

User's Guide

GC × GC-MS-Property-estimation_v1.1.0

Version 1.1.0

©Yasuyuki ZUSHI

2018.7

Please cite the following article when users use this source code.

Zushi Y., Yamatori Y., Nagata, J., Nabi D., [Comprehensive two-dimensional gas chromatography \(GC × GC\)-based property estimation to assess the fate and behavior of complex mixtures: A case study of vehicle engine oil.](#)

Description:

This R source code, which is available from the GitHub repository “GCxGC-MS-Property-estimation”, has been developed to estimate properties from GC × GC data. When GC× GC-MS data are provided in cdf format and the retention time data of the alkanes are provided in csv format as input data, the tool calculates properties, such as $\log K_{O-W}$. When a user wants to use a different column combination from the default setting for the GC × GC-based property estimation (DB1-BPX50), another file of retention-time data with Abraham solvation parameters is required for calibration chemicals. The property profiles of a target chromatographic peaks are provided as a spider chart. The chromatographic region on environmental behaviors, such as long-range transport potential, is depicted and provided as GC × GC-MS data as cdf formatted output.

Author (Contact):

Yasuyuki ZUSHI (yasuyuki.zushi@gmail.com)

Requirement:

Hardware computer (≥ 16 GB RAM)

Software R freely available from <https://cran.r-project.org/>

R package “plotrix”*

R package “pracma”*

R package “RNetCDF”*

*These packages are installed along with the code execution.

How to use:

1. Download the zip file “GCxGC-MS-Property-estimation_v1.1.0”, unzip it, and place the folder where you prefer.
2. **GC × GC-MS data:** Open the “GCxGC-MS-Property-estimation_v1.1.0” folder, and place your GC × GC data in cdf format in the “Input” folder.

Dataset for calibration: Property estimation requires extensive information, such as retention times and descriptors of chemicals, the properties column properties, and system parameters of properties of interest as input data. As a default setting, the information for the measurement condition of the reference¹⁾ (column combination of DB1 as the first column and BPX50 as the second column) have been set up to test this tool.

If you want to change the measurement setting, besides the column combination of DB1-BPX50, the alkane retention times must be measured and input to the “Ni_data_Alkane.csv” file. If you want to change the column combination from DB1-BPX50, the 1.

“Chemical_list_for_LFER_Calibration.csv”, 2. “Column_System_Parameter.csv”, 3. “Exp_property_data.csv”, and 4. “Ni_data_Alkane.csv” files must be updated. The “default.alpha.csv” and “default.prop_coef.csv” file are generated/updated automatically, and are used for some processing. Do not delete these files.

3. Start R software, then open the code “Just_Run_Me.r” included in the folder as a R script file.
4. In the R script, rewrite the folder location path that is suitable for your own environment. Rename an output file name, if you prefer.
5. If you have not installed the R package “plotrix”, “pracma”, and “RNetCDF”, it must be installed as an initial setting. As a default setting, the package installation will proceed. This step can be omitted the next time the code is run.
6. Change or retain default parameters, such as Input_measuremet_data, parameter_calibration, first.column, second.column, Modulation_Period, Sampling_Rate, Phase_Shift, Initial_RT, and the domain.name according to the reference.¹⁾ See comments provided in “Just_Run_Me” for each paramter.
7. After setting all the parameters, click [Edit] => [Run all]. An output file will be produced in the “Output” folder. Several tens of seconds are required to complete the process for the test dataset that was included initially in the folder.
8. In the “Output” folder, the property profile of the target peak, their text format data (csv file), and the GC × GC-MS chromatogram of the focusing domain (cdf file) are generated.

9. If you would like to perform a property estimation for other chromatographic peaks, their retention times must be entered in the “Input_RTs.csv” file in the input folder. Items include the pID, compound name, and formula. The assignment does not necessarily need to be identified, therefore, arbitrary letters will result if a case is not obtained. For the compound name, a unique input name, such as “unknown1”, “unknown2”, and “unknown3”, is used because the names are used for saved file names of property profiles.

The R code, which was developed in this study¹⁾, is available via GitHub (<https://github.com/Yasuyuki-Zushi/GCxGC-MS-Property-estimation>).

Supplementary:

GC × GC property-estimation algorithm, which is used for GC × GC univariant data is available from <https://github.com/jsarey/GCxGC-property-estimation>.²⁾

Reference:

- 1) Zushi Y., Yamatori Y., Nagata, J., Nabi D., Comprehensive two-dimensional gas chromatography (GCxGC)-based property estimation to assess the fate and behavior of complex mixtures: A case study of vehicle engine oil,
- 2) Deedar Nabi, Jonas Gros, Petros Dimitriou-Christidis, and J. Samuel Arey, Mapping environmental partitioning properties of nonpolar complex mixtures by use of GC×GC, Environmental Science & Technology 2014, vol 48, p 6814-6826.