Quack and the QACD (Quantitative Assessment of Compositional Distribution) method

A python-based tool for the processing and statistical analysis of quant optimised X-ray Maps collected by SEM EDS analysis

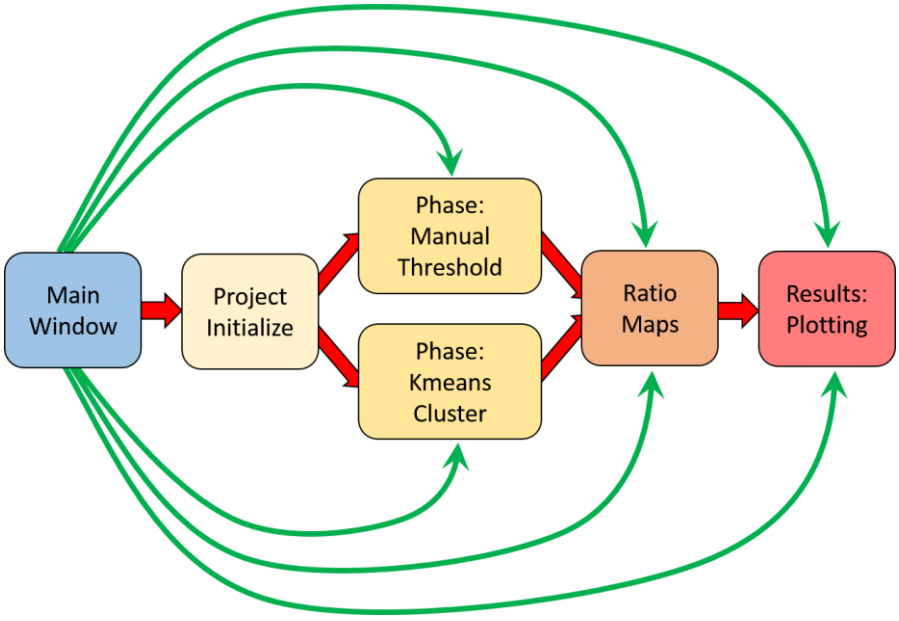
By Matthew Loocke and C. Johan Lissenberg

System Requirements and Data specification

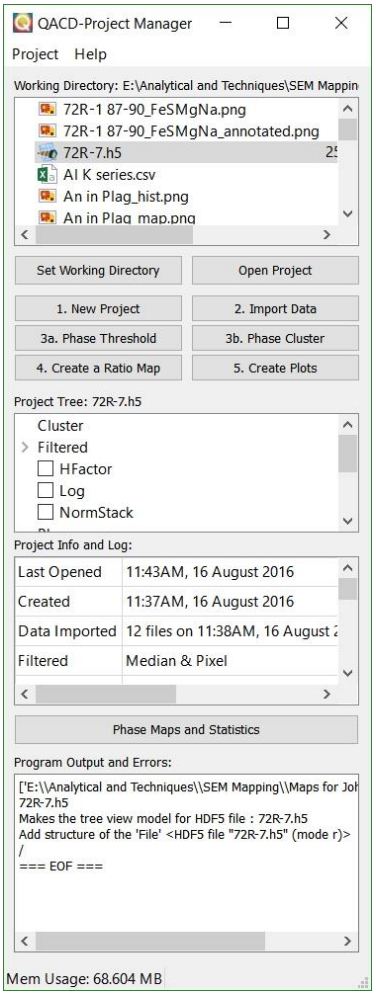
This document is a user manual for the processing and handling of element (X-ray) maps collected by SEM EDS analysis for statistical analysis of chemical populations within analysed samples. The QACD software can be run as a python script executable. The files are available for distribution within the QACD repository on GitHub (link). There are some limitations and initial bugs in the software which are inherent to the Python language. Many of these bugs have already been reported and we are hard at work in correcting them. However, we encourage you to submit a bug report to the GitHub if you do come across one in your day-to-day operation of the software.

The background of the QACD software is outlined in Loocke (Cardiff University PhD thesis, 2016, "*The role of the axial melt lens in crustal accretion at fast-spreading mid-ocean ridges*") and uses a workflow as outlined in Figure 1. This application was born out of frustration with the interactive plotting abilities and, to some extent, the cost of the MatLab platform. The software is coded using the Python 2.7 language and the PyQt4 module for the GUI. The complete list of modules being implemented in the software is as follows: *pytables*, *numpy*, *dask*, *sklearn*, *PyQt4*, *matplotlib*, *PIL*, *psutil*, *scipy*, *pandas*.

The software requires that the data must be background-corrected X-ray intensity values for each pixel in a .csv text file format with no headers; this is a common exportable format in vendor-supplied software for element map collection. Effectively, the .csv file is a text image or matrix of pixel intensities (e.g., akin to a numpy array/matrix). The files for each element should be named '# K series.csv', where # would be the element abbreviation (e.g., 'Mg K series.csv' or 'O K series.csv').



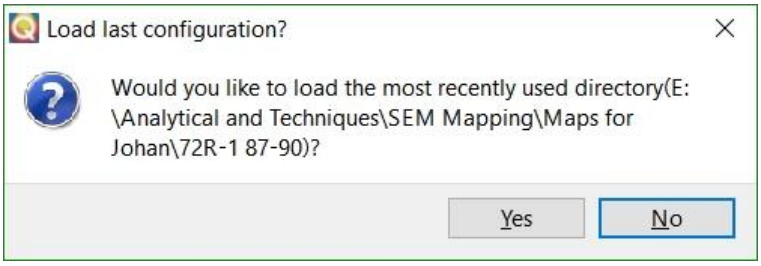
Getting Started



The main window for the QACD software is the 'Project Manager' window. It sits at the centre of the program and is used for launching the various dialogs. The QACD method contains a number of tools for statistical analysis and plotting of element maps and is divided up into 5 main parts:

- Part 1, Project creation: After setting the working directory to the folder containing the desired maps, the user proceeds to create a New Project File.
- Part 2, Import Data: The user chooses which .csv files will be imported, the filtration type for the data, and provides the step size or pixel size.
- Part 3, Phase Analysis: The user can either choose to manually threshold a single phase at a time (more accurate and more control) or stack the maps and perform a Kmeans++ cluster analysis for a range of estimated total number of phases.
- Part 4, Create Ratio Maps: The user has the option to create quantified element or ratio maps (e.g., Anorthite (Ca/Ca+2Na) or Mg# (Mg/Mg+Fe)) and store them in the project.
- Part 5, Create Plots: The user can create map plots and histograms of the element maps with the option to mask the maps for display of specific phases.

Outside of these sections of the software, there are also the options to plot and export phasemaps and statistics resulting from phase cluster analysis and export various maps and arrays from the project as text files.



On initiation of the software, you will be prompted with a window asking if you would like to open the main window with the working directory already set to the last known working directory. This has been created so that if you run out of memory while working and need to restart the software, you will be able to pick up where you left off.

Overview and Reference for Principal Components

The Project Manager

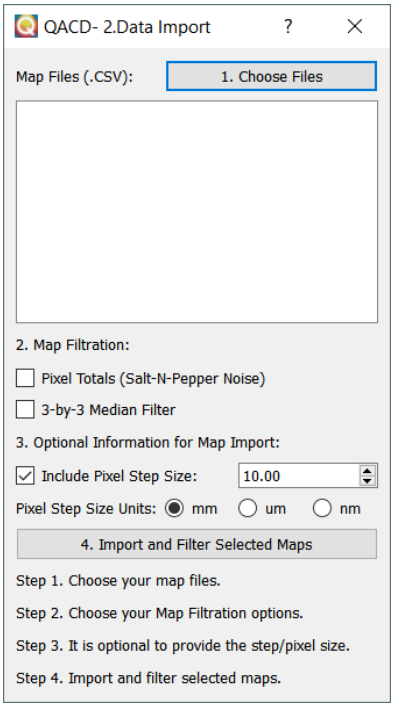
Upon loading the application, the first thing that you will see is the Project Manager Window. This window is the parent for every other dialog in the program, and serves as a tool for exploring the local working directory, project file tree, and the project file log. Aside from the readily visible options within the window, the Project and Help menus in the menubar provide further options and tools alongside a handful of right-click options for the buttons.

The working directory view provides a directory tree view for the directory which was chosen and set with the working directory tool.

The Project tree view provides a file tree view of the groups and nodes within the HDF5 project file (hierarchical data format file, .h5) which is updated contemporaneously with functions that modify the project file.

The Project info and Log view provides a table which displays pertinent information to the modification and saved configuration for each project file. This view maintains a record of the date-time stamp for various steps in the QACD process, as well as metadata for the project file.

Import Data Dialog



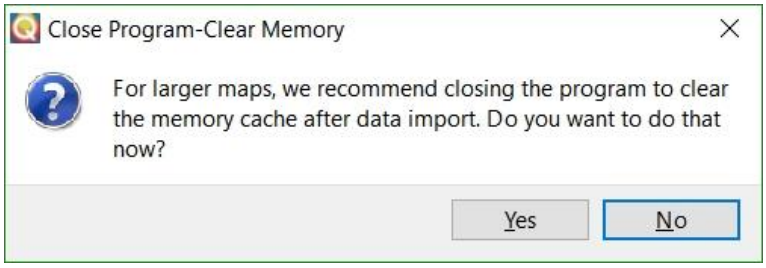
The data import dialog is launched by clicking on the '2. Import Data' button in the Project Manager window. This dialog consists of four steps.

Step 1: The '1. Choose Files' button launches a file dialog. The user selects the .CSV files they want to include in the processing. This will populate the white list box with the corresponding element symbols (e.g., Ca, Mg, Si, O, etc.).

Step 2: The '2. Map Filtration' section provides two checkbox options for filtration of the map data, 'Pixel Totals (Salt-N-Pepper Noise)' and '3-by-3 Median Filter. The user has the option of choosing either both filtration methods or only one. We recommend that you choose both options so as to minimise potential errors in the software alter on. The bugs associated with this step are noted for revision.

Step 3: If it is known, the user can elect to 'include Pixel Step Size' by adjusting the spinbox accordingly and choosing the appropriate units for the value. The default value is 10.

Step 4: The 'Import and Filter Selected Maps' button takes all of the inputs provided by the user and, using cython, sends the data off to be processed accordingly. WARNING, this step needs a large amount of processing, which can cause the software to enter a 'Not Responding' state. This is normal. If this happens at this step, then the software is working fine. We note that with maps of large size (e.g. >2,500,000 total pixels) and/or larger numbers of element maps (e.g. >10) this step can take up to a minute or longer.



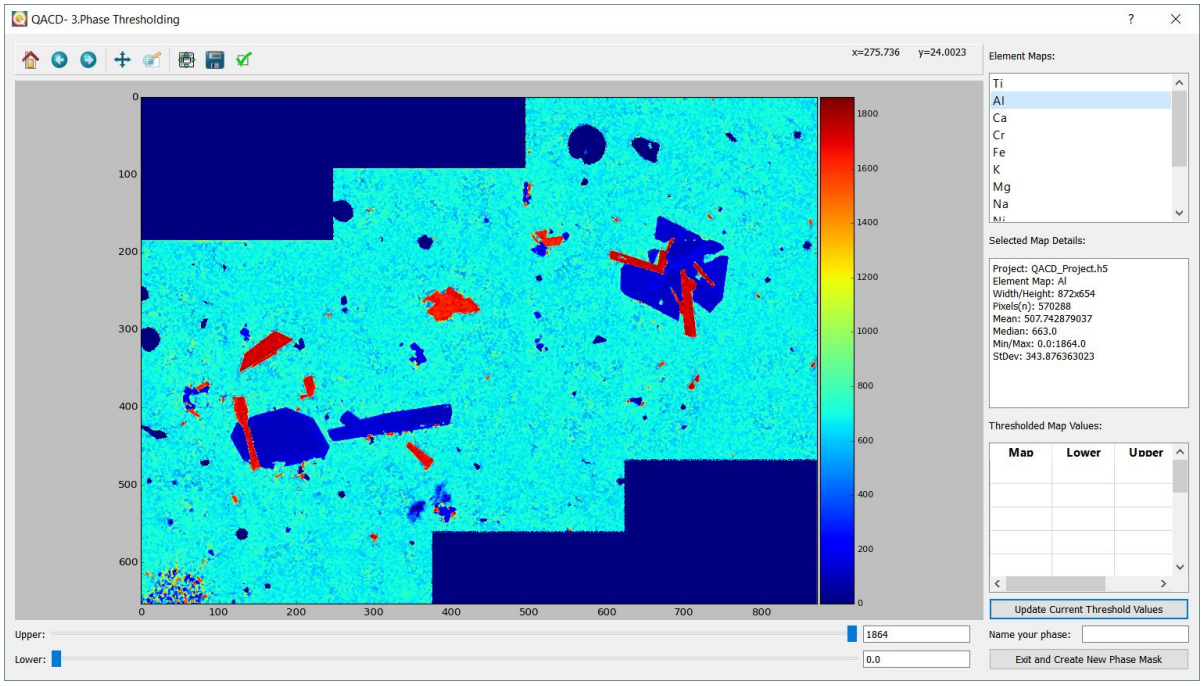
After the maps have finished processing and importing, you will be prompted with the option to close the software, saving the working directory. This is highly recommended due to the memory-heavy nature of the current version of the software. This will then allow for you to carry on with the processing and map creation without having to worry about consuming too much memory. We are also currently working on alternative ways to minimise memory usage and more efficiently recycle python objects in memory.

Phase Threshold Dialog

The phase thresholding dialog is launched by clicking on the '3a. Phase Threshold' button in the Project Manager Window. This dialog contains a figure widget (Matplotlib module), a list of the projects element maps, a map details text view, a threshold table, two threshold sliders, and a line edit for naming the thresholded phase. When an element map is chosen from the list, its map details and metadata are displayed in text view, the map is plotted in the figure widget, and the maps maximum and minimum values are set for the upper and lower sliders respectively. The sliders are used to adjust the minimum and maximum values displayed on the map. When satisfied with the adjusted threshold values for the target phase, the user clicks the 'Update Current Threshold Values' button. This captures the current threshold values and places them in the table view. Then the user chooses another map from the list and repeats the process until they are satisfied.

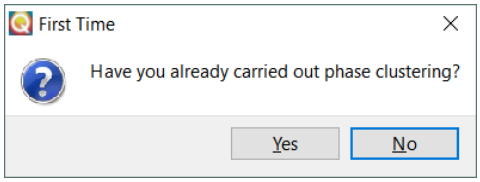
Logically, the minimum number of thresholded map values needed for phase mask calculation is one. This can be more than enough when dealing with phases such as sulphides which may be the only phases to contain sulphur in an element map. When ready, the user types their chosen name for the thresholded phase into the 'Name your phase' line edit and clicks on the 'Exit and Create New Phase Mask' button. The calculation should not take too long, but may lock the program for a brief moment before closing the dialog and returning to the Project Manager window.

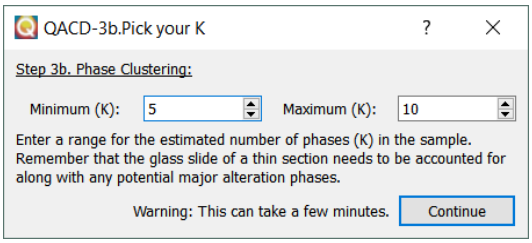
Note that displaying this dialog consumes a decent amount of memory. Due to the nature of certain components within the Matplotlib module which are currently implemented in this software, a percentage of this memory will 'leak' and not be released by the system unless the program is closed. If you are working with very large maps, it may be beneficial to occasionally close the software and reload the project. This is a bug that we are more than aware of, and are currently working on a permanent solution.



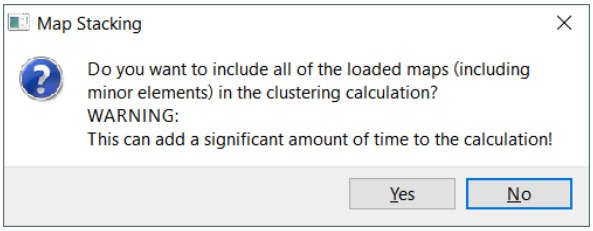
Phase Cluster Dialogs

The phase cluster dialog set is launched by clicking on the '3b. Phase Cluster' button in the Project Manager Window. This launches a chain of dialogs with the most important being the 'Pick your K', 'Map Stacking', 'Results', and 'Edit' dialogs. The user will first be asked if they have carried out a round of clustering already.





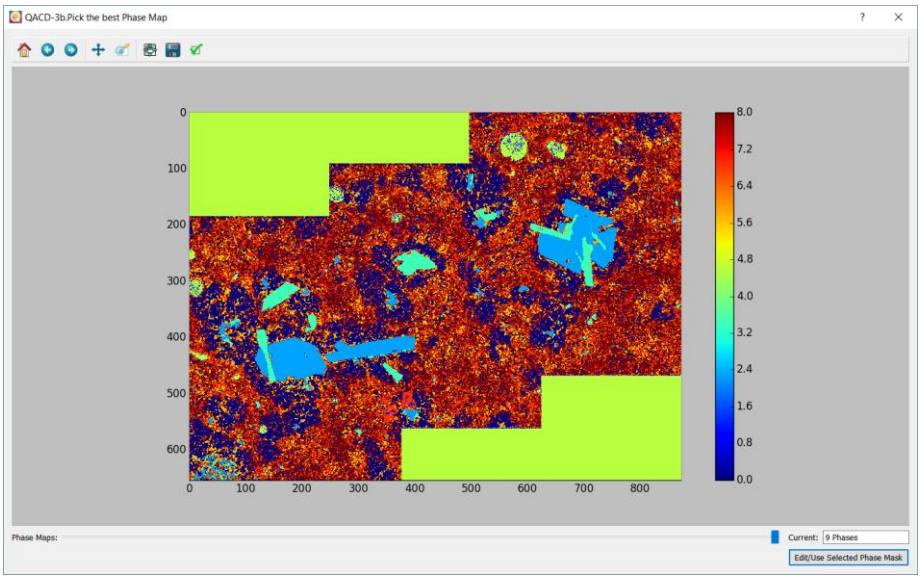
Next, the 'Pick Your K' dialog asks the user to provide a range of expected number of phases. It is important to note that, although you may think there are only 4 phases in the sample, the algorithm will pick up and group things such as the glass slide or epoxy if the samples in a thin section and variations in alteration phases as well as chemical zoning (i.e., if it is distinct). So, it is always better to choose a larger number of phases than expected. This is why we provide a minimum and maximum number for user input. We suggest that you keep these within 3-4 of each other. Any more, and the software will run out of memory during the calculation (memory allocation on a given machine is set by the machine and the administrator, i.e., this is out of our hands).



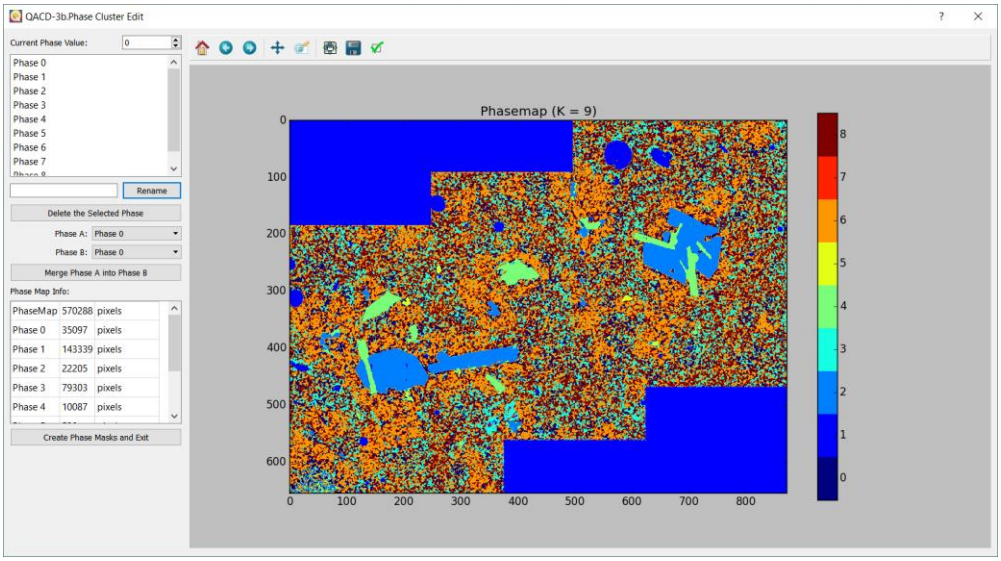
After choosing a K-range, the user has the option of whether or not to include the ENTIRE set of element maps for the project in the clustering calculation. If the user chooses 'No', then the calculation is performed with Ca, Mg, Fe, Si, and Al. If they choose 'Yes', then the calculation is performed with the entirety of the loaded and processed element maps. Once the user selects an option, the calculation will begin. As with the data import step, this calculation can take a little bit and may cause the software to be non-responsive.

> We note that, in most cases dealing with Rock Forming minerals, these maps will work just fine for the cluster analysis. If you are working with smaller maps, then the entire map set option should be fine. Just remember, that it will include maps that may cause the calculation to return weird or unexpected results. We currently plan to add in options which would allow for the user to customise the base map set for cluster calculation by selecting either a subset from the maps or even selecting a 'lithological preset' for the project which will customise aspects of the work flow and calculations according to the selected lithology.

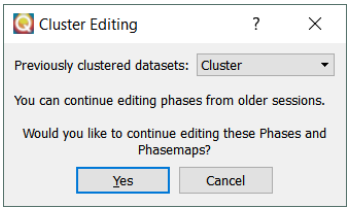
The next 'Results' dialog displays a simple figure widget (Matplotlib) with a slider and a line edit with the currently displayed map name. The user can scroll through the results of the clustering for the range of K values. After settling on the map determined to be the best, the user clicks on the 'Edit\Use Selected Phase Mask' button.



The next, and final, dialog is the ‘Phase Cluster Edit’ dialog. This provides the user with a chance to edit the results of the chosen cluster calculation. A list of the resulting phases is displayed in the top left corner. A phase labelled ‘Removed’ is used as a placeholder for pixels which have been deleted by the user. When a phase is selected from the list, its value in the map displayed in the figure widget is set in the spinbox above the list, and its name displayed in the line edit below the list. When selected, the user can edit the phase name in the line edit and rename the phase by pressing the ‘Rename’ button. Phases that you do not want to include in your final maps (e.g., glass slide, epoxy, or maybe some alteration) can be deleted by selecting them from the list and clicking the ‘Deleted the Selected Phase’ button. The total pixels for the deleted phase will be added to the ‘Removed’ group as a form of metadata. Phases can also be merged together by selecting the desired phases with the ‘Phase A’ and ‘Phase B’ combo boxes and clicking on the ‘Merge Phase A into Phase B’ button. This button does exactly what it says. Phase A will be removed from the record and its pixels will be reassigned to Phase B. Finally, a ‘Phase Map Info’ table is provided which keeps track of the total pixels in the PhaseMap and each phase for the group. Once the user is finished, they click on the ‘Create Phase Masks and Exit’ button.



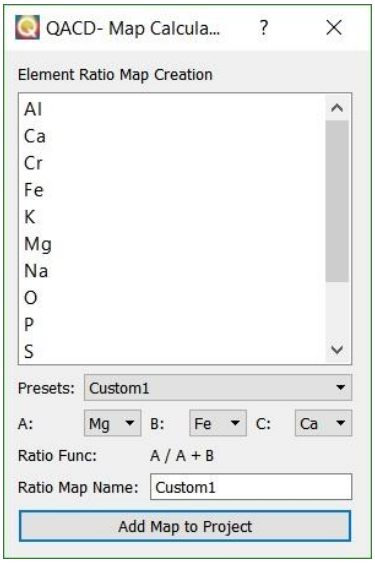
If you want to return and continue editing the results, simply right click on the '3b. Phase Cluster' button in the Project Manager window and choose the 'Edit Previous Cluster Group' option from the context menu. This will pop up the 'Cluster Editing' dialog where the user can choose which Cluster group they want to continue editing.

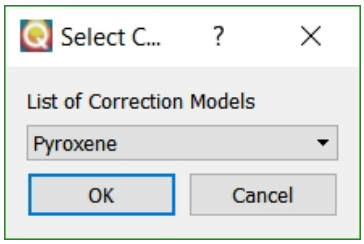


> Note that the kmeans clustering algorithm may result in some subtle and sometimes not-so-subtle differences in the results for a given K value each time that it is implemented. This is why the first dialog of the clustering step asks if the user has performed the calculation previously in for the project file. If they have, then they would choose 'Yes' and a new Cluster group is created for the calculation.

Ratio Map Dialog

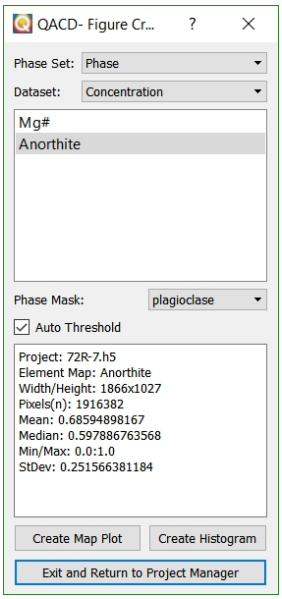
The Ratio Map dialog is spawned by clicking on the '4. Create Ratio Map' button in the Project Manager Window. This step is very simple. The dialog displays a list of the maps which are currently in the project. If the user has already created any ratio maps, then they will show up in the list. The user chooses a ratio map 'preset' from the preset combobox. Unless the chosen preset is a custom preset, the component A and B comboboxes will automatically be chosen and the 'Ratio Func' label and 'Ratio Map Name' line edit will be populated. The name can be changed to suit the user. For set ratio calculations like Mg# and Anorthite, there are certain preset configurations which are used in the calculations. For a custom ratio map, choose one of the custom ratios. The Custom2 ratio is used for a B component which has a valence state of B+1 (e.g., Na or K). There is also a 'Single' element calculator which will quantify a single element according to the internal standard database. We are currently working on expanding the functionality of the Ratio Map calculator towards a more customisable format.





Once you have chosen your ratio, a dialog will pop-up asking for your preferred correction model. This gives you the choice to calculate the ratio according to a general regression of all internal standards, or according to a regression specific to a certain mineral group.

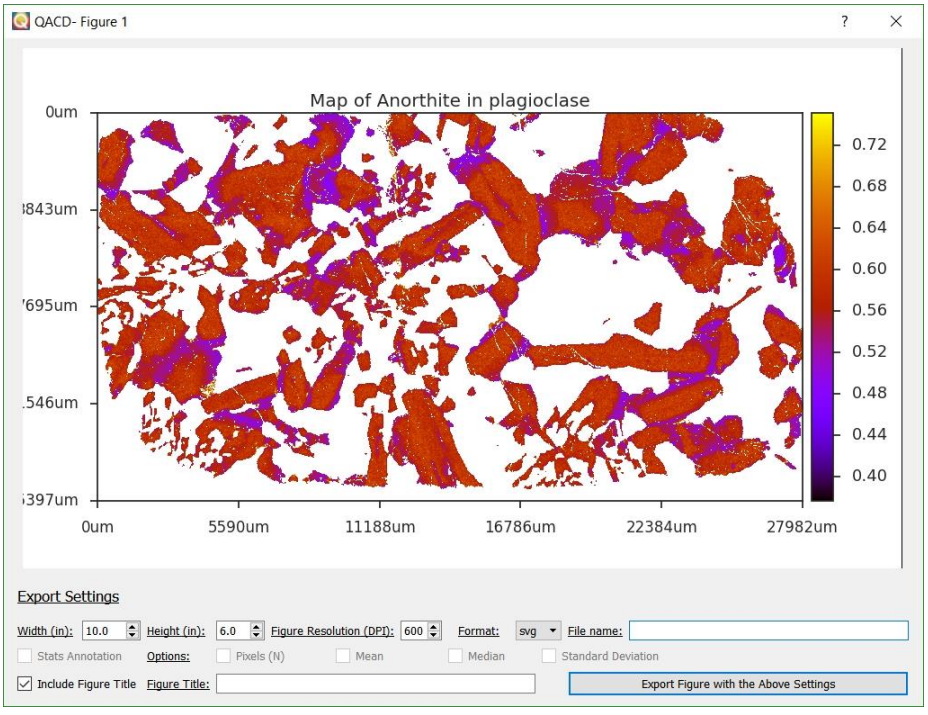
Create Plots Dialogs



The Figure Creation dialog is spawned by clicking on the '5. Create Plots' button in the Project Manager Window. The resulting dialog contains a 'Phase set' combobox at the top for choosing between Cluster groups (i.e., the 'Cluster' options) and thresholded phases (i.e., the 'Phase' option). Maps can be selected for plotting from a list containing all of the element and ratio maps for the project. If the user wants to create plots without phase masks, they should choose the 'No Phase' option from the 'Phase Mask' combobox. This box is automatically populated with the members of whichever phase group is chosen at the top of the dialog. The metadata and statistics for the chosen map are displayed in the 'Map Info' text view. Once the 'Phase Set', element/ratio map, and 'Phase Mask' option have been set, the user can choose to plot either a Map Plot or Histogram of the chosen data by clicking the 'Create Map Plot' or 'Create Histogram' buttons respectively. If the user is finished, they can simply choose the 'Exit and Return to Project Manager' button.

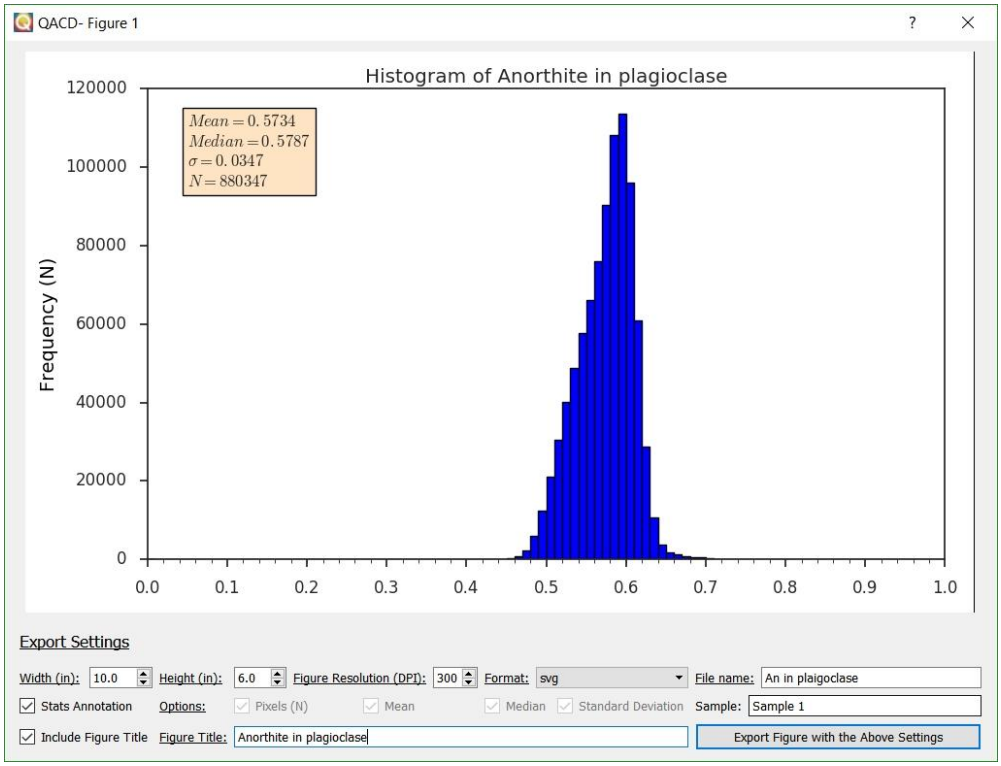
Note: The Map Plot and Histogram Plot creation has been known to eat up memory. We suggest that this step either be saved for last with a given project work flow, or be carried out after closing the program and reloading the project.

The 'Create Map Plot' option spawns a window with an example of the plot resulting from the chosen options. In this window, the user can set the width, height, figure resolution (DPI), file type, file name, and whether or not to include a user defined title on the map before exporting the figure. For most figures, we recommend a minimum DPI of 400.

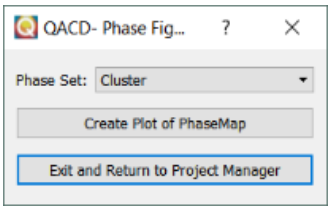


The 'Create Histogram' option spawns a window with an example of the histogram resulting from the chosen options. In this window, the user can set the width, height, figure resolution

(DPI), file type, file name, and whether or not to include a user defined title on the map and the blue annotated stats box on the exported figure. For most histogram figures, we recommend a minimum DPI of 200. The histograms are currently calculated with 200 bins for intensity maps and 100 bins for ratio maps which result in values between 0 and 1. We are working on adding in the option to customise both the number of bins and the x-axis range.

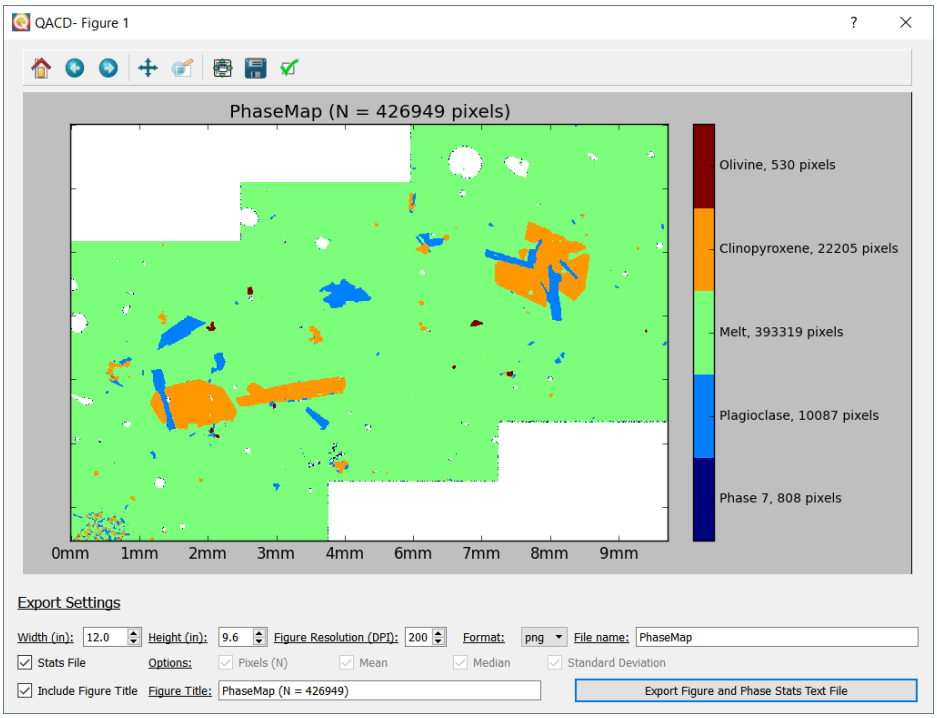


Phase Maps and Statistics Dialog

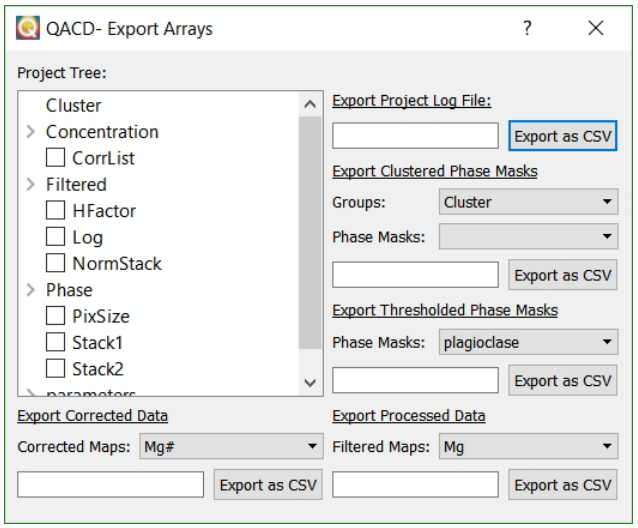


The Phase Maps and Statistics dialog is spawned by clicking on the ‘Phase Maps and Statistics’ button in the Project Manager Window. The first dialog gives the user the option to choose from which Cluster group they would like plot the PhaseMap for. We note that this can only be done for PhaseMaps produced by cluster analysis.

The ‘Create Plot of PhaseMap’ button spawns a dialog akin to the ‘Create Map Plot’ dialog. In this window, the user can set the width, height, figure resolution (DPI), file type, file name, and whether or not to include a user defined title on the map before exporting the figure. The user also has the option to export the phase statistics for the phase map to a CSV text file named ‘PhaseMap_Stats’. For most figures, we recommend a minimum DPI of 400.



Data Export Dialog



The Data Export dialog is spawned by navigating to the Project menu in the menubar of the Project Manager Window, and clicking on the 'Export Maps' option. This window provides a project tree view on the left, allowing the user to navigate and examine the groups and nodes of the current project. To the right are 4 options for exporting data from the project with a custom file name.

Option 1- Export Project Log File: This option allows the user to export a .CSV text formatted file of the Project Log which is displayed in the Project Info and Log table of the Project Manager Window.

Option 2- Export Clustered Phase Masks: This option allows the user to export an individual Phase Mask from a Cluster group. The user simply chooses the Group and the phase, and then gives the file a name. The result is a .CSV text formatted file with Boolean pixels of 1 and 0, with 1 indicating the phase.

Option 3- Export Thresholded Phase Masks: This option allows the user to export an individual thresholded Phase Mask. The user simply chooses the phase and then gives the file a name. The result is a .CSV text formatted file with Boolean pixels of 1 and 0, with 1 indicating the phase.

Option 4- Export Processed Data: This option allows the user to export an individual element map from the project file. All of the maps available for export have been filtered and processed. The user simply chooses the desired map and then gives the file a name. The result is a .CSV text formatted file with type float pixels.

Option 5- Export Calculated/Corrected Data: This option allows the user to export an individual quantified element or ratio map from the project file. The user simply chooses the desired map and then gives the file a name. The result is a .CSV text formatted file with type float pixels.

For pixels which have been tresholded out or assigned a NaN value, the exported maps leave the pixel blank. This can cause problems when trying to load these maps into a program like ImageJ as a text image. We are aware of this bug and it is on our to-do list.