RamanCrystalHunter - *Version 1.0*

-User Guide-

The *RamanCrystalHunter V1.0* software (RCH), including the *RamanCrystalHunter Database* (RCHDB), was conceived and designed by Prof. Fabrizio Nestola, Professor at the Department of Geosciences at the University of Padua, and by Dr. Ivano Rocchetti.

This software is free to download and available to anyone using Raman spectroscopy to study the properties of minerals and related synthetic compounds. Collaboration with researchers from around the world is required to expand the database of Raman spectra and thus improve the applicability of the software. We invite researchers to submit Raman spectra of minerals and/or related synthetic compounds that are not currently included in the database by accessing the submission portal at https://www.fabrizionestola.com/rch. Submitted spectra must be in .asc, .txt or .csv format and accompanied by the following minimum information: mineral name, provenance, bibliographic and analytical references including the author's names. Submitted spectra will be collected, reviewed, and added to the software during periodic updates to the RCHDB.

Acknowledgements

We express our thanks to Prof. Elisa Quintarelli, Dr. Sara Migliorini and Dr. Giacomo Marchioro of the Department of Computer Science (University of Verona) for information related to the automatic search (matching) algorithm used in RCH to identify unknown Raman spectra by comparison with spectra in the RCHDB.

Fabrizio Nestola acknowledges funding from the European Union (ERC, INDIMEDEA, Starting Grant No. 307322). Maxwell C. Day and Martha G. Pamato acknowledge funding from the European Union (ERC, INHERIT, Starting Grant No. 101041620). Raffaele Sassi acknowledges funding from Project SID 2021 (University of Padova). Claudio Mazzoli acknowledges funding from the HYPERION project supported by the European Union Horizon 2020 Research Program (Grant Agreement no. 821054).

Copyright

We are committed to continually improving and developing the RCHDB and the RCH software, to help us do this, we strongly encourage users to provide feedback at https://www.fabrizionestola.com/rch.

Terms and Conditions of Use

The following terms and conditions of use apply to both the RamanCrystalHunter software (RCH) and the RamanCrystalHunterDatabase of Raman spectra (RCHDB).

Users must agree to these terms and conditions of use before:
[1] downloading RCH and RCHDB.
[2] submitting data (Raman spectra) to the RCHDB Project.
[3] joining the RCH Registry.

Development of RCH V1.0 software and encoding of spectral data in the RCHDB was done by Dr. Ivano Rocchetti and Claudio Bendazzoli. Prof. Fabrizio Nestola, Prof. Claudio Mazzoli, Prof. Martha G. Pamato, Prof. Raffaele Sassi and Prof. Davide Novella own all rights reserved to RCH; Dr. Ivano Rocchetti and Dr. Claudio Bendazzoli own the intellectual property rights for RCH.

By accessing the RCHDB, users agree to use the database and RCH software for informational and/or research purposes only. It is expressly prohibited to sell, distribute, publish, circulate or commercially exploit the data contained in the RCHDB or data produced or generated using the RCH software without the express written permission of the data owners and the owners of the RCH and RCHDB property rights. The use of data from the RCHDB and/or RCH software must not convey any right of ownership, title or interest, nor any security title or other interest in any intellectual property right relating to the database and the software, nor in any copy of the database (or data contained within the database) and/or software.

The RCH software and RCHDB database of Raman spectra are open-source and free to download on a strictly non-commercial basis and use of RCH and/or RCHDB does not require a license. Therefore, to ensure further development and support, users of RCH and RCHDB must acknowledge their use in any form of publication by reference to the RCH publication, Nestola et al. 2024 which can be accessed at https://www.fabrizionestola.com/downloads.

Upon download of RCH and RCHDB, users are strongly encouraged to sign-up and join the *RCH Registry* such that they receive news about updates to the software and database. Users are asked to submit their [1] first and last name, [2] email address, and [3] affiliation. All personal data will be kept completely confidential within the secure RCH Registry database and will not be distributed or modified in any way and/or used for any other purpose than to contact RCH users about updates to the RCH and RCHDB. If a member of the RCH Registry chooses to submit data (spectra) to the RCHDB project, they are hereby in agreement that their data will be added to RCHDB and thus accessible to all users of RCH. The authors reserve the right to modify submitted data to ensure it is formatted correctly according to the RCHDB style.

Disclaimer: While every effort is made to ensure that the RCH software is free of bugs and errors, people use it at their own risk. No responsibility whatsoever is taken for either incorrect results or for any physical, mental or other damage arising from use of RCH or RCHDB or from errors in this manual.

Although RCH and several corresponding authors are affiliated with the University of Padua, the software is hosted on the private website of Fabrizio Nestola and published in an independent peer-reviewed journal. It follows that the University of Padua takes no

responsibility whatsoever for either incorrect results or for any physical, mental or other damage arising from use of RCH or RCHDB or from errors in this manual.

Installing RamanCrystalHunter V1.0

Minimum system requirements: Windows 10, 4 Gb RAM, 1920x1080 pixel monitor resolution. If RCH is launched using different monitor resolutions part of the user interface may be cut off. One can change their monitor resolution following the path Settings/System/Display/Scale & layout and then by adjusting the *Display resolution*.

The RCH installation file (RamanCrystalHunter64_Setup.EXE) and the RCHDB file (RCHDB.txt) can be downloaded from https://www.fabrizionestola.com/downloads.

Run the RamanCrystalHunter64_Setup.EXE file to install RCH, once this is done, the path *C:/RamanCrystalHunter* will be created and a shortcut will be automatically added to the user's desktop. Once RCH is launched, click on 

to automatically update the RCHDB (upload the RCHDB.txt file to RCH). After doing this the search menu "Search in RCHDB" becomes active. The RCHDB only needs to be loaded once when RCH is opened for the first time by the user. When updated (revised) databases (RCHDB.txt files) become available, the user will need to reload the updated database (RCHDB.txt file) as described above.

Important Notes: Depending on the security settings of the browser used to download the installation files, the browser may block the download of the RamanCrystalHunter64_Setup.EXE and/or RCHDB.txt files. This is easily resolved by clicking "keep" in the browser download dropdown menu.

Depending on the settings of the user's anti-virus software (e.g., Windows Defender), permission may have to be granted in order to complete installation and/or to launch RCH *V*1.0. Depending on the security settings in the user's operating system, users may have to modify their security/permissions settings by granting permission to the RCH program such that it can access files on the user's computer. This will allow RCH to open spectrum (.txt) files stored anywhere on the user's computer.

If users experience problems downloading or launching the RCH software (or RCHDB) that are associated with security/permissions settings, please contact the authors Fabrizio Nestola, Yanjuan Wang, and/or Maxwell C. Day.

Uninstalling RamanCrystalHunter *V*1.0

It is possible to completely uninstall RCH by running the WDUNINST.EXE program located in the *C:/RamanCrystalHunter* folder. This uninstaller deletes all the data associated with the source location, if you need to keep data, make a copy of the *C:/RamanCrystalHunter/Spectra* folder prior to uninstalling RCH.

Before installing a new version of RCH, ensure that the previous version has been uninstalled and all component files have been deleted.

Using RamanCrystalHunter *V*1.0

The RCH *V*1.0 software allows you to view the Raman spectrum of a mineral in the file formats .asc, .txt, and .csv. The program also allows you to perform operations on the same file to modify the spectrum and an automatic search for similar spectrum can be performed, similar spectra are selected from the RCHDB.

Main Window

When RCH *V*1.0 starts, the following main window appears (Fig. 1), this window can be moved , minimized , or closed using the buttons in the top-right corner.

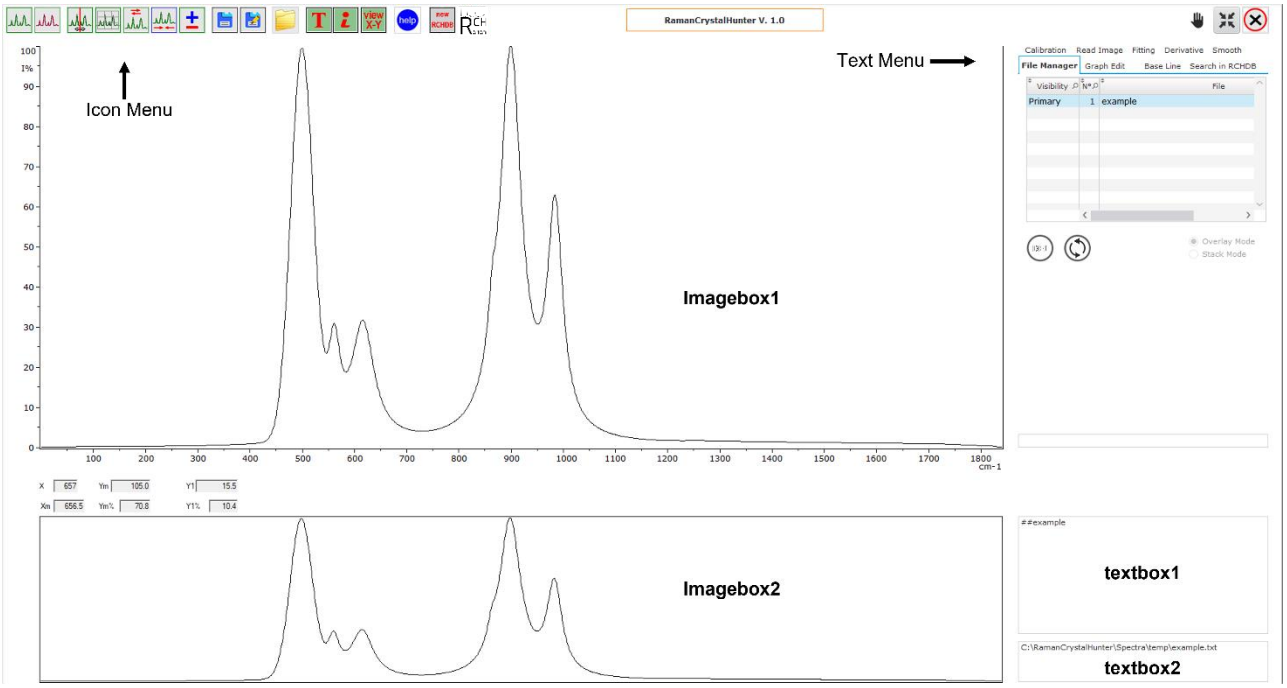

Fig. 1. The graphical user interface of RCH *V*1.0. Once opened, the RCH user-interface window is maximized, to access the taskbar without completely minimizing the program, move the RCH window by dragging it to the desired position.

Opening Files

The spectrum file to be opened (processed) must be in .asc, .txt or .csv format and **it must be saved on the hard disk and not on external devices**. This file is termed the “primary file” and is opened by clicking on , by doing so the spectrum will be displayed in imagebox1 and in imagebox2 (Fig. 1), imagebox2 provides a preview of the spectrum after any given operation is performed during subsequent processing. To replace the primary file, simply click on the same icon and select a different spectrum file.

Six boxes appear under imagebox1 (Fig. 1) when a primary file is loaded which provide the following information corresponding to the position of the mouse pointer in imagebox1:

- **X:** X coordinate detected on the spectrum (cm^{-1}) displayed in imagebox1, **X** is also detected in imagebox2.
- **Xm:** X coordinate of the mouse pointer in imagebox1 (cm^{-1}).
- **Ym:** Y coordinate of the mouse pointer in imagebox1 in the Y scale of the graph (arbitrary units or absorption units related to the spectrum data).
- **Ym%:** Y coordinate percentage of the mouse pointer in imagebox1 in the Y scale of the graph displayed as Y%
- **Y1:** indicates the raw Y coordinate of the spectrum (arbitrary units) corresponding to the position of the mouse pointer (X coordinate).
- **Y1%:** indicates the ordinate of Y1 expressed as a percentage.

To clarify, if the mouse pointer is placed exactly on any part of the spectrum in imagebox1, then **Ym** = **Y1** and **Ym%** = **Y1%**. If the mouse pointer moves to the maximum **Ym** value of the spectrum, **Y1%** = 100 and for the minimum **Ym** value, **Y1%** = 0.

The X coordinate (**X**) is also enabled when you move the mouse over imagebox2 to allow for a more accurate zoom range to be selected (see **Zoom** section). The values of the abscissas are interpolated when the primary file is opened so that they are expressed as integer values (as discrete X coordinates) where $X_{i+1} - X_i = 1\ cm^{-1}$. For this reason, it is possible to view (for the primary file only) a table of the interpolated X and Y coordinates by clicking on 

. This table can be closed by re-clicking on this icon.

***Important Note*:** when the number of pairs of coordinates x_i, y_i that define an input spectrum is greater than the number of pixels comprising imagebox1, there is a certain imprecision with respect to what has been described above as each pixel on the monitor corresponds to more than one coordinate defining the spectrum in the input file. However, when operating under zoom conditions (see **Zoom** section), this inaccuracy decreases (with increasing zoom) until it disappears where each pixel comprising imagebox1 corresponds to only one coordinate value in the spectrum data. This inaccuracy is inverted when the zoom is increased such that the no. coordinates/no. pixels > 1 where each value of X corresponds to several values of Y, in this case **Xm** and **Ym** provide the coordinates of the intermediate points which are plotted as straight-line segments.

The name of the primary file is displayed in the File Manager menu (Fig. 1). Here the icons 

= *Delete* and  = *Visible/Invisible* are shown but are not active as they are not applicable to the primary spectrum (file), they can only be applied to additional spectra added later (see **Add Files** section). Below the File Manager window, there are two textboxes (Fig. 1). In *textbox1*, information related to the mineral spectrum is reported. This information can be changed by clicking on 

(see **Information Storage** section) and filling in the relevant fields. The corresponding file directory for the selected spectrum (file) is indicated in *textbox2*.

When the primary file is opened, if it does not already exist, the temporary (*temp*) folder is created in C:/RamanCrystalHunter/Spectra with the file example.txt inside; the *temp* folder is a service folder inside which all the files that are obtained from processing the primary file are put. In the next work session with the RCH program, if the *temp* folder is not deleted beforehand, all the files in it will be retained and some of them automatically replaced, so it is important for the operator to keep this in mind and take the necessary precautions. The file example.txt represents an artificial spectrum created ad hoc.

Add Files (Spectra)

By clicking on  **,** additional spectra (files) can be loaded in addition to the primary file; each additional spectrum is listed in the File Manager menu. For additional files, the *Delete* and *Visible/Invisible* buttons (see **Opening Files** section) become active in the File Manager menu. When a primary file and at least one additional file are loaded, the *Stack Mode* and *Overlay Mode* become active, selecting the latter will overlap spectra in imagebox1 and selecting the former will stack spectra in imagebox1. When several spectra are loaded, the color of each file name (displayed in the File Manager menu) and the color of the corresponding spectrum (in imagebox1) are the same to facilitate easy visual identification of different spectra.

When multiple spectra are displayed, the X scale in imagebox1 (**X** and **Xm**) is recalculated such that it is correct for all spectra displayed in *Stack* or *Overlay Modes*. The Y scale, as represented in *Stack* or *Overlay Modes*, is recalculated using a common percentage scale and thus the mouse coordinate **Ym** is meaningless and is not shown in either mode. The coordinates **Y1** ... **Yn** and **Y1%** ... **Yn%** (for **n** spectra where **Y1** represents the primary file and **Y2**…**Yn** represent all additional files) appears for each respective spectra, and **Yn** and **Yn%** are color coded based on the color of the corresponding spectrum. When spectra are displayed in *Overlay Mode* using a common percentage Y scale and thus the mouse coordinate **Ym%** is active.

The user can open a maximum of nine files, including the primary spectrum. If you wish to perform addition/subtraction operations on two files, the limit is eight. This allows you to display the spectrum on which these operations were performed and view the addition/subtraction result.

Important Note: Additional spectra will be automatically removed from the File Manager menu when the *Graph Edit, Baseline Correction, Smooth, Fitting* and/or *Derivative* options are activated in the text menu as the associated operations act only on the primary spectrum.

Vertical Cursor

By clicking on  , and moving the mouse cursor over imagebox1, a red vertical cursor is superimposed on the spectrum/spectra in imagebox1. This vertical cursor can be moved and locked with a single click in any position to highlight the X coordinate of the points intercepted by the cursor and the X coordinate will be displayed in blue below the X axis; this is useful to quickly determine peak positions. The vertical cursor can be unlocked by

clicking the mouse button again in imagebox1. The vertical cursor can be deactivated by clicking the icon a second time. When the vertical cursor is enabled, it is not possible to activate the grid and the grid icon is disabled. If the grid is enabled, clicking on  will deactivate the grid.

Draw Grid

By clicking on  , a grid is added to imagebox1. This grid can be removed by clicking this icon again (see **Vertical cursor** section). The grid spacing is automatically generated based on the size (wavenumber range, cm^{-1}) of the spectrum and the degree of zoom (if activated).

Reverse Scale

The X-axis of a given spectrum can be reversed by clicking on  , a reversed spectrum can only be viewed temporarily as all subsequent operations (e.g., active zoom) are applied to spectra such that the lower X value (cm^{-1}) appears on the left side of imagebox1. It is therefore necessary to restore the reversion of the X-axis by clicking on the icon a second time before continuing with processing of the spectrum. While the X-axis of a spectrum is reversed, it is only possible to add additional files and activate the vertical cursor.

Zoom

To activate the zoom function along the X axis, you must indicate with the mouse pointer in imagebox2 the endpoints of the spectral range (cm^{-1}) in which you would like to zoom into. By pressing the left mouse button, you can define one endpoint, then you drag the cursor inside imagebox2, releasing the mouse at the second endpoint of the spectral range the user wishes to zoom into.

The zoom range can be determined (for the primary and additional spectra) by dragging the mouse from left to right and vice versa; if one releases the mouse outside the right-left limits of imagebox2, the zoom range displayed extends from the position of the first mouse click to the limit of imagebox2. If you click only once inside imagebox2, without moving the mouse, a predefined zoom range of 100 pixels right of point at which the mouse was clicked is generated. To deactivate the zoom function, click the icon  .

.

Smoothing

Spectrum smoothing involves the application of a filter that attenuates the high-wavenumber signal in a spectrum that is due to background noise. Therefore, smoothing effectively highlights the characteristic signals (peaks) in the spectrum. The Smooth function acts on the entire spectrum therefore, if the Zoom function is active, it is deactivated upon smoothing; if there are other spectra open in addition to the primary spectrum, these spectra are removed from the File Manager upon smoothing.

To begin smoothing, open the “Smooth” menu by selecting it from the text menu (Fig. 2).

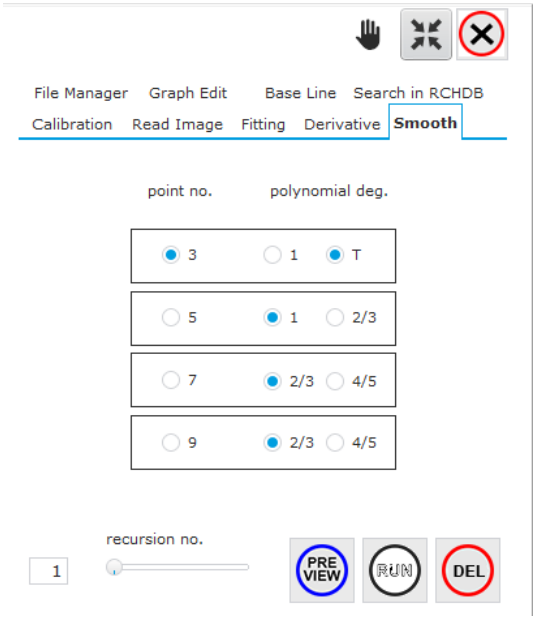

Fig. 2. The Smooth Menu

It is necessary to select the number of points with respect to the approximation polynomial (3, 5, 7 or 9) that will be calculated and, correspondingly, the degree of the polynomial. Then, one must select a recursion number (*i.e.* an integer (≤30) that represents the number of recursions performed during the smoothing calculation). One can then preview the resultant (smoothed) spectrum by clicking PRE VIEW, and the preview of the smoothed spectrum will appear in imagebox2. If the preview is satisfactory, click RUN, otherwise one can redefine the degree of smoothing (adjust smoothing parameters) or click DEL to exit the smoothing menu.

Noise reduction (smoothing) is completed by applying a Savitzky-Golay filter except when a 3-points triangle evaluation is selected (Fig. 2). For a 3-point triangle evaluation, the values of the signal are weighted as follows; for each abscissa x_i, the weight of the value y_i is doubled compared to the first neighbors y_{i-1} and y_{i+1} and therefore values corresponding to y_{i-1} to y_{i+1} are averaged.

A smoothed spectrum will automatically replace the original primary file in the File Manager and will be saved with the suffix _S in the *temp* folder (*C:/RamanCrystalHunter/Spectra/temp)*, which the user can access at any time. One can smooth a spectrum that has already been smoothed if the number of recursions is greater than that used in the initial smoothing procedure. This re-smoothed spectrum will be saved as described above and an additional suffix _S will be added to the file name.

Performing a smoothing operation before editing or calibrating a spectrum is recommended, especially for spectra that are low quality and noisy (see **Graph Edit** and **Calibration** sections).

Addition and Subtraction Operations

One can add two spectra together or subtract one spectrum from another using the addition and subtraction operations, respectively. To begin, click on  which is only active when the File Manager menu is selected. Addition and subtraction operations can be applied to any two spectra loaded in the File Manager. Once the  button is clicked, one must fill in the fields for the coefficient (weight) of both spectra, the number of each spectrum (indicated in the File Manager) that one wishes to involve in the addition or subtraction operation, and the *symbol* (**+** or **-**) of the operation to specify whether a subtraction or addition operation is to be performed. Since the addition and subtraction operations take place with reference to the spectra in which the ordinate Y values are expressed in %, resultant spectra will also have Y coordinates expressed in %.

It is possible to preview the addition or subtraction result by clicking on  . If the result is satisfactory, the user can accept it by clicking on  or reject it by clicking on  . A preview can be rejected by clicking on  before  allowing the user to iteratively adjust the addition and subtraction parameters before any operation is finalized. If you want to exit the *Addition/Subtraction* menu, you can click on  or simply open a different operations menu.

The result is displayed in imagebox1 along with the component spectra and saved in the *temp* folder (*C:/RamanCrystalHunter/Spectra/temp) with the name* “Processed_O.txt” and the File Manager menu as “Processed_O”. Resultant spectra will be saved as “Processed_O”, and subsequent operations performed on a processed spectrum are named ““Processed_O_O” and so on. It is recommended that spectrum files be saved/organized within the *temp* folder to prevent confusion when processing many different spectra.

Addition operation: this operation can be performed in two ways:

[1] ***Common Range*** - the sum of Y% values are only calculated for ranges (X values, cm^{-1}) common to both spectra. This operation highlights common peaks and weakens signal due to unwanted phenomena (e.g., background noise, different background types, *etc*.). In the event that there is no common signal, an alert window appears warning the user that the ranges have no common elements.

[2] ***Entire Range*** - this operation also takes into account spectral ranges that the two spectra do not have in common, therefore this is used for combining spectra with different ranges (X values, cm^{-1}). Signal in the ranges that the two spectra do not have in common will be attenuated proportional to the weighting coefficient selected from the spectrum that contains the uncommon range(s). This type of operation is useful for combining spectra of the same mineral obtained over different spectral ranges (cm^{-1}). If the spectra do not have consecutive ranges, then the undefined ranges will be displayed as straight-line segments. For the addition of signal across common ranges, a result will be obtained as described above. However, the best result is obtained by making sure that the ranges of each spectrum

are consecutive, that is, the maximum X value of the first spectrum corresponds to the minimum X value of the second spectrum + 1 cm^{-1}, and also by appropriately choosing the weighting coefficients. To achieve this, one can edit the input files with an external editor (e.g., blocknote) by deleting overlapping (common) segments of the spectra, *i.e.* data corresponding to identical ranges (cm^{-1}).

Subtraction operation: the subtraction operation for spectra is performed across the range of the minuend spectrum by removing signal that the subtrahend spectrum has in common with the minuend spectrum. This allows the user to eliminate signal (peaks) from the first spectrum that are due to a mineral (or contaminant) different than the mineral of interest. This is particularly useful for analysis of mineral inclusions when one wishes to remove signal due to the mineral encapsulating the inclusion of interest.

Saving Data

Only spectra that correspond to the primary file are automatically saved following processing of the primary file (spectrum) as specified above, one can access these files in the *temp* folder. One can save the primary file manually by clicking on “Save”  or “Save As”  . As each file is saved in the *temp* folder with a particular suffix based on the type of operation that is performed (e.g. _O.txt), it is recommended that one renames files of interest before completing subsequent operations so that files are not overwritten in the *temp* folder. No data (files) are deleted from the *temp* folder upon closing or reopening RCH.

The user can open the source location of the primary file by clicking on  . If the primary file has been processed and thus re-saved in the *temp* folder, clicking on this icon will open its new location (i.e. the *temp* folder).

Information Storage

By clicking on  , a text box appears where one can add, modify, and store information about the mineral and/or spectrum. Information entered in this box will be stored, upon saving, in the first lines of the corresponding file and will appear in the textbox1 (Fig. 1).

***Important Note*:** each line entered in the text box must be preceded by a double number sign “##”, otherwise the information entered by the user will not be saved and thus will not appear upon subsequent reopening of the file.

Graph Edit

The Graph Edit menu can be accessed in the text menu and contains the following operations that may be applied to the primary file regardless of whether the Zoom function is active or not.

Replace: one can delete a region of the spectrum and replace it with a 5-point Bezier curve defined by the points *P*1-*P*5. The endpoints of the region one wishes to delete (*P*1(X1) to

*P*5(X5)) are selected by double clicking at each endpoint in the order X1<X5. Once the endpoints have been selected, three colored reference points (bridles *P*2(X2)-*P*4(X4)) appear on a line segment defined by the selected endpoints *P*1 and *P*5, then one can move the bridles (points *P*2(X2)-*P*4(X4)) to change the shape of the Bezier curve. One can also simulate noise in the Bezier curve by adjusting the *Noise Level* and the *Noise Frequency*. When one is satisfied with the Bezier curve, they can click on 

to complete the replacement and the new edited spectrum will be added to the File Manager menu and saved automatically in the *temp* folder with the suffix _E. Clicking on 

before 

will delete the Bezier curve and restore the original signal in the selected region. If the zoom is active, the Replace operation will act on the zoom region of the spectra in imagebox1 by deactivating the zoom and then making the replacement.

Subtract: using this operation, one can subtract the area between the Bezier curve (defined by P1(X1) to P5(X5)) and the straight-line segment of the spectrum. If the curve is drawn above the reference segment (e.g., a peak), the curve is subtracted, if it is drawn below the reference segment, the curves are summed. This function is useful for attenuating/emphasizing particular signals (peaks) or for locally improving the results of baseline corrections and addition and subtraction operations.

Clicking on 

before 

restores the original primary spectra by canceling the Bezier curve. To make the degree of noise along a peak and the baseline compatible, the user can model the curve with the Replace operation, add noise, and then select Subtract and click on 

. After processing, the edited spectrum will be added to the File Manager menu and saved automatically in the *temp* folder with the suffix _E.

Important Note: the shape of the Bezier curve may result in a small discrepancy with respect to the selected endpoints (*P*1(X1) to *P*5(X5). This occurs when the zoom function is active and depends on whether the X coordinate is discrete. However, the final resultant spectrum is always calculated correctly.

Add function: using this operation, one can add peaks to a primary spectrum using Gaussian, Lorentzian or Voigt functions (peaks). To add a function (peak) to a spectrum, one must first select the vertex position of the peak by clicking once inside imagebox1, this position will be recorded as X and Y coordinates in the “Add function” menu, to change this vertex position click elsewhere in imagebox1. Next, the type of function must be selected, Gaussian, Lorentzian, or Voigt and the FWHM (Full-width-at-half-maximum) must be indicated by the user. For Gaussian and Lorentzian functions, one must define sigma and gamma, respectively, and for Voigt functions one must define sigma and beta. For Voigt functions, beta is defined as ≤ 1 and > 0; as beta approaches 1, the peak is more Lorentzian and when beta approaches 0, the peak is more Gaussian. The default beta value is set to 0.70 as this produces peaks shapes closest to those observed in experiment. The added function takes into account the Y coordinates of the spectrum to which the function is being added. For example, if a Gaussian is added at a position X whose vertex has the value Y = 1000 and the spectrum has the value Y = 200 for the same point X, then a Gaussian peak

with a vertex at Y = 800 is added to the spectrum and all points (Y values) near the peak are increased by the corresponding calculated Y value of the peak.

If you need to modify the parameters of the added peak, you will need to reload the original primary spectrum as it is not possible to preview added peaks before they are added and the primary file is changed. The selected vertex position of the peak must have a Y coordinate greater than the corresponding region of the spectrum with the same X coordinate otherwise, an error will be reported to the user. Once the position (X coordinate) and vertex (Y coordinate) of the peak are selected and the function (peak) type and corresponding constants have been set, click RUN to generate the resultant spectrum in imagebox1. The suffix _E will be added to the resultant file name and will be automatically saved to the *temp* folder.

Baseline Correction

Using this function one can perform a baseline correction by eliminating the area between the spectrum and the X axis which is mainly due to fluorescence phenomena. To ensure the spectrum matching procedure (using the RCHDB) is optimized, it is often necessary to highlight only the characteristic peaks (peaks representative of the mineral) and eliminate (or minimize) peaks that are due to background noise and/or contamination (e.g., by the encapsulating mineral). For this reason, a manual baseline procedure is used instead of an automatic procedure.

A baseline correction can be activated by clicking on the *Base Line* tab in the text menu. First, use the mouse pointer to select the first endpoint of the new baseline. Clicking on the arrows ⇦ ⇨ allows you to move a selected endpoint left or right along the X-axis. Once the position of a point is satisfactory, one can add the point to the baseline by clicking on ADD after which the coordinates of the point will appear in the table in the *Base Line* menu. Once the position of the second point has been selected, a green baseline will appear that intersects both points. Selection of the third point produces parabola baseline that intersects these three points. Selection of a fourth point, and all points thereafter, produces a cubic spline baseline that intersects all baseline points. This baseline is easily optimized by selecting each point in the table and adjusting the X-coordinate of that point as described above. The cubic spline defining the baseline is connected to the endpoints with parables which have the same first-order derivative as the selected endpoints and the same vertices as the abscissa as the endpoints of the spectrum. If you click several times in a row, without adding any of the generated baseline points, all points will remain visible until one clicks ADD , doing this will add only the last selected point and delete all points previously selected.

The selected points can be deleted by highlighting the corresponding line in the table in the *Base Line* menu and then by clicking on DEL. Once all baseline points have been added, click on RUN to subtract the area below the new baseline (green line), the primary spectrum will be replaced (in imagebox1 and imagebox2) with the baseline corrected spectrum and will be renamed in the File Manager with the suffix _N and automatically saved

in the *temp* folder; if necessary, this procedure can be repeated. If a segment of the user-defined baseline is defined by negative Y coordinates (*i.e.* if it crosses the X-axis), that negative portion of that baseline segment is not incorporated in the subtraction calculation and thus will not result in a new primary spectrum that shows negative intensity at any coordinate(s) along the X-axis. One can exit the *Base Line* menu without clicking RUN or delete all selected baseline points by opening the File Manager menu or any other tab in the text menu.

Read Image

Using this function one can transform an image of a spectrum (in .png or .jpg format) into an .txt file that can be uploaded and processed by the RCH software or any other spectrum processing software. This function reads the black pixels that define the spectrum so it is necessary that the trace of the spectrum in the image file is black and saved in black and white, 1-bit file format (B&W 8-bit file format will not work), otherwise the program reports an error. Prior to uploading the image file, the user must remove all aspects of the image not related to the spectrum, *i.e.* the X and Y axes, grids, labels, etc. A spectrum with excessive line thickness may result in background noise that can be eliminated by smoothing. It is important that the curve (spectrum) in the image file has no discontinuity.

To read-in an image file, click on  and choose the .png or .jpg file to open; a thumbnail of the open image appears to the right of the icon in the Read Image menu. A name and the directory path (both of which can be changed) for the converted .txt file will appear in the File name to save box. To ensure the new spectrum and corresponding .txt file are generated correctly it is necessary to specify the minimum and maximum X coordinates (cm^{-1}) of the spectrum in the input image. By clicking RUN, the spectrum will appear in imagebox1 and in the *File Manager* menu with the name defined by the user and an .txt file will be automatically saved in the directory specified by the user. If no directory is specified the .txt file will be saved in the same location that the input image was uploaded from. The Y scale of the generated spectrum is expressed in percentage therefore, Y1 = Y1% (see **Opening Files** section).

Fitting

The Fitting function allows one to fit signal in a Raman spectrum with component functions and evaluate the properties of those functions (peaks) that comprise the spectrum of interest.

***Criteria for choosing the fitting interval*:**
The selected fitting interval (zoom) should be limited to a few hundred cm^{-1} since the calculation times are relatively long (depending on the signal-to-noise ratio). For this reason, the user is given a warning if a large fitting interval is selected, however, this can be ignored. The endpoints of the fitting interval must lie on the baseline or correspond to the minimum Y coordinates of the spectrum if a baseline correction has not been completed. However, we recommend that spectra are baseline corrected prior to fitting if the user wishes to compare peak intensities amongst different spectra. The endpoints of the selected range

must contain peaks that are relatively less intense than those that the user wishes to fit. If the criteria described above are not followed, the fitting may result in an error or may not converge. In any case, the fitting function will highlight discrepancies in the first and last peaks. If an error is reported during the fitting procedure the user must close and then reopen RCH and repeat the fitting procedure.

After opening the Fitting menu, one must define the fitting interval using the Zoom function (see **Zoom section**) and then select the degree of initial smoothing (1 by default), the absolute value of noise in the spectrum (Ynoise), which can be determined by measuring (averaging) the height (Ym) of a peak (or several representative peaks) due to noise. One must also select the beta value of the Voigt function (β-Voigt) between 0 (Gaussian) and 1 (Lorentzian). During the first fitting attempt, it is recommended that one retains the default fitting parameters (zoom = 1, Ynoise = 0, and β-Voigt = 0). If the spectrum contains significant noise, one can increase the Ynoise parameter. Once the spectrum (or region of a spectrum) that one wishes to fit has been selected, click on PRE VIEW to generate a preview of the fit in imagebox2. The fitting result is then confirmed by clicking RUN, or one can click DEL to delete the preview and fitting parameters and to start a new fitting process or switch to other operations. If the user is attempting to fit a noisy spectrum, smoothing will often be required such that RCH does not generate an unrealistically large number of peaks in an attempt to fit all peaks associated with the noise. One can also remedy this issue by determining the degree of noise (Ym, see above) and then entering this value into the Ynoise field.

During the fitting procedure, signal that may be misinterpreted as background noise is eliminated in accordance with the Ynoise value defined by the user. The resultant spectrum after the smoothing correction has been applied is shown in blue, functions (peaks) used to fit the spectrum are shown in green, the resultant fit (i.e. the sum of all added functions) is shown in red. The parameters of each function (green peaks) used to produce the fit (e.g., X, Y, FWHM, and Area, Fig. 3) are calculated and may deviate significantly from the corresponding center (position) of the observed peak, especially for the relatively less intense peaks in the spectrum of interest. In the fitting table (Fig. 3), the column labelled “Ok” allows the user to remove individual components functions from the fit by clicking on the corresponding blue check mark or using the *select/deselect* all button. Once functions have been selected or deselected the user must update the fit by clicking on M-REFINE. The *select/deselect* button is particularly useful when fitting spectra that contain significant amounts of noise. The user can deselect all functions (many of which may correspond to peaks associated with noise) and then select (activate) only those that correspond to characteristic peaks. After updating their selections by clicking M-REFINE, the user can then continue refining functions that fit only characteristic peaks.

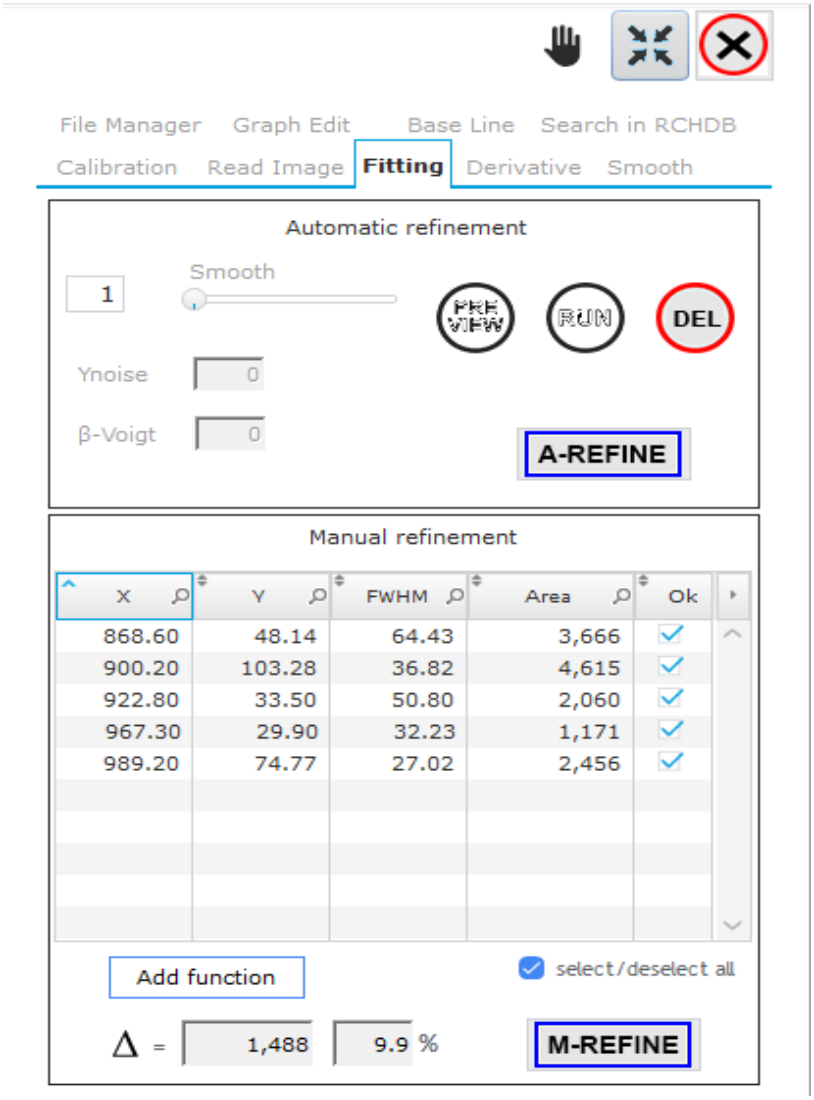

Fig. 3. Automatic and Manual Fitting menu

Important Note: As described above, fitting smaller intervals is recommended as this will decrease the probability that *peak broadening* occurs. Here, for complex spectra with many peaks, component functions may become unrealistically broad to account for peak intensity across large intervals.

Once the fitting parameters have been selected and the fit has been confirmed by clicking RUN, one can refine the fit (refine the parameters of all component functions used in the fit) by clicking A-REFINE. Generally, it is sufficient to repeat the refinement one to three times, but more iterations of refinement may be required for fitting intervals that contain many peaks or relatively broad peaks that are composed of several component functions. If the fitting result is not satisfactory, *i.e.* there is poor agreement between the red and blue curves and/or a high delta value, it is recommended that the user change the smoothing parameter and/or the β-Voigt parameter. It may also be useful to fit fewer peaks at a time (*i.e.* select smaller fitting intervals using the zoom operation). It is important to remember that the fitting procedure is approximate as the X coordinates of the spectrum are discrete.

Therefore, even after the refinement of the parameters of the functions (fits), there may be some degree of discrepancy between the red and blue curve. This discrepancy is accentuated at the extremes of the fitting interval and thus is often observed in the fits of the first and last peak.

When the automatic fitting method leads to an unsatisfactory result or diverges (e.g., *peak broadening* or inconsistent peak shifting occurs), it is possible to refine the fit manually by adjusting the parameters of each function (e.g., X, Y, and FWHM) or by adding functions using the “Add function” button. Any adjustments made to functions in the fitting table are applied to the fit in imagebox1 by clicking on M-REFINE. When adding functions, the user must enter the parameters before they update the fit by clicking on M-REFINE. After a given peak parameter has been manually adjusted or a peak has been added, the user can resume automatic refinement (by clicking on A-REFINE) or continue with manual refinement. The manual fitting method is recommended when fitting only one or two peaks and the automatic fitting method is recommended when fitting larger intervals that contain many (more complex) peaks. Note that only A-REFINE activates the automated fitting algorithm, M-REFINE simply applies (updates) the function parameters modified by the user. Automatic refinement will continue to act on deselected functions but the fitting (and Δ) calculation involves only selected functions.

One can quantify the quality of a fit using the Δ parameter (Fig. 3) which is expressed as absolute and percentage values calculated based on deviation (X^2) between the fit and spectrum (red and blue lines). Smaller Δ parameters represent higher quality fits. The Δ parameter is activated when the *A*- or *M-refine* buttons are clicked. To exit the Fitting menu click on DEL.

By using the dropdown menu beside the “Ok” column in the “Manual Refine” table (Fig. 3), the user can export the parameters of all component peaks (functions). Users can export peak parameters directly in MS Excel, MS Word, and XML file formats. Users can also copy and/or paste all or individual rows such that peak parameters can be opened in other software programs.

Important Note: It is recommended that the β-Voigt parameter be refined by iteratively adjusting it such that the discrepancies between the fit and spectrum for two successive peaks is minimized. Examples of fits where the β-Voigt parameter has been overestimated and underestimated are shown in Fig. 4 and Fig. 5, respectively (observe the trend of the red line = sum of the component functions).

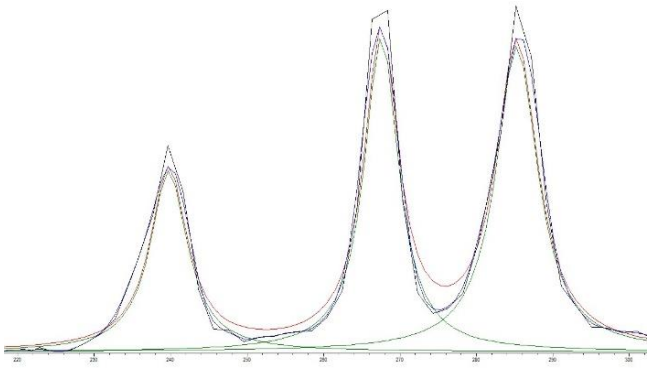

Fig. 4. Fitting example where β-Voigt has been overestimated

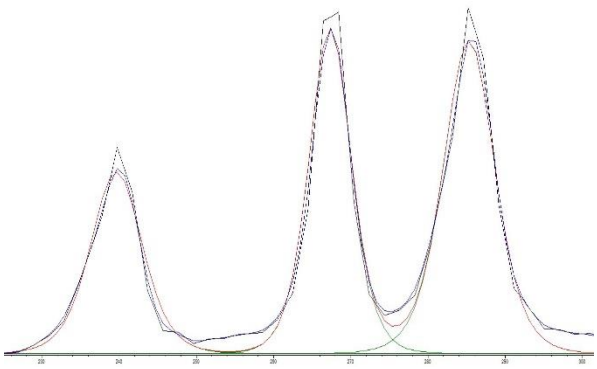

Fig. 5. Fitting example where β-Voigt has been underestimated

Calibration

In cases where the instrument used to collect the spectrum was poorly calibrated, or when transformation of an image to a spectrum resulted in a shift of X coordinates, it is possible to correct the X-axis (apply a correction to all X-coordinates) by clicking on the *Calibration* tab in the text menu. Next, one can click anywhere on the spectrum (imagebox1) to select the X1 coordinate represented by a blue point that can be moved along the X-axis with the right and left arrows. Once the user has selected the position of X1, the corresponding X-coordinated can be added to the table by clicking ADD. All points X1 to Xn selected and added by the user will appear in the Xu (uncorrected X-coordinate) column, then one can enter the correct X-coordinate in the corresponding column Xc (corrected X-coordinate). Points can be selected anywhere along the spectrum, if the user wishes to select the endpoints of the spectra, special care must be taken to ensure the coordinate of the selected point Xn is the same as the endpoint X coordinate of the spectrum.

Once all points have been selected and added, clicking RUN will generate a corrected (calibrated) spectrum in imagebox1 which will also appear in *File Manager* with the suffix _C and will be saved in the *temp* folder accordingly.

Correcting a single coordinate (Xc) will shift the entire spectrum by Xc – Xu. When two points are used in the calibration procedure a linear correction of the spectrum will be performed, when three points are selected a correction according to a cubic function (parabola) will be performed, and when four or more points are selected a correction according to a cubic spline is performed. To obtain the best result, it is advisable to correct the values obtained automatically in the Xu column by first noting the values read by zooming the spectrum or replacing them with the Xm values read in the appropriate box or obtaining them from the X-Y table of the spectrum (using view X-Y). The accuracy of the Calibration operation is limited by the discrete nature of the X-coordinates, *i.e.* X-coordinates are interpolated at steps of 1 cm^{-1}.

Derivatives

The Derivative operation allows one to view the 1st, 2nd and 3rd order derivatives of the primary spectrum in imagebox2 by clicking on the “D1”, “D2”, and “D3” buttons, respectively. To return to viewing the primary spectrum, exit the Derivative menu by opening the File Manager menu. The Derivative operation acts on the entire primary spectrum, if the Zoom is active the Derivative operation disables it. Derivatives can be saved in the *temp* folder as data files (with suffix _D1, _D2, _D3) such that they can be compared to the corresponding spectrum.

Spectrum Matching and Identification using the RCHDB (database)

The user can automatically compare the spectrum of an unknown mineral with those stored in the RCHDB. The RCHDB includes spectra from regular minerals, inclusion minerals, and synthetic phases. The user can compare their unknown spectra to all spectra in the database by selecting the “All spectra” option or search only amongst mineral inclusions in diamond or synthetic phases by selecting the “Inclusions in diamond” or “Synthetic” options, respectively. Users can identify spectra using two different methods called **[1] Spectra** (default) and **[2] Wavenumber**.

[1] Spectra: one must first click on the *Search in RCHDB* tab in the text menu (Fig. 1). If the database is loaded correctly (see **Installing RamanCrystalHunter *V*1.0** section), a list of minerals should appear, then click the *Search* button to filter mineral spectra that are similar to the unknown primary spectrum. It is also possible to search-match for a specific region of the spectrum using the zoom operation, this is useful for filtering minerals that have particular characteristic peaks. The similarity search is performed using the Spectral Angle Mapper (SAM) method (cf. Balaji Vengatesh M. - IJEDR302026).

Once a search has been completed, the results table will show a list of spectra each with a degree of similarity expressed as a percentage (MATCH% column). Spectra are listed in order of decreasing MATCH%, this can be changed by clicking on the column header. Double clicking on the name of a spectrum displays it in imagebox1 along with the unknown (primary) spectrum. All mineral spectra loaded from the database will appear in the *File Manager* menu where one can delete or add mineral spectra to imagebox1. One can increase the accuracy of the SAM method by adjusting the corresponding scroll bar, it may take significantly longer to complete searches when the accuracy is set high. When you return to the *Search in RCHDB* tab, the list retains the search order. It is recommended that users process unknown spectra (e.g., highlight characteristic peaks, perform a baseline correction, calibrate (correct) X-axis if necessary, smooth and reduce background noise, *etc.*) before attempting to search for similar spectra in the RCHDB database.

[2] Wavenumber: One can also search the RCHDB by specifying the parameters of the peaks of interest in the Wavenumber table. For each peak (≤6), the peak center (X-coordinate), Y% and FWHM should be indicated in the Wavenumber table. After the parameters of the peaks of interest have been defined, the user can use the Search command to filter the RCHDB for spectra with similar peaks. Peak parameters can be entered into the Wavenumber table manually or by simply double clicking on a position

(defined by X and Y%) in imagebox1 and then specifying the FWHM for each peak. Peaks can be deleted from the Wavenumber table by clicking on the “Delete values” button. As is the case for the “**Spectra**” search option, once can increase or decrease the accuracy of the search by using the scroll bar. When the accuracy scroll-bar is set to the lowest and highest setting, peaks within 20 and 5 pixels of each user-defined peak positions (X-coordinate) are incorporated in the MATCH% calculation. Therefore, users may want to decrease the accuracy setting to increase the number of similar peaks (and thus reported minerals) that are detected when searching.

If the user wishes to simply view spectra from the database, they can do so by first loading a primary file and then accessing the *Search in RCHDB* tab. If the user does not have a spectrum file to load, they can simply load a blank file (in the appropriate format) and then view spectra from the RCHDB as normal or use the example.txt file in C:/RamanCrystaHunter/Spectra.

***RCHDB drafting criteria*:**

In general, the RCHDB contains one to two spectra for each mineral or more than two spectra for minerals with Raman spectra that show significant variation. For example, three zircon spectra (Zircon_RIR_0664.txt to Zircon_RIR_0666.txt to) are included in the database as each has different characteristic peaks. Spectra that do not show peaks significantly distinct from signal due to background noise are excluded as such spectra cannot be matched to unknown spectra reliably. If spectra of the same mineral, obtained using lasers of different wavelengths, are similar, only one file is included in the RCHDB.

Each spectrum file is accompanied by the following information (shown in textbox 2); [1] Sample number, [2] Name of mineral or synthetic phase, [3] Sample provenance/locality, [4] Source of data, i.e., name of user who submitted and/or collected the spectrum, [5] Collection parameters, i.e., laser wavelength, [6] Mineral formula, [7] bibliographic and/or analytical references, and [8] other information, e.g., “mineral inclusion in diamond”. This information must be included with submitted spectra, more details about how to submit spectra are given below. The RCHDB contains the spectra of minerals which are identified with different characteristic codes (file names) related to the type of spectra (e.g., minerals, inclusion minerals (single or multiphase spectra), and synthetic phases) and the collaborator who provided the data. Inclusion minerals, multi-phase inclusion minerals, and synthetic phases are indicated with the codes _Inc, _Inc(M), and _Syn, respectively. For example, the file “Titanite+wollastonite_QZ_Inc(M)_0082.txt refers to the spectrum of a multi-phase inclusion of titanite and wollastonite submitted by Qiwei Zhang (QZ).

In addition to making regular improvements and updates to the RamanCrystalHunter *V*1.0 software, the authors will regularly update and improve the RCHDB by adding new mineral spectra and by removing/replacing poor quality spectra. We encourage users of the software to submit Raman spectra that are not currently included in the RCHDB or spectra that are of higher quality than those currently included in the database. This can be done by accessing the submission portal at https://www.fabrizionestola.com/submit-spectra.

Contributors

The following collaborators supplied spectra and/or samples for analysis;

Ivano Rocchetti, Edgar Huen, Vittorio Mattioli, Francesco Vecchi, Giorgio Bortolozzi