Supplement of Geosci. Instrum. Method. Data Syst., 14, 193–209, 2025
https://doi.org/10.5194/gi-14-193-2025-supplement

Geoscientific Instrumentation Methods and Data Systems
Open Access
EGU

Supplement of

A free, open-source method for automated mapping of quantitative mineralogy from energy-dispersive X-ray spectroscopy scans of rock thin sections

Miles M. Reed et al.

Correspondence to: Miles M. Reed (miles.reed@wisc.edu)

Step by step guide – Automated mineralogy with QGIS and Orfeo Toolbox

Overview

This document is a step-by-step guide for implementing the method described in the main text for obtaining automated mineral maps from energy dispersive spectroscopy (EDS) scans of rock thin sections. Before starting this method, users will need to have already produced elemental intensity rasters using EDS and to have already installed the necessary software (see below). To help users implement this method for the first time, we recommend applying these steps to the example elemental intensity rasters for one of the rock thin section datasets provided in the supplement.

Software requirements

QGIS (3.28.11 or above) – Windows 10/11, macOS (Intel only), or Linux

Orfeo Toolbox QGIS plugin (8.0 or above)

SAGA GIS QGIS plugin (9.0 or above)

Step 1

a) Identify the mineral phases in your samples with an abundance of at least ~0.1% (i.e., that are likely abundant enough to be successfully mapped). In the example datasets in the supplement (1-13a, 6-3a, and 16-2a), we identified eight mineral phases with abundances greater than 0.1% by examining the raw energy dispersive spectroscopy (EDS) spectra with the Oxford AZtec software package: plagioclase feldspar, quartz, hornblende, biotite, potassium feldspar, Fe-Ti oxides, chlorite, and apatite.

b) Identify the elements that may make for useful input features to the random forest models. For instance, to train a random forest model to identify potassium feldspar, it would be useful to supply the random forest model with maps of elemental intensity of K. For the example datasets in the supplement, the major elements we identified for each mineral phase are listed in the following table.

Table 1. Identified minerals and major elements.

Mineral	Major elements
Plagioclase feldspar	Ca, Na
Quartz	
Hornblende	Mg, Fe, Ca
Biotite	Mg, Fe, K, Ti
Potassium feldspar	K
Fe-Ti oxides	Fe, Ti
Chlorite	Mg, Fe
Apatite	Ca

c) If your sample's raw (unfiltered) elemental intensity rasters are not already formatted as TIF (.tif) image files, create one TIF image file for each elemental intensity raster. This may require coordination with your scanning electron microscope data provider. In the example datasets in the supplement, we did this by exporting elemental intensity rasters for six elements (Ca, Na, Mg, Fe, K, and Ti) from the AZtec EDS environment.

d) Open QGIS and go to 'Project'→ 'Properties'. Select 'No CRS (or unknown/non-Earth projection)'.

e) Import all elemental intensity rasters into QGIS with 'Layer'→ 'Add Layer' → 'Add Raster Layer'.

Step 2

a) Smooth each elemental intensity raster with a circular mean filter.

i) Open the SAGA GIS 'Simple Filter' function.

ii) For 'Filter', select 'Smooth'.

iii) For 'Kernel Type', select 'Circle'.

iv) For 'Radius', select '7' (pixels).

v) For 'Filtered Grid', select 'Save to temporary file'. This will save the output image as a temporary file. If this is not selected, SAGA will save to its own SDAT format, which Orfeo Toolbox cannot use.

vi) Alternatively, if SAGA GIS is not available to the user, other functions in QGIS such as 'r.neighbors' could provide a similar mean filter.

b) Export this temporary file to a TIF image.

i) Go to 'Export' → 'Save as' and save as a GeoTIFF (.tif file).

ii) Select 'Raw data'.

iii) Leave 'CRS', the coordinate reference system, as 'invalid projection'.

Step 3

a) Create a virtual raster from the smoothed elemental intensity rasters.

i) Open the 'gdalbuildvrt' function in QGIS from the Geospatial Data Abstraction Library (GDAL).

ii) Select the smoothed elemental rasters under 'Input Layers' and click 'OK'.

iii) Select the 'Place each input file into a separate band' option to put each smoothed element elemental intensity raster as its own band.
iv) Leave all other options unchanged.
v) Click 'Run' to create the virtual raster. The resulting .vrt file is a container for all the images, which is useful because it is more easily trainable and uses less computer resources.

Step 4

a) Generate training data for the random forest model by creating a series of polygons. Each polygon must encompass a small area within a single mineral phase. In the examples in the main text, we used circular polygons for simplicity. All pixels within the polygon will be considered an example of that mineral. As described in the main text, we created these polygons within a small portion of the thin section, which we refer to as the training area, to avoid training the random forest model on test data, which was needed for the optional calculations of model accuracy. Restricting the training polygons to a small region of the thin section would not be necessary if the user will not be performing an accuracy assessment.

i) Create a training vector polygon dataset in QGIS by clicking 'Layer' → 'Create Layer' → 'New Shapefile Layer'.
ii) Draw a training polygon that encompasses only one mineral phase. For each mineral phase, select the previously chosen combination of elemental intensity rasters that will be useful as a visual aid for delineation of the mineral. For the example datasets in the supplement, see Table 1 of mineral phases listed above at Step 1b.
iii) Save the file as an ESRI shapefile (.shp) with the same unknown CRS. For ease, include a new integer field named 'mineral_id' that contains a number associated with each mineral phase.
iv) Once created, toggle editing by right clicking (Windows and Linux) or Control-clicking (macOS) the file. Once toggled, new features can be added the file with the 'Add Polygon Feature' button.

b) Follow the same steps to generate many training polygons encompassing many more pixel-level examples of the same mineral phase. In the examples in the main text, we generated hundreds of polygons for each mineral phase, and we allowed some overlap between polygons, especially in cases of low abundance minerals (e.g., apatite).

c) Follow the same steps to generate a set of training polygons for every mineral phase.

d) Label each mineral phase with a unique integer within the shapefile's 'mineral_id' field. For ease, keep these numbers constant across all thin section samples.

e) Use the virtual raster and the training data polygons as inputs in the training of the random forest image classification model.

i) Open the 'TrainImagesClassifier' function included with the Orfeo Toolbox package.
ii) For 'Input Image', select the virtual raster.
iii) For 'Input Vector Data', select the training polygons.
iv) For 'Field containing the class integer label for supervision', select the name of the attribute field in the vector data that contains the integer that defines the mineral.
v) For 'Classifier to use for the training', select 'rf'.
vi) Use the default hyperparameters (e.g., "Maximum number of trees in the forest", etc.).

vii) Leave all other options unchanged.

viii) Save the trained model (.mod file) with a unique name that identifies it in relation to the sample.

f) Check the F1 scores outputted from the TrainImagesClassifier function for all mineral phases. F1 scores should be above ~0.95 for all but the least abundant phases (in our case, the F1 score of chlorite was below this threshold for a single sample). If F1 scores are below this level across the board, then the inclusion of other elemental intensity rasters in the virtual raster should be investigated.

Step 5

a) Predict the full mineral map.

i) Open the 'ImagesClassifier' function in Orfeo Toolbox.

ii) For 'Input Image', select the virtual raster for that sample.

iii) For 'Model file', select the .mod file created above in Step 4.

iv) Set 'Number of classes in the model' to the number of trained mineral phases.

v) Set 'Output pixel type' to 'int8' to minimize output file size.

vi) Choose a memorable name for 'Output Image'. This will be the name of the predicted mineral map.

vii) Set 'Confidence map' and 'Probability map' to 'Skip Output', as these are not generated for random forest models.

viii) Click 'Run' to yield the predicted mineral map for the entire thin section.

Step 6

a) Use the SAGA 'Majority/Minority filter' function to apply a majority filter to the predicted mineral map to eliminate spurious noise (e.g., single pixels identified as a given mineral phase encompassed within another mineral phase, or single-pixel rims around grains).

i) For 'Grid', select the predicted mineral map.

ii) For 'Type', select '[0] Majority'.

iii) For 'Kernel Type', select 'Circle'.

iv) For 'Radius', select '10'. (The optimal value for this is sensitive to the resolution of the elemental intensity rasters, so some experimentation may be needed to optimize the accuracy of the predicted mineral maps. In the examples in the main text, we found that a radius of 5-10 pixels optimized the accuracy of the predicted mineral maps.)

v) For 'Filtered Grid', select 'Save to a Temporary File'. Otherwise, in our experience, SAGA functions will only output a SDAT file (.sdat), which is a format specific to SAGA GIS.

vi) Export the temporary file as a TIF file. This yields a majority-filtered mineral map.

b) Clip the majority-filtered mineral map to a polygon outline of the sample. This will eliminate the non-sample portion of the thin section, which will enable accurate calculation of mineral abundances in future steps, if desired. When clipping the map to the sample outline, use GDAL's 'Clip Raster by Mask Layer'.

Optional

a) If desired, convert the raster mineral map to a vector mineral map using the GDAL 'Polygonize' function.

i) For 'Input Layer', select the majority-filtered mineral map.

ii) There should only be 1 band.

iii) Check 'Use 8-connectedness'. This allows for better handling of diagonally-connected pixels.

iv) For 'Name of the field to create', use the same mineral number used in the training data.

v) Save to an ESRI shapefile (.shp and other associated files).