

# RamanCrystalHunter: A new program and database for processing, analysis, and identification of Raman spectra

FABRIZIO NESTOLA<sup>1,2,\*†,‡</sup>, QIWEI ZHANG<sup>3,§</sup>, MAXWELL C. DAY<sup>1</sup>, SOFIA LORENZON<sup>1,4</sup>,  
MARTHA G. PAMATO<sup>1</sup>, IVANO ROCCHETTI<sup>5</sup>, CLAUDIO BENDAZZOLI<sup>6</sup>, DAVIDE NOVELLA<sup>1</sup>,  
CLAUDIO MAZZOLI<sup>1</sup>, RAFFAELE SASSI<sup>1</sup>, D. GRAHAM PEARSON<sup>3</sup>, EVAN M. SMITH<sup>7</sup>, MICHAEL SCOTT<sup>8</sup>,  
ANNA BARBARO<sup>9</sup>, FRANK E. BRENKER<sup>9</sup>, LISA SANTELLO<sup>1</sup>, SIMONE MOLINARI<sup>2,||</sup>, KAI QU<sup>10,11,#</sup>,  
YANJUAN WANG<sup>1,12,\*\*</sup>, RADEK ŠKODA<sup>13</sup>, MATTEO ALVARO<sup>14,††</sup>, MATTIA GILIO<sup>14,‡‡</sup>,  
MARA MURRI<sup>14,§§</sup>, AND ANATOLY V. KASATKIN<sup>15</sup>

<sup>1</sup>Dipartimento di Geoscienze, Università degli Studi di Padova, Via Gradenigo 6, I-35131 Padova, Italy

<sup>2</sup>Museum of Nature and Humankind, Section Alessandro Guastoni, University of Padova, Corso Garibaldi 39, 35121, Padova, Italy

<sup>3</sup>Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, Alberta T6G 2E3, Canada

<sup>4</sup>Dipartimento di Scienze della Terra, Università di Pisa, Via Santa Maria 53, I-56126 Pisa, Italy

<sup>5</sup>MUSE—Museo delle Scienze di Trento, Corso del Lavoro e della Scienza 3, I-38122, Trento, Italy

<sup>6</sup>F-Tech Future Tlc Technologies, Strada Santa Caterina 8, 37134 Verona, Italy

<sup>7</sup>Gemological Institute of America, New York, New York 10036, U.S.A.

<sup>8</sup>DBA White Rose Enterprises, 221 Main Street, Los Alto, California 94022, U.S.A.

<sup>9</sup>Schwiete Cosmochemistry Laboratory, Department of Geosciences, Goethe University, Altenhoferallee 1, 60438 Frankfurt, Germany

<sup>10</sup>Tianjin Center, China Geological Survey, 300170 Tianjin, China

<sup>11</sup>School of Earth Sciences and Engineering, Nanjing University, 210023 Nanjing, China

<sup>12</sup>School of Earth Sciences and Resources, China University of Geosciences, 100083 Beijing, China

<sup>13</sup>Department of Geological Sciences, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic

<sup>14</sup>Dipartimento di Scienze della Terra e dell'Ambiente, Università degli Studi di Pavia, Via Ferrata 7, I-27100 Pavia, Italy

<sup>15</sup>Fersman Mineralogical Museum of the Russian Academy of Sciences, Leninsky Prospekt 18-2, 119071 Moscow, Russia

## ABSTRACT

RamanCrystalHunter (RCH) is a new software program designed to pre-process, analyze, and identify Raman spectra by comparison with spectra in the RamanCrystalHunter Database (RCHDB). The software is free and can be downloaded from the website <https://www.fabrizionestola.com/rch>. RCH is characterized by a simple graphical user interface, making it suitable for both specialist and non-specialist users, and it has been developed mainly for applications in Earth Sciences (processing the spectra of minerals) but can be used to process the Raman spectra of any synthetic or natural inorganic or organic material. RCH allows users to visualize, pre-process (e.g., using smoothing, noise reduction, and baseline correction operations), and analyze (e.g., using fitting or various calculation tools) Raman spectra. Moreover, it is equipped with the RCHDB, a new database of high-quality mineral spectra that can be downloaded for free, along with the RCH program. The RCHDB contains the Raman spectra of minerals (including single- and multi-phase inclusions within mineral hosts, for example, diamonds) and related synthetic compounds, allowing for rapid and accurate identification of unknown spectra. The RCH software includes highly customizable yet efficient and user-friendly methods for processing and analysis of Raman spectra and represents a valuable contribution to the field of Raman spectroscopy, whose applications have expanded greatly in recent years, especially in Earth Sciences. Two practical examples of novel ways in which this software can be used for geoscience applications are presented.

**Keywords:** Raman spectroscopy, software, Raman database, spectrum matching, spectrum fitting, Spectroscopy in Geology: A decade of breakthroughs

## INTRODUCTION

Raman spectroscopy is one of the most powerful “in-situ” non-destructive analytical techniques and is commonly used in many scientific fields. In Earth Sciences, it is used to identify, characterize, quantify, and classify minerals, fluids, and gases [e.g., see reviews by Dubessy et al. (2012) and Pasteris and Beyssac (2020)]. The confocal nature of most modern instruments has made Raman spectroscopy a true three-dimensional micro-analytical technique. In this sense, the technique has proven especially useful when applied to identifying mineral

\* Corresponding author E-mail: [fabrizio.nestola@unipd.it](mailto:fabrizio.nestola@unipd.it). Orcid <https://orcid.org/0000-0002-4875-5125>

† Special collection papers can be found online at our website in the Special Collection section.

‡ Open access: Article available to all readers online. This article is CC-BY-NC-ND.

§ Orcid <https://orcid.org/0000-0002-5053-142X>

|| Orcid <https://orcid.org/0000-0003-4105-4833>

# Orcid <https://orcid.org/0000-0002-4961-9598>

\*\* Orcid <https://orcid.org/0000-0002-6985-2012>

†† Orcid <https://orcid.org/0000-0002-6975-3241>

‡‡ Orcid <https://orcid.org/0000-0002-7299-3377>

§§ Orcid <https://orcid.org/0000-0002-7149-9940>

and fluid inclusions, as exemplified in the first discovery of ringwoodite in diamond (Pearson et al. 2014) or in the study of fluid inclusions trapped within minerals (e.g., Pasteris et al. 1988); it can even be applied in-vivo to investigate marine bio-calcification mechanisms (De Carlo et al. 2019) and in-vitro experiments in the field of antibiotics (Carey and Heidari-Torkabadi 2015).

The simple instrumental setup and minimal sample preparation, coupled with rapid and non-destructive analysis, makes Raman spectroscopy a versatile tool accessible to both specialist and amateur users. In Earth Sciences, Raman spectroscopy has many important applications, including the rapid identification of mineral species and amorphous phases (e.g., fluids, gases, inorganic materials, and organic compounds) and the detection of microscale inclusions within mineral hosts. Although the acquisition of raw spectra is relatively straightforward, it is only the first step in a campaign of spectroscopic study and often specific types of data processing are required depending on what information the researcher wishes to extract from the raw spectra. The most common processing methods involve: (1) comparing a spectrum with a database to identify the mineral (or material) that has been analyzed and (2) fitting of the peaks constituting a spectrum to quantify various chemical and structural aspects of the mineral. Commercial software developed for specific Raman spectroscopy systems, such as OMNIC (Thermo Fisher Inc.), LabSpec 6 (HORIBA Jobin Yvon, Japan), WITec Project (WITec GmbH, Germany), WIRE (Renishaw Inc., U.K.), and LightField (Princeton Instrument Inc.) allow these processing tasks to be performed, in addition to spectra acquisition and visualization, but they are not freely available.

Over the last two decades, non-commercial Raman signal processing software programs have been designed for both geological and non-geological purposes. With regard to software developed for non-geological purposes, programs like Raman-ToolSet (Reisner et al. 2011; Candeloro et al. 2013) and NWUSA (Song et al. 2021) offer methods for spectral pre-processing including noise filtering, background fluorescence subtraction, and smoothing. However, these programs are mainly designed to analyze Raman spectra using statistical techniques and related algorithms, such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Artificial Neural Network (ANN), and Support Vector Machines (SNMs). Although these tasks are useful in many scientific disciplines, they do not allow the fitting of spectra and identification of unknown spectra by comparison with a database, both of which are crucial in Earth Science applications.

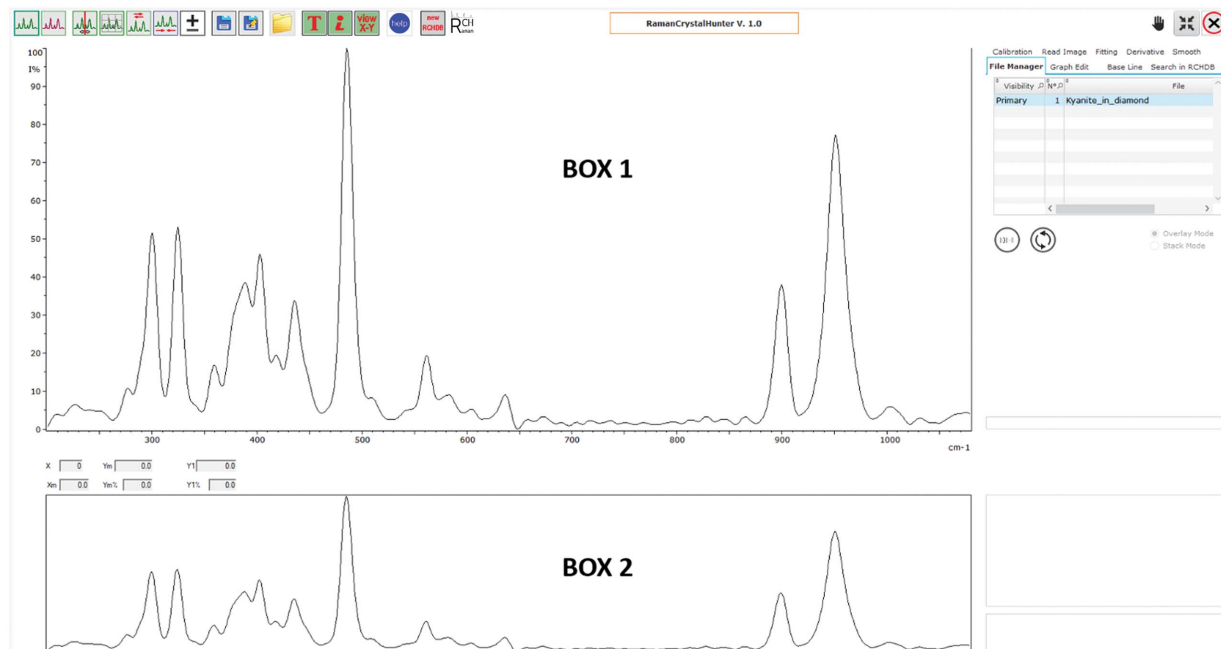
Among software developed for geological purposes, the most common non-commercial Raman software program is CrystalSleuth (Laetsch and Downs 2006). This software was developed in association with the RRUFF Project (Lafuente et al. 2015), which aimed to create an open-access Raman spectroscopy and X-ray diffraction database for mineral identification. CrystalSleuth is characterized by a user-friendly interface and includes methods for background corrections and spectrum identification by matching raw Raman spectra with those in the RRUFF database. Although this program is useful for pre-processing Raman spectra and mineral identification, it cannot perform other useful functions, such as the fitting of spectra.

Spectragryph is another notable software program designed for the acquisition, analysis, and processing of Raman spectra of geologic materials. The basic version of Spectragryph is free, but advanced versions are only available in different commercial packages (Friedrich Menges Software-Entwicklung; <http://spectroscopy.ninja>). This program allows for spectrum fitting and performs other useful calculations, such as spectrum derivatives. However, it may not provide satisfactory fitting results for peaks with relatively low intensities. Although other free-to-use programs are available online that allow for fitting of Raman spectra (and other types of spectra), e.g., Fityk (Wojdyr 2010) and/or open Raman databases [e.g., Raman Open Database (ROD) (El Mendili et al. 2019)], a single open-access program that allows both the fitting of spectra and the comparison of spectra with a free Raman database is currently missing. Moreover, most programs used for identification of Raman spectra cannot reliably process spectra that contain a mixed signal from two or more phases producing mixed spectra. Multi-phase minerals (e.g., mineral intergrowths, exsolution, or retrogression to several phases) are often too small, or the intergrowth occurs at such a small scale that the Raman user cannot reliably verify that they are analyzing each discrete phase, or they cannot even see the different phases. Such mixed spectra can lead to the misidentification of key minerals. Because of these challenges, software capable of deconvoluting (separating) mixed Raman spectra is desperately needed so that multi-phase minerals (or related synthetic compounds) can be accurately identified.

Here, we present RamanCrystalHunter (hereafter RCH), a new Raman software program developed for pre-processing and analyzing Raman spectra to identify geologic samples. This software was designed in conjunction with the RamanCrystalHunter Database (hereafter RCHDB) project, which aims to construct a comprehensive database of high-quality spectra for minerals and other geological materials. This database is fully integrated into RCH and allows users to compare raw spectra with reference spectra in the RCHDB. The RCHDB project will continually develop and improve the RCH software and update the RCHDB in an effort to provide a freely accessible means by which the entire Raman community can process, analyze, and identify spectra. Below, we provide a brief overview of how the RCH program works. A comprehensive user guide and tutorial on RCH can be accessed at <https://www.fabrizionestola.com/rch>. Here, we provide two practical examples of how the RCH software can be used to process and identify geological samples. These examples highlight the unique capabilities of RCH when applied to processing complex spectra collected on mineral inclusions and multi-phase samples.

## RCH USER INTERFACE AND OPERATION

The RCH software is free to use and available from the webpage <https://www.fabrizionestola.com/rch>. Both the RCH software program and the RCHDB database can be downloaded from this webpage. Using the RCH software, users can perform various types of pre-processing, fitting, searching, and other operations through a graphical user interface, making this software easy to use and suitable for users with any level of expertise. The main features of the RCH software are described below.



**FIGURE 1.** Graphical user interface of the RCH software showing the Raman spectrum of a kyanite inclusion in a lithospheric diamond (Nestola et al. 2018). The window indicated as box 1 is the main window where resultant spectra are displayed after processing. The window indicated as box 2 is the preview window allowing the user to preview the results of all operations before they are finalized.

### Loading and viewing spectra

The RCH software currently reads raw spectra saved as .asc, .txt, and .csv files, because every commercial Raman system (program) allows users to export data in these file formats. For each new work session, the first spectrum (data file) loaded is recorded as the “primary spectrum” or “primary file.” All operations selected by the user are performed only on the primary spectrum. After performing specific operations on this spectrum, the result is automatically saved (as a.txt file) in the computer directory C:/RamanCrystalHunter/Spectra/temp. In the graphical user interface, the primary spectrum is displayed in two boxes, box 1 and box 2 [see Fig. 1; the Raman spectrum of a kyanite inclusion in a lithospheric diamond (Nestola et al. 2018)]. Box 1 is the main window for spectrum visualization and where the resultant spectrum is displayed after processing is complete. Box 2 serves as a “preview” window allowing the user to visualize (preview) the result of a specific operation before it is applied to the primary spectrum. In box 1, for each spectrum, the wavenumber (X coordinate as  $\text{cm}^{-1}$ ) and both the absolute Raman intensity (Y1) (i.e., the value collected by the spectrometer) and the normalized Raman intensity (Y1%), expressed as a percentage with respect to the maximum intensity value, are reported. More Raman spectra (up to eight) can be added to the user interface in addition to the primary spectrum. All loaded spectra can be displayed as overlapping (Overlay Mode) or stacked spectra (Stack Mode). Although more than one spectrum can be loaded, all operations are always applied solely to the primary spectrum.

### Pre-processing spectra

Raw Raman spectra often contain signals that overlap with the Raman signal from the analyzed phase(s), including noise

(due to systematic, instrumental, and/or contamination-related issues), photoluminescence, cosmic rays, etc. Consequently, it is often necessary to apply a sequence of filters (pre-processing operations) to improve the quality of the raw data and highlight the Raman signal from the phase(s) of interest thereby optimizing subsequent operations (e.g., spectrum matching). To achieve this attenuation of background noise, a process called smoothing is performed with RCH in two ways: (1) using a triangle evaluation where each intensity value ( $y_i$ ) is averaged with respect to the two closest intensity values ( $y_{i-1}$  and  $y_{i+1}$ ), and (2) using the Savitzky-Golay filter (Savitzky and Golay 1964) where up to nine points are applied to each smoothing interval. Savitzky-Golay filters are commonly used in digital smoothing of spectroscopic data. They are derived directly from the time-domain problem of “smoothing” and are well suited to this purpose (Press and Teukolsky 1990), reducing high-frequency noise and low-frequency signal. To improve the background, i.e., to reduce fluorescence effects and remove signal associated with noise, a baseline correction can be performed. This operation is performed by drawing a line (spline) that follows the baseline of the raw spectra. The spline is a function represented by a continuous line passing through a series of points selected by the user. Applying the user-defined spline (baseline) to a spectrum “flattens” the spectrum baseline, reducing inherent background artifacts and highlighting the main (characteristic) peaks. This is why the characteristic peaks of the spectra must not be included in the spline (baseline) trace.

### Fitting spectra

The fitting operation is the only operation performed to determine the position (Raman shift in wavenumber,  $\text{cm}^{-1}$ ), intensity

(height), full-width at half maximum (FWHM), and area of the component peaks that comprise the primary spectrum. This operation involves fitting each peak of the spectrum with a series of curves (functions) and iteratively improving the fit by refining the position, height, and FWHM of each component function. All component functions are summed to produce a single calculated fit (the aggregate intensity of all component functions), which should approximately reproduce the primary spectrum (or a specific characteristic peak of the primary spectrum) that is being fit. Fitting parameters are calculated from the first, second, and third derivatives, which correspond to the position, height, and FWHM of each component curve, respectively. The first fitting calculation will probably not provide a precise match between the calculated and primary spectrum and subsequent cycles of refinement of the fit are often required. The fitting algorithm interprets the difference between the calculated and primary spectra to determine which properties of each curve are contributing to the discrepancy, e.g., if the position of a given component curve is not perfectly centered or if the height and/or FWHM of a given component curve is under- or overestimated. Depending on the type of discrepancy, the fitting algorithm automatically applies specific correction parameters to improve the fitting during the refinement procedure. Refinement of the fitting parameters is performed iteratively and can be done automatically or manually by the user until a satisfactory fit is obtained. The quality of a fit is calculated as a delta ( $\Delta$ ) value reported both as an absolute and percentage value that quantifies the aggregate discrepancy between the fit and primary spectrum. Smaller  $\Delta$  values correspond to higher quality fits.

To facilitate the fitting operation and to achieve the best possible fit, we recommend that users do not attempt to fit the entire primary spectrum but instead use a smaller user-defined fitting interval (wavenumber range,  $\text{cm}^{-1}$ ) that includes the characteristic peaks of interest. The two endpoints of the selected fitting interval must lie along the baseline of the primary spectrum (Raman intensity  $\sim 0$ ), and thus, it is advisable to perform a baseline correction before the fitting operation is used.

### The RCH Database of Raman spectra (RCHDB)

The RCHDB Raman Project was initiated to create a free, easily accessible, and comprehensive database of Raman spectra of minerals and other geological materials. The database can be downloaded from <https://www.fabrizionestola.com/downloads>. Currently, the database includes Raman spectra collected in the Raman Laboratory at the Department of Geosciences, University of Padua, with additional mineral spectra provided by the authors and other collaborators. The RCHDB database will be regularly improved and updated, and thus, users of the RCH software are encouraged to contribute to the project by submitting Raman spectra of minerals that are not currently included in the database. Users can submit spectra by following the instructions at <https://www.fabrizionestola.com/submit-spectra>.

The RCHDB generally contains one or two Raman spectra for each mineral, but more spectra are included for minerals that have variants and, thus, several unique Raman spectra. A feature of the RCHDB is that it also contains examples of mineral inclusions trapped within diamond. The differential decompression

between the lattice of the included mineral and the host diamond results in an over-pressure (Angel et al. 2022) that can shift the Raman peak positions of both the host diamond and inclusion away from those typically measured at “room” conditions. For such inclusions, it is important to consider spectral shifts in certain Raman bands and an application using this shift is described below (see “Results and discussion” section below). This portion of the RCHDB that contains “in-situ” measurements of inclusions within diamonds that likely exhibit various different over-pressure conditions has been constructed from both the compilation of Smith et al. (2022) and unpublished measurements made at the University of Alberta, the University of Padova, and Goethe University, Frankfurt.

RCH can be used to identify unknown spectra by comparison with spectra included in the RCHDB. Spectrum matching is performed by calculating a match percentage (MATCH%) between the unknown spectrum and all reference spectra in the database. Spectrum matching is computed using the Spectral Angle Mapper (SAM) method (Vengatesh 2013). Here, the degree of similarity between the unknown and reference spectra is determined by calculating the angle between the two spectra and treating them as vectors to match pixels of the unknown spectrum to pixels of the reference spectrum. The RCH software generates a ranking list in which spectra with similar features are listed in order of decreasing match percentage. The user can also compare a single or series of peaks by specifying the position (Raman shift,  $\text{cm}^{-1}$ ), the relative intensity, and the FWHM of each peak. The user can adjust the accuracy of searches (based on entire spectra or individual peaks) by controlling the frequency range over which “matching peaks” are reported. In addition to filtering all spectra in the RCHDB in an attempt to identify one’s spectrum, users can also search in the “inclusion” and “synthetic” databases to refine their search and decrease processing times.

### Other operations

In addition to pre-processing, fitting, and spectrum identification by comparison with the RCHDB, RCH can perform several other useful operations. For example, during the acquisition of a Raman spectrum, optical aberrations of the spectrometer and/or misalignment of the CCD camera can result in distortion of the wavenumber scale (X-axis) of the spectrum. This type of distortion can be corrected using the Calibration operation, which transforms one or more of the original wavenumber values (X-coordinates) of the spectrum to corrected values. Depending on the number of X-coordinates selected by the user, calibration (correction) of the spectrum will involve shifting of the entire spectrum (one point), a linear correction (two points), a parabolic correction (three points), or a correction using a cubic spline function (four or more points).

The Graph Edit operation allows the user to modify the primary spectrum in several different ways depending on the needs of the user. For example, the Replace operation allows for deletion of a pre-defined region of the spectrum and subsequent replacement of this region with a five-point Bezier curve. Users can simulate noise along the Bezier curve using the Noise Level and Noise Frequency tools. The Subtract operation allows the user to improve the background baseline of the primary

spectrum by weakening or accentuating particular peaks in the primary spectrum. The Add function operation allows the user to add additional peaks to the primary spectrum. These peaks can be added as Gaussian, Lorentzian, or Voigt functions by adjusting the values of the  $\sigma$ ,  $\beta$ , and  $\gamma$  parameters.

Using the Addition and Subtraction operations, the user can sum or subtract two spectra. This is useful if users wish to combine different wavenumber ranges of the same mineral spectrum (Addition operation) or to remove peaks produced by phases other than the mineral of interest, e.g., peaks in an inclusion-host system (Subtraction operation). Addition operations can be performed in two ways. The first method, Common Range, sums the signal intensity (expressed in percent, Y%, not absolute intensities) over the common wavenumber range of both spectra. This operation allows users to highlight characteristic peaks of a specific spectrum with respect to the background. The second method, Entire Range, allows the user to combine two spectra collected over different wavenumber regions. In the overlapping wavenumber region of both spectra, intensities are summed, and the intensities are expressed in percent. The signal intensity in the wavenumber regions that do not overlap is recalculated as a function of the relative intensity of the overlapping regions. If the two spectra do not have a common wavenumber region, the region without data will be displayed as a straight-line segment. The Subtraction operation allows the user to subtract the signal (expressed in Y%) of the second spectrum (subtrahend) from the first spectrum (minuend) across overlapping wavenumber regions. Moreover, the Addition and Subtraction operations allow the user to assign a specific “weight” to spectra involved in

addition or subtraction operations to control the degree to which peaks are minimized (or removed) or accentuated.

Using the Derivative operation, the user can calculate the first-, second-, and third-order derivatives of any Raman spectra. This operation allows the user to easily determine the maximum, minimum, and inflection points of the primary spectra. The most unique operation in RCH, Read Image, allows the user to convert an image of a Raman spectrum, in.png or.jpg format, to a spectrum that can be saved as an.asc file. This resultant.asc file can be loaded into any available Raman software system. The input image file must be a 1-bit black and white image (a spectrum defined by black pixels) and must only consist of the spectrum trace and no other elements, such as the reference axes or the tick marks. Additional elements in the image (apart from the spectrum) must be deleted using a graphics processing software package prior to loading the image file into the RCH software. After loading the image file, the minimum ( $X_{min}$ ) and maximum ( $X_{max}$ ) wavenumbers of the spectrum must be indicated by the user so the wavenumber value (X-coordinate) for each Y% value can be calculated. Using the Calibration operation (see above), it is then possible to recalculate the precise minimum and maximum X-coordinates of the spectrum. The intensity of the spectrum generated from an image is expressed in percentage values (Y%), not as absolute values. The Read Image operation can be applied to any type of graph, not just Raman spectra, allowing users to digitalize old spectra or graphical plots in cases where the corresponding data are no longer available.

The general workflow of RCH is provided in Figure 2, which outlines the actions required to load, modify, identify,

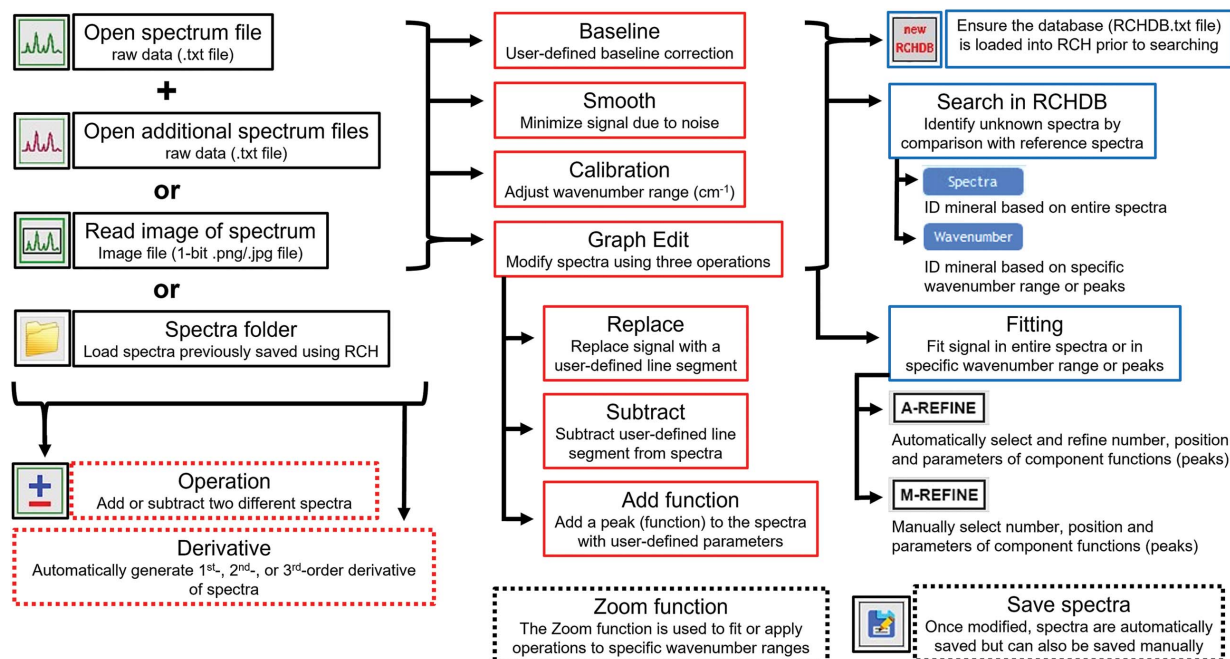


FIGURE 2. Schematic illustrating the general workflow (from left to right) of the RCH software. Black solid boxes describe initial actions for loading/opening spectrum (.txt) files. Red solid boxes describe operations for baseline corrections, smoothing, wavenumber calibration, and graph editing. Blue solid boxes describe operations for fitting and identifying unknown spectra by comparison with reference spectra in the RCHDB. Dashed red boxes describe additional operations for spectrum addition/subtraction and determination of spectrum derivatives. Dashed black boxes describe the zoom function and information about saving spectrum files.

and fit a given Raman spectrum. A more detailed and comprehensive description of the RCH software (and how to use it) is provided in the RCH User-Guide, which is included in the Online Materials<sup>1</sup> and is also available at <https://www.fabrizionestola.com/rch>.

## RESULTS AND DISCUSSION

### Example 1: Geobarometry of a breyite ( $\text{CaSiO}_3$ ) inclusion within a super-deep diamond

Raman spectroscopy is commonly used for in situ analysis of fluid and/or mineral inclusions within mineral hosts, such as diamonds (Nasdala et al. 2004). Diamonds, and their inclusions, are key geological materials that provide a unique opportunity to directly investigate the deepest regions of our planet as diamonds form at depths of 120–130 to ~800–1000 km (e.g., Walter et al. 2022; Day et al. 2023). Mineral inclusions entrapped by diamonds during their crystallization not only provide information about the chemical composition and mineralogy of mantle rocks but also about the depths (and thus the pressures and temperatures) at which they form.

Here, we present an example of how RCH can be used to extract geological information from an inclusion-host system. In this example, we examine the mineral breyite,  $\text{CaSiO}_3$ , which is one of the most commonly observed inclusions in super-deep diamonds (Brenker et al. 2021). Super-deep diamonds are extremely rare (compared to lithospheric diamonds) and form at depths of ~300 to ~800–1000 km (e.g., Walter et al. 2022). To obtain the depth of formation of these very rare diamonds, “elastic geobarometry” is often applied to breyite inclusions (for more details about this method, the reader is referred to Anzolini et al. (2018), Angel et al. (2022), and Genzel et al. (2023)]. Here, we provide an example of

how to apply the pre-processing and fitting operations of RCH to the Raman spectrum of a breyite inclusion entrapped within the super-deep diamond investigated by Anzolini et al. (2018). Precise determination of peak positions in the Raman spectrum of breyite allows the derivation of information that constrains the residual pressure, which in turn, can be used to determine the depth of formation of the diamond-breyite system (Angel et al. 2022).

First, since the original Raman spectrum of the breyite inclusion described by Anzolini et al. (2018) was not published or supplied in digital form, we copied Figure 5 from this publication and saved it as .jpg image file after deleting all text, the X and Y axes, and the numbers that were included in the original figure. Next, we used the Read Image operation to upload the image by clicking on the green icon located in the Read Image menu. Then, we indicated the minimum and maximum wavenumbers of the spectrum that were taken from the original figure (e.g., Fig. 5 in Anzolini et al. 2018 reports minimum and maximum wavenumbers of 100 and 1200  $\text{cm}^{-1}$ , respectively). By clicking on RUN, the raw Raman spectrum of breyite was generated and displayed in box 1 and box 2. At the same time, an .asc file of the spectrum was automatically saved. Following Anzolini et al. (2018), we focused on the three Raman peaks at approximately 657, 977, and 1040  $\text{cm}^{-1}$ . The raw Raman spectrum of the breyite inclusion (Fig. 3) showed significant background signal, and therefore, it was necessary to apply a smoothing correction.

This was achieved by clicking on the Smooth tab, located on the right side of the user interface, and selecting the polynomial function with 9 points and 5 recursions. The Raman spectrum also showed anomalous background intensity in the higher wavenumber region, and thus, a Baseline Correction was required. This was performed by clicking on the Base Line tab and adding a series of points along the baseline of the Raman

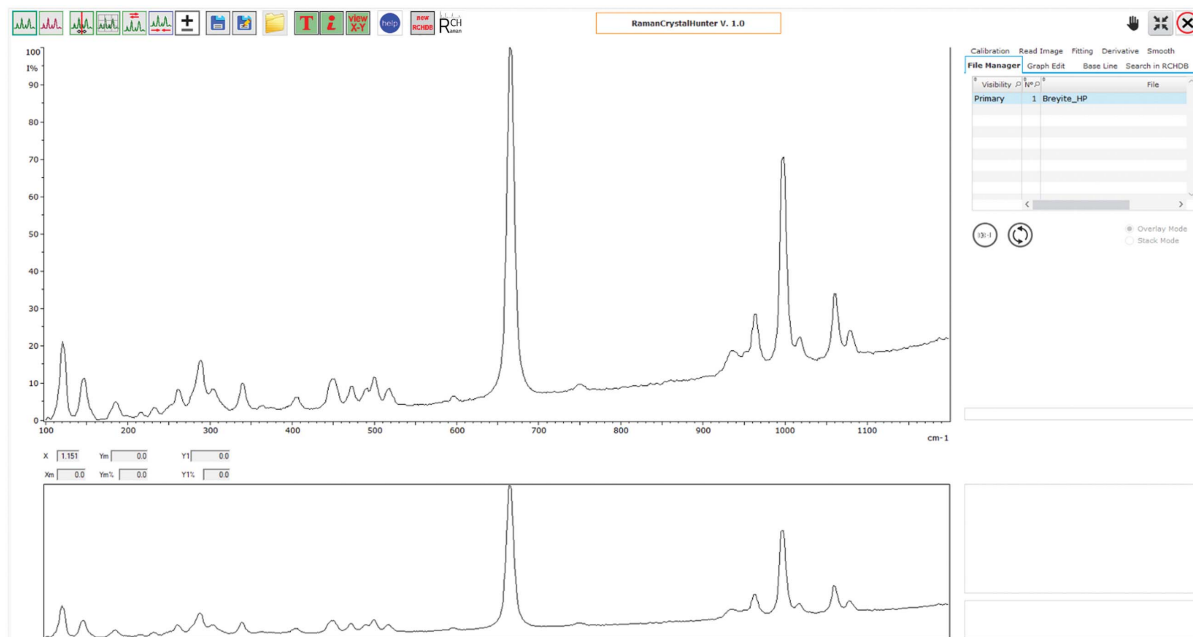


FIGURE 3. The Raman spectrum of breyite,  $\text{CaSiO}_3$  (Brenker et al. 2021) generated using an image of Figure 5 from Anzolini et al. (2018), the image was transformed into a spectrum using the “Read Image” operation.

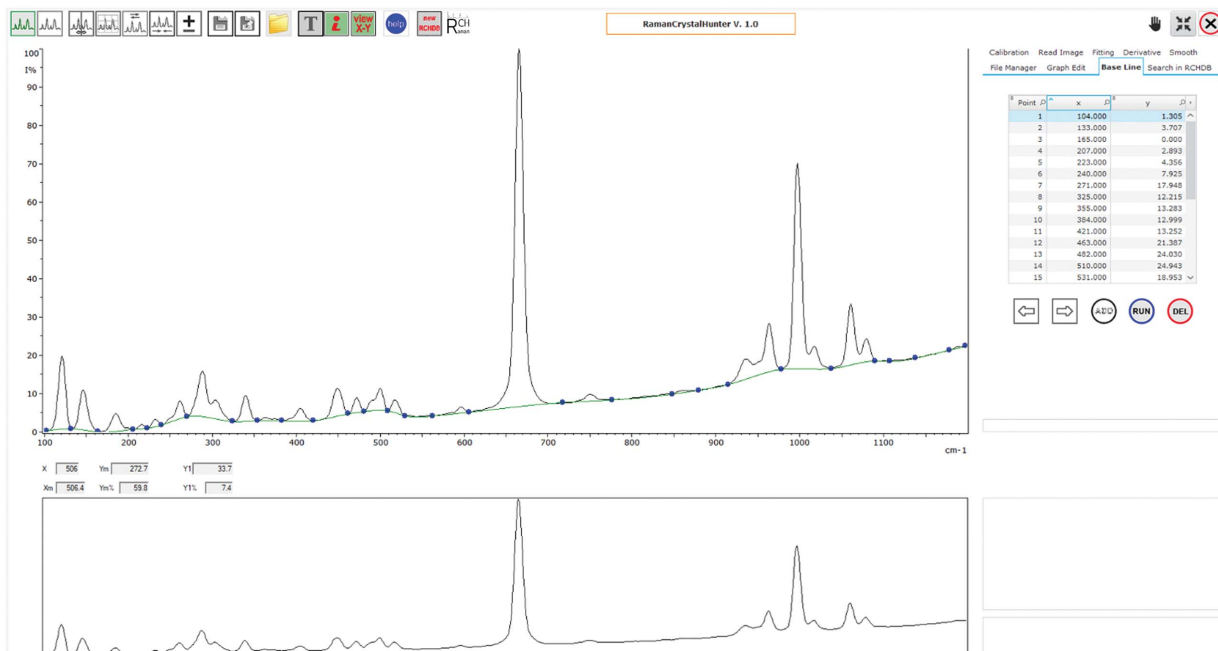


FIGURE 4. The application of single points to the Raman spectrum shown in Figure 2 to apply a baseline correction using the Base Line operation.

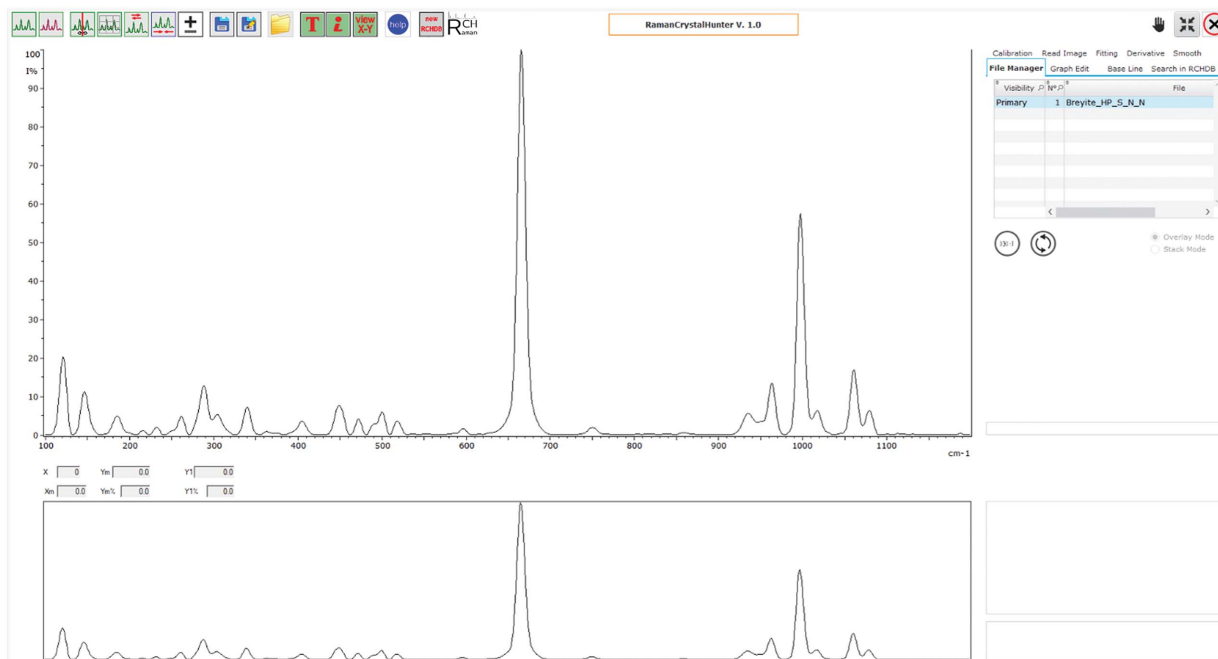
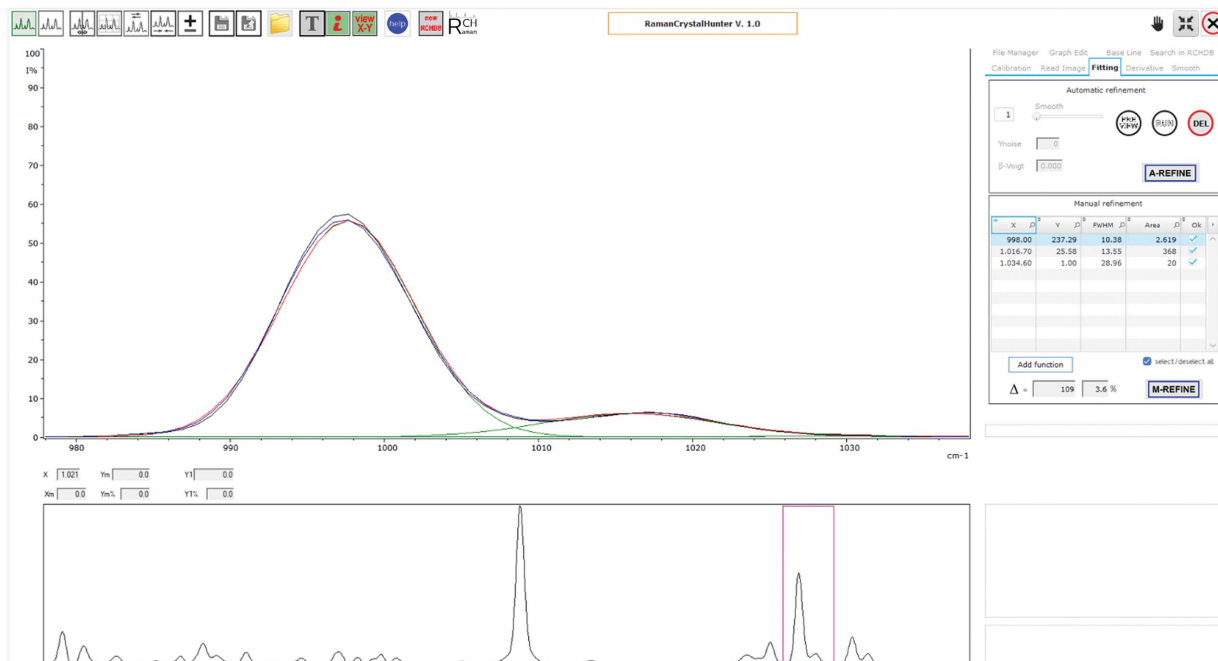


FIGURE 5. The Raman spectrum of breyite after the baseline correction was completed, as explained for Figure 3.

spectrum (Fig. 4). These points define a new baseline for the Raman spectrum below which all intensity will be reduced to ~0. By clicking on RUN, the baseline correction was applied to produce a corrected spectrum in which the background was considerably improved (Fig. 5).

After completion of the pre-processing corrections, the spectrum was ready for further analysis. At this point, if we

could not immediately identify or did not have prior knowledge of the mineral phase corresponding to our spectrum, we could perform a matching search using the RCHDB. This is done by opening the Search in the RCHDB tab and clicking on the Search button to generate a ranking of spectra similar to the breyite spectra listed in order of decreasing match percentage (MATCH%). Alternatively, we can perform a matching search



**FIGURE 6.** The first fitting attempt of the peak in the spectrum of breyite located at  $\sim 1000$   $\text{cm}^{-1}$ . As described by Anzolini et al. (2018), determining the intensity of this peak is crucial to obtain the depth of formation of the inclusion-host system. This first fitting attempt results in a fit [ $\Delta = 109$  (3.6%)] that can be improved and thus requires further processing.

using individual peak positions instead of the entire spectrum, this is done by clicking the Wavenumber button and entering the position (X), intensity (Y%), and FWHM of the peaks we wish to match in the corresponding table.

To apply the elastic geobarometric method to the breyite inclusion, we must first precisely determine the positions of the three main peaks in the spectrum of the breyite inclusion. To do this, we used the fitting operation. First, we selected the wavenumber range of the spectrum that contains these three peaks using the Zoom function. As described above, the two endpoints of the user-defined wavenumber region must lie on the baseline, i.e., have an intensity value of  $\sim 0$ . For our purposes, we need to determine the exact position of the peak at  $\sim 977$   $\text{cm}^{-1}$  as this is the only peak in the spectrum of breyite that is not significantly affected by anisotropy (Anzolini et al. 2018). To do this, we can click the Fitting tab and set the fitting parameters. As a first attempt, we selected the wavenumber region corresponding to the peak at  $\sim 1000$   $\text{cm}^{-1}$  in box 2 [this peak is located at  $999$   $\text{cm}^{-1}$  in Fig. 5 of Anzolini et al. (2018)]; we did not need to change the default fitting parameters and previewed the fit by clicking PREVIEW. By clicking on RUN, the fitting procedure was completed, and the fit shown in Figure 6 was generated. In Figure 6, the blue line represents the original Raman spectrum, the green lines represent the component functions used to fit the primary spectrum, and the red line represents the fitting result. In the Fitting menu, a table was displayed in which X values (peak positions), Y values (peak intensities), peaks widths (FWHM), and peak areas of all component functions were reported for the user-defined fitting interval (wavenumber region). The first fitting result (Fig. 6) was already very satisfactory. The quality of the fit was expressed as a delta value ( $\Delta$ ), which represented the calculated sum ( $X^2$ ) of

differences between the fit and the spectrum across the entire fitting interval. For the fit in Figure 6,  $\Delta = 109$  (3.6%).

To improve the fitting, we clicked A-REFINE several times, until a better fitting result was obtained. The fitting result can be further improved by using the M-REFINE operation (also located in the Fitting menu). This involved adjusting the positions, intensities, and FWHM values of the component functions. Using the fitting procedure described above, we obtained a high-quality fit [ $\Delta = 84$  (2.8%)] in a few minutes, shown in Figure 7.

Anzolini et al. (2018) reported a position of  $999$   $\text{cm}^{-1}$  for this peak (obtained using the Thermo-Scientific OMNIC Software) compared to a peak position of  $997.8$   $\text{cm}^{-1}$  obtained using RCH. Taking into consideration that we did not have the original Raman data from Anzolini et al. (2018) and thus were forced to start from a.jpg image of their spectrum, a difference of  $1.2$   $\text{cm}^{-1}$  should be considered a very satisfying result. Using our result and the Raman pressure dependency for this peak, which is  $5.16 \pm 0.09$   $\text{cm}^{-1}$   $\text{GPa}^{-1}$  for the  $977.1$   $\text{cm}^{-1}$  peak at room pressure (Anzolini et al. 2018), we obtained a residual pressure  $P_{\text{inc}} = \sim 4.01$   $\text{GPa}$ , compared to  $P_{\text{inc}} = 4.26 \pm 0.07$   $\text{GPa}$  reported by Anzolini et al. (2018). Using this  $P_{\text{inc}}$  value we determined the depth of formation of the diamond-breyite system following the same approach used by Anzolini et al. (2018) and the methods described by Angel et al. (2022). We obtained an average  $P_{\text{depth}}$  of  $9$   $\text{GPa}$  ( $\sim 270$   $\text{km}$  depth), in good agreement with the  $P_{\text{depth}}$  of  $8.7 \pm 0.9$   $\text{GPa}$  ( $\sim 260 \pm 30$   $\text{km}$  depth) reported by Anzolini et al. (2018).

This simple example demonstrates how the RCH software can be used to accurately and efficiently process and analyze mineral spectra, providing results comparable to commercial software.



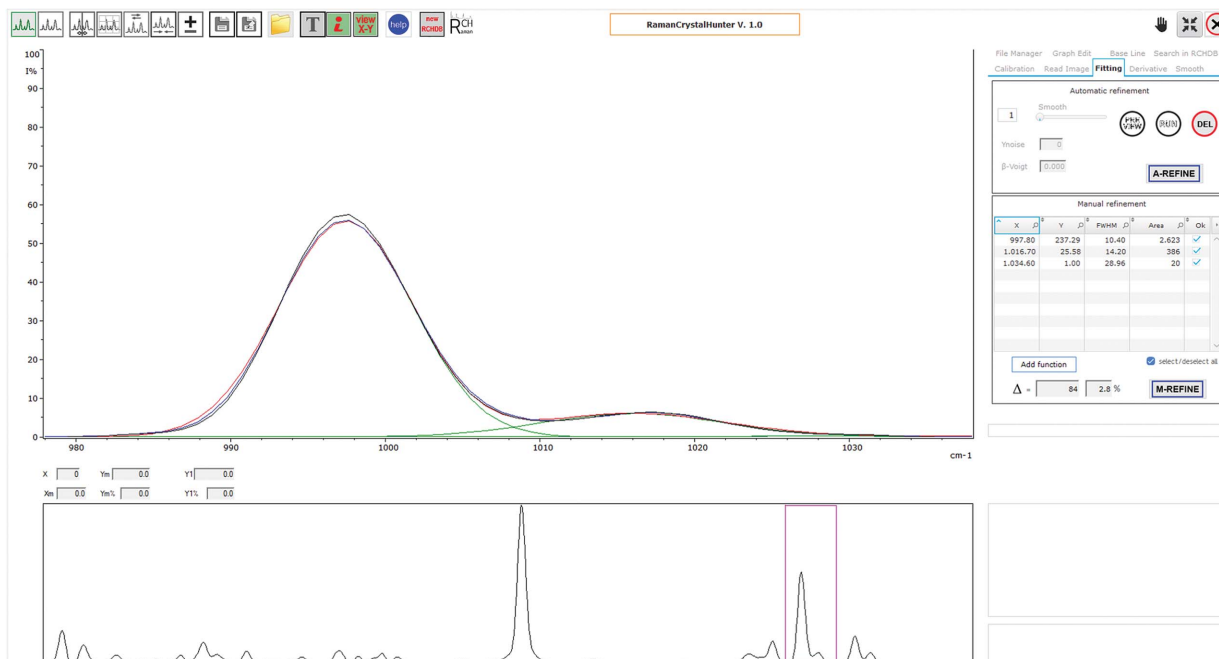


FIGURE 7. The final fitting result of the two peaks in the spectrum breyite as shown in Figure 5 after data processing. The fitting result [ $\Delta = 84$  (2.8%)] is significantly better than the fit in Figure 5 and closely reproduces the data published by Anzolini et al. (2018).

### Example 2: Mineral identification from mixed spectra—a multi-phase inclusion of breyite, coesite, and larnite in diamond

A single Raman spectrum of geologic material may contain signals from several (structurally or chemically) distinct phases. This is often encountered when analyzing composite minerals or synthetic compounds (e.g., polytypic intergrowths, exsolution lamellae, or retrogression of high- to low-pressure phases, etc.) that are too small to be visualized and thus avoided during analysis. Identification of individual phases in a mixed Raman spectrum is notoriously difficult, especially where there is significant overlap in the position of characteristic peaks corresponding to different phases. In most spectrum processing programs, automated phase identification (spectrum matching algorithms) applied to mixed spectra fails, and thus, separating individual mineral spectra is not possible. In the following example, a demonstration of how RCH can be used to process and identify individual phases in a multi-phase inclusion of breyite, coesite, and larnite in diamond is presented.

In Figure 8a, the mixed spectrum of an inclusion of breyite, coesite, and larnite is shown after baseline and smoothing corrections were applied. At first, the user may be unaware that this spectrum is mixed and thus may attempt to identify the spectrum by starting with a simple Spectra search on the entire spectrum using the Search in RCHDB operation. If this is done, the best matches correspond to the spectra of two breyite inclusions in diamond. These spectra (spectra 0004 and 0005, Fig. 8b) account for several of the most intense peaks in the primary spectrum, and the other match results suggested by the software (e.g., serandite, brizziite, rhodonite, Fig. 8b) can be disregarded as these other minerals contain characteristic peaks that are not observed in the primary spectrum and have never been observed

as inclusions in diamond. At this point, the user may conclude that their unknown inclusion contains breyite, without considering the presence of the additional peaks. However, it is likely that the inclusion is composite, comprised of additional phases because several relatively intense peaks in the primary (unknown) spectrum are not observed in the spectrum of breyite. These peaks include a doublet at  $868\text{ cm}^{-1}$  (with a shoulder at  $855\text{ cm}^{-1}$ ) and a peak at  $530\text{ cm}^{-1}$  (Fig. 8b).

To identify the mineral(s) that may correspond to these unknown peaks, a Wavenumber search was performed so only spectra that contained peaks located close to  $868$  and  $530\text{ cm}^{-1}$  were returned. Performing a Wavenumber search for a peak at  $530\text{ cm}^{-1}$  using All spectra returned several minerals (e.g., gonardite, romeite, iowaite, etc., see Fig. 9a) that can be disregarded as they are not observed as inclusions in diamonds and contain many other characteristic Raman peaks not observed in the primary spectrum. To avoid this, the user can perform the same Wavenumber search by filtering only spectra of inclusions in diamond. This is done by selecting the inclusions in diamond option (Fig. 9b) from the database. After performing this type of Wavenumber search, the best matches corresponded to the spectra of two coesite inclusions in diamond (spectra 0013 and 0014, Fig. 9b). The user then performed a similar Wavenumber search on the peak located at  $868\text{ cm}^{-1}$ . The best match corresponded to the spectrum of larnite (spectrum 0035, Fig. 9c). Although some of the other results (e.g., olivine, merwinite, jeffbenite, etc.) are observed as inclusions in diamond, they do not show a shoulder at  $855\text{ cm}^{-1}$  and contain additional characteristic peaks that are not observed in the primary spectrum. Performing an additional search on the  $855\text{ cm}^{-1}$  peak (shoulder) confirmed these results (Fig. 9d). It followed that larnite is the most likely candidate for the peak and shoulder located at  $868$  and  $855\text{ cm}^{-1}$ , respectively.

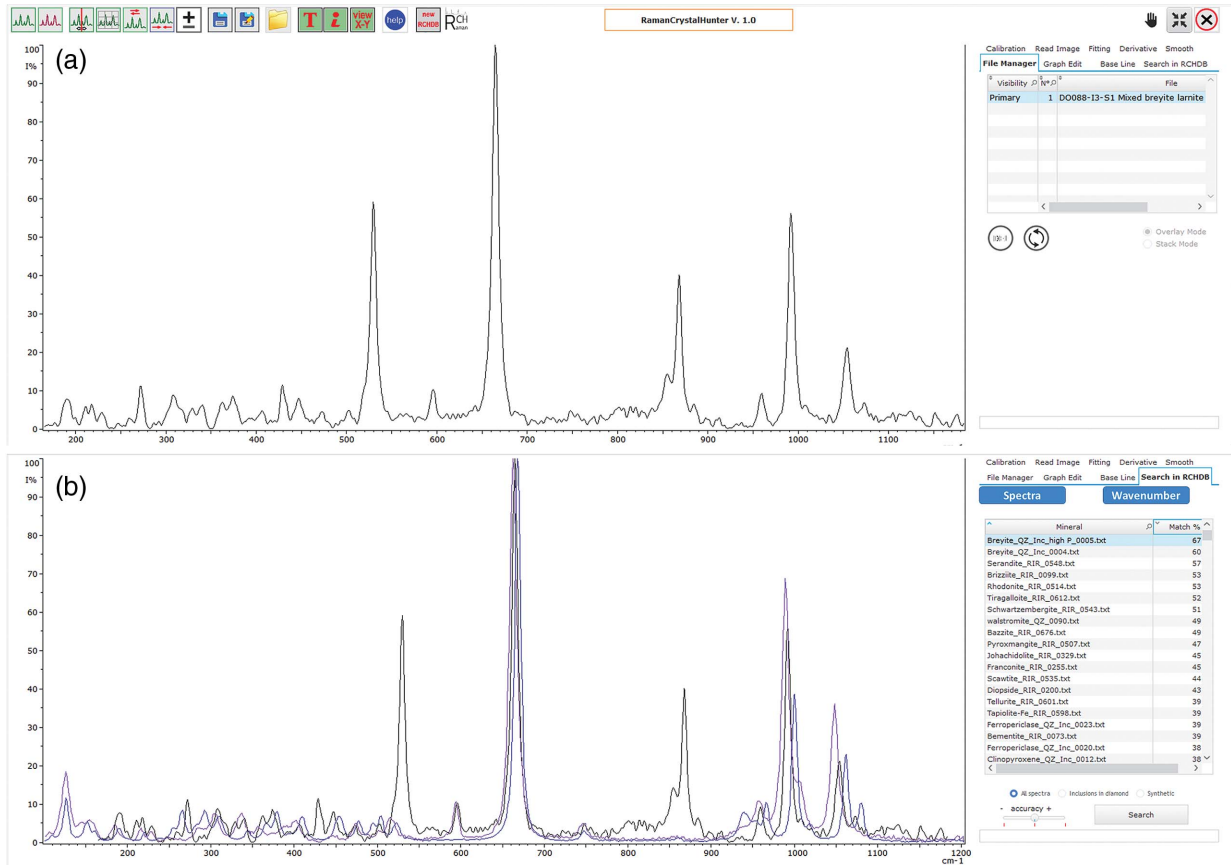


FIGURE 8. (a) The mixed Raman spectrum of a multi-phase inclusion of breyite, coesite, and lamnite in diamond after baseline and smoothing corrections are applied (shown in black). (b) The spectrum match results produced by searching on the entire spectrum (with accuracy set to medium) using the Spectra operation. Two spectra of two breyite inclusions in diamond (samples 0004 and 0005) are shown in blue and purple, respectively in b and produce the best match results but do not account for the presence of additional peaks e.g., at 868/855 and 530 cm<sup>-1</sup>.

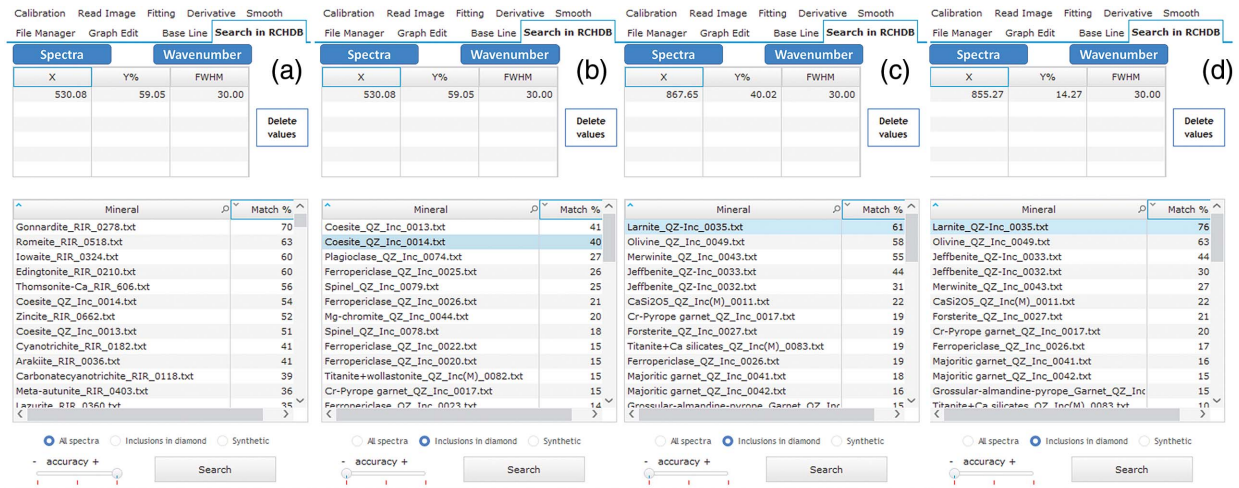
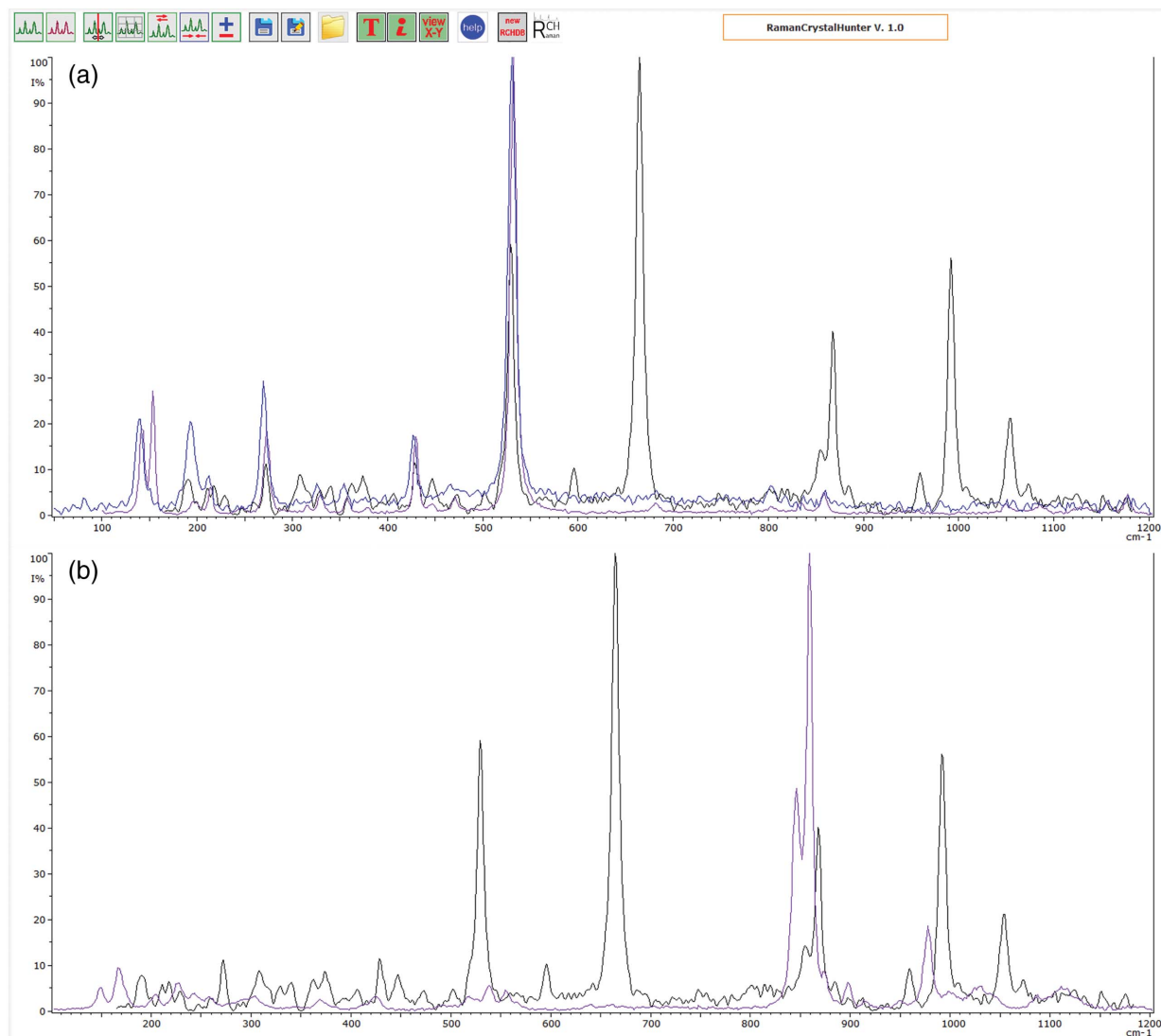


FIGURE 9. The Wavenumber search results for the 530 cm<sup>-1</sup> peak using (a) All spectra in the RCHDB and (b) Inclusions in diamond spectra in the RCHDB. The Wavenumber search results for the (c) 868 cm<sup>-1</sup> peak and (d) 855 cm<sup>-1</sup> shoulder using the Inclusions in diamond spectra in the RCHDB.



**FIGURE 10.** (a) The spectra of two coesite inclusions in diamond (samples 0013 and 0014, shown in blue and purple, respectively). (b) The spectrum of a larnite inclusion in diamond (sample 0035, shown in pink). The primary spectrum (multi-phase inclusion in diamond) is shown in black in **a** and **b**.

The spectra of both coesite (samples 0013 and 0014) and the spectrum of larnite (sample 0035) are shown in Figures 10a and 10b, respectively.

Based on the results obtained above, the user can confidently conclude that the unknown spectrum is from a multi-phase inclusion in diamond composed on breyite, coesite, and larnite. Without the sophistication of the new software we present here, it would be easy to miss the additional phases. At this point, the user can deconvolute the primary spectra as all characteristic peaks have been assigned to different minerals. For example, the user can use the Subtract operation to subtract the spectra of coesite (i.e., Fig. 10a) and larnite (i.e., Fig. 10b) from the primary (mixed) spectrum to produce a new spectrum containing signal due to only the breyite component of the inclusion. Alternatively, if there is some uncertainty as to which of the lower-intensity peaks belong to which minerals (breyite, coesite, or larnite), the user can use the Replace operation to remove only the characteristic (most intense)

peaks in the spectra of coesite and larnite from the primary spectrum to produce a new spectrum of breyite.

Identifying mixed spectra offers valuable, yet often overlooked, insights for Earth Science research. For instance, the co-occurrence of the three phases described in the example above provides information about how the composition of the diamond-forming fluids/melts, from which this diamond crystallized, evolved from a high Ca/Si ratio to a lower Ca/Si ratio, thereby enhancing our understanding of the diamond formation process (Zhang et al. 2024).

## IMPLICATIONS

The principal goal of the RCHDB Project is to develop a free, easily accessible, and user-friendly software program for processing and analyzing Raman spectra of minerals and related geological materials. For these reasons, the RCH software includes

several methods for pre-processing (e.g., smoothing, baseline correction, calibration, etc.), fitting, and spectrum identification integrated into a simple graphical interface. The RCH software and the RCHDB database will be regularly updated through collaboration with users of the software. Therefore, we strongly encourage users to: (1) submit suggestions about how to improve (or add) operations to the RCH software, and (2) submit Raman spectra of minerals (or related geological materials) that are not currently included in the RCHDB database and/or the Raman spectra of minerals that are of higher quality than those included in the database. For more information about how to submit suggestions and spectra, refer to the RCH user guide (Online Materials<sup>1</sup>), which can also be downloaded from <https://www.fabrizionestola.com/rch>.

The RCH software offers an effective means by which a user can perform all operations on Raman spectra that are typically required for Earth Science applications. No other open-access program can perform standard pre-processing, fitting, and spectrum identification operations as efficiently as RCH. It follows that RCH offers a unique solution to users who wish to increase the efficiency of data processing and interpretation. The RCHDB was constructed specifically to allow for accurate spectrum matching (identification) results and thus does not suffer from inadequacies related to poorly described and/or low-quality data, as is the case for other databases in different fields of spectroscopy. The RCHDB Project was constructed in the hope that RCH users will contribute to the RCHDB and thus aid in the ongoing development of the RCH software and, in general, improve the applicability of Raman spectroscopy in Earth Sciences.

#### ACKNOWLEDGMENTS AND FUNDING

M.S. kindly sponsored the entire RCH project. F.N. acknowledges funding from the European Union (ERC, INDIMEDEA, Starting Grant No. 307322). M.C.D. and M.G.P. acknowledge funding from the European Union (ERC, INHERIT, Starting Grant No. 101041620). R.S. acknowledges funding from Project SID 2021 (University of Padova). C.M. acknowledges funding from the HYPERION project supported by the European Union Horizon 2020 Research Program (Grant Agreement no. 821054).

#### REFERENCES CITED

- Angel, R.J., Alvaro, M., and Nestola, F. (2022) Crystallographic methods for non-destructive characterization of mineral inclusions in diamonds. *Reviews in Mineralogy and Geochemistry*, 88, 257–306.
- Anzolini, C., Prence, M., Alvaro, M., Romano, C., Vona, A., Lorenzon, S., Smith, E.M., Brenker, F.E., and Nestola, F. (2018) Depth of formation of super-deep diamonds: Raman barometry of CaSiO<sub>3</sub>-walsstromite inclusions. *American Mineralogist*, 103, 69–74, <https://doi.org/10.2138/am-2018-6184>.
- Brenker, F.E., Nestola, F., Brenker, L., Peruzzo, L., and Harris, J.W. (2021) Origin, properties, and structure of breyite: The second most abundant mineral inclusion in super-deep diamonds. *American Mineralogist*, 106, 38–43, <https://doi.org/10.2138/am-2020-7513>.
- Candeloro, P., Grande, E., Raimondo, R., Di Mascolo, D., Gentile, F., Coluccio, M.L., Perozziello, G., Malara, N., Francardi, M., and Di Fabrizio, E. (2013) Raman database of amino acids solutions: A critical study of extended multiplicative signal correction. *Analyst*, 138, 7331–7340, <https://doi.org/10.1039/c3an01665j>.
- Carey, P.R. and Heidari-Torkabadi, H. (2015) New techniques in antibiotic discovery and resistance: Raman spectroscopy. *Annals of the New York Academy of Sciences*, 1354, 67–81, <https://doi.org/10.1111/nyas.12847>.
- Day, M.C., Pamato, M.G., Novella, D., and Nestola, F. (2023) Imperfections in natural diamond: The key to understanding diamond genesis and the mantle. *La Rivista del Nuovo Cimento della Società Italiana di Fisica*, 46, 381–471, <https://doi.org/10.1007/s40766-023-00045-6>.
- DeCarlo, T.M., Comeau, S., Cornwall, C.E., Gajdzik, L., Guagliardo, P., Sadekov, A., Thillainath, E.C., Trotter, J., and McCulloch, M.T. (2019) Investigating marine bio-calcification mechanisms in a changing ocean with *in vivo* and high-resolution *ex vivo* Raman spectroscopy. *Global Change Biology*, 25, 1877–1888, <https://doi.org/10.1111/gcb.14579>.
- Dubessy, J., Caumon, M.C., and Rull, F. (2012) Preface. *European Mineralogical Union Notes in Mineralogy*, 12, 15–16.
- El Mendili, Y., Vaitkus, A., Merkys, A., Gražulis, S., Chateigner, D., Mathevet, F., Gascoin, S., Petit, S., Bardeau, J.F., Zanatta, M., and others. (2019) Raman Open Database: First interconnected Raman-X-ray diffraction open-access resource for material identification. *Journal of Applied Crystallography*, 52, 618–625, <https://doi.org/10.1107/S1600576719004229>.
- Genzel, P.T., Pamato, M.G., Novella, D., Santello, L., Lorenzon, S., Shirey, S.B., Pearson, D.G., Nestola, F., and Brenker, F.E. (2023) Geobarometric evidence for a LM/TZ origin of CaSiO<sub>3</sub> in a sublithospheric diamond. *Geochemical Perspectives Letters*, 25, 41–45, <https://doi.org/10.7185/geochemlet.2313>.
- Laetsch, T.A., and Downs, R.T. (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. *Program and Abstracts of the 19th General Meeting of the International Mineralogical Association in Kobe, Japan*, p. 8–25.
- Lafuente, B., Downs, R.T., Yang, H., and Stone, N. (2015) The power of databases: the RRUFF project. In T. Armbruster and R.M. Danisi, Eds., *Highlights in Mineralogical Crystallography*, 1–30. De Gruyter.
- Nasdala, L., Brenker, F.E., Glinemann, J., Gasparik, T., Harris, J.W., Stachel, T., and Reese, I. (2004) Residual pressure and strain around mineral inclusions in diamonds. *European Journal of Mineralogy*, 15, 931–935, <https://doi.org/10.1127/0935-1221/2003/0015-0931>.
- Nestola, F., Prence, M., Nimis, P., Sgreva, N., Perritt, S.H., Chinn, I.L., and Zaffiro, G. (2018) Toward a robust elastic geobarometry of kyanite inclusions in eclogitic diamonds. *JGR Solid Earth*, 123, 6411–6423, <https://doi.org/10.1029/2018JB016012>.
- Pasteris, J.D. and Beyssac, O. (2020) Welcome to Raman spectroscopy: Successes, challenges, and pitfalls. *Elements*, 16, 87–92, <https://doi.org/10.2138/gselements.16.2.87>.
- Pasteris, J.D., Wopenka, B., and Seitz, J.C. (1988) Practical aspects of quantitative laser Raman microprobe spectroscopy for the study of fluid inclusions. *Geochimica et Cosmochimica Acta*, 52, 979–988, [https://doi.org/10.1016/0016-7037\(88\)90253-0](https://doi.org/10.1016/0016-7037(88)90253-0).
- Pearson, D.G., Brenker, F.E., Nestola, F., McNeill, J., Nasdala, L., Hutchison, M.T., Matveev, S., Mather, K., Silversmit, G., Schmitz, S., and others. (2014) Hydrous mantle transition zone indicated by ringwoodite included within diamond. *Nature*, 507, 221–224, <https://doi.org/10.1038/nature13080>.
- Press, W.H. and Teukolsky, S.A. (1990) Savitzky-Golay smoothing filters. *Computers in Physics*, 4, 669–672, <https://doi.org/10.1063/1.4822961>.
- Reisner, L.A., Cao, A., and Pandya, A.K. (2011) An integrated software system for processing, analyzing, and classifying Raman spectra. *Chemometrics and Intelligent Laboratory Systems*, 105, 83–90, <https://doi.org/10.1016/j.chemolab.2010.09.011>.
- Savitzky, A. and Golay, M.J.E. (1964) Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36, 1627–1639, <https://doi.org/10.1021/ac60214a047>.
- Smith, E.M., Krebs, M.Y., Genzel, P.T., and Brenker, F.E. (2022) Raman identification of inclusions in diamond. *Reviews in Mineralogy and Geochemistry*, 88, 451–473.
- Song, D., Chen, Y., Li, J., Wang, H., Ning, T., and Wang, S. (2021) A graphical user interface (NWUSA) for Raman spectral processing, analysis and feature recognition. *Journal of Biophotonics*, 14, e202000456, <https://doi.org/10.1002/jbio.202000456>.
- Vengatesh, B.M. (2013) Implementation of Spectral Angle Mapper (SAM) Algorithm on a Graphic processing unit (GPU). *International Journal of Engineering Development and Research*, 1, 122–124.
- Walter, M.J., Thomson, A.R., and Smith, E.M. (2022) Geochemistry of silicate and oxide inclusions in sublithospheric diamonds. *Reviews in Mineralogy and Geochemistry*, 88, 393–450.
- Wojdyr, M. (2010) Fityk: A general-purpose peak fitting program. *Journal of Applied Crystallography*, 43, 1126–1128, <https://doi.org/10.1107/S0021889810030499>.
- Zhang, Q., Timmerman, S., Stachel, T., Chinn, I., Stern, R.A., Davies, J., Nestola, F., Luth, R., and Pearson, D.G. (2024) Sublithospheric diamonds extend Paleoproterozoic record of cold deep subduction into the lower mantle. *Earth and Planetary Science Letters*, 634, 118675, <https://doi.org/10.1016/j.epsl.2024.118675>.

MANUSCRIPT RECEIVED MAY 13, 2024

MANUSCRIPT ACCEPTED AUGUST 2, 2024

ACCEPTED MANUSCRIPT ONLINE AUGUST 12, 2024

MANUSCRIPT HANDLED BY SUZETTE TIMMERMAN

#### Endnotes:

<sup>1</sup>Deposit item AM-25-49457. Online Materials are free to all readers. Go online, via the table of contents or article view, and find the tab or link for supplemental materials.