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

January 12,2018

1. Introduction

MARS 2 (MS-Assisted Resolution of Signal) was developed for pre-processing untargeted GC-MS data files. The greatest strength of MARS 2 was "**resolve once, extract anywhere**". Fig.1 depicts the core ideal of MARS 2. Mass spectra and chromatographic profile were most important features used for qualitative and quantitative analysis. MARS 2 has improved and integrated classical MCR methods, and improved both the robustness and sensitivity of extracted features. It consisted of two main modules, RESOLVE and EXTRACT modules. In RESOLVE module, ITTFA, HELP and MCR-ALS were used to resolve mass spectra from overlapped system. In EXTRACT module, RM (reverse match) was used to detect target zone of interest, and PCO-ITTFA was used to extract pure chromatographic profile form complex system (baseline, retention time shift and overlapped peaks). These two modules have been implemented in python language and compile into user interface.

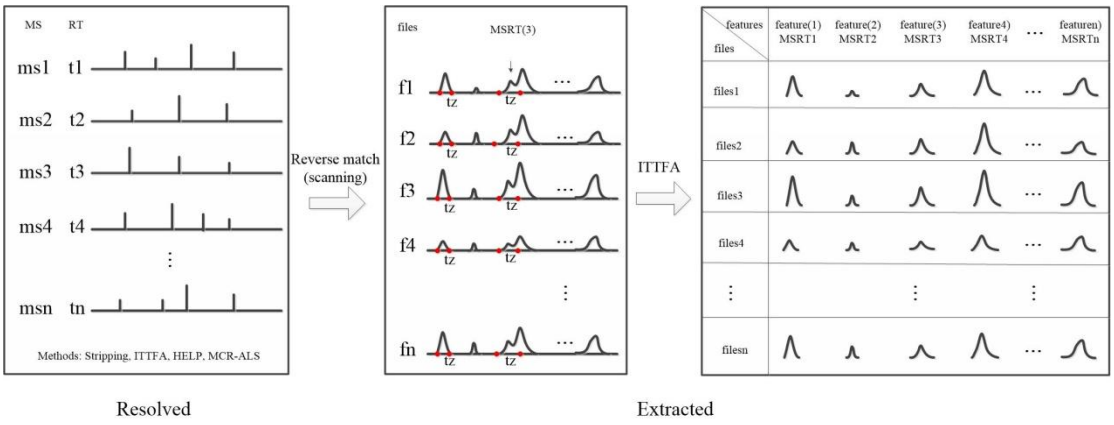

Fig. 1 The core ideal of MARS2.

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 button (Clone or Download) on the right as shown in Figure 2. The whole program was complemented with python. The steps to test MARS2 under python environment: *https://github.com/mapancsu/mars2/READEME.md*. Moreover, a release version can be downloaded at *https://github.com/mapancsu/mars2/releases*. Unzip the downloaded files. Enter the folder MARS2 and then click on win64.exe to install and start MARS2 program.

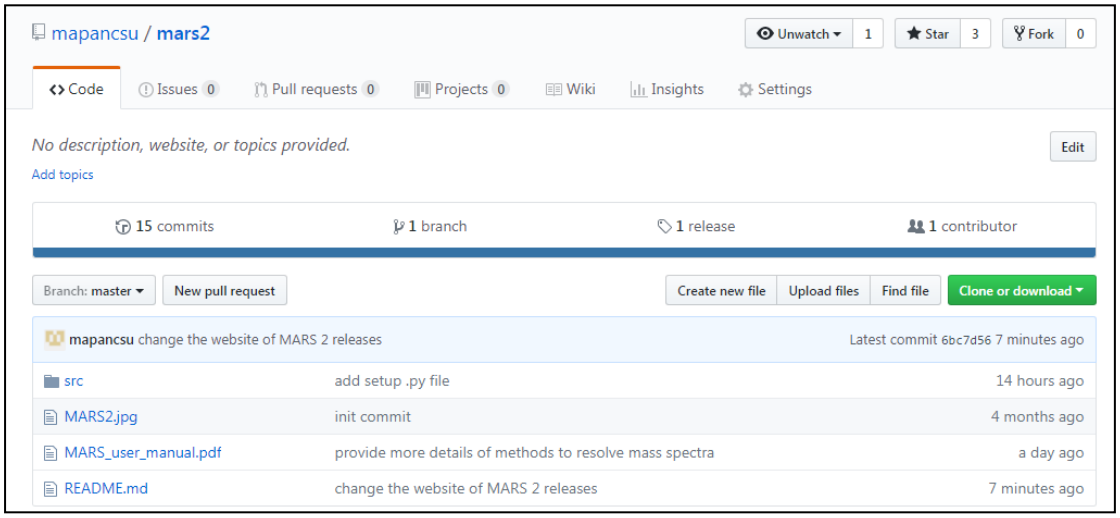

Fig. 2 Download MARS2 from Github.

3. Methods

We will illustrate how to use the MARS2 workflow using six data files. These files and a demo video for users (test MARS2.mp4) was uploaded at *https://drive.google.com/drive/folders/0B72Pu_uQr73IbkVtT2w0Mm1wZ0U?usp=sharing.* The illustration will be introduced as four parts, including File and Projection, Resolve module, Extract module and Export. The overall perspective of user interface was shown in Fig. 3.

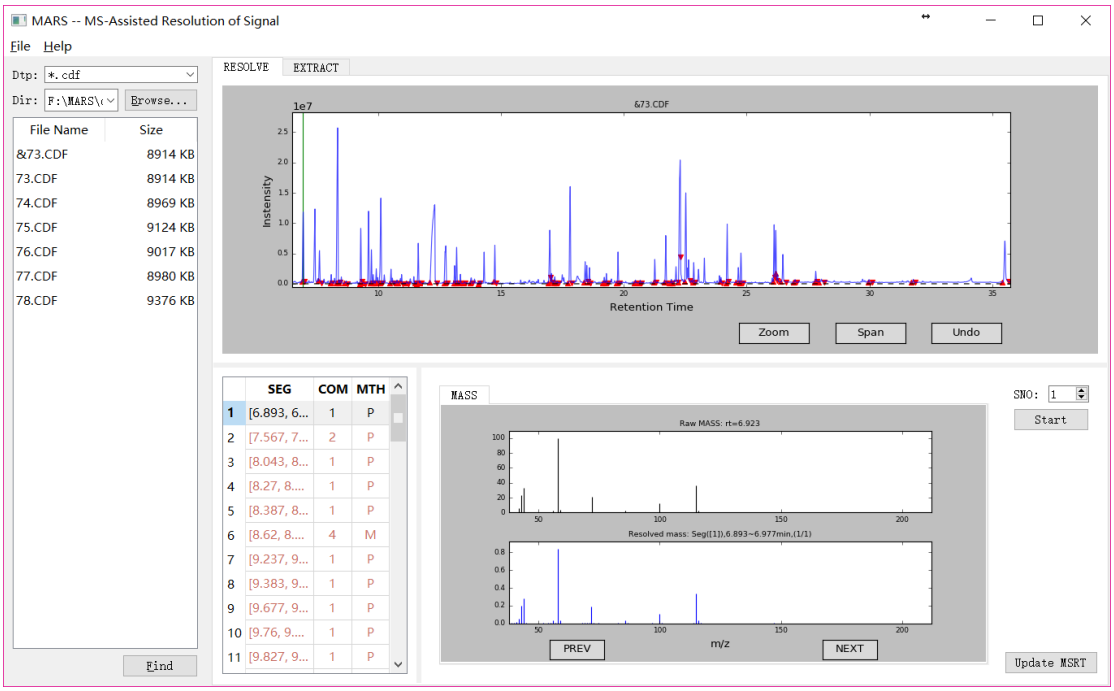

Fig. 3 The overall perspective of user interface.

3.1 Files and Projection

Import files:

MARS2 take cdf format files as input. If you start a new projection, you should import files first. Clicking Browse button (Fig. 4) to switch the work direction or adding cdf files to the displaced direction. The loaded files can only be plotted on EXTRACT widget.

Fig. 4 Import cdf files

Loading projections:

The users can save projection as a pkl format file. You can load unfinished projections anytime. Fig. 5 showed how to save and load projection. When you load a projection and continue to complete it, you must keep the displayed files in the working direction.

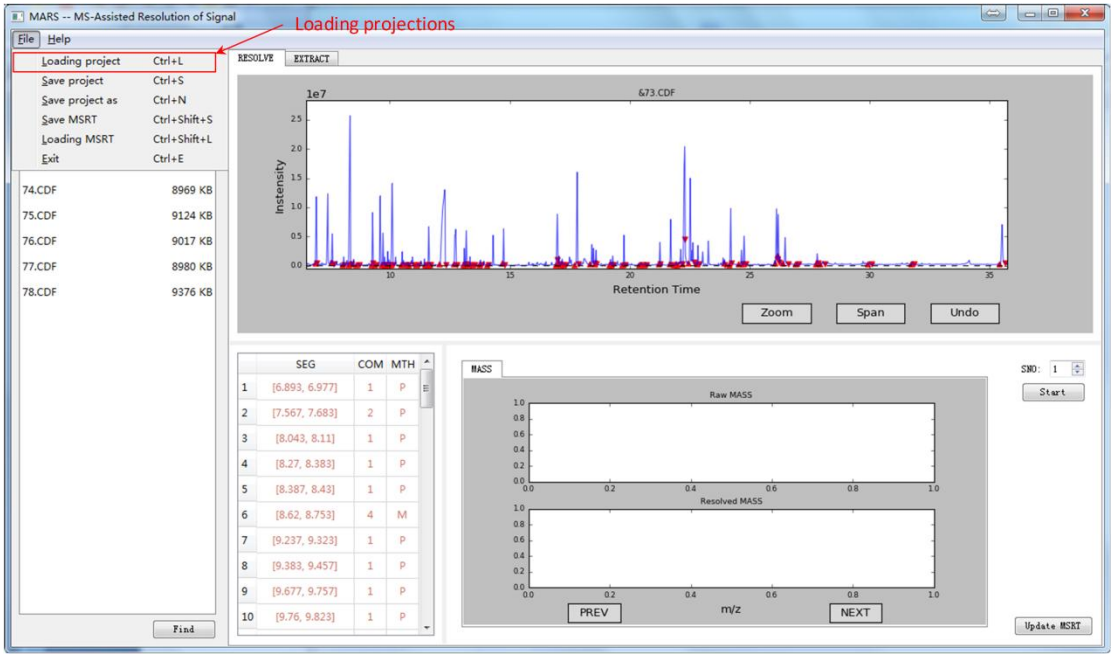

Fig. 5 Loading projections

Loading MSRTs:

The users can save resolved MSRTs (mass spectra and retention time pairs) for future use in same batching experiment. If you have saved MSRTs, you can import saved files into EXTRACT widget and start MARS directly. If you can't get MSRTs, you should set a ref sample and resolve MSRTs from the ref sample in RESOLVE module. Fig. 6 showed how to Load MSRTs and set ref sample. The ref file can be plotted on RESOLVE widget.

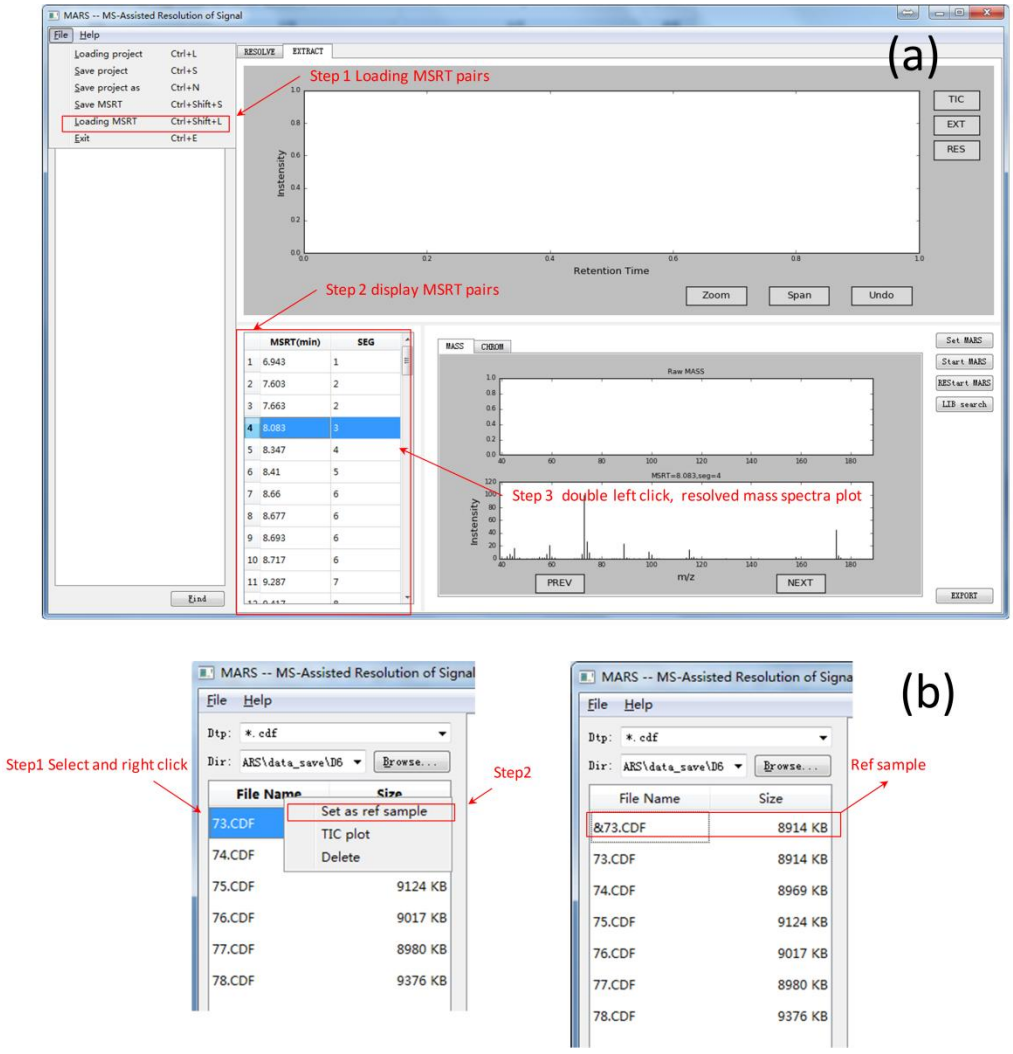

Fig. 6 Loading MSRT and set reference file. (a) loading MSRTs, (b) set reference file

3.2 Resolve Module

MARS 2 has advantage in qualitative analysis, because MARS 2 uses MCR methods to obtain robust mass spectra features. This part showed how to select regions and resolve mass spectra from selected regions.

Select regions:

Any peak cluster can be spanned and selected for manual resolution. The users can select regions by two principles. 1) Checking intensity, intensity of true component significantly higher than adjacent baseline and noise. 2) Checking mass spectra, mass spectra of true component have same ion fragments. An example was shown in Fig. 7.

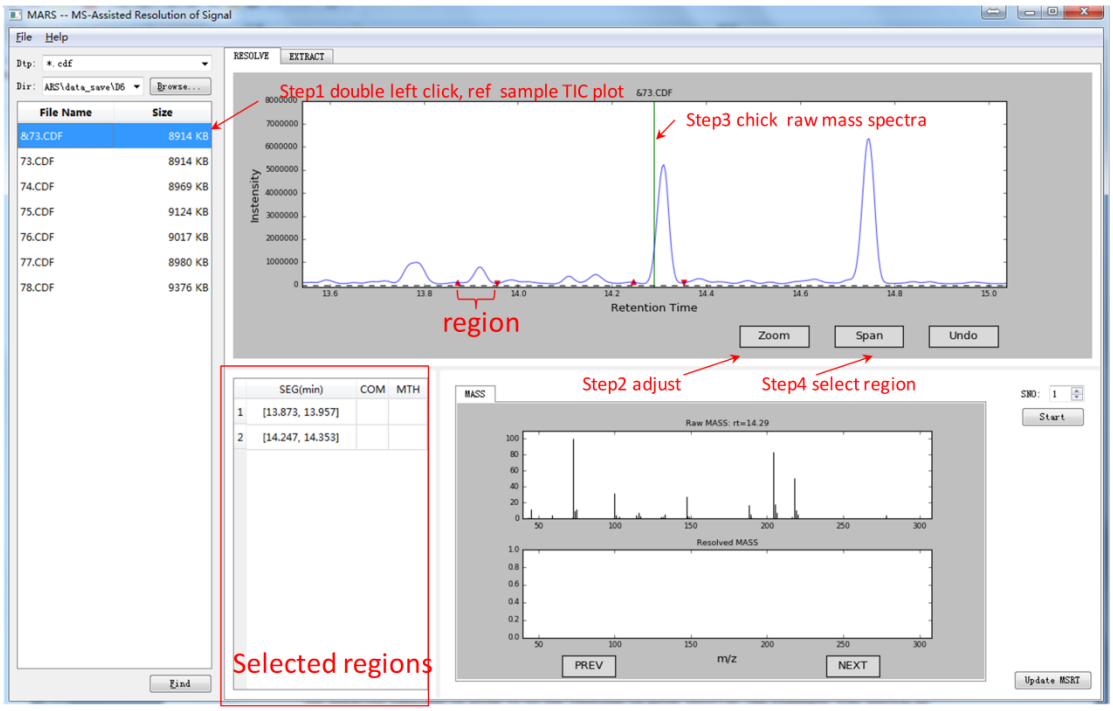

Fig. 7 Examples of select regions.

Resolve mass spectra:

Four methods are used to obtain mass spectra. When CN (component number) = 1, mass spectra can be picked directly at apex position. When CN >= 2 (overlapping peaks), ITTFA, HELP and MCR-ALS are used to resolve pure mass spectra. Fig. 8 shows how to start RESOLVE module.

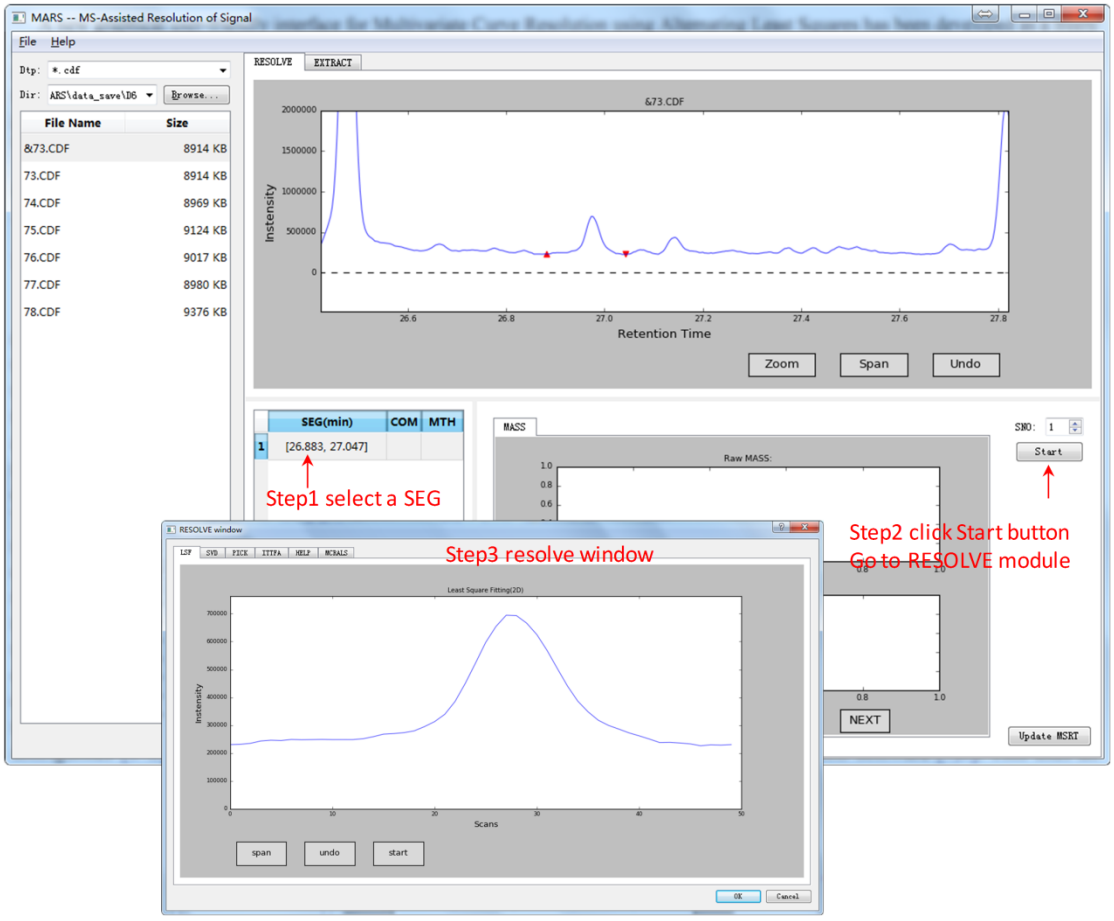

Fig. 8 Start RESOLVE module

LSF: Least square fitting (LSF, 2D) is used to remove baseline of selected regions [1]. LSF uses the adjacent baseline of peak to fit the baseline of peak interval. An example was shown in Fig. 9. When the selected regions have no baseline, this step can be ignored.

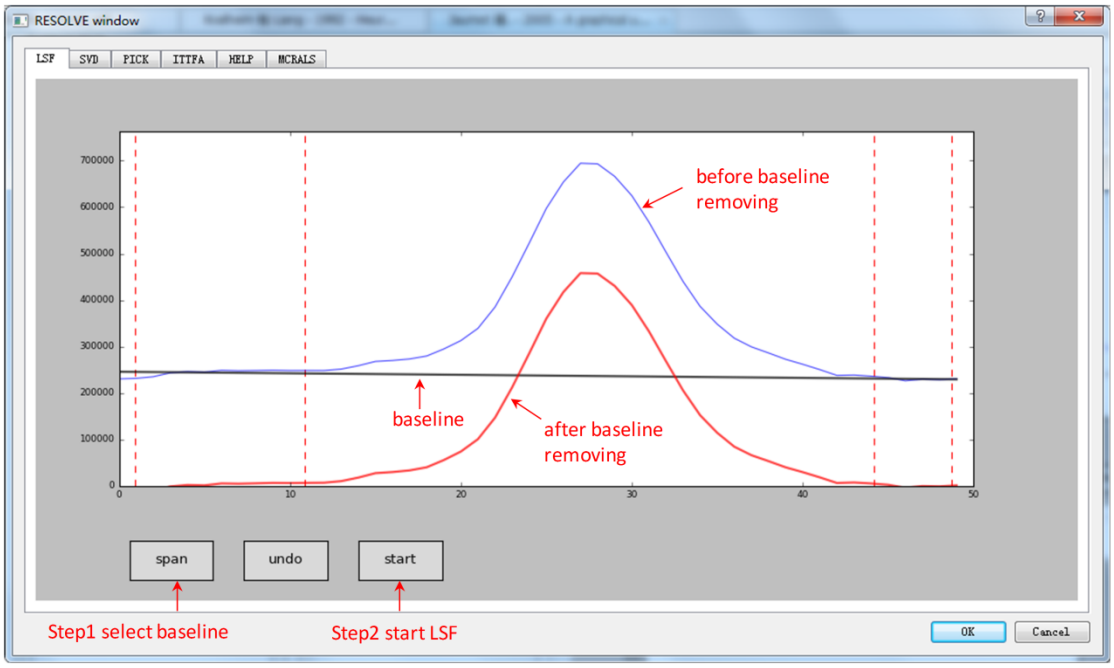

Fig. 9 An example of baseline removing.

SVD: singular value decomposition (SVD) is used to estimate the component number of a given region according to eigenvalues and eigenvectors. The eigenvalue of true component was significantly large than noise and the eigenvector of true component was smoother than noise vector. An example is shown in Fig. 10.

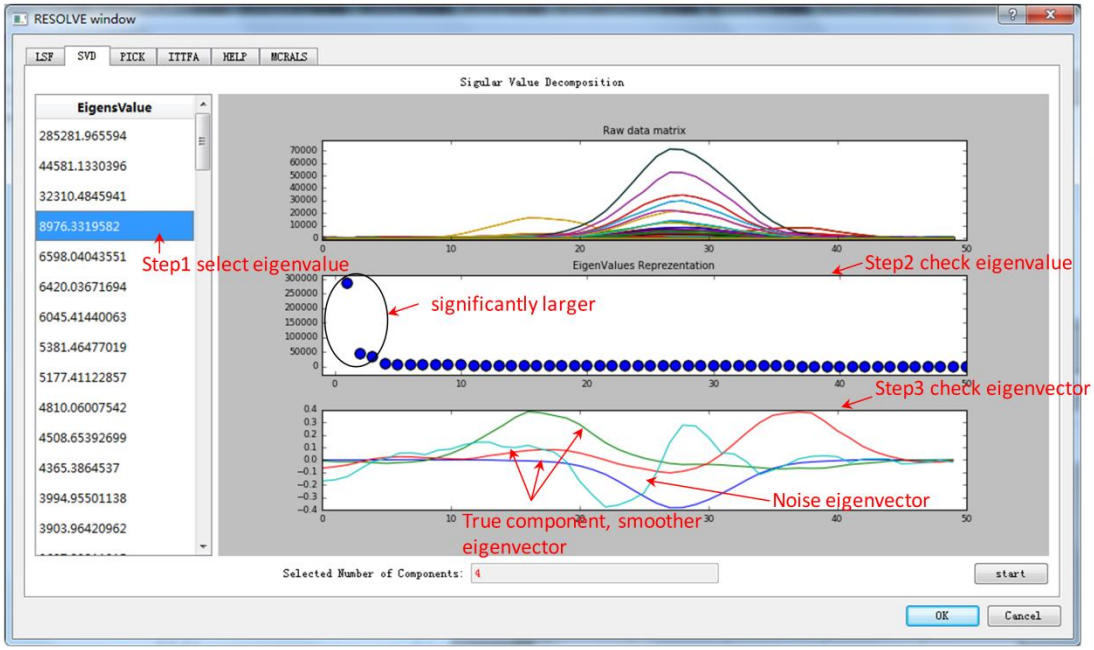

Fig. 10 An example of SVD

PICK: When CN of selected region is 1, it means that all mass spectras of peak region are same. Mass spectra at apex position is least affected by noise. Therefore, mass spectra can be picked directly at apex position. An example is shown in Fig. 11.

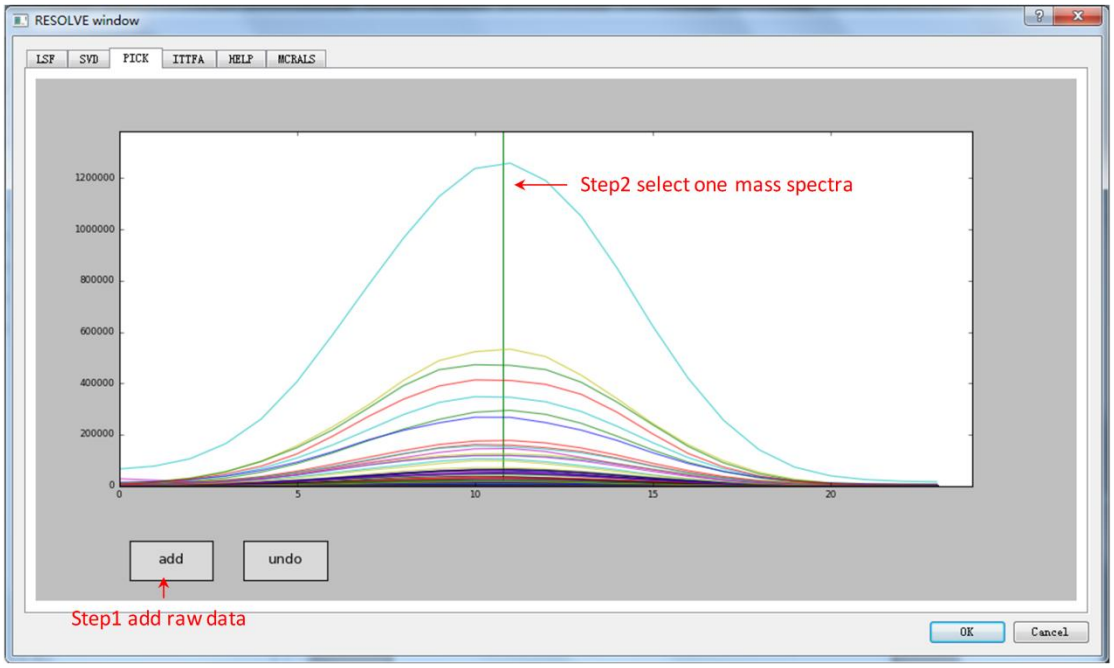

Fig. 11 Pick the pure mass spectrum and retention time (one-component system).

ITTFA: Iterative target transformation factor analysis (ITTFA) was developed by Vandeginste and Gerperline [2]. The major idea of this method is construct delicately an iterative procedure to approach the possible solution of chromatographic profiles or spectra of pure components. The users need to specify the needle position (apex position) as initial estimation of chromatographic profile. An example is shown in Fig. 12.

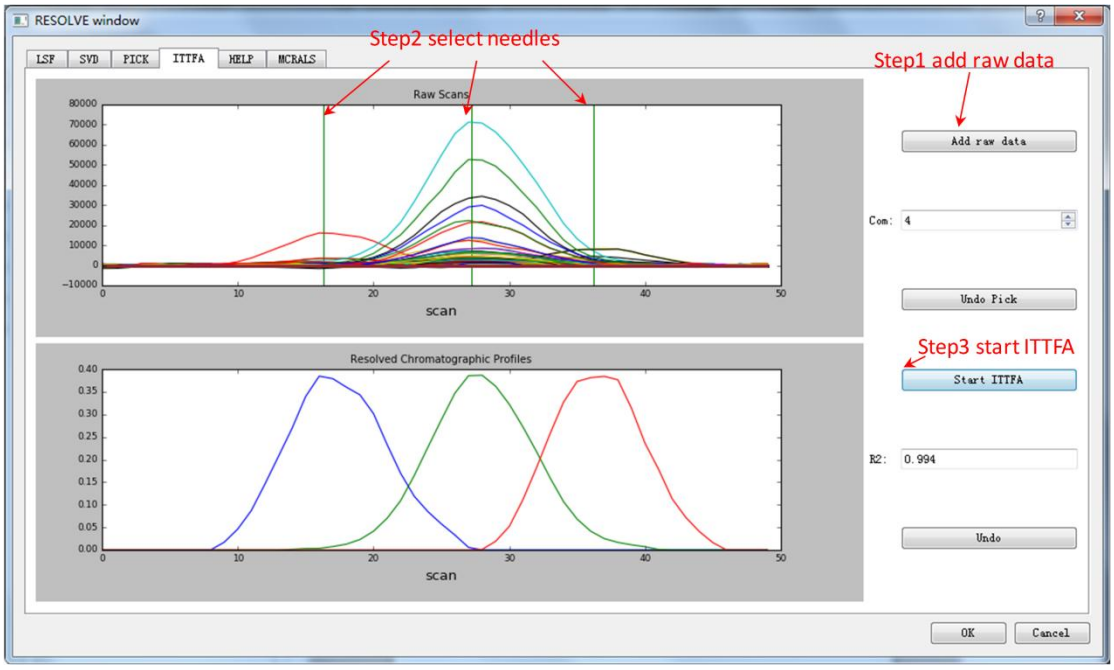

Fig. 12 An example of ITTFA (three-component system)

HELP: Heuristic evolving latent projections was developed by Olav M. Kvalheiml and Yizeng Liang [3–5]. HELP uses latent projection graph (LPG) and eigenstructture tracking analysis (ETA) to find selective information (region of pure component) and zero-concentration regions. Full-rank resolution is used to resolve pure chromatographic profiles. Finally, the component stripping procedure can be followed to continue to resolve the other components. This method is not suitable for embedding peaks. Because the selective information of embedding peaks can not be obtained. An example is shown in Fig. 13.

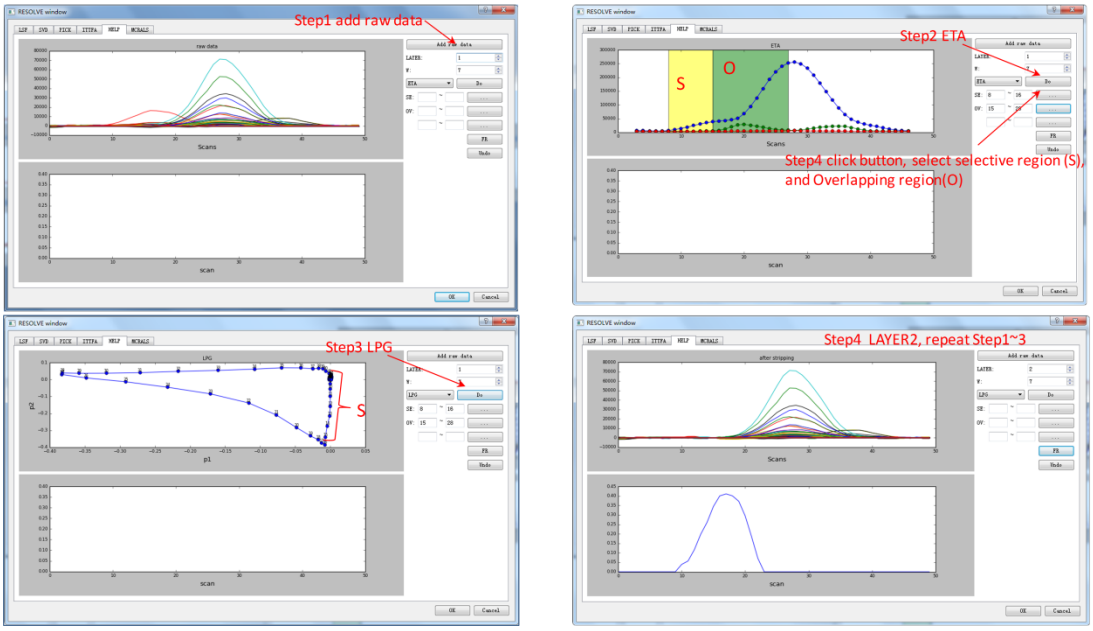

Fig. 13 An example of HELP

MCR-ALS: Multivariate curve resolution-alternating least squares (MCR-ALS) is an iterative MCR method [6–8]. It uses PURE methods to obtain initial 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. An example of MCR-ALS is shown in Fig 14.

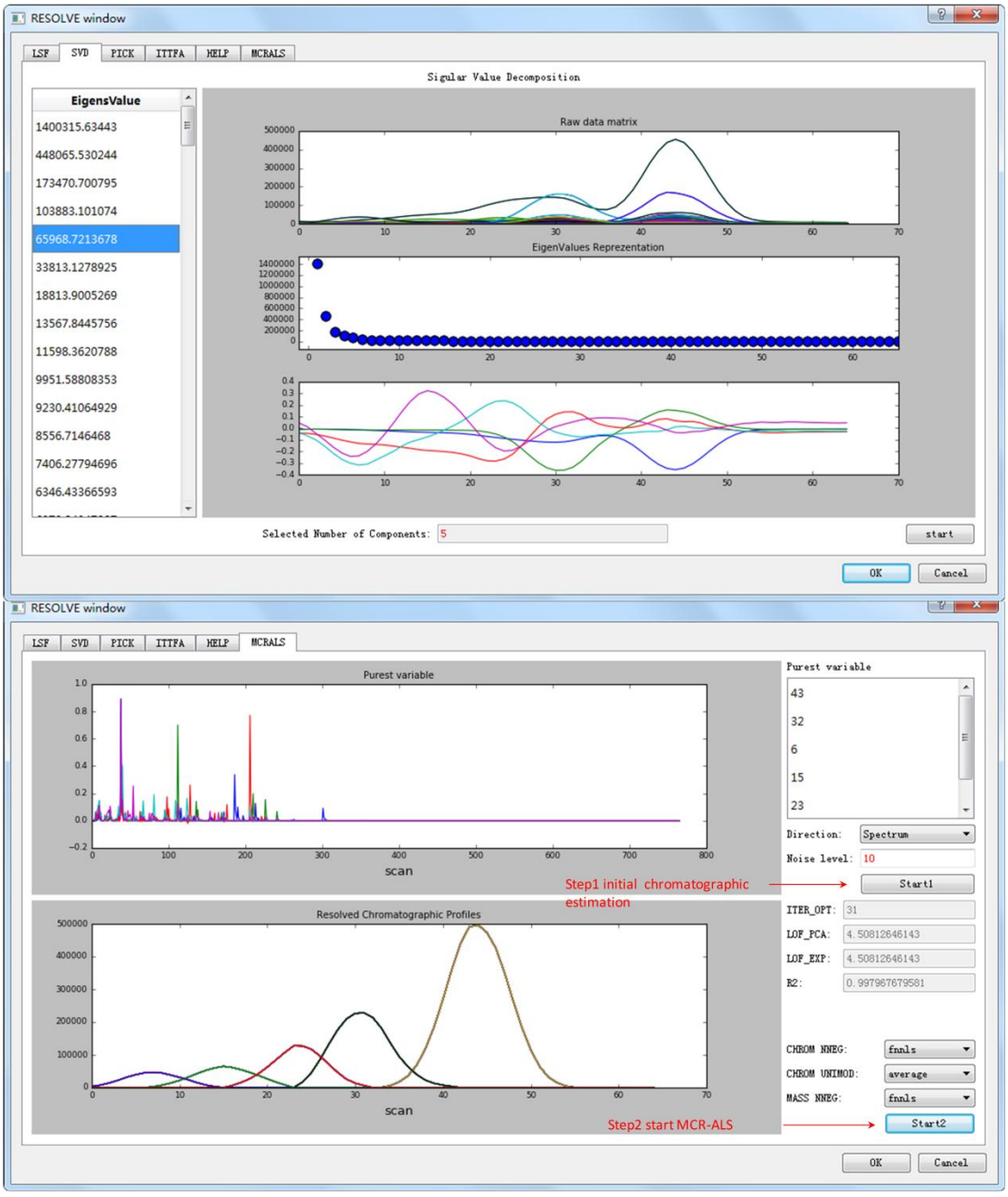

Fig. 14 An example of MCR-ALS.

When the selected regions are resolved, the number in SEG-COM-MTH table turn from black to red (Fig. 15a). The resolved MSRTs can be displaced and compared with raw mass spectrums (Fig. 15a). The resolved MSRTs can be imported into EXTRACT module (Fig. 15b).

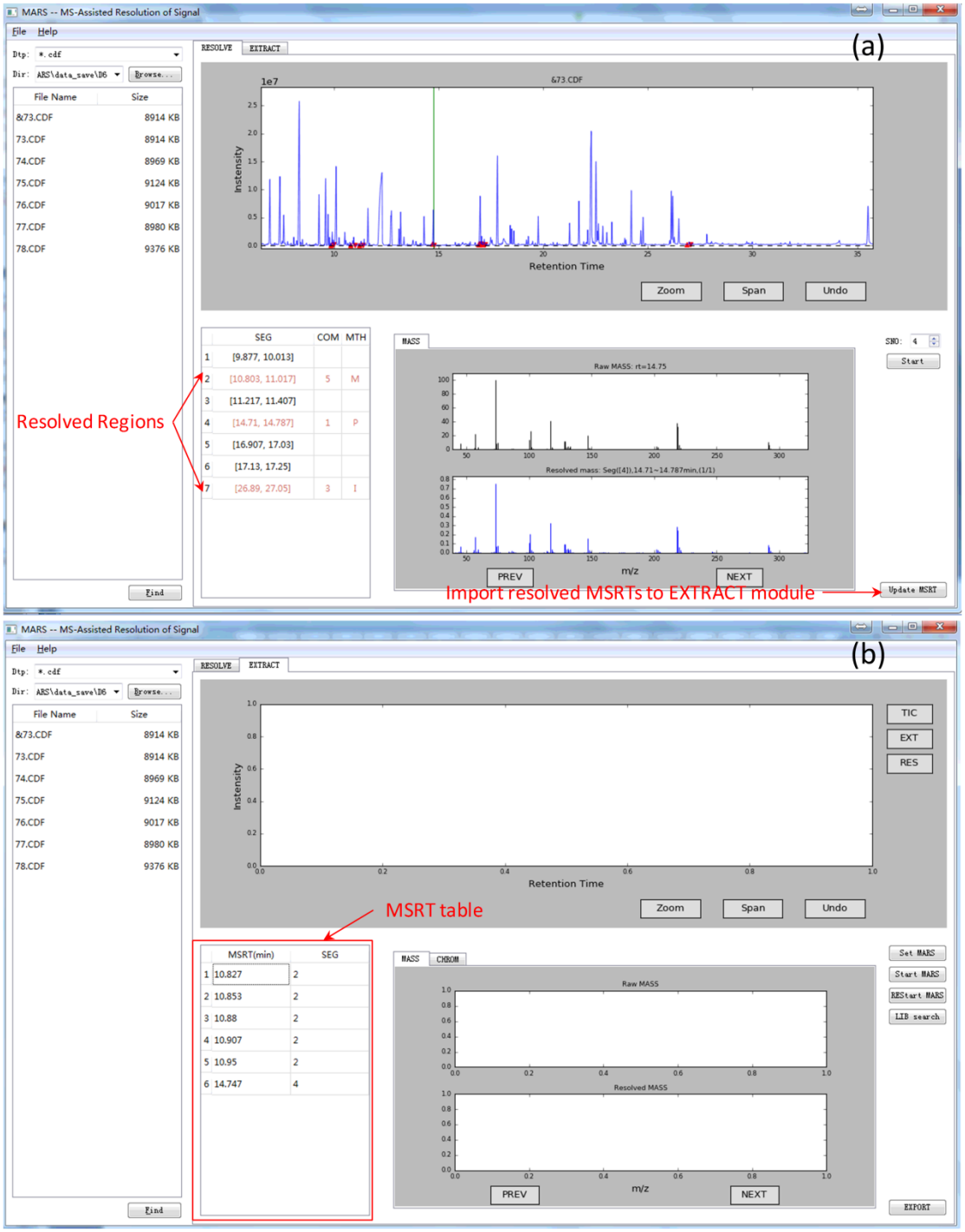

Fig. 15 Import MSRT to EXTRACT module.

3.3 Extract Module

This module uses resolved MSRTs to extract features from other cdf files. RM or MSWFA is used to detect target zones. PCO-ITTFA is used to obtain pure chromatographic profiles. The parameters are shown in Fig. 16.

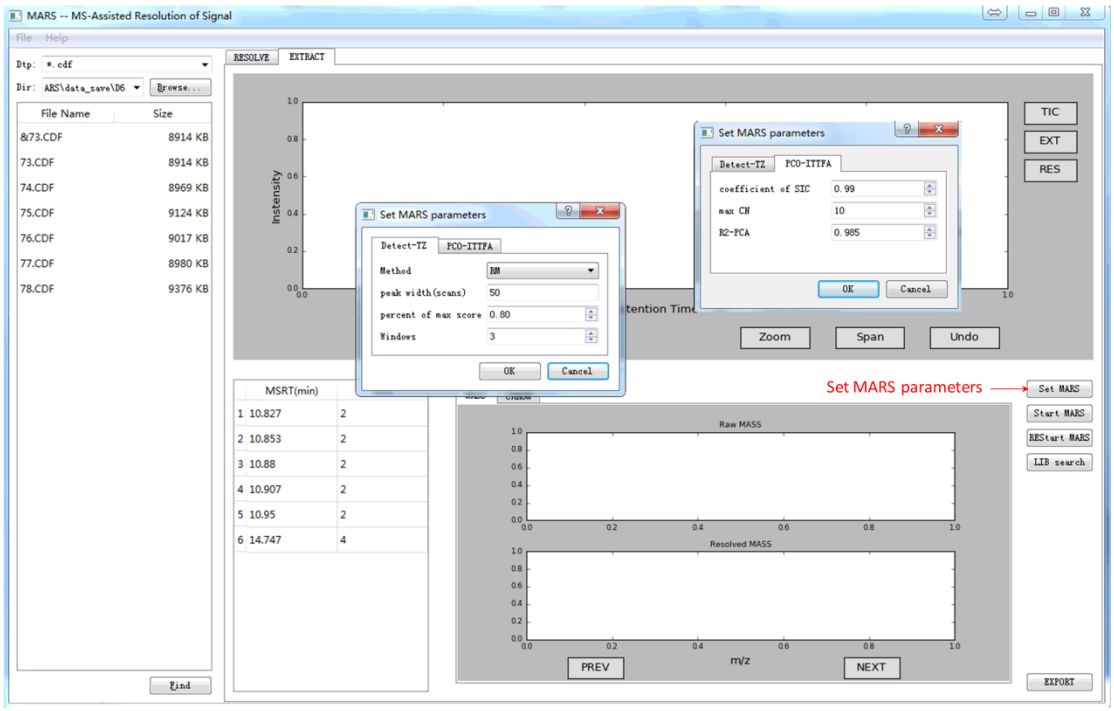

Fig. 16 Set parameters of MARS2.

Detecting target zone:

Method : Detecting target zone (MSWFA, RM)

PW: max allowed peak windows on both sides of RT

Percent of max score:

Window: window of MSWFA.

PCO-ITTFA:

Coefficient of SIC: minimum coefficient value (0.90~0.99) to determine selective ion chromatogram (SIC)

Max CN: maximum component number of optimization (6~10)

R2-PCA: explained variance of principal component analysis (0.985~0.995)

After MSRTs are imported into MSRT table, MARS can be started. Fig. 17 shows how to start MARS.

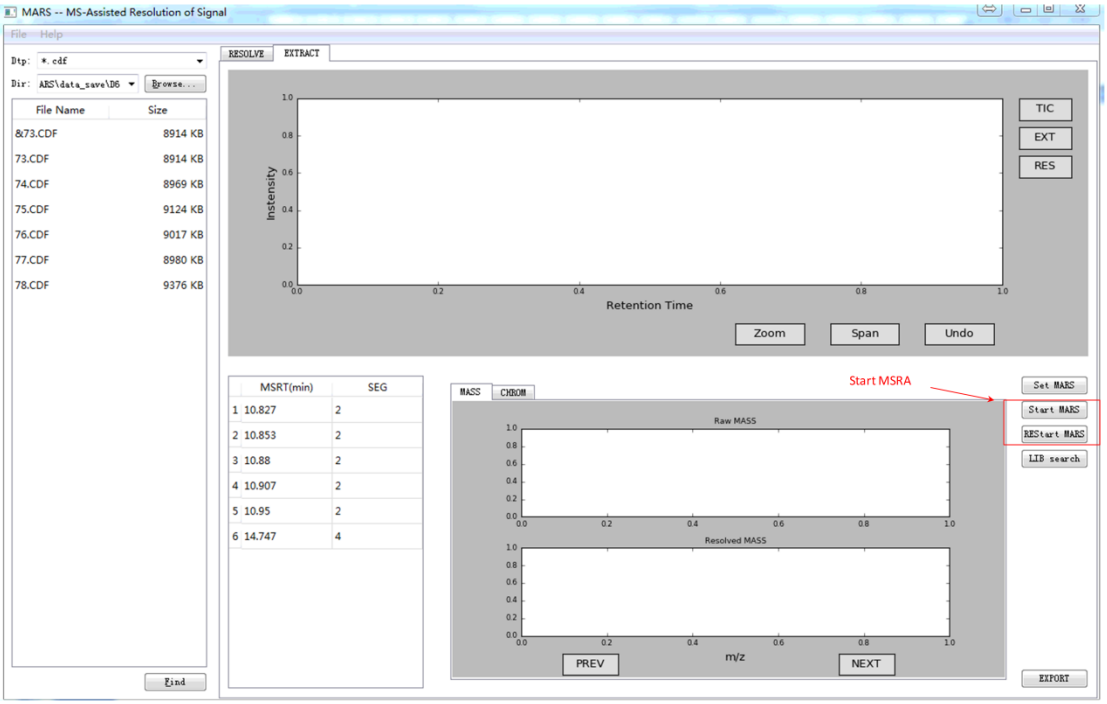

Fig. 17 Extract results of MARS2.

Start MARS: extract features of new added MSRTs in MSRT table.

REStart MARS: extract features of all MSRTs in MSRT table.

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

Checking resolution results by MARS2.

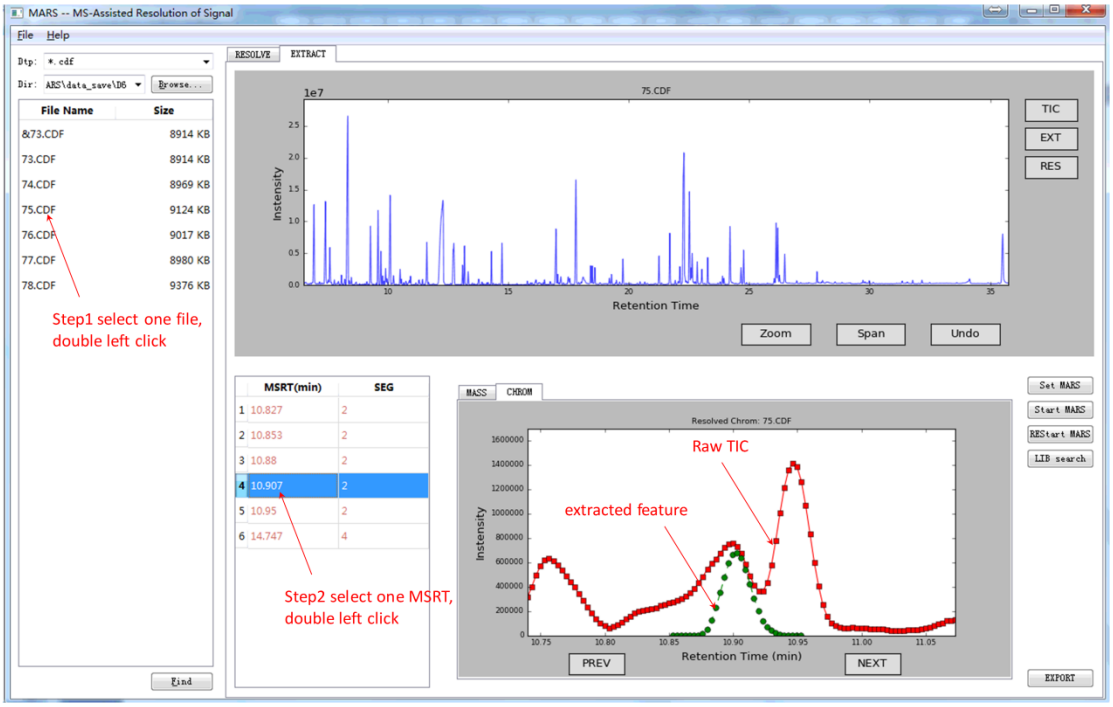

Fig. 18 Resolution results of MARS2

3.4 Library Search

MARS2 compares all resolved mass spectra found against a reference database. This package includes a custom version of the MassBank MS library. Fig. 19 shows how to search library in MARS 2.

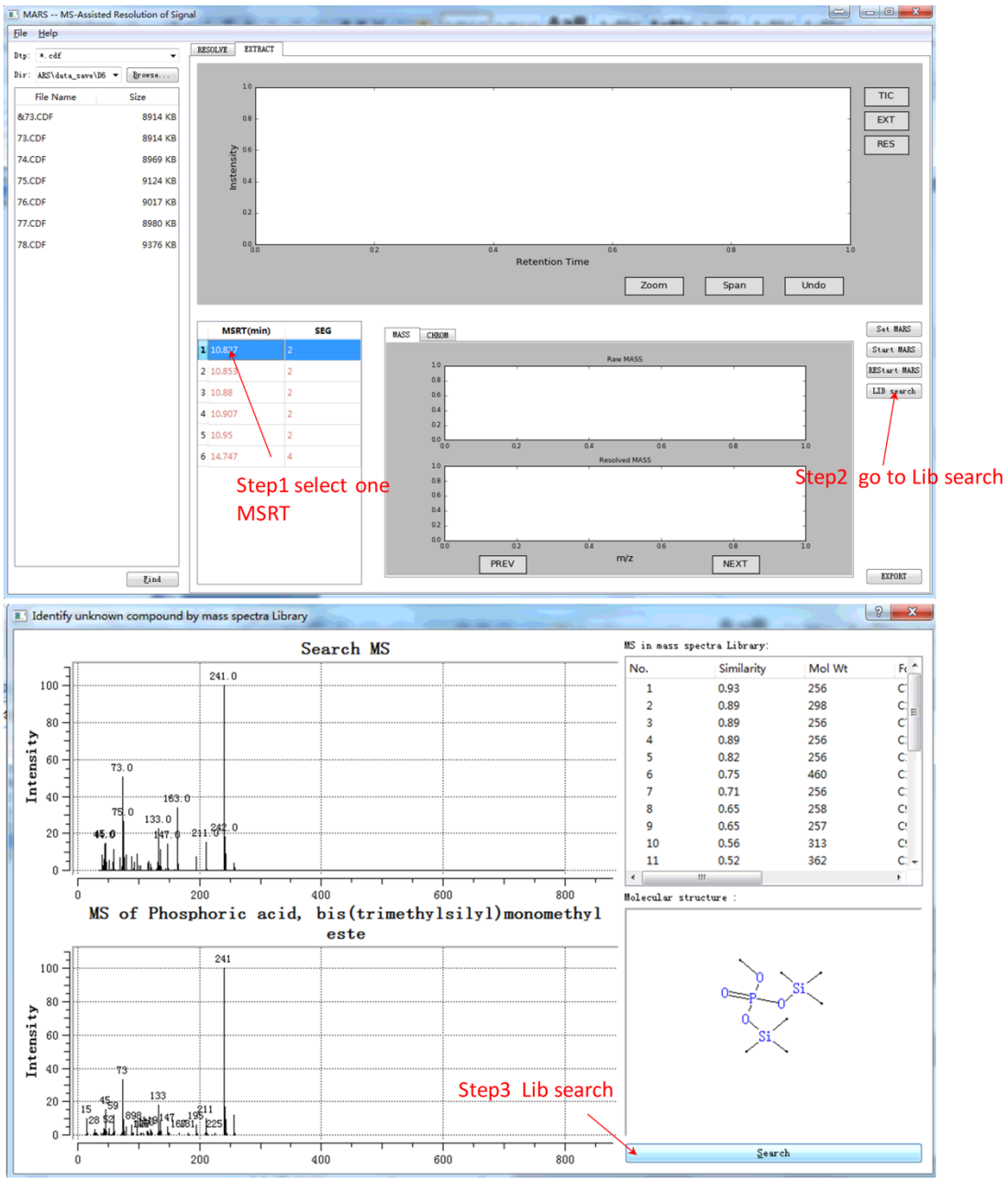

Fig. 19 Library Search

3.5 Export

This part is to export quantitative table and qualitative table (Fig.20).

	A	B	C	D	E	F	G	H	I
1	comID	RT.ref(mi	73.CDF	73.CDF	74.CDF	75.CDF	76.CDF	77.CDF	78.CDF
2	1	10.82667	1337713	1337713	1208876	891453.4	1439650	1203373	1754612
3	2	10.85333	2748007	2748007	2559723	2545649	2886005	1344840	1799088
4	3	10.88	3673454	3673454	3696258	4215921	3826313	3578283	3767119
5	4	10.90667	8511156	8511156	7753452	5497175	7882723	6212998	5444184
6	5	10.95	12693034	12693034	12960651	12002511	10414485	13066601	10671354
7	6	14.74667	56924522	56924522	58615640	55581232	57532812	57538116	59027994

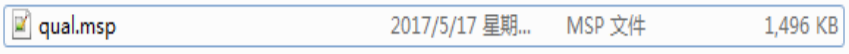

Fig. 20 Quantitative and qualitative table.

When you export results, quan.txt and qual.msp file is saved in selected direction and this file can be import into MS Search v.2.0 for identification analysis (Fig. 21).

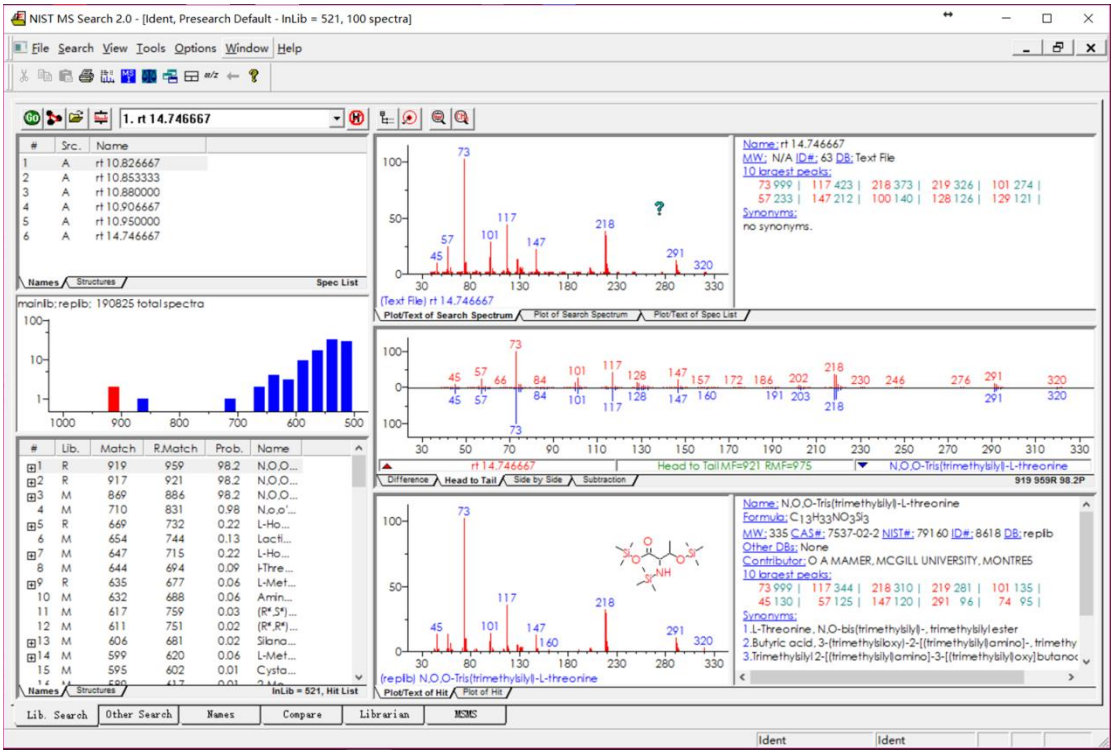

Fig. 21 Identification results of MS Search v2.0.

Reference:

[1] Y.-Z. Liang, O.M. Kvalheim, A. Rahmani, R.G. Brereton, A two-way procedure for background correction of chromatographic/spectroscopic data by congruence analysis and least-squares fit of the zero-component regions: comparison with double-centering, Chemom. Intell. Lab. Syst. 18 (1993) 265–279. doi:10.1016/0169-7439(93)85003-Y.

[2] P.J. Gemperline, A priori estimates of the elution profiles of the pure components in overlapped liquid chromatography peaks using target factor analysis, J. Chem. Inf. Comput. Sci. 24 (1984) 206–212. doi:10.1021/ci00044a004.

[3] Y.-Z. Liang, O.M. Kvalheim, Heuristic evolving latent projections: Resolving hyphenated chromatographic profiles by component stripping, Chemom. Intell. Lab. Syst. 20 (1993) 115–125. doi:10.1016/0169-7439(93)80009-7.

[4] O.M. Kvalheim, Y.Z. Liang, Heuristic evolving latent projections: resolving two-way multicomponent data. 1. Selectivity, latent-projective graph, datascope, local rank, and unique resolution, Anal. Chem. 64 (1992) 936–946. doi:10.1021/ac00032a019.

[5] Y.Z. Liang, O.M. Kvalheim, H.R. Keller, D.L. Massart, P. Kiechle, F. Erni, Heuristic evolving latent projections: resolving two-way multicomponent data. 2. Detection and resolution of minor constituents, Anal. Chem. 64 (1992) 946–953. doi:10.1021/ac00032a020.

[6] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, Chemom. Intell. Lab. Syst. 76 (2005) 101–110. doi:10.1016/j.chemolab.2004.12.007.

[7] R. Tauler, A. Smilde, B. Kowalski, Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution, J. Chemom. 9 (1995) 31–58. doi:10.1002/cem.1180090105.

[8] R. Tauler, Multivariate curve resolution applied to second order data, Chemom. Intell. Lab. Syst. 30 (1995) 133–146. doi:10.1016/0169-7439(95)00047-X.