

# MassAnalyzer 1.0 Instructions

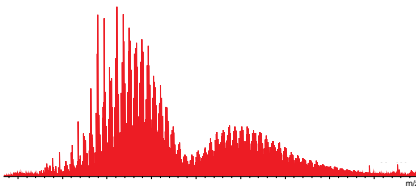
## I. PROBLEM SOLVING

MassAnalyzer v1.0 is a machine learning-based software, which can be used to calculate molecular formulas for nuclear ratio of organic matter on anion mass spectrometry containing nine elements of Carbon(C), Hydrogen (H), Oxygen (O), Nitrogen (N), Sulfur (S), Chlorine (Cl), Bromine (Br), Fluorine (F) and Phosphorus (P). Currently, a anion pattern is supported in MassAnalyzer: M-H.

The program can be accessed through [MassAnalyzer: Machine learning-based mass spectrometry discovery \(sjtu.edu.cn\)](http://MassAnalyzer: Machine learning-based mass spectrometry discovery (sjtu.edu.cn)). Before running this program, please read this description carefully and set the parameters of the calculation, confirm that the input mass spectrometry file format is correct, and leave an email address, and the calculation results will be sent to your mailbox within 24h.

## II. INPUT FILE FORMAT

The input file should be in '.xlsx' format for Excel. The file should contain four columns with the headings 'm/z', 'I', 'Res.', 'S/N'. The title order can be changed but the case and symbols must be correct, where 'm/z' represents the plasma-core ratio, 'I' represents the peak intensity, 'Res.' represents the resolution, and 'S/N' represents the signal-to-noise ratio. If information on resolution and signal-to-noise



ratio is missing, this column is set to 0. For specific Excel forms, please refer to the example download in the web page.

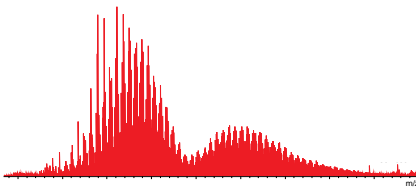
1	m/z	I	Res.	S/N	Note 1 Each column is entitled "m/z", "I", "Res.", "S/N".
2	100.7077	4043925	1346738	0	Note 2 The order of the headings can be changed, but the case must be correct.
3	106.2823	3481865	1285391	0	Note 3 m/z is the plasm-core ratio and I is the peak intensity, these two sets of data must be available.
4	106.9479	3823333	904969	0	Note 4 Res. is the resolution, S/N is the signal-to-noise ratio, if these two sets of data are not available, please set 0.
5	109.0294	4677805	974085	0	Note 5 When uploading, except for the first line of titles, how letters and Chinese characters cannot appear.
6	110.6183	8203261	899352	0	
7	118.5385	7229382	758144	0	
8	118.5419	23600582	956166	0	
9	126.948	4281354	835590	0	
10	126.9512	4113162	835031	0	
11	126.9541	3482762	861577	0	
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					

Figure 1. MassAnalyzer 1.0 input file format.

### III. PARAMETER SETTING

The maximum relative error is the maximum relative error. The mass spectrometer determines the relative error size. The unit selected by this software is ppm, that is, one millionth. The formula is  $\text{Acc} = (\text{experimental value} - \text{theoretical value}) / \text{theoretical value} * 1000000$ . Please enter an integer greater than 0.

Among them, the maximum number of nitrogen atoms, sulfur atoms and chlorine atoms are to resolve the maximum allowable number of N, S, Cl atoms in the molecular formula. You need to enter an integer of not less than 0. It is recommended to be slightly larger than the number of

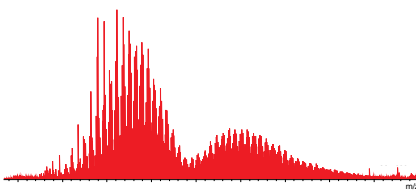


known organic atoms in the experiment. If the sample does not contain S or Cl, enter 0. Bromine (Br), Fluorine (F), and Phosphorus (P) are three optional choices. If it is confirmed that the molecular formula to be analyzed contains any of these elements, check the corresponding element.

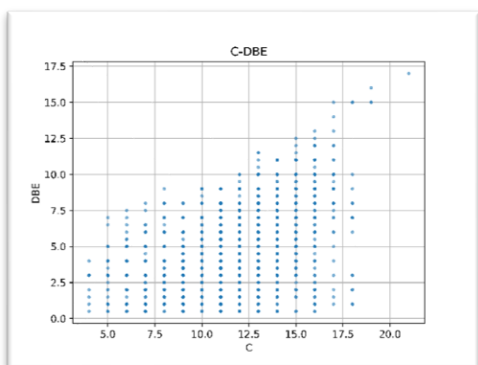
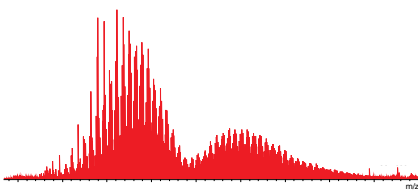
#### IV. USAGE AND OUTPUT FILE DESCRIPTION

The program can only submit one excel file at a time, but different files can be submitted multiple times before the results are sent. After clicking submit to submit, the result will be sent to your entered email address within 24h. Please contact [hbshen@sjtu.edu.cn](mailto:hbshen@sjtu.edu.cn) to obtain the email package password. The results of each file are saved as a compressed package containing 1 excel sheet and 8 corresponding visualizations.

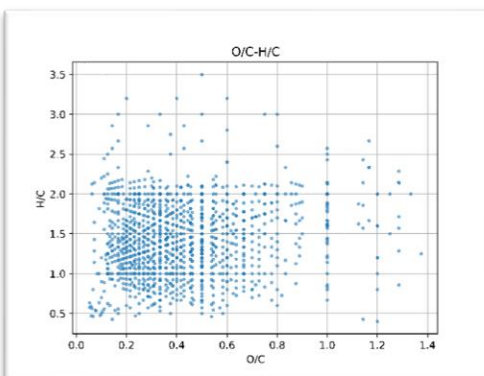
- There are 6 sheets in the Excel table, among which sheet result is the non-isotopic molecular formula, which is also the basis for the following drawing. Sheet resultC, resultS, resultO and resultH are the molecular formula of isotopic peak, Molecularclass is the classification of organic matter.
- The specific meanings and graphic examples of the 8 pictures are as follows:

**Table1.** The meanings of the 8 images outputted by MassAnalyzer

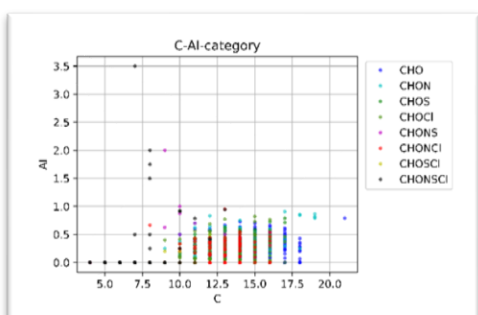
	<b>Result graph file name</b>	<b>Name of result graph</b>	<b>Implication</b>
1	plots_of_DBE_versus_carbon_numbers	C-DBE	Schematic diagram of the relationship between the number of carbon atoms and DBE
2	plots_of_H_C_ratios_versus_O_C_ratios	O/C-H/C	Schematic diagram of the relationship between carbon and hydrogen ratio and carbon and oxygen ratio
3	C-AI-category	C-AI-category	Schematic diagram of the relationship between the number of carbon atoms classified according to category and AI
4	C-Xc-category	C-Xc-category	Schematic diagram of the relationship between the number of carbon atoms classified according to category and Xc
5	mz-HC-category	mz-HC-category	Schematic diagram of the relationship between the ratio of mass and nucleus and the ratio of carbon and hydrogen according to category classification
6	mz-KMD-category	mz-KMD-category	Schematic diagram of the relationship between the ratio of cytoplasm to nucleus and KMD of category classification
7	Van_Krevelen_diagrams_relative_intensity	O/C-H/C-relative intensity	Schematic diagram of C/H ratio and C/O ratio classification based on relative peak strength
8	Van_Krevelen_diagrams_s_n	O/C-H/C-s/n	Schematic diagram of C-H ratio and C-O ratio classification based on SNR



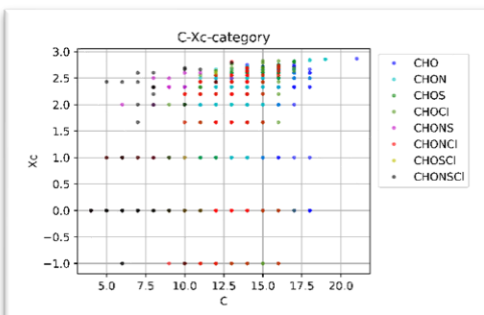
(1) C-DBE



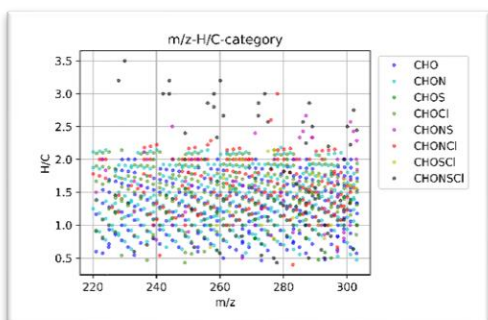
(2) O/C-H/C



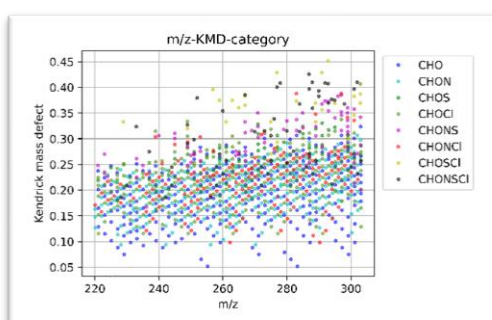
(3) C-AI-category



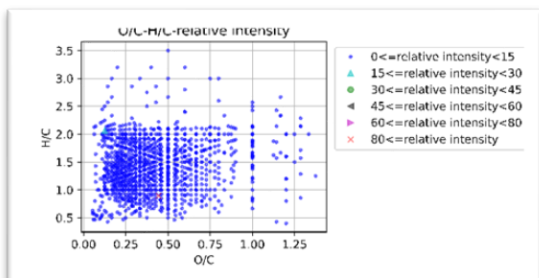
(4) C-Xc-category



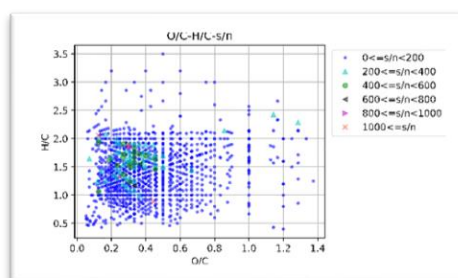
(5) m/z-HC-category



(6) m/z-KMD-category

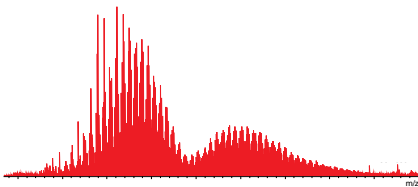


(7) O/C-H/C-relative intensity



(8) O/C-H/C-s/n

**Figure 6.** 8 images of calculated from the input\_sample.



## V. MAINTENANCE AND DEVELOPMENT

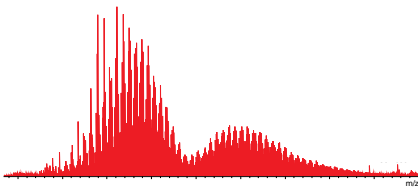
If there is any exception or other problems during operation, please contact us in time and send the input file and parameters to the mailbox. For any suggestions or specific needs, such as adding more elements or drawing more visualizations, please contact us.

## VI. CONTACT

Email: [hbshen@sjtu.edu.cn](mailto:hbshen@sjtu.edu.cn)

Telephone: 021-34205320

Address: Room#2-529, Bld. SEIEE, 800 Dongchuan Road, Shanghai,  
200240, China.



## Appendix: Explanation of Excel Parameters Output by MassAnalyzer Software

### 1. Atomic Masses Used in the Software

C: 12.0000000

H: 1.00782503207

O: 15.9949146

N: 14.003074

P: 30.973761998

S: 31.9720707

Cl: 34.968852682

Br: 78.9183376

F: 18.998403163

e: 0.00054858

C<sup>13</sup>: 13.0033548

H<sup>2</sup>: 2.0141018

O<sup>18</sup>: 17.9991603

S<sup>34</sup>: 33.967867

### 2. Formulas Corresponding to Output Parameters

$$AI_{mod} = \frac{1 + C - 0.5 \cdot O - 0.5 \cdot H}{C - 0.5 \cdot O}$$

$$NOSC = 4 - \frac{(4 \cdot C + H - 2 \cdot O)}{C}$$

$$DBE = \frac{(1 + 2 \cdot C - H + N)}{2}$$

$$AI_{denominator} = C - O - S - N$$

$$AI = \frac{(1 + C - O - S - 0.5 \cdot H)}{AI_{denominator}}$$

$$Xc_{denominator} = DBE - (O + S)$$

$$Xc = \frac{(3 \cdot DBE - 2 \cdot (O + S))}{Xc_{denominator}}$$