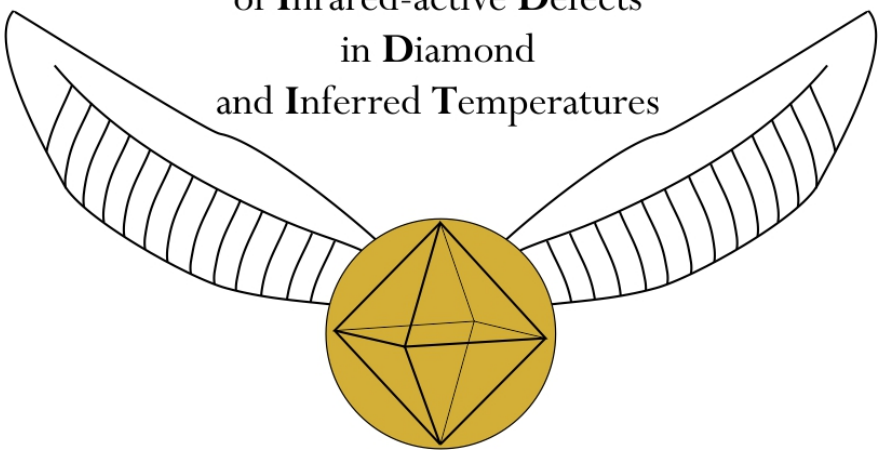
User Manual

for

QUIDDIT

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



Laura Speich

laura.speich@bristol.ac.uk

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:

- (inspection of spectra)
- baseline correction and normalisation
- (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 Usage

2.1 The GUI

To aid user-friendliness, the QUIDDIT package provides a Graphical User Interface (GUI, see figure 2.1) 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.

2.2 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. Development of a routine to open and process ENVI files (a common file format that combines 2-d spatial and spectral data) is currently in progress.

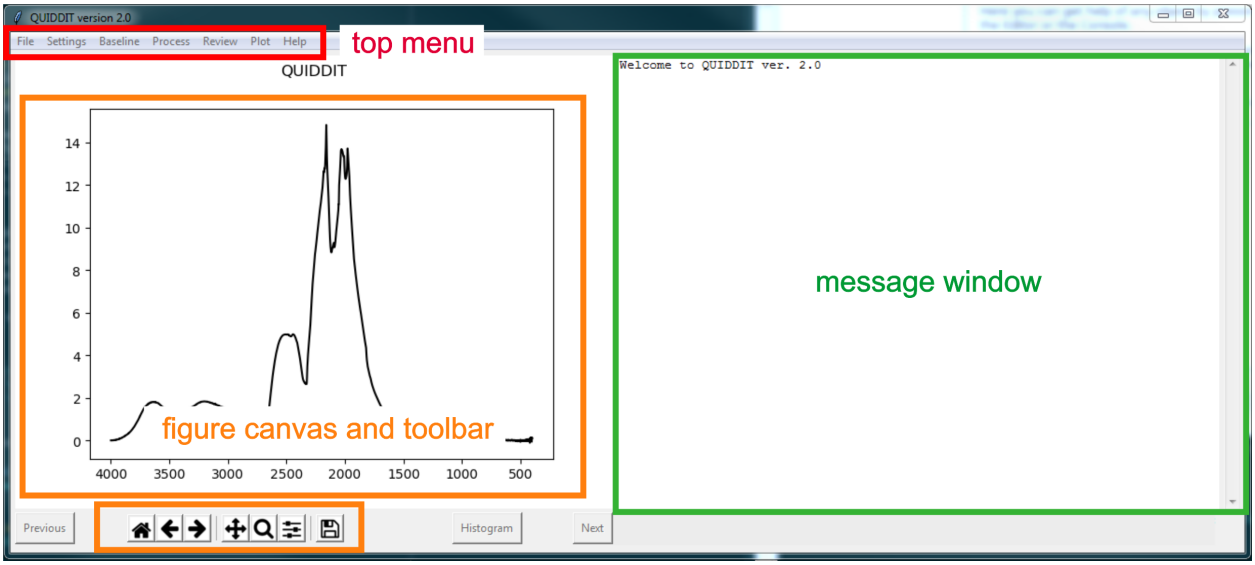


Figure 1: Graphical User Interface (GUI) included in QUIDDIT. Selections are made from the top menu (red), figures are plotted in the figure canvas and can be manipulated using the toolbar below (orange). The message window (green) provides information on the progress of data processing and displays results.

All spectral data is interpolated thus in theory, all data can be processed spectral resolution 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.

2.3 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.

2.4 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 2.4). 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.).

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
- **Tf**: Tailing factor

- **beta**:
- **phi**:
- **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

2.5 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.

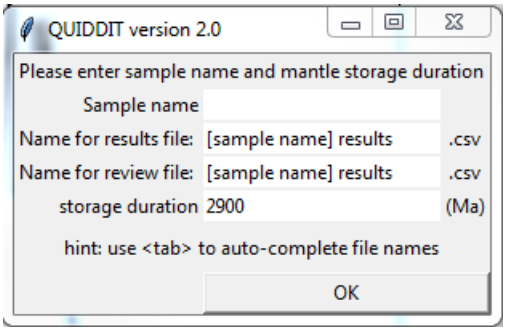


Figure 2: User input prompt. Enter a sample name and press <tab> to auto-complete the names of results and review file.

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.

2.6 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 2.6 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.), as well as a regularity diagram after Woods

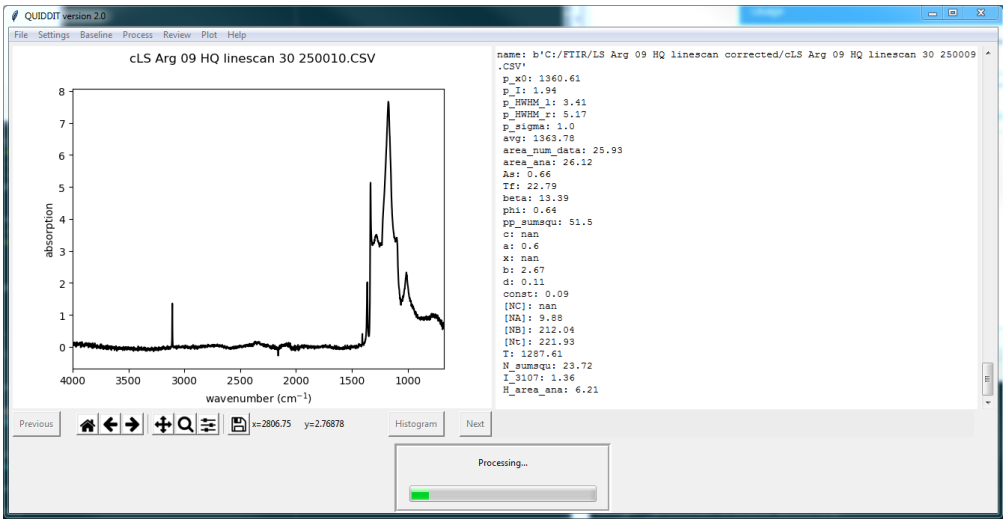


Figure 3: Example of data processing. The figure canvas shows the spectrum that is currently processed, whereas the message window displays some of the results of fitting.

(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 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.

3 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
- Quadplot

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

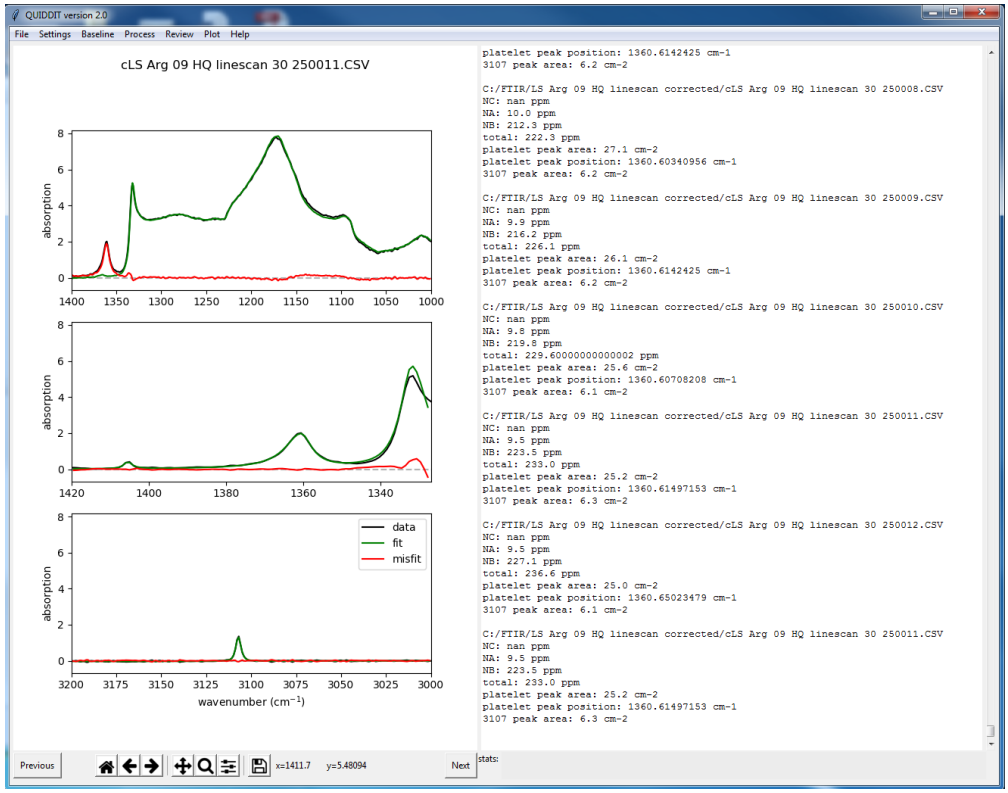


Figure 4: Review fitting in QUIDDIT. The figure canvas shows the three main regions of interest (N aggregation, platelet peak and 3107 cm^{-1} peak) with original data (black), fit (green) and misfit (red) with the name of the spectral file at the top. The message window prints some relevant data corresponding with the spectrum, such as file name, nitrogen concentration and aggregation state, and platelet peak area and position.

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.

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

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)

4 Help

Fur further questions, new bugs, please contact Laura Speich (ls13943@my.bristol.ac.uk) or Simon Kohn (Simon.Kohn@bristol.ac.uk). Please make sure you are using the most up to date version of

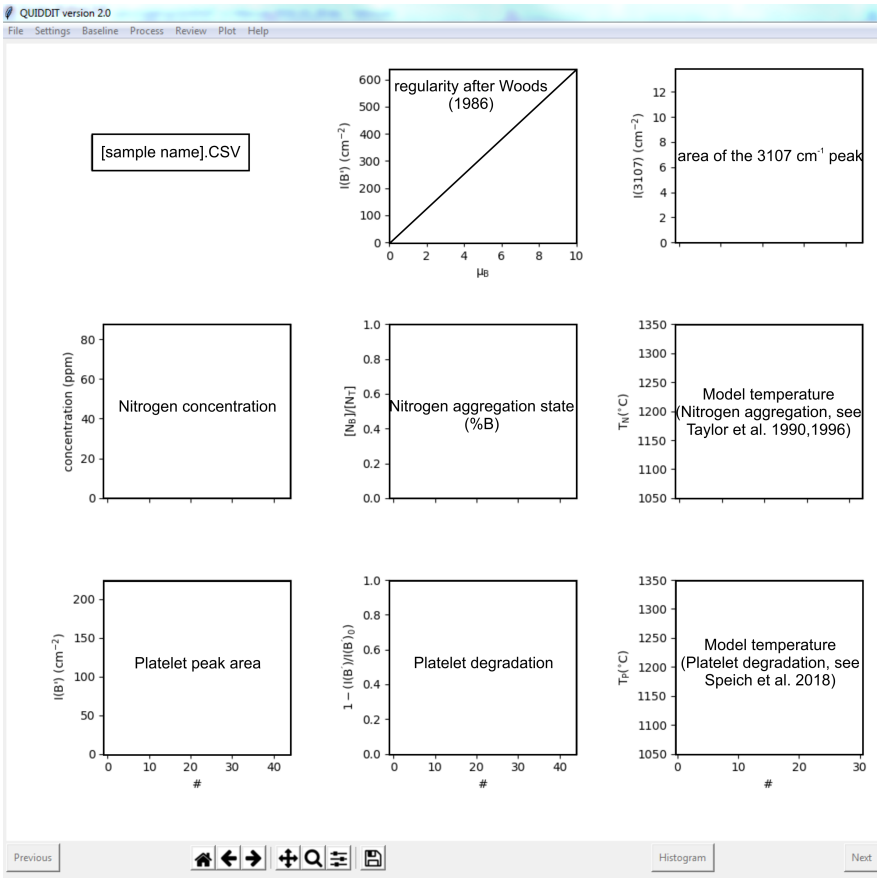


Figure 5: Illustration of parts of the line data plot. The QUIDDIT window was maximised to avoid diagrams overlapping for lack of space.

QUIDDIT and consider consulting the Readme, Known errors and Ipython Notebook files in our GitHub repository (https://github.com/LauraSp/QUIDDIT) first.

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.

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*. doi: 10.1007/s00410-018-1463-4.

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*, 278âĂŞ281:419 – 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. 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.