

XCMS Workshop

The centWave Algorithm

Ralf Tautenhahn

The Scripps Research Institute

5/31/2009

LC/MS data acquisition

HPLC



MS



ESI

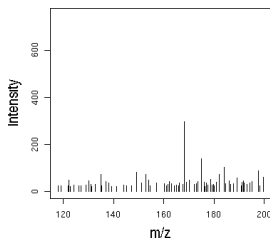
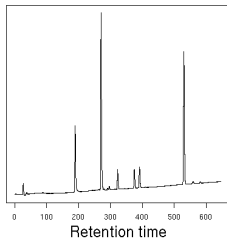


1st Separation
Retention time
depending on e.g.
polarity

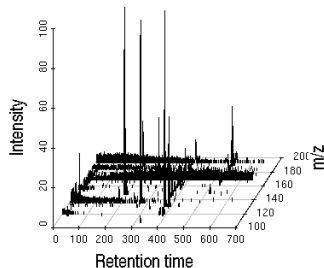
Ionization
Molecules
receive
charge

2nd Separation
Measurement of
mass/charge
(m/z)

Mass spectrum



LC/MS: high performance liquid chromatography (HPLC)/ mass spectrometry (MS)



LC/MS data acquisition

HPLC



1st Separation
Retention time
depending on e.g.
polarity

ESI

Ionization
Molecules
receive
charge

MS



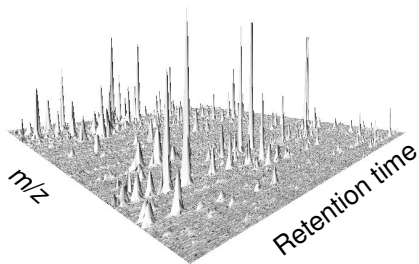
2nd Separation
Measurement of
mass/charge
(m/z)

LC/MS: high performance liquid chromatography (HPLC)/ mass spectrometry (MS)



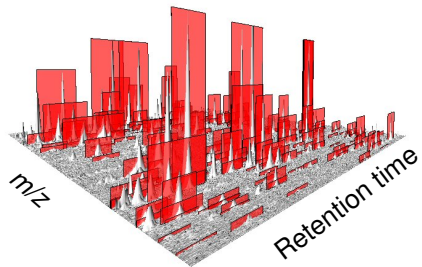
Feature detection

- Feature: 2D-Signal induced by a single ion species (e.g. $[M+H]^+$) of a compound
- Detection of all relevant features
- Suppression of noise-induced signals
- Precise quantification

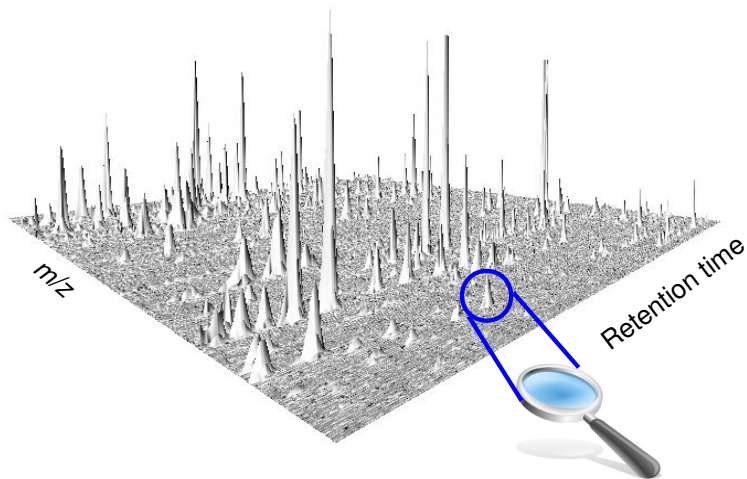


Feature detection

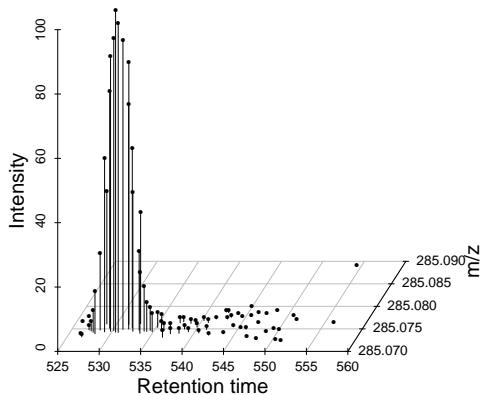
- Feature: 2D-Signal induced by a single ion species (e.g. $[M+H]^+$) of a compound
- Detection of all relevant features
- Suppression of noise-induced signals
- Precise quantification



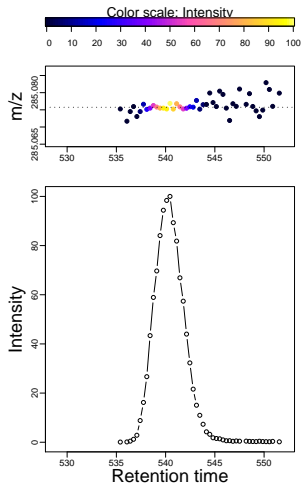
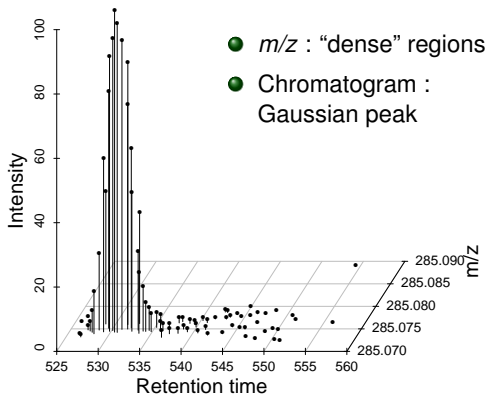
Feature characteristics



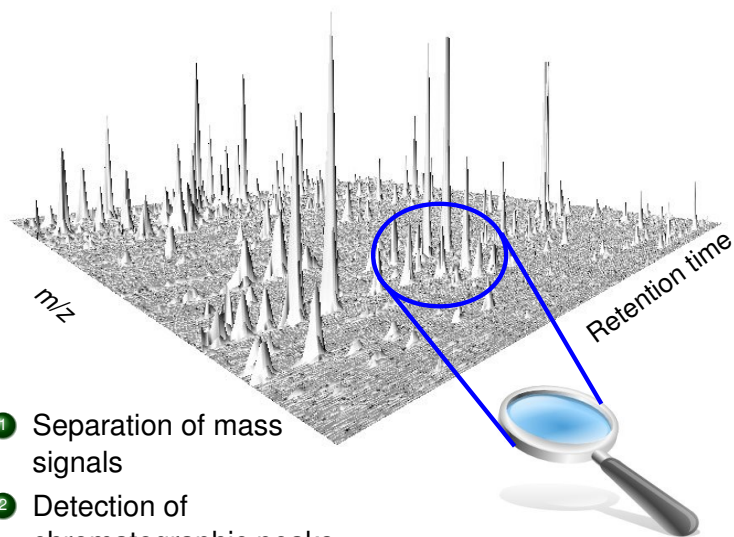
Feature characteristics (centroid mode)



Feature characteristics (centroid mode)



Feature detection

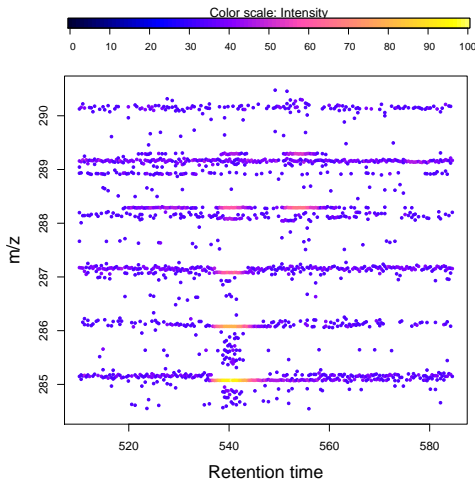


- 1 Separation of mass signals
- 2 Detection of chromatographic peaks

Step 1. Separation of mass signals

Conventional method: Binning

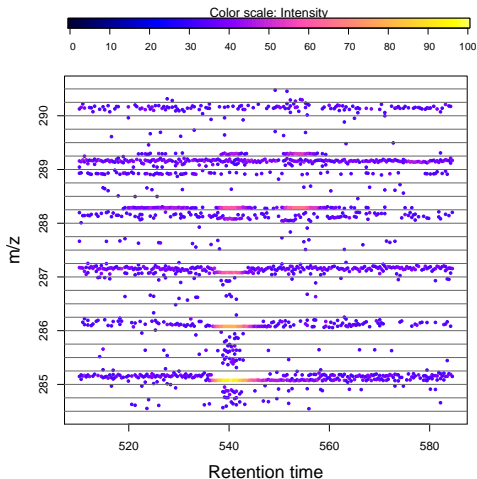
- Partition into bins of fixed width
e.g. $0.1\ m/z$
- works for profile & centroid data
- How do you estimate optimal bin width (*absolute* deviation) ?
 - TOF: fixed resolution \rightarrow
higher deviation in the high
mass range
 - too small \rightarrow split feature
 - too wide \rightarrow possible loss of
S/N ratio



Step 1. Separation of mass signals

Conventional method: Binning

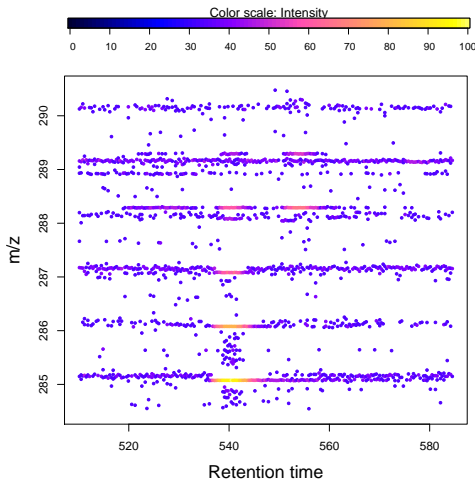
- Partition into bins of fixed width
e.g. $0.1\text{ }m/z$
- works for profile & centroid data
- How do you estimate optimal bin width (*absolute* deviation) ?
 - TOF: fixed resolution → higher deviation in the high mass range
 - too small → split feature
 - too wide → possible loss of S/N ratio



Step 1. Separation of mass signals

New method: ROI detection
(Region Of Interest)

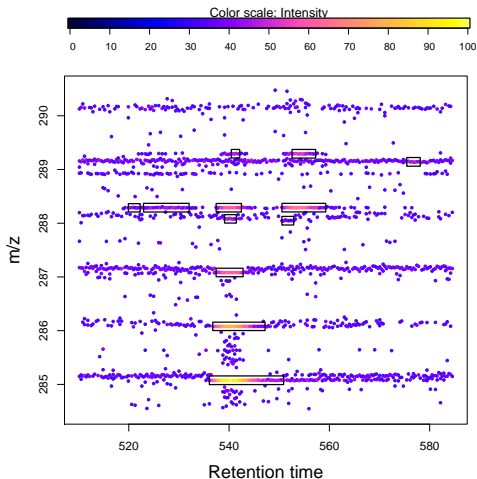
- only centroid data
- *relative* deviation (ppm)
- Examination of the mass spectra in their chronological order
- Calculate deviation of new centroids to m/z mean of existing ROI's
- Create or extend ROI



Step 1. Separation of mass signals

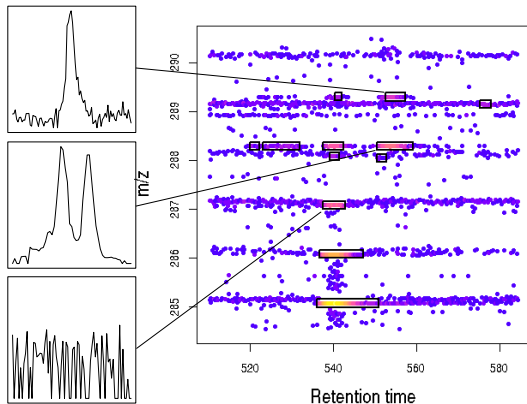
New method: ROI detection
(Region Of Interest)

- only centroid data
- *relative* deviation (ppm)
- Examination of the mass spectra in their chronological order
- Calculate deviation of new centroids to m/z mean of existing ROI's
- Create or extend ROI



Step 2: Detection of chromatographic peaks

- ROI may contain
 - no chromatographic peak (just noise)
 - a single chromatographic peak
 - several chromatographic peaks
- Detection of chromatographic peaks using Continuous Wavelet Transformation (CWT)



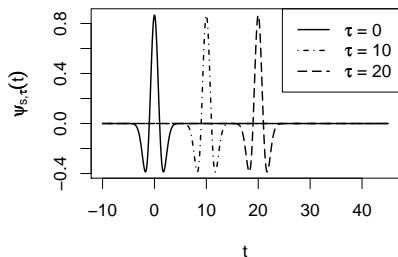
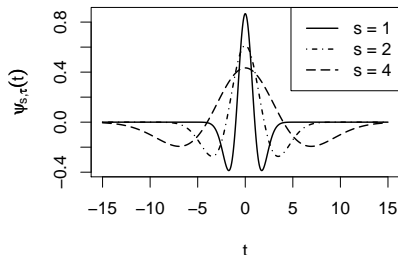
Continuous Wavelet Transformation (CWT)

- Mother wavelet: Mexican Hat
- Scaled and translated mother wavelet

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right)$$

- Continuous Wavelet Transformation

$$W_f(s, \tau) = \int_{-\infty}^{\infty} f(t) \psi_{s,\tau}(t) dt$$



Continuous Wavelet Transformation (CWT)

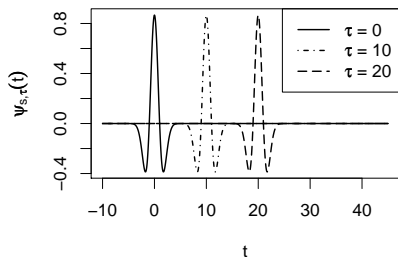
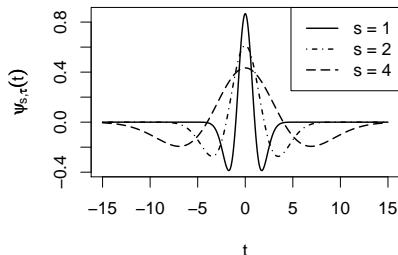
- Mother wavelet: Mexican Hat
- Scaled and translated mother wavelet

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right)$$

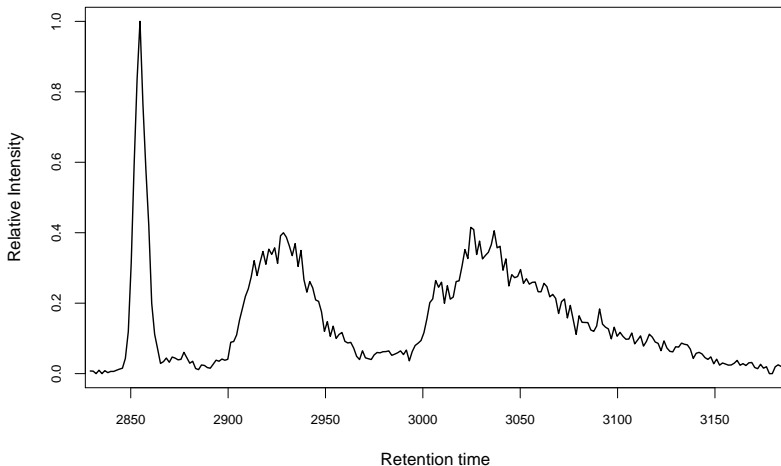
- Continuous Wavelet Transformation

$$W_f(s, \tau) = \int_{-\infty}^{\infty} f(t) \psi_{s,\tau}(t) dt$$

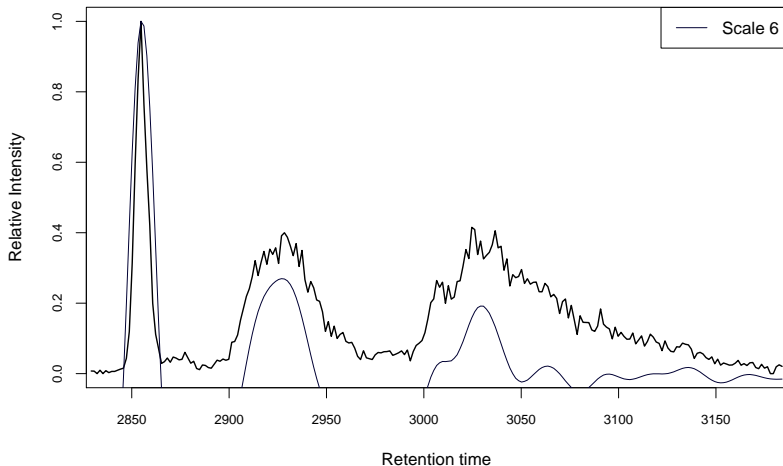
- smooth and enhance the data
- calculate similarity between the scaled wavelet and the data



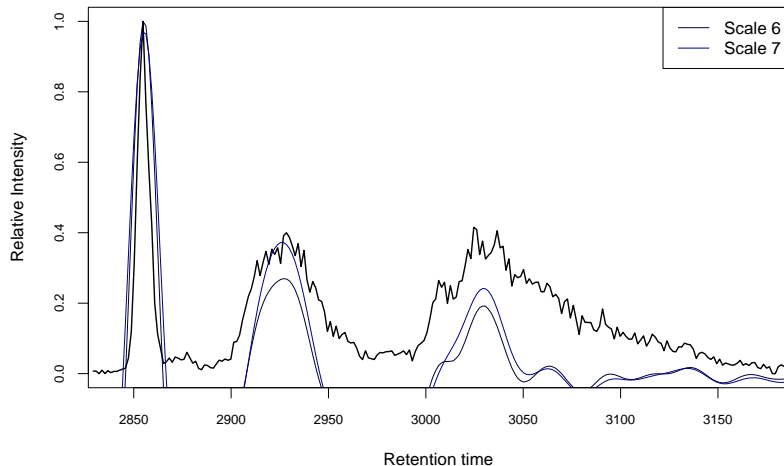
Step 2: Detection of chromatographic peaks



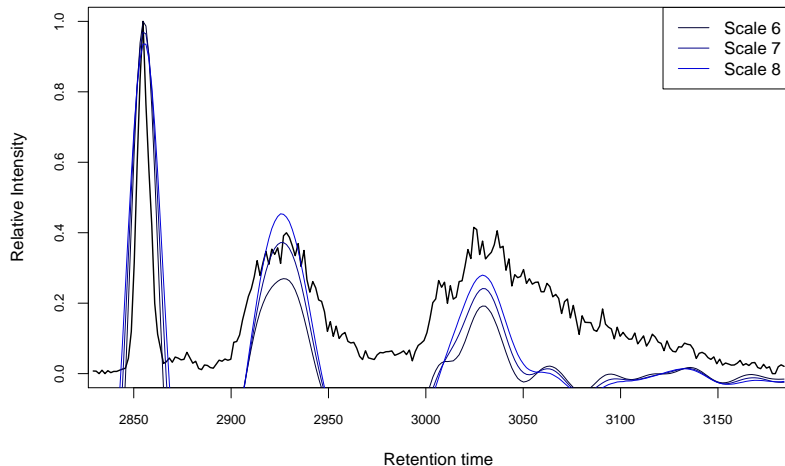
Step 2: Detection of chromatographic peaks



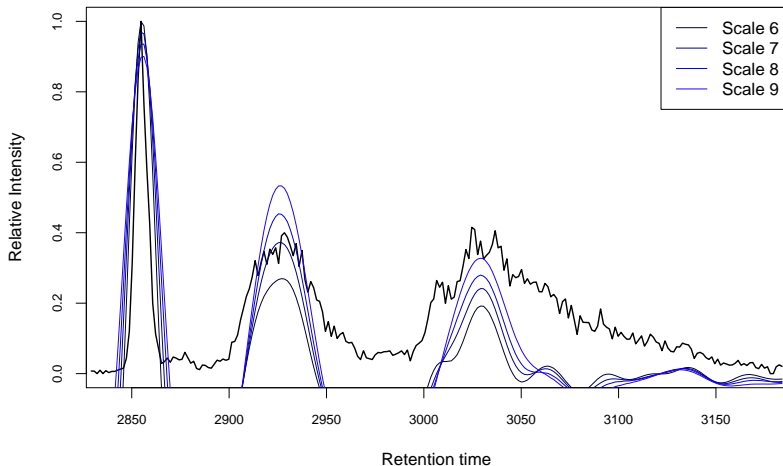
Step 2: Detection of chromatographic peaks



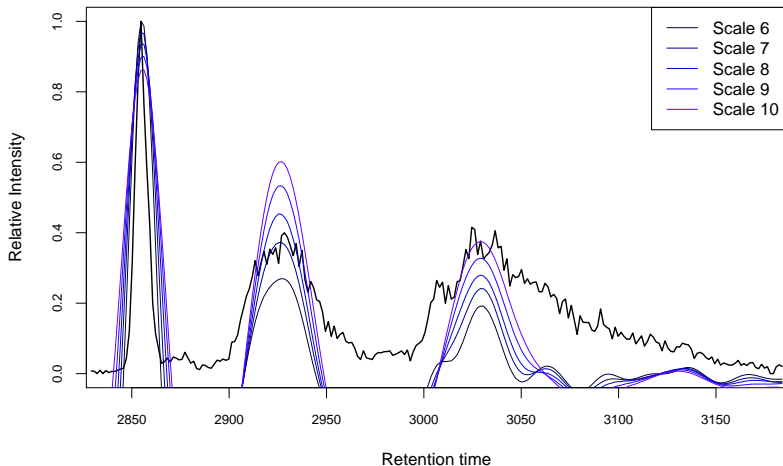
Step 2: Detection of chromatographic peaks



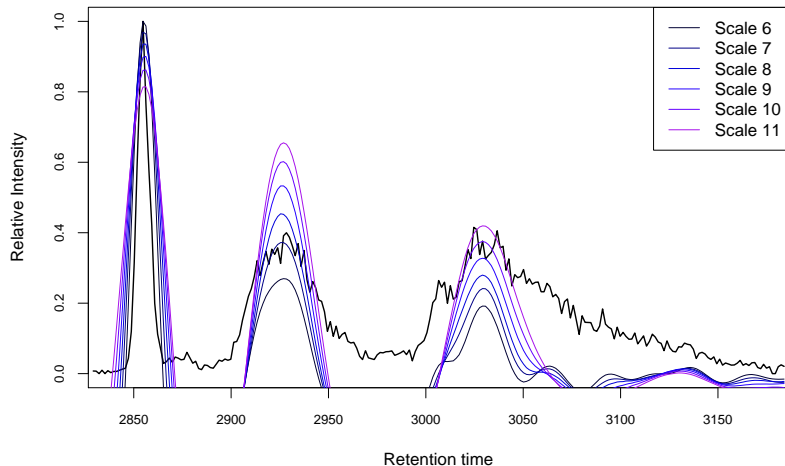
Step 2: Detection of chromatographic peaks



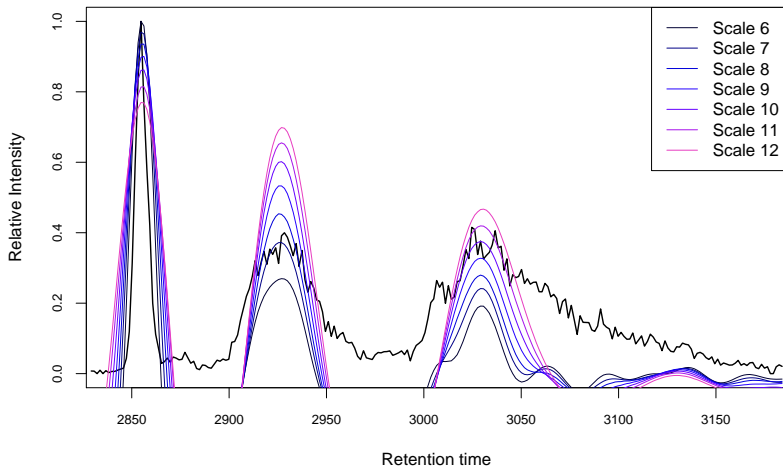
Step 2: Detection of chromatographic peaks



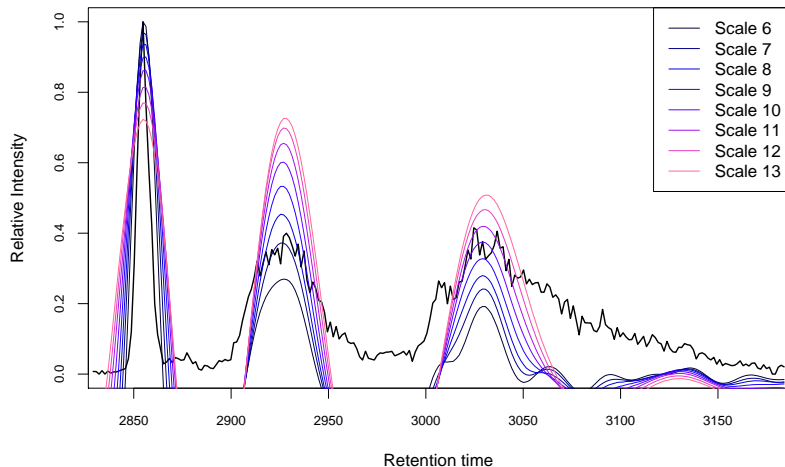
Step 2: Detection of chromatographic peaks



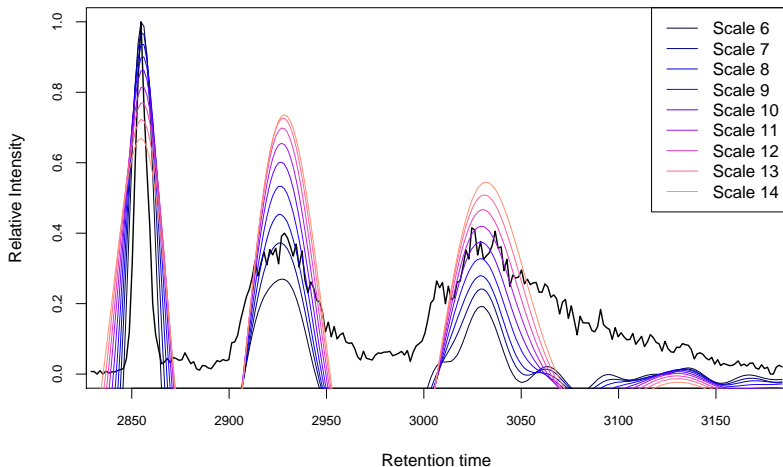
Step 2: Detection of chromatographic peaks



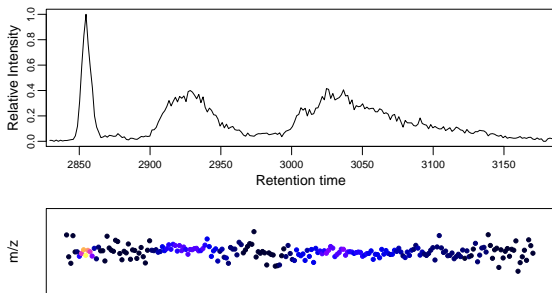
Step 2: Detection of chromatographic peaks



Step 2: Detection of chromatographic peaks

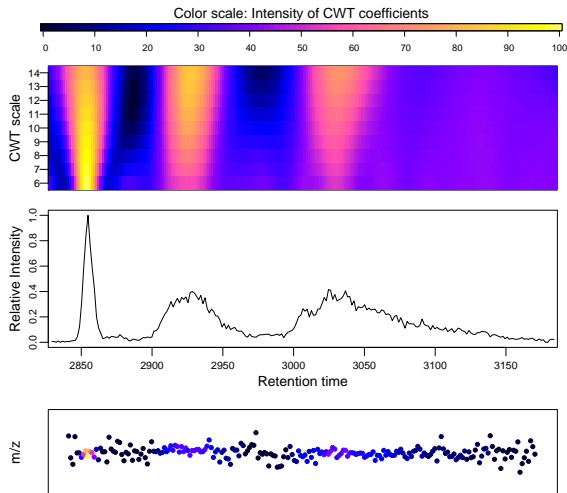


Step 2: Detection of chromatographic peaks



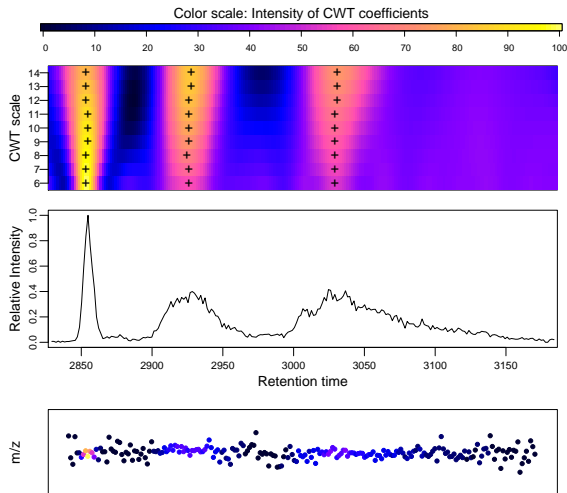
Step 2: Detection of chromatographic peaks

● Application of CWT



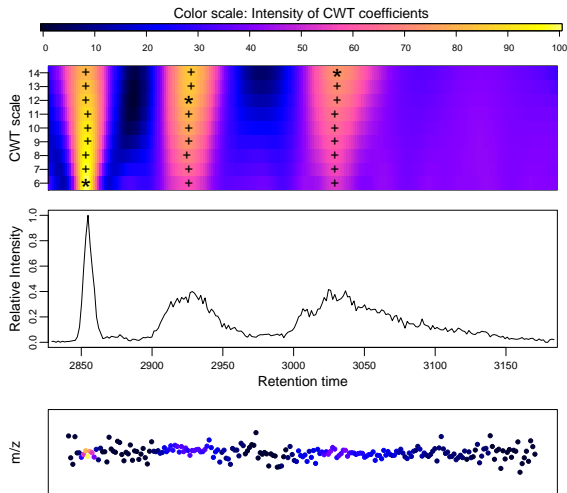
Step 2: Detection of chromatographic peaks

- Application of CWT
- Detect and connect local maxima (+)



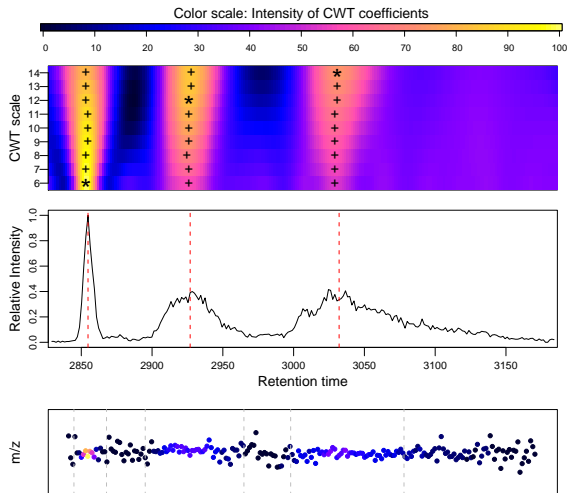
Step 2: Detection of chromatographic peaks

- Application of CWT
- Detect and connect local maxima (+)
- Identify optimal scale (*)



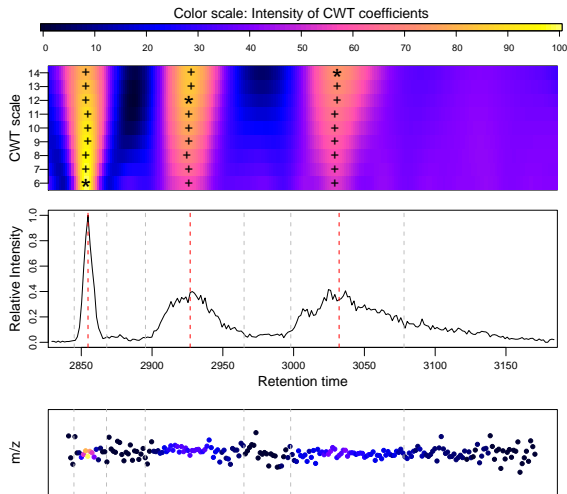
Step 2: Detection of chromatographic peaks

- Application of CWT
- Detect and connect local maxima (+)
- Identify optimal scale (*)
- Identify chromatographic center points



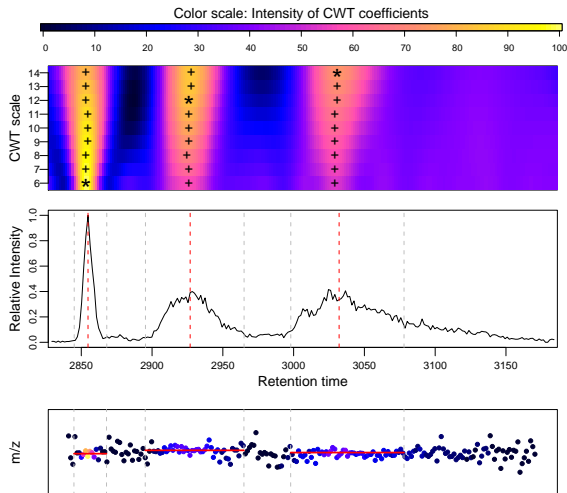
Step 2: Detection of chromatographic peaks

- Application of CWT
- Detect and connect local maxima (+)
- Identify optimal scale (*)
- Identify chromatographic center points
- Identify chromatographic boundaries



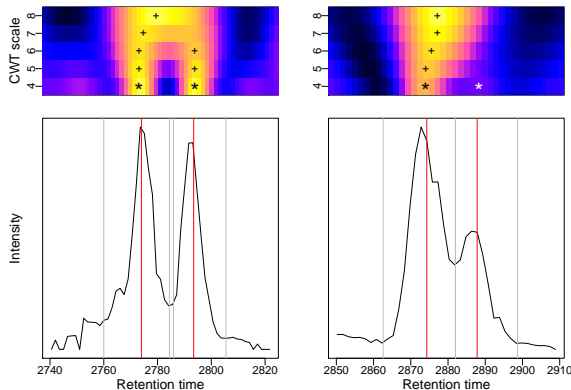
Step 2: Detection of chromatographic peaks

- Application of CWT
- Detect and connect local maxima (+)
- Identify optimal scale (*)
- Identify chromatographic center points
- Identify chromatographic boundaries
- Identify center points in m/z



Step 2: Detection of chromatographic peaks

Separation of chromatographically overlapping features



Parameter settings

- *ppm*
max. m/z deviation in consecutive scans
in ppm (parts per million), e.g. *ppm=30*
- *peakwidth*
Chromatographic peak width range (min,max) in seconds,
e.g. *peakwidth=c(10,60)*

Additional parameters see `?findPeaks.centWave`

Parameter settings

- UPLC, new QTOF
(e.g. Bruker MicroTOF-Q, Agilent QTOF 6210)

```
xs <- xcmsSet(method="centWave",ppm=30, peakwidth=c(5,20))
```

- HPLC, new QTOF
(e.g. Bruker MicroTOF-Q, Agilent QTOF 6210)

```
xs <- xcmsSet(method="centWave",ppm=30, peakwidth=c(10,60))
```

- HPLC, old TOF (e.g. ABI QStar)

```
xs <- xcmsSet(method="centWave",ppm=120, peakwidth=c(10,60))
```

centWave: Summary

- includes density based ROI-detection and chromatographic peak detection using CWT
- especially useful for high resolution massspec like LC/TOF, LC/FTICR, LC/Orbitrap
- spectra need to be in centroid mode

For more details on algorithm & evaluation, see

Tautenhahn R., Böttcher C., Neumann S.

Highly sensitive feature detection for high resolution LC/MS
BMC Bioinformatics, **2008** 9:504

Peak detection for single-spectrum non-chromatography MS data

- Peak detection for single FTICR, SELDI, MALDI spectra

```
xs <- xcmsSet(method="MSW")
```

- Calibration

```
xs <- calibrate(xs)
```

- Alignment

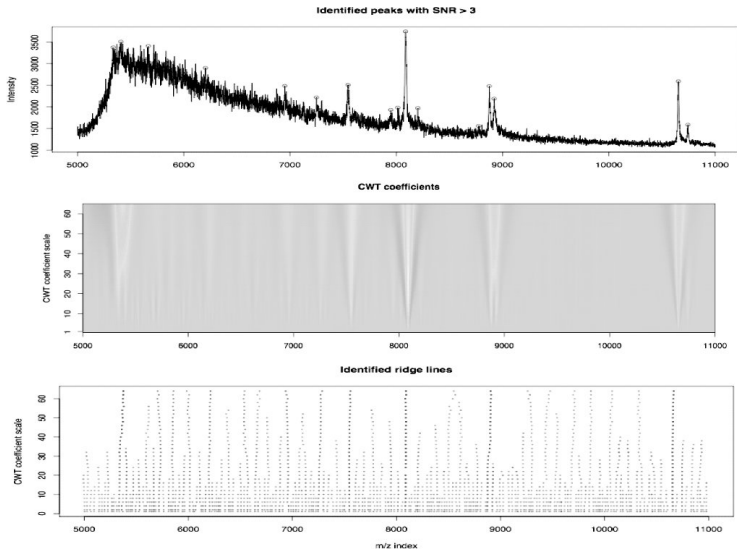
```
xsg <- group(xs, method="mzClust")
```

- diffreport

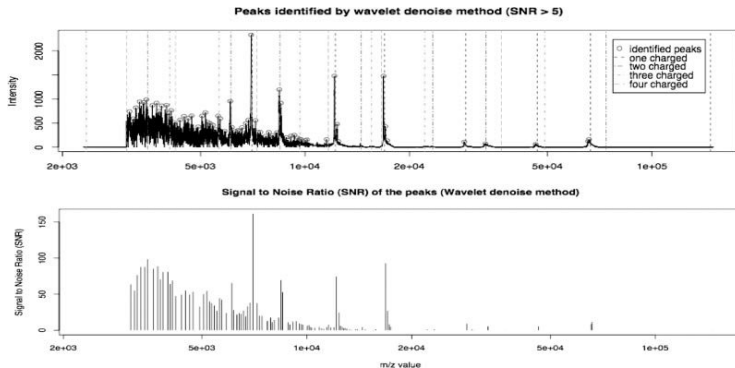
```
diffreport(xsg, ...)
```

including EIC like graphics!

MSW: Peak detection for single spectra



MSW: Peak detection for single spectra



Pan Du , Warren A. Kibbe , and Simon M. Lin

Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching

Bioinformatics, **2006** 22: 2059-2065.

```
xs <- xcmsSet(method="MSW")
```

```
?findPeaks.MSW
```


mzClust: Alignment of single spectra

- group peaks via high resolution alignment
 - avoid boundary problem of binning techniques
- use sliding window approach

Saira A. Kazmi, Samiran Ghosh, Dong-Guk Shin, Dennis W. Hill, and David F. Grant

Alignment of high resolution mass spectra: development of a heuristic approach for metabolomics

Metabolomics, **2006** Vol. 2, No. 2, 75-83

```
xsg <- group(xs, method="mzClust")  
?group.mzClust
```

Example: FTICR spectra (Bruker Apex III 70e)

```
library(xcms)

xs <- xcmsSet(method="MSW",snthr=3, SNR.method='data.mean',
              winSize.noise=500, peakThr=80000, amp.Th=0.005,
              scales=c(1,7))

xsg <- group(xs,method="mzClust", mzppm=5)

xsf <- fillPeaks(xsg,method="MSW")

diffreport(xsf,filebase="xcms-result", eicmax=1000)
```

Multicore/Cluster Usage with MPI

1 Install MPI

- Ubuntu, OpenMPI:

```
# sudo aptitude install libopenmpi-dev
# export MPI_ROOT=/usr/lib/openmpi R:
install.packages(c("Rmpi"),dep=T)
```

- Windows, DeinoMPI: <http://mpi.deino.net/>
<http://www.stats.uwo.ca/faculty/yu/Rmpi/deinomp.html>

2 Feel the infinite power

```
xs <- xcmsSet(..., nSlaves=4)
xs <- xcmsSet(..., nSlaves=16)
```

?xcmsSet

?xcmsPapply

Coming soon: Simple multicore support without MPI by using the R-package *multicore*.