HAL open science

Segmented Curve-Fitting Method for Continuum Removal in Hyperspectral Images

Priyanka Kumari, Sampriti Soor, Amba Shetty, Shashidhar G Koolagudi

To cite this version:

Priyanka Kumari, Sampriti Soor, Amba Shetty, Shashidhar G Koolagudi. Segmented Curve-Fitting Method for Continuum Removal in Hyperspectral Images. 2023. ⟨hal-04104263⟩

HAL Id: hal-04104263

https://hal.science/hal-04104263v1

Preprint submitted on 23 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Autorisation HAL

HAL Authorization

Segmented Curve-Fitting Method for Continuum Removal in Hyperspectral Images

Priyanka Kumari, *Graduate Student Member, IEEE*, Sampriti Soor, *Member, IEEE*, Amba Shetty, and Shashidhar G Koolagudi

***Abstract*—A spectrum in a multiband remotely sensed image is generally a mixture of spectra of different materials present in the scene which can be distinguished by distinct absorption signatures. A mixed spectrum possesses a smooth baseline shape, known as a continuum, that masks the individual spectral features. Continuum can also appear due to instrument artifacts and topographic illumination effects. Eliminating the continuum from a spectrum being analyzed and correctly identifying its unique absorption characteristics is crucial for material identification. In this paper, a new continuum removal technique called *Segmented Curve-Fitting* is proposed, which estimates the continuum extending the approach of using an upper convex hull (UCH), such that, distinct absorption signatures are determined more accurately, resulting in more accurate material identification. Here parabolic curves are fitted on the translated spectrum for each derived segment by UCH to identify more shoulder points inside each segment. The performance of the suggested method is compared with the UCH method using a collection of simulated data of varying complexity as well as a real CRISM TRDR hyperspectral dataset. The identification score is improved by around 8% for the similarity matching method Weighted Sum of Spectrum Correlation and by around 2% for a Convolutional Neural Network. The proposed method has been effectively applied to mineral identification on Earth and Mars hyperspectral data, AVIRIS and CRISM MTRDR, respectively.**

***Index Terms*—Continuum Removal, Segmented Curve-fitting, Hyperspectral Images, Mineral Mapping**

I. Introduction

HYPERSPECTRAL imaging system can detect subtle changes in the reflectance spectrum that are indicative of the presence of specific minerals or other objects, such as clay minerals, which are difficult to detect with other remote sensing techniques [1]. Every object in a scene has its own spectral signature, such as absorption or emission lines, which is a unique pattern comprised of the reflectance values on light intensity across different wavelengths. When light from all the objects in the scene is collected and captured by the hyperspectral imaging system, the individual spectral signatures are combined, and the overall spectral shape of the objects in the scene is determined. Hence the resultant spectrum of the scene represents the underlying spectral information that is not due to individual spectral features, but instead due to the overall spectral shape of the objects in the scene [2], [3]. This spectral mixing in hyperspectral images makes it difficult to separate the spectral information of individual objects. The overall spectral shape resulting from the superposition of the individual spectral signatures forms a smooth background intensity of the scene, which is known as the continuum in hyperspectral imagery [4]–[6]. Moreover, Instrument artifacts and illumination effects from the surface can also cause a continuum to appear [7], [8]. The continuum information is useful in hyperspectral image analysis as it provides a reference for identifying and analyzing individual spectral features in the image.

One of the major challenges for material detection in a hyperspectral image is to predict the continuum of a spectrum accurately. Continuum can be influenced by atmospheric effects, such as atmospheric absorption and scattering, which can affect its global shape [9]. Some materials have complex spectral shapes that can make it challenging to distinguish them from the continuum, which can result in false or missed detections [10], [11]. Different hyperspectral images of a surface area may have different continuums, or even different parts of the same hyperspectral image could have distinct continuums. Because of this, continuums in a hyperspectral image are typically anticipated pixel-by-pixel and are generally considered to be independent of continuums in nearby pixels [12].

Simple distance-based similarity-matching is not suitable for identifying materials in hyperspectral images, as the presence of continuum in the hyperspectral data alters the global shape of an end-member [13] although the absorptions are preserved [figure I]. It requires specialized techniques to effectively separate the continuum from individual objects in the scene, which is often present as an essential pre-processing step in hyperspectral data analysis. The imposed continuum removal (ICR) methods involve fitting a polynomial or a power-law model to the entire spectrum to estimate the spectral baseline assuming a functional form for the continuum [14]. The apparent continuum removal (ACR) methods, on the other hand, select a few local regions of the spectrum that are free of any spectral features assuming those represent the underlying continuum directly, and connect those local regions by some specified methods to estimate the continuum for the full spectrum domain [15]. The spectrum is subtracted from or divided by the estimated continuum to nullify the effects of the continuum [eq. I.1] and thus the unique absorption signatures are enhanced [16]. ICR can be more accurate when the continuum shape is well-known or can be assumed with

Manuscript received on xxx xx, xxxx.

P.K. and A.S. are associated with the Department of Water Resources and Ocean Engineering and S.G.K. is associated with the Department of Computer Science and Engineering respectively in the National Institute of Technology Karnataka, India. S.S. is associated with the School of Computer Engineering at Kalinga Institute of Industrial Technology, India. E-mail ids: priyanka.187am004@nitk.edu.in, sampreetiworkid@gmail.com, amba@nitk.edu.in, koolagudi@nitk.edu.in

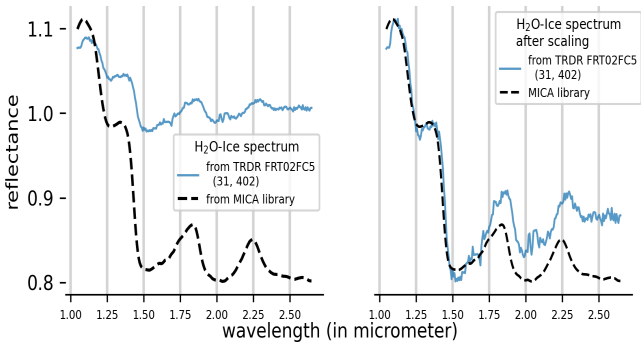


Fig. 1. The difference in global curvatures between the two H_2O-Ice spectra in the left image is evident after the scaling operation in the right image.

high confidence, while ACR can be more flexible and robust to variations in the continuum shape.

$$\begin{aligned} \mathcal{C}_C^{div}(R) &= R/C \\ \mathcal{C}_C^{sub}(R) &= R - C \end{aligned} \tag{I.1}$$

where R is the spectrum being analyzed, C is the estimated continuum and $\mathcal{C}(R)$ is the continuum-removed spectrum ($\mathcal{C}$ being the continuum removing function superscripted with the removal method, and subscripted with the continuum). In this paper, if the removal method is not specified at the superscript level for a continuum removing function $\mathcal{C}$, it is assumed that both subtraction and division can be applied. Similarly, if the estimated continuum is not specified at the subscript level, it is assumed that the continuum is estimated by an upper convex hull. Often a continuum-removed spectrum is normalized for further processing.

There are several methods for continuum removal each with its own advantages and limitations, and the choice of method depends on the specific requirements of the analysis and the characteristics of the data being analyzed. Frequently, an upper hull of the spectrum is used to estimate the continuum. Clark et al. introduced the upper convex hull (UCH) method [5] and the segmented upper hull (SUH) method [17] for this purpose, which is still well used in literature. On the other hand, instead of estimating the continuum and then highlighting the spectral features of the spectrum by removing it, techniques like first and second derivatives [18] operate directly to enhance the spectral signatures. However, other than removing the spectral baseline continuum removal has another apparent purpose, which is to isolate the segments of the spectrum, that can be used to extract the distinct absorption information. By using upper hull methods, the segments of spectra between the reflectance values touching the hull are typically treated as distinct absorption features, however, this is not very effective for complex spectral shapes. In the methods described in [19], [20], the spectrally uninformative portions are considered as part of the continuum, and the absorption shoulders, termed as pin-points, are determined at the extremities of each of such continuum portions. These are then stretched through the spectrum domain by linear interpolations to estimate the full continuum. The spectral features of interest between these pin-points are extracted after continuum removal for material identification. This technique of absorption information retrieval is applied in [21], [22] with prior knowledge of the position of the absorption shoulders. An estimated continuum can over-fit the spectrum if a large number of such segments are taken into account. Piecewise cubic spline methods work in a similar way with the only difference being the portions of a continuum between the pin-points are fitted by cubic curves to provide a smooth fit to the spectrum [4].

In this paper, the idea of piece-wise estimation of the continuum is combined with the imposed approach of UCH. Initially, the shoulder points in the analyzed spectrum are determined by the UCH, and the exclusive segments of the spectrum between two consecutive touch-points of the spectrum and the UCH are processed independently to extract more shoulder points. These second-level shoulder points are identified by fitting parabolic curves in each spline. The resultant continuum is not exactly the linear interpolation through these detected points, rather has a parabolic shape between each pair of consecutive detected points, which is a more realistic estimation of the continuum. Though different researchers vouch for using simple continuum removal techniques like UCH, the proposed method Segmented Curve-Fitting (SCF) separates the distinct absorption features present in a spectrum more accurately, and thus the segmented correlations between library spectra and a test spectrum are determined more correctly. In this paper, we suggested a linear time algorithm to determine the UCH, and using this the proposed method SCF also can be determined in linear time. The proposed method is applied to the objective of mineral identification in hyperspectral data collected from the earth's surface as well as from the Martian surface. To assess the performance of SCF, the accuracy of this aim is compared to the widely used UCH process, with both used as a preprocessing step in a mineral classification framework.

The section organization in this paper are as follows:

In section II the proposed continuum removal technique, Segmented Curve-Fitting is explained in detail with the help of easy-to-understand diagrams. Section III contains the comparison results of the proposed approach with the upper convex hull approach for continuum removal on multiple datasets, with the objective of mineral identification particularly, whereas, section IV contains some experimental results of the proposed method on Nevada Cuprite dataset.

II. Segmented Curve-fit for Continuum Removal

Since the form of the continuum in a hyper-spectral image is considered to have a convex curvature most often, the convex upper hull approach is typically used in literature. However, in reality, the structure of a continuum may not be confined to that and can unevenly affect distinct localized wavelength ranges in a spectrum. The suggested approach specifically addresses this issue by dividing the spectrum into multiple segments and predicting the shape of the continuum for each segment independently. Although this approach disregards the global pattern of the spectrum to some extent, the benefit of this approach is to enhance the weaker band-depths more precisely. The step-by-step process of the proposed method is now outlined in detail, with each step illustrated in figure 2 using an example spectrum of H_2O-Ice from MICA spectral

library [12]. Note that, a continuum removal method can also be applied on a pure spectrum too, to extract its distinct absorption signatures.

❍ ***Step-1 (Segmenting the spectrum):***

The first step of the proposed method is to extract the initial shoulder points from the spectra and by that determine the exclusive partitions in the translated spectrum, on which parabolic curves are fitted to extract more shoulder points and continuums are estimated for each segment. The Upper Convex Hull (UCH) method is used here for the initial segmenting of the spectra. An algorithmic outline for UCH function, UPPERCONVEXHULL, is given below which works in linear time. The same function is used in step-3 as well, to extract shoulder points from each of the initial segments.

Let $\mathscr{W}$ be the wavelength domain for which $\{R_w : \forall w \in \mathscr{W}\}$ be the reflectance values of the spectrum R. The following method describes the process to generate the UCH.

function UPPERCONVEXHULL(P)
  ▷ $P = [(w, R_w) : \forall w \in \mathscr{W}]$, $\mathscr{W}$ is ascendingly sorted
  initialize Empty stack $\mathcal{S}$
  $\mathcal{S}.push(P_1)$, $\mathcal{S}.push(P_2)$
    ▷ P_i is the i-th element from first in P
  for $P_i : i \in 3, 4, ..., P.length$ **do**:
    while True **do**
      if $\angle \mathcal{S}_{-2}\mathcal{S}_{-1}P_i \geq 180°$ **then**:
        ▷ $\mathcal{S}_{-i}$ is the i-th element from last in $\mathcal{S}$
        ▷ $\angle$ is upward angle
        $\mathcal{S}.push(P_i)$
        break
      else
        $\mathcal{S}.pop()$
        if $\mathcal{S}.length = 1$ **then**:
          $\mathcal{S}.push(P_i)$
          break
  return $\mathcal{S}$

Fig. 2(a) shows the example spectrum and its UCH. Fig. 2(b) depicts the continuum-removed spectra by the usual process, that is, dividing the spectrum by its UCH to obtain a translated spectrum, say $T^1 = \mathcal{C}(R)$.

The consecutive wavelengths having translated values lesser than 1 between two 1-valued wavelengths comprise a segment, where the 1-valued wavelengths at the two ends serve as the shoulder-points of the segment. Fig. 2(c) shows such segments within the translated spectra T^1 for the example spectrum.

❍ ***Step-2 (Segment-wise fitting of parabolic curves)***

Specifically, only the segments having at least one local peak are picked to process in this step, and for each picked segment a parabolic curve is fitted on the local peaks/maxima within the segment, considering the curve passes through the two shoulder points at the end of the segment. To obtain a functional form of the fitted curve, rather than the original spectrum, the translated spectrum is used, on which such parabolas facing the positive y-axis can exist as functions.

As the values at the end-wavelengths of a translated spectrum segment are 1, considering these as w_s and w_t, and w be any wavelength within the segment, the general equation

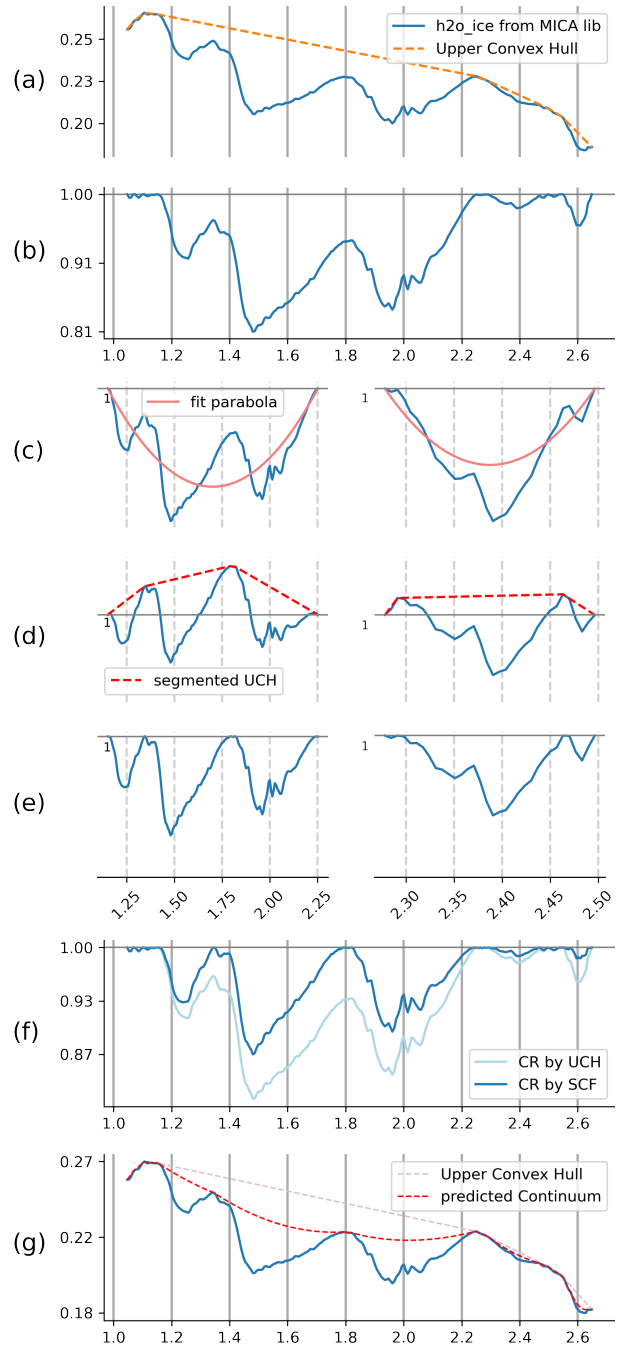


Fig. 2. (a) H_2O-Ice spectrum R from MICA library and the UCH of it; (b) Translated spectrum T^1 ; (c) segments are extracted from T^1 and parabolic curves as in eq.-II.2 are fitted; (d) Translated spectrums T^2_S and corresponding UCHs; (e) Translated spectrums T^3_S ; (f) Merging the segments T^3_S and the unchanged spectra portions, the translated spectra T^1 is also provided for visual comparison; (g) A visual comparison between the predicted continuum and UCH.
Note-1: X-axis represents the wavelength domain. The domain is same for the full spectra in (a), (b), (f) and (g), whereas the domain is same for the spectra segments in (c), (d) and (e). Y-axis in (a) and (g) represents the reflectance values in the input spectra, and the continuum-removed normalized spectra in the others.
Note-2: Only the local peaks for each segment are used to fit the parabolic curves in (c), as these points could be new shoulder points or being part of the estimated continuum could be close to the new shoulder points in (d-e).
Note-3: The continuum estimated by the proposed method in (g) has a parabolic shape between the new shoulder-points extracted in (d).
Note-4: The continuum shown in (g) is not needed to be computed to get the continuum-removed spectrum shown in (f).

of the parabolic curve for a segment S of T^1 becomes

$$C(T^1_S)_w = a(w - w_s)(w - w_t) + 1 \tag{II.1}$$

Let $\mathscr{W}_{S_{max}}$ be the set of wavelengths in segment S having all local maxima, and T^1_w is the translated value at wavelength w in T^1. Then for the fitted parabola

$$a = \frac{\sum\limits_{w_i \in \mathscr{W}_{S_{max}}} T^1_{w_i} p \;-\; \sum\limits_{w_i \in \mathscr{W}_{S_{max}}} p}{\sum\limits_{w_i \in \mathscr{W}_{S_{max}}} p^2}$$

$$\text{where } p = (w_i - w_s)(w_i - w_t) \tag{II.2}$$

Fig. 2(c) shows the fitted curves for the different segments of the translated spectra.

❍ ***Step-3 (Segment-wise Translation)***

The next step is to calculate another segment-wise translation T^2_S from T^1_S (a segmented range S of the translated spectrum T^1) and $C(T^1_S)$ (the parabolic curve fitted to T^1_S), such that, $T^2_S = \mathcal{C}_{C(T^1_S)}(T^1_S)$. Fig. 2(d) depicts T^2_S for distinct segments of the example spectrum.

To keep all the shoulder-points on the continuum-removed spectrum at the same level, another step of continuum removal for each segment by UCH is performed, $T^3_S = \mathcal{C}(T^2_S)$. Fig. 2(d) shows segment-wise UCH and fig. 2(e) shows T^3_S of the corresponding segments.

❍ ***Step-4 (Merging the translated segments)***

The final step of the procedure involves combining the independently translated segments of T^3 with the unpicked portions in step-2 from the initially translated spectrum T^1 on the entire wavelength domain $\mathscr{W}$ to generate the continuum-removed spectrum $\mathcal{C}(\mathscr{W})$. Let $\mathcal{S}$ is the set of segments picked for further translation in step-2, then

$$\mathfrak{C}(R_w) = \begin{cases} {T^3_S}_w, & \text{if } w \in S, S \in \mathcal{S} \\ {T^1}_w & \text{otherwise} \end{cases} \forall w \in \mathscr{W} \tag{II.3}$$

Fig. 2(f) shows the continuum-removed spectra $\mathfrak{C}(R)$ by the proposed method, and in fig. 2(g) a visualization of the predicted continuum is given. The predicted continuum could separate almost all the prominent absorptions which the previous method was unable to.

The usual methods of continuum removal like UCH determine the continuum from the full spectrum first, then eliminate its effect from the spectrum; on the other hand, the proposed method specifically aims to directly remove the effect of the continuum from the spectra without estimating the whole continuum, i.e., the continuum shown in fig. 2(g) is not needed to be computed to get the continuum-removed spectrum shown in fig. 2(f). The continuum-removed spectrum $\mathfrak{C}(R)$ can be calculated in linear computational time, because, the initial segmentation in step-1 takes linear time, and the segments consist of distinct wavelength ranges which are processed further (eq. (II.2-II.3)) in linear computational time independently and each in linear time. Note that, though it seems reasonable to consider only the local peaks to reduce the runtime further but I might not produce the normalized spectra at the end.

Explanation of equation II.2:
$y = a(x - x_t)(x - x_s) + c$
Squared error $J = \sum(y - a(x - x_t)(x - x_s) - c)^2$
With optimum fitting $\frac{\delta J}{\delta a} = 0$
$\implies \sum(y - a(x - x_t)(x - x_s) - c)(x - x_t)(x - x_s) = 0$
let $p = (x - x_t)(x - x_s)$
$\implies \sum yp - \sum ap^2 - \sum cp = 0$
$\implies a = \frac{\sum yp - c\sum p}{\sum p^2}$

III. Performance Evaluation

To evaluate the performance of the proposed continuum removal method SCF, four datasets with different levels of complexity are used. The first three datasets are synthesized from the MICA spectral library, which was created by collecting spectral data from a variety of mineral samples using CRISM spectrometer from different locations of martian surface [12], and the fourth dataset is the recently released labelled TRDR spectra by Plebani et al [23]. For each dataset three evaluation measures are calculated, which are (i) Mean shift of band-minima, (ii) Mean difference in Full Width at Half Maximum (FWHM), and (iii) Identification accuracy. Two methods are used to calculate the accuracy of identification. The first method utilizes a curve-matching procedure to calculate a matching index of a test spectrum to a library spectrum, while the second method employs a shallow convolutional neural network model. The results of SCF are compared with those obtained by UCH to provide a comparative evaluation. This section first provides a detailed description of how the synthetic data were generated and a brief summary of any necessary preprocessing steps. Following that, the evaluation procedures are explained, and finally, the performance of the model is analyzed using the comprehensive results. All the related codes of this work are available at [24].

A. Experimental Setup

1) **Data Specifications:** MICA spectral library contains 31 different mineral spectra from 6 mineral groups like Iron oxides and primary silicates, ices, sulfates, phyllosilicates, carbonates, and other hydrated silicates and halides. To recreate three different degrees of complex datasets that might be in accordance with real-life data, varying magnitudes of shape change and noise are applied to the spectra of this library, also different spectra from it are mixed together to get the first three datasets.

a) **Synthetic Data (dataset- 1, 2, 3):** Primarily a random noise N is generated by the composition of the following.

i. A curvature noise C, which is predicted by adding multiple Gaussian curves of $\Phi(w, 0.65)$, where the wavelength w is the position of the peak in a curve and is picked randomly within the selected range of the wavelengths. If p number of such curves are combined, the resultant curvature (Φ_p) can have a maximum p number of peaks.
ii. A fluctuation noise N, which is predicted by randomly chosen values from a Gaussian distribution $\Phi(0, \sigma^2_N)$ for each wavelength of the spectra domain. If σ_N is higher, the noise fluctuation is higher.
iii. A mixed-spectrum S formed by adding scaled spectra of a number of randomly chosen different minerals other than the mineral to which noise is put. The primary mineral is combined with all of the other minerals in a 60-40 ratio. S_k denotes the mixed-spectrum with $k(\geq 0)$ secondary

spectra, where S_0 indicates no other spectrum is mixed with the primary spectrum.

The general equation of spectrum R with noise is given by,

$$R^N(p, k, \sigma_N) =_{scaled} R + {}_{scaled}(C_p + S_k + N_{\sigma_N}) \quad \text{for } p, \sigma_N, k \geq 0 \quad \text{(III.1)}$$

The prefix $_{scaled}$ denotes a min-max scaling operation on the data. Note that, though the reflectance values in the generated spectrum are not entirely coherent with the original spectrum, that will not affect the evaluation process due to the continuum removal process applied to it. Following are the specification of the first three datasets.

dataset-1	**dataset-2**	**dataset-3**
$p = 1, 2, 3, 4$	$p = 0$	$p = 1, 2$
$k = 0$	$k = 1, 2, 3, 4$	$k = 1, 2$
$\sigma_N = 0$	$\sigma_N = 0$	$\sigma_N = .01, .02, .03, .04$

Dataset-1 can be used to test the performance of the continuum removal process on spectra affected by non-convex continuum that could be generated by the instrumental artifacts and environmental noises in real spectra. Dataset-2, on the other hand, can test the same if there is mixed-spectra. Dataset-3 includes both of these types of noise in a spectrum and can additionally test for fluctuation noise in the spectrum. For each mineral class, 100 samples were generated for each combination of specifications in each dataset. The generation process involved randomness (eq. III.1), resulting in each sample being unique. Figure 3 illustrates the process for creating these three datasets. The final dataset, dataset-4, consists of real spectra captured from the Martian surface by the CRISM sensor.

b) **CRISM TRDR Data (dataset-4):** CRISM (Compact Reconnaissance Imaging Spectrometer for Mars) is a spectral imager that is part of the Mars Reconnaissance Orbiter (MRO) mission, which was launched in 2005. The instrument collects spectral images of Mars' surface, which can be used to identify minerals and other compositional information about the surface. To facilitate data analysis, a high spatial resolution data product called Targeted Remote Sensing Data Record (TRDR) is generated from the CRISM instrument, which is easily downloadable from NASA PDS. Plebani et al in [23] using a hierarchical Bayesian model classified a collection of 592413 TRDR spectra, that is from more than 70 different images, with 39 different labels including some spectra for artifacts and some bland pixels, of which 28 labels are common with MICA library. The evaluation process here is restricted to these minerals only, randomly sampling 200 spectra from each label. In this experiment, only the bands within the 1-2.6 μm wavelength range were considered, from the given bands for the 1-3.47 μm range.

In [25] a preprocessing pipeline is used for mineral identification in CRISM MTRDR (Map-Projected Targeted Reduced Data Record) data. The pipeline includes spectra smoothing, spike removal, continuum removal, and spectra standardization. Similar steps are applied to the datasets used here. To eliminate fluctuation noise, dataset-4 and dataset-3 essentially undergo spectra smoothing and spike removal steps, while all four datasets are standardized and undergo continuum removal. Proper smoothing is crucial for the SCF method,

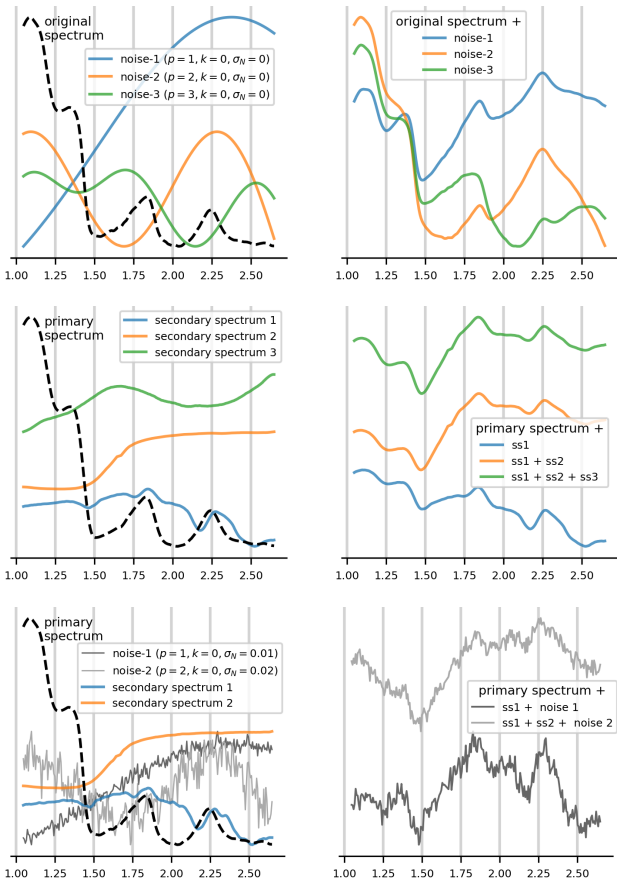


Fig. 3. Example of synthesizing daataset-1 (top row) dataset-2 (middle row) and dataset-3 (bottom from) for different noise parameters are shown using H_2O-Ice spectrum from MICA spectral library.

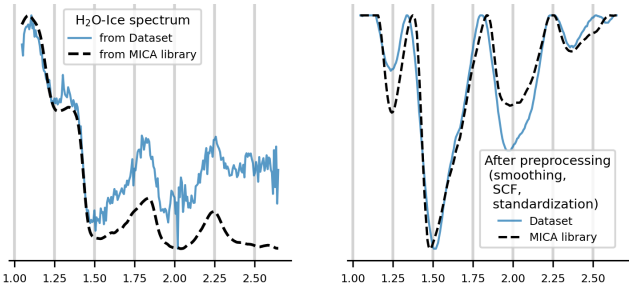


Fig. 4. A preprocessing pipeline that includes SCF as the continuum removal operation can translate the two H_2O-Ice spectrum (from MICA spectral library and CRISM TRDR dataset) in the left image to the right image.

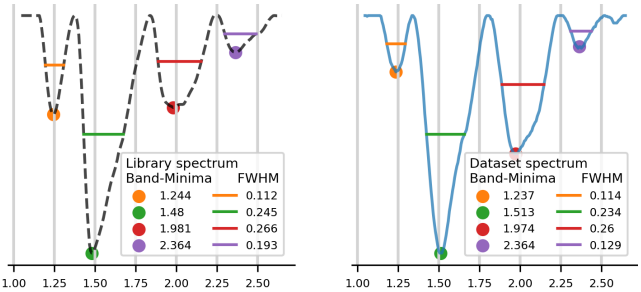


Fig. 5. The band-minima and FWHM for all the prominent segments of the processed spectra given in figure 4

which fits a parabola through local maxima in a segment, and smoothing helps remove unwanted kinks from spectra. Figure 4 illustrates the effect of the preprocessing steps applied on H_2O-Ice spectrum from the MICA library and dataset-4. To analyze the effectiveness of the proposed SCF in comparison with UCH, all the evaluation measures described below are calculated with SCF and UCH each used once in the pipeline while keeping all other steps the same. a

2) **Evaluation Measures:**

a) **Band-minima Score:** A band-minima refers to the minimum reflectance value at a spectral region, generally an absorption . The amalgamation of the positions of band-minima in different absorptions or segments of a spectrum is a useful tool for distinguishing materials as each of them have a unique set of absorptions over the wavelength domain. The percentage of spectra in a dataset for which the mean of shift in the positions of band-minima in a spectrum from those in the corresponding library spectrum is lesser than 0.02 μm is referred to as the band-minima score in this experiment.

b) **FWHM (Full Width at Half Maximum) Score:** In spectroscopy, FWHM is a measurement used to describe the shape of an absorption in a spectrum, that measures the width of an absorption in terms of the range of wavelengths at which the depth of the absorption is half of its maximum value. The FWHM score in this experiment represents the percentage of spectra in a dataset with a mean deviation of FWHM values from the library spectra less than 0.06 μm.

In these experiments, the segments of a continuum-removed spectrum are extracted using the shoulder points of both the test spectrum and the library spectrum. Then, prominent segments are selected by applying a threshold, and the band-minima location and FWHM of each matching segment in the two spectra are compared to determine if they are identical. An example of this process is shown in Figure 5. If the library and test spectra have different numbers of prominent segments, the segments from the spectrum with fewer segments are used as the base for the analysis.

c) **Identification Score:** Identification accuracy refers to the percentage of correctly identified spectra in the experiment. Datasets 1, 2, and 3 use the primary spectrum's class as its true label, while dataset 4 is pre-labeled. In identification procedures, each class in a dataset is typically assigned a value based on a similarity measure or the likelihood of a test spectrum belonging to that class. The test data is then assigned to the class with the highest similarity measure or likelihood. Proper preprocessing is necessary before identification, and the Weighted Sum of Segment Correlation (WSSC) procedure used in this study is highly dependent on it. The identification process can be impacted by slight changes in the preprocessing parameters, which may result in assigning a lower similarity measure or likelihood to the true-label class. As determining the most precise parameters for the preprocessing is beyond the scope of this study, here a general set of parameters as given in [25] is applied to the datasets. The identification score in this experiment is calculated based on the percentage of spectra in a dataset for which the true label is assigned a value within 5% of the highest assigned value for any class. The classification techniques used to identify a spectrum are described below.

3) **Identification Procedures:** In addition to the neural network-based model CNN [26], the accuracy of mineral identification is also determined by a novel spectra-matching method called Weighted Sum of Segment Correlation (WSSC). It shares some similarities with the shape-matching algorithm described in Clark et al.'s work on imaging spectroscopy [27], which uses a condition-driven and sophisticated framework for mineral identification for hyperspectral images. Unlike imaging spectroscopy, the method WSSC relies more on extracting the prominent absorptions without any pre-knowledge about the absorption features in the spectral library, by calculating a match index between any two spectra based on a new correlation measure. WSSC also differs in the process of calculating the spectra correlation coefficient and the match index. This method is described in detail before briefly discussing the specifications of the CNN model.

a) **Weighted Sum of Segment Correlation (WSSC):** This shape-matching technique generates the correlation coefficient for a segment in a preprocessed library spectrum with the corresponding spectra region in the preprocessed test spectrum. A matching index is calculated by summing the weighted correlation for all the segments in the preprocessed library spectrum, which are chosen based on a certain qualification criterion. The mineral type of a test spectrum can be identified by selecting the library mineral with the highest match index.

Let the spectral domain of a segment in the preprocessed library spectrum is $\mathcal{W}$; and in this spectral domain L and T, respectively, are the spectra portion in the preprocessed library spectrum and test spectrum. For T, the most effective translation to match with L is given by eq. III.2, while for L to match with T the same is given by eq. III.3. These equations involve adjusting coefficients a_1 and a_2 to shift the spectra, and b_1 and b_2 to modify the curvature during the translation process.

$$L = a_1 + b_1 * T \quad \text{(III.2)} \qquad T = a_2 + b_2 * L \quad \text{(III.3)}$$

The correlation coefficient $c^{\mathcal{W}}$ between the two segments T and L is calculated as eq. III.4.

$$c^{\mathcal{W}} = \begin{cases} (b_1 * b_2)^{\frac{1}{2}} & \text{if } b_1, b_2 > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{(III.4)}$$

The matching index I, constrained by the non-negativity of $c^{\mathcal{W}}$, does not penalize non-correlated segments in the two spectra being matched. This feature increases the chances of correctly identifying the mineral group, where the exact end-member class is not identified. $W^{\mathcal{W}}$, the weight associated with the segment, is a proportion of the sum of products of $f^{\mathcal{W}}$ and $d^{\mathcal{W}}$ over all the segments of the library spectra, where

If w is a wavelength in $\mathcal{W}$, from eq. III.2 the optimization is $\frac{\partial}{\partial a_1}(\sum_w (L_w - a_1 - b_1 * T_w)^2) = 0$ and $\frac{\partial}{\partial b_1}(\sum_w (L_w - a_1 - b_1 * T_w)^2) = 0$ solving which b_1 is derived as

$$b_1 = \frac{\sum_w (T_w * L_w) - (\sum_w T_w) * (\sum_w L_w)/n}{\sum_w (T_w)^2 - (\sum_w L_w)^2/n}$$

Similarly from eq. III.3 b_2 can be derived as

$$b_2 = \frac{\sum_w (L_w * T_w) - (\sum_w L_w) * (\sum_w T_w)/n}{\sum_w (L_w)^2 - (\sum_w T_w)^2/n}$$

$f^{\mathcal{W}}$ and $d^{\mathcal{W}}$ respectively are the FWHM and the depth at the band-minima of the segment.

$$I = \sum_{\forall \mathcal{W}} (W^{\mathcal{W}} * c^{\mathcal{W}}) \text{ where } W^{\mathcal{W}} = \frac{f^{\mathcal{W}} * d^{\mathcal{W}}}{\sum_{\forall \mathcal{W}} (f^{\mathcal{W}} * d^{\mathcal{W}})} \quad \text{(III.5)}$$

b) **CNN Model:** The experiment employs a shallow sequential CNN model as described in [25]. The model consists of three sets of convolution, max-pool, and drop-out layers, followed by a flatten layer and a dense output layer. The optimal parameter values for the model are determined using hyper-parameter tuning. The number of epochs is set to 50 with a batch size of 150. Leaky ReLU activation is utilized in all layers except the last one, which uses SoftMax activation. The loss function is categorical cross-entropy, and the optimizer is RAdam with a learning rate of 1×10^{-5}. The datasets are split into 60-20-20 ratio for training, validation and testing. The test accuracies are evaluated to assess the performance.

B. Performance Analysis

Table I contains the experimental results of band-minima score and FWHM score from dataset-1, 2 and 3 for different values of p, k and σ_N. The thresholds of 0.02 μm for band-minima score and 0.06 μm for FWHM score are selected based on experiments to match with the accuracy scores presented in table II, which contains the results of identification scores for the same datasets and specifications. Table III contains all these evaluation results of dataset-4. The primary objective of the analysis is to understand the effectiveness of the proposed method on some data with certain characteristics, and therefore, the degree of change in scores for a change in parameters of the data is more important than the exact scores themselves. However, achieving satisfactory high scores on an evaluation metric can provide evidence for the effectiveness of the method. The experimental results provide various insights, which are summarized below.

Table I shows that for both UCH and SCF, as the number of peaks in the estimated instrumental noise (p) increases, the band-minima and FWHM scores decrease (dataset-1 part). On the other hand, as the number of secondary spectra mixed with the primary spectra (k) increases, both the band-minima and FWHM scores increase (dataset-2 part), indicating that a smoother background curvature is formed when more spectra are mixed with the primary spectrum. If both instrumental artifacts (p) and the presence of other spectra (k) distort the primary spectra, as presented in dataset-3 part of table I, the scores become more dependent on k. This means that if both k and p are increased both scores increase. But, similar to the outcome of dataset-1 and 2, for a constant p value, the scores increase as k increases, and, the scores decrease for a constant k value as p value increases. This trend is evident in each row of the dataset-3 part. As environmental noise (σ_N) increases the scores decrease, which is expected as high levels of environmental noise distort local small absorption signatures in the spectra and require objective-specific preprocessing steps to handle them. The applied smoothing may not be sufficient for this task. In all of the specifications considered in these three datasets, the SCF outperformed UCH in terms of both

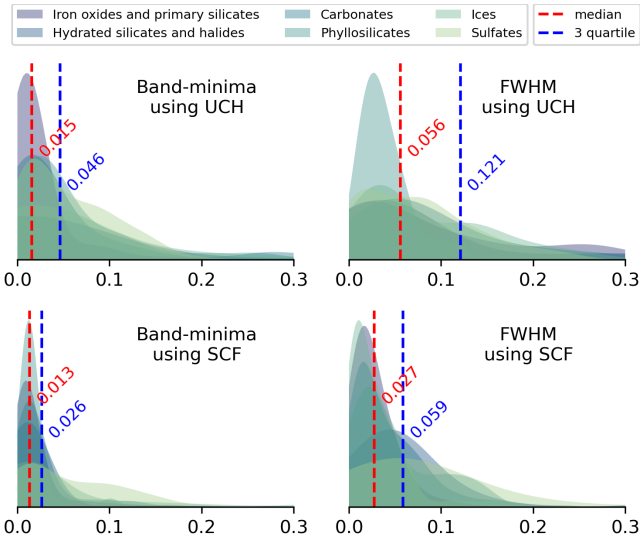


Fig. 6. The density plots of the differences in band-minima and FWHM values between the processed spectra of dataset-4 and the corresponding spectra from the MICA library, using SCF and UCH in preprocessing. The plots are colour-coded by the mineral groups. The lesser quartile-2 (median) and quartile-3 values indicate that the band-minima and FWHM values are more accurate when SCF is applied.

the band-minima and FWHM scores. Specifically, for UCH, the low FWHM scores indicate that UCH is more reliant on global shapes, and if the curvature is distorted by the noises, the shape of the extracted absorption signatures could deviate significantly from the library spectra. Conversely, SCF relies more on local shapes and can compare more accurately with the library spectra by extracting more spectra regions with small absorption signatures.

Table II shows a strong correlation between the accuracy of mineral class identification using the WSSC method and the correct determination of the position of the band-minima. This means that as the accuracy of band-minima determination increases or decreases, so does the accuracy of mineral class identification using the WSSC method. This is because both of these scores rely on accurately extracting the segments and are similarly impacted by the data specifications. The SCF method demonstrated better performance than the UCH method in terms of identification score by the WSSC method for each of the datasets, similar to the band-minima and FWHM scores. This suggests that using a higher number of segments with small spectral ranges in the SCF method can yield more precise information compared to using a lesser number of segments with long spectral ranges in the UCH method. Although the test accuracy scores of the CNN model for both SCF and UCH are relatively higher than WSSC, however unlike WSSC, the accuracy scores obtained by the CNN model do not consistently favour SCF over UCH. This suggests that shallow neural network models can more effectively learn hidden features when the original curvature of the unprocessed spectra is better preserved. UCH preserves the curvature better than SCF by extracting a lesser number of shoulder points.

Table III shows the evaluation metric scores for each mineral group separately. The experiments have indicated that, for both SCF and UCH methods, while there may be some mineral

TABLE I
BAND-MINIMA AND FWHM SCORES IN **data-1 data-2**, AND **data-3** FOR DIFFERENT p, k AND σ_N

CR Method	dataset-1			dataset-2			dataset-3								
								Band-minima				FWHM			
								k=1		k=2		k=1		k=2	
	p	Band-minima	FWHM	k	Band-minima	FWHM	σ_N	p=1	p=2	p=1	p=2	p=1	p=2	p=1	p=2
UCH	1	96.8	73.8	1	65.2	43.6	0.01	69.5	69.2	78.5	74.0	41.3	41.3	47.5	46.5
SCF		**96.9**	**96.0**		**74.3**	**83.4**		**77.5**	**73.2**	**84.4**	**81.3**	**82.3**	**82.3**	**85.9**	**85.7**
UCH	2	90.0	68.6	2	71.6	49.4	0.02	68.4	65.6	73.7	70.8	40.6	39.5	46.0	44.1
SCF		**94.0**	**93.0**		**78.1**	**85.1**		**75.1**	**73.0**	**81.7**	**80.2**	**79.5**	**79.0**	**82.6**	**83.4**
UCH	3	89.0	66.1	3	75.2	54.0	0.03	66.2	62.1	74.7	66.0	40.5	37.1	44.5	40.8
SCF		**92.6**	**91.3**		**80.5**	**88.0**		**74.0**	**69.0**	**81.2**	**73.0**	**75.4**	**76.0**	**81.8**	**78.0**
UCH	4	88.5	65.7	4	78.2	57.2	0.04	65.2	59.0	69.7	63.3	38.3	35.5	44.5	38.0
SCF		**92.0**	**91.0**		**81.1**	**88.4**		**72.1**	**65.5**	**78.7**	**71.2**	**73.5**	**72.0**	**78.8**	**76.0**

TABLE II
MINERAL IDENTIFICATION SCORES IN **data-1 data-2**, AND **data-3** FOR DIFFERENT p, k AND σ_N

CR Method	dataset-1			dataset-2			dataset-3								
								WSSC				CNN			
								k=1		k=2		k=1		k=2	
	p	WSSC	CNN	k	WSSC	CNN	σ_N	p=1	p=2	p=1	p=2	p=1	p=2	p=1	p=2
UCH	1	89.0	**100**	1	70.0	97.5	0.01	70.8	69.0	78.6	76.9	96.6	**94.7**	**99.0**	**98.9**
SCF		**100**	99.8		**77.1**	**98.7**		**77.4**	**75.9**	**88.1**	**86.9**	**98.1**	92.5	**99.0**	98.1
UCH	2	86.3	99.8	2	79.0	**99.5**	0.02	71.0	68.9	82.2	78.8	**95.0**	91.6	**99.2**	**96.9**
SCF		**99.1**	**100**		**85.0**	99.0		**75.3**	**73.9**	**85.3**	**83.6**	93.8	**95.6**	98.6	**98.1**
UCH	3	87.2	**100**	3	83.5	**100**	0.03	66.8	64.4	75.6	72.8	96.2	91.0	98.8	95.8
SCF		**98.5**	**100**		**90.0**	**100**		**73.8**	**70.0**	**83.3**	**78.4**	**97.5**	**92.8**	**99.0**	**96.6**
UCH	4	87.8	**100**	4	89.2	**100**	0.04	66.1	62.0	74.6	68.9	**95.0**	86.8	97.5	95.3
SCF		**98.1**	**100**		**93.9**	**100**		**73.3**	**69.2**	**82.6**	**80.4**	94.3	**88.1**	**98.3**	**95.9**

TABLE III
EVALUATION MEASURES FOR **data-4**

	Iron oxides and primary silicates		Hydrated silicates and halides		Carbonates		Phyllosilicates		Ices		Sulfates	
	UCH	**SCF**	UCH	**SCF**	UCH	**SCF**	UCH	**SCF**	UCH	**SCF**	UCH	**SCF**
band-minima	62.5	**68.4**	85.3	**96.0**	52.3	**65.5**	78.0	**89.6**	78.0	**86.0**	77.0	**80.5**
FWHM	64.0	**71.0**	65.3	**81.3**	73.0	**71.0**	67.2	**88.8**	73.0	**78.0**	63.1	**86.1**
WSSC	67.8	**79.5**	78.2	**87.6**	62.8	**77.0**	84.5	**88.0**	97.0	**98.0**	79.0	**87.0**
CNN	86.5	**86.9**	89.1	**91.3**	78.0	**81.0**	89.3	**92.7**	91.5	**88.7**	**91.6**	90.1

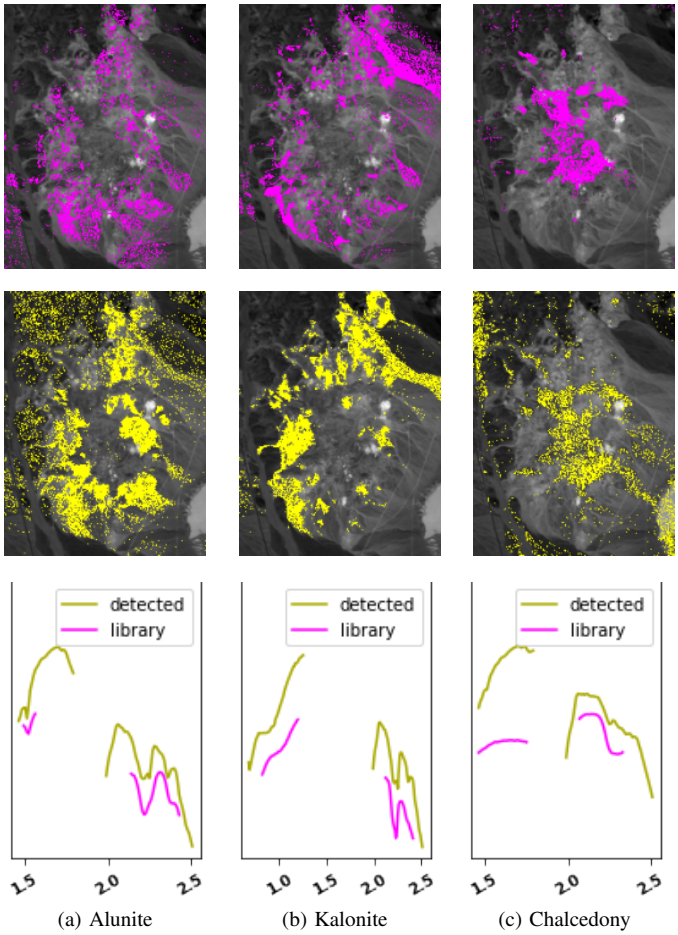
spectra whose true labels are not detected, a large portion of those spectra are still classified into the correct mineral group. This indicates that, at the very least, the WSSC method is able to detect mineral groups with a high degree of accuracy, even if the detection of individual minerals may not be as precise. The relatively high scores for all the mineral groups in table-III compared to the scores in table-II suggest that the synthetic dataset-3 has similar or even more noise than the TRDR dataset of real spectra. This indicates that similar synthetic datasets can be used effectively to determine the efficacy of such methods. The presented accuracy scores of WSSC in the table suggest that WSSC is not very accurate in detecting minerals that have prominent absorption signatures in the form of broad bands or sharp absorptions within a broad band in the spectrum, such as silicates and carbonates. Although the evaluation scores for these mineral groups could be increased with group-specific preprocessing steps. For dataset-4, it was observed that the accuracy scores of CNN obtained by using SCF are not always greater than the scores obtained by using UCH, similar to the results for dataset-3 in table-II. However, for most of the mineral groups, the SCF method yields higher accuracy scores, resulting in an overall improvement in the performance of around 2%. Figure 6 represents the density plots of the calculated mean shift in band-minima and mean change in FWHM for the spectra in dataset-4. It is evident from the figure that SCF detects the band-minima more accurately and preserves the FWHMs more precisely than UCH.

IV. Application on Hyper-spectral Images

In this section, the result of mineral identification on AVIRIS Cuprite data of earth-surface and CRISM MTRDR data of martian-surface are presented, which were obtained by using the previously stated WSSC method to match the wavelength position and shape of characteristic absorption features with the spectra in MICA library. SCF is applied as a preprocessing step on the data to remove the continuum and produce scaled spectra.

A. AVIRIS data [28]

The Cuprite dataset, which can be acquired from the AVIRIS NASA website, is the most important benchmark for the hyperspectral unmixing research that covers the area of Las Vegas, Nevada, USA. Alunite Hill is a volcanic dome in this location that formed about 20 million years ago, that has been altered by hot mineral-rich fluids that were circulating in the subsurface. As a result of this alteration, large deposits of alunite formed in the dome. The dataset comprised 224 channels, with wavelengths between 370 and 2480 nanometers, among which 188 channels are left after eliminating the noisy channels (1-2 and 221-224) and water absorption channels (104-113 and 148-167). A 250×190 pixel area is taken into consideration. This site was initially identified to have 14 different mineral types, although due to the minor variations among related minerals the number of end-members is decreased to 12 [27].

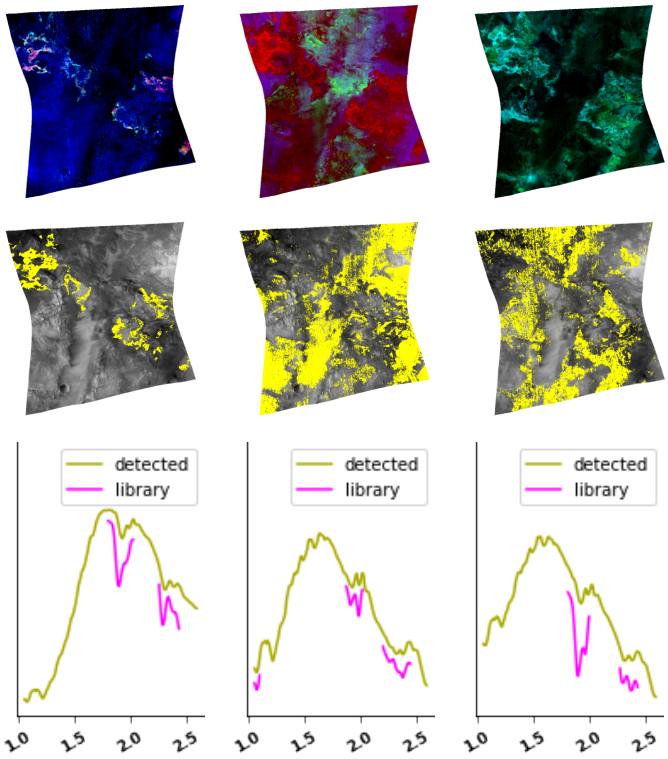


(a) Alunite (b) Kalonite (c) Chalcedony

Fig. 7. Detected dominant minerals in the Alunite hill scene from AVIRIS Cuprite data. Top Row: reference ground-truth as published in [27]; Middle Row: minerals detected by proposed method (SCF in preprocessing for continuum removal, WSSC for identification); Bottom Row: absorption feature matching between mean spectra of detected pixels and spectral library [28].

The Cuprite dataset is known for its discontinuous spectra at specific wavelengths, at 0.65, 1.24, 1.34 and 1.67 μm, which divides the wavelength domain into five distinct regions. To address this issue, the preprocessing steps, including SCF, are applied to each region separately. For WSSC, the match index is calculated by summing the separate match indices of the different regions. Different sets of spectral regions are selected to identify different minerals. For instance, to identify alunite minerals, only the two ranges between 1.34-1.67 μm and 1.68-2.17 μm are considered, as these contain the key absorptions that distinguish alunite from similar minerals.

The spatial distribution map of the most dominant three minerals in the site, that are kaolinite, alunite and chalcedony, as identified by WSSC is presented in figure 7. Kaolinite is a common clay mineral found in many soils and sedimentary rocks. It is mainly composed of aluminium silicate. Kaolinite has a prominent 1.4-μm feature due to vibrations of hydroxyl groups and a doublet at 2.2 μm (cantered at 2.16 and 2.21μm) from a combination of vibrations of Al-OH [29]. Alunite is a hydrous potassium aluminium sulfate mineral. The key spectral features of alunite are similar to Kaolinite along with another prominent absorption at 2.32 μm [27]. Chalcedony is a



(a) Mg-Carbonate (b) HCP (c) Mg-Smectite

Fig. 8. Detected dominant minerals in CRISM MTRDR data FRT93BE. Top Row: Browse Products (left- CR2, red/magenta coloured pixels indicate Mg-Carbonate) (middle- MAF, blue/magenta coloured pixels indicate HCP) (right- PFM, cyan coloured pixels indicate Fe/Mg-Smectite); Middle Row: minerals detected by the proposed method using SCF for continuum removal in preprocessing and WSSC for identification; Bottom Row: absorption feature matching between mean spectra of detected pixels and corresponding spectra from MICA library [12].

type of silica mineral that is found in a variety of forms (agate, onyx, and jasper) and is composed of very fine intergrowths of quartz and moganite. Chalcedony has a distinctive absorption feature at 2.25 μm.

In [27] Clark et al. proposed imaging spectroscopy for mineral classification and applied that particularly on Cuprite data to produce a vast quantity of different mineral maps on this region. The result obtained by the proposed method draws a very prominent visual similarity on the detected dominant minerals as shown in figure 7.

B. CRISM data

MTRDR CRISM data is publicly available through the Planetary Data System (PDS), which is maintained by the NASA Planetary Science Division [30]. The CRISM image FRT93BE (latitude: 19.1N, longitude: 283.5W) captured from the Jezero crater region is used in this study for mineral identification. Jezero Crater is a large impact crater located on the western edge of Isidis Planitia on the planet Mars. Jezero Crater is of great interest to scientists because it contains a delta at the western end barring the presence of clay minerals, which are indicative of the presence of liquid water in the past [31]. The dominant minerals in this region include olivine, pyroxene, and plagioclase feldspar, which are all common minerals in basaltic rocks. The image contains 804×832 pixels with 489 bands covering a wavelength range of 0.43-3.89 μm. However, in this experiment, only the wavelength range of 1-2.6 μm is considered, as various studies have shown that most of the absorptions in different mineral spectra occur in this wavelength range [32].

The detected dominant minerals in the scene are Mg-Carbonate, High-Calcium Pyroxene (HCP) and Mg-Smectite. Mg-Carbonate is a carbonate mineral with strong spectral features at 2.2-2.4 μm and a weaker feature at 2.33 μm, attributed to carbonate ions and hydroxyl/water in the crystal structure, respectively. HCP is a mafic mineral with a deep absorption feature at 0.9 μm due to iron and magnesium, a prominent feature at 1.2 μm due to calcium, and a shoulder feature at 1.6 μm due to electronic transitions. Mg-Smectite is a phyllosilicate mineral with spectral features at 1.8-1.9 μm attributed to OH groups and metal-oxygen bonds, and a characteristic feature at 2.2-2.35 μm due to stretching vibrations of OH groups and Mg-O bonds in the octahedral sheet of the crystal structure. Mg-Smectites in the scene are likely to be combined with Carbonates in a large abundance. The obtained mineral maps are mostly consistent with the browse products provided with the image which are shown in figure 8.

V. Discussions and Conclusion

This paper introduces a new technique called Segmented Curve Fitting (SCF), which is designed to effectively remove non-convex-shaped continuum in spectra. SCF achieves this by extracting more shoulder points of the existing absorptions in the spectra, than the traditional approach of estimating the continuum using an upper convex hull (UCH). This method preserves the resemblance of distinct absorption features between a test spectrum and its corresponding library spectrum more effectively than the previous method, UCH. This is demonstrated by calculating the mean shift of band-minima position and the difference in FWHMs on a set of synthetic datasets with varying degrees of complexity. When it comes to mineral identification, the proposed SCF method outperformed UCH when the identification is based on a spectral region correlation method, Weighted Sum of Segment Correlation (WSSC). However, SCF performed similarly to UCH when mineral identification in the synthetic datasets is done by a convolutional neural network model.

When the experiments were conducted on real CRISM TRDR dataset, mineral identification using the WSSC method with continuum removal by SCF showed an improvement in accuracy of over around 8% compared to using the UCH method for continuum removal, and around 2% if identification was performed using a CNN model, After examining the mineral group-wise accuracies, it was observed that the SCF method may not be very effective in segregating prominent broad-band absorption signatures or sharp absorptions within broad bands. Overall, the effectiveness of the proposed SCF method lies in detecting narrow to medium absorptions from spectra with curvature distortion more effectively than the existing method.

The proposed SCF method could be extended in the following aspects to improve its performance:

- SCF uses all the local maxima in a spectral region to fit a parabolic curve. This process could be made more selective by imposing certain criteria on the maxima.
- The method could be applied recursively to test spectra to extract additional local absorptions until a stop criterion is met.
- The method could be further extended to automatically distinguish broad absorptions. The above suggestions or other approaches could be explored to achieve this goal.

REFERENCES

[1] Sabine Chabrillat, Alexander FH Goetz, Lisa Krosley, and Harold W Olsen, "Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution," *Remote sensing of Environment*, vol. 82, no. 2-3, pp. 431–445, 2002.

[2] Peter M. Atkinson, Eulogio Pardo-Iguzquiza, and Mario Chica-Olmo, "Downscaling cokriging for super-resolution mapping of continua in remotely sensed images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 2, pp. 573–580, 2008.

[3] Anthony M. Filippi and John R. Jensen, "Effect of continuum removal on hyperspectral coastal vegetation classification using a fuzzy learning vector quantizer," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, pp. 1857–1869, 2007.

[4] Roger N Clark and Paul G Lucey, "Spectral properties of ice-particulate mixtures and implications for remote sensing: 1. intimate mixtures," *Journal of Geophysical Research: Solid Earth*, vol. 89, no. B7, pp. 6341–6348, 1984.

[5] Roger N Clark and Ted L Roush, "Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications," *Journal of Geophysical Research: Solid Earth*, vol. 89, no. B7, pp. 6329–6340, 1984.

[6] Michael F Goodchild, "Integrating gis and remote sensing for vegetation analysis and modeling: methodological issues," *Journal of Vegetation Science*, vol. 5, no. 5, pp. 615–626, 1994.

[7] Jinkai Zhang, Benoit Rivard, and Arturo Sanchez-Azofeifa, "Derivative spectral unmixing of hyperspectral data applied to mixtures of lichen and rock," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 9, pp. 1934–1940, 2004.

[8] Curtis E Woodcock and Alan H Strahler, "The factor of scale in remote sensing," *Remote sensing of Environment*, vol. 21, no. 3, pp. 311–332, 1987.

[9] AN French, JM Norman, and MC Anderson, "A simple and fast atmospheric correction for spaceborne remote sensing of surface temperature," *Remote Sensing of Environment*, vol. 87, no. 2-3, pp. 326–333, 2003.

[10] Nina Kristine Boesche, Christian Rogass, Christin Lubitz, Maximilian Brell, Sabrina Herrmann, Christian Mielke, Sabine Tonn, Oona Appelt, Uwe Altenberger, and Hermann Kaufmann, "Hyperspectral ree (rare earth element) mapping of outcrops—applications for neodymium detection," *Remote Sensing*, vol. 7, no. 5, pp. 5160–5186, 2015.

[11] David A Neave, Martin Black, Teal R Riley, Sally A Gibson, Graham Ferrier, Frances Wall, and Sam Broom-Fendley, "On the feasibility of imaging carbonatite-hosted rare earth element deposits using remote sensing," *Economic Geology*, vol. 111, no. 3, pp. 641–665, 2016.

[12] Christina E. Viviano, Frank P. Seelos, Scott L. Murchie, Eliezer G. Kahn, Kimberley D. Seelos, Howard W. Taylor, Kelly Taylor, Bethany L. Ehlmann, Sandra M. Wiseman, John F. Mustard, and M. Frank Morgan, "Revised CRISM spectral parameters and summary products based on the currently detected mineral diversity on mars," *Journal of Geophysical Research: Planets*, vol. 119, no. 6, pp. 1403–1431, June 2014.

[13] Mario Parente, Heather D Makarewicz, and Janice L Bishop, "Decomposition of mineral absorption bands using nonlinear least squares curve fitting: Application to martian meteorites and crism data," *Planetary and Space Science*, vol. 59, no. 5-6, pp. 423–442, 2011.

[14] A.J. Brown, "Spectral curve fitting for automatic hyperspectral data analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1601–1608, 2006.

[15] Michael A McCraig, Gordon R Osinski, Edward A Cloutis, Roberta L Flemming, Matthew RM Izawa, Vishnu Reddy, Sherry K Fieber-Beyer, Loredana Pompilio, Freek van der Meer, Jeffrey A Berger, et al., "Fitting the curve in excel®: Systematic curve fitting of laboratory and remotely sensed planetary spectra," *Computers & geosciences*, vol. 100, pp. 103–114, 2017.

[16] Saeid Asadzadeh and Carlos Roberto de Souza Filho, "A review on spectral processing methods for geological remote sensing," *International journal of applied earth observation and geoinformation*, vol. 47, pp. 69–90, 2016.

[17] Roger N Clark and Trude VV King, "Automatic continuum analysis of reflectance spectra," in *JPL proceedings of the 3rd airborne imaging spectrometer data analysis workshop*, 1987.

[18] E Ben-Dor, "The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process," *Remote Sensing of Environment*, vol. 61, no. 1, pp. 1–15, July 1997.

[19] MS Rice, EA Cloutis, JF Bell Iii, DL Bish, BH Horgan, SA Mertzman, MA Craig, RW Renaut, B Gautason, and B Mountain, "Reflectance spectra diversity of silica-rich materials: Sensitivity to environment and implications for detections on mars," *Icarus*, vol. 223, no. 1, pp. 499–533, 2013.

[20] Ralph E. Milliken and John F. Mustard, "Quantifying absolute water content of minerals using near-infrared reflectance spectroscopy," *Journal of Geophysical Research*, vol. 110, no. E12, 2005.

[21] MA Craig, EA Cloutis, V Reddy, DT Bailey, and MJ Gaffey, "The effects of grain size,¡ 10 μm-4.75 mm, on the reflectance spectrum of planetary analogs from 0.35-2.5 μm," in *39th Annual Lunar and Planetary Science Conference*, 2008, number 1391, p. 2082.

[22] Tasha L Dunn, Timothy J McCoy, JM Sunshine, and Harry Y McSween Jr, "A coordinated spectral, mineralogical, and compositional study of ordinary chondrites," *Icarus*, vol. 208, no. 2, pp. 789–797, 2010.

[23] Emanuele Plebani, Bethany L Ehlmann, Ellen K Leask, Valerie K Fox, and M Murat Dundar, "A machine learning toolkit for crism image analysis," *Icarus*, vol. 376, pp. 114849, 2022.

[24] Sampriti Soor, "Continuum removal," https://github.com/SampritiSoor/continuumRemoval, [Online].

[25] Priyanka Kumari, Sampriti Soor, Amba Shetty, and Shashidhar G. Koolagudi, "A fully-automated framework for mineral identification on martian surface using supervised learning models," *IEEE Access*, vol. 11, pp. 13121–13137, 2023.

[26] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[27] Roger N Clark, Gregg A Swayze, K Eric Livo, Raymond F Kokaly, Steve J Sutley, J Brad Dalton, Robert R McDougal, and Carol A Gent, "Imaging spectroscopy: Earth and planetary remote sensing with the usgs tetracorder and expert systems," *Journal of Geophysical Research: Planets*, vol. 108, no. E12, 2003.

[28] Robert O Green, Michael L Eastwood, Charles M Sarture, Thomas G Chrien, Mikael Aronsson, Bruce J Chippendale, Jessica A Faust, Betina E Pavri, Christopher J Chovit, Manuel Solis, et al., "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris)," *Remote sensing of environment*, vol. 65, no. 3, pp. 227–248, 1998.

[29] Roger N. Clark, Trude V. V. King, Matthew Klejwa, Gregg A. Swayze, and Norma Vergo, "High spectral resolution reflectance spectroscopy of minerals," *Journal of Geophysical Research*, vol. 95, no. B8, pp. 12653, 1990.

[30] Justin N. Maki, "Mer 1 mars microscopic imager radiometric rdr ops v1.0," 2004.

[31] Vivienne H. Payne, Eli J. Mlawer, Karen E. Cady-Pereira, and Jean-Luc Moncet, "Water vapor continuum absorption in the microwave," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2194–2208, 2011.

[32] J Carter, F Poulet, S Murchie, and JP Bibring, "Automated processing of planetary hyperspectral datasets for the extraction of weak mineral signatures and applications to crism observations of hydrated silicates on mars," *Planetary and Space Science*, vol. 76, pp. 53–67, 2013.

Priyanka Kumari is a Senior Research Fellow at Department of Water Resources and Ocean Engineering in National Institute of Technology, Karnataka. She received Bachelor of Technology degree in Electrical and Electronics Engineering in the year 2013 from BPUT, Rourkela. She received her M.Tech degree in Remote Sensing in the year 2015 from BIT Mesra, Ranchi. Her research interests include Remote Sensing, Geographic Information Systems, and Machine Learning.

Sampriti Soor received Bachelor of Technology degree in Information Technology from West Bengal University of Technology in 2011 and Master of Engineering degree in Software Engineering from Jadavpur University, Kolkata in 2015. He was awarded the PhD degree in Computer Science by the Indian Statistical Institute in 2022. He is currently working as an Assistant Professor at the Kalinga Institute of Industrial Technology, Bhubaneswar. His research interests are in the fields of Data Science, Image Processing, and Mathematical Morphology.

Amba Shetty has completed her bachelor's from Mangalore University, Karnataka, India in Civil Engineering and Master's from K.R.E college, Mangalore, Karnataka in Depertment of Water Resources and Ocean Engineering. She completed her doctoral studies at the K.R.E college, Mangalore, Karnataka in the year 2006. Her research interests are Hydrological Modelling, Hyperspectral Remote sensing, GIS, Surface Water Hydrology, Microwave remote sensing application in soil moisture and flood studies. Currently, she is working in the National Institute of Technology Karnataka, Surathkal, India as a Professor.

Shashidhar G. Koolagudi has completed his bachelor's from Karnataka University, Dharwad, Karnataka, India in Computer Science and Engineering and masters from the Indian Institute of Engineering Science and Technology, Shibpore, Howrah in Computer Science and Information Technology. He completed his doctoral studies at the Indian Institute of Technology, Kharagpur, West Bengal in the year 2011. His research interests are Machine Learning, Speech, and audio signal processing. Currently, he is working in the National Institute of Technology Karnataka, Surathkal, India as an associate professor.