MARS User Manual

Version 2

Prof Hongmei Lu Group

School of Chemistry and Chemical Engineering

Central South University

Mapan_csu@yahoo.com

https://github.com/mapancsu

September 7,2017

1. Introduction

MARS2 (MS-Assisted Resolution of Signal) was developed for pre-processing untargeted GC-MS data files. It consisted of two main modules, resolve and extract modules. Fig.1 depicts the core ideal of MARS2. In resolve module, MARS2 integrated classical multivariate curve resolution (MCR) methods, including ITTFA, HELP and MCR-ALS. All MCR methods can obtain robust mass spectrum, which was used for compound identification. In extract module, RM (reverse match) and ITTFA were used to obtain chromatographic profiles. MARS2 has improved both the robustness and sensitivity of extracted features. All of the computing modules have been implemented in python language and compile into user interface.

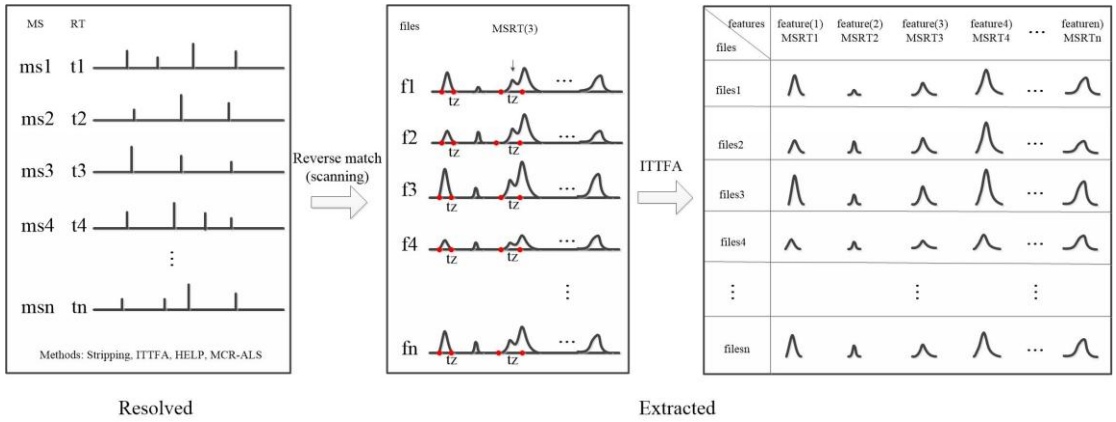

Fig. 1 The core ideal of MARS2.

2. Download and Installation

The software currently runs on Windows operating systems only. The software can be downloaded from Github at https://github.com/mapancsu/mars2. Click the green Clone or Download button on the right as shown in Figure 2.

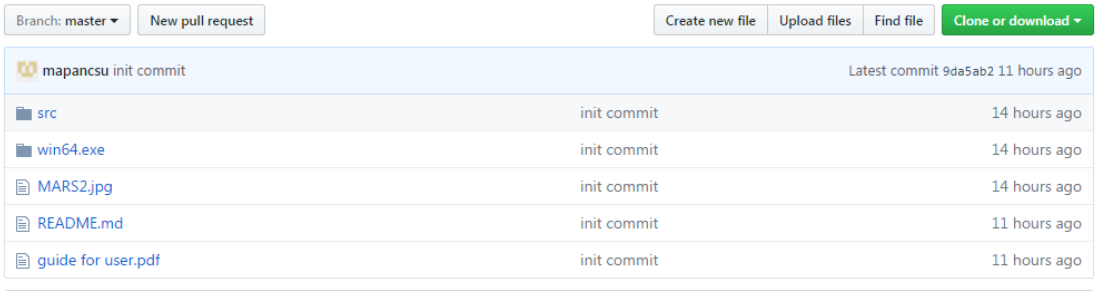

Fig. 2 Download MARS2 from Github.

Unzip the downloaded files. Enter the folder MARS2/win64.exe and then click on win64.exe to install and start MARS2 program. The whole program was complemented with python and the steps to test MARS2 under python environment can be followed at https://github.com/mapancsu/mars2/READEME.md.

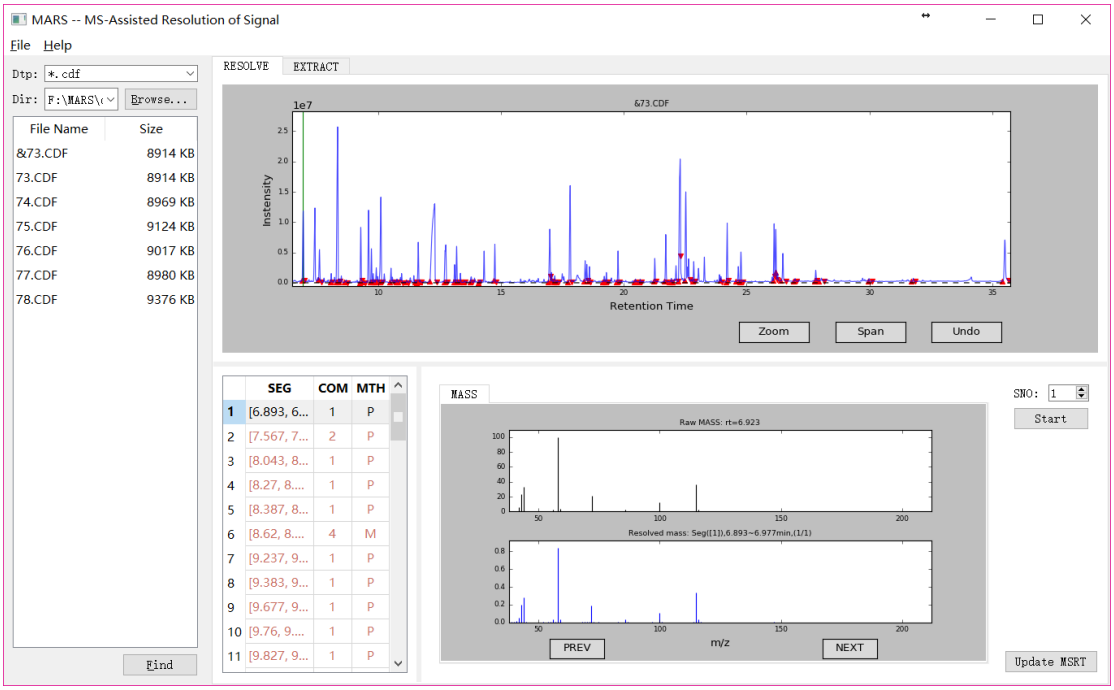

Fig. 3 The overall perspective of user interface.

3. Methods

We will illustrate how to use the MARS2 workflow using six data files. Totally six features were selected for illustration, including one-component system with baseline, two-component

overlapping peaks, and three-component overlapping peaks. The illustration will be introduced as four parts, including Input files, Resolve module, Extract module and Export. The overall perspective of user interface was shown in Fig. 3.

3.1 Input Files

MARS2 take cdf format files as input. Clicking Browse button to switch the work direction or

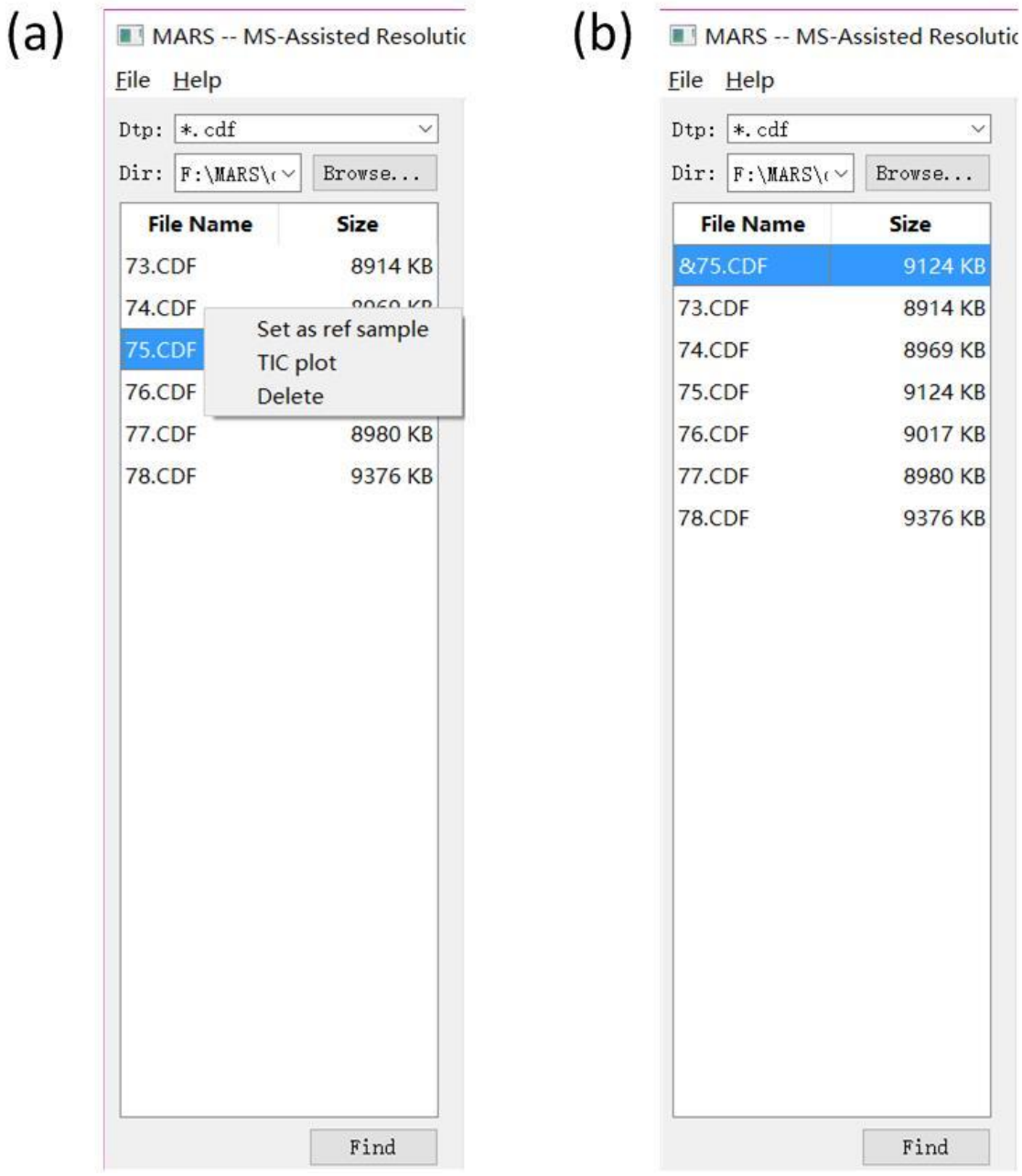

Fig. 4 The overall perspective of user interface.

adding cdf files to the displaced direction. In addition, when importing saved projections, the direction must switch to the absolute path of cdf files. First of all, right click cdf file and set a reference files (&75.CDF). Reference file can only be plotted on Resolve widget and others can only be plotted on Extract widget.

Browse… button ⟶ change working direction

Find button ⟶ displace existed cdf files in working direction

3.2 Resolve Module

Mass spectrum and retention time (MSRT) pairs are vital for batch processing. Here, any peak cluster can be spanned and selected for manual resolution. From Fig. 5, three systems are spanned. To obtain MSRT pairs, four solutions can be available as follow:

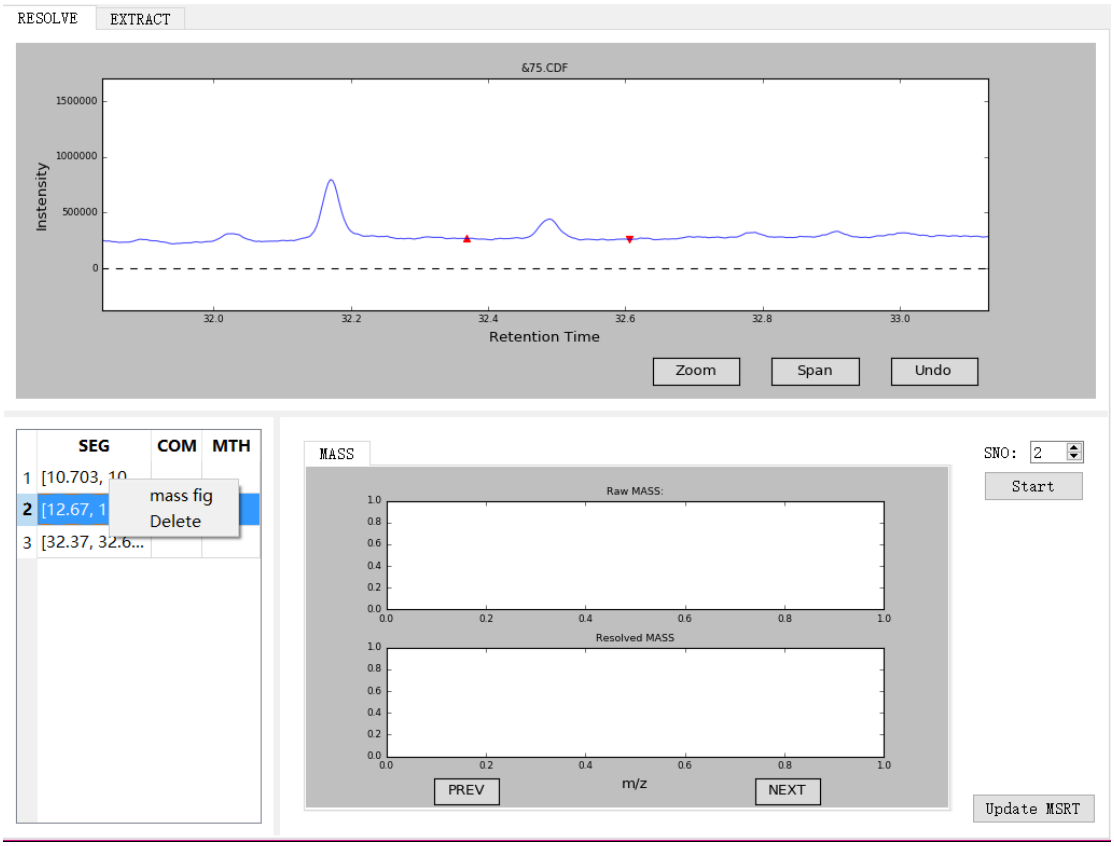

Fig. 5 Span segment of peak clusters.

(1) (LSF) ⟶ SVD ⟶ Pick

(2) (LSF) ⟶ SVD ⟶ ITTFA

(3) (LSF) ⟶ SVD ⟶ HELP

(4) (LSF) ⟶ SVD ⟶ MCR-ALS

LSF: least square fitting (2D) is optional step in aforementioned four flows.

span ⟶ select baseline region

start ⟶ run LSF

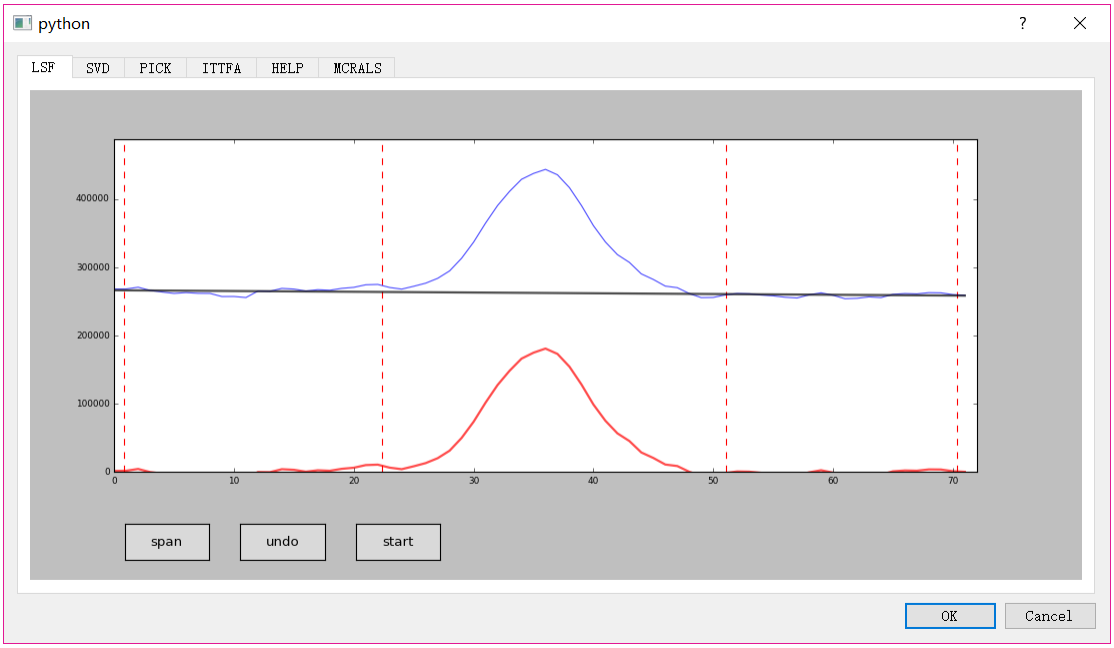

Fig. 6 An example of baseline removing.

SVD: singular value decomposition is used to estimate the number of component.

start ⟶ run SVD

click eigensvalue⟶ select right component number (the eigensvalue of true components are large than others significantly and their vector are smother than other vectors)

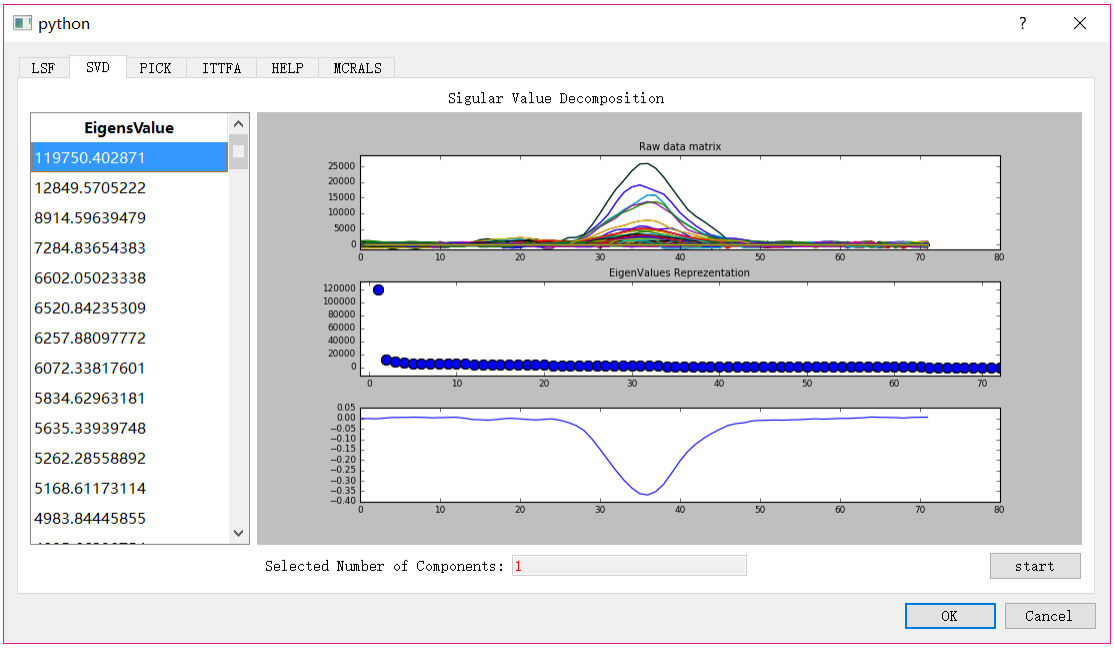

Fig. 7 An example of SVD

Pick (P): when all components own their selection region (pure mass spectrum region), their MSRTs are picked directly.

add ⟶ input data

Right click mouse ⟶ pick position with pure mass spectrum

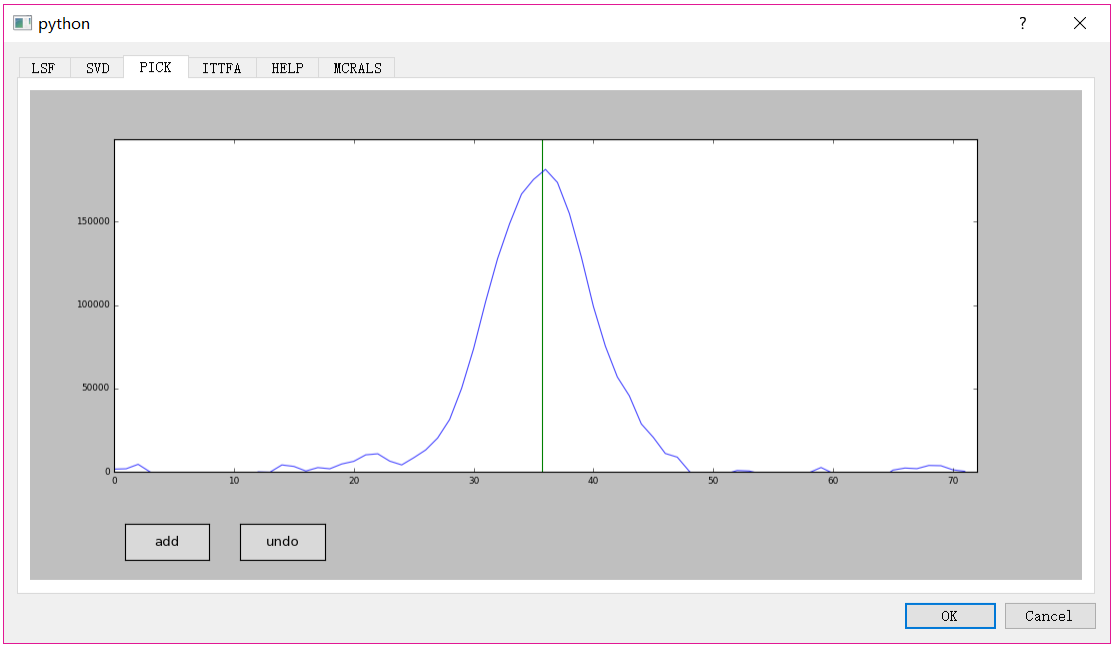

Fig. 8 Pick the pure mass spectrum and retention time (one-component system).

ITTFA (I) : ITTFA is a MCR method, which need to specify the apex of all components.

Add raw data ⟶ input data

Start ITTFA ⟶ run ITTFA to obtain pure chromatographic profiles. Fnnls is used to obtain pure mass spectrums.

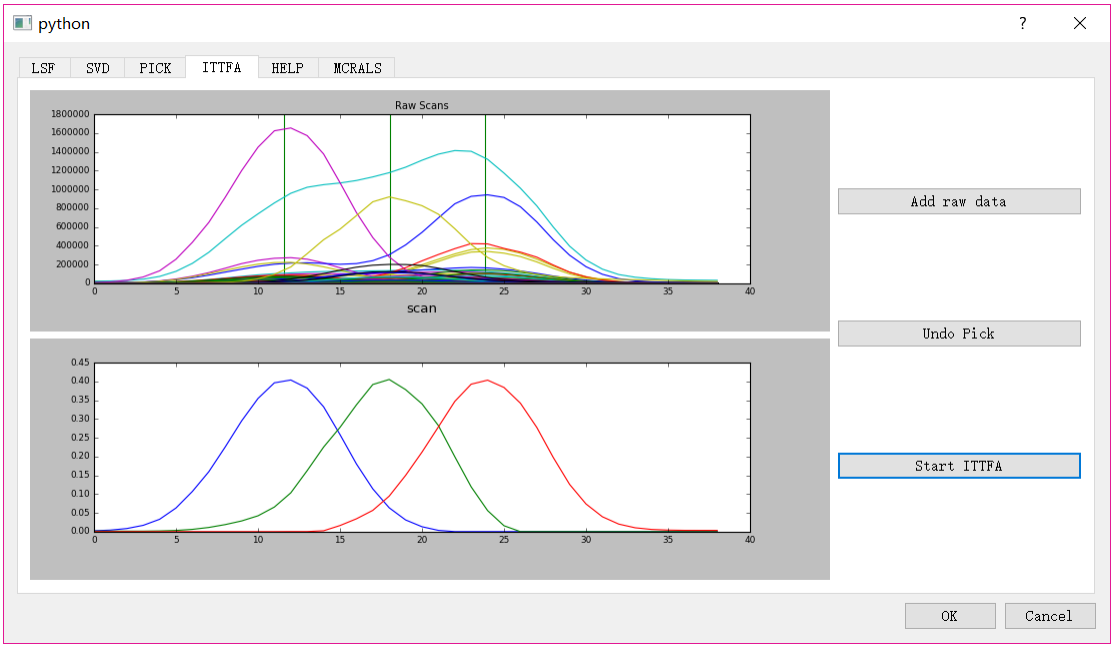

Fig. 9 An example of ITTFA (three-component system)

HELP (H): HELP is MCR method, which rely on selection regions and zero-concentration regions to perform Full Rank Analysis (FRA) for pure chromatographic profiles. Moreover, the stripping strategy is applied from outer to core.

Add raw data ⟶ input data (Fig. 10)

Do ⟶ ETA ⟶ find selection region and overlapping region (Fig. 11)

Do ⟶ LPG ⟶ find selection region (Fig. 12)

Whole region = selection region + overlapping region + zero-concentration region

… button ⟶ select region by mouse or input by users

FR button ⟶ run full rank analysis

LAYER ⟶ jump from i to i+1

W: the window of ETA

SE: selection region

OV: overlapping region

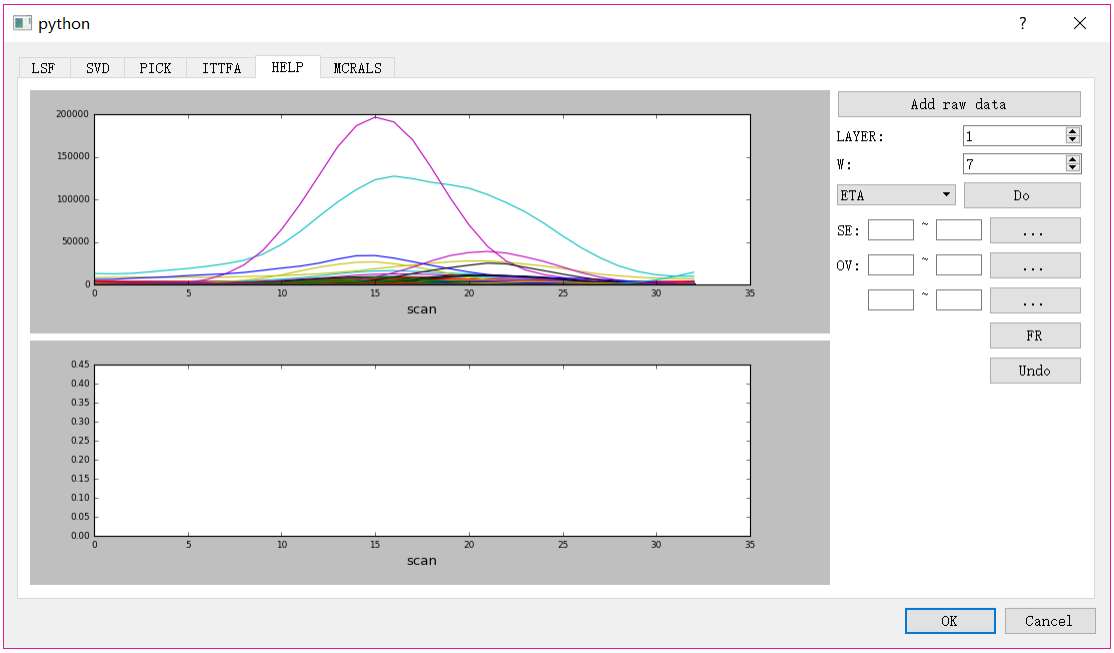

Fig. 10 An example of HELP

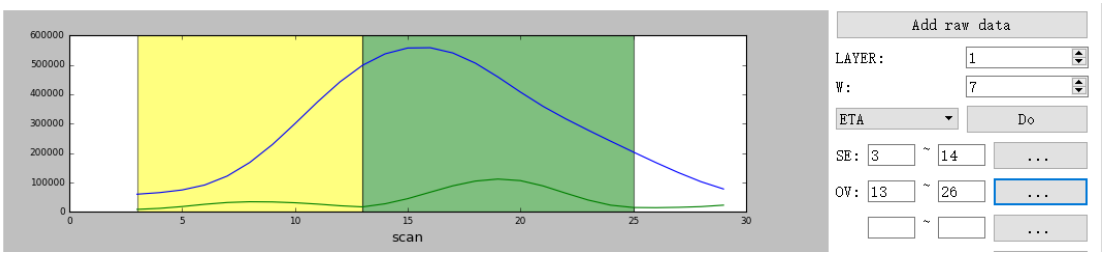

Fig. 11 Eigenstructure tracking analysis (ETA).

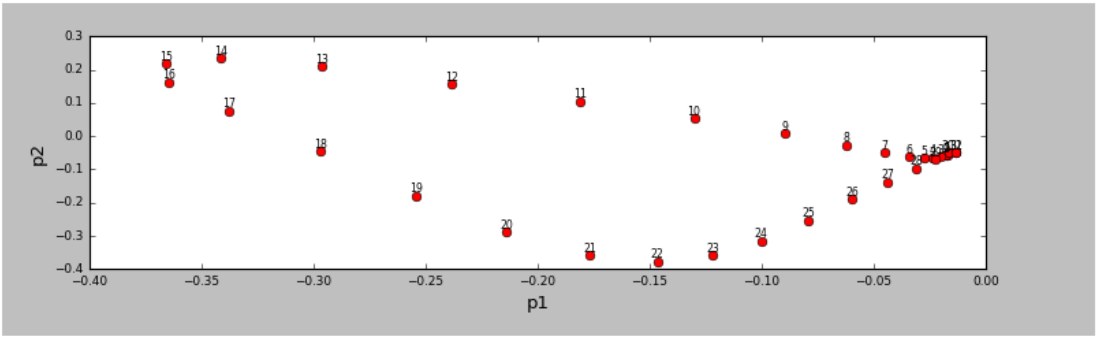

Fig. 12 Latent projection graph (LPG)

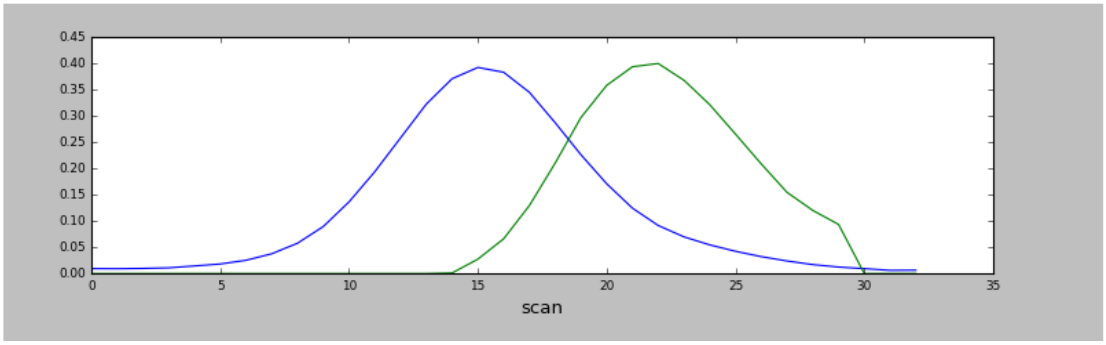

Fig. 13 Resolved pure chromatographic profiles by HELP.

MCR-ALS (M): MCR-ALS is an iterative MCR method. It can resolve multi-component overlapping system (more than three components). It uses PURE methods to obtain initial

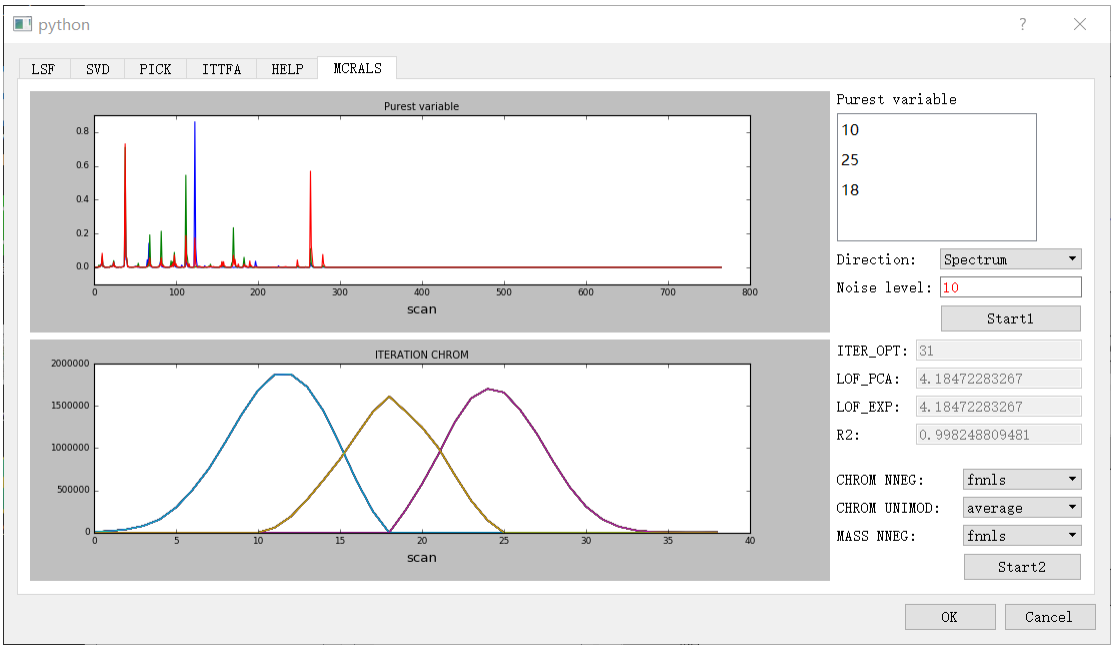

Fig. 14 An example of MCR-ALS.

estimation of mass spectrum or chromatographic profiles. During iteration, non-negetive, unimodal are used to constraint chromatographic profiles, and non-negetive is used to constraint mass spectrums. (Fig. 14)

Start1 ⟶ run PURE methods for initial estimation

Start2 ⟶ run MCR-ALS

When the selected segments are resolved, the number in QtableWidget turn from black to red. The resolved MSRTs can be displaced and compared with raw mass spectrums. (Fig. 15)

Update MSRT ⟶ add resolved MSRTs to EXTRACT module.

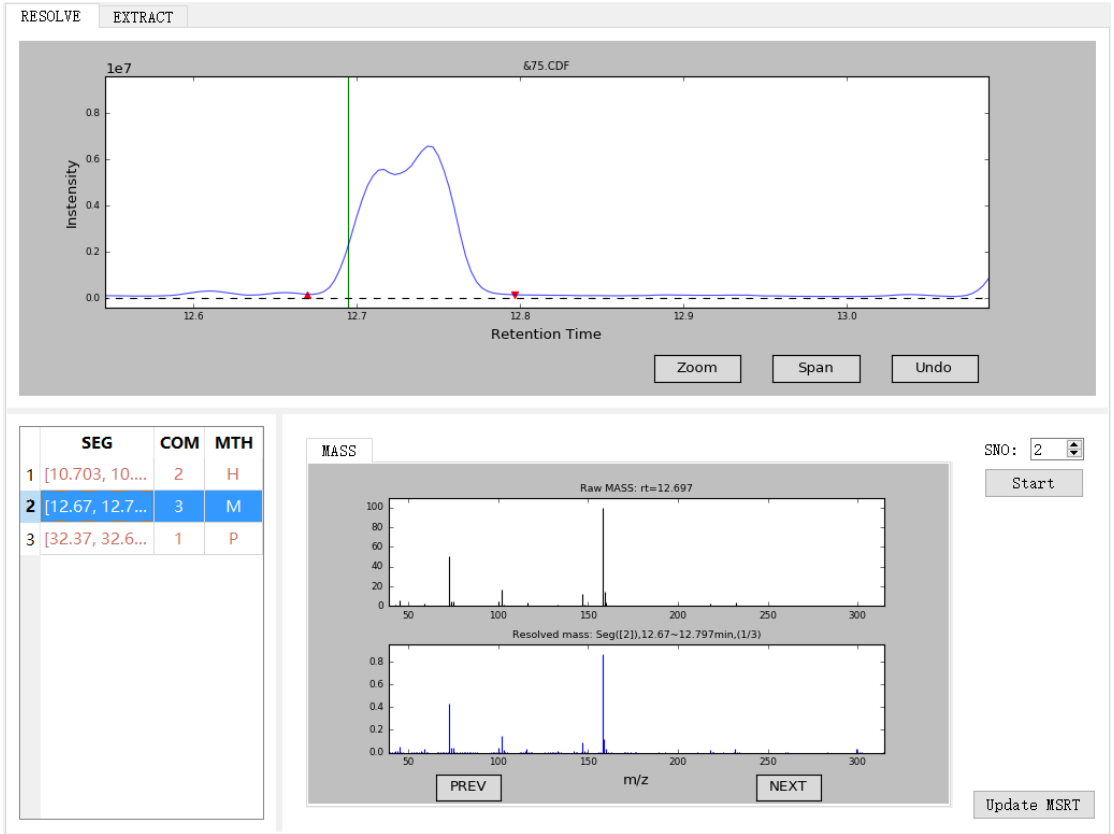

Fig. 15 Resolved results of three systems.

3.3 Extract Module

This module uses resolved MSRTs to extract features from batch cdf files. RM or MSWFA is used to search the target zones. ITTFA is used to obtain pure chromatographic profiles.

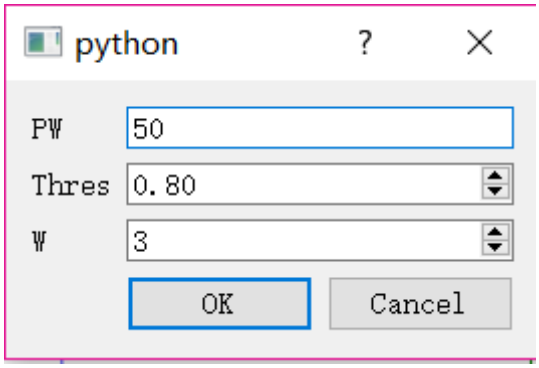

Fig. 16 Set parameters of MARS2.

Set MARS ⟶ Set parameters of MARS2 (Fig. 16)

PW: max allowed peak windows (search range of RM, about 101 scan points)

Thres: threshold of score from RM or MSWFA, to find target zone.

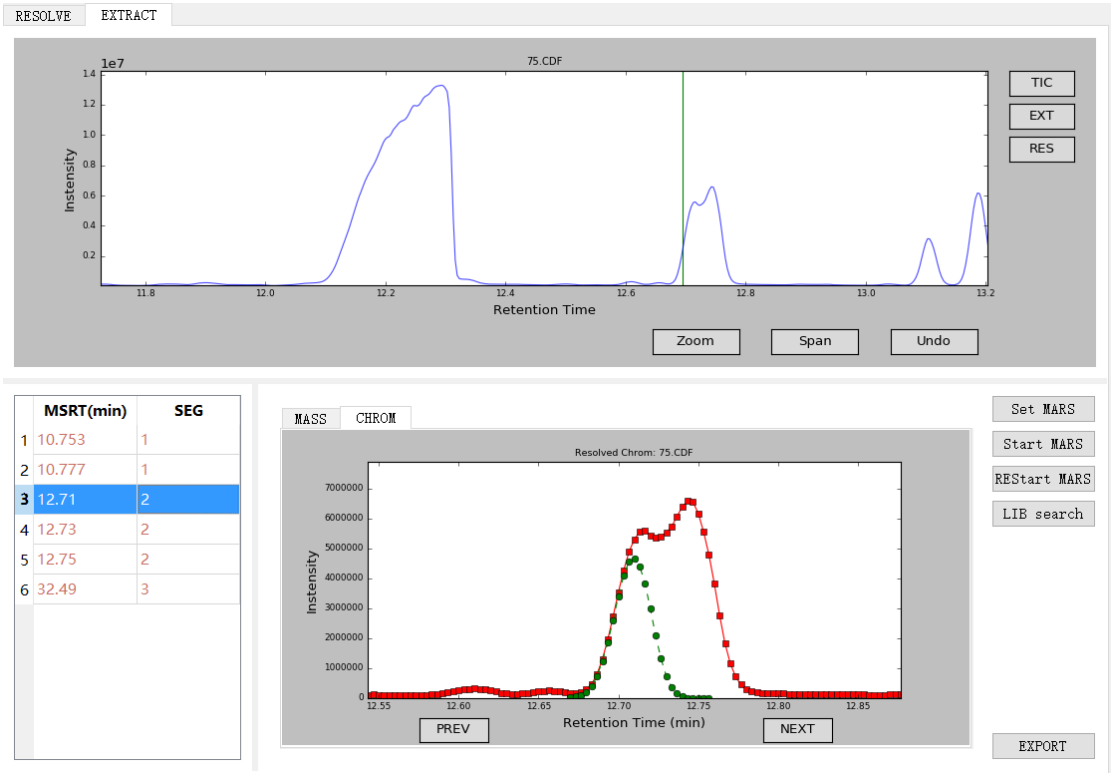

Fig. 17 Extract results of MARS2.

Start MARS ⟶ extract features of new added MSRTs in MSRT table (black). (Fig. 17)

REStart MARS ⟶ extract features of all MSRTs in MSRT table (black and red). (Fig. 17)

When all files are processed, the number in the MSRT table is changed from black to red.

3.4 Export

This part is to export quantitative table and qualitative table.

EXPORT ⟶ export quantitative table and qualitative table (Fig. 18 and 19)

	A	B	C	D	E	F	G	H	I
1	comID	RT.ref(min	&75.CDF	73.CDF	74.CDF	75.CDF	76.CDF	77.CDF	78.CDF
2	1	10.75333	4814012	4754277	5137662	4814012	4643099	4644071	4883736
3	2	10.77667	2177513	2586422	2360930	2177513	2182555	2292012	2542952
4	3	12.71	39650989	37924033	38728102	39650989	37834406	37152317	39333078
5	4	12.73	39886605	37348640	40612310	39886605	41599733	39604344	37525237
6	5	12.75	42734364	45948447	43612164	42734364	43353297	44807218	46609394
7	6	32.49	1938989	1089591	1005024	1938989	3326330	313040.3	744558.2

Fig. 18 Quantitative table of six components from three systems.

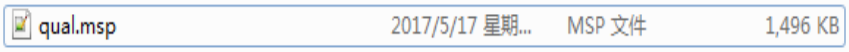

Fig. 19 Qualitative table of six components from three systems.

qual.msp file is saved in work direction and this file can be import into MS Search v.2.0 for identification analysis.