

README file for software to accompany

P. D. Asimow, L. C. Stein, J. L. Mosenfelder and G. R. Rossman
Quantitative polarized infrared analysis of trace OH in populations
of randomly oriented mineral grains
Submitted to American Mineralogist, March 2005

1. This README accompanies and documents two programs, pbgi and pdc. pbgi is for subtracting backgrounds from infrared spectra. pdc is for determining orientation of the polarization vector in measured spectra using standard spectra in the principal directions on the same mineral, and for synthesizing principal spectra from a set of randomly oriented spectra. See the paper for details.
2. Each program is distributed as C source code and as executables compiled for Windows, MacOS X, and Linux. The source code may be freely modified and redistributed, but please cite the above paper for all uses and ensure that a comparable message accompanies any modified versions.
3. It is best to run pbgi from a terminal window. The way to get a terminal window varies among operating systems. On Linux, any shell or command window will suffice. On MacOS, under Applications/Utilities, use the Terminal program. On Windows, on the Start Menu under Programs/Accessories you will find the "Command Prompt" program, which brings up a command line window.
4. The versions that are being distributed with the paper are version 1.0. Updates will be posted at <http://minerals.gps.caltech.edu>. Please send bug reports to Paul Asimow at asimow@gps.caltech.edu.
5. How to run pbgi
 - a. If you just invoke pbgi from the command line with no arguments, it prints a help summary with a list of all the command line arguments and a short description of each.
 - b. If you run pbgi -i, you get an interactive mode, where you will be prompted to set each of the options that control the behavior of the program. This way you don't have to remember the meaning of any of the other command line options.
 - c. The program requires at least one input file with a spectrum to be backgrounded. This is a plain text file, csv format, with a wavenumber, absorbance pair on each line. There must be a comma between the wavenumber and the absorbance; other whitespace characters are ignored. Any line-endings (DOS, Windows, Mac, Unix) are OK. There are no header lines. The file examplespectrumraw.csv included with this distribution is a legal input file. It is best if the filename ends with .csv for all inputs.
 - d. You always get an output file matching each input file with either the background or the background-subtracted spectrum (see the -bgs option below) in the same format as the input; the filenames are generated by

inserting .bg or .bgs before the .csv. There are also options to generate an omnibus output file with all the backgrounds or background-subtracted spectra (-o) generated in a given execution, and to concatenate all the input spectra into a single file with columns labelled by the input filenames (-oi).

e. All the options:

-i

puts you in interactive mode. Default is non-interactive. The rest of the command line is still read, but the answers to prompts can override command-line options.

-bgs

tells the program to output background-subtracted spectra. The default is to output the backgrounds instead. You could then subtract them yourself elsewhere. If -bgs is active, the output filenames end with .bgs.csv; otherwise .bg.csv .

-o omnibusfilename

Activates output of a file into which all the backgrounds or background-subtracted spectra from a given run are concatenated. Each column is labeled by the filename of the input. Default is no omnibus output file.

-oi omnibusinputfile

Activates output of a file into which all input spectra used are echoed. Each column is labelled by the filename of the input. Note, the name of this option is perhaps a bit misleading - this file is written, not read, and it is not a way to get a bunch of spectra input to the program all at once. Default is no omnibus input file.

-p degree

Flags the program to use a polynomial fit of the specified degree to the selected anchor points rather than a cubic spline. We have had better luck with the cubic spline fit, but the polynomial option is included as an option. Default is cubic spline.

-1 lowwavenumber

-2 highwavenumber

Indicates the wavenumber range over which background fitting is performed. Data outside this range are copied as-is into the background-subtracted output file or set to zero in the background output file. The default range is 1100 to 3970 cm^{-1} .

-a #anchors

Specified how many anchor points are located for fitting the background to the spectrum. The program begins with a line connecting the datapoint at lowwavenumber to the datapoint at highwavenumber. It then searches for the point on the spectrum between these ends that is farthest below this line. It adds this as an anchor and then searches for the point that is furthest below the piece-wise linear fit through these three points. This is iterated until #anchors are found. Default 25. Values up to 100 give a closer fit to the spectrum, but may overfit noise in some cases.

`-z low1 high1 [-z low2 high2] ...`

This is one way of dealing with artifactual peaks, for example from organic contamination. Within each range specified by a low and high wavenumber limit, the difference between the data and a line between the endpoints is added to the background. This has the result that the background-subtracted spectrum is linear between the endpoints. Peaks that occur in these ranges are therefore taken to be background features, not data. This operation is done AFTER the anchor search. Use this option for positive peaks - See the `-pz` option for negative peaks (e.g. from bad correction of air CO₂ absorption). Specify as many `-z` regions as you like. Default is no `-z` regions.

`-pz low1 high1 [-pz low2 high2] ...`

Another way of dealing with artifacts. In this case, the region between the low and high wavenumber limits specified is replaced with a line BEFORE the anchor search. This is needed when the spectrum contains negative background peaks, like from undercorrection for CO₂ in air along the optical path. Without the `-pz` option, the anchor search will pick the bottoms of these negative peaks as anchors and will not therefore fit the desired background. Removing these peaks with `-pz` will prevent this problem. Specify as many `-pz` regions as you like. Default is no `-pz` regions.

`-f fixedanchor1 [-f fixedanchor2] ...`

This option inserts a fixed anchor, i.e. a wavenumber value at which the background is forced to equal the data spectrum, and the background-subtracted spectrum is forced to be zero (exactly zero for cubic spline fitting, optimized towards zero in polynomial fitting). Use this option when the automatic anchor search does not do what you want. For example, if your background has a concave-down region, you will need some fixed anchors to pull the background up into the concavity. Default is on `-f` points.

`-x low1 high1 [-x low2 high2] ...`

This specified forbidden regions for the automatic anchor search. It will not pick any anchors between the specified low and high limits. This is another way of dealing with troublesome regions like negative background peaks. Specify as many `-x` regions as you like. Default is no `-x` regions.

`-g linearpoint1 [-g linearpoint2] ...`

Specifies a location where the cubic spline between the anchors immediately to either side will be replaced by a linear fit instead. This helps deal with locations where the cubic spline is "too wiggly". Note that the property of the cubic spline is such that introducing linear regions does not introduce any kinks - the first derivative remains continuous at the anchor points bounding a linear region. `-g` has no effect during polynomial fitting. Default is no `-g` points.

`-Z epoxyfilename`

Activates subtraction of a spectrum given in `epoxyfilename`, and scaled as specified by the `-Zp` parameter or its default. This is used to remove the complete spectrum of a known contaminant, for example mounting epoxy, instead of individual contamination peaks as with `-z` and `-pz` above. This assumes that all the spectral features due to contamination scale linearly, which may not be a very good assumption. The spectrum given in

epoxyfilename should be background-subtracted already! Please ensure that the wavenumbers sampled in the epoxy spectrum are exactly the same as those in the data spectrum being subtracted. By default this feature is off.

-Zp peaklocation

Indicates the wavenumber at which the amplitude of the contamination spectrum is estimated. The height of the data spectrum above the background calculated before epoxy subtraction is compared to the absolute height at this wavenumber in epoxyfilename, and the entire epoxy spectrum is scaled by the ratio of these and then subtracted from the background-subtracted output or added to the background output. Default 3925.126. If the data and epoxy spectra do not have a sample at exactly this wavenumber, the nearest one is used.

infile1 [infile2] [infile3] ...

All other arguments on the command line are taken to be input files. See above for the format.

f. The following command string, for example, does a good job on examplespectrumraw.csv:

```
pbg -bgs -a 100 -g 1800 -pz 2300 2400 -z 2820 3050 examplespectrumraw.csv
```

This should produce the background-subtracted spectrum given in examplespectrumraw.bgs.csv.

6. How to run pdc

a. If you just invoke pdc from the command line with no arguments, it prints a help summary with a list of all the command line arguments and a short description of each.

b. If you run pdc -i, you get an interactive mode, where you will be prompted to set each of the options that control the behavior of the program. This way you don't have to remember the meaning of any of the other command line options.

c. The program requires an input file with all the spectra to be analyzed. This is a plain text file, csv format, with the following specification.

- Optional header lines at the beginning, for example to label the spectra. The number of header lines is set by the users with the -hl flag (default is 2). These are not processed, they are simply copied to the output file.
- A line of sample thicknesses, in the same units to which your standards are normalized. So, if the standards are given as 1 mm thick, then give unknown sample thicknesses in mm. Note, this line should start with a comma, since the first column in subsequent lines is the wavenumber, and the samples start at the second column. Hence whatever precedes the first comma on the thickness line is ignored.
- All subsequent lines give wavenumber and then the background-

subtracted absorbance of each unknown spectrum pair on each line.

All spectra must be sampled at the same wavenumbers. Spectra can be either normalized or not (see the -n flag below), but all must be the same option.

See the example in examplepdcinput.csv. Any line-endings (DOS, Windows, Mac, Unix) are OK.

d. It also needs a standard file. This file, also csv format, contains the E||a, E||b, and E||c standard absorbance spectra, background-subtracted and normalized. There are no header lines, and the format of each line is wavenumber, a, b, c.

e. You get an output file giving the orientation angles determined for each spectrum in the input file, the thickness ratios, and the goodness of fit; the filename is generated by inserting .decon before the .csv. The header lines are copied from the input file in order to label the orientations the same way the spectra were labeled in the input file.

f. If the -y synthfile option is set there is another output file with the synthetic principal spectra, csv format. It also gives the integrals of these spectra between the limits given, and the water contents estimated from these integrals.

g. All the options:

-i

Puts you in interactive mode. Default is non-interactive. The rest of the command line is still read, but the answers to prompts can override command-line options.

-t

Indicates fixed thickness in the orientation fitting. Otherwise thickness is taken as a fitting parameter, and the ratio of best-fitting to nominal thickness is reported in the output file. Default is thickness fitting on.

-n

Indicates that the input spectra are normalized to a thickness of 1 of the units in which thickness is given, i.e. the same normalization as the standards. In this case they are denormalized to the measured thickness before fitting. The default is non-normalized input.

-s stdfilename

Gives the filename of the standards csv file, see above.

-o outputfilename

Specifies the outputfilename. By default .decon is added to the input filename before the .csv.

-hl headerlines

Dictates how many lines occur in the input file before the line of sample thicknesses. These are not read, they are merely copied to the output file.

-1 lowwavenumber

-2 highwavenumber

These indicate the range over which the unknown spectra are compared to the standard spectra to determine their orientation. The default is 1200 to 2200 cm^{-1} .

-x low1 high1 [-x low2 high2] ...

These indicates regions between low and high limits that are ignored in the orientation fitting. If there are contamination peaks in the silicate region used for fitting, their effect can be neutralized with this option. Specify as many -x regions as you like. Default is none.

-y synthfile

Turns on the option to generate synthetic principal-axis spectra and gives the filename where they are written. The synthfile also contains the integrals of the principal axis spectra, and the water contents converted from each orientation and total, using the Bell et al. calibration factor of 0.188 ppm-cm.

-int low high

Gives the wavenumber range for integration of the synthetic principal axis spectra. Default 3000 to 3750 cm^{-1} .

infile

The filename of the input file with the unknown spectra. See above.

h. The following command string, for example, takes examplepdcinput.csv and does the orientation and synthesis exercises using the standards file grr997.csv:

```
pdc -t -n -s grr997.csv -hl 1 -x 1720 1740 -y examplesynth.csv examplepdcinput.csv
```

This should produce the orientations given in examplepdcinput.decon.csv and the synthetic spectra given in examplesynth.csv.