Documentation for the Gros-Eilers-Arey code to perform baseline correction of GC×GC chromatograms as implemented in Matlab

Version 1.0.1

Jonas Gros, Paul H. C. Eilers, and J. Samuel Arey, 2016.

BY USING THE CODES, THE USER AGREES TO THE LICENSE TERMS STATED IN THE FILE LICENSE.txt

Please cite the following articles when publishing any results obtained by use of this software:

- Eilers, P. H. C., "Parametric time warping", Anal. Chem. 2004, vol 76, p 404–411.

- Gros, J.; Reddy, C. M.; Aeppli, C.; Nelson, R. K.; Carmichael, C. A.; Arey J. S., "Resolving biodegradation patterns of persistent saturated hydrocarbons in weathered oil samples from the *Deepwater Horizon* disaster", Environ. Sci. Technol. 2014, vol 48, num 3, p 1628-1637.

1   The purpose of the algorithm

The Matlab code is designed to correct the baseline in chromatograms obtained with comprehensive two-dimensional gas chromatography (GC×GC) coupled to a univariate detector, for example a flame ionization detector (FID), or an electron capture detector (ECD). The use of these baseline-correction codes was shown to perform favorably for the quantification of individual analytes when combined with the watershed algorithm as implemented in the GC Image software (Gros et al. 2014 and Samanipour et al. 2015).

Other baseline-correction methods may be more appropriate for different needs (see discussion in Gros et al. 2014).

Note: this Matab code will not work with multivariate data (mass spectrum detector).

2   What the Matlab code does

It applies the one-dimensional Eilers 2004 baseline-correction code independently to each second dimension modulation.

3   Organization of the model file directory. Where to find what.

The model code is organized as follows. From the base directory of the program two folders are present, called `users/`, and `model_code/`.

These two folder names should not be changed.

The user should only need to operate from within the folder called `users/`. Normally, nothing should be changed or adjusted in the folder called `model_code/`.

Within the folder called `users/`, the organization of folders and files is user-defined. The user can define directory paths with the following two model variables in the file `main.m`:

    A) `input_path`. This variable indicates the directory path location of the input files. The input path variable is set in the file called `main.m`, and it assumes that `main.m` is located in the directory `users/`. The `input_path` variable also assumes that the indicated directory exists.

    B) `output_path`. This variable indicates the directory path location of the output files.

<u>Note</u>: both `input_path` and `output_path` should be relative paths, starting with `users/`, which itself should be situated within the program base directory.

<u>Note 2</u>: The operating system must allow Matlab to write files within the `output_path` directory. For example, on windows computers, **<u>do not</u>** locate the base directory within the `C:\Program Files` folder.

4   Steps for use of the algorithm

4.1   Prepare input files

The model requires a chromatogram. The required structure of this file is:

By default, this chromatogram should be a csv file containing one long column vector of signal intensity values, separated by a comma (,) or semi-colon (;) (as can be exported from GC Image). The length of the file is assumed a multiple of the product of the sampling rate multiplied by the modulation period, and any last additional values would be ignored. Other file types might be used (two-dimensional csv files in the GC Image format, or multi-column csv files exported from ChromaTOF, in which case the column labeled "S1" or "TIC" is imported).

4.2   Adjust parameters in `main.m`

Adjust the parameter settings that appear in `main.m`. This file can be read and modified from within Matlab or using a generic text editor. This is the only Matlab file that you need to adjust. Most of these parameters are self-explanatory and/or discussed above.

The three tuning parameters of the Eilers code are discussed below.

- `lambda`: this is the usual tuning parameter. Typically, lambda will be between 4 and 8 (corresponding to 10^4 to 10^8). Different values have been evaluated and discussed (Gros et al. 2014 and Samanipour et al. 2015), and values of 4 or 5 were performing the best in those works.

- `p`: Usually 0.001 to 0.02. Usually 0.001 according to Eilers 2004 for one-dimensional data; 0.02 has been used by Gros et al. 2014 and Samanipour et al. 2015 for GC×GC chromatograms.

- `d`: 2 by default (not to be changed).

In order to obtain a reasonable result with these baseline-correction codes, modification of the parameters lambda or p might be required.

5 Name and contents of the output file

The baseline-corrected chromatogram is saved as a one-dimensional column vector in a csv file (this step takes some time; an estimate of this time is displayed in Matlab's main window).

The name of the file is the same as the name of the target chromatogram file, with the string "_BSLN_CORR" appended.

6 Figure displayed

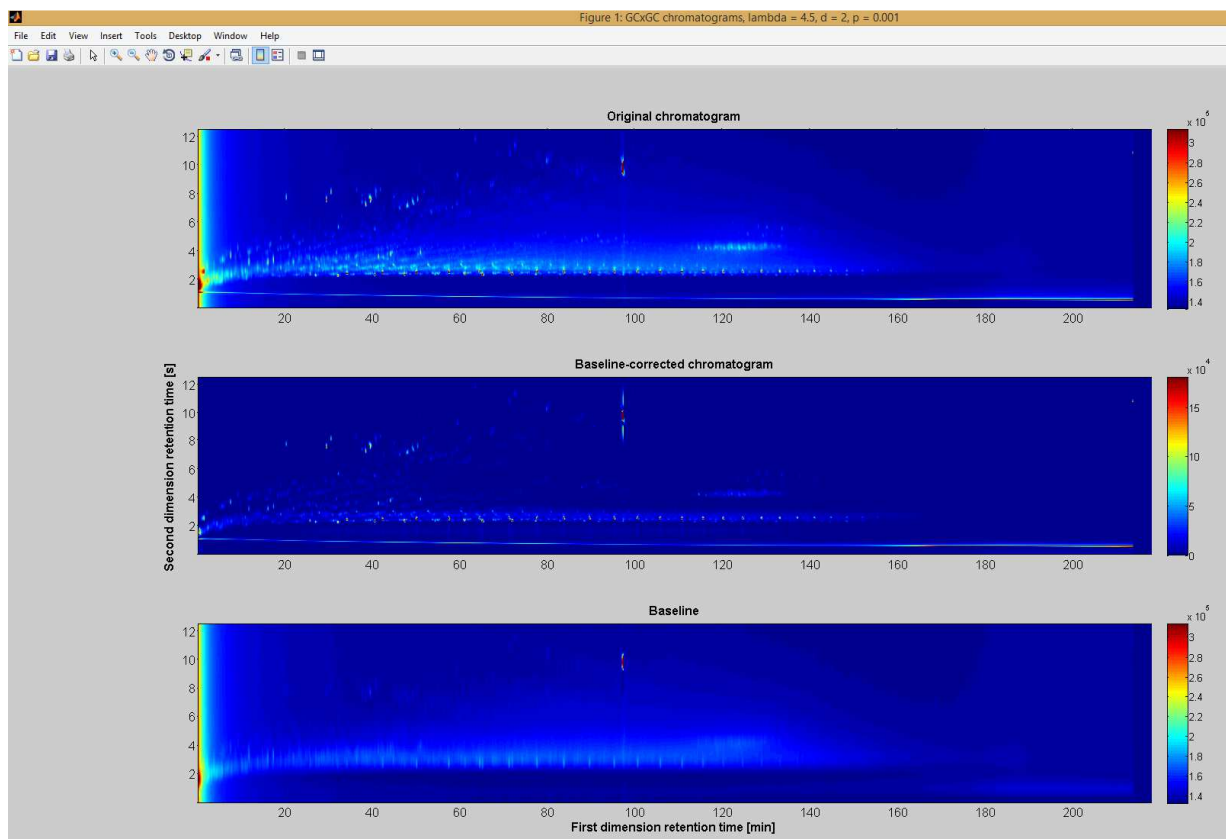After the completion of the baseline correction, one figure is displayed:

Figure 1: Figure displayed by the code. Here, as an example, the figure obtained for the example chromatogram "Grane_crude_oil.csv", a crude oil GC×GC–FID chromatogram provided in the folder input (this figure can be obtained by running the original codes without any edit). The figure contains three panels: the original chromatogram; the baseline-

corrected chromatogram; the baseline (from top to bottom). The values of the three parameters lambda, p, and d are provided in the figure name. Note that color axis of panels a and c do not start at zero, contrary to the color axis of panel b.

7   References

- Eilers, P. H. C., "Parametric time warping", Anal. Chem. 2004, vol 76, p 404–411.

- Gros, J.; Reddy, C. M.; Aeppli, C.; Nelson, R. K.; Carmichael, C. A.; Arey J. S., "Resolving biodegradation patterns of persistent saturated hydrocarbons in weathered oil samples from the *Deepwater Horizon* disaster", Environ. Sci. Technol. 2014, vol 48, num 3, p 1628-1637.

- Samanipour, S.; Dimitriou-Christidis, P.; Gros, J.; Grange, A., Arey, J. S., "Analyte quantification with comprehensive two-dimensional gas chromatography: Assessment of methods for baseline correction, peak delineation, and matrix effect elimination for real samples", J. Chrom. A 2015, vol 1375, p 123-139.

Contacts:

For questions, problems, or bug reports, feel free to contact Jonas Gros (gros.jonas@gmail.com) or J. Samuel Arey (arey@alum.mit.edu). (Or Paul Eilers (p.eilers@erasmusmc.nl) for the one-dimensional code.)