User Manual

for

QUIDDIT

Quantification
of **I**nrared-active **D**efects
in **D**iamond
and **I**nferred **T**emperatures
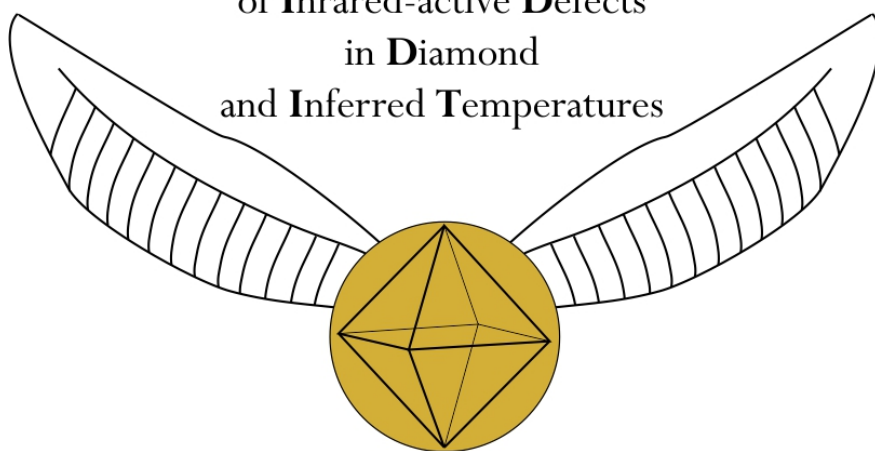
Laura Speich

laura.speich@bristol.ac.uk
https://github.com/LauraSp/QUIDDIT

1   Introduction

"QUIDDIT" stands for Quantification of Infrared active Defects in Diamond and Inferred Temperatures. It is a tool for quick automated spectral deconvolution of infrared (IR) spectral data of diamonds. QUIDDIT was developed as a research tool with a focus on the main spectral feature associated with the platelet defect but provides additional output on nitrogen impurity concentration and aggregation state and the main peak associated with the N_3VH defect found at 3107 cm^{-1}. Because processing is fast, QUIDDIT allows for processing of large amounts of data (this will depend on the type of computer used but the main part of processing typically takes less than one second per spectrum, baseline correction is even quicker).

The strategy for data processing with QUIDDIT is usually as follows:

- (optional: inspection of spectra)

- baseline correction and normalisation

- (optional: inspection of baseline corrected spectra)

- processing (N aggregation including model temperature, platelet peak including model temperature, N_3VH)

- review of fitting (best used for linescan data)

- plotting of results (if data is in the form of a linescan or map)

2   Installation

2.1   Installing Python

In order to run QUIDDIT, you will need a working version of Python 3. I recommend installing an integrated development environment (IDE) that includes the most commonly used libraries. To run QUIDDIT, you will need:

- **SciPy** (https://www.scipy.org/, Jones et al. 2001-present)

- **NumPy** (http://www.numpy.org/)

- **matplotlib** (https://matplotlib.org/, Hunter 2007)

- **Tkinter** (https://docs.python.org/2/library/tk.html)

- **webbrowser** (https://docs.python.org/2/library/webbrowser.html)

- **Spectral Python (Spy)** (http://www.spectralpython.net/)

Except for Spectral Python, all of these are part of the most common IDE for Python (help for the installation of SPy can be found in section 2.2 of this document). The instructions in this manual were created for use with **Spyder** (which is part of the Anaconda package), so I recommend installing Anaconda and running scripts with Spyder for users not familiar with coding.

To install **Anaconda**, visit https://www.anaconda.com/download/. Chose your operating system and follow the on-screen instructions.

2.2 Installing QUIDDIT

Download the QUIDDIT package from https://github.com/LauraSp/QUIDDIT and unpack the zip file. (To download, select "Clone or download" on the repository website)

To install Spectral Python (Spy), you will have to open a command window (type "cmd" into the search function in your start menu if you are on a Windows system) and navigate to the directory that contains your Python (Anaconda) installation. The Anaconda installer will ask where to install Anaconda. If you are using Microsoft Windows, it will most likely be located in either of the following locations

- `C:\ProgramFiles\Anaconda3\Scripts`

- `C:\Users[Username]\AppData\Local\Continuum\Anaconda3\Scripts`

- `C:\ProgramData\Anaconda3\Scripts`

You will need to use the cd [path] command to do so. Once you have navigated to the Scripts folder, type pip install spectral. Spectral Python should be installed in your system, this should only take a few minuutes. You might be prompted to enter your password during installation.

3 Usage

To run QUIDDIT, open your python IDE of choice and find and run the "QUIDDIT GUI" file in the downloaded repository (you may need to unzip the files in the repository first). The GUI can also be run using IDLE (a very basic standard Python IDE) or from the command line.

If you are using Anaconda, start Spyder (either directly from the start menu if you are on a Windows system or by starting the Anaconda Navigator first and then selecting "Launch Spyder"). Spyder may take a few minutes to start. Once it is finished, you can open the QUIDDIT script by selecting "File", then "Open..." from the start menu or by clicking the "Open File" icon. Navigate to the QUIDDIT folder you have downloaded and unzipped and select the file "QUIDDIT_GUI.py". This is usually the only file you will have to open. To run the script, select "Run", then "Run" from the top or click the single green arrow icon or hit the F5 key. If this is the first time you are running a script, you might be asked to agree to some run settings. You can confirm the defaults by clicking "Ok".

3.1 The GUI

To aid user-friendliness, the QUIDDIT package provides a Graphical User Interface (GUI, see figure 3.1) that is described in this manual. The **top menu** is a dropdown menu that starts all data processing and allows to change some basic settings. These will be discussed in the following sections of this manual. For a documentation of the computational methods, see the Jupyter Notebook files provided with the QUIDDIT GitHub repository and Speich et al. (2017, in progress).

Below the top menu, a **figure canvas** can be found which is used for displaying spectra, results of data processing (for both line and map data) and review results. It is equipped with a **toolbar** at the bottom that allows manipulating the graphs. The **message window** to the right prints messages regarding data processing and results and error messages.
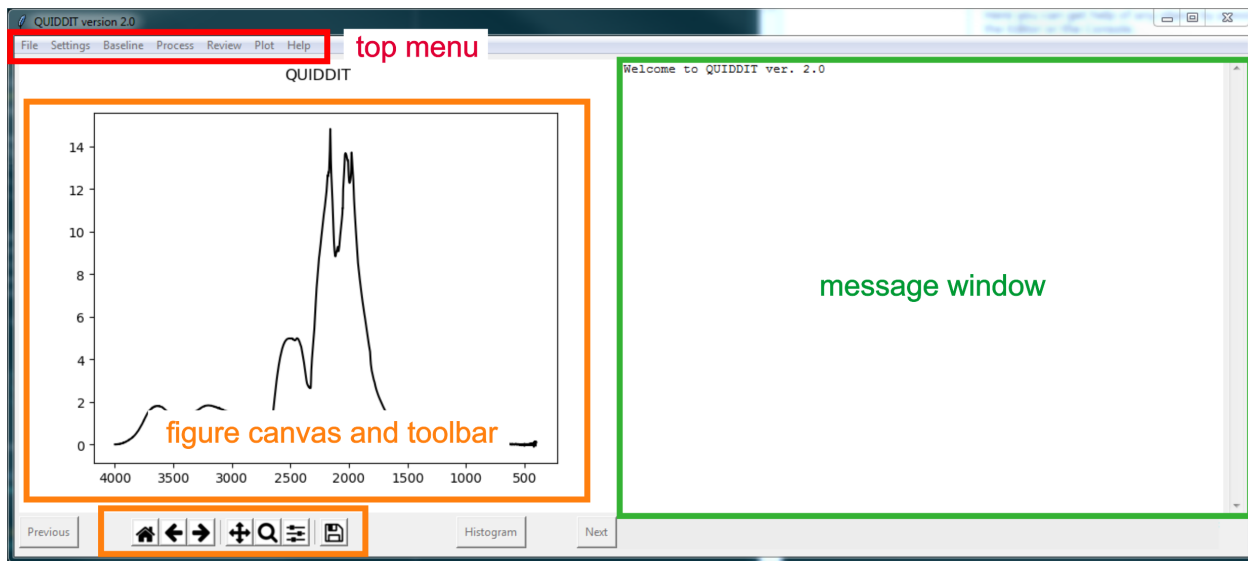
Figure 1: Graphical User Interface (GUI) included in QUIDDIT. Selections are made from the top menu (red), figures are plotted in the figure canvas and can be manipulated using the toolbar below (orange). The message window (green) provides information on the progress of data processing and displays results.

3.2 Supported Data Formats

Currently, QUIDDIT supports only data provided as CSV files. Each spectrum should be in a separate two-column CSV file with wavenumber values in the first column and absorption values in the second column. Individual values need to be separated by commas. QUIDDIT allows opening and reading of multiple or all files within a directory to allow batch processing.

ENVI files containing maps can be converted to CSV files using QUIDDIT. ENVI is a format for hyperspectral data (combining spectral and spatial information) that is supported by many manufacturers of IR spectrometers. It might be necessary to convert the map file into ENVI format using the spectrometer software. The ENVI file can then be split into multiple CSV files (one file per spectrum contained in the map) by selecting "File" and "convert ENVI" from the top menu in QUIDDIT. The user will be prompted to select a .dat file first and then the corresponding .hdr file. These files contain the measured data and information on how to read and interpret the data, respectively. The header (.hdr) file can be inspected using a text editor, such as Notepad and consists of rows of "key words" such as "samples" (number of columns of spectra in the map), "lines" (number of rows in the map), "bands" (number of absorption measurements per spectrum - spectral resolution) and "pixel size". QUIDDIT will read both these files and use the information contained in the header to convert the data in the .dat file into CSV files. The user will also be prompted to select a directory for the resulting spectra to be stored in. The conversion can take several minutes depending on the size of the map. In some cases, the header file was found to miss the "byte order" key word. Byte order depends on the operating system the file was created on, so QUIDDIT will ask for this information and add it to the header file if necessary.

3

All spectral data is interpolated. Thus in theory, all data can be processed regardless of spectral resolution. However, the reliability of results depends on spectral resolution, so we recommend 2-4 cm^{-1}.

All results (including baseline corrected spectra) are stored in CSV files that can either be read by QUIDDIT or opened in most standard spreadsheet software (such as Microsoft excel) for data manipulation and plotting, ensuring easy access. For displaying results, QUIDDIT will try to sort spectra according to the file name. For line data, the last 4 digits are used to put spectra in order; for map data, the expected file format is "X 1234.5 Y 1234.5.CSV", i.e. the x and y position of each spectrum within the 2-dimensional map is included in the file name. However, this is only relevant for plotting results, not for processing. If your file names are different, you will still be able to process the data.

3.3   Baseline Correction and Normalisation

The first step in spectral deconvolution is usually baseline correction. In QUIDDIT this also includes normalisation to 1 cm diamond thickness and can be achieved by selecting "Baseline" from the top menu. A pop-up window should appear, prompting the user to select measured spectra in CSV format. When the user has made a selection, another pop-up window requests a location for the corrected CSV files to be stored. You may chose to store the corrected spectra in one folder with the original files which is why QUIDDIT adds a character "c" to the beginning of the CSV file to mark corrected spectra.

3.4   Processing Data

For spectral deconvolution, select "Process" from the top menu. A pop-up window will appear, asking for spectra to be processed (corrected and in CSV format). Then, the user is prompted to input a sample name and mantle storage duration in million years (see figure 3.4). After the sample name is entered, the names for results and review file can be auto-completed by pressing the <tab> key. These names will be used to generate two CSV files to store results which can be used for displaying and reviewing data in QUIDDIT. The spectrum currently being processed will be displayed in the figure canvas, while a bar at the bottom of the window tracks overall progress. The message window prints some important results after processing, such as nitrogen concentration and model temperature.

Absorption coefficients for A- and B-centres are taken from Boyd et al. (1994, 1995). Model temperatures are determined from both nitrogen aggregation (calculated according to Taylor et al., 1990, 1996) and platelet degradation (calculated according to Speich et al., 2018).

As mentioned above, results are stored in two separate CSV files: "[sample name] results.csv" and "[sample name] review.csv" that can both be opened and inspected using a simple text editor or spreadsheet software such as Microsoft excel. Or, they can also be read by QUIDDIT for further processing. It is important to note that opening CSV files in excel and saving any changes might result in the file becoming unreadable by QUIDDIT. Therefore, it is recommended to save an unchanged copy of the CSV file or save a separate copy in .xlsx/.xls or similar format for manipulation in excel.

The "results" file contains important output on nitrogen concentrations and aggregation state, the platelet peak and the 3107 cm^{-1} peak and is read when the user prompts QUIDDIT to plot data. The "review" file contains all information necessary to reconstructs the fits achieved by QUIDDIT.

This includes most information reported in the "results" file but additional data on baselines and, for example, the 1405 cm^{-1} peak is retained as well. An explanation of the data columns reported in both files can be found below.

"results" file:

- **p_x0**: position of the platelet peak

- **p_I**: intensity of the platelet peak

- **p_HWHM_l** and **p_HWHM_r**: left (low wavenumber side) and right (high wavenumber side) half widths at half maximum of the platelet peak

- **p_sigma**: Lorentzian contribution to the platelet peak (values close to 0 indicate a Gaussian peak shape, values close to 1 indicate a Lorentzian peak shape)

- **avg**: the average (mean) wavenumber of the platelet peak. The distance to p_x0 can be used as a measure of peak symmetry

- **area_num_data** and **area_ana**: area of the platelet peak calculated numerically from the original data ("num") and analytically from the fitted pseudo-Voigt function ("ana")

- **As**: Asymmetry factor (ratio of half widths measured at 10% of the total peak height)

- **Tf**: Tailing factor (ratio of the full width measured at 5% of the total peak height over the left half width at 5% of the total peak height)

- **beta**: integral breadth (ratio of peak area over peak height)

- **phi**: form factor (ratio of full width at half maximum over integral breadth)

- **pp_sumsqu**: minimised sum of squared differences between the measured platelet peak and the fitted pseudo-Voigt function

- **c, a, x, b, d, const.**: absorption coefficients of C-, A-, X-, B- and D-centres. A constant ("const.") can be included in the fit as well

- **[NC], [NA], [NB], [NT]**: concentration of nitrogen in C-, A and B-centres and total nitrogen concentration

- **[T]**: temperature determined from the rate of A to B nitrogen aggregation (°C)

- **N_sumsqu**: minimised sum of squared differences between the measured nitrogen region (1000-1400 cm^{-1}) and the sum of the fitted components

- **I_3107**: intenisty of the 3107 cm^{-1} peak

- **H_area_ana**: analytically determined area of the 3107 cm^{-1} peak

Additional columns in the **"review" file:**

- **H1405_x0, H1405_I, H1405_HWHM_l, H1405_HWHM_r, H1405_sigma**: position (x0), intensity (I), left and right half widths (HWHM_l and -r) and Lorentzian contribution of the hydrogen-related peak at 1405 cm^{-1}

- **B_x0, B_I, B_HWHM_l, B_HWHM_r, B_sigma**: position (x0), intensity (I), left and right half widths (HWHM_l and -r) and Lorentzian contribution of the peak at ca. 1332 cm^{-1} that is part of the spectrum of B-centres

- **p_s2n**: signal-to-noise ratio within the platelet peak region

- **psv_c**: optional constant baseline value that can be added to the platelet peak fit

- **H_bg_a, -b, -c, -d**: parameters used in subtracting a third order polynomial baseline of the form $f(x) = ax^3 + bx^2 + cx + d$ from the region around the 3107 cm^{-1} peak

- **H_x0, H_I, H_HWHM_l, H_HWHM_r, H_sigma**: position (x0), intensity (I), left and right half widths (HWHM_l and -r) and Lorentzian contribution of the 3107 cm^{-1} peak

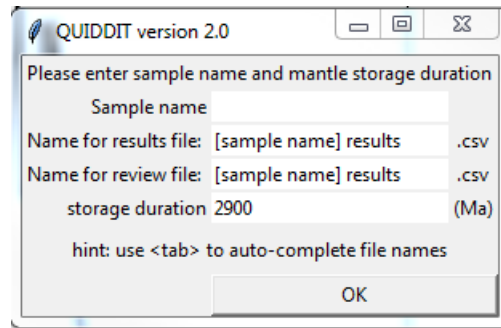Figure 2: User input prompt. Enter a sample name and press <tab> to auto-complete the names of results and review file.

3.5 Review

This part of QUIDDIT provides a review of all fitting that is carried out in the processing, thus allowing the user to evaluate fits and potentially reject data. Since the routine will attempt to open all selected spectral files, this is most efficient for linescan or point data. However, some spectrometer software (such as Omnic Picta) allows the extraction of a linescan from pre-existing map data. Since a linescan that consists of hundreds of spectra can be processed within minutes in QUIDDIT, extracting and processing a linescan or another type of sample subset of the overall map is recommended.

Reviewing results is done by selecting "Review" from the top menu. The user will be prompted to select the (baseline corrected) spectra and corresponding review file. Three subregions of each spectrum will then be displayed in the figure canvas; the nitrogen region (1000 to 1400 cm^{-1}), the platelet peak region (1330 to 1420 cm^{-1}) and the region of the 3107 cm^{-1} absorption (3000 to 3200 cm^{-1}). Measured data is shown in black, the corresponding best fit that was determined by processing is shown in green and the misfit in red (data minus fit). The user can cycle through spectra by using the "Next" and "Previous" buttons at the bottom of the window. Corresponding results regarding the three regions of interest are printed in the message window for the active spectrum.
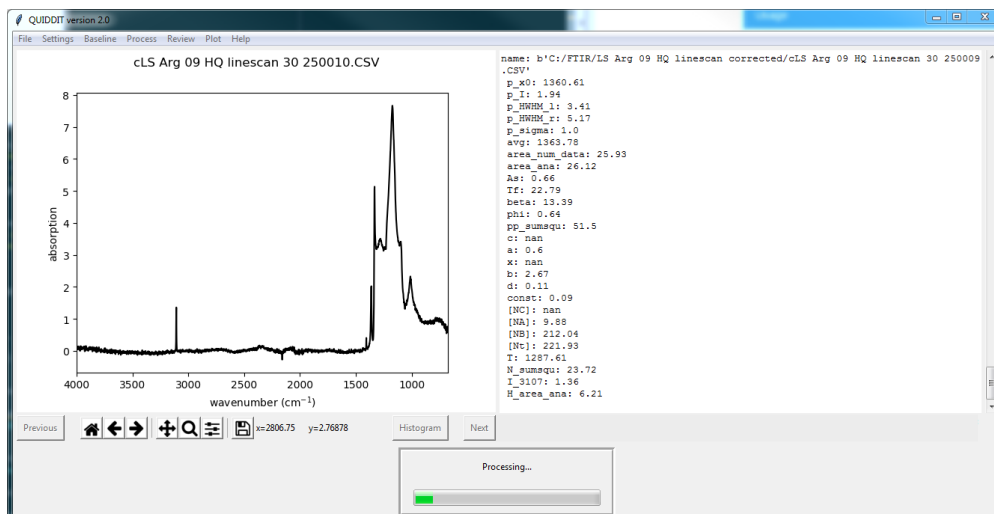
Figure 3: Example of data processing. The figure canvas shows the spectrum that is currently processed, whereas the message window displays some of the results of fitting.

3.6 Plotting Data

At any point, data can be plotted in the figure canvas for inspection. This includes spectra (before and after baseline correction) and results. The plots can be manipulated using the toolbar below the figure canvas. This includes zooming/panning and saving images in various different formats, such as jpg and png. Next to the toolbar, the x and y position of the cursor will be shown while the cursor is positioned within a plot.

Plotting Spectra To plot a single or multiple spectra, select "Plot", then "Plot single spectra" from the top menu. Then highlight all spectra you would like to plot and click "Open". The first spectrum you have selected will appear in the figure canvas to the left, the file name will be displayed above the spectrum. You can cycle through all spectra by clicking the "Previous" and "Next" buttons at the bottom. The toolbar below the figure canvas allows you to zoom and save images.

Plotting line data To plot results from processing linescans, select "Plot", then "Plot line data". You will be prompted to select a results file. Figure 3.6 is an overview of the types of diagrams displayed in the figure canvas. This includes nitrogen concentration (N in A- and B-centres and total=A+B), aggregation state, model temperature calculated based on nitrogen aggregation (Taylor et al., 1990, 1996), the area of the 3107 cm^{-1} peak, platelet peak area, degradation and model temperature based on platelet degradation (Speich et al., 2018), as well as a regularity diagram after Woods (1986). It may be necessary to maximise the QUIDDIT window at this point to avoid overlap between the separate diagrams.

Plotting map data To plot results from processing 2-dimensional maps, select "Plot", then "Plot map data". Select a results file to be read. Depending on the size of the map and the machine you
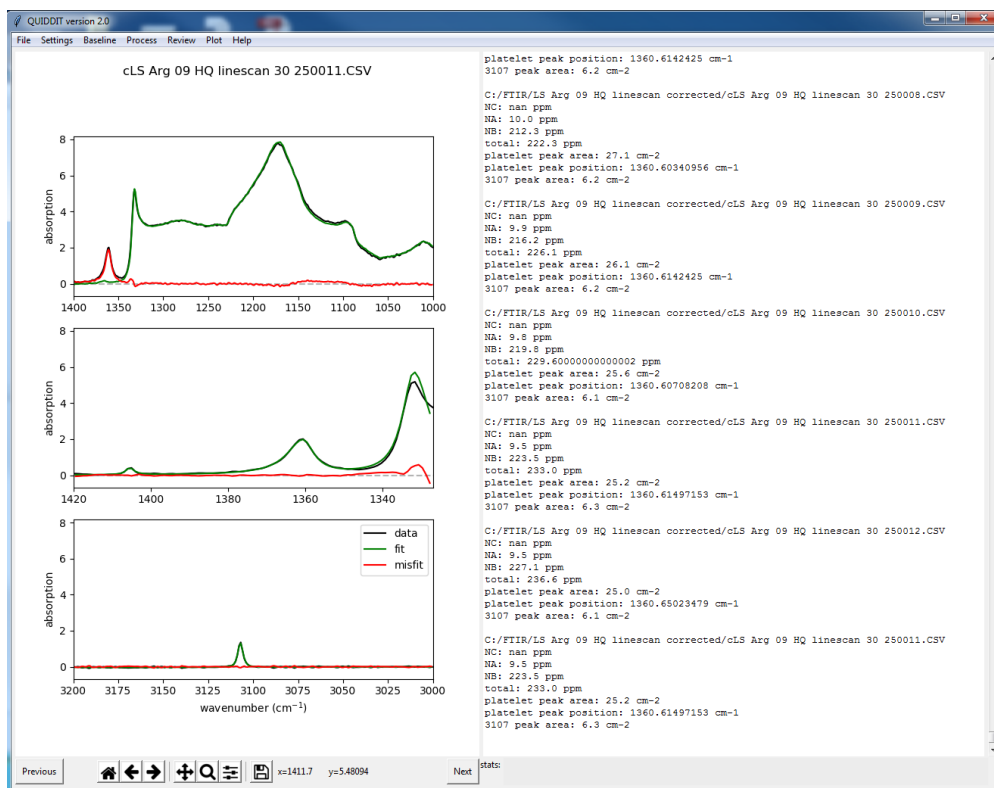
Figure 4: Review fitting in QUIDDIT. The figure canvas shows the three main regions of interest (N aggregation, platelet peak and 3107 cm^{-1} peak) with original data (black), fit (green) and misfit (red) with the name of the spectral file at the top. The message window prints some relevant data corresponding with the spectrum, such as file name, nitrogen concentration and aggregation state, and platelet peak area and position.

are using, reading the file and generating maps might take a few seconds. The first map that is displayed by default is of total nitrogen concentration. You can cycle through all other maps by using the "Previous" and "Next" buttons at the bottom. If you would like to change the range of the colour scale, select "Histogram" at the bottom. A new window will be opened that shows a histogram of the current map to allow the user to select a sensible range of values as a basis for the colour scale. Click "Redo map" to re-generate the map image.

4   Additional functionality

QUIDDIT provides limited additional functionality some of which has not been integrated into the GUI. This includes the following

- N-fit widget
- peak fit widget
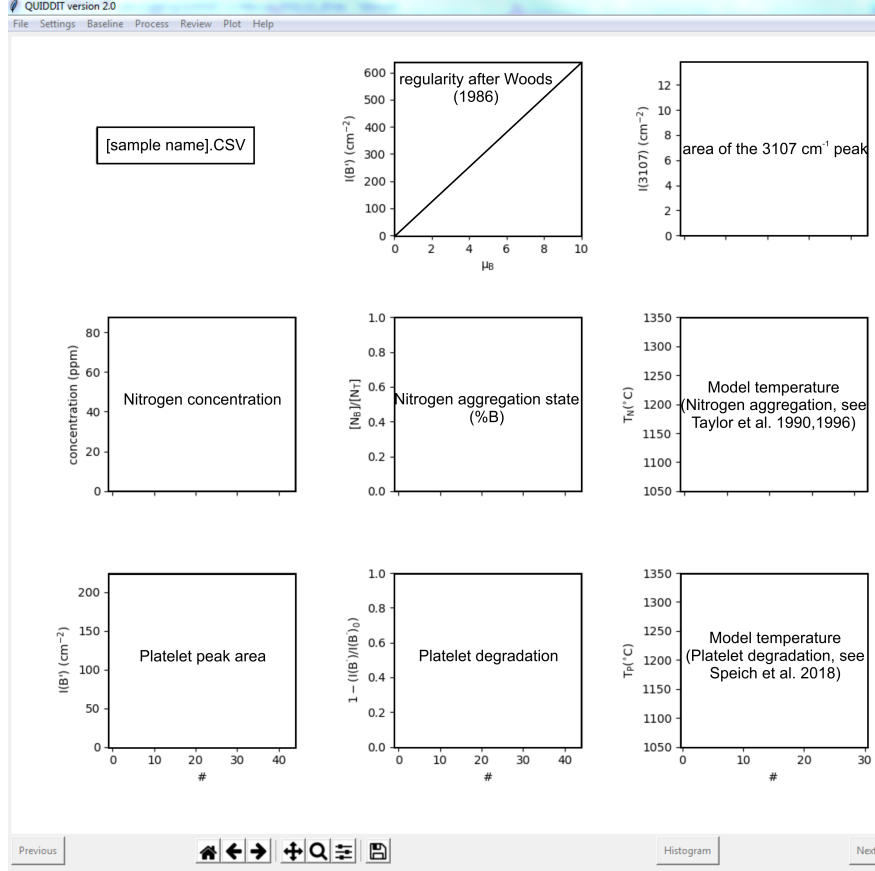- two-stage aggregation modelling

8

Figure 5: Illustration of parts of the line data plot. The QUIDDIT window was maximised to avoid diagrams overlapping for lack of space.

- Quadplot

4.1 Fitting widgets

The two "widget" applications allow the user to select a spectrum to be fitted by hand. They can be accessed by selecting "Manual fit" from the top menu. This can be useful if, for instance, nitrogen concentration is low, so the automated fitting routine that is used in the main part of QUIDDIT does not produce satisfactory results. It also allows the user to examine the effect of a change in parameters of the fit so an insight into the accuracy of a particular fit can be gained.

If "Fit N region manually" is selected, the user will be prompted to open a (baseline corrected) spectrum in CSV format first and then provide a mantle storage duration in Ma. The latter is only required for temperature calculations from the A to B defect reaction and should be ignored if C is present. If "Fit peak manually" is selected, in addition to the spectrum to be processed, the user is asked to provide an approximate value for the position of the maximum of the peak to be fitted.

In both cases, QUIDDIT will attempt to fit the selected spectral area automatically and display the measured data (black dots), fit (green line) and mismatch (red line) in the diagram area. Sliding

blue bars at the bottom can be dragged with the mouse to change the fit. The mismatch will change accordingly. Corresponding values of peak position, height, half widths, Lorentzian contribution or c, a, x, b, d and const. (absorption coefficients of the contributing end-members in the N-region and an optional constant baseline value) are shown to the right to each slider. Vertical red bars indicate values obtained from the automatic fit for each slider. They can be restored by using the "reset" button.
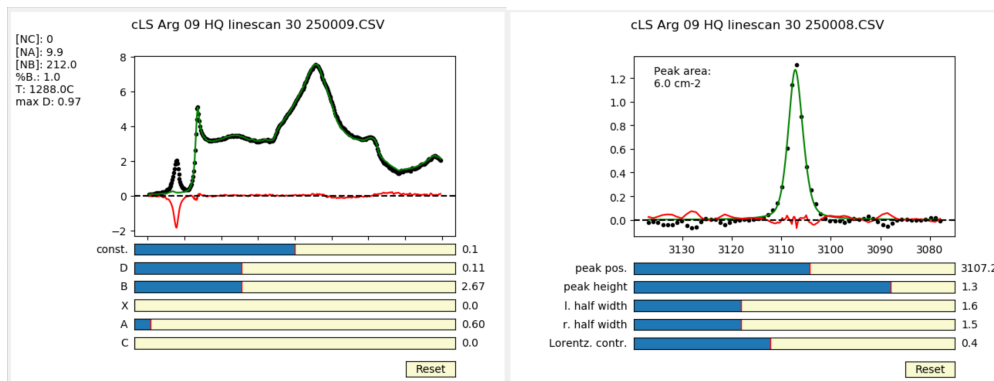

Figure 6: Manual fitting in QUIDDIT. Left: fitting of the N-region (1000-1400 cm^{-1}), right: fitting of a single peak of choice. When selecting "Manual fit" from the top menu, the user will be prompted to enter a mantle storage duration if fitting the N-region or the approximate peak position if fitting a single peak. QUIDDIT will try to fit the selected area and display the measured spectrum (black dots), suggested fit (green line) and misfit (red line) in the plotting canvas. Depending on the type of manual fit chosen, the user can manipulate a number of sliders below the diagram (blue) by dragging to change the fit. Corresponding values of peak position, height, half widths, Lorentzian contribution or c, a, x, b, d and const. (absorption coefficients of the contributing end-members in the N-region and an optional constant baseline value) are displayed next to each slider.

4.2 Two-stage aggregation modelling

The concept of two-stage modelling for zoned diamonds was first introduced in Kohn et al. (2016). The idea is to use the known total duration of aggregation (from inclusion and eruption age dating) and the measured nitrogen concentration and aggregation state to constrain the thermal history of the diamond. This feature can be found under "2-stage-modelling" in the top menu. The user will be prompted to enter the required data for core and rim (note: please enter aggregation state as a proportion of 1, e.g. 40% IaB = 0.4). QUIDDIT will vary the duration of the two annealing stages (keeping the total annealing time constant at the duration supplied by the user) and calculate temperatures for the core and rim portion of the diamond that would result in the observed aggregation state.

Results will be shown in the form of a plot in the figure canvas but will also be stored in a .CSV file which can be found in the folder that contains your QUIDDIT code files. If using Spyder, you can check the location of your current working directory at the top right of the screen. The file will be named automatically according to the input ("two-st_[ ]Ma_cNT[ ]_cagg[ ]_rNT[ ]_ragg[ ].CSV"). The figure generated by QUIDDIT shows core (red) and rim (blue) temperatures for each dura-

tion of the first annealing stage (i.e. the time of rim growth). The crossover between the two is the point in time at which the rim grew if the temperature was constant throughout the entire pre-exhumation history of the diamond (see figure 4.2).
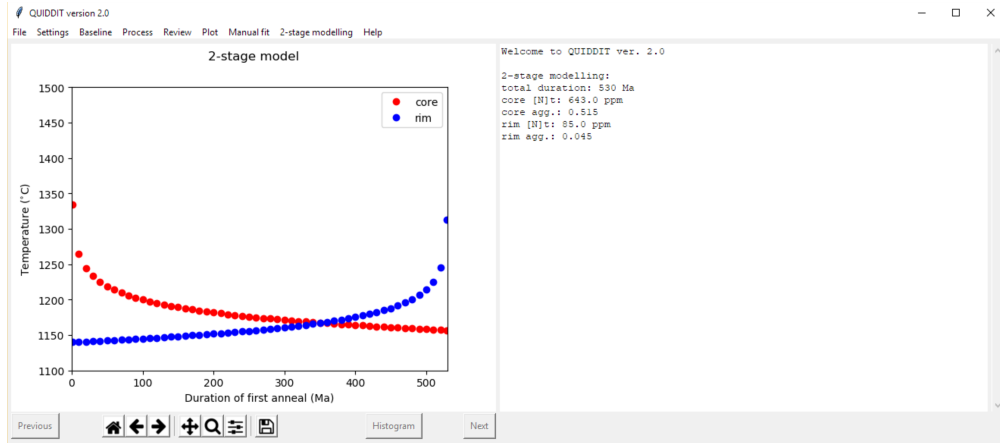

Figure 7: Two-stage modelling in QUIDDIT. Dots show temperatures for the core (red) and rim (blue) zones of the diamond per assumed duration of the first annealing stage (time elapsed before rim growth). The crossover (ca. 350 Ma and 1170°C in the example shown) shows the timing of rim growth if constant temperature is assumed.

4.3 Quadplot

The "Quadplot" application allows plotting of the four diagrams presented in Speich et al. (2018):

1. $I(B')$ vs μ_B (area of the platelet peak vs absorption due to B-centres)

2. $I(B')$ vs x_0 (area vs position of the platelet peak)

3. sym vs x_0 (symmetry vs position of the platelet peak)

4. $FWHM$ vs x_0 (width vs position of the platelet peak)

Note: Quadplot can not currently be accessed from the GUI but work on GUI integration is in progress.

5 Help

For further questions, new bugs, please contact Laura Speich (ls13943@my.bristol.ac.uk). Please make sure you are using the most up to date version of QUIDDIT and consider consulting the Readme, Known errors and Ipython Notebook files in our GitHub repository first (https://github.com/LauraSp/QUIDDIT).

References

S. R. Boyd, I. Kiflawi, and G. S. Woods. The relationship between infrared-absorption and the A defect concentration in diamond. *Philosophical Magazine B-Physics of Condensed Matter Statistical Mechanics Electronic Optical and Magnetic Properties*, 69(6):1149–1153, 1994. ISSN 0141-8637. doi: 10.1080/01418639408240185.

S. R. Boyd, I. Kiflawi, and G. S. Woods. Infrared-Absorption by the B-nitrogen aggregate in diamond. *Philosophical Magazine B-Physics of Condensed Matter Statistical Mechanics Electronic Optical and Magnetic Properties*, 72(3):351–361, 1995. ISSN 0141-8637. doi: 10.1080/13642819508239089.

J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3): 90–95, 2007. ISSN 1521-9615. doi: 10.1109/MCSE.2007.55.

E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001-present. URL http://www.scipy.org/. [Online; accessed 2016-08-15].

Simon C. Kohn, Laura Speich, Christopher B. Smith, and Galina P. Bulanova. FTIR thermochronometry of natural diamonds: A closer look. *Lithos*, 265:148 – 158, 2016. ISSN 0024-4937. doi: 10.1016/j.lithos.2016.09.021. The Nature of Diamonds and Their Use in Earth's Study.

L. Speich, S.C. Kohn, R. Wirth, G.P. Bulanova, and C.B. Smith. The relationship between platelet size and the B′ infrared peak of natural diamonds revisited. *Lithos*, 278281:419 – 426, 2017. ISSN 0024-4937. doi: 10.1016/j.lithos.2017.02.010.

L. Speich, S. C. Kohn, G. P. Bulanova, and C. B. Smith. The behaviour of platelets in natural diamonds and the development of a new mantle thermometer. *Contributions to Mineralogy and Petrology*, 173:39, 2018. doi: 10.1007/s00410-018-1463-4.

L. Speich, S.C. Kohn, G.P. Bulanova, and C.B. Smith. QUIDDIT - quantification of infrared active defects in diamond and inferred temperatures. *Computers & Geosciences*, in progress.

W. R. Taylor, A. L. Jaques, and M. Ridd. Nitrogen-defect aggregation characteristics of some Australasian diamonds - time-temperature constraints on the source regions of pipe and alluvial diamonds. *American Mineralogist*, 75(11-12):1290–1310, 1990. ISSN 0003-004X.

W. R. Taylor, D. Canil, and H. J. Milledge. Kinetics of Ib to IaA nitrogen aggregation in diamond. *Geochimica Et Cosmochimica Acta*, 60(23):4725–4733, 1996. ISSN 0016-7037. doi: 10.1016/S0016-7037(96)00302-X.

G. S. Woods. Platelets and the infrared-absorption of Type-Ia diamonds. *Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences*, 407(1832):219–238, 1986. ISSN 1364-5021. doi: 10.1098/rspa.1986.0094.