

TDK-thesis

Sebastián Ignacio Arriagada Silva

Elitist Bacterial Memetic Algorithm and Fuzzy Logic Integration for Spectral Deconvolution and Mineral Classification in the NIR-SWIR Range

EÖTVÖS LORÁND UNIVERSITY

FACULTY OF INFORMATICS

ARTIFICIAL INTELLIGENCE DEPARTMENT



Author:

Sebastián Ignacio Arriagada Silva

Computer Science MSc

IV. grade

Supervisor:

Márk Domonkos

PhD Student

Budapest, 2024

Abstract

Mineral identification from spectral data is challenging due to complex absorption features and data variability. This thesis introduces the Elitist Bacterial Memetic Algorithm for Spectral Deconvolution (EBMA-SD), integrating spectral decomposition with fuzzy logic-based classification. EBMA-SD decomposes near-infrared (NIR) and short-wave infrared (SWIR) reflectance spectra into a background envelope minus the sum of Voigt profiles, represented as chromosomes of bacteria derived from database samples. Focusing on the highest-performing bacteria, the algorithm improves solutions through memetic local search and horizontal chromosome transfer.

A fuzzy logic system classifies decomposed spectra by comparing extracted features with a reference mineral database, effectively managing uncertainties inherent in spectral data. The methodologies were evaluated using 367 samples from 78 minerals in the USGS database. The EBMA-SD achieved high convergence for 58% of USGS minerals, utilizing a diverse database with only one sample per mineral. Furthermore, in cases where the mineral was absent from the database, it predicted the mineral type for 70% of the samples. The proposed algorithm effectively generalizes across more than 82% of mineral types, including carbonates, phosphates, and subgroups of silicates. However, it faces challenges with oxides, sulfates, and minerals like quartz, as these lack distinctive peaks within the analyzed range.

Acknowledgements

I would like to express my deepest gratitude and dedicate this work to my loved ones. To my family, who, through their immense effort, have always encouraged me to persevere in the face of adversity, offering their unconditional love and support, and enabling me to seize this unique opportunity to live abroad. Above all, I dedicate this to my uncle, Victor Hugo Silva Herrera, who sadly passed away while I was here, but whose warmth and spirit will forever remain with us.

I also wish to extend heartfelt thanks to the remarkable persons I had the privilege of meeting during this journey, particularly my partner, Bernadett Ignácz, who has been an indispensable pillar, making it possible for me to fully embrace and enjoy this experience.

Moreover, I am profoundly grateful for the invaluable guidance and support of my advisor, Domonkos Márk, whose academic mentorship and support made this work possible. I am equally thankful to professors like Dr. Janos Botzheim, whose inspiring lectures reignited my passion for academia.

Finally, I would like to express my sincere appreciation to *Graiph*'s team for their steadfast support from the very beginning, which has enabled me to grow and develop both personally and professionally.

Contents

1	Introduction	4
1.1	Overview	4
1.2	Motivation	6
1.3	Problem Statement	6
1.4	Proposed Solution	7
2	Literature Review	9
2.1	Spectroscopy in Mineralogy	9
2.1.1	Fundamental Principles	9
2.1.2	Instrumentation Considerations	11
2.1.3	Sample considerations	13
2.2	Spectral Deconvolution	16
2.2.1	Continuum Removal	17
2.2.2	Deconvolution Components	18
2.2.3	Current Deconvolution Techniques	20
2.3	Bacterial Algorithms	21
2.3.1	Bacterial Evolutionary Algorithms	21
2.3.2	Bacterial Memetic Algorithms	23
2.4	Elitist Strategy in Genetic Algorithms	23
2.5	Fuzzy Logic	24
2.5.1	Fundamental Concepts	24
2.5.2	Mathematical Formulation	25
2.5.3	Fuzzy Logic in Evolution Adaptive Parameter Control	25
2.5.4	Fuzzy Logic in Mineral Classification	26
2.6	Geology	27
2.6.1	Actinolite	27
2.6.2	Alunite	28

2.6.3	Amphibole (Hornblende)	28
2.6.4	Calcite	28
2.6.5	Chlorite	28
2.6.6	Dickite	28
2.6.7	Dolomite	29
2.6.8	Epidote	29
2.6.9	Garnet (Grossular)	29
2.6.10	Gypsum	29
2.6.11	Goethite	29
2.6.12	Halloysite	30
2.6.13	Hematite	30
2.6.14	Hydrated Silica	30
2.6.15	Illite	30
2.6.16	Jarosite	30
2.6.17	Kaolinite	30
2.6.18	Muscovite	31
2.6.19	Nontronite	31
2.6.20	Olivine	31
2.6.21	Phlogopite	31
2.6.22	Prehnite	31
2.6.23	Pyroxene	32
2.6.24	Saponite	32
2.6.25	Talc	32
2.6.26	Wollastonite	32
3	Specification and Design	33
3.1	Algorithm Overview	33
3.2	Bacterial Embedding Structure	34
3.2.1	Gene Categories	35
3.3	Objective Function	36
3.4	Bacterial Operations	37
3.4.1	Local Search	37
3.4.2	Gene Transfer	38
3.5	Fuzzy Logic for Gene Transfer and Local Search Boundaries	38

3.5.1	Fuzzy Variables and Membership Functions	38
3.5.2	Fuzzy Rules for Action Potential	39
3.5.3	Fuzzy Inference System for Gene Transfer	39
3.5.4	Action Strategy Based on Action Potential	40
3.6	Final Solution	40
3.7	Initial Population Generation	41
3.8	Algorithm Pseudocode	42
3.9	Algorithm Complexity	42
3.10	Evaluation Methods and Performance Metrics	43
3.11	Dataset Creation	44
4	Implementation and Usage	45
4.1	Software and Tools	45
4.2	Objective Function Visualization	45
4.3	Local Search Visualization	46
4.4	Gene Transfer Visualization	48
4.5	Fuzzy Logic Visualization	49
4.6	Final Solution Visualization	50
5	Experimental Results and Conclusions	52
5.1	Deconvolution Results	52
5.2	Classification Results	53
5.3	Final Conclusion	54
5.4	Limitations and Future Work	55
A	Interface for Database Creation	56
B	Visualization of Carbonate Minerals	58
C	Detailed Confusion Matrix	60
	Bibliography	60
	List of Figures	70
	List of Tables	73
	List of Algorithms	74

Chapter 1

Introduction

1.1 Overview

Mineral identification using spectral data in the visible to near-infrared (VNIR) and short-wave infrared (SWIR) ranges is vital for various geological applications, including resource exploration, environmental monitoring, and planetary science [1, 2]. Each mineral exhibits unique spectral signatures due to its specific chemical composition and crystal structure, enabling geologists to discern mineral types based on their spectral reflectance or absorption features over a range of wavelengths.

However, accurate interpretation of spectral data is challenging. Complex absorption features and overlapping spectral signatures often occur due to variations in grain size [3], mineral mixtures [4], surface coatings [1], and environmental conditions [4]. These factors introduce variability and noise into the spectral data, complicating the mineral identification process.

Supervised learning algorithms have been employed to automate this process; however, they typically require large, labeled datasets to perform effectively [5, 6]. In mineral spectroscopy, obtaining such extensive labeled data is impractical due to the vast diversity of minerals and the associated costs of sample collection and preparation. Consequently, alternative spectral analysis methods that require fewer data, such as spectral deconvolution, have garnered interest.

Spectral deconvolution involves isolating absorbed energy and decomposing it into smaller components, thereby compacting the data and enhancing human interpretability. Despite these advantages, classical approaches to spectral deconvolution rely heavily on expert knowledge and are time-consuming [7]. Recent advancements, such as Exponential

Gaussian Optimization (EGO) [8], have introduced automated selection processes based on predefined shape databases and iterative fitting techniques. These innovations facilitate the direct implementation of algorithms without the need for prior expert knowledge, and the decomposed spectra can subsequently serve as inputs for classifiers. Nevertheless, there remains a need to balance expert knowledge and automation to improve both efficiency and accuracy.

Advancements in computational intelligence offer novel approaches to solving complex optimization problems. Evolutionary algorithms, particularly Bacterial Evolutionary Algorithms (BEAs), have demonstrated effectiveness in handling challenging tasks by encoding solutions into the genes of individuals and enhancing them through processes that mimic natural bacterial evolution [9]. Building upon this, Bacterial Memetic Algorithms (BMAs) augment BEAs by incorporating local search techniques that utilize gradient-based optimization, thereby accelerating convergence and yielding higher-quality solutions [10]. Additionally, fuzzy logic has proven successful in guiding the convergence of these evolutionary systems by providing adaptive control mechanisms [11].

In this context, we propose the Elitist Bacterial Memetic Algorithm for Spectral Deconvolution (EBMA-SD), which builds upon the principles of BMAs with modifications in bacterial operations and the application of fuzzy rules to guide local search and gene transfer processes. By utilizing a small and diverse database, we aim to obtain more robust solutions in the deconvolution process and leverage these results to infer mineral classification. This approach not only enhances computational efficiency but also improves the interpretability of the results for geologists by providing clearer spectral component analysis. Furthermore, it demonstrates the potential of using BEAs in spectral deconvolution applications.

Following this overview, we provide a summary of the problem and the proposed solution. In Chapter 2, Literature Review, we examine relevant studies on spectroscopy in mineralogy, spectral preprocessing techniques, bacterial evolutionary algorithms, and fuzzy logic systems. Chapter 3, Specification and Design, details the specifications of the proposed EBMA-SD algorithm, highlighting how it integrates fuzzy logic into bacterial memetic frameworks. In Chapter 4, Implementation and Usage, we describe the development process, parameter settings, and practical guidelines for utilizing the EBMA-SD algorithm effectively. Finally, in Chapter 5, Experimental Results and Conclusions, the results of applying the EBMA-SD algorithm, with all samples of minerals of interest excluded to evaluate its generalization capabilities, are presented.

1.2 Motivation

The depletion of near-surface, high-grade mineral deposits compels the mining industry to venture into deeper and more geologically complex environments, as well as to exploit lower-grade ores [12]. This shift increases the complexity and cost of exploration and extraction processes. Accurate mineral identification becomes critical in this context for several reasons:

- **Economic Viability:** Precise characterization of mineral compositions enhances ore valuation and optimizes metallurgical processing, reducing operational costs [13].
- **Safety Considerations:** Detecting minerals with lower mechanical strength is essential for geotechnical assessments, preventing structural failures and ensuring the safety of mining operations [14].
- **Environmental Impact:** Improved identification allows for better waste management and mitigation strategies, minimizing the environmental footprint of mining activities [15].

Despite the importance, current spectral analysis methods are inadequate for these challenges due to their limitations in handling complex and variable spectral data with limited labeled samples.

1.3 Problem Statement

The fundamental problem in mineral identification from VNIR and SWIR spectral data is the accurate decomposition and interpretation of complex spectral signatures. Existing methods face several key challenges:

- **Overlapping Spectral Features:** Minerals often have overlapping absorption bands, making it difficult to distinguish between them using traditional methods [1].
- **Spectral Variability:** Factors such as particle size, surface roughness, and mixtures introduce variability that complicates analysis [3].
- **Limited Labeled Data:** Supervised algorithms require extensive labeled datasets, which are difficult to obtain in mineral spectroscopy due to practical constraints [5].

- **Interpretability:** Traditional spectral decomposition techniques may yield multiple plausible solutions, some of which lack physical or chemical validity, leading to misinterpretation [8].
- **Feature Preservation:** Existing algorithms may achieve good curve-fitting but at the cost of altering or omitting key spectral features essential for accurate mineral identification.

These challenges highlight the need for a new approach that can effectively decompose spectral data into meaningful components, preserve critical spectral features, and operate with limited labeled data.

1.4 Proposed Solution

This research proposes an integrated algorithm that combines spectral deconvolution with an Elitist Bacterial Memetic Algorithm (EBMA) and fuzzy logic to enhance mineral identification from spectral data. The key components of the proposed solution are as follows:

1. **Few Samples Manual Spectral Decomposition:** Guided by established literature, an initial manual decomposition of spectra from the USGS spectral library [16] is conducted to ensure that the initial components are coherent and physically meaningful.
2. **Elitist Bacterial Memetic Algorithm (EBMA):** The EBMA is employed to optimize the spectral decomposition process. It incorporates:
 - **Local Search (Levenberg-Marquardt Method):** Enhances solution precision by performing gradient-based optimization in non-linear least squares problems [17]. In this context, it fine-tunes the genes of the bacterial population to minimize the difference between the observed and reconstructed spectra.
 - **Gene Transfer:** A bidirectional gene transfer mechanism is implemented, following a phenotypic elitist approach. In our approach, the best-performing mineral (elite) is selected, and the rest of the population is used to enhance this elite individual.

- **From Population to Elite:** The elite bacterium searches the rest of the population for genes that can help it improve its fitness.
 - **From Elite to Population:** The elite individual shares its genes with population members that contributed to its genetic makeup, thereby expanding the search space to include other minerals and their mixtures.
3. **Fuzzy Logic System:** A fuzzy logic system guides the gene transfer policy and defines the boundaries for local searches during the evolutionary process. It prioritizes the conservation of spectral features that are strongly supported by the literature while allowing modifications to less critical features.
4. **Geological Interpretation:** The results provide a decomposition of the spectra along with information about the most similar minerals, mineral groups, and mineral types from the database used. Additionally, even if a mineral is not present in the database, the system offers important information about which elements could be responsible for the observed absorptions.

This integrated approach aims to overcome the limitations of existing methods by ensuring physically valid spectral decompositions, preserving essential spectral features, and providing interpretable results that can assist geologists in decision-making processes.

Chapter 2

Literature Review

This chapter provides an overview of the fundamental concepts that underpin this research proposal. It begins with an explanation of spectrometry, covering its basic principles and the factors that influence spectrometric measurements to assess potential data variability. Following this, the chapter discusses preprocessing techniques and spectral deconvolution methods essential for accurate data analysis. The discussion then introduces bacterial algorithms and the elitist evolutionary approach, highlighting their significance and application within this study. Subsequently, a review of fuzzy logic is presented, emphasizing its applications in similar problems and its relevance to the current research. Finally, the chapter concludes with a bibliographic review of minerals, focusing on their most extensively studied spectral peaks, particularly those cataloged in the USGS library.

2.1 Spectroscopy in Mineralogy

2.1.1 Fundamental Principles

The fundamental principles of spectroscopy in mineralogy are rooted in three primary processes: emission, absorption, and reflection. While emission spectroscopy examines the radiation released by excited mineral samples, absorption and reflectance spectroscopy focus on the attenuation and scattering of incident radiation, as illustrated in Figure 2.1. These interactions produce characteristic spectral signatures that serve as fingerprints for mineral identification and characterization [3].

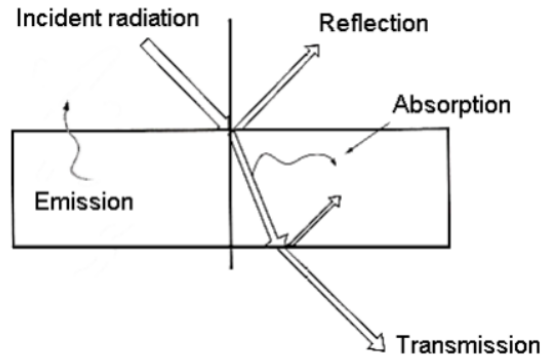


Figure 2.1: Interaction of radiation with matter, illustrating reflection, transmission, absorption, and emission processes [18].

In the VNIR-SWIR range, two key absorption processes dominate mineral spectra:

- **Electronic Transitions:** Primarily observed in the VNIR region (400–1300 nm), these transitions arise from crystal field effects and charge transfer processes involving transition metal ions such as Fe^{2+} , Fe^{3+} , and Ti^{3+} . These processes typically produce broad absorption features due to the wide range of possible electronic states and transitions available within the crystal lattice. For example, the Actinolite spectrum in Figure 2.2 shows a broad absorption band near 1 μm , primarily attributed to Fe^{2+} ions [19].
- **Vibrational Processes:** Dominant in the SWIR region (1300–2500 nm), these processes arise from fundamental molecular vibrations, overtones, and combination tones. In contrast to electronic transitions, vibrational features are typically sharper and more distinct because they are more tightly constrained by the mineral's crystal structure. The lattice often allows only specific vibrational modes while blocking others, limiting the available energy states and resulting in narrower absorption bands. This structural selectivity provides detailed information about the mineral's composition, bonding environment, and hydroxyl content [1].

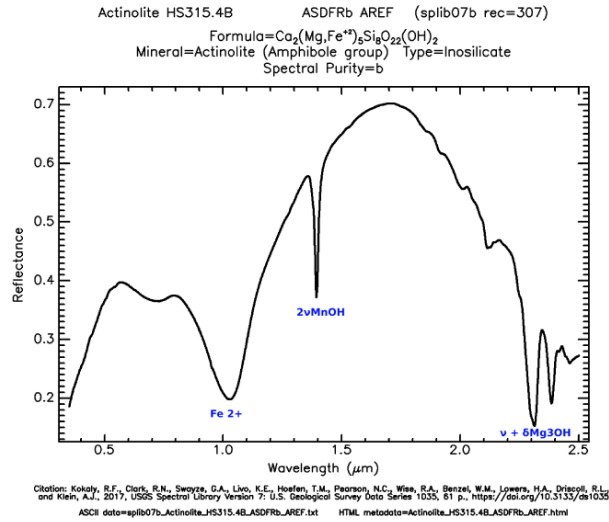


Figure 2.2: Reflectance spectrum of Actinolite, showing absorption features related to Fe^{2+} , Mn-OH, and Mg-OH in the VNIR-SWIR range. Adapted from [20].

2.1.2 Instrumentation Considerations

When capturing data, it is essential to acknowledge that the characteristics of the sensor can significantly influence the appearance of spectral curves. Some key factors that affect the quality and interpretability of the obtained reflectance curve include:

- **Spectral Resolution:** Spectral resolution refers to the ability of a spectrometer to resolve fine details in the spectrum. As can be seen in Figure 2.3, higher spectral resolution enables the identification of narrow absorption features, which are critical for distinguishing between different minerals [21].

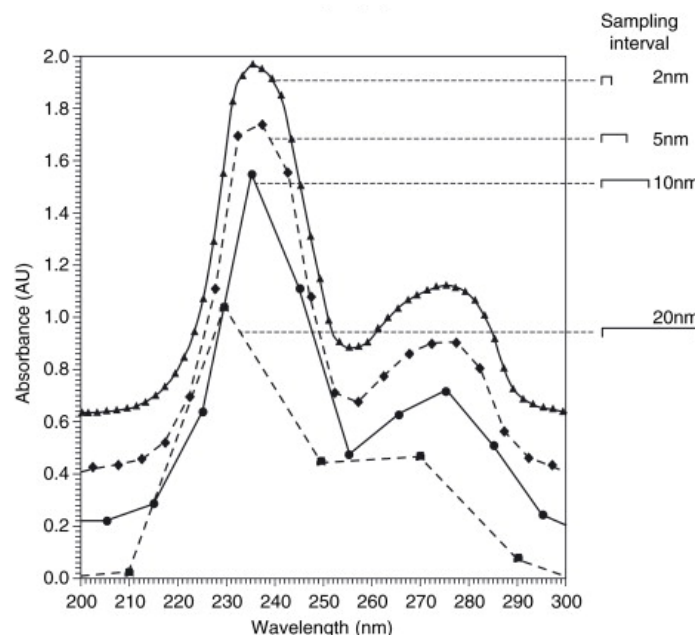


Figure 2.3: Absorbance spectra at various sampling intervals, demonstrating the effect of spectral resolution on peak identification [21].

- **Signal-to-Noise Ratio:** A low SNR can obscure important spectral features and introduce uncertainties in the data, making it difficult to distinguish between different minerals. Noise reduction techniques and careful data processing are necessary to mitigate these issues [22]. In Figure 2.4, you can see how noise can cover the signal if it is very weak.

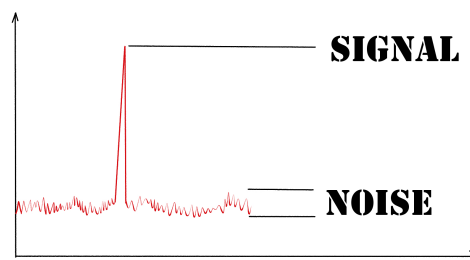


Figure 2.4: Illustration of the Signal-to-Noise Ratio (SNR), showing how low SNR can obscure spectral features [23].

- **Sensor Calibration:** Regular calibration of the hyperspectral instrument is necessary to maintain measurement accuracy. This involves using standard reference materials with known spectral properties to correct for instrument-specific biases

and drift over time [24], as shown in Figure 2.5. Calibration is performed under uniform illumination and in complete darkness.

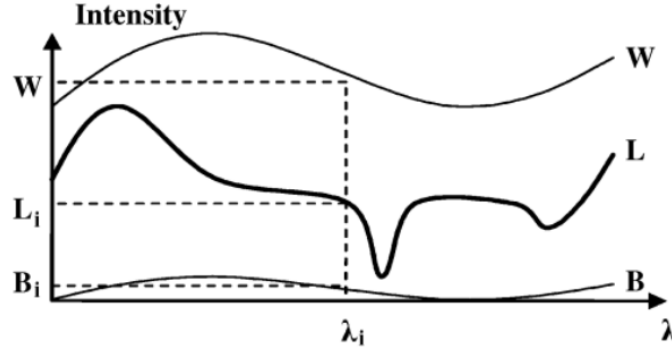


Figure 2.5: Sensor calibration in hyperspectral imaging, displaying intensity levels for different conditions: L - light reflected from the target, B - baseline in darkness, W - standard material with a reference light source [25].

2.1.3 Sample considerations

The condition of the sample to be analyzed will also cause variations in the obtained spectral curve. Some of the most important factors include:

- **Grain Size:** The size of the mineral grains influences the path length of light within the material, affecting absorption and scattering processes. Larger grains can increase absorption path lengths, enhancing absorption features, but the center of absorption bands is generally not affected by grain size, as shown in Figure 2.6.

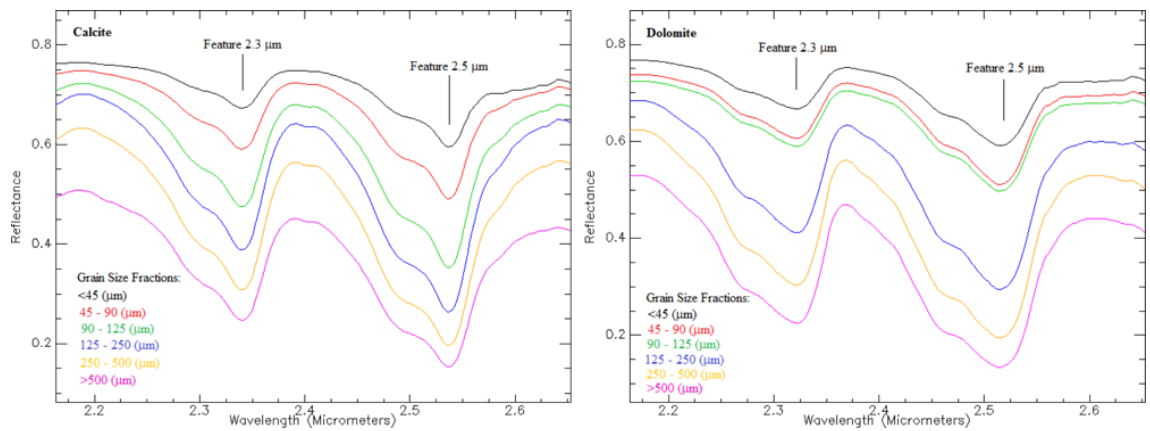


Figure 2.6: Reflectance spectra demonstrating the effect of grain size on spectral characteristics for two carbonates: Calcite (left) and Dolomite (right) [26].

- **Crystallinity:** The crystallinity of minerals can significantly influence observed spectral features. The amplitude of the absorption features increase as the crystallinity of minerals decreases, since bond lengths show greater variation in poorly ordered crystal structures [27]. For specific minerals like those in the kaolin group, the relative intensity of absorption bands directly relates to their crystallinity [28]. The characteristic doublet at 2159 nm and 2209 nm in kaolin can be decomposed into four absorption bands, with changes in the shape of this doublet indicating variations in structure and crystallinity [29], as shown in Figure 2.7. Additionally, crystallinity affects mineral identification - smectites and mica-like minerals of the same octahedral composition show different locations of OH absorption bands based on crystallinity, with smectites absorbing at shorter wavelengths (ca. 2200 nm) compared to more crystalline illite absorbing at longer wavelengths (ca. 2220 nm) [30, 31].

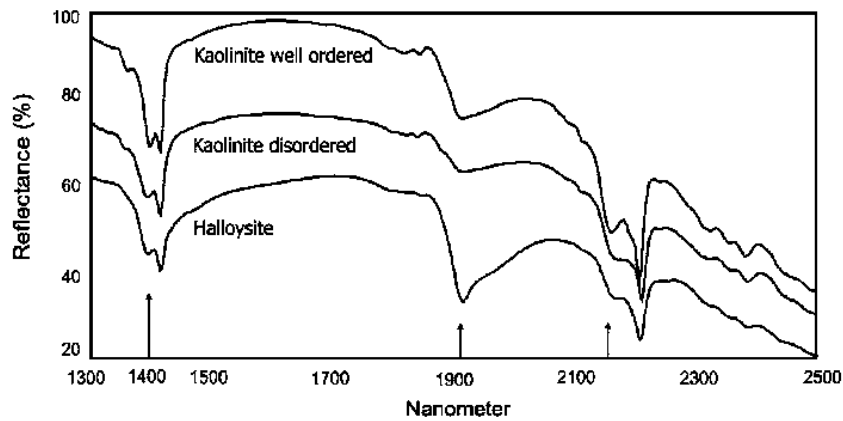


Figure 2.7: Comparison between a well-ordered kaolinite sample with high crystallinity and a sample of disordered kaolinite and halloysite, both with low crystallinity [32].

- **Mineral Mixtures:** The spectral features of individual minerals can overlap, making it challenging to distinguish between them, as shown in Figure 2.8.
 - **Homogeneous Mixtures:** In a homogeneous mixture, the minerals are uniformly distributed, and the resultant spectrum is a weighted average of the individual spectra. This can result in blended absorption features, which may obscure the signatures of the constituent minerals.
 - **Heterogeneous Mixtures:** Heterogeneous mixtures contain minerals with varying proportions and distributions. This can lead to complex spectral features with varying depths and shapes of absorption bands. Advanced spectral

unmixing techniques are often required to deconvolve the spectra and identify the individual components.

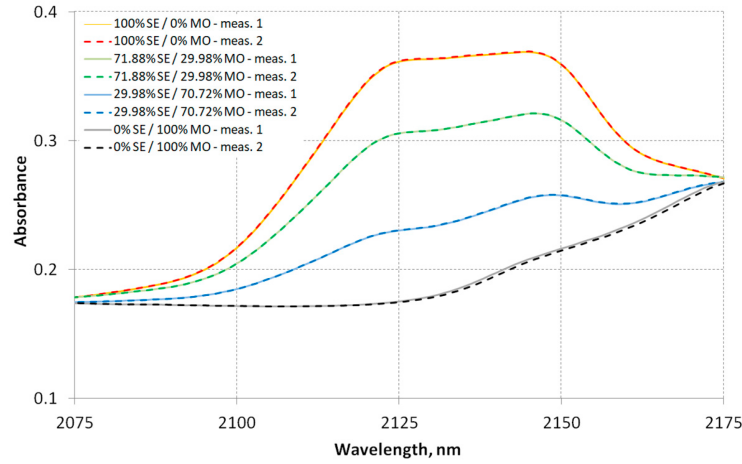


Figure 2.8: Absorption spectra of heterogeneous mixtures showing variations in band depth and shape [33].

- **Surface Coatings:** Surface coatings, such as dust, organic matter, or oxidation layers, can modify the reflectance spectra of minerals. These coatings can introduce additional absorption features or alter the intensity of existing ones, as shown in Figure 2.9.

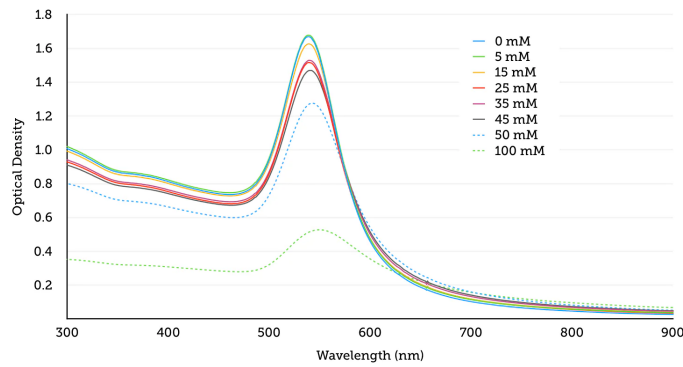


Figure 2.9: Impact of surface coatings on reflectance spectra, with changes in absorption intensity due to dust or organic matter [34].

- **Moisture:** The presence of moisture can influence the reflectance spectra by introducing additional absorption features related to water. These features, particularly in the near-infrared region, can overlap with mineral absorption bands, complicating the spectral analysis. Moisture can also affect the overall reflectance, reducing the intensity of the spectral features, as shown in Figure 2.10.

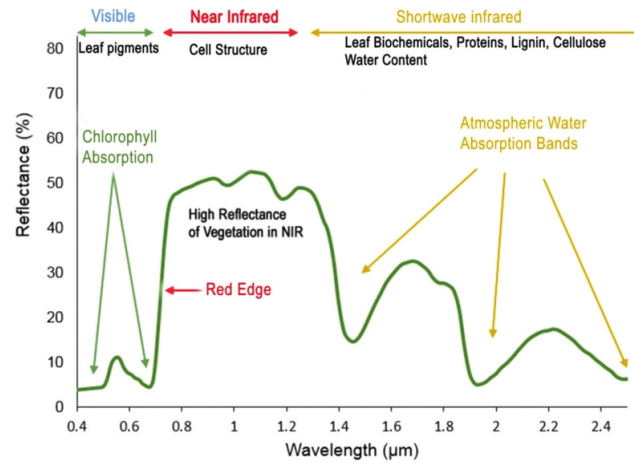


Figure 2.10: Multispectral analysis showing different regions of reflectance for vegetation, including chlorophyll absorption and water content bands [35].

- Temperature:** Variations in temperature can influence the vibrational states of molecules within minerals, thereby affecting their absorption features. Higher temperatures can cause broadening of the absorption bands and shifts in their positions. However, at room temperature, these effects are typically minimal [36], as illustrated in Figure 2.11.

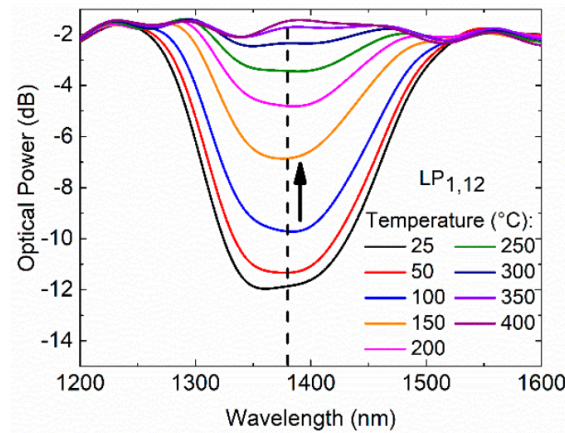


Figure 2.11: Temperature effects on absorption spectra, illustrating band broadening and shifts at elevated temperatures [37].

2.2 Spectral Deconvolution

Spectral deconvolution is an essential process in reflectance spectroscopy, offering a powerful technique for knowledge-driven expert applications. This technique isolates absorption features attributed to minerals and disentangles complex spectral data into meaningful components, enabling the identification and characterization of their chemical

and physical properties. The process is illustrated in Figure 2.12, where the reflectance curve of calcite and its deconvolved absorption features are shown. It consists of two main steps: the removal of the continuum and the deconvolution of the absorption curve into its individual components.

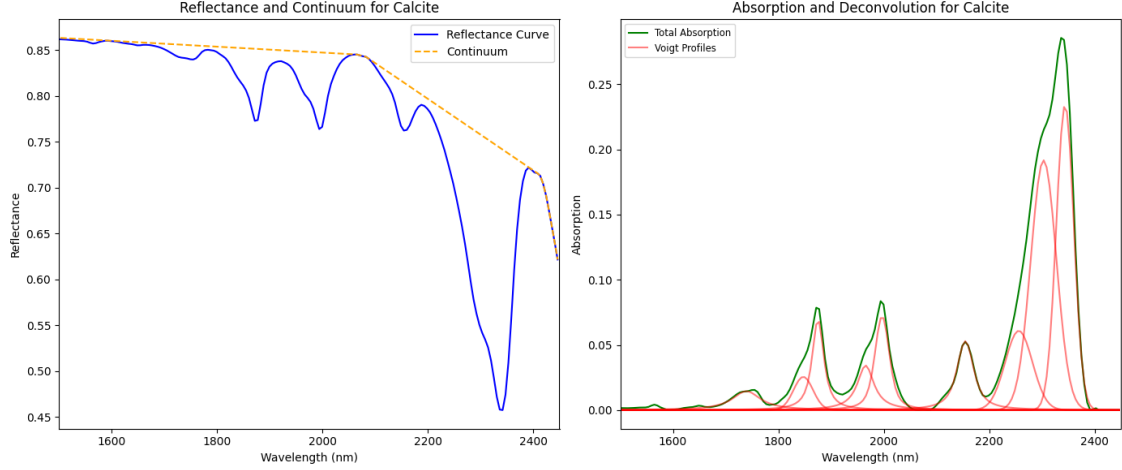


Figure 2.12: Left: Reflectance curve of calcite with its continuum (linear interpolation). Right: Absorption curve obtained by subtracting the continuum, with Voigt profile deconvolution shown.

2.2.1 Continuum Removal

The continuum in a reflectance spectrum represents the baseline shape, influenced by factors such as grain size, surface roughness, and illumination conditions. Accurate removal of the continuum is a prerequisite for reliable deconvolution, as it isolates absorption features from the background spectrum.

Several methods exist for continuum removal:

- **Geometric Approaches:** These include convex hull techniques and linear interpolation methods. Convex hull techniques fit a geometric envelope above the spectrum, while linear interpolation connects selected points on the spectrum with straight lines to approximate the continuum. Although geometric approaches are straightforward and computationally efficient, they are sensitive to noise and may introduce artifacts in the processed data [38, 39].
- **Parametric Models:** These involve fitting polynomial functions, such as linear or quadratic polynomials, or using physical models to represent the continuum. Polynomial fitting offers flexibility in capturing the continuum's shape, with higher-order polynomials providing more adaptability at the cost of potential overfitting.

Modified Gaussian Models (MGM) have demonstrated the utility of quadratic polynomials for continuum modeling, showing better fits for complex spectral baselines [39].

In this work, we employ linear interpolation within the geometric approaches for continuum removal due to its simplicity. However, it is important to note that more sophisticated methods, such as polynomial fitting, are often recommended for improved accuracy and reliability [38, 39].

2.2.2 Deconvolution Components

Spectral deconvolution involves decomposing spectral features into their constituent mathematical distributions, typically represented by Gaussian and Lorentzian functions.

Gaussian Distribution

Gaussian distributions are often used to represent electronic transitions in spectral data. They are characterized by their symmetric bell-shaped curves, which are defined by the following equation:

$$G(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-x_0)^2}{2\sigma^2}}, \quad (2.1)$$

where:

- σ is the standard deviation, determining the width of the peak.
- x_0 is the position of the peak center.

Gaussian distributions are suitable for modeling features where the broadening mechanisms are primarily due to Doppler effects, leading to symmetric absorption lines.

Lorentzian Distribution

Lorentzian distributions better represent vibrational processes and are characterized by their peak shape, which has longer tails compared to Gaussian distributions. The Lorentzian function is defined as:

$$L(x) = \frac{\gamma}{\pi((x-x_0)^2 + \gamma^2)}, \quad (2.2)$$

where:

- γ is the Lorentzian half-width at half-maximum (HWHM), determining the peak's width.
- x_0 is the position of the peak center.

The half-width at half-maximum (HWHM) γ is the distance from the center of the peak to the point where the function reaches half of its maximum value. It quantifies the broadening of the peak due to interactions such as pressure broadening.

Lorentzian distributions are particularly effective in modeling spectral features where pressure broadening or other interactions lead to asymmetrical or wider absorption lines.

Voigt Profile

The Voigt profile, a convolution of Gaussian and Lorentzian distributions, provides a robust framework for modeling absorption features influenced by both Doppler (Gaussian) and pressure (Lorentzian) broadening effects [40, 41, 42]. It is mathematically defined using the convolution operator $*$ as:

$$V(x; \sigma, \gamma) = G(x; \sigma) * L(x; \gamma), \quad (2.3)$$

which expands to:

$$V(x; \sigma, \gamma) = \int_{-\infty}^{\infty} G(t; \sigma) L(x - t; \gamma) dt. \quad (2.4)$$

Here, G and L represent the Gaussian and Lorentzian components defined in Equations (2.1) and (2.2), respectively.

The full Voigt profile can be expressed as [43]:

$$f(x) = \frac{\gamma}{\sigma\sqrt{2\pi^3}} \int_{-\infty}^{+\infty} \frac{e^{-y^2/2}}{((x-y)^2 + (\gamma/\sigma)^2)} dy. \quad (2.5)$$

In spectroscopic applications, it is common to express the Voigt profile in terms of wavenumber ω and peak height h [44]:

$$I_V(\omega, h, \sigma, \gamma) = hJ \int_{-\infty}^{\infty} \frac{e^{-(t/\sigma)^2}}{1 + \left(\frac{\omega - t}{\gamma}\right)^2} dt, \quad (2.6)$$

where:

- I_V is the absorbance intensity given by the Voigt profile.

- h is the maximum peak height (maximum absorbance).
- ω is the centralized wavenumber ($\omega = \sigma - \omega_0$).
- ω_0 is the wavenumber at peak maximum.
- σ is the Gaussian width parameter (standard deviation).
- γ is the Lorentzian width parameter (HWHM).
- J is a normalization factor: $J = \left[\pi \int_{-\infty}^{\infty} \frac{e^{-(t/\sigma)^2}}{1 + \left(\frac{t}{\gamma}\right)^2} dt \right]^{-1}$.

For computational efficiency, the Voigt profile can be expressed in terms of the complex error function $W(z)$ [45]:

$$I_V(\omega, h, \sigma, \gamma) = h \frac{\text{Re}[W(z)]}{\text{Re}[W(x=0, iy)]}, \quad (2.7)$$

where $z = x + iy$, $x = \frac{\omega}{\sigma}$, $y = \frac{\gamma}{\sigma}$, and $\text{Re}[\cdot]$ denotes the real part of a complex number.

2.2.3 Current Deconvolution Techniques

Spectral deconvolution can be categorized based on the models and algorithms employed. Current techniques include:

- **Gaussian and Lorentzian Models:** These classical approaches use symmetric profiles to fit absorption features. While effective for simple spectra, they are limited in handling asymmetrical or overlapping features [46].
- **Modified Gaussian Models (MGM):** MGM extends Gaussian models by introducing parameters for asymmetry and saturation. This model is particularly suitable for analyzing short-wave infrared (SWIR) spectra, where absorptions often exhibit asymmetric shapes [47].
- **Exponential Gaussian Optimization (EGO):** The EGO model further refines MGM by incorporating additional parameters for flexibility in fitting complex absorption features. It is commonly used for minerals exhibiting highly non-linear spectral characteristics. The authors warn that large asymmetry values can lead to unrealistic absorption shapes [48, 8].

- **Sparse Approximation Algorithms:** Methods like Non-Negative Orthogonal Matching Pursuit (NNOMP) apply sparsity constraints to select the most representative spectral features. These algorithms are computationally efficient and are increasingly used in automated analyses [49].
- **Elitist Bacterial Memetic Algorithm for Spectral Deconvolution (EBMA-SD):** This is our proposed method, introducing a novel approach that integrates a phenotypical elitist strategy into Bacterial Memetic Algorithms and replaces random mutation by leveraging a small, curated database of spectral components derived from expert knowledge. This effectively balances automation with expert insight, offering convergence more aligned with expert knowledge, and allows the transfer of metadata into the components, improving the human interpretation of the results.

2.3 Bacterial Algorithms

2.3.1 Bacterial Evolutionary Algorithms

Bacterial Evolutionary Algorithms (BEAs) are optimization techniques inspired by microbial evolution and horizontal gene transfer [9]. This technique was first proposed to design fuzzy systems by encoding the components of these systems within the chromosomes of bacteria. They diverge from traditional genetic algorithms by focusing on asexual reproduction and utilizing mechanisms where one individual can donate genes to another, rather than relying on parent-based generational reproduction. The key components of BEAs include *Bacterial Mutation* and *Gene Transfer*.

Bacterial Mutation

Bacterial mutation simulates rapid bacterial reproduction and mutation, creating clones with random mutations and selecting the best-performing ones. This process introduces genetic diversity into the population, allowing BEAs to explore a wide range of potential solutions and avoid premature convergence on suboptimal solutions.

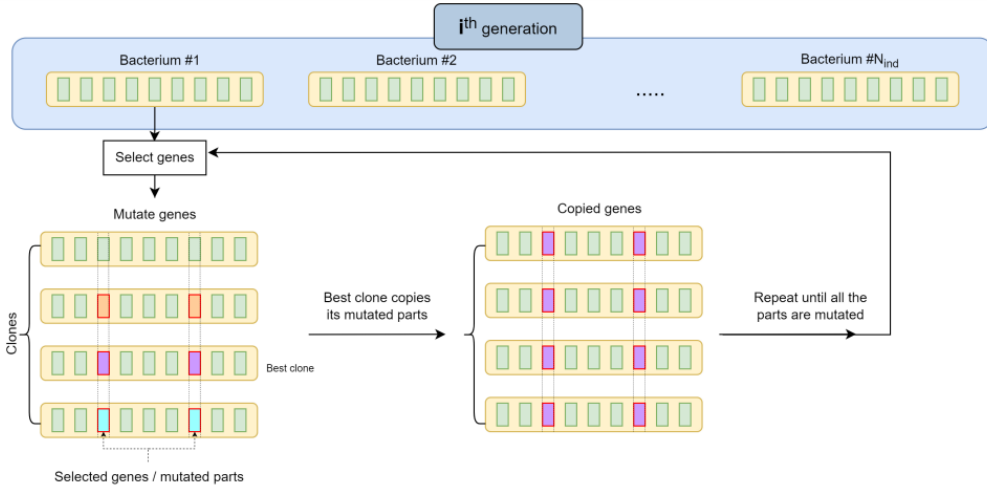


Figure 2.13: Illustration of the i -th generation in BMA, showing genes mutation in a bacterium by selecting genes, mutating them, and propagating the best clone across the remaining copies of the bacterium [50].

Although mutation is a fundamental component of bacterial algorithms, it will not be utilized in this proposal. The primary role of mutation is to enhance global search by exploring the solution space randomly. However, in our specific problem, the goal is not broad exploration but rather the focused investigation of solution spaces closely related to specific minerals.

Instead of employing a random global search through mutation, we will adopt a more localized search strategy. By leveraging a diverse initial population of minerals, this approach ensures that we effectively target all relevant areas with high potential for solutions, thereby maintaining diversity while focusing the search on mineral-specific regions of the solution space.

Gene Transfer

In BEAs, one individual is selected to transfer genes to another, mimicking the exchange of genetic material between bacteria. To perform gene transfer, the bacterial population is sorted into two halves: the half with individuals of higher fitness is called the superior half, while the other half is known as the inferior half. A chromosome is randomly selected from the superior half and designated as the source chromosome, while another chromosome is randomly chosen from the inferior half and designated as the destination chromosome. According to specific criteria, a ‘good’ segment or gene from the source chromosome is transferred to the destination chromosome as shown in Figure 2.14.

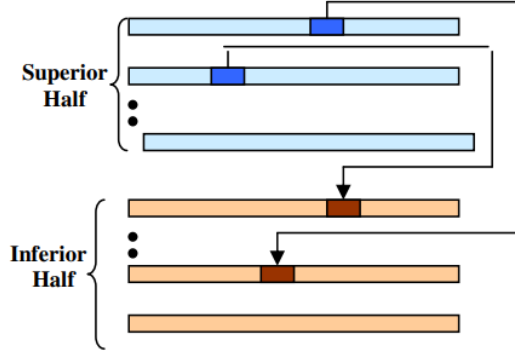


FIGURE 2. SCHEME FOR THE GENE TRANSFER OPERATION.

Figure 2.14: Scheme for the gene transfer operation [51].

2.3.2 Bacterial Memetic Algorithms

Bacterial Memetic Algorithms (BMAs) enhance BEAs by incorporating local search techniques to refine solutions [10]. The memetic component allows for fine-tuning of candidate solutions, improving convergence and solution quality (see Figure 2.13).

Mathematically, the local search can be represented as:

$$b^* = \text{LocalSearch}(b') \quad (2.8)$$

where b' is a bacterium after mutation, and b^* is the locally optimized bacterium.

2.4 Elitist Strategy in Genetic Algorithms

In genetic algorithms with parents, an elitist strategy typically refers to allowing the best individuals to pass directly to the next stage without applying operations such as crossover and mutation, and the rest of the new population is then filled with *offspring* created through crossover and mutations [52]. When referring to bacterial algorithms, these do not have an explicit *offspring* since there are no generations, but rather modified copies of a bacterium. These algorithms inherently possess an elitist component by preserving the best version of all copies [53].

Therefore, in this study, we will address elitism in bacteria in a different manner. Our focus is on exploring the space of possibilities within specific zones that represent particular minerals. Since the good performance of a solution increases the likelihood of a mineral being present, we will consider the best individual as the elite. The rest of the individuals will attempt to contribute genes that improve its performance, with the

reciprocal benefit of also being able to obtain genes from the elite. This bidirectional gene transfer will allow us to explore options for mineral mixtures effectively, as illustrated in Figure 2.15.

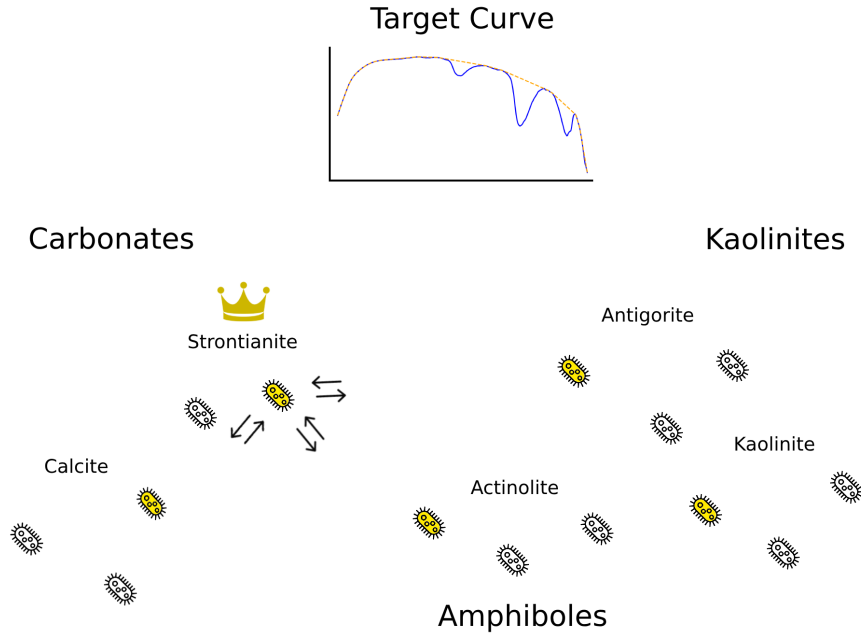


Figure 2.15: Simplified representation of the elitist system. Given a target curve, the best bacterium from each mineral is selected and shown in yellow. The mineral with the best bacterium is chosen as the elite, represented in the figure with a crown. Afterwards, gene swapping is allowed between the elite and the rest of the population.

2.5 Fuzzy Logic

2.5.1 Fundamental Concepts

Fuzzy set theory extends classical set theory by allowing elements to possess varying degrees of membership, rather than adhering to a strict binary inclusion [54]. Unlike traditional sets, where an element is either a member or not, a fuzzy set employs a membership function that assigns each element a value between zero and one, representing its degree of membership. This allows for mathematical modeling of imprecise or uncertain information, making it particularly valuable for problems involving complex decision-making or pattern recognition.

2.5.2 Mathematical Formulation

A fuzzy set A in a universe of discourse X is characterized by a membership function $\mu_A(x)$, where:

$$\mu_A(x) : X \rightarrow [0, 1] \quad (2.9)$$

The membership value $\mu_A(x)$ represents the degree to which element x belongs to set A . The basic operations on fuzzy sets extend classical set operations to handle varying degrees of membership [54]:

- **Intersection:** The intersection of two fuzzy sets A and B , denoted $A \cap B$, is characterized by the membership function:

$$\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)).$$

This operation selects the minimum degree of membership between A and B for each element x , representing the degree to which x belongs to both sets simultaneously.

- **Union:** The union of two fuzzy sets A and B , denoted $A \cup B$, is characterized by the membership function:

$$\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)).$$

This operation selects the maximum degree of membership between A and B for each element x , representing the degree to which x belongs to either or both sets.

- **Complement:** The complement of a fuzzy set A , denoted $\bar{A}$, is characterized by the membership function:

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x).$$

This operation represents the degree to which x does not belong to set A , effectively inverting the membership grade.

2.5.3 Fuzzy Logic in Evolution Adaptive Parameter Control

Recent research has shown the effectiveness of fuzzy logic controllers in dynamically adjusting evolutionary algorithm parameters [55]. These systems typically consist of:

- **Input Variables:** Monitor population diversity and fitness improvements
- **Fuzzy Rules:** Map evolutionary states to parameter adjustments
- **Output Functions:** Generate precise parameter values through defuzzification

For example, in [11], crossover and mutation rates are adjusted based on fuzzy rules that evaluate current population performance. When fitness improvements stagnate, mutation rates may increase to escape local optima.

2.5.4 Fuzzy Logic in Mineral Classification

In spectral mineral identification, Rialland et al. [56] developed a fuzzy classification system that compares absorption positions between unknown spectra and reference database spectra. Their system processes:

- Estimated absorption positions (μ_e) of unknown spectrum
- Uncertainties on these positions (σ_{μ_e})
- Reference database positions (μ_d) for known minerals

The classification employs a coincidence function to evaluate similarity:

$$f(x) = \sum_i \exp\left(\frac{-(x - \mu_{ei})^2}{2\sigma_{\mu_{ei}}^2}\right) \quad (2.10)$$

where the function is limited to 1 when multiple Gaussians overlap. Their system uses two key metrics for mineral identification:

- **Similarity degree (S):** Maximum coincidence function value (range [0,1])
- **Matched positions percentage (M_{pos}):** Proportion of database positions with coincidence > 0.1

The fuzzy classification process:

1. Evaluates main and secondary absorptions separately
2. Uses membership functions for S (Low/High) and M_{pos} (Low/Medium/High)
3. Generates identification score (0-10) through fuzzy inference

4. Prioritizes main absorptions over secondary features

For final classification, their system:

- Requires near to 100% matched main absorptions
- Distinguishes between mineral mixtures (>10nm maximum position distance) and similar minerals (<10nm)
- Handles overlapping features and variable positions
- Enables identification of mineral combinations

Their approach demonstrated promising results in handling complex mineral identification challenges. However, the authors noted potential improvements could be made by incorporating additional spectral parameters beyond peak positions to achieve better discrimination between minerals. Furthermore, special consideration is needed for minerals with single characteristic absorption features or those lacking distinctive peaks, as their current methodology may not provide sufficient discrimination in these cases [56].

2.6 Geology

Understanding the geological context is essential for the effective application of spectral decomposition and fuzzy classification algorithms in mineralogical studies. Minerals are the fundamental components of rocks and are defined as naturally occurring, inorganic solids with a definite chemical composition and an ordered atomic structure [57]. This section provides a brief overview of the main USGS mineral types relevant to our study, grouped by the mineral type specified in the database. Not all minerals will be designated in this section due to their large number, but we will address the most studied minerals.

2.6.1 Actinolite

Actinolite exhibits an Mg-OH absorption feature between 2314–2324 nm, characteristic of its Mg-rich composition. Additionally, it shows secondary peaks at 2380–2390 nm associated with combination bands of Mg-OH vibrations. These absorption features arise from Mg-OH bond vibrations within its crystal structure [58].

2.6.2 Alunite

Alunite displays absorption peaks at 1473 nm for K-rich varieties due to OH stretching and bending combinations, and at 1491 nm for Na-rich varieties, shifted due to different cations. Additional absorption features are observed at 1770 nm and 2170 nm, related to SO₄ and OH vibrations. These features result from OH and sulfate (SO₄) vibrational modes, with cation type affecting the peak positions [27].

2.6.3 Amphibole (Hornblende)

Hornblende shows an Mg-OH absorption feature between 2326–2350 nm, indicative of its composition. It also exhibits secondary features at 2390–2410 nm due to Fe-OH vibrations. These absorption features result from Mg-OH and Fe-OH vibrations in its amphibole structure, influenced by the Mg/Fe ratio [58].

2.6.4 Calcite

Calcite is characterized by an absorption peak at 2350 nm due to overtones and combination bands of C–O stretching vibrations. Weaker combination modes are observed near 1990 nm and 1870 nm. These absorption features arise from overtones and combination bands of the fundamental vibrations of the carbonate ion (CO₃²⁻) [59, 60].

2.6.5 Chlorite

Chlorite exhibits an Al-OH absorption around 1400 nm, which can overlap with other minerals, causing spectral complexities. It also shows Fe-OH absorption near 2250 nm, with Fe-rich chlorite peaking at 2256–2265 nm, intermediate chlorite at 2250–2255 nm, and Mg-rich chlorite at 2240–2249 nm. Additionally, Mg-OH absorption near 2350 nm shifts depending on Mg content. These features result from vibrational overtones and combination bands of Fe-OH and Mg-OH bonds, shifting with the Fe/Mg ratio [58, 61].

2.6.6 Dickite

Dickite, a polymorph of kaolinite, displays O-H vibrational absorption bands at 1381 nm and 1412 nm, which are sharper compared to kaolinite. It also shows an Al-OH absorption at 2178 nm. These features arise from O-H bond vibrations in a more ordered crystal structure compared to kaolinite [62].

2.6.7 Dolomite

Dolomite shows absorption features between 2300–2340 nm due to overtones and combination bands of carbonate vibrations. These features are similar to those of calcite but are slightly shifted to shorter wavelengths due to Mg substitution in the crystal lattice [60].

2.6.8 Epidote

Epidote exhibits an Fe-OH absorption feature at 2255 nm and Mg-OH absorption features between 2335–2342 nm, with positions shifting based on Mg content. Additional features are present at 1540 nm and 1835 nm, associated with overtones of OH-related vibrations. These absorption features arise from Fe-OH and Mg-OH vibrational overtones within the silicate structure [58, 63].

2.6.9 Garnet (Grossular)

Grossular garnet shows crystal field transitions of Fe^{2+} and Fe^{3+} around 450 nm and 950 nm, respectively. It also exhibits an OH-related absorption feature at 1400 nm and combination bands involving cation-OH vibrations at 2350 nm. These features are influenced by Fe electronic transitions and hydroxyl group vibrations [64].

2.6.10 Gypsum

Gypsum shows a prominent absorption peak around 1750 nm, associated with the first overtone of H_2O combination bands and SO_4 overtones. This feature arises from the vibrations of water molecules and sulfate groups within the crystal structure, specifically from the combination of the overtone of the H_2O bending mode and SO_4 stretching vibrations unique to gypsum. This absorption peak is optimal for gypsum discrimination due to its lower interference from other soil components compared to other features [65].

2.6.11 Goethite

Goethite displays electronic transitions of Fe^{3+} ions at 540 nm and 920 nm. Additionally, it has OH and H_2O combination bands at 1410 nm, 1800 nm, and 1930 nm. These absorption features arise from Fe^{3+} electronic transitions and OH group vibrations [66].

2.6.12 Halloysite

Halloysite exhibits O-H stretching vibrations at 1389 nm, H₂O absorption at 1913 nm, and structural OH vibrations at 2138 nm. These features result from O-H and H₂O vibrations in its tubular structure [62].

2.6.13 Hematite

Hematite shows charge transfer transitions involving Fe³⁺ at 860 nm and crystal field transitions at 940 nm. These electronic transitions contribute to its characteristic red coloration [60].

2.6.14 Hydrated Silica

Hydrated silica presents a broad absorption near 2210 nm due to Si-OH vibrational overtones, with weaker features at 1390 nm and 1930 nm associated with H₂O and Si-OH combination bands. These features arise from silanol (Si-OH) group vibrations and molecular water within its amorphous structure [67].

2.6.15 Illite

Illite exhibits an OH stretching vibration around 1405 nm, an H₂O combination band at 1910 nm, and an Al-OH bending vibration at 2205 nm. Secondary features at 2350 nm and 2440 nm arise from combination bands involving cation-OH vibrations. These absorption features result from vibrational modes of OH and Al-OH groups [68].

2.6.16 Jarosite

Jarosite displays an OH vibrational absorption at 1470 nm and combination bands involving OH and sulfate groups between 1762–1764 nm. These features are due to OH vibrations and sulfate group combinations, similar to alunite [27].

2.6.17 Kaolinite

Kaolinite shows O-H stretching vibrations between 1340–1409 nm and a doublet of Al-OH stretching and bending vibrations at 2160 nm and 2200 nm. High-crystallinity

kaolinite has sharper and more distinct absorption features compared to disordered kaolinite, where the doublet in the Al-OH region is less distinct. These features arise from O-H bond vibrations and Al-OH group vibrations, with crystallinity affecting feature sharpness [62].

2.6.18 Muscovite

Muscovite shows an Al-OH absorption feature between 2195–2215 nm and secondary combination bands at 2340 nm. These features result from Al-OH stretching and bending vibrations, serving as a key diagnostic for identifying muscovite in spectral data [61].

2.6.19 Nontronite

Nontronite exhibits an OH stretching vibration at 1420 nm, an H₂O combination band at 1910 nm, an Fe-OH bending vibration at 2290 nm, and an Fe-OH combination band at 2410 nm. These absorption features arise from vibrational modes of OH and Fe-OH in its layered silicate structure [67].

2.6.20 Olivine

Olivine displays absorption features due to Fe²⁺ crystal field transitions between 1000–1080 nm, with positions shifting based on the Mg/Fe ratio. Higher Mg content shifts these features to shorter wavelengths. Additionally, the presence of Mg affects the crystal field environment, further influencing the absorption band positions [60].

2.6.21 Phlogopite

Phlogopite shows an Mg-OH absorption feature between 2310–2312 nm, associated with a high Mg# of 0.90–0.95. This feature results from Mg-OH bond vibrations, reflecting its high magnesium content relative to iron [69].

2.6.22 Prehnite

Prehnite exhibits an overtone of OH stretching vibrations at 1476 nm, OH combination bands at 2230 nm and 2282 nm, and combination bands involving OH and lattice vibrations at 2350 nm. These absorption features arise from OH stretching and bending vibrations within its crystal structure [58].

2.6.23 Pyroxene

Pyroxene presents Fe^{2+} crystal field transitions in M2 sites between 1000–1100 nm and in M1 sites around 2000 nm. The positions of these features shift depending on cation substitutions, such as the presence of Ca and Mg, affecting the band positions [60].

2.6.24 Saponite

Saponite displays an Mg-OH absorption feature at 2310 nm and weaker hydration features at 1390 nm and 1910 nm. These features result from Mg-OH vibrations characteristic of Mg-rich smectite, along with structural hydroxyls and interlayer water [67].

2.6.25 Talc

Talc shows a primary Mg-OH absorption feature at 2310 nm and a doublet at 2280 nm due to Mg-OH combination bands. Additionally, it has overtones from carbonate groups and H_2O at 2000 nm and 2370–2380 nm. These absorption features arise from Mg-OH bond vibrations in its trioctahedral sheet silicate structure and carbonate overtones [66].

2.6.26 Wollastonite

Wollastonite presents Si-O overtone vibrations at approximately 1350 nm and minimal absorption features at shorter wavelengths due to the lack of transition metals. These features result from silicate bond vibrations, with the absence of significant electronic involving transition metals [64].

Chapter 3

Specification and Design

This chapter outlines the structure of the proposed EBMA-SD algorithm. We begin by defining the problem that the algorithm aims to solve and describing the representation of bacteria within the system. Following this, we present our objective function and detail the bacterial operations and fuzzy rules utilized. The generation of the initial population is then explained, along with an estimation of the algorithm's complexity. The algorithm is subsequently described in pseudocode to provide a clear procedural overview. Finally, the chapter concludes with a description of the methodology used to create the initial database and how the results are obtained.

3.1 Algorithm Overview

The EBMA-SD algorithm, as depicted in Figure 3.1, leverages Bacterial Genetic Algorithms to embed the representation of spectral deconvolution. Memetic algorithms are employed to explore the solution space through gradient-based optimization, enhancing the efficiency and precision of the search. Instead of performing global search via random mutations, as is usual in traditional BGAs, this approach utilizes a diverse mineral database to explore possible mineral combinations. A fuzzy logic system guides gene modifications during the evolutionary process, playing a pivotal role by determining when to delete a gene, when to fine-tune it, and the extent of such modifications. Altogether, this integrated approach is designed to decompose complex spectral data into meaningful components, preserving essential spectral features and providing interpretable results that facilitate accurate mineral identification.

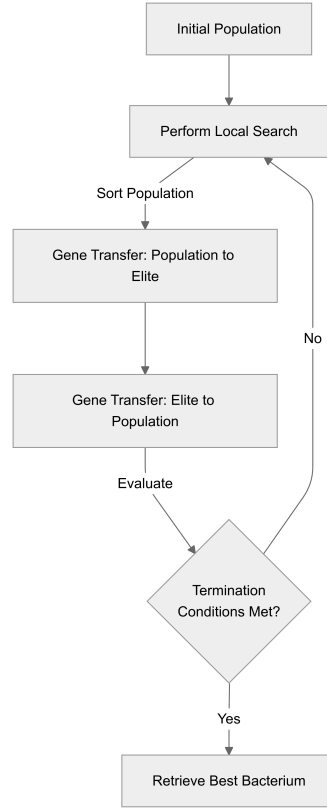


Figure 3.1: Illustration of the EBMA-SD algorithm flow.

3.2 Bacterial Embedding Structure

To achieve an effective decomposition, we propose a hierarchical bacterial evolutionary algorithm (BEA) that models the absorption curve as a sum of Voigt profiles, each representing an individual absorption feature. Our approach organizes the genetic representation into a hierarchical structure consisting of genes, chromosomes, and bacteria, as illustrated in Figure 3.3.

Definiton 1. Gene (Voigt Profile): Each gene g_i is associated with a single Voigt profile $V_i(\lambda)$, characterized by the four parameters defining the Voigt profile function (2.7), along with a fifth parameter used as a label in the database.

- ω_i : Central wavelength of the absorption feature.
- h_i : Height (amplitude) of the absorption feature.
- γ_i : Lorentzian width parameter of the Voigt profile.
- σ_i : Gaussian width parameter of the Voigt profile.
- k_i : Category of the gene (1 to 5).

Definiton 2. Chromosome: A chromosome c is a collection of related genes $\{g_1, g_2, \dots, g_n\}$ (see Figure 3.2) that together encode specific segments or features of the overall absorption curve. This grouping reflects the functional or spectral relatedness of the genes, analogous to how a genome organizes genes with related biological functions.

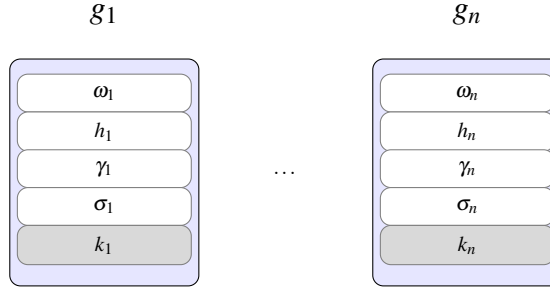


Figure 3.2: Representation of the chromosome as a collection of genes containing Voigt Profile parameters and an additional label parameter k .

Definiton 3. Bacterium: A bacterium b_j represents the decomposition of the target absorption curve into a continuum component minus the summation of the Voigt profiles embedded within its chromosome parameters. Specifically, the absorption curve of bacterium b_j is expressed as:

$$b_j(\lambda) = C(\lambda) - \sum_{i=1}^{G_j} V_i(\lambda), \quad (3.1)$$

where:

- $C(\lambda)$ is the continuum component of the absorption curve.
- G_j is the total number of genes in the chromosome of bacterium b_j .
- $V_i(\lambda)$ denotes the Voigt profile of the i -th gene in bacterium b_j .

3.2.1 Gene Categories

To enhance the physical plausibility of the decomposition and guide the optimization process, we classify the genes (Voigt profiles) into five categories based on their significance and documentation in the literature:

1. **Category 1:** Main absorption features that are well-documented in the literature and consistently observed in the USGS spectral database. These peaks are crucial for mineral identification.

2. **Category 2:** Absorptions that are not individually documented but consistently appear alongside primary peaks in all USGS samples of a specific mineral.
3. **Category 3:** Absorption features that sometimes appear, likely due to contaminations such as metal ions.
4. **Category 4:** Peaks associated with water content in the sample, which can vary depending on environmental conditions.
5. **Category 5:** Voigt profiles with no recognizable pattern, possibly due to noise or unidentified components.

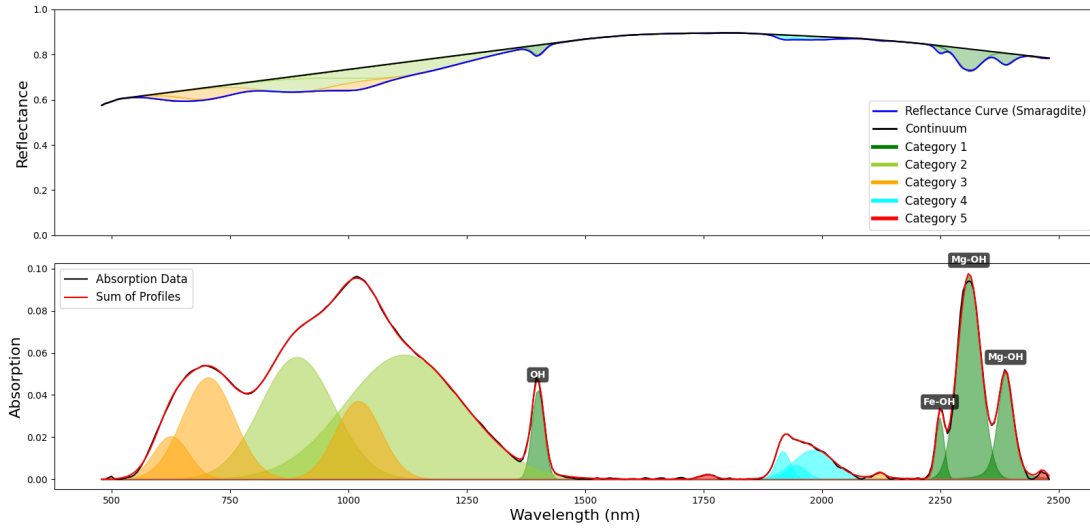


Figure 3.3: The top graph shows the original reflectance curve of a Smaragdite sample and its linear continuum. The bottom graph depicts its Voigt profile decomposition: Category 1 peaks are the most relevant, Category 2 peaks are present in all samples, while Categories 3 to 5 show greater variability.

3.3 Objective Function

When decomposing the absorption curve, an approximation that exceeds the target suggests an overestimation or misinterpretation. Therefore, the proposed solution must always be less than or equal to the target absorption:

$$\sum_{i=1}^{G_j} V_i(\lambda) \leq S(\lambda), \quad \forall \lambda \in [\lambda_{\min}, \lambda_{\max}], \quad (3.2)$$

where:

- $S(\lambda)$ is the target absorption curve.
- $V_i(\lambda)$ are the individual Voigt profiles.
- G_j is the total number of genes (Voigt profiles) in the chromosome of bacterium b_j .

The objective function $F(b_j)$ for each bacterium b_j is designed to evaluate the quality of the approximation while penalizing overestimation. We define the objective function as:

$$F(b_j) = E(b_j) + \alpha \cdot O(b_j), \quad (3.3)$$

where:

- $E(b_j)$ is the squared distance (error) between the target absorption curve $S(\lambda)$ and the bacterium's approximation $B(\lambda)$:

$$E(b_j) = \int_{\lambda_{\min}}^{\lambda_{\max}} [S(\lambda) - B(\lambda)]^2 d\lambda. \quad (3.4)$$

- $O(b_j)$ is the overestimation penalty, calculated as:

$$O(b_j) = \int_{\lambda_{\min}}^{\lambda_{\max}} \max[B(\lambda) - S(\lambda), 0] d\lambda. \quad (3.5)$$

- α is a weighting factor that controls the emphasis on the overestimation penalty.

3.4 Bacterial Operations

The bacterial operations are designed to maintain the diversity of the bacterial population while exploring the solution space effectively. We introduce a hierarchical mutation strategy that operates at different levels: gene, chromosome, and bacterium.

3.4.1 Local Search

To start, a local search is performed on all bacteria using the selected genes and the target absorption curve. This involves adjusting the parameters using the Levenberg-Marquardt optimization algorithm with reduced boundaries.

- **Primary Genes (Category 1):** Can change their size (height h_i and width parameters γ_i , σ_i) but not significantly their central position ω_i . If a primary gene needs to be changed too much or deleted, it is likely that the bacterium does not represent the correct mineral.
- **Companion Genes (Category 2):** Similar to primary genes, they can adjust size and width parameters but have limited changes to their central positions.
- **Less Significant Genes (Categories 3 to 5):** Have more flexibility to change their parameters, including central position ω_i , and can be deleted or merged as needed.

3.4.2 Gene Transfer

Following the local search, we calculate the fitness vector for each bacterium and assign a fitting score to each gene according to the average of the fitness vector in the gene's influence range. The genes are ranked by fitness, and decisions about gene transfer are guided by fuzzy logic rules, as detailed in the next section.

3.5 Fuzzy Logic for Gene Transfer and Local Search Boundaries

This section describes the integration of fuzzy logic into the EBMA-SD, with membership functions and fuzzy rules designed based on expert knowledge.

3.5.1 Fuzzy Variables and Membership Functions

To implement fuzzy logic in our algorithm, we define the following fuzzy variables, which consider the quality of the curve approximation and the significance of the peak, to determine how abrupt the changes in the genes can be:

1. **Gene Fitness Level (GFL):** (*Antecedent*) This variable measures how effectively a gene contributes to fitting the target curve. It is categorized as *Poor*, *Fair*, or *Good*.
2. **Gene Importance (GI):** (*Antecedent*) This variable assesses the significance of a gene, classified from *Very Low* to *Very High*, corresponding to gene categories 1 through 5.

3. **Action Potential (AP):** (*Consequence*) Based on the antecedent variables, this variable determines the degree to which a gene should be modified or deleted. Its levels range from *Very Low* to *Very High*.

Each fuzzy variable is associated with a membership function that maps numerical values to degrees of membership in linguistic terms.

3.5.2 Fuzzy Rules for Action Potential

We define a set of fuzzy rules that use the fuzzy variables to guide gene transfer decisions. The Table 3.1 provides a clear understanding of how various combinations of gene fitness and importance affect the action potential. For example:

- If a gene has a *Poor* fitness level and a *Very Low* importance, the action potential is *Very High*, suggesting that the gene should be considered for deletion or replacement.
- If a gene has a *Fair* fitness level and a *High* importance, the action potential is *Medium*, indicating that the gene's parameters may need adjustment.
- If a gene has a *Good* fitness level and a *Very High* importance, the action potential is *Very Low*, implying that the gene should be maintained.

GI \ GFL	Poor	Fair	Good
Very Low	Very High	High	Medium
Low	High	Medium	Low
Medium	Medium	Medium	Low
High	Low	Medium	Very Low
Very High	Very Low	Low	Very Low

Table 3.1: Action Potential (Consequence) Based on Gene Fitness Level and Gene Importance (Antecedents). The table maps *Gene Importance* (GI) against *Gene Fitness Level* (GFL) to determine the corresponding *Action Potential*.

3.5.3 Fuzzy Inference System for Gene Transfer

The fuzzy inference system (FIS) processes the inputs (gene fitness levels and importance) through the rule base to produce outputs that guide gene transfer actions. The steps involved are:

1. **Fuzzification:** Convert crisp input values of the numerical gene fitness scores into fuzzy degrees of membership using the membership functions.
2. **Rule Evaluation:** Apply the fuzzy rules by computing the degree to which each rule's antecedent conditions are satisfied.
3. **Aggregation:** Combine the outputs of all rules that pertain to a particular action.
4. **Defuzzification:** Convert the aggregated fuzzy output back into a crisp decision.

3.5.4 Action Strategy Based on Action Potential

Once the Action Potential (AP) level is determined, the EBMA-SD algorithm applies specific gene transfer strategies to adjust the gene parameters accordingly. The table 3.2 outlines the allowable local search boundaries and corresponding actions for each AP level.

AP Level	Central Wavelength Shift	Height Variation	Width Variation	Additional Actions
Very Low	± 5 nm	$\pm 25\%$	$\pm 15\%$	-
Low	± 10 nm	$\pm 50\%$	$\pm 25\%$	-
Medium	± 15 nm	$\pm 75\%$	$\pm 35\%$	-
High	± 25 nm	0% to +75%	$\pm 50\%$	Allow deletion of gene if height reaches zero
Very High	-	-	-	Search for better genes from another bacterium

Table 3.2: Gene Transfer Strategies Based on Action Potential (AP) Levels

Regions with errors where no Voigt profile is nearby are considered as having a very high AP level, and thus a Voigt profile will be searched for among other bacteria.

3.6 Final Solution

The final bacterium will be the best-performing one according to the objective function 3.3 after meeting the stopping criteria. This bacterium will hold the solution iden-

tified as the most suitable based on the database used. The classification of the mineral will involve extracting the mineral from the final bacterium and verifying if it contains the primary peaks of another bacterium. If so, it will be classified as a mixture of both minerals.

3.7 Initial Population Generation

The initial population (P_0) of bacteria is generated by taking spectral data from the USGS spectral library [70]. Each mineral's spectrum is decomposed into Voigt profiles (genes) representing individual absorption features. These genes have been manually labeled and pre-categorized based on their significance (Categories 1 to 5).

Bacteria are grouped by minerals, and within each mineral group, all bacteria share the same number of Category 1 and 2 peaks, ensuring consistency in representing the essential features of that mineral. Categories 3 to 5 genes can vary among bacteria within the same mineral.

For each mineral group:

1. **Internal Optimization:** Apply the EBMA-SD algorithm using the target absorption curve as the objective. Each bacterium within the group is optimized to best fit the target curve, focusing on adjusting Categories 3 to 5 genes while maintaining only one gene per shared Category 1 and 2 peaks.
2. **Selection of Representative Bacterium:** From each mineral group, select the bacterium that achieves the best fit to the target curve. This bacterium represents the mineral in the main algorithm.

The initial population for the main EBMA-SD algorithm thus consists of the best representative bacterium from each mineral group.

3.8 Algorithm Pseudocode

Algorithm 1 Elitist Bacterial Memetic Algorithm for Spectral Decomposition with Fuzzy Logic (EBMA-SD)

```

1: Input: Target absorption curve  $S(\lambda)$ , maximum iterations  $G_{\max}$ 
2: Initialize population  $P_0$  with best bacteria from each mineral group      ▷ Initial
   Population Generation
3: Set iteration counter  $g = 1$ 
4: while not converged and  $g \leq G_{\max}$  do
5:   for each bacterium  $b_j$  in  $P_0$  do
6:     Perform local search on  $b_j$ 
7:     Evaluate objective function  $F(b_j)$                                 ▷ Equation ??
8:   end for
9:   Select elite bacterium  $b^*$  with the best fitness
10:  Gene Transfer from Rest to Elite
11:  for each bacterium  $b_j$  in  $P_0$ ,  $b_j \neq b^*$  do
12:    Identify genes in  $b_j$  that can improve  $b^*$ 
13:    Use fuzzy logic to decide on gene transfer to  $b^*$ 
14:  end for
15:  Gene Transfer from Elite to Rest
16:  for each bacterium  $b_j$  in  $P_0$ ,  $b_j \neq b^*$  do
17:    Identify deficiencies in  $b_j$ 
18:    Use fuzzy logic to decide on gene transfer from  $b^*$  to  $b_j$ 
19:  end for
20:  Update population  $P_0$ 
21:  Increment iteration counter  $g = g + 1$ 
22: end while
23: Output: Best bacterium  $b^*$  and classification results

```

3.9 Algorithm Complexity

The computational complexity of the EBMA-SD algorithm is influenced by several factors, including the number of bacteria in the population, the number of iterations, and the complexity of the operations performed on each bacterium. Below, we break down the complexity associated with each major component of the algorithm:

- **Fitness Evaluation:** Each fitness evaluation involves computing the error $E(b_j)$ and the overestimation penalty $O(b_j)$ for a bacterium b_j . These computations require integrating over the wavelength range, which is discretized into W points. Therefore, the complexity for evaluating the fitness of a single bacterium is $O(K)$. Since there are N bacteria in the population, the total complexity for fitness evaluations per iteration is $O(N \times K)$.

- **Mutation and Gene Transfer Operations:** Mutation involves adjusting the parameters of genes within each bacterium's chromosome. The complexity of mutation depends on the number of genes per bacterium, denoted as G . For gene transfer operations, which include transferring genes between bacteria based on fuzzy logic decisions, the complexity is also proportional to G . Consequently, the complexity for mutation and gene transfer per bacterium is $O(G)$, leading to a total complexity of $O(N \times G)$ per iteration.
- **Local Search:** The local search process, such as the Levenberg-Marquardt optimization algorithm, is applied to refine the parameters of each bacterium. The complexity of the local search depends on the number of parameters being optimized, typically proportional to G . Assuming the local search converges in a constant number of steps, the complexity per bacterium remains $O(G)$, resulting in an overall complexity of $O(N \times G)$ per iteration.
- **Fuzzy Logic Evaluation:** The fuzzy inference system evaluates rules based on gene fitness and importance to determine action potentials for gene transfer. The number of fuzzy rules is fixed and independent of N or G , leading to a constant complexity $O(1)$ per gene. Thus, for all genes in the population, the complexity is $O(N \times G)$ per iteration.
- **Overall Complexity:** Combining all components, the total computational complexity per iteration is:

$$O(N \times K) + O(N \times G) + O(N \times G) + O(N \times G) = O(N \times (K + G))$$

Therefore, the overall complexity per iteration is linear with respect to both the population size N and the number of genes G , scaled by the discretization of the wavelength range W .

3.10 Evaluation Methods and Performance Metrics

As the objective of this algorithm is to fit the target absorption curve using meaningful components for effective classification, we will analyze the feature selection process alongside metrics derived from the confusion matrix to evaluate classification performance.

To confirm that the algorithm effectively converges on the expected solution, we will evaluate a scenario where a sample of a mineral absent from our database is processed. This assessment aims to determine how closely the algorithm's approximation aligns with a manually generated solution when using a diverse database devoid of samples of that mineral. The process involves systematically excluding samples of each mineral and applying the model to the excluded sets. The results will reveal which minerals exhibit unique peaks that the model finds challenging to accurately infer.

3.11 Dataset Creation

The initial database is created using spectral data from the USGS spectral library [16]. An interface was internally developed to facilitate the visualization, manual adjustment, and labeling of the genes. The number and category of Voigt profiles were determined through an iterative process based on the data found in the literature discussed in the Section Geology, along with the observation of patterns within the dataset. Metadata such as sample name, formula, mineral, and mineral group were also added to the bacteria. Figure B.1 provides a visualization of the Carbonate group, illustrating an example of the input reflectance data, absorption curve, and bacterium representation.

Chapter 4

Implementation and Usage

This chapter details the implementation of the methodologies discussed in Chapter 3. We begin by describing the development of a Python-based interface for managing the algorithm’s data structures. Subsequently, we provide examples illustrating the objective function, graphical representations of the local search, and the gene transfer processes. Additionally, we visualize the fuzzy inference systems and present examples of generating initial populations and the final solution.

4.1 Software and Tools

To facilitate the creation of the database, data management, and visualization, we developed a custom desktop application (see Figure A.1). This application allows users to load both the USGS database and SGU drill cores from `.raw` files, along with the corresponding wavelength data. The interface enables the graphical definition of initial parameters p_0 for curve fitting using SciPy’s `curve_fit` function in Levenberg-Marquardt (LM) mode. Additionally, it allows for the labeling of each Voigt profile according to its category and associated phenomenon.

4.2 Objective Function Visualization

Because the components of the deconvolution represent the use of energy to excite features within the sample material, and since these components are additive, any exceeding of the absorption curve implies an overestimation of the energy used in that region, which is physically incorrect. To address this, we empirically investigated the effect of the

weighting factor α in the objective function (Equation 3.3). As illustrated in Figure 4.1, using $\alpha = 1$ (effectively not penalizing overestimation) results in the fitted values exceeding the target curve, particularly at the edges. This occurs because the reduction in underestimation error compensates for the error introduced by overestimation. We found that setting α between 3 and 4 yields solutions that closely align with the expected physical behavior.

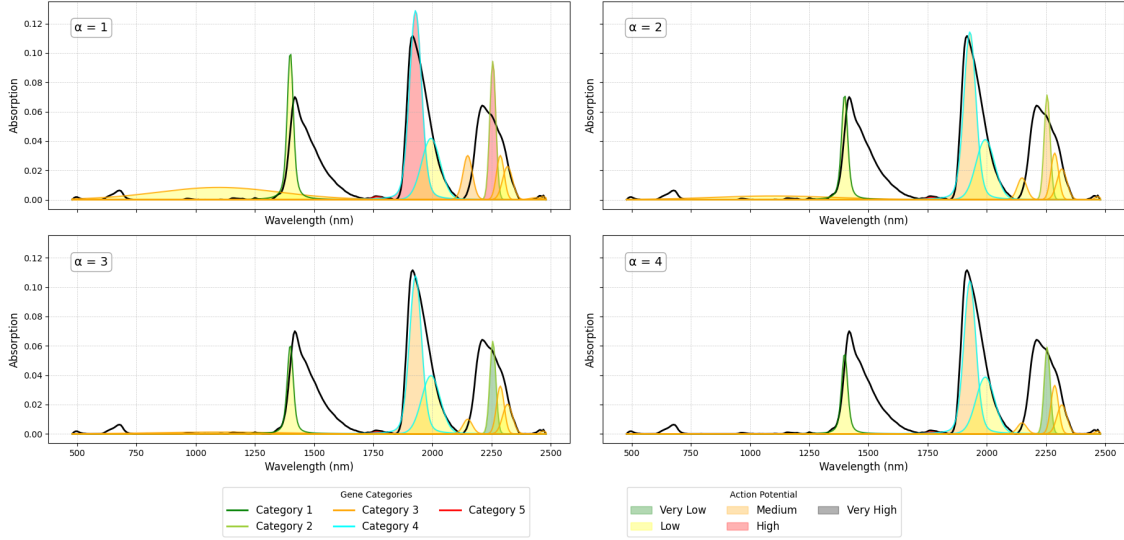


Figure 4.1: Effect of the weighting factor α in the objective function on fitting the Voigt profiles during the local search. Higher values of α prevent the approximation from exceeding the target curve.

4.3 Local Search Visualization

To generate the initial population, a local search is conducted, focusing exclusively on adjusting the heights of the Voigt profiles within the bacterial population. This approach utilizes the objective function to effectively approximate the absorption curve (see Figure 4.2). By restricting the fitting process, it becomes possible to distinguish minerals with closely situated absorption peaks. For instance, in Figure 4.2, the results are displayed for the three minerals that best match the target, an Actinolite sample. The closest match is another Actinolite sample, followed by two minerals belonging to the same Amphibole group. While all samples exhibit similar characteristic peaks, yielding a generally good fit, the Actinolite sample achieves the most accurate alignment due to the better correspondence of its peak centers and shapes.

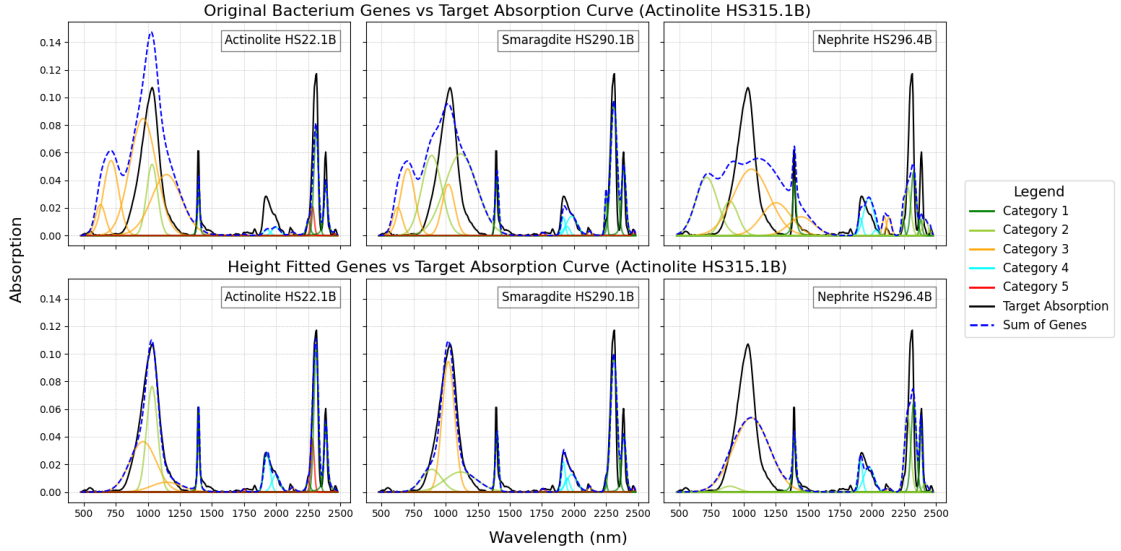


Figure 4.2: Initial LM optimization by adjusting gene heights to fit the target Actinolite HS315.1B. The comparison includes Actinolite HS22.1B, Smaragdite HS290.1B, and Nephrite HS296.4B, demonstrating the effectiveness of restricted fitting in distinguishing closely related minerals.

In the main optimization loop, we adjust all four parameters of the Voigt profiles: central wavelength, height, and the width parameters (γ and σ). To limit computational complexity, we consider a maximum of six Voigt profiles at a time, and these profiles must be consecutive. Genes with very high Action Potential (indicating a poor fit and low importance) are eliminated. For the remaining genes, the local search boundaries are specified based on their Action Potential, which is influenced by the gene's category, as illustrated in Figure 4.3. Genes that typically vary within samples of a mineral are granted greater freedom for adjustment. Conversely, genes representing specific, well-documented features—often associated with vibrational effects—are modified minimally, even if they do not yield an optimal fit.

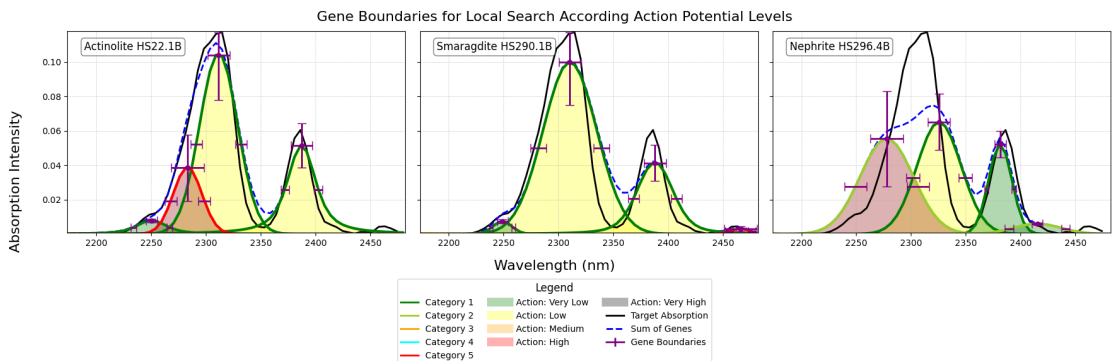


Figure 4.3: Visualization of the Action Potential of the Voigt profiles and the corresponding central wavelength boundaries during the local search.

For instance, the absorption peak near 2305 nm is characteristic of hydroxide minerals and is associated with Mg and Fe atoms and O-H bond vibrations. Small variations in its central wavelength can be significant. In Figure 4.3, we observe that the model achieves a good fit for most peaks, indicating the presence of Mg and Fe in the sample. However, the misalignment of the peak at 2305 nm suggests that the sample may not be exactly the same mineral or may have a different crystal structure compared to the reference bacterium used for approximation.

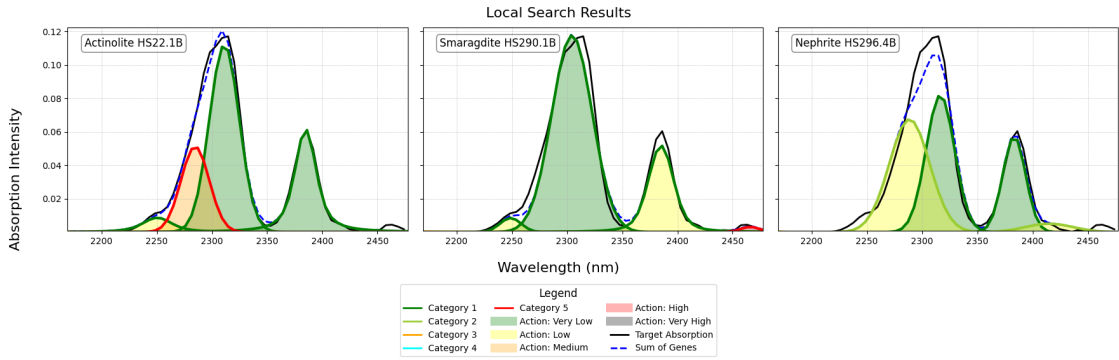


Figure 4.4: Visualization of the results after the local search optimization.

Following the local search, we compute the Action Potential of the adjusted genes. Genes that initially had high Action Potential but have improved to a medium range or lower are retained. Those that remain with high Action Potential are removed prior to initiating the gene transfer phase.

4.4 Gene Transfer Visualization

By using a weighting factor α greater than 1 in the objective function, the solutions from the local search tend to underestimate the absorption curve. To correct this, we proceed segment by segment, determining whether to add a new gene from another bacterium or overwrite an existing one in regions where the fit is inadequate. To justify adding a new gene, the error in the segment must be at least 0.001, ensuring that the new gene's contribution is significant.

In Figure 4.5, the elite bacterium (representing the best current solution) searches for genes from other bacteria to improve fitness in specific regions highlighted in purple. This process embodies the gene transfer mechanism guided by fuzzy logic rules, facilitating the incorporation of beneficial genetic material from various sources. In this case, the

result was the incorporation of the water-related gene near 1450 nm. Additionally, a peak at 2115 nm was included as it narrowly surpassed the threshold.

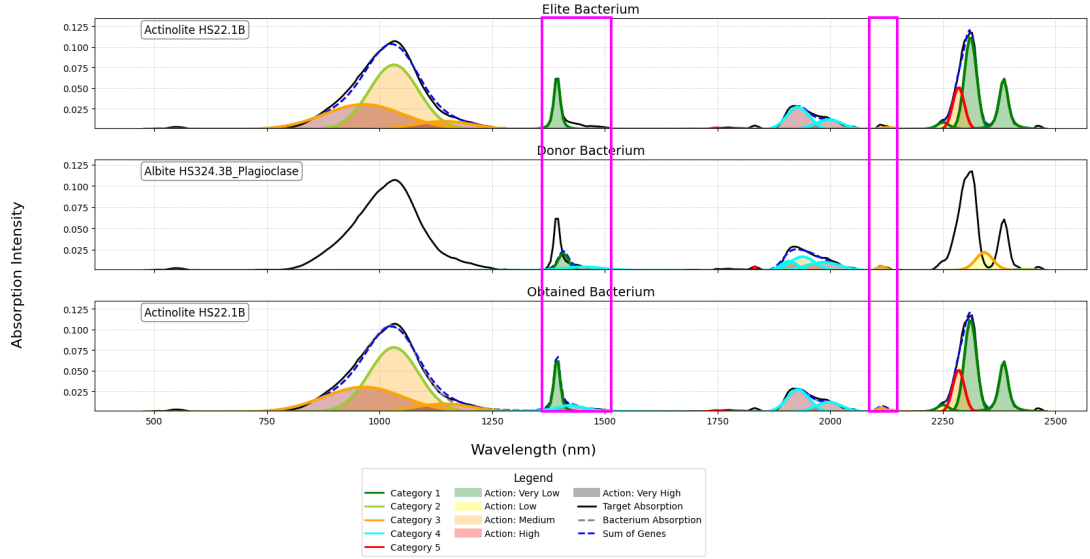


Figure 4.5: The elite bacterium searching for genes in other bacteria in regions highlighted in purple, where the fitness is insufficient and greater than the threshold.

4.5 Fuzzy Logic Visualization

Fuzzy logic serves as a fundamental component of our algorithm by regulating the flexibility of gene parameter adjustments based on their relative importance. It functions as a dynamic buffer for parameter boundaries during optimization. For highly significant peaks, such as those categorized as Category 1 genes, the permissible variation in parameters is minimal to preserve the consistency of critical spectral features. This ensures that minerals closely aligned with the sample achieve an optimal fit early in the process.

Figure 4.6 depicts the influence of gene category on the allowable parameter adjustments during optimization. Genes assigned lower importance (e.g., Categories 3 to 5) are afforded broader parameter boundaries, enabling more extensive modifications. In contrast, essential genes (Categories 1 and 2) are subject to tightly constrained boundaries.

It can also be noticed that peaks within Categories 1 and 2 exhibit increased boundaries only under Fair Gene Level Fitness conditions. For Good or Poor fitness levels, the adjustments are more restricted. This design minimizes the use of these critical absorption features for solution-space exploration, instead prioritizing lower-importance categories to drive exploratory adjustments.

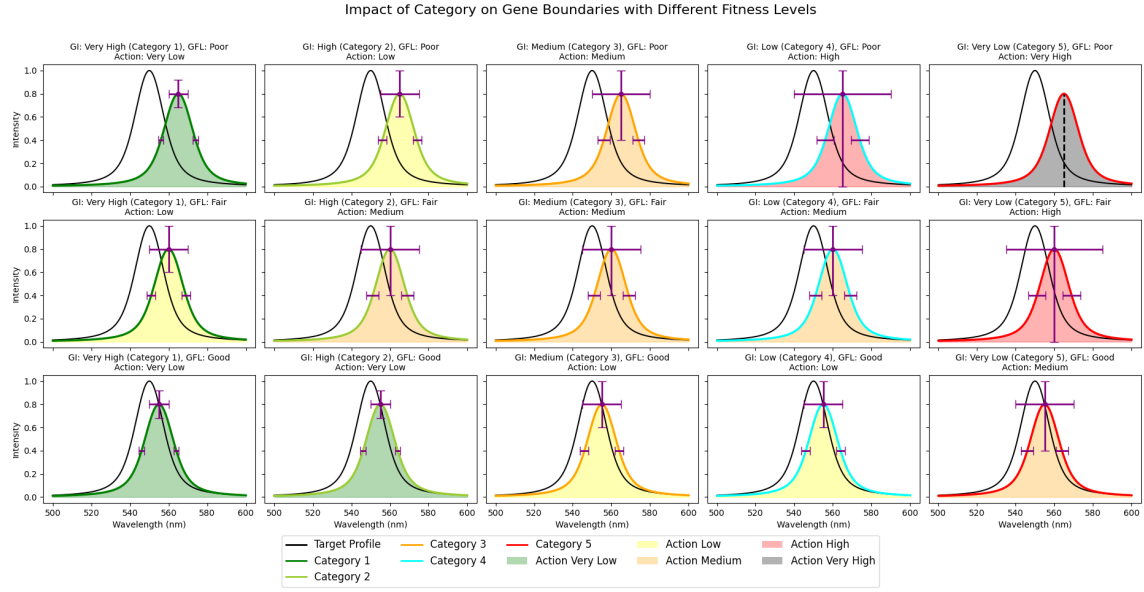


Figure 4.6: Impact of gene category on permissible parameter boundaries during the local search, as defined by the Action Potential.

4.6 Final Solution Visualization

The algorithm culminates in a ranking of bacteria based on their ability to approximate the target absorption curve. If a sample from the database corresponds to the target, there is a high probability of obtaining it as the top-ranking mineral among the extensive set of samples. This can be observed in Figure 4.7, where another sample of the same mineral was retrieved, exhibiting a gene pattern highly similar to the original.

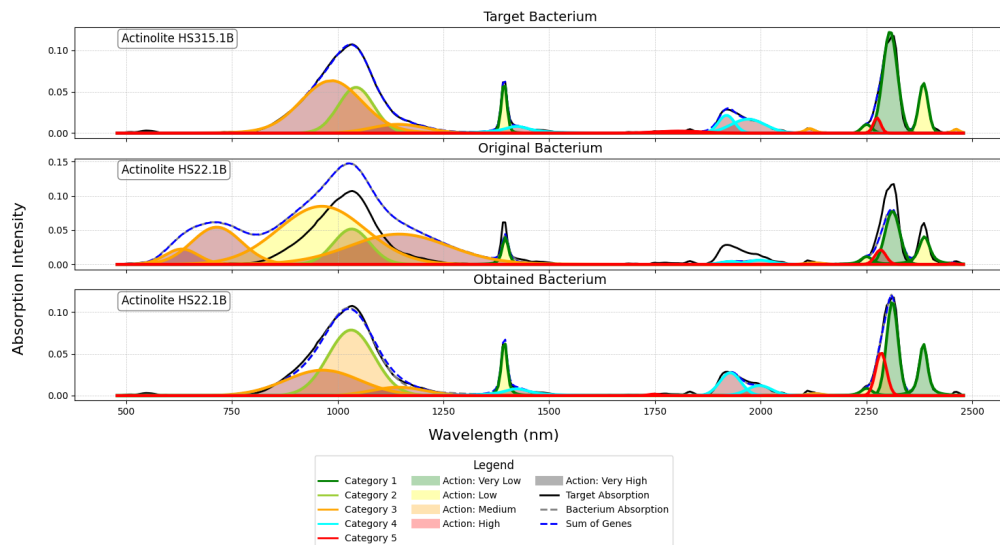


Figure 4.7: The target absorption curve (top), followed by the original elite bacterium and its updated state after applying the EBMA-SD algorithm.

However, even if an exact sample of the same mineral is not present in the database, the minerals with the next-best fitness scores can still provide valuable insights into the potential family or mineral group. For example, as shown in Figure 4.2, which illustrates the top three minerals after the initial local search, all three top-ranking bacteria belong to the same mineral group. If one sample were absent, another would take its place, offering clues about the type of mineral and its elemental composition.

Chapter 5

Experimental Results and Conclusions

Because there are many minerals in the database, we will provide more details about those used by Rialland et al. [8, 56] to compare the benefits offered by our proposal.

Table 5.1: Some minerals of interest in the USGS spectral library [8].

Mineral - Group	Sample
Calcite - Carbonate	splib07b_Calcite_GDS304_75-150um_ASDFRb_AREF
Dolomite - Carbonate	splib07b_Dolomite_HS102.4B_ASDNGb_AREF
Gypsum - Sulphate	splib07b_Gypsum_HS333.4B_ASDFRa_AREF
Kaolinite - Phyllosilicate	splib07b_Kaolinite_CM9_ASDNGb_AREF
Nontronite - Smectite	splib07b_Nontronite_NG-1.a_ASDNGb_AREF

5.1 Deconvolution Results

For this scenario, we analyzed each sample in Table 5.1, excluding all instances of that mineral from the database. For Calcite, the closest mineral found was Strontianite HS272.1B, another carbonate from the Aragonite group. This result is consistent, as a good fit was achieved, preserving the characteristic peaks of carbonates.

For Dolomite, the initial result was Serpentine due to its characteristic magnesium peaks around 1100 nm. However, it lacked a key carbonate peak near 1800 nm, leading to the selection of genes from Calcite. This allowed for the incorporation of magnesium peaks, resulting in magnesium-bearing Calcite, a geologically coherent outcome. Calcite outperformed Serpentine because Serpentine’s characteristic peak near 1400 nm could not be eliminated, making Calcite a better complement.

For Nontronite, the initial match was a Diopside sample. Both are silicates and share iron absorption features. However, the 1400 nm peak was not close enough, and

Diopside's water absorption features were not prominent. This led to the selection of genes from Gypsum for water absorption and an increase in the Fe-OH peak around 2380 nm. This result suggests a hydrated mineral with Fe-OH bonds.

The Kaolinite sample quickly converged with Halloysite NMNH106237, a mineral from the Kaolinite group. It preserved the central features of the double absorption around 2160 and 2208 nm but exhibited a sharper shape, indicating differences in crystallinity. Gypsum also converged rapidly with Carnallite, as both are hydrated minerals.

5.2 Classification Results

To evaluate the scenario where at least one sample of a mineral is present in the database, we excluded all but one sample of each mineral. The detailed results are presented in the appendix. On average, the correct sample was identified as the best match in a significant percentage of cases. Figure 5.1 summarizes the results by mineral type. As shown, oxides and some sulfates were often confused with silicates. This is because these groups primarily exhibit metal absorption features, which are less distinctive and can resemble multiple minerals.

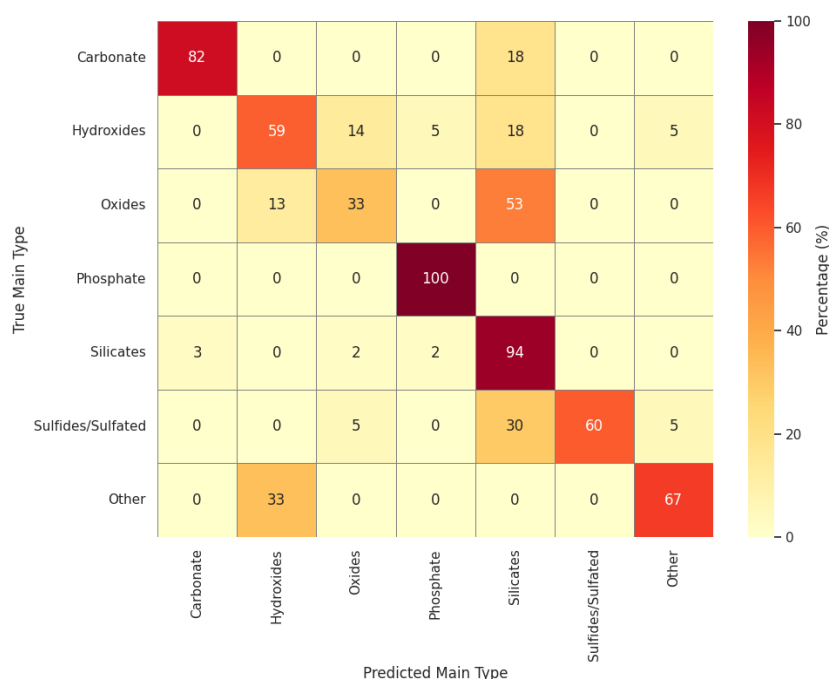


Figure 5.1: Classification confusion matrix by mineral type.

5.3 Final Conclusion

This work presents a comprehensive and innovative approach to addressing the challenges of mineral identification using spectral data, focusing on the development and application of the Elitist Bacterial Memetic Algorithm for Spectral Deconvolution (EBMA-SD). By integrating bacterial memetic algorithms with fuzzy logic, the research achieves a balance between computational efficiency and interpretability, which is critical in geoscientific applications.

The EBMA-SD algorithm is designed to decompose complex spectral data into meaningful components, leveraging a curated database of Voigt profiles. This methodology ensures precise spectral decomposition while preserving key spectral features essential for accurate mineral classification. A notable innovation lies in the incorporation of fuzzy logic to guide gene transfer and local optimization processes, enhancing the adaptability and physical plausibility of the spectral solutions.

Empirical results underscore the robustness of the algorithm, particularly its ability to classify minerals accurately with limited labeled data—a common constraint in geological studies. The method effectively addresses longstanding challenges, such as overlapping spectral features, variability in spectral data, and the need for interpretable results. Validation on the USGS spectral database demonstrates high classification accuracy across diverse mineral types, although the algorithm encounters difficulties with minerals like oxides, sulfates, and quartz, which lack distinctive spectral peaks within the studied range.

The study identifies promising avenues for future research, including refining the handling of minerals with indistinct features and expanding the algorithm's database to enhance its generalization capabilities. These improvements could extend its applicability to broader geological contexts, such as resource exploration, environmental monitoring, and planetary sciences.

In conclusion, this work represents a different approximation in mineral classification methodologies. It exemplifies how the integration of computational intelligence with domain-specific knowledge can address complex real-world problems, providing a robust framework for spectral analysis in mineralogical and geological studies.

5.4 Limitations and Future Work

This study identifies several avenues for future refinement and development. The continuum calculation method warrants further optimization to enhance accuracy and reliability. Additionally, a more in-depth investigation into peak shifts, electronic transitions, and the organization of genes into chromosomes based on their type or physical origin is crucial to advancing the methodology. While this research underscores the potential of bacterial algorithms in mineral spectroscopy, additional studies and experimental validation are needed to fully realize and refine their capabilities. A promising direction for future work involves the parallel analysis of complete hyperspectral images using swarm-based algorithms. This approach could allow the most representative mineral to dominate and propagate its genetic information across the two-dimensional spectral space, further improving the efficiency and accuracy of mineral classification.

Appendix A

Interface for Database Creation

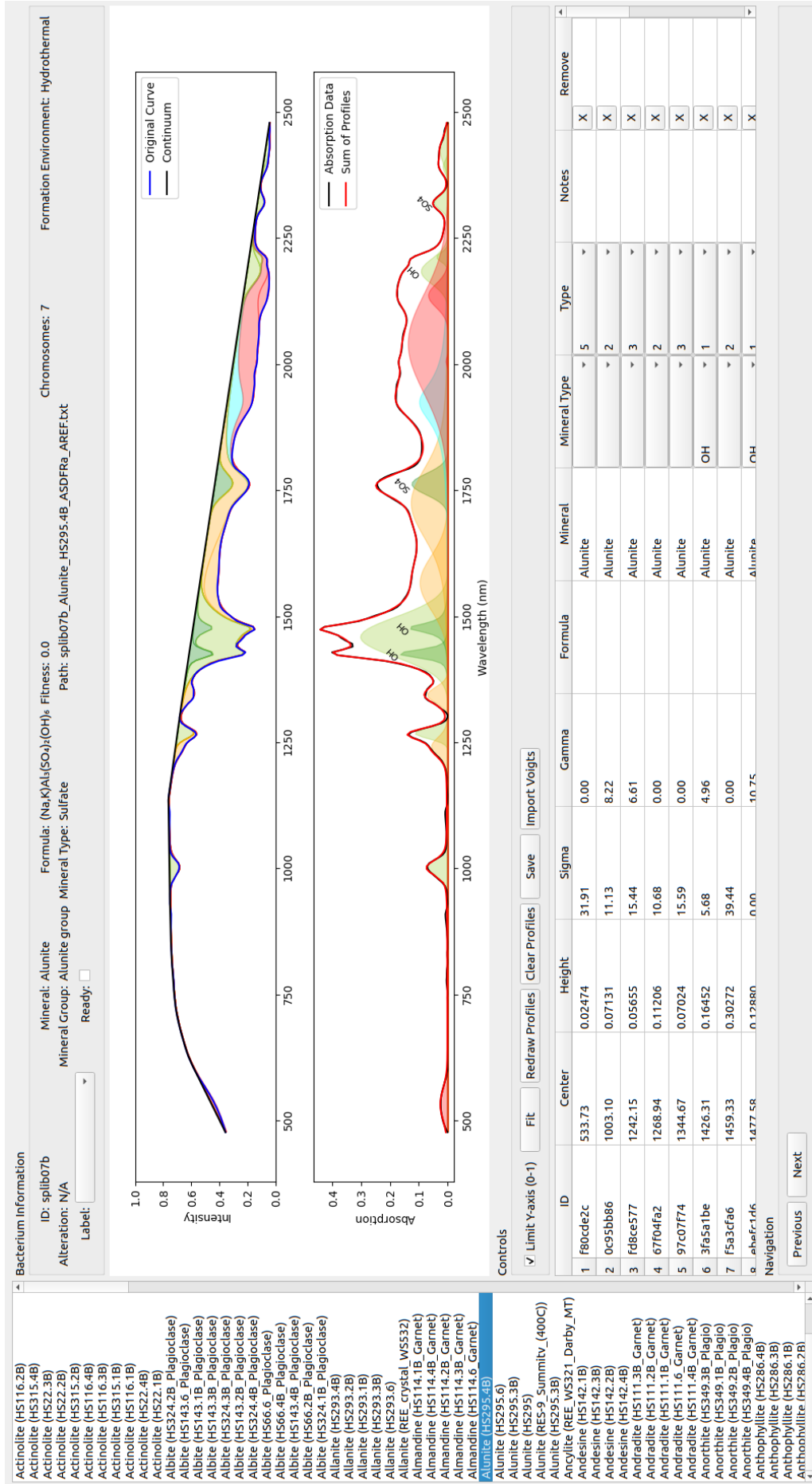


Figure A.1: Screenshot of the interface used to create the database.

Appendix B

Visualization of Carbonate Minerals

Spectral Curves and Gene Representation - Mineral Type: Carbonate

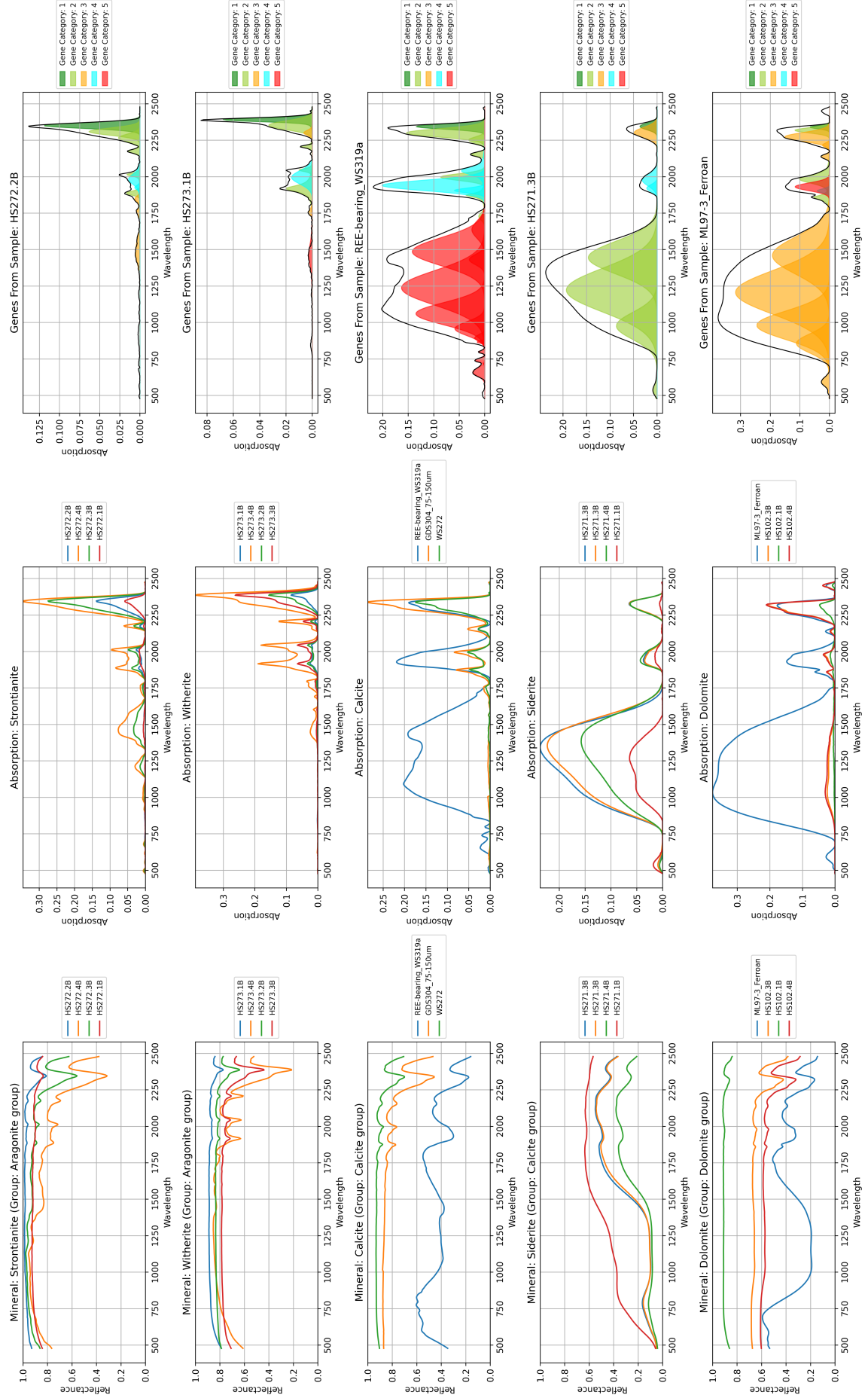


Figure B.1: Visualization of carbonate minerals: the first column shows the reflectance curve, the middle displays the absorption curve, and on the right is the representation of the genes in bacterial format, with genes color-coded according to their category.

Appendix C

Detailed Confusion Matrix

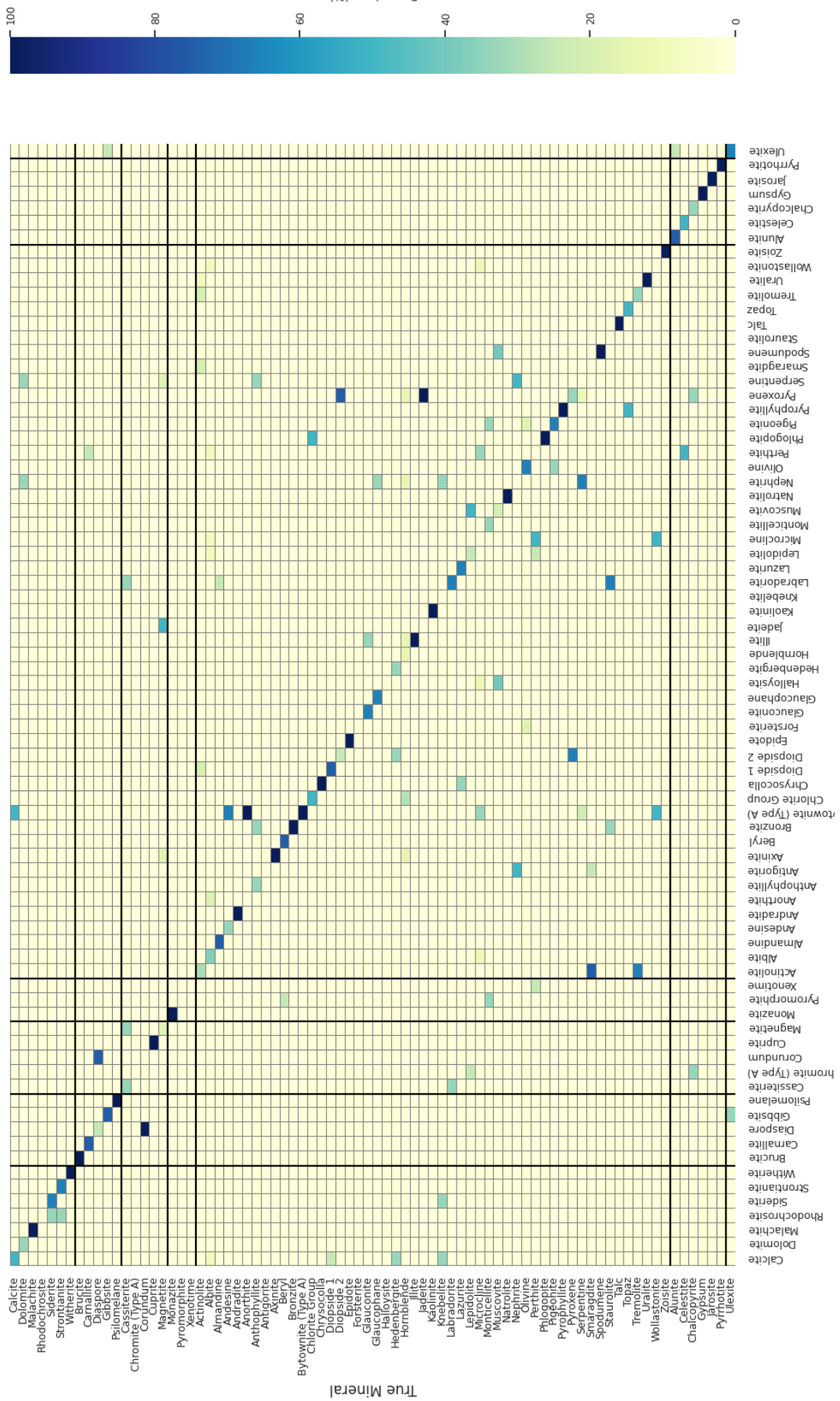


Figure C.1: Detailed confusion matrix resulting from the mineral classification using only one sample of each mineral in the database.

Bibliography

- [1] Roger N. Clark. Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy. Ed. by Andrew N. Rencz. John Wiley & Sons, 1999, pp. 3–58.
- [2] Fred A. Kruse. “Mapping surface mineralogy using imaging spectrometry”. In: Remote Sensing 4.1 (2012), pp. 155–179. DOI: 10.3390/rs4010155.
- [3] Fuan Tsai and William Philpot. “Derivative analysis of hyperspectral data”. In: Remote Sensing of Environment 66.1 (1998), pp. 41–51.
- [4] Freek Van der Meer and et al. “Remote sensing and mineral exploration”. In: International Journal of Applied Earth Observation and Geoinformation 13.1 (2012), pp. 1–5. DOI: 10.1016/j.jag.2011.11.002.
- [5] P. Kumari et al. “A Fully-Automated Framework for Mineral Identification on Martian Surface Using Supervised Learning Models”. In: IEEE Access 11 (2023). DOI: 10.1109/ACCESS.2023.3243061.
- [6] Y. Liang et al. “Unbalanced Geologic Body Classification of Hyperspectral Data Based on Squeeze and Excitation Networks at Tianshan Area”. In: IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium. 2020. DOI: 10.1109/IGARSS39084.2020.9323795.
- [7] Freek Van der Meer. “Analysis of spectral absorption features in hyperspectral imagery”. In: International Journal of Applied Earth Observation and Geoinformation 8.1 (2006), pp. 3–17. DOI: 10.1016/j.jag.2005.06.007.
- [8] R. Rialland et al. “Improved deconvolution of mineral reflectance spectra”. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14 (2021), pp. 9711–9726.
- [9] Norberto Nawa and Takeshi Furuhashi. “Bacterial Evolutionary Algorithm for Fuzzy System Design”. In: (Dec. 1998). DOI: 10.1109/ICSMC.1998.725020.

- [10] J. Botzheim et al. “Fuzzy rule extraction by bacterial memetic algorithms”. In: International Journal of Intelligent Systems 24.3 (2009), pp. 312–339.
- [11] Krzysztof Pytel. “Fuzzy logic applied to mutation size in evolutionary strategies”. In: Evolutionary Intelligence 17 (2024), pp. 2433–2451.
- [12] Richard Schodde. “Challenges and Opportunities for Under-Cover Exploration in Australia”. In: UNCOVER Summit. Accessed from MinEx Consulting presentations. MinEx Consulting. Mar. 2014. URL: <https://www.minexconsulting.com/publications>.
- [13] Barry A. Wills and James Finch. Wills’ Mineral Processing Technology: An Introduction to the Practice 8th. Oxford, UK: Butterworth-Heinemann, 2015. ISBN: 9780080970530.
- [14] Barry H.G. Brady and E. T. Brown. Rock Mechanics for Underground Mining. 3rd. Dordrecht, Netherlands: Springer, 2006. ISBN: 9781402020643. DOI: 10.1007/1-4020-2064-3.
- [15] Bernd G. Lottermoser. Mine Wastes: Characterization, Treatment and Environmental Impacts. 3rd. Berlin, Germany: Springer, 2010. ISBN: 9783642124181. DOI: 10.1007/978-3-642-12419-8.
- [16] Roger N Clark et al. USGS digital spectral library splib06a. Tech. rep. US Geological Survey, 2007.
- [17] Donald W. Marquardt. “An algorithm for least-squares estimation of nonlinear parameters”. In: Journal of the Society for Industrial and Applied Mathematics 11.2 (1963), pp. 431–441. DOI: 10.1137/0111030.
- [18] Kari Hyll. “Infrared emittance of paper - Method development, measurements, and application”. PhD thesis. Jan. 2012. DOI: 10.13140/RG.2.1.3457.8965.
- [19] Roger G. Burns. Mineralogical Applications of Crystal Field Theory. 2nd Edition. Cambridge, UK: Cambridge University Press, 1993.
- [20] Raymond F Kokaly et al. USGS spectral library version 7. Tech. rep. US Geological Survey, 2017.
- [21] A.V. Medvedovici and V. David. “Spectrophotometry | Diode Array”. In: Encyclopedia of Analytical Science. Second Edition. Elsevier, 2005.

- [22] J. W. Boardman. “Geometric and Radiometric Corrections of Hyperspectral Data: Improving Signal-to-Noise Ratio for Mineral Mapping”. In: Proceedings of the 8th JPL Airborne Earth Science Workshop. NASA Jet Propulsion Laboratory. Pasadena, CA, 1999.
- [23] Postel. Signal-to-Noise Ratio Explained. Accessed: October 27, 2024. 2024. URL: <https://ghicn.org/postels-blog-oct-27-signal-to-noise-ratio/>.
- [24] Fred A Kruse et al. “The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data”. In: Remote sensing of environment 44.2-3 (1993), pp. 145–163.
- [25] Artzai Picón et al. “Fuzzy Spectral and Spatial Feature Integration for Classification of Nonferrous Materials in Hyperspectral Data”. In: IEEE Transactions on Industrial Informatics 5.4 (2009), pp. 483–494. DOI: 10.1109/TII.2009.2031238.
- [26] Nasrullah Zaini, Freek Van der Meer, and Harald Van der Werff. “Effect of Grain Size and Mineral Mixing on Carbonate Absorption Features in the SWIR and TIR Wavelength Regions”. In: Remote Sensing 4.4 (2012), pp. 987–1003. ISSN: 2072-4292. DOI: 10.3390/rs4040987. URL: <https://www.mdpi.com/2072-4292/4/4/987>.
- [27] Carsten Laukamp et al. “Mineral Physicochemistry Underlying Feature-Based Extraction of Mineral Abundance and Composition from Shortwave, Mid and Thermal Infrared Reflectance Spectra”. In: Minerals 11.4 (2021), p. 347.
- [28] HH Murray. “Correlation of Paper-Coating Quality with Degree of Crystal Perfection of Kaolinite”. In: Clays Clay Miner 4 (1955), pp. 31–40.
- [29] JK Crowley. “Near-Infrared Reflectance Spectra of Mixtures of Kaolin-Group Minerals: Use in Clay Mineral Studies”. In: Clays Clay Miner 36 (1988), pp. 310–316.
- [30] J Madejová. “FTIR Techniques in Clay Mineral Studies”. In: Vibrational Spectroscopy 31.1 (2003), pp. 1–10.
- [31] JL Post and PN Noble. “The Near-Infrared Combination Band Frequencies of Dioctahedral Smectites, Micas, and Illites”. In: Clays Clay Miner 41 (1993), pp. 639–644.

- [32] Jörg Bowitz and Angela Ehling. “Non-destructive infrared analyses: A method for provenance analyses of sandstones”. In: *Environmental Geology* 56 (Dec. 2008), pp. 623–630. DOI: 10.1007/s00254-008-1361-6.
- [33] Piotr Przybylek. “Determination of Mineral Oil Concentration in the Mixture with Synthetic Ester Using Near-Infrared Spectroscopy”. In: *Energies* 16.17 (2023), p. 6381. DOI: 10.3390/en16176381.
- [34] nanoComposix. *Silica Surface*. Accessed: October 27, 2024. 2024. URL: <https://nanocomposix.com/pages/silica-surface>.
- [35] Anamaria Roman and Tudor-Mihai Ursu. “Multispectral Satellite Imagery and Airborne Laser Scanning Techniques for the Detection of Archaeological Vegetation Marks”. In: *Landscape Archaeology on the Northern Frontier of the Roman Empire at Porolissum*. Ed. by Coriolan Horațiu Opreanu and Vlad-Andrei Lăzărescu. Cluj-Napoca: Mega Publishing House, 2016, pp. 141–173. ISBN: 978-606-543-787-6. URL: <https://www.researchgate.net/publication/315797574>.
- [36] John W. Salisbury et al. “Infrared (2.1–25 μm) Spectra of Minerals”. In: *Johns Hopkins University Press* (1991). Study on temperature effects on mineral spectra.
- [37] Duarte Viveiros et al. “Temperature Stability and Spectral Tuning of Long Period Fiber Gratings Fabricated by Femtosecond Laser Direct Writing”. In: *Sensors* 20.14 (2020), p. 3898. DOI: 10.3390/s20143898.
- [38] Roger N. Clark and Ted L. Roush. “Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications”. In: *Journal of Geophysical Research: Solid Earth* 89.B7 (1984), pp. 6329–6340. DOI: 10.1029/JB089iB07p06329.
- [39] M. Miyamoto et al. “Evaluation of a curve-fitting method for diffuse reflectance spectra in the UV-Visible-NIR wavelength region”. In: *Polar Science* 3 (2009), pp. 110–116. DOI: 10.1016/j.polar.2009.06.004.
- [40] Byong-kyu Shin. “An Analytical Method for the Deconvolution of Voigtian Profiles”. In: *Applied Magnetic Resonance* 47.8 (2016), pp. 783–796. DOI: 10.1007/s00723-016-0769-5.

- [41] Gregory V. Morrow. “Deconvolution of Spectral Voigt Profiles Using Inverse Methods and Regularization”. Available at: <https://sites.math.washington.edu>. PhD thesis. University of Washington, 2010.
- [42] Mahzia. “Competition between Lorentzian and Gaussian Width in Pure Oxygen Absorption Spectrum at 1270 nm Band”. In: Journal of Quantitative Spectroscopy and Radiative Transfer 192 (2017), pp. 103–112. DOI: 10.1007/s12596-017-0409-y.
- [43] J. Humlíček. “Optimized computation of the Voigt and complex probability functions”. In: Journal of Quantitative Spectroscopy and Radiative Transfer 27.4 (1982), pp. 437–444. DOI: 10.1016/0022-4073(82)90078-4.
- [44] J. T. Reilly et al. “Analysis of FT-IR spectroscopic data: The Voigt profile”. In: Spectrochimica Acta Part A: Molecular Spectroscopy 48.10 (1992), pp. 1459–1479. DOI: 10.1016/0584-8539(92)80150-6.
- [45] Xianglei Huang and Yuk L. Yung. “A common misunderstanding about the Voigt line profile”. In: Journal of the Atmospheric Sciences 60.13 (2003), pp. 1630–1632. DOI: 10.1175/1520-0469(2003)060<1630:ACMATV>2.0.CO;2.
- [46] V.L. Mulder, M. Plötze, S. de Bruin, et al. “Quantifying Mineral Abundances of Complex Mixtures by Coupling Spectral Deconvolution of SWIR Spectra (2.1–2.4 μm) and Regression Tree Analysis”. In: Geoderma 207–208 (2013), pp. 279–290. DOI: 10.1016/j.geoderma.2013.05.011.
- [47] Heather D. Makarewicz, Mario Parente, and Janice L. Bishop. “Deconvolution of VNIR Spectra Using Modified Gaussian Modeling (MGM) with Automatic Parameter Initialization (API) Applied to CRISM”. In: IEEE International Conference on Image Processing. 2009.
- [48] Ronan Rialland et al. “Improved Deconvolution of Mineral Reflectance Spectra”. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2021). DOI: 10.1109/JSTARS.2021.3110008.
- [49] Thi Thanh Nguyen et al. “Non-negative orthogonal greedy algorithms”. In: IEEE Transactions on Signal Processing 67.21 (2019), pp. 5643–5658. DOI: 10.1109/TSP.2019.2943225.
- [50] J. Botzheim. Lectures on Computational Intelligence. Figure extracted from lecture materials. 2024.

- [51] Swagatam Das, Archana Chowdhury, and Ajith Abraham. “A Bacterial Evolutionary Algorithm for Automatic Data Clustering”. In: Proceedings of the World Congress on Nature & Biologically Inspired Computing (NaBIC). Coimbatore, India: IEEE, 2009, pp. 1036–1042.
- [52] Algorithm Afternoon. Elitist Genetic Algorithm. Accessed: 2024-04-27. n.d. URL: [https : / / algorithmafternoon . com / genetic / elitist _ genetic _ algorithm/](https://algorithmafternoon.com/genetic/elitist_genetic_algorithm/).
- [53] M. F. Hatwágner and A. Horváth. “Maintaining Genetic Diversity in Bacterial Evolutionary Algorithm”. In: Acta Universitatis Sapientiae, Informatica 3.2 (2011). Supported by the National Development Agency and the European Union within the project TAMOP 4.2.2-08/1-2008-0021 at Széchenyi István University, titled "Simulation and Optimization — basic research in numerical mathematics.", pp. 175–187. DOI: 10.2478/v10166-011-0011-0.
- [54] Lotfi A Zadeh. “Fuzzy sets”. In: Information and Control 8.3 (1965), pp. 338–353.
- [55] YH Song et al. “Improved genetic algorithms with fuzzy logic controlled crossover and mutation”. In: UKACC International Conference on CONTROL’96 (1996), pp. 140–144.
- [56] R Rialland et al. “Identification of minerals from hyper-spectral imaging based on a fuzzy logic approach”. In: Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imaging XXVII Vol. 11727. SPIE. 2021, pp. 108–119.
- [57] OpenGeology. Minerals. Accessed: 2024-10-28. 2023. URL: [https : / / opengeology.org/textbook/3-minerals/](https://opengeology.org/textbook/3-minerals/).
- [58] Mohammad S. Abweny et al. “Short-Wavelength Infrared (SWIR) spectroscopy of low-grade metamorphic volcanic rocks of the Pilbara Craton”. In: Journal of African Earth Sciences 117 (2016), pp. 124–134.
- [59] Roger N. Clark et al. “High Spectral Resolution Reflectance Spectroscopy of Minerals”. In: Journal of Geophysical Research 95.B8 (Aug. 1990), pp. 12653–12680. DOI: 10.1029/89JB03415.
- [60] Roger N Clark. “Spectroscopy of rocks and minerals, and principles of spectroscopy”. In: Manual of remote sensing. Vol. 3. John Wiley and Sons New York, 1999, pp. 3–58.

- [61] Jonathan Cloutier, Stephen J. Piercey, and Jonathan Huntington. “Mineralogy, Mineral Chemistry and SWIR Spectral Reflectance of Chlorite and White Mica”. In: Minerals 11.5 (2021), p. 471. DOI: 10.3390/min11050471.
- [62] Andrea García-Vicente et al. “Field Spectroscopy Applied to the Kaolinite Polytypes Identification”. In: Proceedings of the 2nd International Electronic Conference on Mineral Science. MDPI. Switzerland: MDPI, 2021, pp. 1–7. DOI: 10.3390/iecms2021-09353. URL: <https://doi.org/10.3390/iecms2021-09353>.
- [63] Mathew Kamau, Chris Hecker, and Caroline Lievens. “Use of Short-Wave Infrared Reflectance (SWIR) Spectroscopy to Characterize Hydrothermal Alteration Minerals in Olkaria Geothermal System, Kenya”. In: Proceedings of the 45th Workshop on Geothermal Reservoir Engineering. Stanford University. Stanford, California, 2020, pp. 1–15. DOI: 10.1146/annurev.ea.06.050178.001305. URL: <https://pangea.stanford.edu/ERE/pdf/IGAstandard/SGW/2020/Kamau.pdf>.
- [64] Yongsik Jeong et al. “Spectral characteristics of minerals associated with skarn deposits: a case study of Weondong skarn deposit, South Korea”. In: Geosciences Journal 19.3 (2015), pp. 375–388. DOI: 10.1007/s12303-015-0043-0.
- [65] Robert Milewski et al. “Assessment of the 1.75 μ m absorption feature for gypsum estimation using laboratory, air- and spaceborne hyperspectral sensors”. In: International Journal of Applied Earth Observation and Geoinformation 77 (2019), pp. 69–83.
- [66] Mahesh Kumar Tripathi and Himanshu Govil. “Evaluation of AVIRIS-NG hyperspectral images for mineral identification and mapping”. In: Heliyon 5.e02931 (2019), pp. 1–12. DOI: 10.1016/j.heliyon.2019.e02931. URL: <https://doi.org/10.1016/j.heliyon.2019.e02931>.
- [67] Nancy K McKeown et al. “Characterization of phyllosilicates observed in the central Mawrth Vallis region, Mars, their potential formational processes, and implications for past climate”. In: Journal of Geophysical Research: Planets 114 (2009), E00D10.

- [68] Mark P. Simpson and Andrew J. Rae. “Short-wave infrared (SWIR) reflectance spectrometric characterisation of clays from geothermal systems of the Taupō Volcanic Zone, New Zealand”. In: Geothermics 73 (2018), pp. 74–90. DOI: 10.1016/j.geothermics.2018.01.006.
- [69] YongHwi Kim et al. “Application of SWIR Spectrometry to the Determination of Biotite Compositions in Hydrothermally Altered Units of the Yangyang Iron-Oxide-Apatite (IOA) Deposit, South Korea”. In: Ore Geology Reviews Accepted Manuscript (2018). DOI: 10.1016/j.oregeorev.2018.06.022.
- [70] Roger N. Clark et al. “Imaging Spectroscopy: Earth and Planetary Remote Sensing with the USGS Tetracorder and Expert Systems”. In: Journal of Geophysical Research: Planets 108.E12 (2003), pp. 5.1–5.44. DOI: 10.1029/2002JE001847. URL: <https://doi.org/10.1029/2002JE001847>.

List of Figures

2.1	Interaction of radiation with matter, illustrating reflection, transmission, absorption, and emission processes [18].	10
2.2	Reflectance spectrum of Actinolite, showing absorption features related to Fe^{2+} , Mn-OH, and Mg-OH in the VNIR-SWIR range. Adapted from [20].	11
2.3	Absorbance spectra at various sampling intervals, demonstrating the effect of spectral resolution on peak identification [21].	12
2.4	Illustration of the Signal-to-Noise Ratio (SNR), showing how low SNR can obscure spectral features [23].	12
2.5	Sensor calibration in hyperspectral imaging, displaying intensity levels for different conditions: <i>L</i> - light reflected from the target, <i>B</i> - baseline in darkness, <i>W</i> - standard material with a reference light source [25]. . . .	13
2.6	Reflectance spectra demonstrating the effect of grain size on spectral characteristics for two carbonates: Calcite (left) and Dolomite (right) [26]. . .	13
2.7	Comparison between a well-ordered kaolinite sample with high crystallinity and a sample of disordered kaolinite and halloysite, both with low crystallinity [32].	14
2.8	Absorption spectra of heterogeneous mixtures showing variations in band depth and shape [33].	15
2.9	Impact of surface coatings on reflectance spectra, with changes in absorption intensity due to dust or organic matter [34].	15
2.10	Multispectral analysis showing different regions of reflectance for vegetation, including chlorophyll absorption and water content bands [35]. . .	16
2.11	Temperature effects on absorption spectra, illustrating band broadening and shifts at elevated temperatures [37].	16
2.12	Left: Reflectance curve of calcite with its continuum (linear interpolation). Right: Absorption curve obtained by subtracting the continuum, with Voigt profile deconvolution shown.	17

2.13	Illustration of the i -th generation in BMA, showing genes mutation in a bacterium by selecting genes, mutating them, and propagating the best clone across the remaining copies of the bacterium [50].	22
2.14	Scheme for the gene transfer operation [51].	23
2.15	Simplified representation of the elitist system. Given a target curve, the best bacterium from each mineral is selected and shown in yellow. The mineral with the best bacterium is chosen as the elite, represented in the figure with a crown. Afterwards, gene swapping is allowed between the elite and the rest of the population.	24
3.1	Illustration of the EBMA-SD algorithm flow.	34
3.2	Representation of the chromosome as a collection of genes containing Voigt Profile parameters and an additional label parameter k	35
3.3	The top graph shows the original reflectance curve of a Smaragdite sample and its linear continuum. The bottom graph depicts its Voigt profile decomposition: Category 1 peaks are the most relevant, Category 2 peaks are present in all samples, while Categories 3 to 5 show greater variability.	36
4.1	Effect of the weighting factor α in the objective function on fitting the Voigt profiles during the local search. Higher values of α prevent the approximation from exceeding the target curve.	46
4.2	Initial LM optimization by adjusting gene heights to fit the target Actinolite HS315.1B. The comparison includes Actinolite HS22.1B, Smaragdite HS290.1B, and Nephrite HS296.4B, demonstrating the effectiveness of restricted fitting in distinguishing closely related minerals.	47
4.3	Visualization of the Action Potential of the Voigt profiles and the corresponding central wavelength boundaries during the local search.	47
4.4	Visualization of the results after the local search optimization.	48
4.5	The elite bacterium searching for genes in other bacteria in regions highlighted in purple, where the fitness is insufficient and greater than the threshold.	49
4.6	Impact of gene category on permissible parameter boundaries during the local search, as defined by the Action Potential.	50
4.7	The target absorption curve (top), followed by the original elite bacterium and its updated state after applying the EBMA-SD algorithm.	50

5.1	Classification confusion matrix by mineral type.	53
A.1	Screenshot of the interface used to create the database.	57
B.1	Visualization of carbonate minerals: the first column shows the reflectance curve, the middle displays the absorption curve, and on the right is the representation of the genes in bacterial format, with genes color-coded according to their category.	59
C.1	Detailed confusion matrix resulting from the mineral classification using only one sample of each mineral in the database.	61

List of Tables

3.1	Action Potential (Consequence) Based on Gene Fitness Level and Gene Importance (Antecedents). The table maps <i>Gene Importance</i> (GI) against <i>Gene Fitness Level</i> (GFL) to determine the corresponding <i>Action Potential</i> .	39
3.2	Gene Transfer Strategies Based on Action Potential (AP) Levels	40
5.1	Some minerals of interest in the USGS spectral library [8].	52

List of Algorithms

1	Elitist Bacterial Memetic Algorithm for Spectral Decomposition with Fuzzy Logic (EBMA-SD)	42
---	---	----