This is the peer-reviewed, final accepted version for American Mineralogist, published by the Mineralogical Society of America.
The published version is subject to change. Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2023-9227. http://www.minsocam.org/

Revision 2

3759 words

Raman Match: Application for Automated Identification of Minerals from Raman Spectroscopy Data

Berrada, Meryem[1]*, McFall, Alan[1], and Chen, Bin[1]

[1]Hawai'i Institute of Geophysics and Planetology, University of Hawaii at Manoa.
2500 Campus Rd, Honolulu, HI 96822
*Corresponding author

Abstract

Raman spectroscopy is a rapid, nondestructive analysis technique that is used in various scientific disciplines, including chemistry, materials science, and biology. The analysis of Raman spectra and the identification of specific substances in unknown samples can be complex and time-consuming tasks due to the large database of Raman spectra. The *Raman Match* application was developed to simplify and automate the sample identification process through a search and match method. The application integrates the well-established RRUFF Raman database with the Python programming language and provides a user-friendly graphical interface to load Raman spectra, identify and fit peaks, match peaks to the reference libraries, visualize the results, and generate publication-ready figures. The application offers a swift and automated method for mineral identification using Raman spectroscopy in both laboratory and field settings, as well as during planetary exploration missions to extraterrestrial environments with constraints on time and resources.

Keywords: matching algorithm, peak identification, spectrum analysis, mineralogy, Raman database, python

Introduction

Raman spectroscopy, discovered experimentally by C. V. Raman in 1928 (Raman & Krishnan, 1928), is a powerful vibrational spectroscopy technique for mineralogical analysis commonly used in terrestrial laboratories. Vibrational spectroscopies, Raman and infrared, have the ability of positively identifying a chemical. In Earth and planetary science, Raman spectroscopy serves as an invaluable nondestructive tool for characterizing the chemical and mineralogical composition of planetary bodies remotely and for studying phase transitions in controlled laboratory environments (Angel et al., 2012; Beyssac, 2020; Caracas et al., 2023; Chou & Wang, 2017; Goncharov, 2012; Jehlička & Culka, 2022; Klein et al., 2004; Marshall et al., 2010; Mattioda et al., 2023; Qi et al., 2023; Sharma et al., 2003; Tu & Chang, 2012; Yamamoto & Hagiwara, 2022). This analytical tool is also widely used in other fields such as oncology and chemistry. (Das & Agrawal, 2011; Fenn et al., 2011; Jermyn et al., 2016; Shvalya et al., 2020). Raman spectroscopy yields insights into the chemical composition, crystal structure, and molecular vibrations of minerals, allowing for the detection and monitoring of phase transitions by analyzing shifts or alterations in Raman peaks (Caracas et al., 2023; Goncharov, 2012).

Raman spectroscopy entails the examination of vibrational and rotational modes of molecules present in a specimen. When photons interact with a molecule, two distinct scattering processes occur: Rayleigh scattering, which preserves the initial photon

characteristics, and Raman scattering, causing changes in photon properties due to altered energy levels (Cialla-May et al., 2019). These energy shifts are reflected as alterations in wavelength (λ), inversely related to energy, and are typically represented as wavenumber ($\sigma = 1/\lambda$) in Raman spectra. This visualization facilitates an understanding of the relationship between energy and wavelength, driven by Planck's constant and the speed of light. By measuring $\Delta\sigma$, Raman spectrometers reveal the wavelength of inelastically scattered photons, unraveling vibrational modes, chemical composition, and crystal structures within the sample (Pelletier, 2003). The vibrational modes of a molecule depend on the molecular structure, masses of atoms involved in the vibration, and the chemical bonds between various atoms (Ferraro et al., 2003). This results in a unique spectrum for each chemical, which can distinguish polymorphs of isochemical materials. Raman analysis has several advantages over passive spectroscopy, most notable being the sharpness of the spectral features. This allows the precise detection of specific species, especially in the presence of mineral mixtures. The Raman spectra of samples contain a wealth of molecular fingerprint information that can be used to identify inorganic and organic chemicals, biomarkers, biominerals, hydrous and anhydrous minerals, and new chemical phases based on the vibrational frequencies, relative intensities, and number of Raman lines in the spectra (Degen & Newman, 1993; Durben et al., 1991; Kawamoto et al., 2004; Wang et al., 1995). It can also be used to characterize polymorphs of minerals and organics (Gauldie et al., 1997; Girlando et al., 2016; Marckmann & Whalley, 2004).

Effective preparation of samples is an essential aspect of Raman spectroscopy measurements, and considerations must be made regarding the nature and physical state of the sample. Solid samples may require grinding to obtain representative portions,

while geological or mineralogical specimens might require thin sections or polished surfaces. Liquids can be directly analyzed, but container material selection is essential to prevent interference. The handling of gaseous samples can be complex, often requiring specialized cells. To mitigate the fluorescence that can obscure Raman signals, careful attention must be paid, although reducing laser power and employing suitable filters usually help to resolve this problem (Wei et al., 2015). Additionally, it is worth noting that a time-gated system can be used to further address fluorescence-related issues and enhance the quality of Raman spectroscopy data. Time-gating allows for the selective detection of Raman signals at specific time intervals, effectively separating them from background fluorescence, and thus improving the accuracy and reliability of the measurements. This approach demonstrated its value in samples susceptible to in fluorescence, making it a potent method in Raman spectroscopy (Misra et al., 2005).

Within a Raman spectrum, peaks correspond to specific vibrational modes of molecules, with their characteristics extending beyond the molecular vibrational modes. The features of the peaks, including the full-width at half-maximum (FWHM), conveys valuable information (Conner et al., 2023; Hong et al., 2022; Obraztsova et al., 1998; Schiferl et al., 1997). A narrow FWHM indicates a well-defined and distinct peak, implying a relatively pure molecular vibration unaffected by external factors. Conversely, a broad FWHM suggests influences from multiple factors, such as molecular interactions, impurities, or variations in molecular environments. The vibrational modes present in Raman spectra reflect not only the molecular vibrations within the mineral or material but also the broader structural components and arrangements inherent to the crystal lattice. Minerals possess complex and unique

crystalline structures composed of atoms or ions arranged in specific patterns and orientations. These crystallographic arrangements lead to characteristic lattice dynamics, influencing the propagation of vibrations throughout the mineral. As vibrations travel through the crystal lattice, they interact with the structural components, lattice defects, and boundaries, inducing distinct spectral features. The interplay among these factors dictates the observed peaks in Raman spectra. For instance, variations in lattice parameters, crystal symmetry, or the presence of defects within the mineral can results in shifts, splitting, or broadening of Raman peaks, providing valuable insights into the structural integrity and properties of the mineral.

Analyzing peak properties employing various models (e.g., Gaussian, Lorentzian, Voight, Pseudo-Voight) and accounting for baseline correction aids in comparing spectra with reference Raman spectra (Ciris, 2023; Ida et al., 2000; Sparavigna, 2023; Young & Wiles, 1982). In essence, the Gaussian distribution (**Eq. 1**), often referred to as the bell curve, takes parameters such as amplitude (A), mean (μ), and standard deviation (σ) to represent data that follows a normal distribution. The Gaussian function is widely employed because of its bell-shaped curve, which resembles many natural distributions. It is particularly useful for modeling symmetric peaks and is computationally efficient. Gaussian functions help to describe peaks characterized by Gaussian broadening, representing well-defined and symmetrical peaks.

$$f(x)_{Gaussian} = A \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

The Lorentzian distribution (**Eq. 2**), also known as the Cauchy distribution, is characterized by its sharp peak and long tails. Relevant parameters include amplitude, mean, and gamma (γ), which controls the width and shape of the distribution.

Lorentzian functions are employed to describing peaks with tails that extend further from the central peak, making them suitable for modeling asymmetric peaks and handling broader line shapes. These functions are beneficial for capturing spectral lines affected by Lorentzian broadening, prevalent in various spectroscopic techniques.

$$f(x)_{Lorentzian} = \frac{A \cdot \gamma^2}{(x-\mu)^2 + \gamma^2} \quad (2)$$

The Voigt function (**Eq. 3**) combines the characteristics of both Gaussian and Lorentzian distributions. It is often used to fit complex spectral data especially in spectroscopy, and considers the amplitude, mean, standard deviation, and gamma via the Faddeeva function wofz(z), where z is $\frac{x-\mu+i\gamma}{\sigma\sqrt{2}}$. Voigt functions are particularly useful in scenarios where a peak exhibits a combination of both broadening types.

$$f(x)_{Voigt} = Re\left[\frac{A \cdot wofz(z)}{\sigma\sqrt{2\pi}}\right] \quad (3)$$

Finally, the Pseudo-Voigt function (**Eq. 4**) is a mixture model that blends both Gaussian and Lorentzian profiles, where the fraction parameter (η) determines the contribution of each model.

$$f(x)_{Pseudo-Voigt} = \eta \cdot f(x)_{Lorentzian} + (1-\eta) \cdot f(x)_{Gaussian} \quad (4)$$

A fraction of 0 corresponds to a pure Gaussian, while a fraction of 1 corresponds to a pure Lorentzian. The weighted combination allows for flexibility in adjusting the shape of the fitted peak, providing a more adaptable model for a broad range of line shapes encountered in experimental data. It offers a good approximation to complex spectral lines and is widely used for peak fitting due to its flexibility and ability to describe a variety of peak shapes.

The baseline correction in the peak fitting process follows the Asymmetrically Reweighted Penalized Least Squares (arPLS), see the Python function arpls(y) in **Eq.5** (Baek et al., 2015).

$$\text{arPLS}(y) = \text{argmin}_z\{\|W \cdot (y - z)\|_2^2 + \lambda\|D \cdot z\|_2^2\} \quad (5)$$

where $\boldsymbol{W}$ is a diagonal matrix with elements *w,* representing the weight of each data point signal in *y*. *D* is a finite difference matrix that approximates the second derivative, and $\boldsymbol{\lambda}$ is a tuning parameter. The algorithm iteratively updates the weights *w* using a sigmoidal weighting scheme until convergence, where argmin$_z$ denotes the argument that minimizes the expression. The sigmoidal weighting scheme is defined as $\boldsymbol{w_i} = \frac{\mathbf{1}}{\mathbf{1}+\mathbf{exp}\left(\mathbf{2}*\frac{\boldsymbol{d_i}-(\mathbf{2}\boldsymbol{s}-\boldsymbol{m})}{\boldsymbol{s}}\right)}$, for each data point *i,* using the standard deviation (*s*), the negative residuals (*d*), and their mean (*m*). This weighting scheme assigns smaller weights to data points with larger negative residuals, effectively downweighting outliers. The output of the arPLS function is *z*, representing the baseline correction signal. The method takes three parameters as input: lam, ratio, and itermax. The first parameter controls the regularization strength. It is a hyperparameter that balances the smoothness of the baseline estimate. The second parameter determines the convergence criterion. If the change in the weighting factor between iterations is below this ratio, the algorithm stops. The third parameter specifies the maximum number of iterations. The user may modify the default parameters of the background subtract, or baseline correction, along with the default parameters for the peak fitting models.

Reference Raman spectra are consolidated within the well-established RRUFF™ Project database (Lafuente et al., 2015), encompassing high-quality spectral data from well-characterized minerals. These serve as standards for mineralogists,

geoscientists, and gemologists, facilitating mineral identification on Earth and aiding planetary exploration efforts. However, interpreting Raman spectra and extracting meaningful information can be challenging due to the complexity of the data and the various databases available. In addition, visualization tools available in the literature are typically limited to the associated spectrometers. For example, Horiba LabSpec is a widely used Raman analysis software that accompanies Horiba Raman spectrometers (HORIBA Scientific, n.d.). It offers comprehensive tools for data acquisition, processing, and visualization, although there is no matching process available. Similarly, Thermo Fisher Scientific Raman spectrometers are accompanied by the OMNIC software (Thermo Fisher Scientific Raman, n.d.). The software provides various features for spectral data processing, including spectral subtraction, spectral search from a limited database, and quantitative analysis. Another common software is OPUS, used with Bruker Raman spectrometers (Bruker, n.d.). Similar to the previously mentioned software, users can apply different data treatments, such as a baseline correction, and perform spectral fitting. Other software such as WiRE, Renishaw's Raman spectroscopy software, enables data acquisition, processing, and analysis (Renishaw, n.d.). It includes features such as baseline correction, peak identification, and chemometric analysis. Finally, the software called CrystalSleuth leverages the extensive RRUFF database to provide a list of potential mineral matches based on the full Raman spectrum analysis of a sample. The software displays the top three matches, ranking them according to their percentage match with the sample spectrum, thereby offering a comparative view of the closest mineralogical matches. This method primarily focuses on the overall spectral pattern, considering the collective features of the spectrum for comparison and match ranking. To our knowledge, only the Renishaw

Empty Modelling tools and CrystalSleuth (Laetsch & Downs, 2006) integrate the RRUFF database, while the other software programs have unique sample mapping capabilities. To address the challenges in analysis and to account for home-built Raman spectroscopy systems which do not work well with commercial software programs, the *Raman Match* application developed in this study provides a comprehensive set of tools which include baseline correction, peak fitting, and peak matching for Raman spectroscopy analysis, along with automated mineral identification capability from Raman data. The *Raman Match* application adopts a distinct approach from previous software tools by examining individual peaks within the Raman spectrum. This is beneficial for samples with complex compositions or composite minerals. It can be used to identify potential matches by analyzing individual peaks and synthesize the results to determine the top three most probable mineral matches. This approach improves the ability to discern and identify a wider range of mineral compositions from a given spectrum, offering a more nuanced analysis, especially for samples with mixed mineralogical content. *Raman Match* represents a departure from traditional methodologies used in mineral identification via Raman spectroscopy, which heavily rely on manual and time-consuming literature reviews of expected minerals.

Method and Approach

Raman Match is a Python-based application which transforms complex spectroscopic data into actionable insights. Users are provided with instructions in the Instruction tab, tools to manually identify the experimental data in the “Manual” tab, and tools to automatically match with reference Raman spectra in the “Auto” tab. The application enables users to identify substances, understand molecular structures, and

make informed decisions based on their research findings. The architecture of the application is summarized in the Supplementary Materials. *Raman Match* is designed with a user-friendly graphical interface to facilitate easy data analysis. It is built on the foundation of several key components including user-friendly graphical interface, customization options, data import and export, data processing and analysis tools, automation, compatibility, documentation and support, version control and updates, collaboration and sharing, and speed, which work harmoniously to provide users with a seamless experience. Users can load Raman spectroscopy data files (i.e., .txt or .csv files), typically containing information about the intensity of Raman scattering at various wavelengths. Once the data file is loaded, the application employs various mathematical and analytical techniques to process it. For example, it can identify the peaks in the data using an intensity threshold value and the standard deviation on peak positions. In the Manual tab (see **Figure 1**), users may use any of the fitting models and corresponding parameters to analyze the peaks identified, after accounting for the baseline correction if necessary. The best peak fitting parameters, along with the FWHM method results, are dynamically added to the table in the Manual tab. The user may save the results with the "Save Table" button. One of the primary functions of the application is to match the obtained peaks with known molecular signatures stored in a Structured Query Language (SQL) for Light database. This database, based on the RRUFF project database, acts as a vast library of known molecular signatures, enabling the application to make rapid and accurate identifications. In the "Manual" tab, the user may select files from the database via the search bar to be displayed in the bottom figure for reference. In the Auto tab (see **Figure 2**), the application looks for similarities between individual peak positions in the experimental data and the RRUFF database.

The user may guide the accuracy of the matching process by specifying keywords (e.g., Si, O, Quartz) to look for or to avoid in the database. Following **Eq.6**, the matching algorithm loops through the entries of the database looking for elements to "Look for" or "Avoid" according to the user entries. The first condition of the loop identifies the minerals for which there is a match with the peak position. The second condition verifies if the identified mineral remains an option after keeping any of the elements to look for but removing all those to avoid.

$$
\begin{aligned}
&For\ any\ peak\ in\ data\text{:}\\
&conditon_1\text{:}\begin{cases} True & if\ peak_{data} = peak_{mineral\ in\ database} \\ False & if\ peak_{data} \neq peak_{mineral\ in\ database} \end{cases}\\
&condition_2\text{:}\begin{cases} True & if\ wanted \backslash avoid\ \neq \emptyset \\ False & if\ wanted \backslash avoid = \emptyset \end{cases}\\
&mineral\ count = \begin{cases} total + 1 & if\ condition_1 \wedge condition_2 \\ total & otherwise \end{cases}
\end{aligned}
\tag{6}
$$

The insights derived from the data analysis are presented in a clear and visual manner: The bottom figure in the "Auto" tab displays the top three files with the highest matching score across the complete experimental data, while the right figure displays the top three files with the highest matching score at each identified peak. This interactivity empowers users to tailor the analysis to their specific needs (i.e., via the peak dentification threshold, standard deviation, peaks to be removed, and elements to look for or avoid). The application allows users to save and manage their analysis results, ensuring that valuable findings are not lost.

Results and Discussion

Ensuring the reliability and accuracy of the automated mineral identification process of the application is essential for producing credible and reproducible research

results. *Raman Match* employs several quality control measures to validate and enhance its performance. One of the key quality control aspects is the accurate identification of Raman peaks in the experimental data. The software uses peak-finding algorithms to locate spectral features. Users may modify the peak identification intensity threshold to ensure that the algorithms reliably detect peaks corresponding to mineral vibrations, rather than noise in the experimental data. Unknown peaks may also be ignored by interacting with the peak removal feature in both the Manual and Auto tabs. To account for variations in data acquisition, the software incorporates spectrum normalization and baseline subtraction techniques. These techniques ensure that the identified peaks are not influenced by variations in signal intensity, laser power, or other experimental factors.

Quality control involves testing the software with a diverse set of Raman spectra from known mineral samples. A known mineral sample, unoriented forsterite crystal, is included in the package of the software. The experimental data for this crystal is automatically matched correctly to forsterite files from the RRUFF database in a few seconds, without the guidance from the user. The validation process confirms the software's ability to correctly identify minerals. The software is also tested on a sample of Hawaiian vesicular basalt collected on the island of Oahu (**Figure 3**). Basalt is a volcanic rock primarily composed of plagioclase feldspar, pyroxene, olivine, and sometimes magnetite. The automated analysis of the sample suggests a strong presence of forsterite and possibly inderite and scolecite. Olivine minerals such as forsterite, and zeolite minerals like scolecite, which can form in vesicles within basalt, are valid results for the sample. To interpret the matching results, it is important to take into the laser beam size, which is controlled by the objective lens magnification, and the measured

grain size. While the peak at 422 cm^{-1} matches well with many minerals, forsterite is not listed among the top 3 matches. However, the *Raman Match* allows a closer examination of the Forsterite reference data, and the output results spreadsheet suggests that forsterite could be a better match, as the sample spectrum also matches other characteristic peaks of forsterite. Similarly, the peak at 605 cm^{-1} is labelled as Titanite, but the bar graphs also show a match to Scolecite as well, which in more appropriate considering the common minerals found in Hawai'i (De Carlo & Swoboda, 2017). The other minerals are not expected in pure basalt and may be traces of the surrounding laboratory materials at the University of Hawai'i. In **Figures 4 & 5**, examples of guided and unguided matching are showcased for unoriented crystals of calcite and quartz, respectively. **Figure 4** shows accurate matching of the calcite reference files with the two selected peaks. **Figure 5** shows that quartz is 100% match for each of the selected peaks, as it was included in the user's input of elements to "Look for".

Users can easily review the identified minerals, peak locations, and associated information, which allows verification of the software's performance. The software also includes error handling mechanisms to deal with unexpected scenarios and if any issues or anomalies are encountered during the analysis, the software generates informative error messages. This feature ensures that users are alerted to potential problems, such as incorrect file format loaded or no available matches. In case errors occur during the matching process, the process is terminated, and the "Match" button becomes available to use. Finally, the software is accompanied by detailed documentation, in the Instruction tab and in the application package, that explains its functionality, usage, and potential troubleshooting steps. Users can reach out to the corresponding author for

support to address any questions or concerns. A flowchart summarizing the tools available for the user is illustrated in **Figure 6**.

Implications

The *Raman Match* application represents a significant advancement in the realm of Raman data analysis, offering a versatile open-source platform that streamlines the intricate tasks of data processing, peak identification, and mineral and chemical recognition. Based by the flexibility and robustness of Python and its associated libraries, this software introduces a user-friendly interface, ensuring that researchers can navigate and interpret Raman spectra with remarkable ease and efficiency. Beyond its intuitive interface, the application boasts interactive features, synchronization capabilities, and visually informative representations that not only simplify the analytical process but also enrich the user experience.

This software holds promise as a valuable tool beyond the laboratory setting, such as in space missions, where the instantaneous analysis of Raman data can significantly enhance mission efficiency. Raman spectroscopy may be utilized to identify mineral compositions on distant celestial bodies or to assess the composition of materials encountered during space exploration (Cho et al., 2021; Edwards et al., 2021; Hickman-Lewis et al., 2022; Rull et al., 2017, 2022). In such missions, time is often of the essence, and the ability to analyze Raman data rapidly and accurately can be critical. The compatibility of *Raman Match* with rover-based instrumentation or handheld Raman spectrometers, as a simple Python-based code, can empower astronauts or researchers on-site to rapidly analyze encountered materials and make informed decisions without the delay of sending samples back to Earth for analysis.

Raman Match's capacity to automate and expedite the analysis process positions it as an indispensable asset for real-time decision-making. In addition, *Raman Match* has the capability to catalyze breakthroughs in various scientific fields heavily reliant on home-built Raman spectroscopy systems, including materials science, chemistry, and pharmaceutical analysis. By simplifying the analysis workflow and offering automated chemical and mineral identification features, this software empowers researchers to delve deeper into the intricacies of their studies. For instance, *Raman Match*'s ability to automate analysis processes aligns with the increasing need for autonomous exploration tools in remote and challenging environments. Modifications and improvements are continuously monitored by the corresponding author, such as implementation of the recently published NASA Raman database (Mattioda et al., 2023). The source code for *Raman Match* is made available, and users are encouraged to implement their own preferences and contribute to the software's development.

Acknowledgements

M.B. would like to thank Jade Comellas and Ari Essunfeld, and Dr. Shiv K. Sharma, for their enriching discussions and valuable input regarding features and usage of the software. This work was supported by funds to M.B. from the Natural Sciences and Engineering Research Council of Canada (PDF – 567732 – 2022) and the Fonds de Recherche Nature et Technologies du Quebec (B3X 317379) and funds to B.C. from National Science Foundation (NSF) grants (EAR-2127807, EAR-1829273) , and NASA grant 80NSSC22K0138. The authors declare no competing interests. The manuscript was improved by comments from the Editorial team and reviewers.

RamanMatch.exe may be operated on Windows without additional system requirements, while the source code may be operated across platforms including macOS and Linux. The source code relies on various Python libraries and packages, including NumPy, SciPy, Matplotlib, Pandas, Tkinter, OpenPyXL, and Wofz. Users running the source code, instead of the executable file, should have these dependencies installed to run the software successfully. The source code programming language is Python, written in English, and published under the GNU GENERAL PUBLIC LICENSE by Meryem Berrada. All relevant documents are available from September 26th, 2023, on GitHub: https://github.com/meryemberradauwo/RamanMatch.

REFERENCES

Angel, S. M., Gomer, N. R., Sharma, S. K., & McKay, C. (2012). Remote Raman Spectroscopy for Planetary Exploration: A Review. *Applied Spectroscopy*, *66*(2), 137–150. https://doi.org/10.1366/11-06535

Baek, S.-J., Park, A., Ahn, Y.-J., & Choo, J. (2015). Baseline correction using asymmetrically reweighted penalized least squares smoothing. *Analyst*, *140*(1), 250–257. https://doi.org/10.1039/C4AN01061B

Beyssac, O. (2020). New Trends in Raman Spectroscopy: From High-Resolution Geochemistry to Planetary Exploration. *Elements*, *16*(2), 117–122. https://doi.org/10.2138/gselements.16.2.117

Bruker. (n.d.). OPUS - Vibrational Spectroscopy Software .

Caracas, R., Mohn, C., & Li, Z. (2023). Predicting HP-HT Earth and Planetary Materials. In L. Bindi & G. Cruciani (Eds.), *Celebrating the International Year of Mineralogy: Progress and Landmark Discoveries of the Last Decades* (pp. 131–151). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-28805-0_6

De Carlo, E. H., & Swoboda, B. (2017). Mineral Localities in Hawaii. *Rocks & Minerals*, *92*(3), 218–237. https://doi.org/10.1080/00357529.2017.1283658

Cho, Y., Böttger, U., Rull, F., Hübers, H.-W., Belenguer, T., Börner, A., et al. (2021). In situ science on Phobos with the Raman spectrometer for MMX (RAX): preliminary design and feasibility of Raman measurements. *Earth, Planets and Space*, *73*(1), 232. https://doi.org/10.1186/s40623-021-01496-z

Chou, I.-M., & Wang, A. (2017). Application of laser Raman micro-analyses to Earth and planetary materials. *Journal of Asian Earth Sciences*, *145*, 309–333. https://doi.org/https://doi.org/10.1016/j.jseaes.2017.06.032

Cialla-May, D., Schmitt, M., & Popp, J. (2019). Theoretical principles of Raman spectroscopy, *4*(6). https://doi.org/doi:10.1515/psr-2017-0040

Ciris, P. (2023). Information theoretic evaluation of Lorentzian, Gaussian, Voigt, and symmetric alpha-stable models of reversible transverse relaxation in cervical cancer in vivo at 3 T. *Magnetic Resonance Materials in Physics, Biology and Medicine*, *36*(1), 119–133. https://doi.org/10.1007/s10334-022-01035-1

Conner, K., Sharma, S., Uchiyama, R., Tanaka, K., Murakami-Sugihara, N., Shirai, K., & Kahng, S. (2023). Raman analysis of octocoral carbonate ion structural disorder along a natural depth gradient, Kona coast, Hawai‘i, *108*(5), 999–1013. https://doi.org/doi:10.2138/am-2022-8406

Das, R. S., & Agrawal, Y. K. (2011). Raman spectroscopy: Recent advancements, techniques and applications. *Vibrational Spectroscopy*, *57*(2), 163–176. https://doi.org/https://doi.org/10.1016/j.vibspec.2011.08.003

Degen, I. A., & Newman, G. A. (1993). Raman spectra of inorganic ions. *Spectrochimica Acta Part A: Molecular Spectroscopy*, *49*(5), 859–887. https://doi.org/https://doi.org/10.1016/0584-8539(93)80110-V

Durben, D. J., Wolf, G. H., & McMillan, P. F. (1991). Raman scattering study of the high-temperature vibrational properties and stability of CaGeO3

399 perovskite. *Physics and Chemistry of Minerals*, *18*(4), 215–223.
400 https://doi.org/10.1007/BF00202573/METRICS
401 Edwards, H. G. M., Jehlička, J., & Culka, A. (2021). Portable Raman
402 Spectroscopy in Field Geology and Astrobiology Applications. In *Portable*
403 *Spectroscopy and Spectrometry* (pp. 377–400).
404 https://doi.org/https://doi.org/10.1002/9781119636489.ch39
405 Fenn, M. B., Xanthopoulos, P., Pyrgiotakis, G., Grobmyer, S. R., Pardalos, P.
406 M., & Hench, L. L. (2011). Raman Spectroscopy for Clinical Oncology.
407 *Advances in Optical Technologies*, *2011*, 213783.
408 https://doi.org/10.1155/2011/213783
409 Ferraro, J. R., Nakamoto, K., & Brown, C. W. (2003). Introductory Raman
410 spectroscopy - Chapter 7, 434.
411 Gauldie, R. W., Sharma, S. K., & Volk, E. (1997). Micro-raman spectral study of
412 vaterite and aragonite otoliths of the coho salmon, Oncorhynchus kisutch.
413 *Comparative Biochemistry and Physiology Part A: Physiology*, *118*(3),
414 753–757. https://doi.org/10.1016/S0300-9629(97)00059-5
415 Girlando, A., Masino, M., Brillante, A., Toccoli, T., & Iannotta, S. (2016).
416 Raman Identification of Polymorphs in Pentacene Films. *Crystals 2016,*
417 *Vol. 6, Page 41*, *6*(4), 41. https://doi.org/10.3390/CRYST6040041
418 Goncharov, A. F. (2012). Raman Spectroscopy at High Pressures. *International*
419 *Journal of Spectroscopy*, *2012*, 617528.
420 https://doi.org/10.1155/2012/617528
421 Hickman-Lewis, K., Moore, K. R., Hollis, J. J. R., Tuite, M. L., Beegle, L. W.,
422 Bhartia, R., et al. (2022). In Situ Identification of Paleoarchean
423 Biosignatures Using Colocated Perseverance Rover Analyses: Perspectives
424 for In Situ Mars Science and Sample Return. *Astrobiology*, *22*(9), 1143–
425 1163. https://doi.org/10.1089/ast.2022.0018
426 Hong, M., Dai, L., Hu, H., Zhang, X., & Li, C. (2022). High-Temperature and
427 High-Pressure Phase Transition of Natural Barite Investigated by Raman
428 Spectroscopy and Electrical Conductivity. *Frontiers in Earth Science*, *10*.
429 Retrieved from
430 https://www.frontiersin.org/articles/10.3389/feart.2022.864183
431 HORIBA Scientific. (n.d.). HORIBA LabSpec 6 Spectroscopy Suite Software.
432 HORIBA France SAS.
433 Ida, T., Ando, M., & Toraya, H. (2000). Extended pseudo-Voigt function for
434 approximating the Voigt profile. *Journal of Applied Crystallography*, *33*(6),
435 1311–1316. https://doi.org/10.1107/S0021889800010219
436 Jehlička, J., & Culka, A. (2022). Critical evaluation of portable Raman
437 spectrometers: From rock outcrops and planetary analogs to cultural
438 heritage – A review. *Analytica Chimica Acta*, *1209*, 339027.
439 https://doi.org/https://doi.org/10.1016/j.aca.2021.339027
440 Jermyn, M., Desroches, J., Aubertin, K., St-Arnaud, K., Madore, W.-J., De
441 Montigny, E., et al. (2016). A review of Raman spectroscopy advances with
442 an emphasis on clinical translation challenges in oncology. *Physics in*
443 *Medicine & Biology*, *61*(23), R370. https://doi.org/10.1088/0031-
444 9155/61/23/R370

Kawamoto, T., Matsukage, K. N., Nagai, T., Nishimura, K., Mataki, T., Ochiai, S., & Taniguchi, T. (2004). Raman spectroscopy of cubic boron nitride under high temperature and pressure conditions: A new optical pressure marker. *Review of Scientific Instruments*, *75*(7), 2451–2454. https://doi.org/10.1063/1.1765756

Klein, V., Popp, J., Tarcea, N., Schmitt, M., Kiefer, W., Hofer, S., et al. (2004). Remote Raman spectroscopy as a prospective tool for planetary surfaces. *Journal of Raman Spectroscopy*, *35*(6), 433–440. https://doi.org/https://doi.org/10.1002/jrs.1168

Laetsch T, Downs R (2006) Software For Identification and Refinement of Cell Parameters From Powder Diffraction Data of Minerals Using the RRUFF Project and American Mineralogist Crystal Structure Databases. Abstracts from the 19th General Meeting of the International Mineralogical Association, Kobe, Japan.

Lafuente, B., Downs, R. T., Yang, H., & Stone, N. (2015). The power of databases: the RRUFF project. In T. Armbruster & R. M. Danisi (Eds.), *Highlights in Mineralogical Crystallography* (pp. 1–30). Berlin, Germany: W. De Gruyter.

Marckmann, J. P., & Whalley, E. (2004). Vibrational Spectra of the Ices. Raman Spectra of Ice VI and Ice VII. *The Journal of Chemical Physics*, *41*(5), 1450–1453. https://doi.org/10.1063/1.1726088

Marshall, C. P., Edwards, H. G. M., & Jehlicka, J. (2010). Understanding the Application of Raman Spectroscopy to the Detection of Traces of Life. *Astrobiology*, *10*(2), 229–243. https://doi.org/10.1089/ast.2009.0344

Mattioda, A. L., Gavilan, L., Ricketts, C. L., Najeeb, P. K., Ricca, A., & Boersma, C. (2023). The NASA Raman spectroscopic database: Ramdb version 1.00. *Icarus*, 115769. https://doi.org/https://doi.org/10.1016/j.icarus.2023.115769

Misra, A. K., Sharma, S. K., Chio, C. H., Lucey, P. G., & Lienert, B. (2005). Pulsed remote Raman system for daytime measurements of mineral spectra. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *61*(10), 2281–2287. https://doi.org/10.1016/J.SAA.2005.02.027

Obraztsova, E. D., Fujii, M., Hayashi, S., Kuznetsov, V. L., Butenko, Yu. V., & Chuvilin, A. L. (1998). Raman identification of onion-like carbon. *Carbon*, *36*(5), 821–826. https://doi.org/https://doi.org/10.1016/S0008-6223(98)00014-1

Pelletier, M. J. (2003). Quantitative Analysis Using Raman Spectrometry. *Applied Spectroscopy*, *57*(1), 20A-42A. https://doi.org/10.1366/000370203321165133

Qi, X., Ling, Z., Liu, P., Chen, J., Cao, H., Liu, C., et al. (2023). Quantitative Mineralogy of Planetary Silicate Ternary Mixtures Using Raman Spectroscopy. *Earth and Space Science*, *10*(5), e2023EA002825. https://doi.org/https://doi.org/10.1029/2023EA002825

Raman, C. V., & Krishnan, K. S. (1928). A New Type of Secondary Radiation. *Nature*, *121*(3048), 501–502. https://doi.org/10.1038/121501c0

Renishaw. (n.d.). Windows®-based Raman Environment (WiRE) software.

Rull, F., Maurice, S., Hutchinson, I., Moral, A., Perez, C., Diaz, C., et al. (2017). The Raman Laser Spectrometer for the ExoMars Rover Mission to Mars. *Astrobiology*, *17*(6–7), 627–654. https://doi.org/10.1089/ast.2016.1567

Rull, F., Veneranda, M., Manrique-Martinez, J. A., Sanz-Arranz, A., Saiz, J., Medina, J., et al. (2022). Spectroscopic study of terrestrial analogues to support rover missions to Mars – A Raman-centred review. *Analytica Chimica Acta*, *1209*, 339003. https://doi.org/https://doi.org/10.1016/j.aca.2021.339003

Schiferl, D., Nicol, M., Zaug, J. M., Sharma, S. K., Cooney, T. F., Wang, S.-Y., et al. (1997). The diamond C13/12C isotope Raman pressure sensor system for high-temperature/pressure diamond-anvil cells with reactive samples. *Journal of Applied Physics*, *82*(7), 3256–3265. https://doi.org/10.1063/1.366268

Sharma, S. K., Lucey, P. G., Ghosh, M., Hubble, H. W., & Horton, K. A. (2003). Stand-off Raman spectroscopic detection of minerals on planetary surfaces. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *59*(10), 2391–2407. https://doi.org/https://doi.org/10.1016/S1386-1425(03)00080-5

Shvalya, V., Filipič, G., Zavašnik, J., Abdulhalim, I., & Cvelbar, U. (2020). Surface-enhanced Raman spectroscopy for chemical and biological sensing using nanoplasmonics: The relevance of interparticle spacing and surface morphology. *Applied Physics Reviews*, *7*(3), 031307. https://doi.org/10.1063/5.0015246

Sparavigna, A. C. (2023). Tsallis and Kaniadakis Gaussian Functions, Applied to the Analysis of Diamond Raman Spectrum, and Compared With Pseudo-Voigt Functions. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4495547

Thermo Fisher Scientific Raman. (n.d.). OMNIC.

Tu, Q., & Chang, C. (2012). Diagnostic applications of Raman spectroscopy. *Nanomedicine: Nanotechnology, Biology and Medicine*, *8*(5), 545–558. https://doi.org/https://doi.org/10.1016/j.nano.2011.09.013

Wang, A., Jolliff, B. L., & Haskin, L. A. (1995). Raman spectroscopy as a method for mineral identification on lunar robotic exploration missions. *Journal of Geophysical Research*, *100*(E10). https://doi.org/10.1029/95JE02133

Wei, D., Chen, S., & Liu, Q. (2015). Review of Fluorescence Suppression Techniques in Raman Spectroscopy. *Applied Spectroscopy Reviews*, *50*(5), 387–406. https://doi.org/10.1080/05704928.2014.999936

Yamamoto, J., & Hagiwara, Y. (2022). Precision evaluation of nitrogen isotope ratios by Raman spectrometry. *Analytical Science Advances*, *3*(9–10), 269–277. https://doi.org/https://doi.org/10.1002/ansa.202200020

Young, R. A., & Wiles, D. B. (1982). Profile shape functions in Rietveld refinements. *Journal of Applied Crystallography*, *15*(4), 430–438. https://doi.org/10.1107/S002188988201231X

LIST OF FIGURE CAPTIONS

Figure 1: Manual tab - Example of experimental data analysis on an unoriented forsterite crystal. The top figure displays the experimental data. The bottom figure displays the reference file taken from the search bar linked to the RRUFF database. The table on the right-hand-side dynamically displays the fitting results. Each figure is equipped with a toolbox allowing the user to interact with the figure. The area at the bottom of the software dynamically displays error messages or updates.

Figure 2: Auto tab - Example of experimental data analysis on an unoriented forsterite crystal. The top figure displays the experimental data. The bottom figure displays the top three reference files from the RRUFF database. The figure on the right-hand-side displays the matching results for each identified peak. Each figure is equipped with a toolbox allowing the user to interact with the figure. The area at the bottom of the software dynamically displays error message or updates.

Figure 3: Example of automated analysis on a Hawaiian vesicular basalt collected in Oahu showing strong forsterite peaks, with traces of possibly inderite and scolecite.

Figure 4: Example of automated analysis without guidance by the user on a calcite crystal. The matching process identified calcite as the best match.

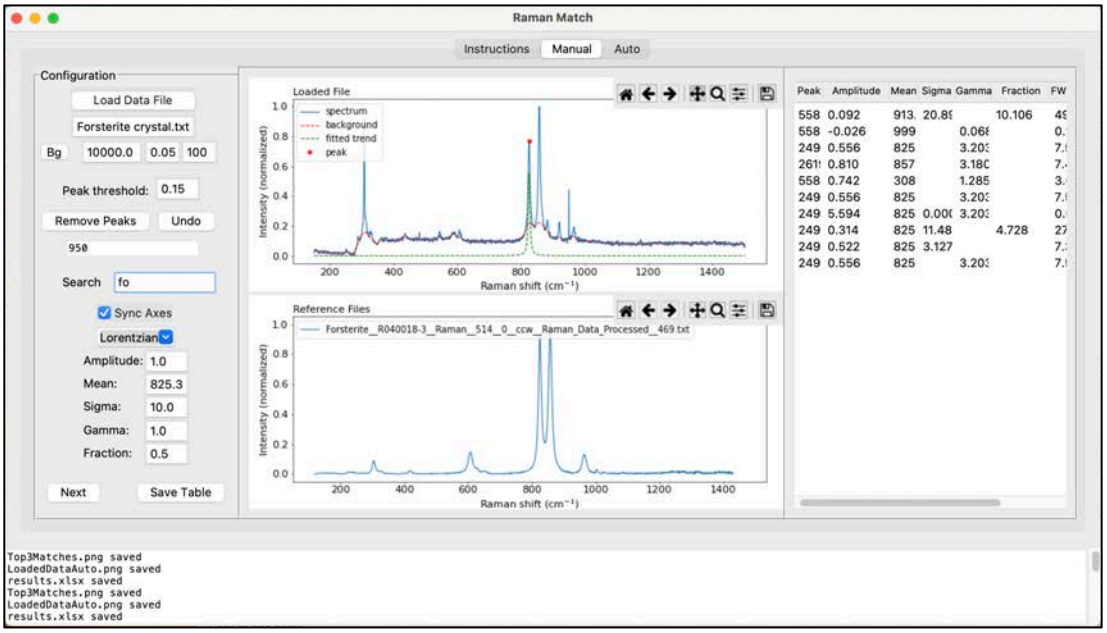
Figure 5: Example of automated analysis with guidance by the user on a quartz crystal. The matching process identified quartz as the best match as expected.

Figure 6: Summary of the interactive tools available for the user in the “Manual” and “Auto” tab. The instruction tab provides the user with a similar description of events.

This is the peer-reviewed, final accepted version for American Mineralogist, published by the Mineralogical Society of America.
The published version is subject to change. Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2023-9227. http://www.minsocam.org/

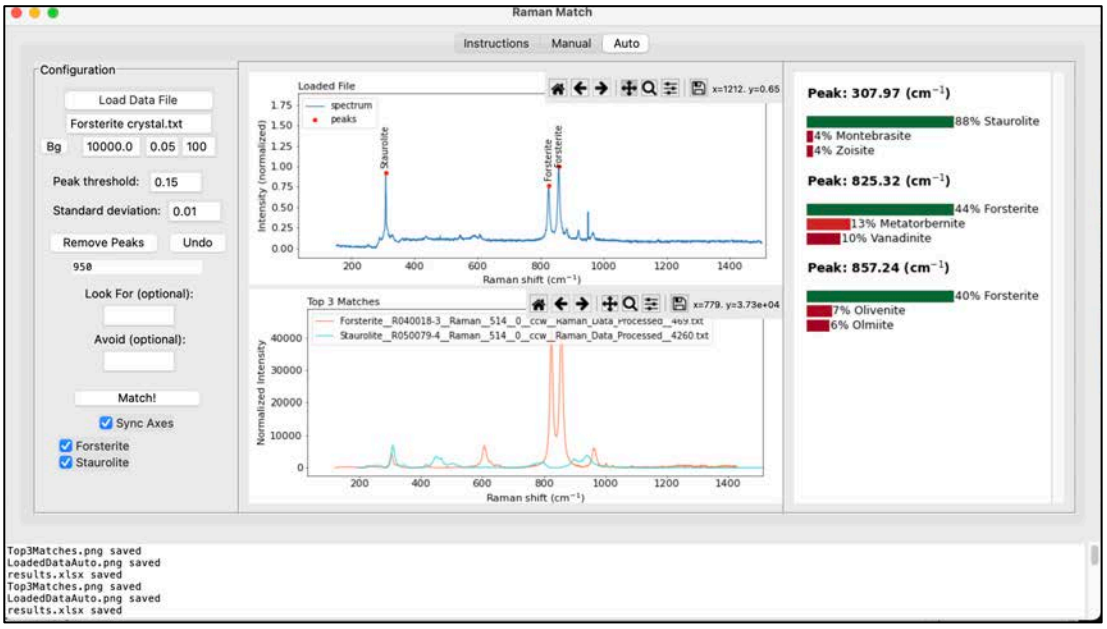
Revision 2

562 **Figure 1:**



563

564 **Figure *2***:

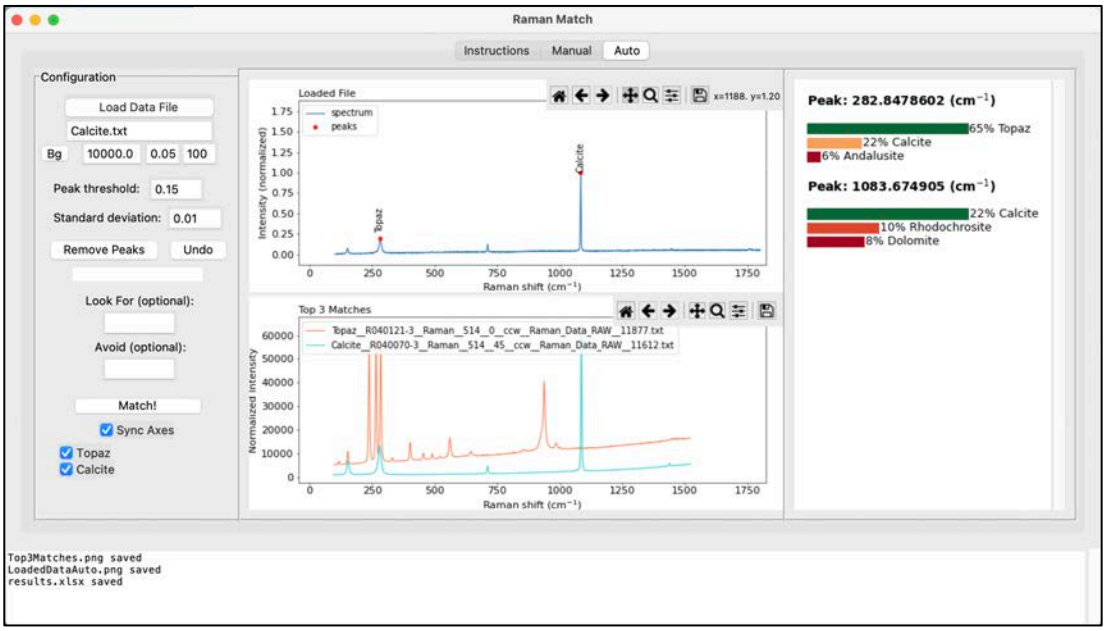


565

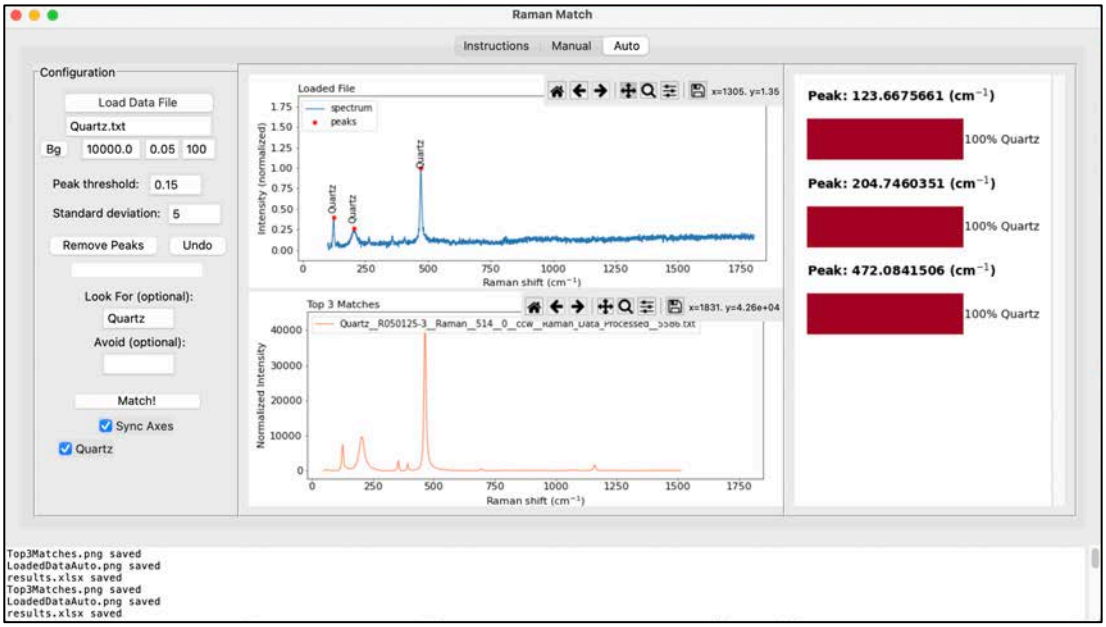
566

Figure 3:

Figure 4:

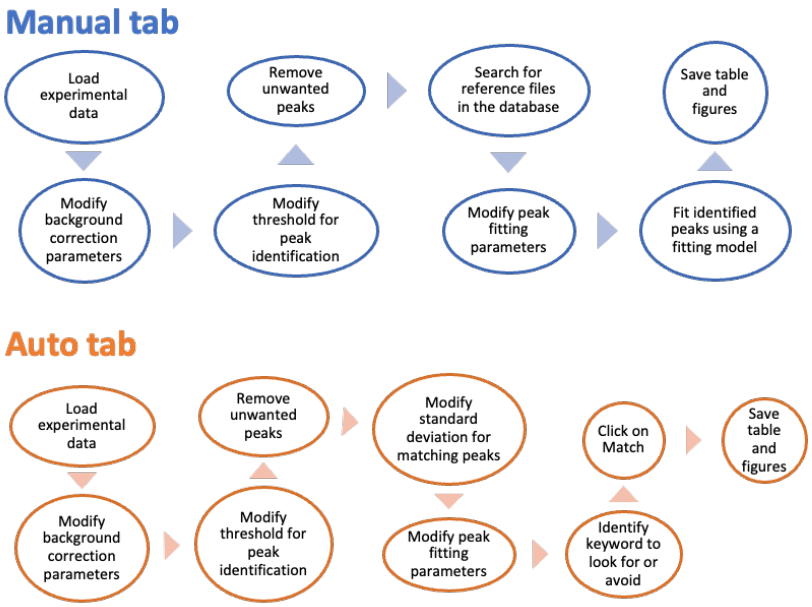


571 **Figure 5:**



572

573 **Figure 6:**



574