User's Guide

# GCxGC-MS-Property-estimation_v1.1.0

Version 1.1.0

©Yasuyuki ZUSHI

2018.6

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

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,

**Description:**

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

**Author (Contact):**

Yasuyuki ZUSHI (yasuyuki.zushi@gmail.com)

**Requirement:**

Hardware computer ($\geqq$16GB 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 of "GCxGC-MS-Property-estimation_v1.1.0", unzip it, and place the folder where you prefer.

2. <u>GCxGC-MS data</u>: Open the "GCxGC-MS-Property-estimation_v1.1.0" folder, and place your GCxGC data as cdf format in the folder of "Input".

   <u>Dataset for calibration</u>: The property estimation required several information, such as retention time and descriptor of chemicals, their properties column properties, and system parameters of interested properties as input data. As a default setting, all these information for the measurement condition of [ref] (column combination of DB1 as $1^{st}$ column and BPX50 as $2^{nd}$ column) have been set up to simply experience this tool.

   If you want to change the measurement setting except for the column combination of DB1-BPX50, it is required to measure the retention time of alkanes which should be input to the file of "Ni_data_Alkane.csv". If you want to change the column combination from the DB1-BPX50, it is required to update the file of 1. "Chemical_list_for_LFER_Caliration.csv", 2. "Column_System_Parameter.csv", 3. "Exp_property_data.csv", 4. "Ni_data_Alkane.csv". The file of "default.alpha.csv" and "default.prop_coef.csv" are automatically generated/updated, and used for certain processing, therefore, don't try to delete.

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 fit for your own environment. Rename an output file name, if you prefer.

5. If you have not install R package "plotrix"," pracma" and "RNetCDF", required to install it as an initial setting. As a default setting, the package installation will be proceeded. Next time to run the code, you can skip this process.

6. Change or keep default parameters, such as Input_measuremet_data, parameter_calibration, first.column, second.column, Modulation_Period, Sampling_Rate, Phase_Shift, Initial_RT, and domain.name according to the [ref] and comments in "Just_Run_Me".

7. After setting all the parameter, click [Edit]=>[Run all], then output files will be produced in the "Output" folder. To complete the process for the test dataset that initially included in the folder, it takes several tens of seconds.

8. In the "Output" folder, property profile of target peak, their text format data (csv file), GCxGC-MS chromatogram of focusing domain (cdf file) are generated.

9. If you want to perform the property estimation for other chromatographic peaks, input their retention times in the file of "Input_RTs.csv" in the input folder. The item of pID, compound name, formula, Assignment need not necessarily to be identified, therefore, just input arbitrary letters if the case they are not obtained. For the compound name, input unique name, such as "unknown1", "unknow2", "unknown3", because they are used for saved file name of property profile.

The R code, which is developed in the study[1], is available via GitHub ().

**Supplementary:**

GCxGC property estimation algorithm, which is used for GCxGC 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.