Chromatographia (2016) 79:1247–1255
DOI 10.1007/s10337-016-3155-4

ORIGINAL

Recursive Wavelet Peak Detection of Analytical Signals

Xia Tong[1] · Zhimin Zhang[1] · Fanjuan Zeng[1] · Chunyan Fu[1] · Pan Ma[1] · Ying Peng[1] · Hongmei Lu[1] · Yizeng Liang[1]

Received: 30 April 2016 / Revised: 22 July 2016 / Accepted: 27 July 2016 / Published online: 13 August 2016

Abstract A novel algorithm, entitled recursive wavelet peak detection (RWPD), is proposed to detect both normal and overlapped peaks in analytical signals. Recursive peak detection is based on continuous wavelet transforms (CWTs), which can be used to obtain initial peak positions even for overlapped peaks. Genetic algorithm (GA) and Gaussian fitting are used to refine peak parameters (peak positions, widths, and heights). Finally, area of peaks can be calculated by numeric integration. Simulated and ultra-high performance liquid chromatographic ion trap time-of-flight mass spectrometry (UPLC-IT-TOF-MS) data sets have been analyzed by RWPD, MassSpecWavelet, and peakfit package by Tom O'Haver. Results show that RWPD can obtain more accurate positions and smaller relative fitting errors than MassSpecWavelet and peakfit, especially in overlapped peaks. RWPD is a convenient tool for peak detection and deconvolution of overlapped peaks, and it has been developed in R programming language and is available at https://github.com/zmzhang/RWPD.

Keywords Peak detection · Overlapped peaks · Continuous wavelet transforms · Genetic algorithm

Introduction

Accurate peak detection of chromatographic signals is critical to further data analysis, and it is a fundamental step for both qualitative and quantitative analysis in practical applications. LC–MS is one of the most powerful tools for analyzing complex samples, but the deconvolution of peaks in extracted ion chromatograms (EIC) is challenging to the existing software tools [1]. Determination of peak parameters was often impacted seriously by overlapped peaks. Several techniques have developed for peak detection, which often follow two strategies, derivative and pattern matching [2]. Derivative-based peak detection uses the first derivative of a peak has a zero-crossing at its local maximum or the second derivative has a negative region to determine a peak. To avoid false positives, the threshold on slope or amplitude in zero crossings and negative regions is often imposed, so that those thresholds exceeds a predetermined minimum can be retained [3, 4]. Peakfit package relies on the first derivative to find peaks and resolve signals [5]. The famous pattern matching of peak detection is MassSpecWavelet [6]. It is based on CWT and maintains a low false positive rate. Zhang has applied CWT-based peak detection in baselineWavelet [7], alignDE [8], MSPA [9], and CAMS [10], and then, an improved peak detection method entitled MSPD [11] has been proposed recently. Comparing with other methods, Cromwell [12], Limpic [13], Lms [14], and PROcess [15], CWT provides the best average performance [16]. As the complexity increasing, MassSpecWavelet is still sophisticated, whereas derivative-based methods require more preprocessing, such as baseline correction and smoothing. When analyzing complex samples with analytical instruments, overlapped peaks always appear which is difficult to extract quantitative information accurately from overlapped peaks. MassSpecWavelet may fail to detect these peaks in them, which means that vital information may be lost and error is unavoidable.

The peak model is significant to the deconvolution of overlapped peaks. The mathematical models should be

✉ Zhimin Zhang
zmzhang@csu.edu.cn

[1] Institute of Chemometrics and Intelligent Instruments, College of Chemistry and Chemical Engineering, Central South University, Changsha 410083, People's Republic of China

sufficient flexible enough to fit different shapes of peaks. In the literature, peak models mainly include exponentially modified Gaussian (EMG), polynomial modified Gaussian function (PMG), hybrid of Gaussian and truncated exponential functions (EGHs), and bi-Gaussian mixture model [17–19]. These models are designed to fit limited peak shapes, such as asymmetric, fronting, and tailing peaks. In addition, the number of parameters of them is more than four and cannot easily be determined, and thus, the versatility of them may be suffered. For these reasons, they are not benefit to achieve automatic detection of overlapped peaks [20, 21].

This study aims to develop an automatic peak detection method for both normal and overlapped peaks in analytical signals. CWT-based pattern matching is utilized for peak detection. It can not only directly apply to the raw chromatograms without baseline correction and signal smoothing but also identify each peak accurately [6]. The segments of overlapped peaks in analytical signals are extracted to perform deconvolution. Genetic algorithm (GA) is then used to optimize positions and widths of overlapped peaks to obtain optimal solutions in acceptable time, and the balance between population sizes and iterations is adjusted by grid searching. Combining with the results of GA, Gaussian fitting and trapezoidal integration are employed to calculate peak heights and peak areas of each fitting curve. To obtain exactly peak parameters, the baseline can be corrected by linear model [22] or airPLS [23] if necessary. After baseline correction, RWPD is applied to this baseline-corrected signal. If residual signal is large after deducting detected peaks from raw signal, there may exist undetected peaks. Then, CWT peak detection is performed recursively with residual signal until the residue is small enough. When it cannot detect new positions or can detect the approximation positions with last time, this process will be terminated. If new peak positions are detected, then repeat the above steps to obtain better fitting results. The flow chart describing the architecture of RWPD is shown in Fig. 1.

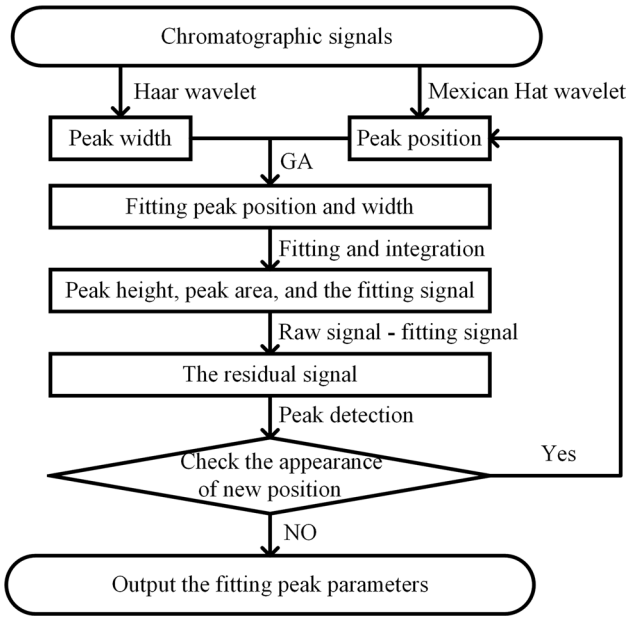

Fig. 1 Flow chart describing the framework of RWPD

Theory

Recursive Peak Detection via Continuous Wavelet Transforms

Identification of peak positions is a critical step in analysis of analytical signal. One of the most popular techniques is based on CWT. CWT can analyze signal at some special frequency or sets of frequencies (scales), and it has been widely applied in peak detection [24–30].

Wavelet theory is based on a series of basic functions which are continuously differentiable and zero mean. Mother wavelet is represented as follows:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right),\ a \in R^{+}, b \in R, \tag{1}$$

where a and b represent the dilation (or scaling) and translation parameter, respectively. The CWT can be represented as:

$$C(a,b) = \int_{-\infty}^{+\infty} s(t)\psi_{a,b}(t)\mathrm{d}t, \tag{2}$$

where $s(t)$ denotes given signal, and $C(a,b)$ represents 2D matrix of wavelet coefficients [31]. The Mexican hat wavelet is similar to Gaussian and Lorentzian functions, and it is symmetrical and has one major positive peak. Therefore, it is selected as mother wavelet and described mathematically as:

$$\psi(x) = \left(\frac{2}{\sqrt{3}}\pi^{-\frac{1}{4}}\right)(1-x^{2})e^{-\frac{x^{2}}{2}}, \tag{3}$$

where $\psi(x)$ represents the Mexican hat wavelet.

The peak identification process can be divided into four steps: (1) identify the ridges by linking the local maxima in 2D matrix of CWT coefficients; (2) define of the signal to noise ratio; (3) identify the peaks based on the ridges lines; and (4) refine the peak parameters estimation.

The peak width can be estimated roughly according to its optimal scale in wavelet space. It is based on CWT using the Haar wavelet function to improve the SNR during the derivate calculation [7]. In Fig. 2, the initial estimation of peak positions and widths by Mexican hat wavelet and Haar wavelet is marked as solid square points and circles.

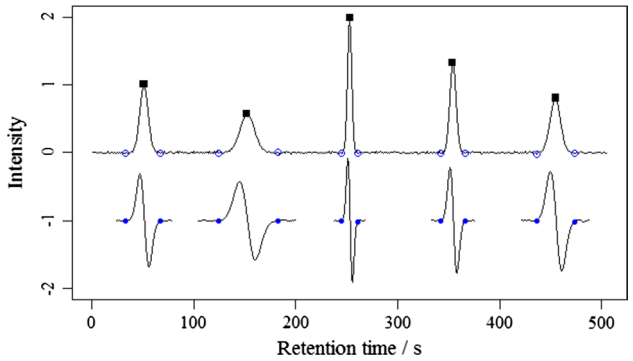

Fig. 2 Detect peak positions by Mexican hat wavelet and mark them with solid squares. Estimate peak widths by calculating the derivative using Haar wavelet. For each peak, its derivative by the optimal Haar wavelet has been shown at the *bottom* of this figure

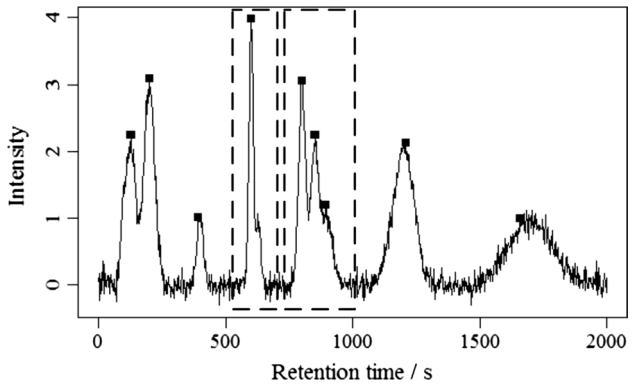

Fig. 3 Results of peak detection of simulated data. The *solid squares* are the estimation of peak positions by CWT with Mexican hat wavelet. The segments in the *dotted boxes* are overlapped peaks and they will be deconvolved by RWPD

The peak detection method based on CWT such as MassSpecWavelet can detect most peaks in signal (please see Fig. 3), and the results are robust and accurate. However, the main defect of MassSpecWavelet is that it cannot handle overlapped peaks well. If there exist relatively weak peaks overlapped with strong peaks on their both sides, MassSpecWavelet may not detect the weak peaks. To address it, the solution has been proposed in this study as follows. Peak positions of raw analytical signal are detected by CWT, and GA and Gaussian fitting can fit the detected peaks (please see the next section for more details). Then, deducting fitted signal from raw signal, the residual signal includes undetected peaks but without or smaller influence from overlapped large peaks. It can be applied to search whether there are undetected peaks in residual signal. These results combine with first fitting results can determine all peak positions in raw signal and repeat above procedure until no new peaks have been detected. GA, Gaussian fitting, and integral are used to get the peak widths, heights, and areas. This procedure is called recursive wavelet peak detection (RWPD), which can remedy the major defect of MassSpecWavelet. By RWPD, each peak in analytical signal, weak, strong or overlapped peaks, can be determined effectively.

Deconvolve Overlapped Peaks by Genetic Algorithm

GA was based on the famous evolutionary rule of Darwin, which is survival of the best. In 1975, J. Holland introduced GA, and it was used in mathematics, physics, and chemistry. In chemistry, it was used to predict the chromatographic retention time in LC [32–36], and peak alignment of 1H-NMR and IR spectra [32, 37].

As an optimization method, GA has several advantages over other local searching techniques: (a) simple, efficient, and accurate in computation. (b) Global optimization method to avoid local optimization. (c) Select the most suitable solution from series of optimal solutions. (d) Solve different search space, as continuity, discrete or existence of derivation [38, 39]. Owing these advantages, GA is suitable for optimizing parameters of each peak in overlapped peaks.

The advantage of GA over peakfit package is that GA can set the boundary for each parameter to narrow the search range. The initial input parameters are the estimation of peak positions and widths in previous step. The difference between the fitting signal $\hat{y}$ and raw signal y is regarded as the error. To minimize the error, the negative of least squares norm (L-2) of it is regarded as fitness function and can be calculated as:

$$\text{Fitness function} = -(||y - \hat{y}||^2). \tag{4}$$

Baseline Correction for Better Quantification

It is difficult to directly find accurate peak areas of the overlapped peaks with the existence of a non-zero baseline. To calculate exact peak areas by integration, two algorithms of baseline correction techniques are introduced.

The baseline, which is simulated linearly using starting and ending data points, is called polynomial fitting method. These points are 10 % of the total points that are selected, respectively. The linear model is applied to these points to fit a line, which is defined as baseline. Thus, the signal without baseline can be calculated by deducting the baseline from the raw signal [22].

When signal is simple, polynomial fitting method can be selected to remove baseline. However, if baseline drift is complex, the polynomial fitting method performs poorly, more flexible methods, such as adaptive iteratively reweighted penalized least squares (airPLS) [23], MPLS [40], SirQR [41], and ATEB [42], can replace it to remove baselines. airPLS is simple but flexible, valid, and fast algorithm for estimating baseline. The parameters in airPLS can

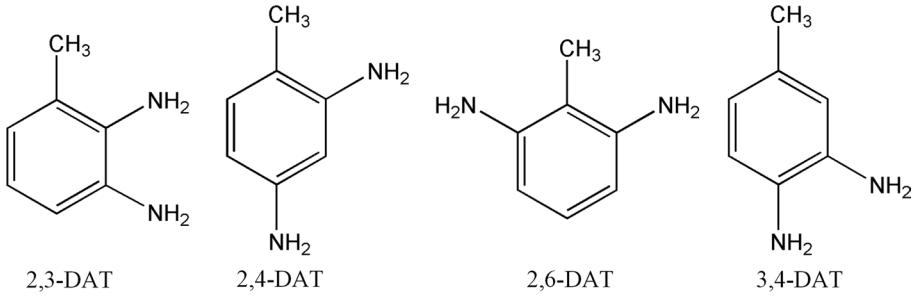

Fig. 4 Chemical structures of *2,3*-DAT, *2,4*-DAT, *2,6*-DAT, and *3,4*-DAT

be set as follows. The lambda is adjustable parameters, and the larger lambda, the baseline is smoother. When baseline is similar to the linear function, it can be set 10^2; if the baseline is quadratic function, it is 10^4 generally. It has been applied to chromatograms, Raman spectra, and NMR signals, and its performance is better than other baseline correction methods.

Extract Important Features of Peaks

Least squares fitting and Gaussian fitting can be used together to infer peak heights as Eq. (5) shows. In these two equations, X denotes the computation results by Gaussian model; y denotes the measurement data; h denotes the corresponding peak height:

$$\begin{aligned} y &= Xh \\ h &= (X'X)^{-1}X'y. \end{aligned} \tag{5}$$

The area of peak can be calculated by trapezoidal integration on the multiplication of estimated height and Gaussian peak model.

Experimental

Both simulated data and real data are used to benchmark the performance of RWPD, and the segments of overlapped peaks in them are extracted to resolve orderly.

Simulated Data

The simulated data consists of Gaussian peaks by adding to 1 % random noise to the data in Fig. 3. The resolutions of the triplet overlapped peaks near 850 and the doublet overlapped peaks near 600 are shown in Sects. 4.1 and 4.5, respectively.

LC–MS Data Set of Isomers of Diaminotoluenes

2,4-Diaminotoluene(*2,4*-DAT) is widely used as intermediates in the synthesis of dyes, and it has carcinogenicity and genotoxic activity [43–45]. To avoid the interference of the non-banned structural isomers (*2,3*-DAT, *2,6*-DAT and *3,4*-DAT) on the determination of *2,4*-DAT [46], an effective LC–MS method combining with RWPD was established. It also solves effectively the false positive problem in the analysis of *2,4*-DAT. The chemical structures of DATs isomers are shown in Fig. 4.

Chemicals and Reagents

2,3-DAT, *2,4*-DAT, *2,6*-DAT, and *3,4*-DAT were purchased from Dr. Ehrenstorfer GmbH (Germany). HPLC-grade acetonitrile was from Merck (Germany). LC–MS grade formic acid was purchased from Sigma (America). HPLC-grade methyl alcohol was purchased from Merck (Germany). The water used in all test was treated in a Milli-*Q* water purification system (Millipore, Bedford, MA, USA).

Preparation of Standard Solutions

Standard stock solutions of drugs (including *2,3*-DAT, *2,4*-DAT, *2,6*-DAT, and *3,4*-DAT) were dissolved in methyl alcohol. Mixture standard solutions were prepared by mixing stock solutions and diluting appropriately with methyl alcohol. The concentrations of them were 2.04, 1.94, 2.18, and 2.45 μg mL^{-1}, respectively.

Apparatus

All sample analyses carried out on a UPLC-IT-TOF-MS system (Shimadzu, Tokyo, Japan). LC experiments were conducted on a Shimadzu (Kyoto, Japan) ultrahigh performance liquid chromatography (UPLC) system consisting of a solvent delivery pump (LC-30AD), an auto-sampler (SIL-30AC), a DGU-20A5R degasser, a photodiode array detector (SPD-M20A), a communication base module (CBM-20A), and a column oven (CTO-30A). Chromatographic separation was carried out on a column of Shim-pack XR-ODS (1.6 μm, 2. 0 mm I.D. × 75 mm) using a gradient elution consisting of mobile phase A (0.1 % formic acid) and mobile phase B (acetonitrile). The gradient was as follows: 0–3 min, a linear gradient from 5 % B to 10 % B; 3–8 min,

a linear gradient to 90 % B; 8–8.01 min a linear gradient back to 5 % B. The injection volume was 5 μL, the flow rate was 0.4 L min^{-1}, and PDA detection was performed from 190 to 800 nm. The sample chamber in the autosampler was maintained at 4 °C, while the column was set at 40 °C. The whole analysis lasted 10 min.

Mass spectral data for the compounds were obtained using a Shimadzu ITTOF mass spectrometer. It was equipped with an electrospray ionization (ESI) source operated in the positive ionization mode. Liquid nitrogen was used as nebulizing gas at a flow rate of 1.5 L min^{-1}, drying gas (N_2) pressure 0.1 MPa. The interface and detector voltages were set at 4.5 and 1.56 kV, respectively. The CDL voltage sets at constant mode (optimized by autotuning), and its temperature was 200 °C. Mass spectrometry was conducted in the full scan and automatic multiple stage fragmentation scan modes over an *m*/*z* range of 100–500 for MS^1. The ion accumulation time was set at 10 ms. Argon was used as the collision gas. Trifluoroacetic acid (TFA) sodium solution was used as the standard sample for calibrating the instrument against the entire mass range (*m*/*z* 100–2000). Data processing was performed using the LC–MS Solution software (version 3.70).

LC–MS Data Set of FaahKO

The faahKO package consist quantitated LC/MS peaks from the spinal cords of six wild-type and six FAAH knockout mice. The data are a subset of the original data from 200 to 600 *m*/*z* and 2500–4500 s, and it is collected in positive ionization mode. The extraction ion chromatographic (EIC) in Fig. 7a is a sample in FAAH knockout mice, and its *m*/*z* range is between 429.0 and 429.5. The EIC in Fig. 7c is a sample in wild type, and its *m*/*z* range is between 575 and 575.5 [47, 48].

Results and Discussion

RWPD can fit peak parameters of each peak in the signal. Here, undetected peaks and overlapped peaks are selected to test the performance of it.

Results and Comparisons with Previous Methods on Simulated Data Set

The peakfit package is capable of measuring peak positions and heights accurately; however, peak widths and areas are accurate only when peak shapes are approximate Gaussian or Lorentzian. The comparison of fitting results of simulated data by RWPD and peakfit is shown in Fig. 5 and Table 1.

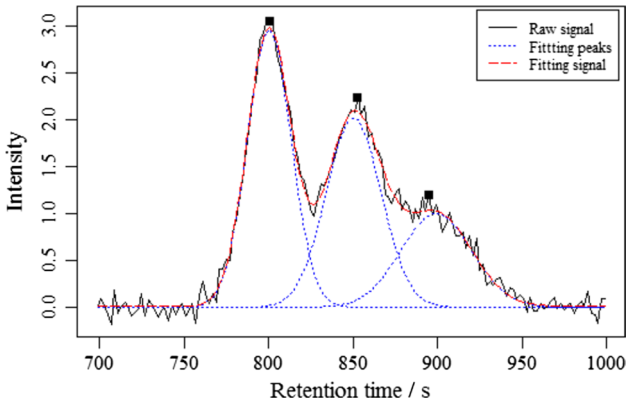

Fig. 5 Deconvolution results of simulated data by RWPD. The *solid squares* denote the detect peak positions by Mexican hat wavelet. *Black lines*, *red dashed lines,* and *blue dotted lines* represent raw signals, fitting signals, and fitting peaks, respectively

Table 1 Comparison of the estimation of peak parameters among expected, RWPD and peakfit package

No. peak	Method	Peak position	Peak height	Peak width	Peak area
1	Expected	800.0000	3.0000	30.0000	95.8020
	RWPD	800.3400	2.9584	30.5743	96.2834
	Peakfit	800.0400	3.0628	29.3150	95.5830
2	Expected	850.0000	2.0000	40.0000	85.1574
	RWPD	850.7994	2.0315	39.2496	84.8775
	Peakfit	850.1500	1.9881	41.0140	86.8040
3	Expected	900.0000	1.0000	50.0000	53.2234
	RWPD	900.2079	0.9927	47.5860	50.2855
	Peakfit	901.3000	0.9699	46.8610	48.3760
Fitting error (%)	RWPD	1.3019			
	Peakfit	3.2315			

From Table 1, it is found that the fitting error of RWPD is lower than peakfit. Comparing the estimation of strong peaks, both methods have good performance, but when estimating weak peaks, RWPD is more accurate and closer to the expected than peakfit.

Results of LC–MS Data Sets

The molecular ions of *m*/*z* 123.0912 for DATs isomer were analyzed using RWPD, and four peaks were observed, and the fitting signal is consistent with EIC in Fig. 6. Each fitting peaks in EIC is corresponding to different structural isomers, and they are *3,4*-DAT, *2,3*-DAT, *2,6*-DAT, *2,4*-DAT, respectively. Each sample in DATs was analyzed by LC–MS, and the retention time was consistent with fitting peaks in EIC. The quantitative analysis is based on

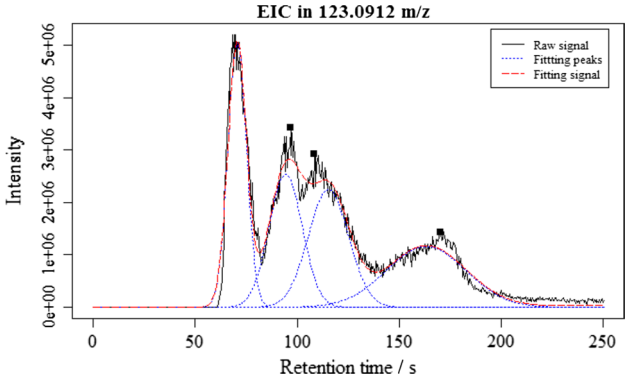

Fig. 6 Deconvolution of DATs EIC into Gaussian peak shapes by RWPD. Each peak with different elution times correspond to different chemical isomeric structures. The *lines* and *solid squares* have the same meaning as given in Fig. 5

integrating the area under the curve to estimate the relative abundance.

The EICs of faahKO data sets were extracted by XCMS package. The overlapped peaks in EICs of LC–MS data set are also applied to benchmark the performance of RWPD, and the results are shown in Fig. 7. One can observe that the fitting signal match pretty well with the raw (or baseline corrected) signal from the deconvolution results, which means that almost all the peaks information in the overlapped peaks has been correctly extracted.

The Choice of Peak Model

The Gaussian and Lorentzian are used as peak models to fit overlapped peaks in chromatograms and spectra, respectively. They have less parameters, and can give a reasonable fit to most experimental peaks in this study. Although signals are more complex and always impacted by random noise or baseline drift, these effects can be dealt with our method. Based on the characteristics of the signals, Generally, Gaussian is used as peak model of chromatograms; Lorentzian is used to fit spectra. In the cases of chromatographic peak serious distortion, tailing, and heavy overlapping in real application, these models may fail to solve them and other functions may be implemented.

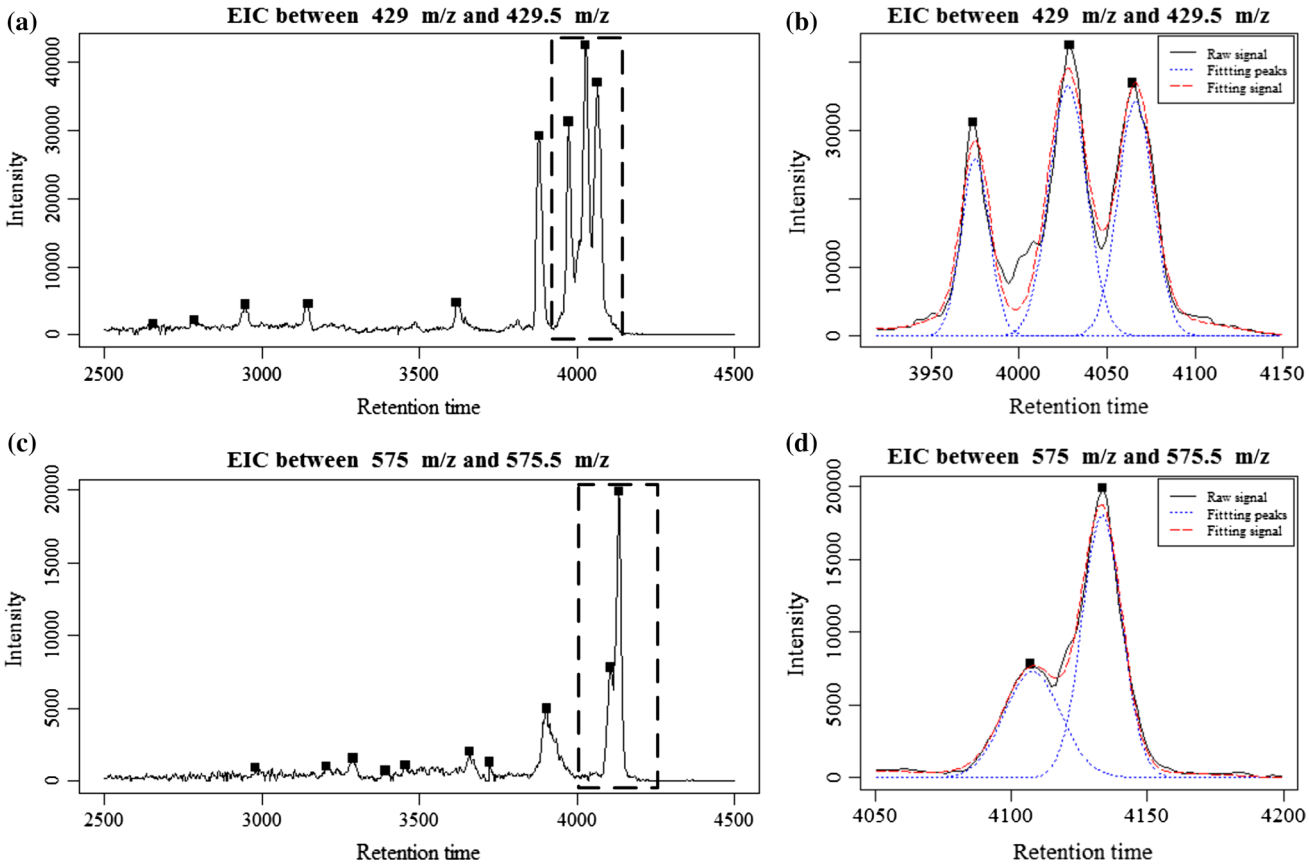

Fig. 7 Deconvolution results of EICs by RWPD. The segments in the *dotted boxes* are overlapped peaks, and they will be deconvolved by RWPD. The *lines* and *solid squares* have the same meaning as given in Fig. 5

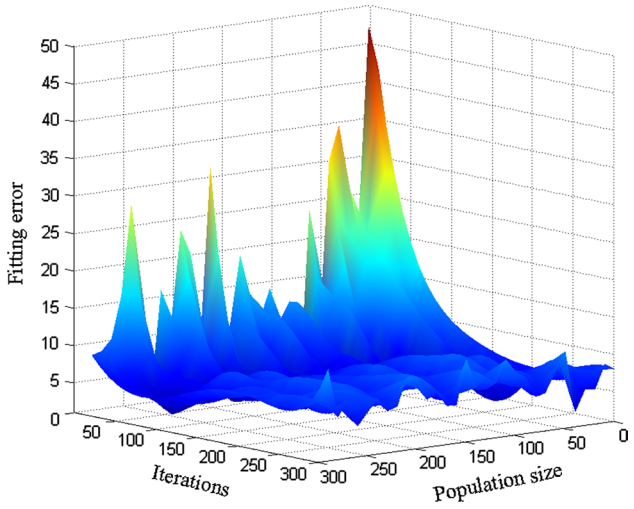

Fig. 8 Fitting error and its relationship with population sizes and iterations

Balance between Accuracy and Computation Speed

The choice of population sizes and iterations can affect the final results of GA. The lower number of population sizes or iterations may not search the optimal solution before the process stopped. In theory, the greater population sizes or iterations can keep the population diversity, and the fitness results are closer to the true values and error is lower. However, this may take up lots of computational resources, and leads to reduce the search efficiency. The suitable population sizes and iterations should be determined to achieve accurate solution within acceptable time. The iterations are the convergence criteria of GA, when the maximum number of iterations has been reached, and GA is terminated.

The relationship between population sizes, maximum iterations, and fitting error can be seen in Fig. 8. The fitting error is a dependent variable of population sizes and maximum iterations. It is calculated with different iterations and population sizes. Both of them range from 1 to 300, and the interval is 10. The data have been smoothed to obtain clearer trends between them. The relationships are obvious: with the increase in population sizes and iterations, the fitting error is reducing obviously; especially, population sizes and iterations are less than 100 and 150, respectively. It has been tested the relationship between them with many overlapped peaks. When population sizes and iterations equal 100 and 150, respectively, this method can achieve acceptable fitting error.

Recursive Wavelet Peak Detection of Overlapped Signal

RWPD can extract each peak from the overlapped signal. In simulated data, there is overlapped peak near 600, and the weak peak cannot be detected in Fig. 3. The residual signal after the first iteration is shown in Fig. 9b. It is detected by CWT again until no new position is appearing. The peak positions and widths can be determined by combining the new positions with the results in first iteration. Then Gaussian fitting and other methods have been used to obtain the accurate fitting results. As shown in Fig. 9c, the undetected peak can also be detected by RWPD. Table 2 shows the results by comparison the first and second iterations, and the fitting results are more accurate. The fitting signal matches better with the raw signal, while the fitting error is rapidly decreasing.

Conclusion

In this study, we present a practical peak detection method by seamlessly combining RWPD and heuristic optimization. RWPD has been proposed for peak positions estimation of overlapped peaks. It is significantly better than the traditional peak detection method based on CWT. Heuristic optimization has been used to optimize the important features of peaks including positions, widths, heights, and

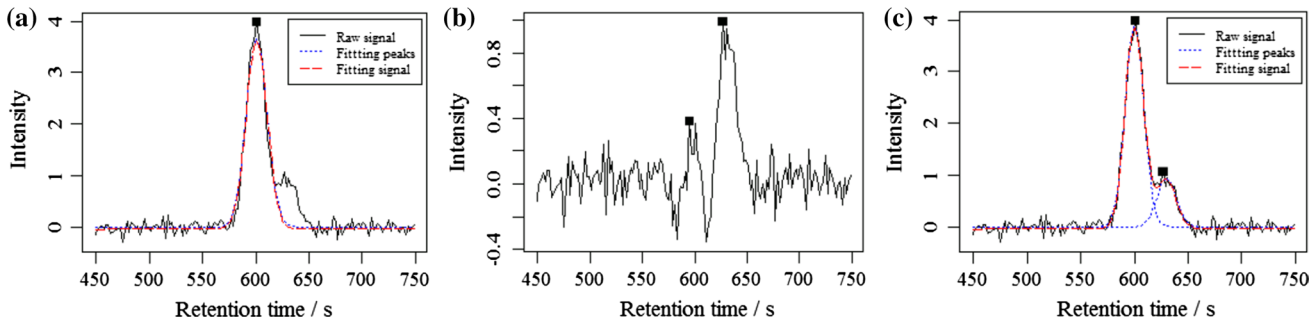

Fig. 9 Results of RWPD of simulated data in *dotted boxes* in Fig. 3. **a** In the first iteration, only strong peaks are detected and fitted; **b** detecting peak positions in residual signal; **c** combining with the results of (**a**, **b**), fitting each peak again. *Black lines* denote the residual signal in (**b**). The *lines* and *solid squares* have the same meaning as given in Fig. 5

Table 2 Comparison of the fitting results of simulated data in the first and second iterations by RWPD

No. peak	Iterations	Peak position	Peak height	Peak width	Peak area
1	1	600.7811	3.6486	23.1839	90.0408
	2	600.0782	3.9022	20.3308	84.4494
2	1	–	–	–	–
	2	629.6234	0.9405	21.5917	21.6151
Fitting error (%)	1	91.4853			
	2	25.8413			

areas. The initial value and boundary of each parameter to be optimized can be obtained from RWPD. By investigating the results of simulated and LC–MS data set, one can observe that RWPD has more accurate positions and smaller relative errors than MassSpecWavelet and peak fit, especially in overlapped peaks. It means that our method is suitable for extracting features of scientific interest from complex analytical signals.

Acknowledgments This study was supported by the National Nature Foundation Committee of People's Republic of China (Grants No. 21275164, Grants No. 21375151 and Grants No. 21305163), Hunan Provincial Natural Science Foundation of China (Grants No. 14JJ3031), National Instrumentation Program of China (No. 2011YQ03012407) and China Postdoctoral Science Foundation (No. 2014M552146). The studies meet with the approval of the university's review board. We are grateful to all employees of this institute for their encouragement and support of this research.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animal performed by any of the authors.

References

1. Kronewitter SR, Slysz GW, Marginean I, Hagler CD, LaMarche BL, Zhao R, Harris MY, Monroe ME, Polyukh CA, Crowell KL, Fillmore TL, Carlson TS, Camp DG 2nd, Moore RJ, Payne SH, Anderson GA, Smith RD (2014) GlyQ-IQ: glycomics quintavariate-informed quantification with high-performance computing and GlycoGrid 4D visualization. Anal Chem 86:6268–6276
2. Lopatka M, Vivo-Truyols G, Sjerps MJ (2014) Probabilistic peak detection for first-order chromatographic data. Anal Chim Acta 817:9–16
3. Vivo-Truyols G, Torres-Lapasió JR, Van Nederkassel AM, Heyden YV, Massart DL (2005) Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals Part I: peak detection. J Chromatogr A 1096:133–145
4. Vivo-Truyols G, Torres-Lapasió JR, Van Nederkassel AM, Heyden YV, Massart DL (2005) Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals Part II: peak model and deconvolution algorithms. J Chromatogr A 1096:146–155
5. Tom OH (2015) A pragmatic introduction to signal processing with applications in scientific measurement. http://terpconnect.umd.edu/~toh/spectrum/index.html. Accessed 12 Mar 2015
6. Du P, Kibbe WA, Lin SM (2006) Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. Bioinformatics 22:2059–2065
7. Zhang Z-M, Chen S, Liang Y-Z, Liu Z-X, Zhang Q-M, Ding L-X, Ye F, Zhou H (2009) An intelligent background-correction algorithm for highly fluorescent samples in Raman spectroscopy. J Raman Spectrosc 41:659–669
8. Zhang Z-M, Chen S, Liang Y-Z (2011) Peak alignment using wavelet pattern matching and differential evolution. Talanta 83:1108–1117
9. Zhang Z-M, Liang Y-Z, Lu H-M, Tan B-B, Xu X-N, Ferro M (2012) Multiscale peak alignment for chromatographic datasets. J Chromatogr A 1223:93–106
10. Zheng Y-B, Zhang Z-M, Liang Y-Z, Zhan D-J, Huang J-H, Yun Y-H, Xie H-L (2013) Application of fast Fourier transform cross-correlation and mass spectrometry data for accurate alignment of chromatograms. J Chromatogr A 1286:175–182
11. Zhang Z-M, Tong X, Peng Y, Ma P, Zhang M-J, Lu H-M, Chen X-Q, Liang Y-Z (2015) Multiscale peak detection in wavelet space. Analyst 140:7955–7964
12. Coombes KR, Tsavachidis S, Morris JS, Baggerly KA, Hung M-C, Kuerer HM (2005) Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform. Proteomics 5:4107–4117
13. Mantini D, Petrucci F, Pieragostino D, Del Boccio P, Di Nicola M, Di Ilio C, Federici G, Sacchetta P, Comani S, Urbani A (2007) LIMPIC: a computational method for the separation of protein MALDI-TOF-MS signals from noise. BMC Bioinformatics 8:101
14. Yasui Y, Pepe M, Thompson ML, Adam B-L, Qu Y, Potter J, Winget M, Thornquist M, Feng Z (2003) A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection. Biostatistics 4:449–463
15. Li X, Gentleman R, Lu X, Shi Q, Iglehart J, Harris L, Miron A (2005) SELDI-TOF mass spectrometry protein data. In: Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S (eds) Bioinformatics and computational biology solutions using R and bioconductor. Springer, China
16. Yang C, He Z, Yu W (2009) Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis. BMC Bioinformatics 10:4–17
17. Caballero RD, Garcıa-Alvarez-Coque MC, Baeza-Baeza JJ (2002) Parabolic-Lorentzian modified Gaussian model for describing and deconvolving chromatographic peaks. J Chromatogr A 954:59–76
18. Buys TS, De Clerk K (1972) Bi-Gaussian fitting of skewed peaks. Anal Chem 69:3822–3831
19. Nikitas P, Pappa-Louisi A, Papageorgiou A (2001) On the equations describing chromatographic peaks and the problem of the deconvolution of overlapped peaks. J Chromatogr A 912:13–29
20. Li J-W (2002) Comparison of the capability of peak functions in describing real chromatographic peaks. J Chromatogr A 952:63–70
21. Marco VBD, Bombi GG (2001) Mathematical functions for the representation of chromatographic peaks. J Chromatogr A 931:1–30
22. Zeng Z-D, Chin S-T, Hugel HM, Marriott PJ (2011) Simultaneous deconvolution and re-construction of primary and secondary

overlapping peak clusters in comprehensive two-dimensional gas chromatography. J Chromatogr A 1218(16):2301–2310
23. Zhang Z-M, Chen S, Liang Y-Z (2010) Baseline correction using adaptive iteratively reweighted penalized least squares. Analyst 135:1138–1146
24. Galiana-Merino JJ, Pla C, Fernandez-Cortes A, Cuezva S, Ortiz J, Benavente D (2014) EnvironmentalWaveletTool: continuous and discrete wavelet analysis and filtering for environmental time series. Comput Phys Commun 185:2758–2770
25. Liu Y, Cai W-S, Shao X-G (2013) Intelligent background correction using an adaptive lifting wavelet. Chemometr Intell Lab 125:11–17
26. Shao XG, Cai WS, Sun PY (1998) Determination of the component number in overlapping multicomponent chromatogram using wavelet transform. Chemometr Intell Lab 43:147–155
27. Shao XG, Leung AK-M, Chau FT (2003) Wavelet: a new trend in chemistry. Acc Chem Res 36:276–283
28. Shao XG, Gu H, Wu JH, Shi YY (2000) Resolution of the NMR spectrum using wavelet transform. Appl Spectrosc 54:731–738
29. Shao XG, Cai WS, Sun PY, Zhang MS, Zhao GW (1997) Quantitative determination of the components in overlapping chromatographic peaks using wavelet transform. Anal Chem 69:1722–1725
30. Jiao L, Gao S, Zhang F, Li H (2008) Quantification of components in overlapping peaks from capillary electrophoresis by using continues wavelet transform method. Talanta 75:1061–1067
31. Mohammadpour K, Sohrabi MR, Jourabchi A (2010) Continuous wavelet and derivative transform applied to the overlapping spectra for the quantitative spectrophotometric multi-resolution of triamterene and hydrochlorothiazide in triamterene-H tablets. Talanta 81:1821–1825
32. Nikitas P, Pappa-Louisi A, Papageorgiou A (2007) Simple algorithms for fitting and optimisation for multilinear gradient elution in reversed-phase liquid chromatography. J Chromatogr A 1157:178–186
33. Bolanča T, Ukić Š, Novak M, Rogošić M (2014) Computer assisted method development in liquid chromatography. Croat Chem Acta 87:111–122
34. Zhang XT, Zhu HY, Zhang HB (1998) Robust grey model based on genetic algorithms and its application to prediction for chromatographic retention. Chemometr Intell Lab 44:197–203
35. Shao XG, Chen ZH, Lin XQ (2000) Resolution of multicomponent overlapping chromatogram using an immune algorithm and genetic algorithm. Chemometr Intell Lab 50:91–99
36. Holland JH (1992) Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT press, Cambridge
37. Larsen FH, van den Berg F, Engelsen SB (2006) An exploratory chemometric study of1H NMR spectra of table wines. J Chemometr 20:198–208
38. Garcia-Talavera M, Ulicny B (2003) A genetic algorithm approach for multiplet deconvolution in γ-ray spectra. Nucl Instrum Meth Phys Res A 512:585–594
39. Lovedy Singh L, Gartia RK (2014) Glow-curve deconvolution of thermoluminescence curves in the simplified OTOR equation using the hybrid genetic algorithm. Nucl Instrum Meth B 319:39–43
40. Li Z, Zhan DJ, Wang JJ, Huang J, Xu QS, Zhang ZM, Zheng YB, Liang YZ, Wang H (2013) Morphological weighted penalized least squares for background correction. Analyst 138:4483–4492
41. Liu XB, Zhang ZM, Sousa PFM, Chen C, Ouyang ML, Wei YC, Liang YZ, Chen Y, Zhang CP (2014) Selective iteratively reweighted quantile regression for baseline correction. Anal Bioanal Chem 406:1985–1998
42. Liu XB, Zhang ZM, Liang YZ, Sousa PFM, Yun YH, Yu L (2014) Baseline correction of high resolution spectral profile data based on exponential smoothing. Chemometr Intell Lab 139:97–108
43. Brennan RJ, Schiestl RH (1997) Diaminotoluenes induce intrachromosomal recombination and free radicals in Saccharomyces cerevisiae. Mutat Res-Fund Mol M 381:251–258
44. Nakayama K, Kawano Y, Kawakami Y, Moriwaki N, Sekijima M, Otsuka M, Yakabe Y, Miyaura H, Saito K, Sumida K, Shirai T (2006) Differences in gene expression profiles in the liver between carcinogenic and non-carcinogenic isomers of compounds given to rats in a 28-day repeat-dose toxicity study. Toxicol Appl Pharmacol 217:299–307
45. Toyoda-Hokaiwado N, Inoue T, Masumura K, Hayashi H, Kawamura Y, Kurata Y, Takamune M, Yamada M, Sanada H, Umemura T, Nishikawa A, Nohmi T (2010) Integration of in vivo genotoxicity and short-term carcinogenicity assays using F344 gpt delta transgenic rats: in vivo mutagenicity of 2,4-diaminotoluene and 2,6-diaminotoluene structural isomers. Toxicol Sci 114:71–78
46. Dong LL, Shion H, Davis RG, Terry-Penak B, Castro-Perez J, Van Breemen RB (2010) Collision Cross-Section Determination and Tandem Mass Spectrometric Analysis of Isomeric Carotenoids Using Electrospray Ion Mobility Time-of-Flight Mass Spectrometry. Anal Chem 82:9014–9021
47. Smith CA Saghatelian et al. (2004) FAAH knockout LC/MS data. http://bioconductor.org/packages/devel/data/experiment/manuals/faahKO/man/faahKO.pdf Accessed 31 July 2012
48. Saghatelian A, Trauger SA, Want EJ, Hawkins EG, Siuzdak G, Cravatt BF (2004) Assignment of Endogenous Substrates to Enzymes by Global Metabolite Profiling. Biochemistry 43:14332–14339