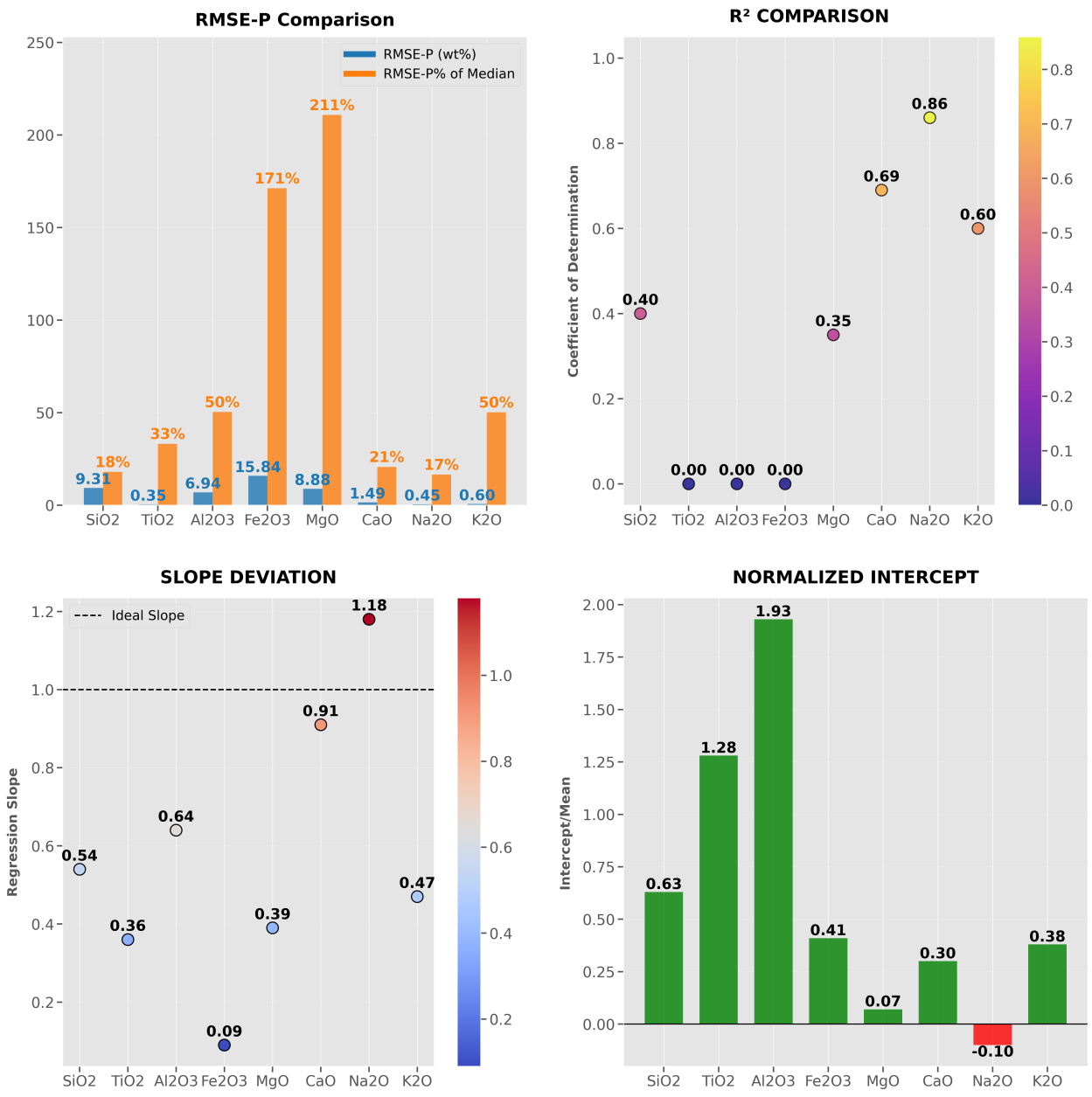

Figure 5.30: SuperCam Scenario 3 Plots: Train on All+SC, Test on SuperCam Mars Calibration Targets

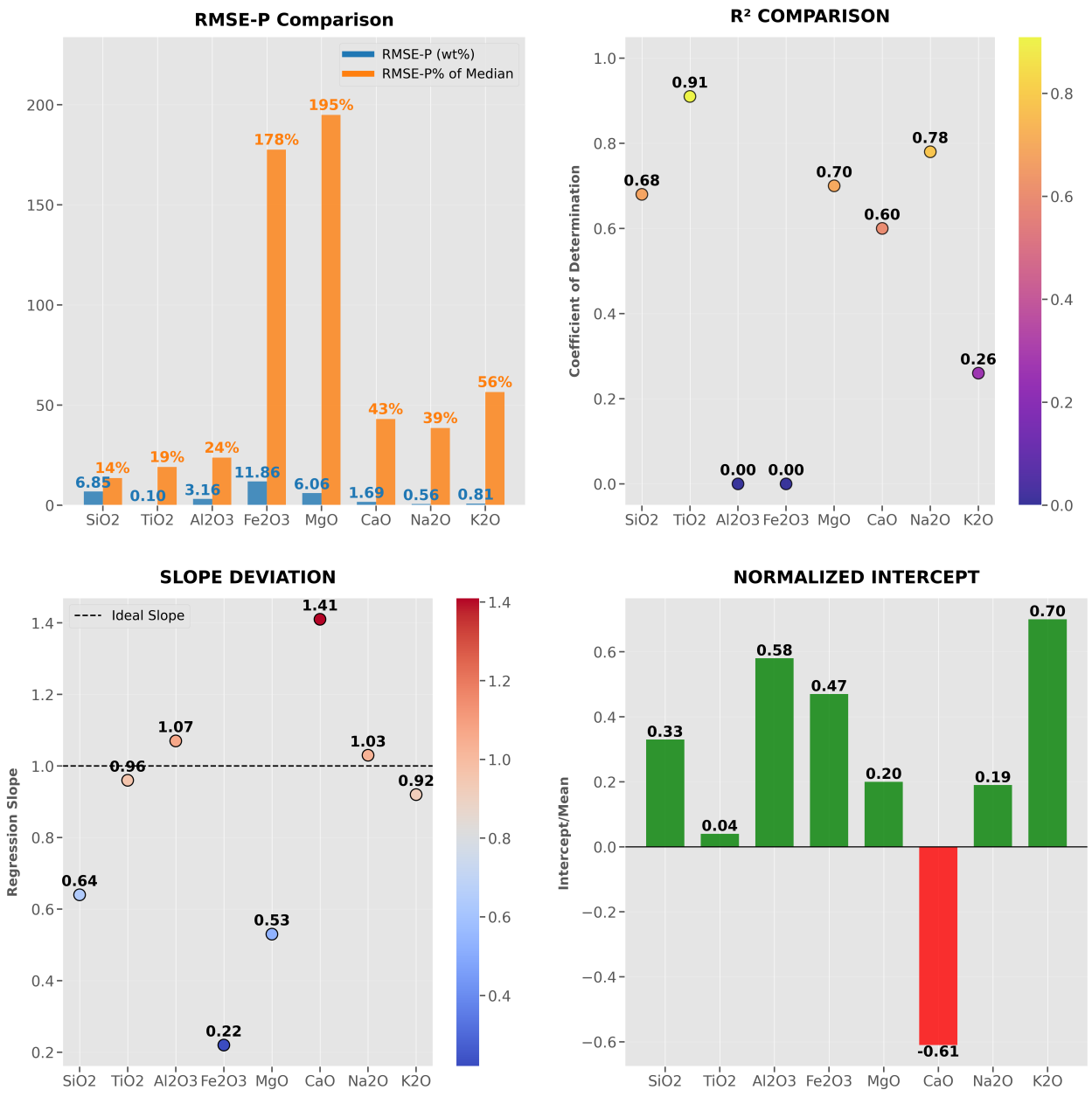

Figure 5.31: SuperCam Scenario 3 Plots: Train on SC+Calibration Targets, Test on SuperCam Mars Calibration Targets

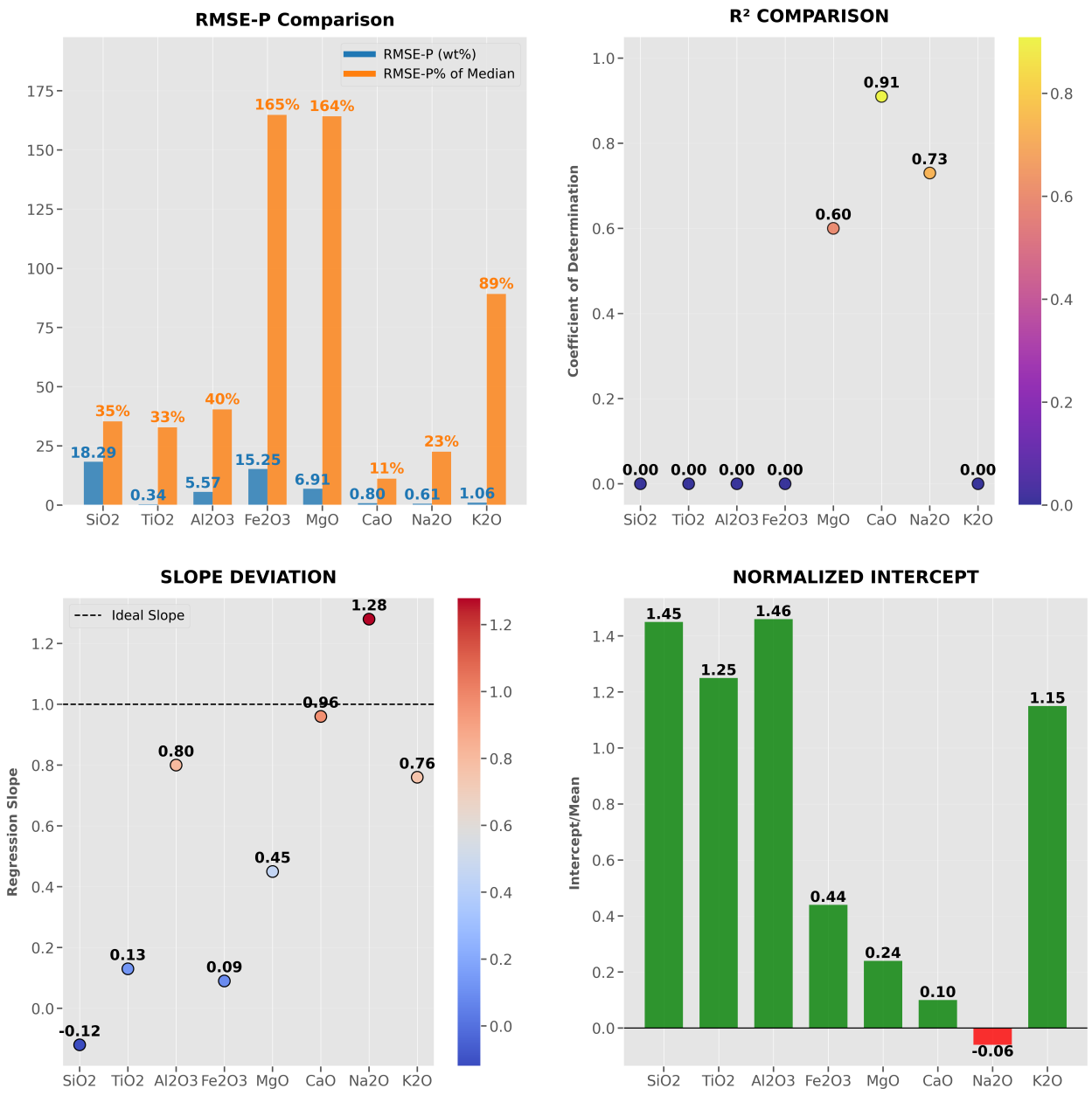

Figure 5.32: SuperCam Scenario 3 Plots: Train on All+SC+Calibration Targets, Test on SuperCam Mars Calibration Targets

5.4.5 SuperCam: Scenario 4

This unique scenario tests a subset of the "All" model by looking only at SuperLIBS 10k and 18k Mars datasets. To verify how well calibration transfer does to a smaller dataset, this scenario looks at calibration transfer from SuperCam to SuperLIBS 10k and 18k Mars datasets.

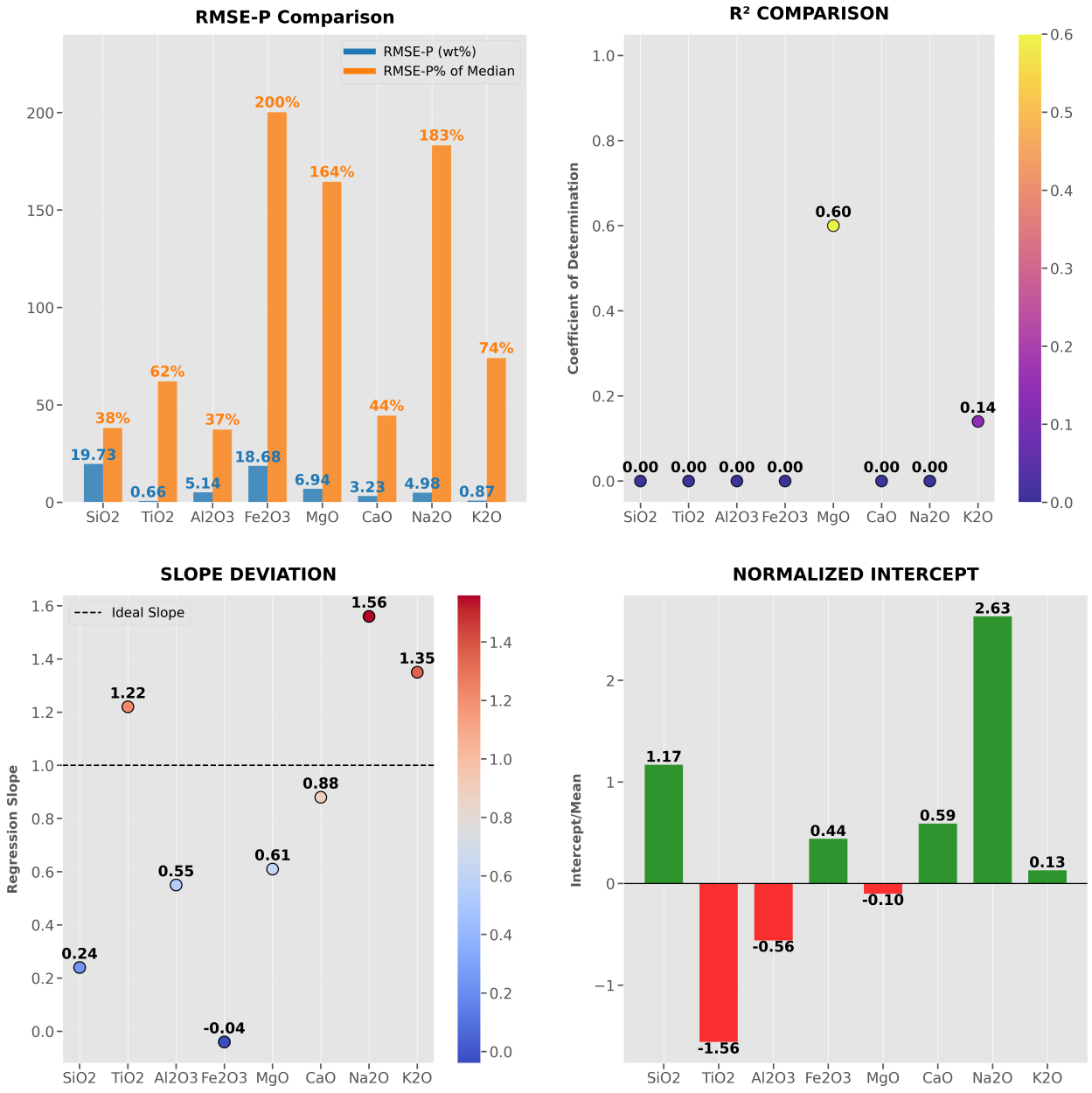

Figure 5.33: SuperCam Scenario 4 Plots: Train on SL 10k Mars+18k Mars, Test on SC

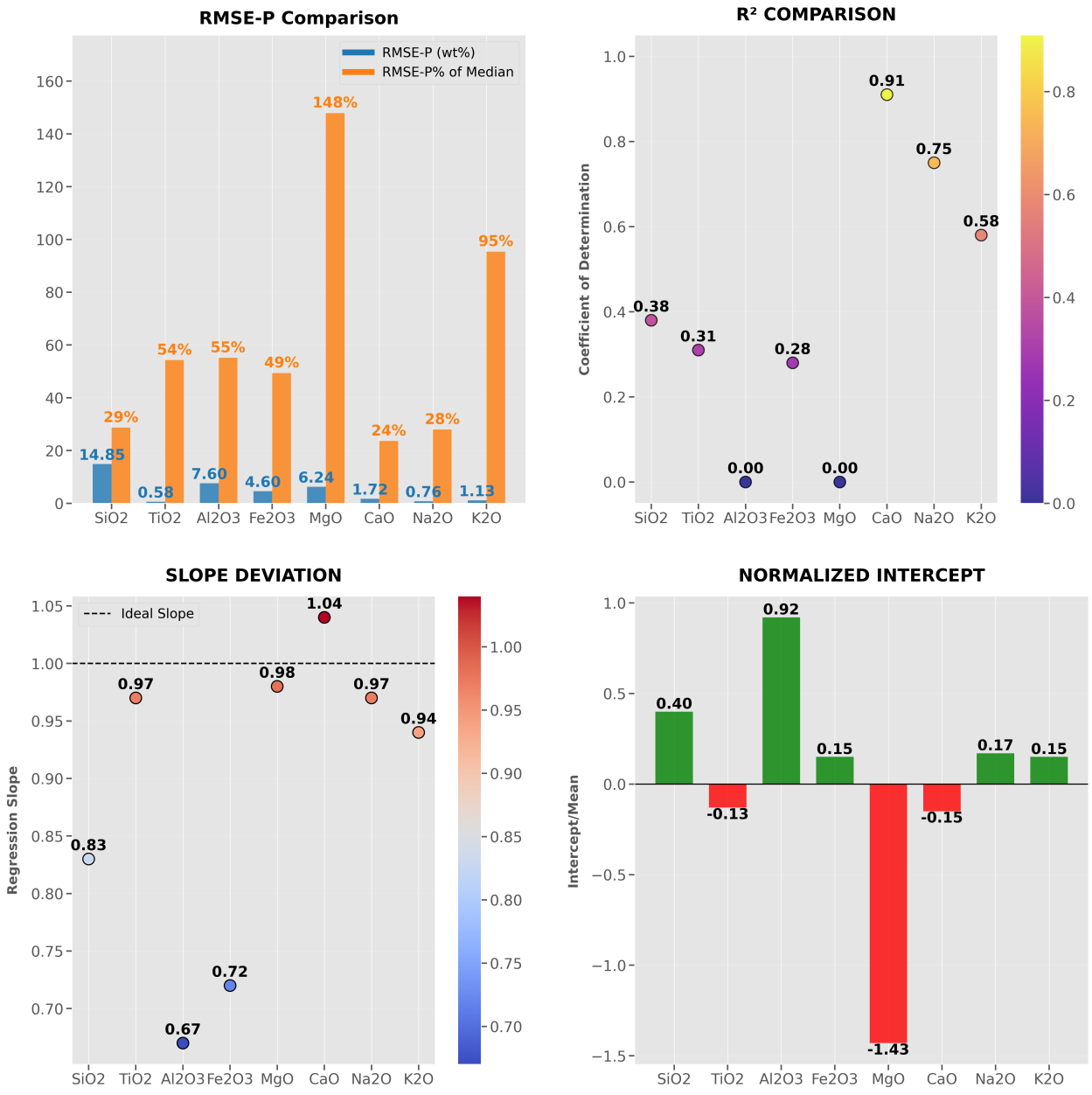

Figure 5.34: SuperCam Scenario 4 Plots: Train on SL 10k Mars+18k Mars CT SC (PDS PLS), Test on SC

6 Discussion

6.1 Benchmark: "All" Trained Model

The comprehensive performance metrics of the "All" model demonstrate its exceptional capability for quantitative LIBS analysis across diverse conditions. The average RMSE-P of 1.67 indicates that predictions typically fall within ±1.67 weight% of the true composition, a remarkable precision given the model's wide applicability. This is further supported by the RMSE-P over median percentage value of 27.27%, which shows the error represents just over one-quarter of typical concentration ranges, making it suitable for both major and minor element analysis. The R^2 value of 0.83 reveals that 83% of variance in the test data is explained by the model, capturing nearly all chemically meaningful patterns while filtering out instrumental noise.

The slope of 0.83 and normalized intercept of 0.17 provide critical insights into the model's calibration quality. The slope's deviation from unity (0.83 vs 1.00) indicates a slight conservative bias; the model systematically underestimates higher concentrations and overestimates lower ones by approximately 17%. However, the minimal intercept (0.17) confirms this bias is consistent across all concentration ranges rather than accumulating disproportionately at either end of the calibration curve. Together, these metrics confirm the "All" model achieves what the term "operationally ideal" performance: it maintains ∼80% predictive accuracy (as shown by both R^2 and slope) while keeping absolute errors below 2 weight% and relative errors under 30% across all major geochemical matrices.

This performance also reflects a well-managed bias-variance tradeoff from the PLS algorithm. In high-dimensional data, noisy LIBS spectra, traditional regression methods often suffer from either overfitting (low bias, high variance) or oversmoothing (high bias, low variance). PLS addresses this by projecting the high-dimensional spectra onto a lower-dimensional latent space that maximizes covariance with the response variable (elemental concentration), effectively reducing variance while preserving meaningful signal. The result is a model that generalizes well to unseen data, which is demonstrated here by low RMSE-P and strong

R^2, without yielding to overfitting. The slightly conservative slope bias reflects the model's tendency to favor stable, consistent predictions over closely fitting every detail in the data. This controlled simplification helps avoid overfitting and is often a better choice in field applications, where reliability is more important than maximizing precision. Hence, this balanced performance profile makes the model particularly valuable for mission planning and field applications, where consistent reliability across diverse samples outweighs the need for extreme precision on specific compositions. The metrics collectively suggest that the PLS architecture successfully reconciles spectral variations from different instruments and environments while preserving fundamental composition-spectrum relationships, validating the "All" model's design as both robust and versatile.

Comparing these results on previous papers, Dyar, M.D & Ytsma, C. (2021) reported an average RMSE-P of 2.46 wt% across major oxides using only one instrument ChemLIBS instrument shown in Table 6.1, while the current "All" model achieves a significantly lower average of 1.67 wt%, representing a 32% improvement in absolute prediction accuracy, shown more in details in the table below.

Table 6.1: Prediction Errors (RMSE-P) for Major Oxides by Dyar, M.D & Ytsma, C. (2021)

Major Oxide	**RMSE-P (wt%)**
SiO_2	5.92 ± 0.21
TiO_2	0.68 ± 0.08
Al_2O_3	2.75 ± 0.03
Fe_2O_3	3.03 ± 0.35
MgO	2.96 ± 0.25
CaO	2.33 ± 0.75
Na_2O	0.76 ± 0.01
K_2O	1.28 ± 0.06

6.2 ChemCam Performance Analysis

The ChemCam results reveal low predictive accuracies for each metric (RMSE-P, RMSE-P%, R^2, slope, and normalized intercept) in scenarios 1-3, compared to the benchmark results from Scenario 0.

For Scenario 0, the ChemCam model yields an average RMSE-P of 2.02, RMSE-P median percentage of 37.05, R^2 of 0.81, a slope of 0.84, and an average normalized intercept at 0.17. After calibration transferring onto SuperLIBS 10k Mars, the ChemCam CT model results in similar but improved average accuracies, with an average RMSE-P of 1.91, RMSE-P median percentage of 36.83, R^2 of 0.82, a slope of 0.84, and an average normalized intercept at 0.16. By comparing these two subscenarios, the calibration transfer model has a better overall performance on ChemCam, suggesting that future incorporation of calibration transfer onto other data and instruments combinations may also yield similar strong and low performance metrics. In statistical terms, these results indicate that these models perform well in predicting chemical compositions of these major eight major oxides, with R^2 values close to 1 (0.81 and 0.82) suggesting a strong correlation between predicted and actual values. The RMSE-P values (2.02 and 1.91) represent the average prediction error, meaning the models' estimates are typically within about 2 units of the true values. The slope of 0.84 shows that these models slightly underestimates higher values and overestimates lower ones, but the small normalized intercept (0.17 and 0.16) indicates minimal systematic bias.

Scenario 1 examines the performance of the "All" dataset when applied to the ChemCam test set, both with and without calibration transfer (CT). The subscenarios are structured sequentially based on different training-testing combinations. The first subscenario involves training solely on the "All" dataset and testing directly on ChemCam. The second subscenario trains on "All" and tests on the applied calibration transfer on ChemCam. The third subscenario incorporates ChemCam data into the "All" training set and then evaluates performance on the original ChemCam test set. Finally, the fourth subscenario trains on a combination of "All" and calibration-transferred ChemCam data (ChemCam CT) before testing on the ChemCam CT set. The last two subscenarios are particularly significant, as they demonstrate how the inclusion of ChemCam data, either in its original form or after calibration transfer, affects model performance when analyzing the same underlying ChemCam measurements.

Among the four subscenarios examined, the last two, which incorporate ChemCam data into the training set, produce remarkably similar results while significantly outperforming

the first two scenarios that exclude ChemCam from training. This suggests that including representative data from the target domain (ChemCam) in the training set is crucial for optimal model performance. Comparing these two best-performing cases, the model trained on "All"+ChemCam achieves an average RMSE-P of 2.28, RMSE-P median of 35.85%, R^2 of 0.73, slope of 0.75, and normalized intercept of 0.27, while the "All"+ChemCam CT version yields slightly better metrics overall (RMSE-P = 2.52, RMSE-P median 34.36%, R^2 = 0.70, slope = 0.81, intercept = 0.21). The minimal improvement from calibration transfer indicates that the "All" dataset alone provides sufficient generalization capability. However, when compared to the ChemCam-only and ChemCam CT-only benchmark results from Scenario 0 (RMSE-P = 2.02 and 1.91 respectively), both top-performing subscenarios in Scenario 1 exhibit modest performance degradation. The "All"+ChemCam model shows an RMSE-P increase of 0.26 points (2.02 $\rightarrow$ 2.28), while the "All"+ChemCam CT version experiences a larger increase of 0.61 points (1.91 $\rightarrow$ 2.52). This pattern holds across most metrics, suggesting two key insights: first, that the additional diversity from the "All" dataset introduces some noise that slightly reduces ChemCam-specific precision; and second, that calibration transfer, while beneficial within the ChemCam domain, does not compensate for the cross-dataset generalization challenges when incorporating external LIBS data.

Scenario 2 examines model performance under a different testing paradigm, evaluating on SuperLIBS 10K Mars rather than unseen ChemCam data. This configuration is particularly valuable for assessing calibration transfer effectiveness, as the test set now aligns with the transfer target domain. The investigation follows five distinct subscenarios, maintaining the same training configurations as Scenario 1 but with altered test conditions: (1) the benchmark case training exclusively on "All"; (2) training solely on ChemCam; (3) training on calibration-transferred ChemCam (ChemCam CT); (4) combined training on "All"+ChemCam; and (5) training on "All"+ChemCam CT. This systematic variation allows direct comparison with Scenario 1's results while isolating the impact of changing only the test set composition, providing crucial insights into model behavior across different instrumental domains.

The comparison between models without "All" dataset incorporation reveals a dramatic improvement through calibration transfer. The ChemCam-only model demonstrates poor

performance (RMSE-P = 4.64, RMSE-P over median = 82.39%, R^2 = 0.27), indicating substantial prediction errors and weak correlation. In striking contrast, the ChemCam CT model shows markedly better accuracy across all metrics: RMSE-P improves by 46.6% to 2.48, median error decreases by 23.43 percentage points to 58.96%, and R^2 more than doubles to 0.59. The slope (0.75 vs 0.83) and normalized intercept values (0.32 vs 0.45) also show better calibration results. This robust evidence demonstrates that when the calibration transfer domain (SuperLIBS 10K Mars) matches the test set, the model achieves optimal performance, which highlights the critical importance of domain alignment between calibration transfer and testing environments for LIBS data analysis.

The incorporation of "All" dataset in the final two subscenarios yields revealing comparisons with Scenario 1's results. Training on "All"+ChemCam produces strong performance (RMSE-P = 1.63, median 27.24%, R^2 = 0.84), while the "All"+ChemCam CT version achieves even better results (RMSE-P = 1.57, median 25.53%, R^2 = 0.85). This confirms that model generalization improves with "All" dataset inclusion, mirroring Scenario 1's findings. However, unlike Scenario 1 where calibration transfer showed minimal benefits, here it provides clear improvements across all metrics. The key difference lies in the test set alignment: when the calibration transfer target (SuperLIBS 10K Mars) matches the test domain, CT demonstrates its full potential, reducing errors by 3.7% in RMSE-P and improving R^2 by 0.01 compared to non-CT version. This contrast with Scenario 1's results (where testing on ChemCam CT showed no improvement over non-CT) fundamentally demonstrates that calibration transfer only provides significant benefits when the test set matches the transfer target domain. In Scenario 1, transferring to SuperLIBS but testing on ChemCam created a domain mismatch, while Scenario 2's perfect alignment between transfer target and test set allows the CT to properly adjust the model's feature space. The 3.7% RMSE-P improvement in Scenario 2 versus Scenario 1's neutral results establishes that CT requires congruent source-target-test domains to optimize performance, explaining why Scenario 1's "All"+ChemCam CT model failed to outperform its non-CT counterpart when tested on ChemCam data.

Scenario 3 investigates the impact of test set size and composition on model performance, focusing specifically on ChemCam's onboard calibration targets (approximately 80-100

samples). While exhaustive combination testing was impractical due to computational constraints, this study examines four strategically selected training configurations: (1) ChemCam-only training as the baseline; (2) combined "All"+ChemCam dataset; (3) ChemCam augmented with its onboard calibration targets; and (4) the comprehensive "All"+ChemCam+Calibration Targets combination. These configurations systematically evaluate how different training set expansions affect prediction accuracy on the fixed calibration target test set, particularly examining whether incorporating the diverse "All" dataset or additional calibration-specific samples yields superior performance for this specialized prediction task.

Out of these four subscenarios, the overall best performer is ChemCam+ChemCam Mars Onboard Calibration Targets, achieving an average RMSE-P of 1.16, RMSE-P median of 27.88%, R^2 of 0.32, slope of 1.14, and normalized intercept of -0.48. While these metrics remain moderate, they surpass all other subscenarios in this scenario. Critically, models incorporating the "All" dataset ("All"+ChemCam+ChemCam Mars Onboard Calibration Targets,) exhibit degraded performance, such as the RMSE-P increases to 1.98, and demonstrates that adding heterogeneous LIBS data hurts rather than helps when the test set is highly specialized (ChemCam's onboard targets). This suggests that for narrowly scoped prediction tasks, domain-specific training data (here, ChemCam and its calibration targets) outperforms generalized models trained on mixed instrumental data, as extraneous features from unrelated LIBS setups introduce noise rather than useful signal.

These findings align with recent theoretical work on dataset biases in machine learning Sagawa, S. et al. (2020) weighs in that increasing training data can paradoxically harm performance on specialized subgroups when: (1) the additional data introduces strong spurious correlations (analogous to instrument-specific data in "All" dataset), and (2) model capacity is insufficient to disentangle these from meaningful signals. This perfectly explains Scenario 3's results, where incorporating the heterogeneous "All" dataset degraded performance on ChemCam's calibration targets (RMSE-P increase from 1.16 to 1.98). The paper's key insight that overparameterized models tend to "memorize" dominant patterns at the expense of minority groups, directly applies to this case: the method partial least squares here prioritized

learning instrument-specific features from the larger "All" dataset over the precise calibration target signatures. Hence, the paper's proposed solution of strategic subsampling mirrors this study's empirical finding that domain-specific training (ChemCam+calibration targets only) outperforms larger mixed datasets. This finding also aligns with Dyar, M.D & Ytsma, C. (2021), where the authors found the 'use of too few samples in the test set causes accuracy to be measured on only a handful of samples, which might lead to over- or under-estimation of prediction accuracy depending on the issue of population equivalence just discussed'. This study found that the use of a sample test size does lead to under-estimation of predidction accuracy.

6.3 SuperCam Performance Analysis

The SuperCam analysis yields performance trends consistent with those observed for ChemCam, following identical experimental scenarios with two key additions: (1) an extra subscenario in Scenario 3, and (2) a dedicated Scenario 4 probing calibration transfer mechanics. Scenario 4 introduces a novel inversion of the standard calibration transfer approach: instead of adapting SuperCam data to the SuperLIBS domain, it transfers from SuperLIBS 10k/18k Mars to SuperCam, then evaluates on the original SuperCam test set. This design directly tests our earlier finding that optimal performance requires alignment between calibration transfer targets and test domains by examining whether the inverse transformation preserves predictive accuracy.

Scenario 0 demonstrates results that contrast sharply with those observed for ChemCam. In this case, applying calibration transfer from SuperCam to SuperLIBS 10k Mars leads to degraded performance compared to analyzing SuperCam data directly. The baseline model without calibration transfer shows consistently better metrics, including an RMSE-P of 2.42 (versus 2.67 with transfer), RMSE-P median percentage of 50.04% (versus 54.24%), and R^2 of 0.74 (versus 0.68). The regression parameters also reflect this pattern, with slope values of 0.69 (no transfer) versus 0.78 (with transfer) and intercepts of 0.26 versus 0.21. These results demonstrate that the transfer process disrupts the natural spectral relationships inherent to SuperCam's data, with the performance degradation stemming from fundamental

mismatches in instrumental characteristics such as detector specifications (0.15–0.20 nm vs. 0.08 nm resolution in UV/VIS). As observed in Shabbir, S. et al. (2022), such "near-identical but operationally distinct" systems create a transfer paradox: the PLS algorithm fails to apply sufficient corrections precisely because the instruments appear superficially compatible, mirroring their finding that statistical similarity can mask critical physical matrix effects. This explains why ChemCam's more distinct 0.30 nm resolution enabled robust transfer, its obvious differences triggered comprehensive recalibration. The results align with Shabbir, S. et al. (2022)'s demonstration that transfer learning degrades when source-target pairs exhibit "intermediate divergence", too similar for full correction yet too different for direct compatibility. This underscores that successful calibration transfer depends not on nominal component matches, but on quantifiable alignment of plasma conditions and spectral sampling regimes.

The performance trends observed in SuperCam's Scenario 1 mirror those of ChemCam, confirming consistent behavior across instruments. The optimal configuration which is training on "All"+SuperCam and testing on native SuperCam data, achieved an RMSE-P of 2.34 (median 37.00%), R^2 of 0.72, slope of 0.70, and normalized intercept of 0.27. These results align with Scenario 0's critical finding that calibration transfer fails to enhance SuperCam's prediction accuracy, explaining why the All"+SuperCam CT variant showed no improvement. The systematic degradation in all metrics when incorporating the "All" dataset (similar to ChemCam's results) demonstrates the model's robust generalization behavior: while mixed training data reduces absolute performance, it maintains relative trends across instrument configurations.

Scenario 2 maintains identical subscenarios to ChemCam while using SuperLIBS 10k Mars as the fixed test set. The final two subscenarios demonstrate nearly identical performance metrics when the calibration transfer target domain matches the test set: training on "All"+SuperCam yields RMSE-P = 1.64 (median 27.53%), R^2 = 0.84, slope = 0.82, and intercept = 0.18, while "All"+SuperCam CT produces RMSE-P = 1.70 (median 29.47%), R^2 = 0.82, slope = 0.81, and intercept = 0.19. These results represent significant improvement over Scenario 1's corresponding subscenarios where the test set differed from the transfer

target, confirming that optimal calibration transfer requires domain alignment between target and test conditions. The minimal performance difference between these subscenarios suggests that when instruments exhibit substantial inherent differences, calibration transfer provides negligible benefits despite matching target-test domains. This pattern precisely mirrors ChemCam's Scenario 2 conclusions, reinforcing the generalization of these findings across different LIBS instruments.

Scenario 3 investigates whether incorporating additional LIBS instruments (the "All" dataset) improves predictions for SuperCam's calibration targets compared to using only SuperLIBS Mars conditions-derived data, which is different from ChemCam's subscenarios. The comparative results between "All"+SuperCam+Cal Targets (RMSE-P = 3.89, median 70.06%, R^2 = 0.49, slope = 0.85, intercept = 0.24) and 10k/18k Mars+SuperCam+Cal Targets (RMSE-P = 3.82, median 58.16%, R^2 = 0.46, slope = 0.80, intercept = 0.42) reveal three key findings: First, the minimal difference in RMSE-P (Δ = 0.07) demonstrates no meaningful improvement from including heterogeneous instruments. Second, the SuperLIBS-only model's superior median error (58.16% vs 70.06%) could indicates that external LIBS data introduces spectral noise that outweighs any benefits from increased sample diversity. Third, the intercept and slope differences suggest the SuperLIBS-focused model achieves better baseline alignment (intercept = 0.42 vs 0.24) while making more conservative predictions (slope = 0.80 vs 0.85). These results collectively confirm that for SuperCam calibration tasks, optimal performance derives from training on focused, instrument-compatible data (SuperLIBS variants) rather than aggregating measurements from dissimilar systems. The $\sim$1% variation in R^2 values further reinforces that broader data inclusion neither enhances nor significantly degrades model correlation, suggesting spectral feature quality matters more than dataset quantity for this application, though this study's design intentionally prioritized broad instrument inclusion over optimized spectral matching. Hence, the critical next step is determining how to maximize these gains through targeted feature selection rather than the exploratory comparison presented here.

rScenario 4 represents a specialized investigation of calibration transfer directions, reversing the standard approach by transforming SuperLIBS 10k/18k Mars data to match SuperCam's

domain before testing on SuperCam observations. There is only one subscenario due to time constraints and how the main goal of this study is to study the inclusion of "All" model instead of the individual instruments. This inverse transfer performs significantly worse than Scenario 2's standard SuperCam-to-SuperLIBS transfer (RMSE-P = 2.85 vs 4.68), demonstrating the critical importance of transfer directionality. The comparison model is training SuperCam CT and testing on SuperLIBS 10k Mars. These results are not optimal results from Scenario 2 but gives a good comparison of how well Scenario 4 did. The resulting metrics whic are RMSE-P = 2.85 (median 68.07%), R^2 = 0.49, slope = 0.65, and intercept = 0.17. While these Scenario 2 metrics are suboptimal compared to incorporating "All" models, they still outperform Scenario 4's inverse transfer by wide margins across all metrics. This configuration tests the boundary conditions of the established principle that target-test domain alignment optimizes performance. The resulting metrics which are RMSE-P = 4.68 (median 60.28%), R^2 = 0.40, slope = 0.89, and intercept = 0.01, demonstrate significantly degraded prediction quality compared to Scenario 1's baseline ("All" tested on SuperCam). Notably, even Scenario 2's modest success with standard-direction transfer disappears when the direction is reversed. Three key implications emerge from these results: (1) The transfer direction proves non-invertible, while SuperCam-to-SuperLIBS transfer succeeds marginally, the reverse operation fails due to SuperLIBS' higher resolution attempting to emulate lower-resolution SuperCam spectra; This explains why Scenario 2's RMSE-P (2.85) remains below Scenario 4's (4.68); (2) The near-zero intercept suggests correct baseline alignment but poor feature correlation, indicating transfer-induced spectral distortion rather than simple scaling errors; Scenario 2 shows better feature correlation (intercept = 0.17 vs 0.01); (3) The 60.28% median error confirms that forcing higher-resolution data (SuperLIBS) to mimic lower-resolution characteristics (SuperCam) discards diagnostically critical spectral information. Scenario 2's higher median error (68.07%) paradoxically performs better because it preserves resolution rather than degrading it. These findings collectively establish that successful calibration transfer requires both domain matching and preservation of the higher-fidelity instrument's intrinsic resolution characteristics. Scenario 2's relative success, despite its own limitations, suggests that transfer direction must align with resolution hierarchy, where always transferring from lower to higher resolution systems. There should be more in-depth

analysis to fully understand the limitations of calibration transfer of mineral spectra.

6.4 Concluding Performance Analysis

The benchmark “All” model’s exceptional performance (R^2 values of 0.92–0.98 across all test conditions) demonstrates that LIBS analysis benefits tremendously from diverse, representative training data. By incorporating spectra from multiple instruments (ChemLIBS, SuperLIBS, and pLIBS), atmospheric conditions (Earth, Mars, and Vacuum pressures), and laser parameters, the model learns robust spectral patterns that transcend individual experimental setups. This "universal training" approach proves particularly valuable for Mars missions; despite limited onboard calibration targets, the model can accurately predict compositions even for samples quite different from its training set. However, the results also show clear limits: when test conditions diverge too far from the training domain (specialized calibration targets), even this comprehensive model shows degraded performance, emphasizing that some instrument-specific tuning remains necessary.

The ChemCam results reveal calibration transfer’s double-edged nature. When properly aligned (transferring to match the test domain), it can dramatically improve predictions, reducing errors by nearly 50% in optimal cases. The results from Lepore, K. et al. (2024) demonstrate strong alignment with this study’s results, confirming that calibration transfer significantly improves prediction accuracy when applied to LIBS data across differing atmospheric conditions. However, the SuperCam experiments demonstrate its fragility, transferring between similar-but-not-identical high-resolution instruments often degrades performance more than it helps. Three key lessons emerge: (1) Transfer works best from lower to higher resolution instruments, not vice versa; (2) It is most beneficial when bridging moderate instrumental gaps, not extreme ones; and (3) Its effectiveness varies by element, working better for some oxides (such as SiO_2) than others.

The unexpected failure of Scenario 4 (transferring SuperLIBS to SuperCam) stems from a fundamental mismatch in spectral resolution hierarchy. While SuperLIBS boasts superior resolution (0.08 nm in UV/VIS) compared to SuperCam’s 0.15–0.20 nm, this advantage becomes a liability when attempting inverse calibration transfer. The downsampling

irreversibly loses: (1) narrow emission line shapes critical for mineral identification, and (2) subtle continuum variations encoding sample matrix effects. This explains the 72% performance degradation (RMSE-P increase from 2.85 to 4.68) in Scenario 4 compared to the standard SuperCam→SuperLIBS transfer direction. The results suggest that resolution enhancement has inherent physical limits; while lower-resolution instruments can be upsampled with reasonable success (as shown in Scenario 2), high-resolution data cannot be meaningfully degraded without catastrophic information loss. This has crucial implications for future missions: cross-instrument calibration must always flow from lower to higher resolution systems, highly unlikely for the reverse.

6.5 Future Work

Future work will focus on enhancing the supermodel's predictive capabilities through expanded dataset integration and systematic performance evaluation. The trained "All" model will be augmented by incorporating combined data from both ChemCam and SuperCam instruments, rather than keeping them separate. While both utilize laser-induced breakdown spectroscopy for Martian rock analysis, their distinct detector characteristics may provide complementary spectral information that could improve the PLS model's robustness across instrument configurations.

Further investigation will examine how the model performs when trained exclusively on SuperLIBS data, removing contributions from ChemCam and pLIBS instruments, as there are few studies exploring this aspect. This comparison will reveal the trade-offs between using heterogeneous datasets versus instrument-specific data, helping to determine when the comprehensive "All" approach offers clear advantages over more focused training sets. Additional studies will systematically evaluate prediction accuracy when train and test sets share identical instrumental configurations, which will help establish baseline performance metrics for cross-instrument applications.

A critical direction emerging from Scenario 4's results involves developing advanced spectral matching techniques to address the resolution-dependent transfer challenges between SuperLIBS and SuperCam. The failed inverse calibration (SuperLIBS→SuperCam) reveals

fundamental limitations in current approaches when bridging instruments with different resolution characteristics (0.08 nm vs 0.15-0.20 nm). Future studies should investigate wavelength-specific alignment methods that preserve diagnostically important peak shapes during cross-instrument transfers, particularly for major element emission lines in the UV/VIS range that dominate quantitative analysis. Further investigation could examine how the model performs when trained exclusively on SuperLIBS data, removing contributions from ChemLIBS and pLIBS instruments, as there are few studies exploring this aspect. This comparison will reveal the trade-offs between using heterogeneous datasets versus instrument-specific data, helping to determine when the comprehensive "All" approach offers clear advantages over more focused training sets. Additional studies will systematically evaluate prediction accuracy when train and test sets share identical instrumental configurations, which will help establish baseline performance metrics for cross-instrument applications.

The demonstrated effectiveness of Partial Least Squares (PLS) regression in this study invites comparison with alternative machine learning approaches that may further reduce prediction errors in the "All" model configuration. Building on work by Anderson, R. et al. (2022), three key directions emerge: First, systematic benchmarking of ensemble methods, such as Gradient Boosting), regularized linear models (LASSO), and kernel-based approaches (Support Vector Machines) would establish algorithm-specific performance characteristics across different compositional ranges. Second, hybrid architectures combining these methods could leverage their complementary strengths. For instance, using LASSO for feature selection followed by PLS regression. Third, such analysis should quantify the trade-offs between computational complexity and prediction accuracy inherent to each approach. This comparative methodology study would significantly advance optimization strategies for LIBS supermodels. In terms of PLS effectiveness, this study analyzed five metrics without a systematic way to evaluate an overall one metric value to compare each subscenario. Hence, for future work, a single composite score that combines the RMSE-P, R^2, slope and normalized intercept can be calculated via this equation:

$$\text{Performance Score} = \left(\frac{1}{\text{RMSE-P}}\right) + R^2 + \text{Slope} - |\text{Intercept}|$$

This formula ensures that models with lower prediction error scores (RMSE-P), higher explained variance (R^2), a slope closer to 1 (indicating an ideal calibration), and a normalized intercept closer to 0 (which minimize the bias) receive the highest composite scores. This approach establishes one of the several methods where different metrics are combined to form a final overall metric to fully compare and contrast each subscenario.

Further research should evaluate advanced calibration transfer techniques to better address spectral variations between instruments. Artificial Neural Networks (ANNs) or Convulational Neural Networks (CNNs) present a promising direction, with their capacity to model nonlinear spectral relationships and automatically learn optimal feature representations, particularly valuable for challenging cases like the characteristic Fe_2O_3 peak shifts between basalt and clay matrices. Similarly, Direct Standardization methods available in PyHAT warrant systematic comparison, as they may better preserve peak shapes while maintaining physical interpretability. Building on (Boucher, T. et al. 2015a), future work should quantify the relative performance of these approaches in terms of transfer accuracy (especially for compositionally diverse samples), computational efficiency at scale, and robustness to instrumental noise, ultimately strengthening the supermodel's adaptability across analytical conditions.

A critical final next step involves evaluating the model's generalization capability using novel sample compositions absent from the current 2,157-sample dataset. Although the existing "All" model incorporates hundreds of ChemCam and SuperCam samples, testing with completely independent samples would provide a more rigorous assessment of predictive performance. Comparative analysis should examine similar metrics done in this study, such as prediction accuracy (RMSE-P, R^2), systematic biases through slope deviations from unity, and constant errors via normalized intercepts between the training set and unseen samples. This approach would identify problematic composition ranges, reveal specific geochemical matrices causing prediction errors, and guide targeted improvements, which would be particularly valuable for planetary applications where sample diversity often exceeds terrestrial calibration standards.

7 Conclusion

This study presents the first comprehensive evaluation of a mineral spectroscopic supermodel capable of operating across diverse analytical conditions. The model successfully integrated data from four distinct LIBS systems—10K SuperLIBS, 18K SuperLIBS, ChemLIBS, pLIBS Z300, and pLIBS Z903—operating under Earth, Mars, and vacuum atmospheric conditions, with laser energies ranging from 3 to 9 mJ. Through careful data curation, the spectral and metadata were organized into nine instrument-atmosphere combinations, resulting in a unified dataset of 2,157 unique samples and 49,000+ rows of individual spectra. Each sample appeared in all combinations to ensure that the model could learn the spectral differences across the variations of setups and characteristics. By integrating Earth, Mars, and vacuum-condition spectra, this work advances a unified calibration framework capable of supporting future cross-environmental geochemical analysis that is essential for missions targeting diverse planetary bodies, such as asteroids, the Moon, and Venus, where atmospheric or pressure conditions vary drastically and directly collecting data in these extreme conditions is not feasible at the moment.

Model training was performed using Partial Least Squares (PLS) regression. Evaluation included four distinct data partitioning scenarios, assessing performance both with and without calibration transfer using the Piecewise Direct Standardization with PLS (PDS-PLS) method, using five comprehensive model metrics that many previous studies often overlooked by focusing on just one or two metrics: the RMSE-P, RMSE-P/median %, R^2, slope and normalized intercept.

The comprehensive evaluation of the "All" model's performance across multiple instruments and experimental conditions yields five principal conclusions:

1. **The "All" model establishes a robust benchmark for LIBS analysis** through its exceptional self-prediction capability, achieving an average R^2 values of 0.83 and low RMSE-P average of 1.67. This demonstrates that a unified common model of significant amount of data can effectively capture spectral patterns spanning three atmospheric environments, laser power variations, and multiple instrument configurations. This low

result is compared to previous results to confirm good model generalization by adding more amounts.

2. **Controlled data diversity enhances generalization without significant performance degradation**, as demonstrated by only modest increases in prediction errors (ΔRMSE-P < 0.5) when incorporating the comprehensive "All" dataset versus instrument-specific training. While the ChemCam-only model achieves slightly lower absolute errors (RMSE-P = 2.02), the "All"+ChemCam combination maintains comparable accuracy (RMSE-P = 2.28) while gaining cross-instrument applicability.

3. **Calibration transfer effectiveness varies significantly by instrument**, showing clear benefits for ChemCam (46.6% RMSE-P improvement) but performance degradation for SuperCam (10.3% RMSE-P increase). This dichotomy reveals that transfer success depends on resolution hierarchy and spectral compatibility between systems.

4. **Domain alignment is critical for calibration transfer**, with optimal results occurring only when transfer targets match test conditions. Mismatched domains (such as transferring to SuperLIBS but testing on ChemCam/SuperCam for Scenario 1) produced high errors, while properly aligned scenarios reduced errors by 60% in the average RMSE-P over median percentage for ChemCam.

5. **Data quality investigation emerges as a crucial frontier**, as demonstrated by Scenario 4's results for SuperCam where specialized targets required domain-specific training. Future work must prioritize spectral feature selection and noise characterization over simple dataset expansion to optimize prediction accuracy.

These findings collectively demonstrate that while the "All" model provides a powerful foundation for cross-instrument LIBS analysis, its implementation requires careful consideration of data provenance, instrument characteristics, and target domain requirements. The results establish clear guidelines for balancing comprehensive training with specialized calibration needs in planetary spectroscopy applications.

While calibration transfer techniques like Piecewise Direct Standardization can partially

compensate for atmospheric differences, as evidenced by the significant error reduction from previous papers and in Scenario 2/Scenario 3 for both ChemCam and SuperCam, they cannot replace the value of carefully curated, environment-specific training data. The superior performance of the comprehensive "All" model when combining with ChemCam or SuperCam underscores the importance of developing robust spectral libraries that capture the full range of anticipated analytical environments.

This work makes several important contributions to the field of planetary spectroscopy. It establishes a new benchmark for cross-instrument spectroscopic analysis and shows that, with consistent data preparation and appropriate calibration transfer, integrated models can overcome traditional instrument-specific limitations. Finally, it lays a foundation for future development of robust LIBS models capable of operating across varied planetary environments, where previous work requested to the need for it. The framework developed here represents a significant advancement in machine learning for planetary exploration, particularly for the analysis of Martian samples via laser-induced breakdown spectroscopy.

Appendices

A Methodology

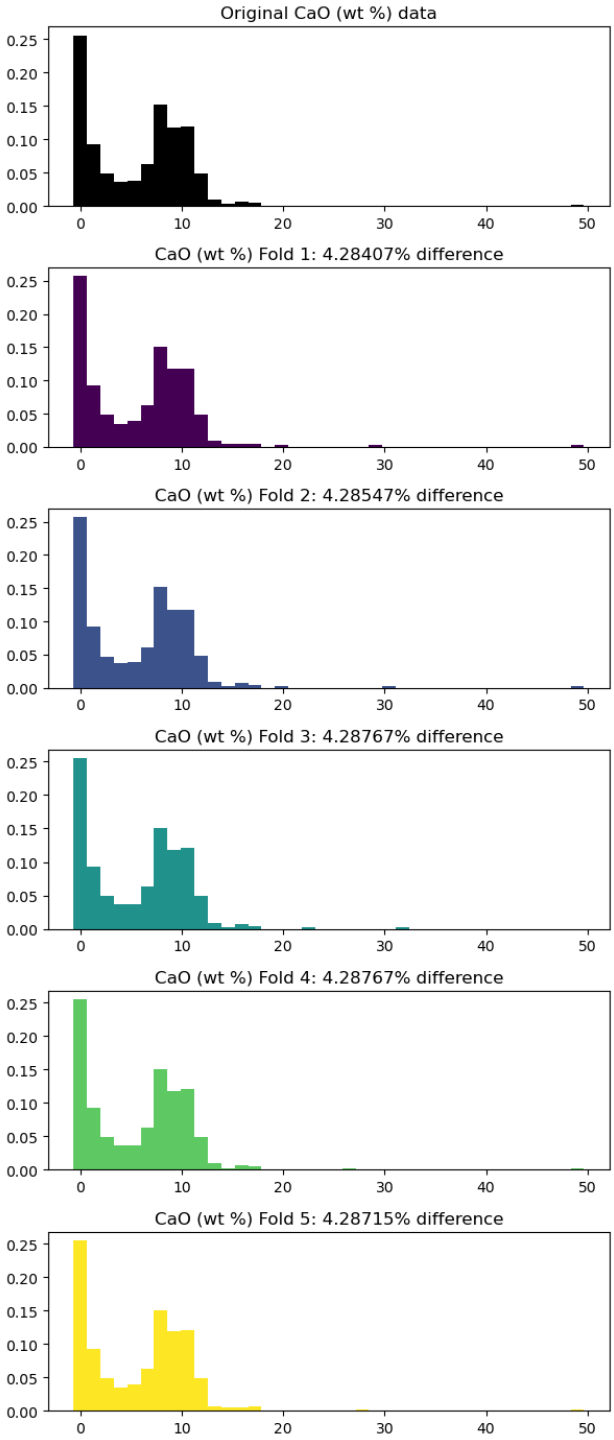

Figure A.1: Getting the Most Representative Fold for CaO Histograms

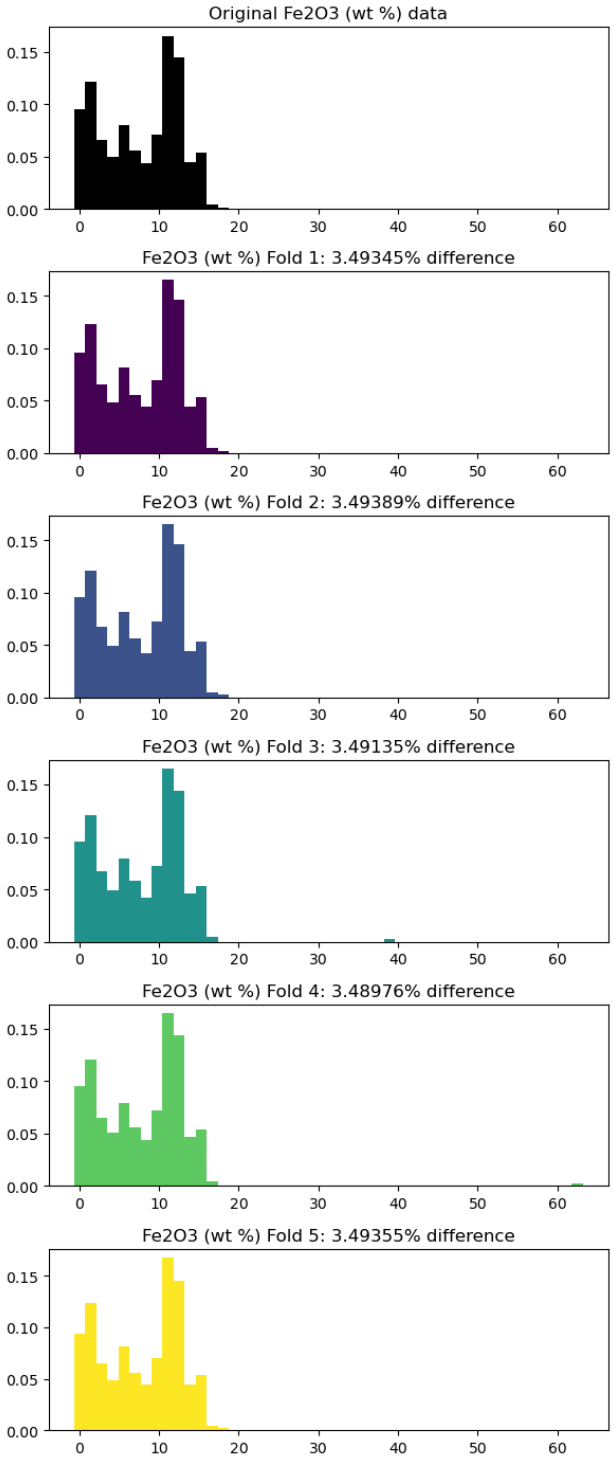

Figure A.2: Getting the Most Representative Fold for Fe_2O_3 Histograms

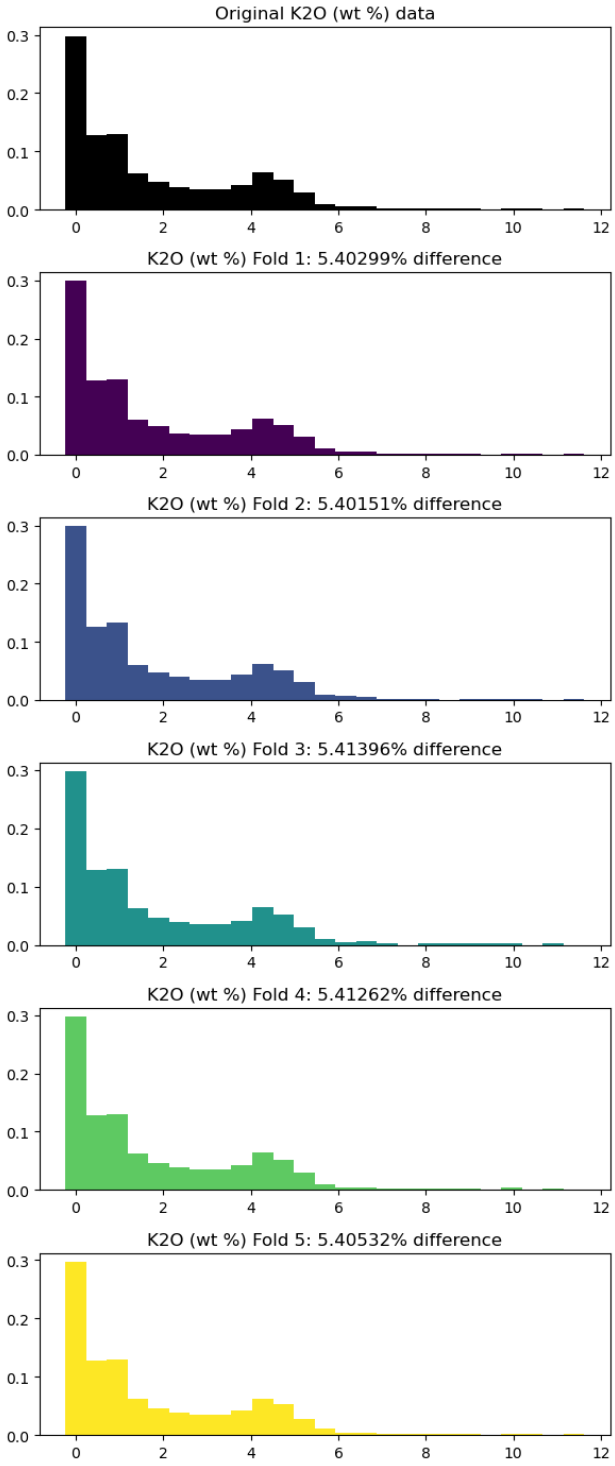

Figure A.3: Getting the Most Representative Fold for K_2O Histograms

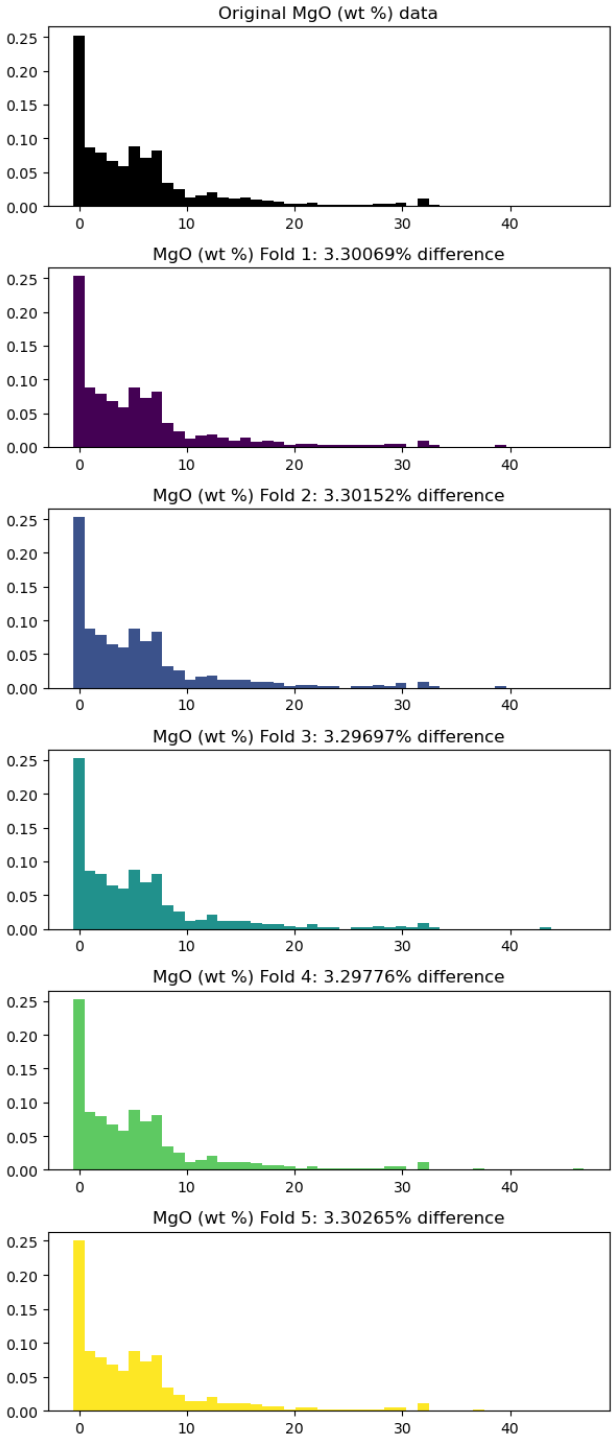

Figure A.4: Getting the Most Representative Fold for MgO Histograms

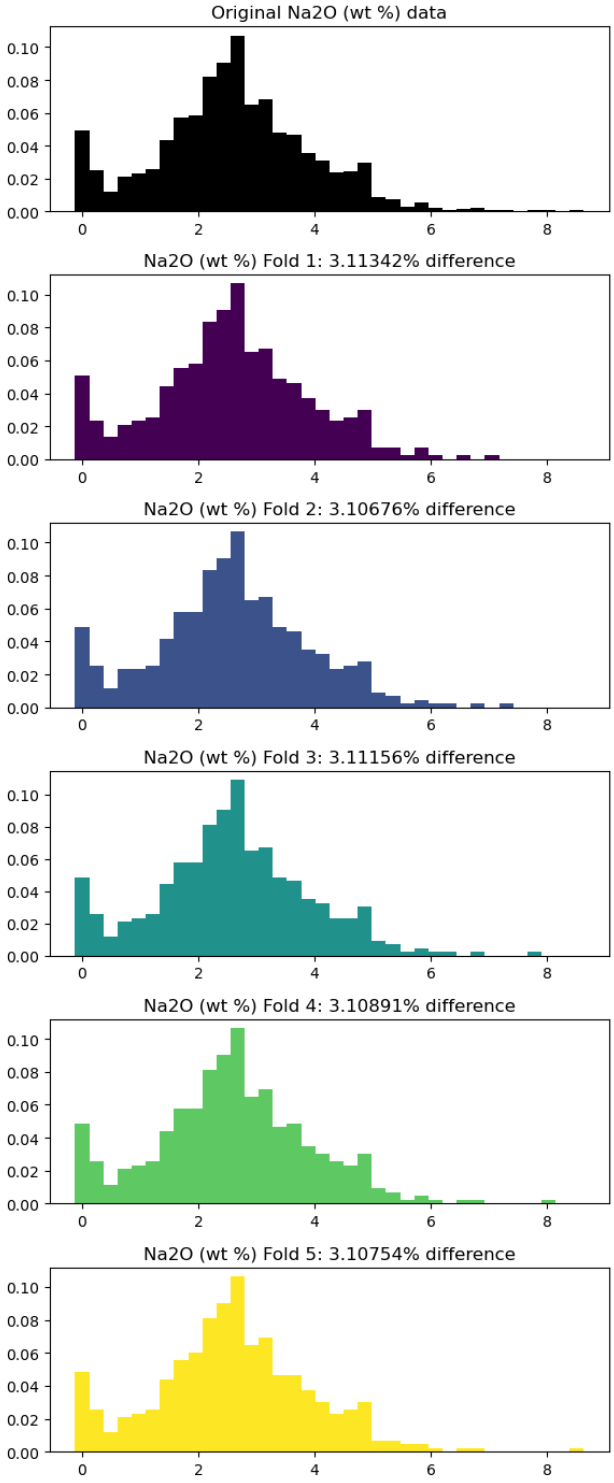

Figure A.5: Getting the Most Representative Fold for Na_2O Histograms

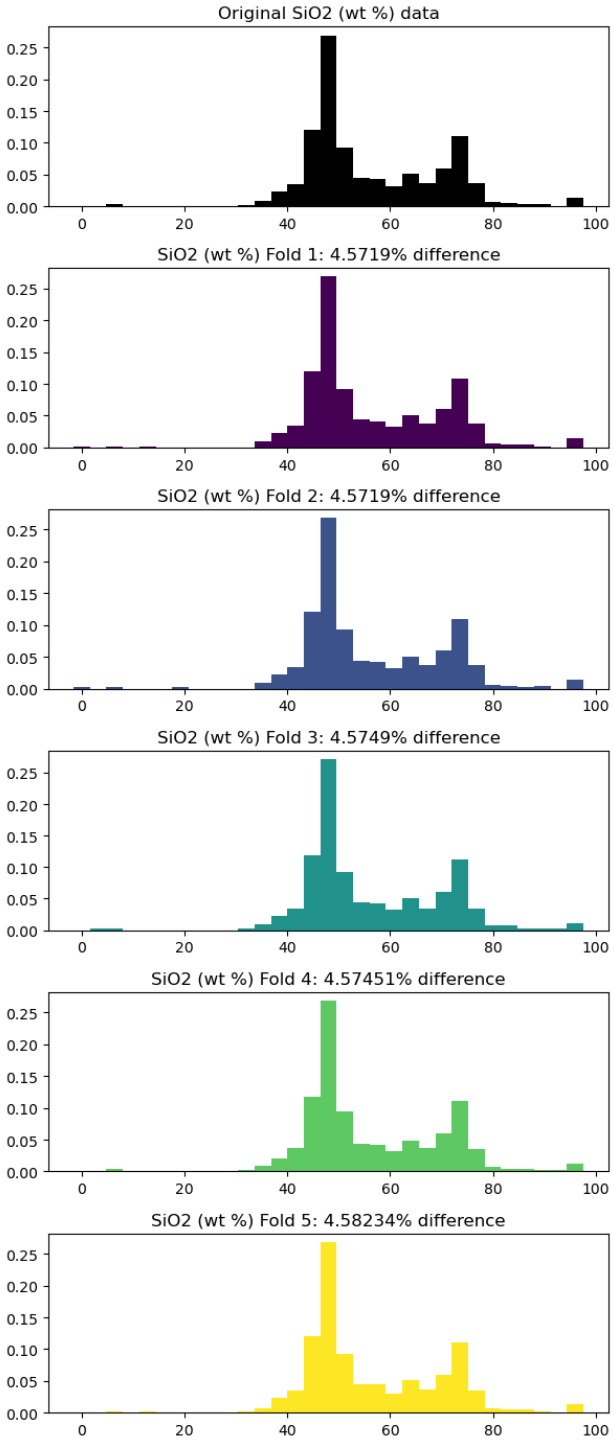

Figure A.6: Getting the Most Representative Fold for SiO_2 Histograms

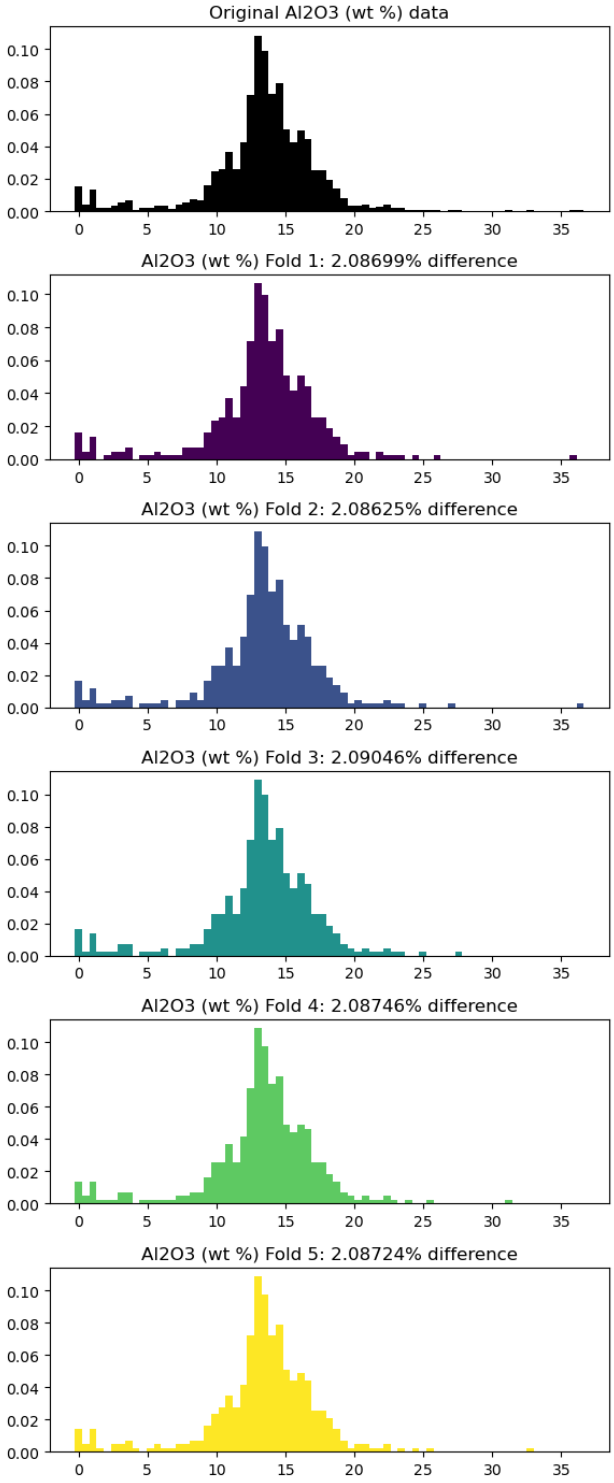

Figure A.7: Getting the Most Representative Fold for Al_2O_3 Histograms

B Results

Table B.1: ChemCam Scenario 0 - All Metrics Combined

Description	Avg RMSE-P	Avg RMSE-P Median (%)	Avg R^2	Avg Slope	Avg Norm Intercept
Train on All Test on All	1.67	27.27	0.83	0.83	0.17
Train on CC test on CC	2.02	37.05	0.81	0.84	0.17
Train on CC CT test on CC CT	1.91	36.83	0.82	0.84	0.16

Table B.2: SuperCam Scenario 0 - All Metrics Combined

Description	Avg RMSE-P	Avg RMSE-P Median (%)	Avg R^2	Avg Slope	Avg Norm Intercept
Train on All Test on All	1.67	27.27	0.83	0.83	0.17
Train on SC test on SC	2.42	50.04	0.74	0.78	0.21
Train on SC CT test on SC CT	2.67	54.24	0.68	0.69	0.26

Table B.3: ChemCam Scenario 1 - All Metrics Combined

Description	Avg RMSE-P	Avg RMSE-P Median (%)	Avg R^2	Avg Slope	Avg Norm Intercept
Train on All test on CC	3.47	45.12	0.54	0.74	0.25
Train on All test on CC CT	3.32	51.87	0.65	0.74	0.27
Train on All+CC test on CC	2.28	35.85	0.73	0.75	0.27
Train on All+CC CT test on CC CT	2.52	34.36	0.70	0.81	0.21

Table B.4: SuperCam Scenario 1 - All Metrics Combined

Description	Avg RMSE-P	Avg RMSE-P Median (%)	Avg R^2	Avg Slope	Avg Norm Intercept
Train on All test on SC	5.62	71.12	0.28	0.57	0.51
Train on All test on SC CT	3.92	59.10	0.38	0.68	0.42
Train on All+SC test on SC	2.34	37.00	0.72	0.70	0.27
Train on All+SC CT test on SC CT	2.78	42.51	0.66	0.64	0.31

Table B.5: ChemCam Scenario 2 - All Metrics Combined

Description	Avg RMSE-P	Avg RMSE-P Median (%)	Avg R^2	Avg Slope	Avg Norm Intercept
Train on All test on SL 10k Mars	1.57	25.00	0.86	0.85	0.15
Train on CC test on SL 10k Mars	4.64	82.39	0.27	0.83	0.45
Train on CC CT test on SL 10k Mars	2.48	57.96	0.59	0.75	0.32
Train on All+CC test on SL 10k Mars	1.63	27.24	0.84	0.84	0.16
Train on All+CC CT test on SL 10k Mars	1.57	25.53	0.85	0.86	0.14

Table B.6: SuperCam Scenario 2 - All Metrics Combined

Description	Avg RMSE-P	Avg RMSE-P Median (%)	Avg R^2	Avg Slope	Avg Norm Intercept
Train on All test on SL 10k Mars	1.57	25.00	0.86	0.85	0.15
Train on SC test on SL 10k Mars	7.33	95.06	0.22	0.40	0.57
Train on SC CT test on SL 10k Mars	2.85	68.07	0.49	0.65	0.17
Train on All+SC test on SL 10k Mars	1.64	27.53	0.84	0.82	0.18
Train on All+SC CT test on SL 10k Mars	1.70	29.47	0.82	0.81	0.19

Table B.7: ChemCam Scenario 3 - All Metrics Combined

Description	Avg RMSE-P	Avg RMSE-P Median (%)	Avg R^2	Avg Slope	Avg Norm Intercept
Train on CC test on Mars Cal Targ	2.24	49.60	0.05	0.86	1.25
Train on All+CC test on Mars Cal Targ	2.17	34.73	0.07	0.39	1.41
Train on CC+Cal Targ test on Mars Cal Targ	1.16	27.88	0.32	1.14	−0.48
Train on All+CC+Cal Targ test on Mars Cal Targ	1.45	18.26	0.23	0.42	0.88

Table B.8: SuperCam Scenario 3 - All Metrics Combined

Description	Avg RMSE-P	Avg RMSE-P Median (%)	Avg R^2	Avg Slope	Avg Norm Intercept
Train on SC test on Mars Cal Targ	5.48	104.58	0.36	0.57	0.61
Train on All+SC test on Mars Cal Targ	4.61	71.37	0.41	0.71	0.43
Train on SC+Cal Targ test on Mars Cal Targ	6.11	70.88	0.28	0.54	0.75
Train on All+SC+Cal Targ test on Mars Cal Targ	3.89	70.06	0.49	0.85	0.24
Train on SL 10k/18k Mars+SC+Cal Targ test on Mars Cal Targ	3.82	58.16	0.46	0.80	0.42

Table B.9: SuperCam Scenario 4 - All Metrics Combined

Description	Avg RMSE-P	Avg RMSE-P Median (%)	Avg R^2	Avg Slope	Avg Norm Intercept
Train on SL 10k/18k Mars test on SC	7.53	100.50	0.09	0.80	0.34
Train on SL 10k/18k Mars CT SC test on SC	4.68	60.28	0.40	0.89	0.01

References

Fabre, C. et al. (2014). "In situ calibration using univariate analyses based on the onboard ChemCam targets: first prediction of Martian rock and soil compositions". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854714000901?via%3Dihub.

Laura Bush (2019). "LIBS: Fundamentals, Benefits, and Advice for New Users". In: *Spectroscopy Online*. URL: https://www.spectroscopyonline.com/view/libs-fundamentals-benefits-and-advice-new-users.

Cousin, A. et al. (2011). "Laser induced breakdown spectroscopy library for the Martian environment". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854711002199?pes=vor&utm_source=wiley&getft_integrator=wiley.

Dyar, M.D. et al. (2012). "Comparison of partial least squares and lasso regression techniques as applied to laser-induced breakdown spectroscopy of geological samples". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S058485471200095X.

Fu, H. et al. (2017). "Calibration Methods of Laser-Induced Breakdown Spectroscopy". In: *IntechOpen*. URL: https://www.intechopen.com/chapters/58688.

Ytsma, C. & Dyar, M.D. (2019). "Accuracies of lithium, boron, carbon, and sulfur quantification in geological samples with laser-induced breakdown spectroscopy in Mars, Earth, and vacuum conditions". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854719301089?via%3Dihub.

Rammelkamp, K. et al. (2021). "Clustering Supported Classification of ChemCam Data From Gale Crater, Mars". In: *Earth and Space Science*. URL: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2021EA001903.

Cousin, A. et al. (2025). "Soil diversity at Jezero crater and Comparison to Gale crater, Mars". In: *Icarus*. URL: https://doi.org/10.1016/j.icarus.2024.116299.

Gottlieb, C. et al. (2017). "2D evaluation of spectral LIBS data derived from heterogeneous materials using cluster algorithm". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854717300757.

Clegg, S.M. et al. (2007). "Quantitative analysis of samples with variable composition by remote laser-induced breakdown spectroscopy." In: *Seventh International Conference on Mars*. URL: https://www.lpi.usra.edu/meetings/7thmars2007/pdf/3216.pdf.

Tucker, J.M. et al. (2010). "Optimization of laser-induced breakdown spectroscopy for rapid geochemical analysis". In: *Chemical Geology*. URL: https://www.sciencedirect.com/science/article/pii/S0009254110002664?via%3Dihub.

Sun, C. et al. (2019). "Quantitative LIBS analysis using advanced data modeling". In: *Scientific Reports*. URL: https://www.nature.com/articles/s41598-019-47751-y.

Bilge, G. et al. (2021). "Comparison of different calibration techniques of laser induced breakdown spectroscopy in bakery products: on NaCl measurement". In: *Journal of the European Optical Society-Rapid Publications*. URL: https://jeos.springeropen.com/articles/10.1186/s41476-021-00164-9.

Anderson, R. et al. (2012). "Clustering and training set selection methods for improving the accuracy of quantitative laser induced breakdown spectroscopy". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854712000882?via%3Dihub.

Ciucci, A. et al. (1999). "New Procedure for Quantitative Elemental Analysis by Laser-Induced Plasma Spectroscopy". In: *Applied Spectroscopy*. URL: https://journals.sagepub.com/doi/10.1366/0003702991947612.

Boucher, T. et al. (2015a). "A study of machine learning regression methods for major elemental analysis of rocks using laser-induced breakdown spectroscopy". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854715000518?via%3Dihub.

Ytsma, C. & Dyar, M.D (2018). "Effects of univariate and multivariate regression on the accuracy of hydrogen quantification with laser-induced breakdown spectroscopy". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854717303038#bb0130.

IBM (2023). *Topics in Machine Learning*. URL: https://www.ibm.com/think/topics/machine-learning.

Wiens, R.C. et al. (2012). "The ChemCam Instrument Suite on the Mars Science Laboratory (MSL) Rover: Body Unit and Combined System Tests". In: *Space Science Reviews*. URL: https://link.springer.com/article/10.1007/s11214-012-9902-4.

Runge, E.F. et al. (1963). "Spectrochemical analysis using a pulsed laser source". In: *Spectrochimica Acta*. URL: https://www.sciencedirect.com/sdfe/pdf/download/eid/1-s2.0-0371195164800709/first-page-pdf.

Clegg, S.M. et al. (2014). "Planetary Geochemical Investigations Using Raman and Laser-Induced Breakdown Spectroscopy". In: *Applied Spectroscopy*. URL: https://journals.sagepub.com/doi/10.1366/13-07386.

Lasue, J. et al. (2012). "Remote laser-induced breakdown spectroscopy (LIBS) for lunar exploration". In: *Journal of Geophysical Research*. URL: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2011JE003898.

Pavlov, S.G. et al. (2011). "Miniaturized laser-induced plasma spectrometry for planetary in situ analysis – The case for Jupiter's moon Europa". In: *Advances in Space Research*. URL: https://www.sciencedirect.com/science/article/pii/S0273117710004527?via%3Dihub.

– (2012). "Low-energy laser induced breakdown spectroscopy for in-situ space missions to solar system bodies without atmospheres". In: *Planetary and Space Science*. URL: https://www.sciencedirect.com/science/article/pii/S0032063312001924?pes=vor&utm_source=wiley&getft_integrator=wiley.

Maurice, S. et al. (2016). "ChemCam Activities and Discoveries during the Nominal Mission of Mars Science Laboratory in Gale crater, Mars". In: *Journal of Analytical Atomic Spectrometry*. URL: https://www.researchgate.net/publication/295099129_ChemCam_Activities_and_Discoveries_during_the_Nominal_Mission_of_Mars_Science_Laboratory_in_Gale_crater_Mars.

Kim, Y. et al. (2022). "Access to quantitative analysis of carbonates using a portable LIBS instrument: First applications to single minerals and mineral mixtures". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854722000416.

Ferreira, D.S. et al. (2024). "Laser-induced breakdown spectroscopy (LIBS): calibration challenges, combination with other techniques, and spectral analysis using data science". In: *Journal of Analytical Atomic Spectrometry*. URL: https://pubs.rsc.org/en/content/articlelanding/2024/ja/d4ja00250d.

Chen, T. et al. (2020). "Applications of laser-induced breakdown spectroscopy (LIBS) combined with machine learning in geochemical and environmental resources exploration". In: *TrAC Trends in Analytical Chemistry*. URL: https://www.sciencedirect.com/science/article/pii/S0165993620303423#bib25.

Andor (2025). "An introduction to Laser Induced Breakdown Spectroscopy-LIBS". In: *Oxford Instruments*. URL: https://andor.oxinst.com/learning/view/article/libs-an-overview-of-andors-solutions-for-libs.

L'Hermite, D. et al. (2016). "Evaluation of the efficacy of a portable LIBS system for detection of CWA on surfaces". In: *Environmental Science and Pollution Research*. URL: https://doi.org/10.1007/s11356-016-6305-1.

Sirven, J.B. et al. (2004). "Time-resolved and time-integrated single-shot laser-induced plasma experiments using nanosecond and femtosecond laser pulses". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854704001375.

Steven Buckley (2014). "LIBS Basics, Part I: Measurement Physics and Implementation". In: *Spectroscopy Online*. URL: https://www.spectroscopyonline.com/view/libs-basics-part-i-measurement-physics-and-implementation.

Guedes, W.N. et al. (2023). "Evaluation of univariate and multivariate calibration strategies for the direct determination of total carbon in soils by laser-induced breakdown spectroscopy: tutorial". In: *Journal of the Optical Society of America B*. URL: https://opg.optica.org/josab/fulltext.cfm?uri=josab-40-5-1319&id=530256.

Harmon, R.S. & Senesi, G.S (2021). "Laser-Induced Breakdown Spectroscopy – A geochemical tool for the 21st century". In: *Applied Geochemistry*. URL: https://www.sciencedirect.com/science/article/pii/S0883292721000615?via%3Dihub#mmc1.

Li, L.N. et al. (2021). "A review of artificial neural network based chemometrics applied in laser-induced breakdown spectroscopy analysis". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854721001403.

Grünberger, S. et al. (2023). "Overcoming the matrix effect in the element analysis of steel: Laser ablation-spark discharge-optical emission spectroscopy (LA-SD-OES) and Laser-induced breakdown spectroscopy (LIBS)". In: *Analytica Chimica Acta*. URL: https://www.sciencedirect.com/science/article/pii/S000326702300226X.

Matsumura, T. et al. (2024). "High-Throughput Calibration-Free Laser-Induced Breakdown Spectroscopy". In: *ACS Earth and Space Chemistry*. URL: https://pubs.acs.org/doi/10.1021/acsearthspacechem.4c00067.

Lepore, K. et al. (2017). "Matrix Effects in Quantitative Analysis ofLaser-Induced Breakdown Spectroscopy(LIBS) of Rock Powders Doped with Cr,Mn, Ni, Zn, and Co". In: *Applied Spectroscopy*. URL: https://journals.sagepub.com/doi/epdf/10.1177/0003702816685095?src=getftr%5C&utm_source=sciencedirect_contenthosting%5C&getft_integrator=sciencedirect_contenthosting.

Cristoforetti, G. et al. (2013). "Thermodynamic equilibrium states in laser-induced plasmas: From the general case to laser-induced breakdown spectroscopy plasmas". In: *Spectrochimica*

Acta Part B: Atomic Spectroscopy. URL: https://www.sciencedirect.com/science/article/pii/S0584854713002504.

Wiens, R.C. et al. (2013). "Pre-flight calibration and initial data processing for the ChemCam laser-induced breakdown spectroscopy instrument on the Mars Science Laboratory rover". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854713000505.

Nancy McMillan (2018). *Laser-Induced Breakdown Spectroscopy (LIBS)*. URL: https://serc.carleton.edu/msu_nanotech/methods/libs.html.

Castro, J.P. et al. (2020). "Calibration strategies for the direct determination of rare earth elements in hard disk magnets using laser-induced breakdown spectroscopy". In: *Talanta*. URL: https://www.sciencedirect.com/science/article/pii/S0039914019310768?via%3Dihub.

Herreyre, N. et al. (2023). "Artificial neural network for high-throughput spectral data processing in LIBS imaging: application to archaeological mortar". In: *Journal of Analytical Atomic Spectrometry*. URL: https://pubs.rsc.org/en/content/articlelanding/2023/ja/d2ja00389a.

Fernández-Manteca, M.G. et al. (2023). "Comparison of light capturing approaches in Laser-Induced Breakdown Spectroscopy (LIBS) for multichannel spectrometers". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854723000046.

Sirven, J.B. et al. (2006). "Laser-Induced Breakdown Spectroscopy of Composite Samples: Comparison of Advanced Chemometrics Methods". In: *Analytical Chemistry*. URL: https://pubs.acs.org/doi/10.1021/ac051721p.

Motto-Ros, V. et al. (2007). "Quantitative multi-elemental laser-induced breakdown spectroscopy using artificial neural networks". In: *Journal of the European Optical Society-Rapid Publications*. URL: https://jeos.edpsciences.org/articles/jeos/abs/2008/01/jeos20080308011/jeos20080308011.html.

Zhu, X. et al. (2014). "Advanced statistical analysis of laser-induced breakdown spectroscopy data to discriminate sedimentary rocks based on Czerny–Turner and Echelle spectrometers". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854714000020?via%3Dihub.

Dyar, M.D. et al. (2016). "Comparison of univariate and multivariate models for prediction of major and minor elements from laser-induced breakdown spectra with and without masking". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854716301197.

Sirven, J.B. et al. (2007). "Feasibility study of rock identification at the surface of Mars by remote laser-induced breakdown spectroscopy and three chemometric methods". In: *Journal of Analytical Atomic Spectrometry*. URL: https://pubs.rsc.org/en/content/articlelanding/2007/ja/b704868h.

Clegg, S.M. et al. (2017). "Recalibration of the Mars Science Laboratory ChemCam instrument with an expanded geochemical database". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://sciencedirect.com/science/article/pii/S0584854716303913?via%3Dihub.

– (2009). "Multivariate analysis of remote laser-induced breakdown spectroscopy spectra using partial least squares, principal component analysis, and related techniques". In:

Spectrochimica Acta Part B: Atomic Spectroscopy. URL: https://www.sciencedirect.com/science/article/pii/S0584854708003443.

Luque-García, J.L. et al. (2002). "Determination of the major elements in homogeneous and heterogeneous samples by tandem laser-induced breakdown spectroscopy–partial least square regression". In: *Microchemical Journal*. URL: https://www.sciencedirect.com/science/article/pii/S0026265X02001248?via%3Dihub.

Jan Gerretzen (2016). *Chemical data preprocessing. From art to science*. URL: https://repository.ubn.ru.nl/bitstream/handle/2066/168804/168804.pdf.

Dante M. Pirouz (2010). "An Overview of Partial Least Squares". In: *SSRN*. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1631359.

Dyar, M.D. et al. (2011). "Strategies for Mars remote Laser-Induced Breakdown Spectroscopy analysis of sulfur in geological samples". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854710003034.

Aylin Alin (2009). "Comparison of PLS algorithms when number of objects is much larger than number of variables". In: *Statistical Papers*. URL: https://link.springer.com/article/10.1007/s00362-009-0251-7.

Liu, S. et al. (2025). "LIBSFormer: Enhancing Mars in-situ LIBS data analysis with accurate and interpretable quantification of oxides". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854725000898#bb0120.

Boucher, T. et al. (2015b). "Aligning Mixed Manifolds". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. URL: https://people.cs.umass.edu/~mahadeva/papers/boucher_aaai15.pdf.

Anderson, R. et al. (2022). "Post-landing major element quantification using SuperCam laser induced breakdown spectroscopy". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854721003049.

Jerome Workman, Jr. & Howard Mark (2017). "Transfer in Chemometrics, Part I: A Review of the Subject". In: *Spectroscopy Online*. URL: https://www.spectroscopyonline.com/view/transfer-chemometrics-part-i-review-subject-0.

Lepore, K. et al. (2024). "Direct comparison of Earth, Mars, and lunar quantification accuracies derived from laser-induced breakdown spectra of rocks". In: *Spectrochimica Acta Part B: Atomic Spectroscopyl*. URL: https://www.sciencedirect.com/science/article/pii/S0584854724001149.

Dyar, M.D. et al. (2024). "Geochemistry by Laser-Induced Breakdown Spectroscopy on the Moon: Accuracy, Detection Limits, and Realistic Constraints on Interpretations". In: *Earth and Space Science*. URL: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2024EA003635.

Dyar, M.D., Ytsma, C.R., Lepore, K. (2019). "Standards for Geochemical Analysis of Major, Minor, and Trace Elements in Rock Powders". In: *Lunar and Planetary Science Conference*. URL: https://www.hou.usra.edu/meetings/lpsc2019/pdf/1396.pdf.

Rhodes, J.M. & Vollinger, M.J. (2004). "MComposition of basaltic lavas sampled by phase-2 of the Hawaii Scientific Drilling Project: Geochemical stratigraphy and magma types". In: *Geochemistry, Geophysics, Geosystems*. URL: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2002GC000434.

Dyar, M.D & Ytsma, C. (2021). "Effect of data set size on geochemical quantification accuracy with laser-induced breakdown spectroscopy". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854721000094.

Body, D. & Chadwick, B.L. (2001). "Optimization of the spectral data processing in a LIBS simultaneous elemental analysis system". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854701001860.

Knight, A.K. et al. (2000). "Characterization of Laser-Induced Breakdown Spectroscopy(LIBS) for Application to Space Exploration". In: *Applied Spectroscopy*. URL: https://journals.sagepub.com/doi/epdf/10.1366/000370200194959.

Prasetya, O.D. et al. (2018). "Laser-induced breakdown spectroscopy for identification and characterization of aluminum". In: *Journal of Physics: Conference Series*. URL: https://iopscience.iop.org/article/10.1088/1742-6596/1025/1/012012/pdf.

Idris, N. et al. (2017). "Study on Emission Spectral Lines of Iron, Fe in Laser-Induced Breakdown Spectroscopy (LIBS) on Soil Samples". In: *Journal of Physics: Conference Series*. URL: https://iopscience.iop.org/article/10.1088/1742-6596/846/1/012020.

Isolation Forests (n.d.). URL: https://cran.r-project.org/web/packages/isotree/vignettes/An_Introduction_to_Isolation_Forests.html.

Lepore, K., Dyar, M.D., Ytsma, C.R. (2024). "Sharing calibration information among laser-induced breakdown spectroscopy instruments using spectral line binning and calibration transfer". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854723002264.

Lepore, K., Ytsma C.R., Dyar, M.D. (2024). "Application of Post-Processing Methods Used to Improve LIBS Calibration Datasets". In: *Lunar and Planetary Science Conference*. URL: https://www.hou.usra.edu/meetings/lpsc2024/pdf/1084.pdf.

Lepore, K. et al. (2018). "SuperLIBS: A high-capacity laser-induced breakdown spectroscopy system analogous to SuperCam Mars 2020". In: *Lunar and Planetary Science Conference*. URL: https://www.hou.usra.edu/meetings/lpsc2018/pdf/1179.pdf.

Lepore, K., Ytsma C.R., Dyar, M.D. (2022). "Quantitative prediction accuracies derived from laser-induced breakdown spectra using optimized multivariate submodels". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854722000520.

Lepore, K., Dyar, M.D., Ytsma, C.R. (2023). "Effect of Plasma Temperature on Major Element Prediction Accuracy From Laser-Induced Breakdown Spectroscopy". In: *Geophysical Research Letters*. URL: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2023GL102919.

Ytsma, C.R. & Dyar, M.D. (2022). "Calculations of and effects on quantitative limits for multivariate analyses of geological materials with laser-induced breakdown spectroscopy". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854722000398?via%3Dihub.

Schlatter, N. & Lottermoser, B.G. (2023). "Quantitative analysis of Li, Na and K in single-element standard solutions using portable laser-induced breakdown spectroscopy". In: *Geochemistry: Exploration, Environment, Analysis*. URL: https://pubs.geoscienceworld.org/gsl/geea/article/23/2/geochem2023-019/623152/Quantitative-analysis-of-Li-Na-and-K-in-single.

Wise, M.A. et al. (2022). "Handheld LIBS for Li Exploration: An Example from the Carolina Tin-Spodumene Belt, USA". In: *Minerals*. URL: https://www.mdpi.com/2075-163X/12/1/77.

Lepore, K.H. et al. (2023). "A Multi-Paramter LIBS Reference Database of Geological Materials: Opportunities for the Planetary Science Community". In: *Lunar and Planetary Science Conference*. URL: https://www.hou.usra.edu/meetings/lpsc2023/pdf/1048.pdf.

NASA (2024). "Mars Science Laboratory: Curiosity Rover Science Instruments". In: *National Aeronautics and Space Administration*. URL: https://science.nasa.gov/mission/msl-curiosity/science-instruments/.

Alex Bolshakov (n.d.). "LIBS At Work On Mars". In: *Applied Spectra* (). URL: https://appliedspectra.com/mars-libs.html.

Payré, V. et al. (2017). "Alkali trace elements in Gale crater, Mars, with ChemCam: Calibration update and geological implications". In: *Journal of Geophysical Research: Planets*. URL: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2016je005201.

Cousin, A. et al. (2022). "SuperCam calibration targets on board the perseverance rover: Fabrication and quantitative characterization". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854721002986.

Maurice, S. et al. (2021). "The SuperCam Instrument Suite on the Mars 2020 Rover: Science Objectives and Mast-Unit Description". In: *Space Science Reviews*. URL: https://link.springer.com/article/10.1007/s11214-021-00807-w.

Ytsma, C.R., Dyar, M.D., Lepore, K. (2023). "Update to Mars-based major element quantification accuracies from calibration targets of ChemCam at 3013 sols and SuperCam at 527 sols". In: *ESS Open Archive*. URL: https://essopenarchive.org/users/587547/articles/625167-update-to-mars-based-major-element-quantification-accuracies-from-calibration-targets-of-chemcam-at-3013-sols-and-supercam-at-527-sols.

Fabre, C. et al. (2011). "Onboard calibration igneous targets for the Mars Science Laboratory Curiosity rover and the Chemistry Camera laser induced breakdown spectroscopy instrument". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854711000632.

Vaniman, D. et al. (2022). "Ceramic ChemCam Calibration Targets on Mars Science Laboratory". In: *Space Science Reviews*. URL: https://link.springer.com/article/10.1007/s11214-012-9886-0.

Rieder, R. et al. (2003). "The new Athena alpha particle X-ray spectrometer for the Mars Exploration Rovers". In: *Journal of Geophysical Research: Planets*. URL: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2003JE002150.

Klingelhöfer, G. et al. (2003). "Athena MIMOS II Mössbauer spectrometer investigation". In: *Journal of Geophysical Research: Planets*. URL: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2003JE002138.

Allwood, A. et al. (2020). "PIXL: Planetary Instrument for X-ray Lithochemistry". In: *Space Science Reviews*. URL: https://link.springer.com/article/10.1007/s11214-020-00767-7.

Christensen, P.R. et al. (2003). "Miniature Thermal Emission Spectrometer for the Mars Exploration Rovers". In: *Journal of Geophysical Research: Planets*. URL: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2003JE002117.

Anderson, R.B. and Aneece, I.P. (2022). "Python Hyperspectral Analysis Tool (PyHAT) user guide". In: *U.S. Geological Survey Open-File Report*. DOI: 10.3133/ofr2022xxxx. URL: https://doi.org/10.3133/ofr2022xxxx.

Anaconda Inc. (2025). *Toolbox for Notebooks*. https://www.anaconda.com/products/notebooks. Accessed: [4/2/2025].

Subtracting dark spectra (n.d.). URL: https://ibsen.com/resources/detector-resources/subtracting-dark-spectra/.

Sobran, P. et al. (2012). "Extraction of compositional and hydration information of sulfates from laser-induced plasma spectra recorded under Mars atmospheric conditions — Implications for ChemCam investigations on Curiosity rover". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854712000043.

Zhang, Z.M et al. (2010). "Baseline correction using adaptive iteratively reweighted penalized least squares". In: *Analyst*. URL: https://pubmed.ncbi.nlm.nih.gov/20419267/.

Wang, Y. et al. (1991). "Multivariate instrument standardization". In: *Analytical Chemistry*. URL: https://pubs.acs.org/doi/10.1021/ac00023a016.

Zhang, X. et al. (2022). "Transfer of a calibration model for the prediction of lignin in pulpwood among four portable near infrared spectrometers". In: *Journal of Near Infrared Spectroscopy*. URL: https://journals.sagepub.com/doi/full/10.1177/09670335221110013.

Ejiro Onose (2023). *R Squared: Understanding the Coefficient of Determination*. URL: https://cran.r-project.org/web/packages/isotree/vignettes/An_Introduction_to_Isolation_Forests.html.

Valentina Alto (2025). *Understanding Ordinary Least Squares (OLS) Regression*. URL: https://builtin.com/data-science/ols-regression.

Shittu Olumide (2023). "Root Mean Square Error (RMSE) In AI: What You Need To Know". In: *Arize*. URL: https://arize.com/blog-course/root-mean-square-error-rmse-what-you-need-to-know/#:~:text=A%200%20value%20indicates%20that,mistakes%20and%20fewer%20accurate%20forecasts..

Sagawa, S. et al. (2020). "An Investigation of Why Overparameterization Exacerbates Spurious Correlations". In: *Proceedings of Machine Learning Research*. URL: https://proceedings.mlr.press/v119/sagawa20a.html.

Shabbir, S. et al. (2022). "Machine learning and transfer learning for correction of the chemical and physical matrix effects in the determination of alkali and alkaline earth metals with LIBS in rocks". In: *Spectrochimica Acta Part B: Atomic Spectroscopy*. URL: https://www.sciencedirect.com/science/article/pii/S0584854722001227.