Paramounter.R User Manual

(Version 2, 2022-01-28)

Jian Guo[1], Tao Huan[1,*]

[1] Department of Chemistry, Faculty of Science, University of British Columbia, Vancouver Campus, 2036 Main Mall, Vancouver, V6T 1Z1, BC, Canada

[*] Author to whom correspondence should be addressed:

Dr. Tao Huan
Tel: (+1)-604-822-4891
E-mail: thuan@chem.ubc.ca
Internet: https://huan.chem.ubc.ca/

- Paramounter.R is an R script for optimizing universal parameters for processing LC-MS-based metabolomics data.
- The program is written in the language 'R' and is publicly available at https://github.com/HuanLab/Paramounter.git
- Please see below for detailed instructions on using the Paramounter.R code.

1) **File preparation.** User needs to create a folder to store the files used for parameter optimization. It is recommended to use 3-4 QC data files from the beginning, middle, and end of the sample analysis if it is a large cohort study. All the QC mzXML files need to be put in the folder user has created. As shown in **Figure 1**, it illustrates the details of how the files should be organized in the corresponding folders. If user intends to use mzML files, please change the R code line 19 in part 1 and line 21 in part 2 to "filename <- list.files(pattern = ".mzML")".

Figure 1. The organization of the files used for Paramounter.

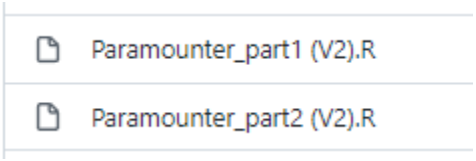


Figure 2. Paramounter part 1 and part 2 on the GitHub.

2) Download the R-scrip "Paramounter_part1 (V2).R" and "Paramounter_part2 (V2).R" separately from Github (https://github.com/HuanLab/Paramounter.git) as shown in **Figure 2**.
3) **R package installation.** In R-studio, user needs to first install libraries "xcms", "MSnbase", "dplyr", "ggplot2", "gridExtra" if they are not previously installed. (R Version 4.0 or above, all the packages should be updated to the newest available version)
4) **Parameter setting for Paramounter part 1.** After all the required libraries are successfully installed. User needs to first set the directory, smooth value, and massSDrange for Paramounter part 1. The "directory" is where the mzXML files are located. The "smooth" value is 0 by default. However, if user is processing data generated in full-scan with spectral acquisition rate higher than 2 Hz, it is recommended to smooth out the raw data to achieve a more accurate parameters estimation. The recommended smooth values are identical to the spectral rate (e.g. smooth = 4 for spectral rate = 4 Hz). The "massSDrange" value is 2 by default. User can adjust this value if they want the mass tolerance calculated by the standard deviation of mass difference multiplied by a different value (e.g. 3).
5) **Run Paramounter part 1**. In R-studio, click on "→Source" in the top right corner of the R-studio interface to start running the code of part 1. In the end, Paramounter part 1 generates a plot in the bottom right corner showing the distribution of mass accuracy in the dimension of *m/z*. User can pick the ppmCut by visualizing the distribution. The ppmCut is the maximum ppm value within which the high-density points should be located (e.g. here 40 ppm is selected as the ppmCut in **Figure 3**).

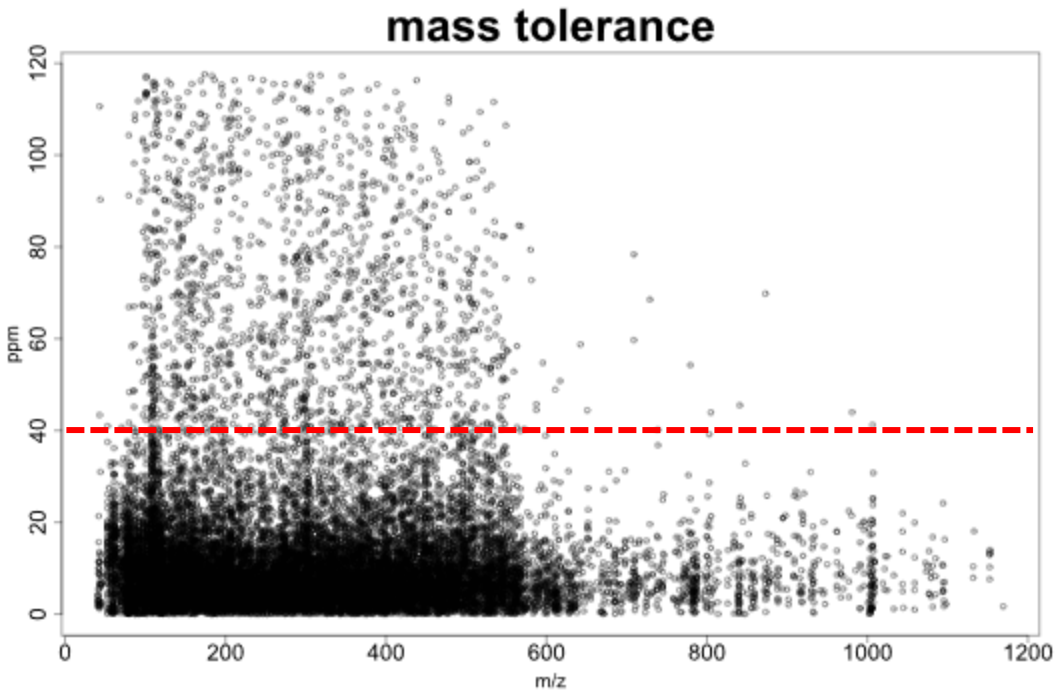


Figure 3. The mass tolerance distribution generated by Paramounter part 1.

6) **Parameter setting for Paramounter part 2.** After running the scrip of Paramounter part 2 and determining the ppmCut from part 1, user can open Paramounter part 2 and set the parameters. The "directory" is set the same as part 1. The "software" should have the corresponding name of the software user wants to optimize the parameters for. User can select either XCMS, MSDIAL, and MZMINE2 for specific parameters, or Universal for universal parameters. "ppmCut" is set to be the value user determines from part 1. "smooth" is set the same value with part 1. "massSDrange" is also set the same value with part 1.
7) **Run Paramounter part 2**. In R-studio, click on "→Source" in the top right corner of the R-studio interface to start running the code of part 2. After the code finishes running, a folder with the name "Universal parameters for the software selected by user" is created within the directory. The folder contains one png file and one pdf file as shown in **Figure 4**.

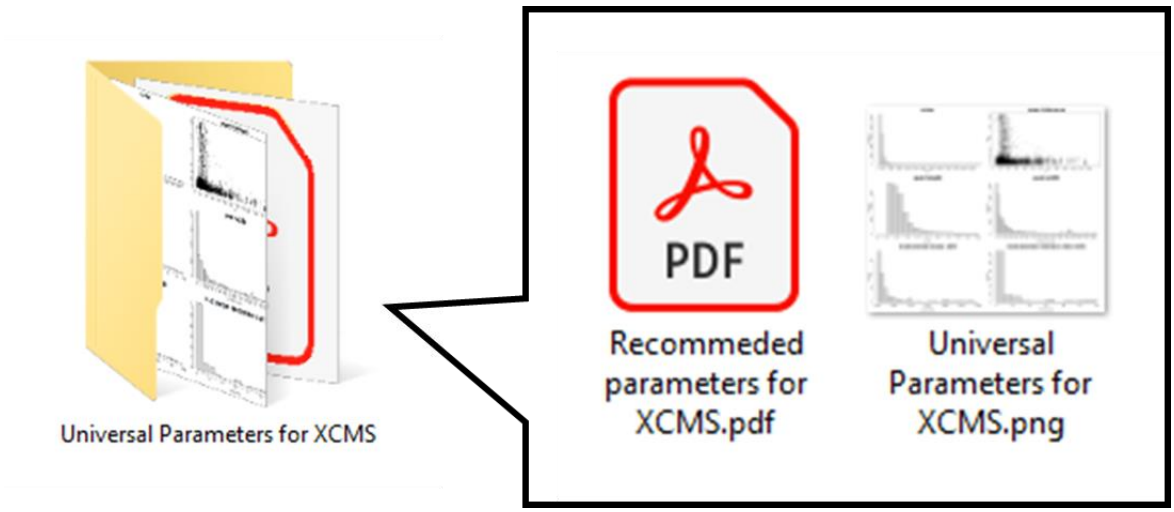
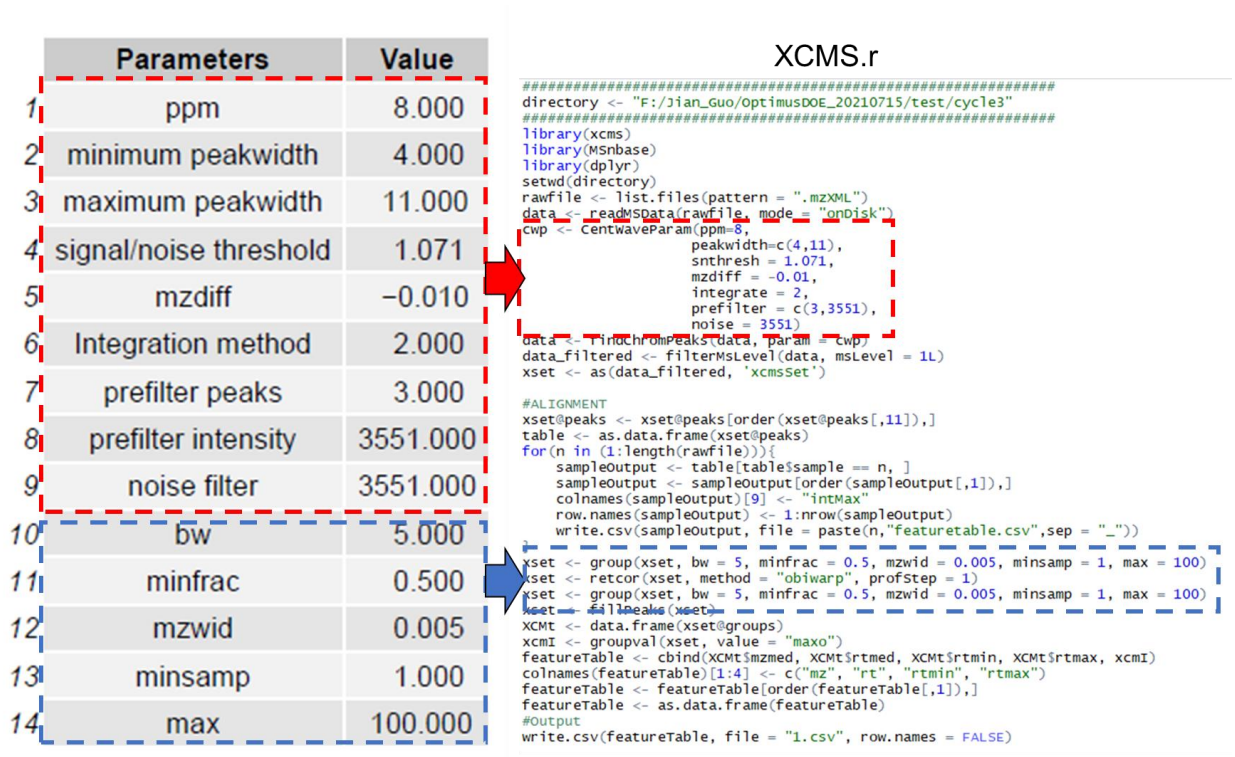


Figure 4. The output of Paramounter part 2.

8) **How to use the results**. User can directly open the pdf file and import the recommended values into the corresponding data processing software. As exemplified in **Figure 5**, user can copy the recommended values for XCMS into either XCMS online or XCMS.R. User can also open the complimentary png file to visualize the distribution of each universal parameters for the raw data.
9) **Test different data processing software**. If user wants to generate the universal parameters for all three data processing software, they do not need to change the software name and run the code all over again. User can simply input "ALL" to software and three folders with recommended parameters for XCMS, MSDIAL, and MZmine2 will be generated simultaneously. User can also input "Universal" to software to only obtain the universal parameters instead of specific recommended parameters for each software.



	Parameters	Value
1	ppm	8.000
2	minimum peakwidth	4.000
3	maximum peakwidth	11.000
4	signal/noise threshold	1.071
5	mzdiff	−0.010
6	Integration method	2.000
7	prefilter peaks	3.000
8	prefilter intensity	3551.000
9	noise filter	3551.000
10	bw	5.000
11	minfrac	0.500
12	mzwid	0.005
13	minsamp	1.000
14	max	100.000

Figure 5. How to copy the recommended parameters into XCMS.R.