Segmented Curve-Fitting Method for Continuum Removal in CRISM MTRDR data

Priyanka Kumari[1], Sampriti Soor[2], Amba Shetty[3], Shashidhar G Koolagudi[4]

[1] Department of Water Resources and Ocean Engineering, NIT Karnataka, India, 575025 - singh.priyanka854@gmail.com
[2] Center for Intelligent Cyber Physical Systems, IIT Guwahati, India, 781039 - sampreetiworkid@gmail.com
[3] Department of Water Resources and Ocean Engineering, NIT Karnataka, India, 575025 – amba@nitk.edu.in
[4] Department of Computer Science and Engineering, NIT Karnataka, India, 575025 - koolagudi@nitk.edu.in

Keywords: Continuum Removal, Segmented Curve-fitting, Hyperspectral Images, Mineral Mapping.

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 are crucial for material identification, traditionally achieved by the apparent continuum removal methods like using an Upper Convex Hull (UCH). Nevertheless, most of these methods struggle when baseline curvature exceeds certain limits, often combining distinct absorptions. In this paper, a new apparent continuum removal technique called *Segmented Curve-Fitting* (SCF) is proposed, which requires no prior information about the spectrum but excels in accurately extracting distinct absorptions, even in the presence of significant curvature. The performance of SCF is compared with UCH and a few other apparent continuum removal methods previously used in literature, 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 1.5% for a Convolutional Neural Network. Furthermore, an SCF-based mineral identification framework demonstrates its effectiveness in identifying the dominant minerals on CRISM MTRDR hyperspectral data collected from different locations on the Martian surface.

1. Introduction

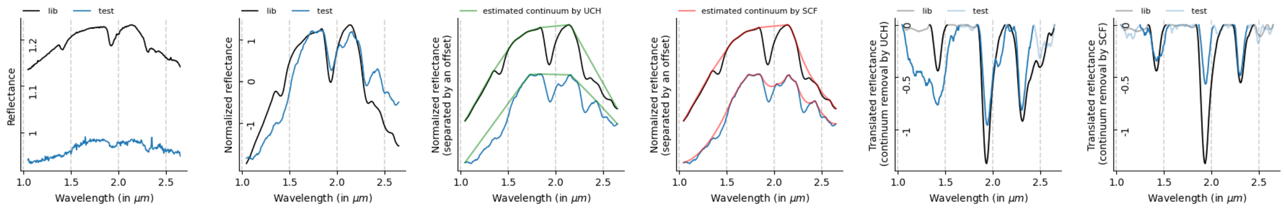
Hyperspectral imaging systems can capture subtle changes in reflectance that reveal specific minerals or materials that are hard to identify using other remote sensing methods (Chabrillat et al., 2002). Each material has a unique “spectral signature,” which is a pattern of reflectance values across different wavelengths showing characteristic positions and shapes of absorption features. When a hyperspectral system captures these reflectances, the combined spectral signatures of all materials in a target area are stored within each pixel. This creates a smooth background intensity, called the “continuum” in hyperspectral imagery (Clark and Lucey, 1984, Clark and Roush, 1984).

Hyperspectral sensors, due to their design, require multiple image frames to cover large areas, which are later combined into spatially connected mosaics. However, these mosaics often show varying continuum characteristics across and even within frames (Kumari et al., 2023b). While some hyperspectral data may have straightforward continuums that need minimal adjustment, others, particularly those with materials having complex spectral shapes, have more challenging nonlinear continuums. These can be influenced by factors like scattering, atmospheric conditions, or sensor variations (Zhang et al., 2004, Zhao et al., 2015). Accurately estimating this continuum is important as it serves as a baseline reference for identifying and analyzing distinct spectral features (Parente et al., 2011).

Continuums in a hyperspectral image are typically estimated pixel-by-pixel and are generally considered to be independent of continuums in nearby pixels (Viviano et al., 2014). However, a simple distance-based similarity approach doesn’t work well, as it considers the entire spectrum, rather than only the absorption signatures, which can be affected by the continuum. The continuum can shift the overall shape of a spectrum, making it challenging to identify the material correctly, even when the main absorption features remain the same (Parente et al., 2011). Figures 1a-1b show a comparison of a sample Mg-Smectite spectrum from a CRISM TRDR image and its corresponding spectrum in the MICA library, highlighting differences in their global curvature and reflectance scales. After aligning their scales through normalization, these curvature differences become even more visually apparent.

Continuum removal has two primary roles: mitigating spectral baseline effects from environmental or instrumental sources (imposed continuum removal - ICR) and isolating absorption features (apparent continuum removal - ACR). The ICR methods assume a functional form, often using polynomial or power-law models to estimate the baseline (Brown, 2006). ICR can be effective when the continuum shape is predictable. ACR methods, on the other hand, estimate the continuum by identifying regions without absorption features, i.e., the blunt regions, and by linking the endpoints of these (McCraig et al., 2017). Traditional ACR methods face limitations in handling spectra with low-intensity signals or significant baseline curvature.

In ACR continuums are commonly estimated using a hull over the spectrum, where the shoulder points are identified as the locations where the hull touches with the spectrum. The spectrum is either subtracted by or divided by the estimated continuum [eq. 1] to nullify the effects of the continuum. All the shoulder-points in a continuum-removed spectrum remain at the same level. Figure 1c illustrates the Upper Convex Hulls over the library and test Mg-Smectite spectra, with the corresponding continuum-removed spectra displayed in figure 1e, showcasing



(a) Raw Mg-Smectite Spectra (b) After smoothing and Scaling (c) UCH on the spectra (d) Continuum by SCF (e) Continuum removal by UCH (f) Continuum removal by SCF

Figure 1. **(a)** Mg-Smectite spectra, with one sourced from the MICA spectral library and another from a CRISM TRDR dataset; **(b)** The disparity in global curvatures between the two Mg-Smectite spectra are noticeable after the scaling operation; **(c)** The estimated continuums using Upper Convex Hulls (UCH) on the preprocessed library and test spectrum in (b); **(d)** The estimated continuums by the proposed Segmented Curve Fitting (SCF) method; **(e)** Continuum removed spectra obtained by subtracting the UCH from the preprocessed spectra in (c); **(f)** Continuum removed spectra of the spectra in (b) by the proposed method SCF. In (e) and (f), extracted absorptions (considering those larger than a specific threshold) in the continuum-removed spectra are highlighted with darker colours. **Note-1:** The extracted absorptions in the continuum-removed library and test spectra by SCF in (f) exhibit more resemblance than those by UCH in (e). **Note-2:** The figures in the top and bottom row are presented with actual scales on the y-axis, while those in the middle row have the same scale as in (b), although are separated by an offset for visual clarity.

how absorption signatures are automatically extracted between pairs of consecutive shoulder points if the absorption size exceeds a threshold. On the other hand, figure 1f depicts the absorptions detected by the proposed method, showing a better similarity between the absorptions extracted from the library spectrum and the test spectrum.

Existing ACR methods fall into two categories: those without prior spectral information and those with known absorption positions. For the former, *Upper Convex Hull* (UCH) (Clark and Roush, 1984) remains popular due to its simplicity and effectiveness within certain limits of continuum curvature. *Segmented Upper Hull* (SUH) restricts changes in slope and thus can struggle with non-convex continuums (Clark and King, 1987). Recent advancements, like scale-space filtered alpha hulls and *Geometric Hulls* (GH) (Mielke et al., 2015), address some of these limitations by iteratively detecting shoulder points, yet they risk overfitting and excessive segmentation. In cases where absorption shoulder locations are known, techniques like cubic splines and reference spectral background removal have proven effective (Clark and Lucey, 1984, Zhao et al., 2015). While functional within specific curvature ranges, these traditional methods may struggle to preserve individual absorption features when the continuum shape varies significantly.

An effective ACR method should be resilient to the variations in the continuum shape. This study introduces a novel ACR method, Segmented Curve-Fitting (SCF), which effectively addresses these challenges without requiring prior spectral information. SCF uses an initial set of shoulder points derived from UCH, followed by segmentation into parabolic fits that refine the continuum estimation in each segment. Unlike typical ACR methods that connect shoulder points with straight lines, SCF applies parabolic curves, offering flexibility to handle both concave and convex shapes within any portion of the test spectrum. This approach adapts a piece-wise technique like SUH but without the slope constraints, providing a more realistic continuum estimation that better preserves absorption features. To validate SCF’s effectiveness, we tested it against traditional ACR methods like UCH, SUH, and geometric hulls (GH) using a hyperspectral dataset. SCF’s performance was assessed through similarity measurements of band-center positions and Full Width at Half Maximum (FWHM) values of absorption features between test spectra and true labels in spectral libraries. Mineral identification accuracy was evaluated using classification methods, including a Convolutional Neural Network (CNN) and shape-based matching methods, confirming SCF’s resilience to continuum shape variations and its improved feature extraction capabilities.

The paper is organized as follows: Section 2 covers relevant notations and metrics; Section 3 provides a detailed explanation of SCF; Section 4 compares SCF with previous continuum removal approaches for mineral identification, and Section 5 presents experimental results on Martian CRISM MTRDR hyperspectral data.

2. Notations

A spectrum and its corresponding wavelength domain are represented as vectors denoted by bold lowercase letters. Consider a spectrum, denoted as $\mathbf{r}$, spanning the wavelength domain $\mathbf{w}$ with wavelengths sorted in ascending order. For a given wavelength w ($\in \mathbf{w}$) the corresponding element in the spectrum is $\mathbf{r}_{(w)}$. Similarly, for a wavelength portion λ of $\mathbf{w}$ (where λ contains some consecutive wavelengths of $\mathbf{w}$), the corresponding spectra portion is denoted as $\mathbf{r}_{(\lambda)}$. A subset of a wavelength portion (which may not necessarily contain consecutive wavelengths) is denoted by λ', while a collection of such wavelength portions λ is represented by $\mathbf{\Lambda}$. For set of wavelengths λ' ($\subseteq \lambda$ or $\subseteq \mathbf{w}$), the corresponding spectrum elements are denoted by $\mathbf{r}_{(\lambda')}$. When a spectrum is not subscripted with an index, a specific wavelength, or a set of wavelengths, it indicates that the spectrum is considered for the entire wavelength domain. On the other hand, if a spectrum or a curve (let $\mathbf{t}$) is defined over a wavelength portion λ only, rather than the entire wavelength domain, the spectrum is denoted with a superscripted λ (e.g. $\mathbf{t}^{(\lambda)}$).

Let $\mathbf{c}$ be an estimated continuum of $\mathbf{r}$, then as mentioned in section **??** the continuum removed spectrum can be calculated by the operation $\mathscr{C}_{\mathbf{c}}(\mathbf{r})$ using division or subtraction as given in eq. 1.

$$\begin{aligned} \mathscr{C}_{\mathbf{c}}^{div}(\mathbf{r}) &= \mathbf{r}/\mathbf{c} \\ \mathscr{C}_{\mathbf{c}}^{sub}(\mathbf{r}) &= \mathbf{r}-\mathbf{c} \end{aligned} \tag{1}$$

In this paper, if the removal method is not specified at the superscript level for a continuum removing function $\mathscr{C}$, it is assumed that both subtraction and division can be applied. Similarly, if a continuum $\mathbf{c}$ is not specified at the subscript level, it is assumed that the continuum is estimated by an upper convex hull.

In section 3 while describing the proposed Segmented Curve-Fitting method three translated spectra $\mathbf{t}^1$, $\mathbf{t}^2$ and $\mathbf{t}^3$ are utilized. Specifically, $\mathbf{t}^1$ is defined over the entire wavelength domain, while $\mathbf{t}^2$ and $\mathbf{t}^3$ are defined over specific wavelength portions. Furthermore, a fitted curve, denoted as $\mathbf{p}$ is defined for each of these wavelength portions.Some additional notations are introduced in section 4.1.3 pertaining to different evaluation measures for assessing a preprocessed test spectrum $\mathbf{t}$ in comparison to the corresponding preprocessed library spectrum $\mathbf{l}$.

3. Segmented Curve-Fitting for Continuum Removal

Traditional continuum removal methods assume convex curvature in hyperspectral data, often estimates the continuum as the UCH. However, natural spectral continua may vary, with non-convex regions that affect local wavelengths unevenly. Our proposed method SCF addresses this issue by segmenting the spectrum and fitting continuum shapes for each segment independently. Although this approach may overlook the global shape of the spectrum, it enables a more precise enhancement of weaker band depths. The SCF methodology is demonstrated in Figure 2 with a sample H_2O-Ice spectrum from the MICA spectral library (Viviano et al., 2014). This method is effective for analyzing pure spectra as well, extracting distinct absorption signatures.

❍ ***Step-1 (Segmenting the spectrum):*** Initially, shoulder points in the spectrum are identified to partition the spectrum into segments, where parabolic curves are fitted to further refine shoulder points and estimate the continuum. The UCH method is employed here to segment the spectrum, as outlined in Algorithm 1. The UCH algorithm operates in linear time, using general stack operations *push*, *pop*, and *length* (Cormen et al., 2022).

Algorithm 1 Pseudocode to calculate the UCH

```
function UPPERCONVEXHULL($\mathbf{r}, \mathbf{w}$)
    initialize Empty stack $\mathcal{S}$
    $\mathcal{S}.push(1)$, $\mathcal{S}.push(2)$
    for $i \in 3, 4, ..., \mathbf{r}.length$ do:
        if $\frac{\mathbf{r}_{\mathcal{S}_{-1}} - \mathbf{r}_{\mathcal{S}_{-2}}}{\mathbf{w}_{\mathcal{S}_{-1}} - \mathbf{w}_{\mathcal{S}_{-2}}} \geq \frac{\mathbf{r}_i - \mathbf{r}_{\mathcal{S}_{-2}}}{\mathbf{w}_i - \mathbf{w}_{\mathcal{S}_{-2}}}$ then
                ▷ $\mathcal{S}_{-k}$ is the $k$-th element from top in $\mathcal{S}$
            $\mathcal{S}.push(i)$
        else
            $\mathcal{S}.pop()$
            if $\mathcal{S}.length == 1$ then
                $\mathcal{S}.push(i)$
    return $linear\_interpolation(\mathcal{S}, \mathbf{r}, \mathbf{w})$
```

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, as

$$\mathbf{t}^1 = \mathscr{C}(\mathbf{r}) \tag{2}$$

Let m be the highest value in $\mathbf{t}^1$ ($m = 1$ for $\mathscr{C}^{div}$ and $m = 0$ for $\mathscr{C}^{sub}$ in eq. 1). A wavelength segment λ characterized by lesser than m values between two m-valued wavelengths in $\mathbf{t}^1$ contains an absorption in $\mathbf{r}$, where the m-valued wavelengths at the two ends contain the shoulder-points of the absorption. Fig. 2(c) shows such spectrum segments of $\mathbf{t}^1$ for the example spectrum. When a parabolic curve $\mathbf{p}^{(\lambda)}$ is fitted through a spectrum segment $\mathbf{t}^1_{(\lambda)}$, by minimizing the quadratic error while ensuring it passes through the end-points of the segment, some points within the segment lie above this curve. If $\mathbf{t}^1_{(\lambda)}$ is divided or subtracted by $\mathbf{p}^{(\lambda)}$ to generate a translated spectra segment $\mathbf{t}^{2(\lambda)}$, translated points of those remain above the m level in $\mathbf{t}^{2(\lambda)}$ too. Now, on $\mathbf{t}^{2(\lambda)}$ another continuum removal operation can be performed to extract more shoulder points.

Now, the parabolic curve can be fitted through various sets of points in a segment of $\mathbf{t}^1$. The fitting points can be all the points in the segment, all the local maxima, all the local minima, or some points selected on some conditions. Alternatively, a lower convex hull also can be used for this purpose. As in the experiments better results are observed if the curves are fitted through the local maxima, in this paper all the local maxima are considered as the fitting points. By this, the method doesn't rely on specific thresholds for selecting the fitting points, making it robust across different types of data. Note that, in the example shown in figure 2, a detected segment around $2.6\mu m$ is not considered for further processing as it does not have any local maxima. Specifically, only the segments having at least one local peak are selected for further processing. The words 'maxima' and 'minima' in this paper indicate both singular and plural forms.

❍ ***Step-2 (Segment-wise fitting of parabolic curves)*** 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. Let λ be a wavelength segment picked for further processing and $\mathbf{\Lambda}$ be the set of all picked wavelength segments. Considering w_s and w_t as the two end-wavelengths of λ, the general equation of the parabolic curve fitted in λ becomes

$$\mathbf{p}^{(\lambda)}_{(w)} = a(w - w_s)(w - w_t) + m \text{ for } w \in \lambda \tag{3}$$

The value of a in 3 can be calculated from $\mathbf{t}^1_{(\lambda')}$ minimizing the quadratic error as following, where λ' ($\subset \lambda$) contains the wavelengths of the local maxima in $\mathbf{t}^1_{(\lambda)}$.

$$\begin{aligned}
&\text{let } p = (w - w_s)(w - w_t) \\
&\text{Error } J = \sum_{w \in \lambda'} (\mathbf{t}^1_{(w)} - ap - m)^2 \\
&\text{With optimum fitting } \frac{\delta J}{\delta a} = 0 \\
&\implies \sum_{w \in \lambda'} (\mathbf{t}^1_{(w)} - ap - m)p = 0 \\
&\implies \sum_{w \in \lambda'} \mathbf{t}^1_{(w)} p \;-\; a \sum_{w \in \lambda'} p^2 \;-\; m \sum_{w \in \lambda'} p = 0 \\
&\implies a = \frac{\sum_{w \in \lambda'} \mathbf{t}^1_{(w)} p \;-\; m \sum_{w \in \lambda'} p}{\sum_{w \in \lambda'} p^2}
\end{aligned} \tag{4}$$

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 for all the segments λ in $\mathbf{\Lambda}$ that generates $\mathbf{t}^{2(\lambda)}$ from $\mathbf{t}^1_{(\lambda)}$ and $\mathbf{p}_{(\lambda)}$. To extract additional shoulder-points, another step of continuum removal is performed on each of the translated segments $\mathbf{t}^{2(\lambda)}$ by UCH.

$$\mathbf{t}^{2(\lambda)} = \mathscr{C}_{\mathbf{p}^{(\lambda)}}(\mathbf{t}^{1}_{(\lambda)}) \qquad (5)$$

$$\mathbf{t}^{3(\lambda)} = \mathscr{C}(\mathbf{t}^{2(\lambda)}) \qquad (6)$$

Fig. 2(d) shows $\mathbf{t}^{2(\lambda)}$ and UCH for the selected segments, and fig. 2(e) shows $\mathbf{t}^{3(\lambda)}$ for each of these segments. Note that, to ensure continuity in the continuum-removed spectra, i.e., to keep all the shoulder-points on the continuum-removed spectrum at the same level, in all continuum removal operations ($\mathscr{C}$), employing the same translation method is necessary, that is, employing either only divisions or only subtractions.

❍ ***Step-4 (Merging the translated segments)*** The final step of the procedure involves combining the independently translated segments of $\mathbf{t}^3$ with the unpicked portions in step-2 from the initially translated spectrum $\mathbf{t}^1$ on the entire wavelength domain $\mathbf{w}$ to generate the continuum-removed spectrum.

$$SCF(\mathbf{r}, \mathbf{w})_{(w)} = \begin{cases} \mathbf{t}^{3(\lambda)}_{\ (w)}, & \text{if } \exists \lambda \in \mathbf{\Lambda} \mid w \in \lambda \\ \mathbf{t}^{1}_{\ (w)} & \text{otherwise} \end{cases} \quad \forall w \in \mathbf{w} \qquad (7)$$

Fig. 2(f) shows the continuum-removed spectrum 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 and SUH determine the continuum from the full spectrum first, then eliminate its effect from the spectrum by subtraction or division operation as given in eq 1; on the other hand, the proposed method directly removes the effect of the continuum from the spectra by eq 7 without estimating the continuum particularly, 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 $SCF(\mathbf{r}, \mathbf{w})$ 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. (4-7)) in linear computational time independently and each in linear time. Note that, while it may appear reasonable to limit the computation by considering only the local peaks from $\mathbf{r}$ to further reduce the runtime, this approach would not be suitable for detecting the shoulder points correctly.

4. Performance Evaluation

The performance of the proposed SCF continuum removal method is evaluated using a labeled dataset of TRDR spectra (Plebani et al., 2022). To assess SCF's effectiveness in preserving absorption features, *Band-center* and *FWHM scores* are calculated, whereas Identification accuracy is evaluated through an *Identification score* by comparing the test spectra with true spectra in the MICA spectral library, which includes 31 mineral spectra from six mineral groups such as iron oxides, primary silicates, ices, sulfates, phyllosilicates, carbonates, and hydrated silicates and halides. The test spectra and reference spectra from the library undergo a preprocessing pipeline (Kumari et al., 2024), including continuum removal, allowing performance comparisons with other continuum removal methods. Identification is conducted using two approaches: the Weighted Sum of Segmented Correlation (WSSC) curve-matching method and a shallow CNN model, showcasing SCF's performance with different classification methods.

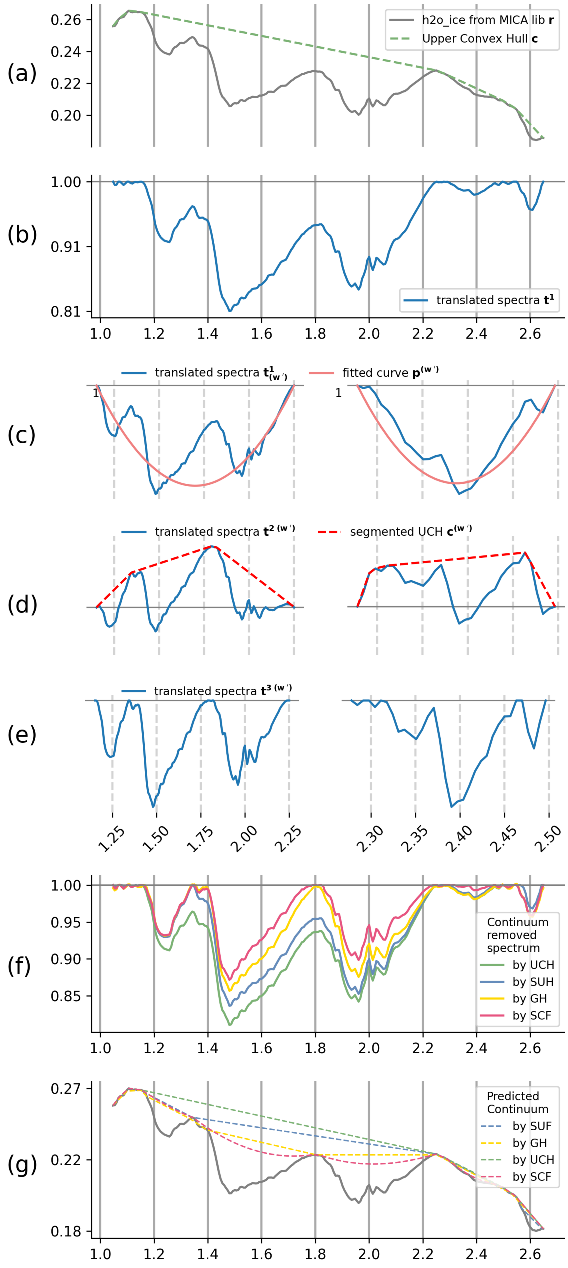


Figure 2. **(a)** H_2O-Ice spectrum R from MICA library and the UCH of it; **(b)** Translated spectrum $\mathbf{t}^1$ using division method of eq 1; **(c)** segments are extracted from $\mathbf{t}^1$ and parabolic curves as in eq.-3 are fitted; **(d)** Translated spectrums $\mathbf{t}^{2\ (\lambda)}$ from 5 and corresponding UCHs; **(e)** Translated spectrums $\mathbf{t}^{3\ (\lambda)}$ from eq.-6; **(f)** Merging the segments from $\mathbf{t}^3$s and the unchanged spectra portions from $\mathbf{t}^1$ to get continuum removed spectra by SCF method. The continuum removed spectra by UCH, SUH and GH are also provided for visual comparison; **(g)** A visual comparison between the predicted continuums by the proposed method, UCH, SUH and GH.

Note: X-axis represents the wavelength domain. The domain is same for the full spectra in (a), (b), (f) and (g); and same for the spectra segments in (c), (d) and (e). Y-axis in (a) and (g) represents the reflectances in the input spectra, and the continuum-removed normalized spectra in the others.

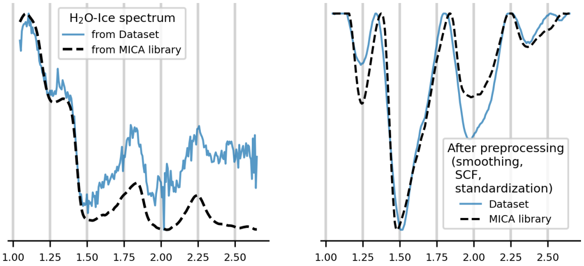


Figure 3. A preprocessing pipeline that includes SCF as the continuum removal operation can translate the two H_2O-Ice spectrum (from MICA spectral library and CRISM TRDR dataset) in the left image to the right image.

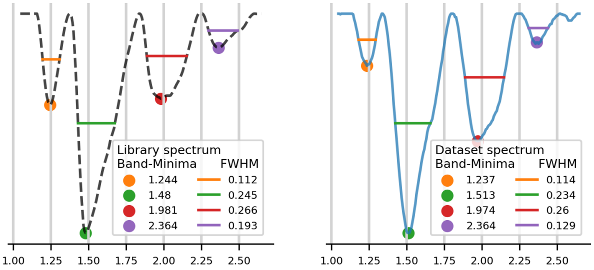


Figure 4. The band-minima and FWHM for all the prominent segments of the processed spectra given in figure 3

4.1 Experimental Setup

4.1.1 Data Specifications CRISM's Targeted Remote Sensing Data Record (TRDR) provides high spatial resolution spectral data of Mars' surface, which is available via NASA PDS (Justin N. Maki, 2004). Plebani et al in (Plebani et al., 2022) classified a large collection of 592,413 TRDR spectra from over 70 images into 39 categories, of which 28 labels correspond with those in the MICA library. For this evaluation, 200 spectra per label are randomly sampled from this dataset, focusing on the wavelength range 1-2.6 μm from the available 1-3.47 μm as most of the unique identifiable absorptions of the minerals appear in this range and beyond this range data are often very noisy.

The preprocessing pipeline used for CRISM MTRDR data (Kumari et al., 2023a) includes smoothing, spike removal, continuum removal, and standardization. Smoothing and spike removal are applied to reduce noise in the dataset, as for SCF, effective smoothing is essential to accurately fit parabolas through local maxima within a segment. Figure 3 illustrates the effect of the preprocessing steps applied on H_2O-Ice spectrum from the MICA library and TRDR dataset.

4.1.2 Identification Procedures For mineral identification, two distinct approaches are employed in this study: a CNN-based classification model and a spectral matching technique, the Weighted Sum of Segment Correlation (WSSC). The CNN model, adapted from (Kumari et al., 2023a), is a shallow sequential network that utilizes convolutional layers to learn spatial features from spectral data. It is trained and validated on labeled TRDR spectra and tested for accuracy on held-out samples, helping to assess the performance of the preprocessing pipeline and continuum removal methods. The WSSC method, a spectral matching approach, calculates a match index between test and library spectra by correlating prominent absorptions without requiring prior knowledge of the test spectrum. While similar to shape-matching algorithms such as those described by (Clark et al., 2003), WSSC focuses on the segmented absorption features, providing an adaptable and robust alternative to standard methods in hyperspectral analysis. By applying both CNN and WSSC, the study evaluates SCF's performance across machine learning and classical matching techniques, offering insights into the effectiveness of each approach for hyperspectral mineral identification.

4.1.3 Evaluation Measures A band-minima refers to the position of minimum reflectance value at a spectral region, generally an absorption, in a non-translated spectrum, that is before the continuum removal operation on it. On the contrary, the band-center is the position of the minimum reflectance value in an absorption after the translation. There can be a minor shift between the band-minima and band-center positions for absorptions. The amalgamation of the positions of band-center in different absorptions or segments of a spectrum is a useful identifier for distinguishing materials as each of them has a unique set of positions and shapes of the absorptions over the wavelength domain. In spectroscopy, FWHM (Full Width at Half Maximum) is a measurement used to describe the shape of absorption in a translated spectrum, that measures the width of absorption in terms of the range of wavelengths at which the depth of the absorption is half of its maximum value. As the datasets in the experiments are either TRDR/MTRDR data, all characterized by a nearly constant spectral resolution (0.00655 μm) within the 1-2.6 μm wavelength range, the accuracies of absorption detection are assed using thresholds related to the spectral resolution. Specifically, a 3-band difference (approximately 0.02 μm) from the expected position is considered to evaluate the correct detection of band-center and a 6-band difference (approximately 0.06 μm) for FWHM. These thresholds are chosen through experiments to ensure they do not significantly impact mineral identification accuracy.

From both the preprocessed test spectrum $\mathbf{t}$ and the corresponding library spectrum $\mathbf{l}$, the absorptions are extracted first. Subsequently, tiny absorptions are filtered out using a threshold. The positions of the band-center and the FWHMs in the remaining absorptions are compared to assess their similarity.

Let $\mathcal{A}^{\mathbf{t}}$ and $\mathcal{A}^{\mathbf{l}}$ be the set of absorptions in $\mathbf{t}$ and $\mathbf{l}$ respectively. Let $BM(\mathcal{A}^{\mathbf{l}}_k)$ and $FWHM(\mathcal{A}^{\mathbf{l}}_k)$ represent the band-center position and FWHM respectively in the k-th absorption in $\mathbf{l}$. Similarly, $BM(\mathcal{A}^{\mathbf{t}}_k)$ and $FWHM(\mathcal{A}^{\mathbf{t}}_k)$ respectively are the band-center position and FWHM in the k-th absorption in $\mathbf{t}$. An example of this is shown in figure 4.

Through a one-to-one mapping based on the nearest band-center positions a bijection of the highest possible cardinality is established from $\mathcal{A}^{\mathbf{l}\prime}$ ($\subseteq \mathcal{A}^{\mathbf{l}}$) to $\mathcal{A}^{\mathbf{t}\prime}$ ($\subseteq \mathcal{A}^{\mathbf{t}}$). Let i-th absorption in $\mathcal{A}^{\mathbf{l}}$ is mapped with b_i-th absorption in $\mathcal{A}^{\mathbf{t}}$. With this, the two measures band-center Score and FWHM Score are defined as below.

Band-center Score The mean-shift in the band-center position in $\mathbf{t}$ is defined as,

$$M_{BM}(\mathbf{t}) = \frac{\sum_{\forall \mathcal{A}^{\mathbf{l}}_i \in \mathcal{A}^{\mathbf{l}\prime}} |BM(\mathcal{A}^{\mathbf{l}}_i) - BM(\mathcal{A}^{\mathbf{t}}_{b_i})|}{|\mathcal{A}^{\mathbf{l}\prime}|} \tag{8}$$

$|.|$ in the numerator denotes the absolute difference and in the denominator denotes the cardinality.

For a set of test spectra $\mathcal{T}$ the band-center score is obtained by the fraction of the set for which the mean-shift in the positions of band-center is lesser than 0.02 μm. Formally,

$$\frac{count(M_{BM}(\mathbf{t}) \leq 0.02),\ \forall \mathbf{t} \in \mathcal{T}}{|\mathcal{T}|} \tag{9}$$

FWHM Score Similarly, the mean-change in the FWHM in $\mathbf{t}$ is defined as,

$$M_{FWHM}(\mathbf{t}) = \frac{\sum_{\forall \mathcal{A}_i^{1} \in \mathcal{A}^{1\prime}} |FWHM(\mathcal{A}_i^{1}) - FWHM(\mathcal{A}_{b_i}^{\mathbf{t}})|}{|\mathcal{A}^{1\prime}|} \tag{10}$$

and FWHM score of $\mathcal{T}$ is obtained by,

$$\frac{count(M_{FWHM}(\mathbf{t}) \leq 0.06),\ \forall \mathbf{t} \in \mathcal{T}}{|\mathcal{T}|} \tag{11}$$

Identification Score Identification accuracy measures the percentage of correctly identified spectra, calculated using the Weighted Sum of Segment Correlation (WSSC) method. Effective preprocessing, particularly smoothing, enhances the performance of WSSC by ensuring high correlation between segment features in test and library spectra. The identification score is the percentage of test spectra for which the true-label class receives a match index within 5% of the highest score assigned to any class.

4.2 Performance Analysis

The Upper Convex Hull and the Segmented Upper Hull, as defined in the literature, do not have the provision of customizability. In contrast, the Geometric Hull, although presented in the literature to consider all local minima in an intermediate step, tends to overfit a real input spectrum if all local minima are considered due to its iterative nature and presence of fluctuations in a real spectrum, resulting in a limited ability to detect absorption features, particularly for the minerals having distinguishable broad absorptions. Lastly, SCF, described in Section 3 as a non-parametric approach, exhibits various forms based on factors such as whether selective or all local minima are used to fit the curve, the minimum required number of local maxima in a segment for consideration in further processing, whether a recursive approach is employed to detect more shoulder points, and so on. To prevent overfitting in GH, only the local minima beyond a specified threshold depth from its shoulder points are taken into consideration. SCF also, in the crude form does not perform optimally when dealing with minerals that exhibit a broad absorption feature like mafic minerals and halides. For such minerals specifically, SCF considers a segment for further division only if it contains at least 3 local minima.

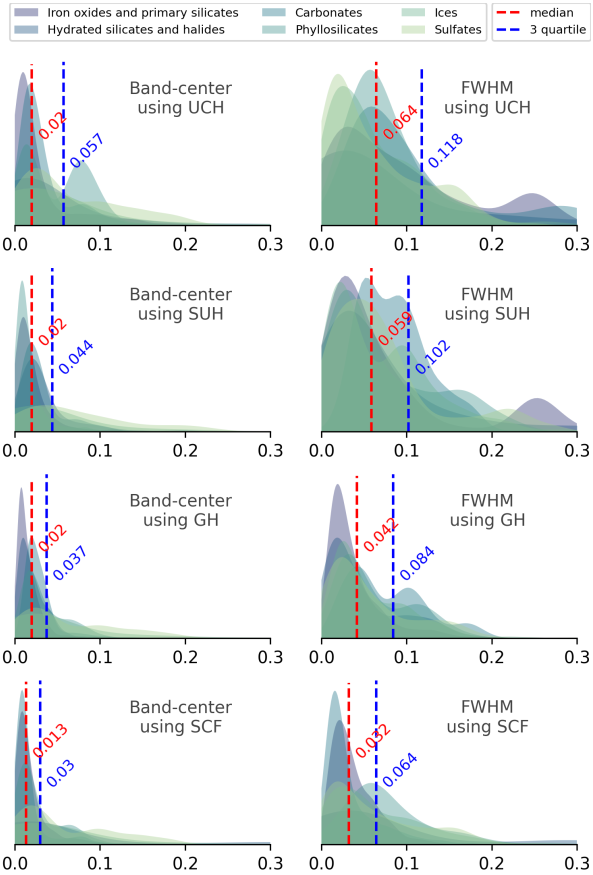


Figure 5. The density plots of the differences in band-center and FWHM values between the processed spectra of dataset-4 and the corresponding spectra from the MICA library, using UCH, SUH, GH and SCF in preprocessing. The plots are colour-coded by the mineral groups. The lesser quartile-2 (median) and quartile-3 values indicate that the band-center and FWHM values are more accurate when SCF is applied.

Table 1 shows the evaluation metric scores for each mineral group separately. The experiments have indicated that, for each method, while there may be some mineral 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 with SCF to detect the absorption can detect mineral groups with a high degree of accuracy, even if the detection of individual minerals may not be as precise. 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 spectral segments, termed as broad-bands, or sharp absorptions within a broad band in the spectrum, such as silicates and carbonates. 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, SUH, and GH. However, for

Table 1. Evaluation measures (in %) using TRDR dataset

Mineral classes	Band-center Score				FWHM Score				Identification Score (WSSC)				Identification Score (CNN)			
	UCH	SUH	GH	SCF	UCH	SUH	GH	SCF	UCH	SUH	GH	SCF	UCH	SUH	GH	SCF
Primary silicates	72.5	54.0	**78.8**	78.4	64.0	61.2	**79.9**	71.0	77.8	59.7	**81.8**	79.5	86.0	86.5	**87.8**	86.9
Hydrated silicates	85.3	66.3	66.3	**96.0**	65.3	68.8	67.8	**81.3**	78.2	72.9	65.3	**86.7**	89.1	89.3	88.3	**91.3**
Phyllosilicates	78.0	58.8	76.8	**89.6**	67.2	80.3	78.8	**88.8**	84.5	65.8	81.7	**88.0**	89.3	90.8	86.8	**92.7**
Sulfates	77.0	58.6	77.6	**80.5**	63.1	78.8	77.8	**86.1**	79.0	72.2	84.0	**87.0**	**91.6**	90.8	90.8	90.1
Carbonates	52.3	**67.0**	61.1	65.5	**73.0**	**73.0**	60.0	71.0	62.8	**77.0**	61.7	75.0	78.0	79.8	78.8	**81.0**
Ices	78.0	65.7	75.1	**86.0**	73.0	80.0	77.7	**78.0**	97.0	91.3	92.7	**98.0**	**91.5**	89.8	89.6	89.7
Overall	74.2	60.0	68.8	**82.0**	65.9	73.4	76.6	**82.7**	77.9	72.8	76.8	**86.5**	90.1	88.8	88.5	**91.5**

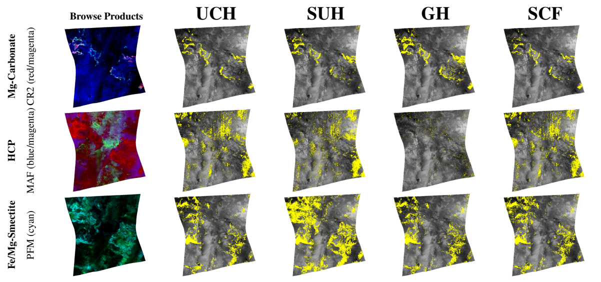
most of the mineral groups, the SCF method yields higher accuracy scores, resulting in an overall improvement in the performance of around 1.5%, where as by the described method WSSC, where local shape-matching is considered, the betterment is more than 8%. Figure 5 represents the density plots of the calculated mean shift in band-center and mean change in FWHM for the spectra in dataset-4. It is evident from the figure that SCF detects the band-centers of the absorptions more accurately and preserves the shape of the absorptions (FWHMs) more precisely similar to the library than the compared methods UCH, SUH, and GH. The following is the summary of the experiments:

- SCF relies on local spectral shapes, and exhibits enhanced noise resilience in terms of preserving more precise spectral information like band-center and FWHM of the absorption features.
- There is a strong correlation between the preservation of band-center and FWHM and the mineral identification accuracy. Preprocessing the spectra with SCF continuum removal yielded satisfactory accuracy gains over the other methods for mineral identification using the spectra matching technique WSSC and also using a shallow CNN model. Performance gains were consistent across the mineral groups.
- While WSSC with SCF may encounter challenges in accurately identifying minerals with broad or sharp absorptions within broad bands, the methods generally maintain accuracy in identifying mineral groups, even if some mineral labels are detected incorrectly.

5. Application on Hyperspectral Images

This section presents the results of mineral identification on CRISM MTRDR data of the Martian surface, achieved by employing the WSSC method to match the characteristic absorption feature positions and shapes with spectra from the MICA library. The SCF method was used as a preprocessing step to remove the continuum and scale the spectra. The CRISM MTRDR data are publicly available via NASA's Planetary Data System (PDS) (Justin N. Maki, 2004). In this study, two CRISM MTRDR images, FRT93BE (latitude: 19.1°N, longitude: 283.5°W) from Jezero Crater and FRT13D1F (latitude: 28.9°S, longitude: 166.7°W) from Columbus Crater, were used to assess the applicability of SCF for mineral identification across distinct Martian regions. Jezero Crater, located on the western edge of Isidis Planitia in Mars' northern hemisphere, and Columbus Crater, a 119-kilometer-wide feature in Terra Sirenum in the southern hemisphere, are of scientific interest due to the presence of clay minerals, suggesting historical liquid water (Payne et al., 2011, Wray et al., 2011). Jezero Crater's mineralogy is dominated by olivine, pyroxene, and plagioclase feldspar, while Columbus Crater predominantly contains pyroxenes with traces of gypsum, a rare Martian mineral. FRT93BE contains 804×832 pixels, and FRT13D1F has 800×648 pixels, each with 489 bands covering a wavelength range of 0.43–3.89 μm. However, consistent with other CRISM data used in section 4, only the 1–2.6 μm wavelength range was analyzed in this study.

FRT93BE (from Jezero Crater region of Mars)



FRT13D1F (from Columbus Crater region of Mars)

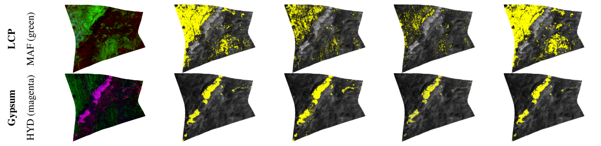


Figure 6. Detected dominant minerals in CRISM MTRDR data FRT93BE and FRT13D1F. For each mineral, the pixels with a match index of more than 0.7 with the corresponding library spectrum are highlighted. The first column contains the browse products with colours indicated for visual verification.

The primary minerals identified in FRT93BE include Mg-Carbonate, High-Calcium Pyroxene (HCP), and Mg-Smectite. Mg-Carbonate exhibits spectral features at 2.2–2.4 μm and 2.33 μm due to carbonate ions and hydroxyl groups, respectively. HCP has a strong absorption at 0.9 μm, a calcium-related feature at 1.2 μm, and a shoulder at 1.6 μm. Mg-Smectite shows features at 1.8–1.9 μm from OH groups and metal-oxygen bonds, with a distinctive feature at 2.2–2.35 μm from OH and Mg-O vibrations. Mg-Smectites in FRT93BE likely coexist with carbonates in significant abundance. In FRT13D1F, Low-Calcium Pyroxene (LCP) and gypsum are the dominant minerals. LCP exhibits a broad absorption around 1.8 μm, and gypsum reveals distinct absorptions between 1.4–1.5 μm and a sulfate group feature at 2.1 μm, resulting from water molecule oscillations.

Figure 6 displays the mineral distributions in both images as determined by the continuum removal and WSSC framework discussed in section 4.1.3. To evaluate SCF's accuracy in detecting absorption features and matching them with library spectra, results are compared with those obtained using other continuum removal methods, such as UCH, SUH, and GH. Pixels matching the library spectra with indices above 0.7 are highlighted in figure 6. This approach enables soft classification, allowing pixels to associate with multiple mineral classes, which is suitable for hyperspectral data like CRISM where pixel spectra often exhibit mixtures. As shown, SUH and GH methods are limited in detecting minerals with broad absorption features, while UCH and SCF consistently align most closely with the mineral distributions in browse products across all detected minerals.

6. Discussions and Conclusion

This paper presents a novel apparent continuum removal method called Segmented Curve-Fitting (SCF), specifically designed to address non-convex continuum shapes in spectra in linear time. SCF effectively extracts a greater number of shoulder points from spectra, resulting in a closer alignment between the extracted absorptions and those in the corresponding library spectra compared to traditional methods such as UCH and SUH.

To evaluate SCF's performance, we measured the mean shift in band-center positions and the differences in full-width at half-maximum (FWHM) of extracted absorption features using the TRDR dataset. Experiments on real CRISM TRDR data demonstrated that mineral identification using WSSC with SCF-based continuum removal improved accuracy by approximately 8% over the UCH method and about 1.5% when employing a CNN model. However, the analysis of mineral group-wise accuracies suggested that SCF may be less effective at distinguishing prominent broad-band absorption signatures or sharp absorptions within broader bands. Overall, the SCF method excels in detecting narrow to medium absorptions from spectra exhibiting curvature distortion, outperforming existing methods.

Future enhancements to the SCF method could focus on the following areas:

- **Enhancing adaptability and precision:** Currently, SCF utilizes all local maxima within a spectral region for parabolic curve fitting. Future research could investigate selective criteria for choosing these maxima during the fitting process.
- **Broad absorption detection:** The method could be improved to automatically identify broad absorptions, which are not currently addressed, thereby expanding SCF's capability to detect a wider range of spectral features.
- **Spectral unmixing applications:** Given SCF's ability to capture distinct spectral features, there is potential for its application in spectral unmixing tasks.

References

Brown, A., 2006. Spectral curve fitting for automatic hyperspectral data analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6), 1601-1608.

Chabrillat, S., Goetz, A. F., Krosley, L., Olsen, H. W., 2002. Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution. *Remote sensing of Environment*, 82(2-3), 431–445.

Clark, R. N., King, T. V., 1987. Automatic continuum analysis of reflectance spectra. *JPL proceedings of the 3rd airborne imaging spectrometer data analysis workshop*.

Clark, R. N., Lucey, P. G., 1984. Spectral properties of ice-particulate mixtures and implications for remote sensing: 1. Intimate mixtures. *Journal of Geophysical Research: Solid Earth*, 89(B7), 6341–6348.

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

Clark, R. N., Swayze, G. A., Livo, K. E., Kokaly, R. F., Sutley, S. J., Dalton, J. B., McDougal, R. R., Gent, C. A., 2003. Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems. *Journal of Geophysical Research: Planets*, 108(E12).

Cormen, T. H., Leiserson, C. E., Rivest, R. L., Stein, C., 2022. *Introduction to algorithms*. MIT press.

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

Kumari, P., Soor, S., Shetty, A., Koolagudi, S. G., 2023a. A Fully-Automated Framework for Mineral Identification on Martian Surface Using Supervised Learning Models. *IEEE Access*, 11, 13121-13137.

Kumari, P., Soor, S., Shetty, A., Koolagudi, S. G., 2023b. Mineral classification on Martian surface using CRISM hyperspectral data: a survey. *Journal of Applied Remote Sensing*, 17(4), 041501–041501.

Kumari, P., Soor, S., Shetty, A., Koolagudi, S. G., 2024. MICAnet: A Deep Convolutional Neural Network for mineral identification on Martian surface. *The Egyptian Journal of Remote Sensing and Space Sciences*, 27(3), 501–507.

McCraig, M. A., Osinski, G. R., Cloutis, E. A., Flemming, R. L., Izawa, M. R., Reddy, V., Fieber-Beyer, S. K., Pompilio, L., van der Meer, F., Berger, J. A. et al., 2017. Fitting the curve in Excel®: Systematic curve fitting of laboratory and remotely sensed planetary spectra. *Computers & geosciences*, 100, 103–114.

Mielke, C., Boesche, N. K., Rogass, C., Kaufmann, H., Gauert, C., 2015. New geometric hull continuum removal algorithm for automatic absorption band detection from spectroscopic data. *Remote Sensing Letters*, 6(2), 97–105.

Parente, M., Makarewicz, H. D., Bishop, J. L., 2011. Decomposition of mineral absorption bands using nonlinear least squares curve fitting: Application to Martian meteorites and CRISM data. *Planetary and Space Science*, 59(5-6), 423–442.

Payne, V. H., Mlawer, E. J., Cady-Pereira, K. E., Moncet, J.-L., 2011. Water Vapor Continuum Absorption in the Microwave. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6), 2194-2208.

Plebani, E., Ehlmann, B. L., Leask, E. K., Fox, V. K., Dundar, M. M., 2022. A machine learning toolkit for CRISM image analysis. *Icarus*, 376, 114849.

Viviano, C. E., Seelos, F. P., Murchie, S. L., Kahn, E. G., Seelos, K. D., Taylor, H. W., Taylor, K., Ehlmann, B. L., Wiseman, S. M., Mustard, J. F., Morgan, M. F., 2014. Revised CRISM spectral parameters and summary products based on the currently detected mineral diversity on Mars. *Journal of Geophysical Research: Planets*, 119(6), 1403–1431. https://doi.org/10.1002/2014je004627.

Wray, J., Milliken, R., Dundas, C. M., Swayze, G. A., Andrews-Hanna, J., Baldridge, A., Chojnacki, M., Bishop, J., Ehlmann, B., Murchie, S. L. et al., 2011. Columbus crater and other possible groundwater-fed paleolakes of Terra Sirenum, Mars. *Journal of Geophysical Research: Planets*, 116(E1).

Zhang, J., Rivard, B., Sanchez-Azofeifa, A., 2004. Derivative spectral unmixing of hyperspectral data applied to mixtures of lichen and rock. *IEEE Transactions on Geoscience and Remote Sensing*, 42(9), 1934–1940.

Zhao, H., Zhang, L., Zhang, X., Liu, J., Wu, T., Wang, S., 2015. Hyperspectral feature extraction based on the reference spectral background removal method. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6), 2832–2844.