Special note:

一种新的校准算法，称为 SFA-MS(基于质谱信息的子窗口因子分析法)，被应用于非靶向代谢组数据中。通过与其他方法的比较，该算法更加准确地校准谱峰，且不改变峰形，特别适用于重叠峰簇和信噪比低的谱峰。

算法操作与步骤

1. 收集不同批次的数据，并保存为“mat”格式。
2. 应用“load”导入数据。
3. 建立参考色谱和校准图谱。设定参数，不同参数的意义如下:

parameters	meaning
phi=20;	% SNR for peak detection
l=6;	% level for DWT , Discrete Wavelet Transform (DWT)
lambda=10e4;	% Smoothness for fitted baseline
valve=0.9;	% Eigenvalue evaluation criteria
max_shift=20;	% Maximum distance allowed to move
AMWFA_wz=3;	% The size of the mass spectrum window.

4. 采用 airPLS 算法进行背景扣除

```
%baseline correction
[sa,ssa]=airPLS(Data260.TIC, lambda,2,0.05);
[xa,xxa]=airPLS(Data247.TIC, lambda,2,0.05);
```

5. SFA-MS 操作是自动执行的，其中 Haar CWT 和 FFT 交叉相关用于峰值检测和候选飘移点数计算。当然，所有的色谱图都要预先进行基线校正和平滑处理。对应的函数(在函数“SFAMS.m”中)是:

```
% smoothing are executed beforehand for all chromatograms
[xs,xln]=wavelet_denoising(xtic,l,6,'soft');
[ss,sln]=wavelet_denoising(reftic,l,6,'soft');
%peak detection and candidate shifts calculation
peaks=peak_detection(xtic,3,xln,phi);
peaks=peak_clustering(xtic,peaks,1);
```

6. 应用子窗口因子分析法(SFA)对色谱相对应的质谱信息进行比对，确定最佳漂移点数。

   相应的函数是：

```
[R,h]=SWFA(X,Y,Noise_threshold,AMWFA_wz,valve)
[xn,peaks,CoCe,shiftvalue]=SFAMS(reftic,refmz,xtic,xmz,phi,max_shift,valve,AMWFA_wz,l)
```

7. 最佳漂移点数确定后，移动待校的峰区域，且在非峰区域进行线性插值，从而维持峰形不变，也使校正后的色谱图最大程度地维持原始状态。

Special note:

In the present work, an alignment algorithm is developed based on Subwindow Factor Analysis and Mass Spectral information (SFA-MS). Compared with other methods, this new algorithm aligns the peaks more accurately without changing their shapes, especially for the overlapping peak clusters and low S/N ratio peaks.

Algorithm operation steps and methods:

1. Collecting the sub-matrices of interest from different datasets, and “mat” format should be saved.

2. Using “load” to import data.

3. Establishing chromatograms (Reference; To be aligned)

   Setting parameters:

PARAMETERS	MEANING
PHI=20;	% SNR for peak detection
L=6;	% level for DWT , Discrete Wavelet Transform (DWT)
LAMBDA=10E4;	% Smoothness for fitted baseline
VALVE=0.9;	% Eigenvalue evaluation criteria
MAX_SHIFT=20;	% Maximum distance allowed to move
AMWFA_WZ=3;	% The size of the mass spectrum window.

4. airPLS algorithm is utilized to proceed background subtraction.

```
%baseline correction
[sa,ssa]=airPLS(Data260.TIC, lambda,2,0.05);
[xa,xxa]=airPLS(Data247.TIC, lambda,2,0.05);
```

5. SFA-MS operation is automatically executed.

   After the pretreatment of baseline correction and smoothing, Haar CWT and FFT cross correlation are used for peak detection and candidate shifts calculation, respectively.

   The corresponding function (‘SFAMS.m’) is:

```
% smoothing are executed beforehand for all chromatograms
[xs,xln]=wavelet_denoising(xtic,l,6,'soft');
```

[ss,sln]=wavelet_denoising(reftic,l,6,'soft');

%peak detection and candidate shifts calculation

peaks=peak_detection(xtic,3,xln,phi);

peaks=peak_clustering(xtic,peaks,1);

6. The MS fitting degree of the common components between the reference chromatogram and the raw chromatogram is determined by the Subwindow Factor Analysis (SFA). When the MS information between reference and raw peaks is identical, the corresponding moving points are the optimum shifts.

   The corresponding function is:

   [R,h]=SWFA(X,Y,Noise_threshold,AMWFA_wz,valve)

   [xn,peaks,CoCe,shiftvalue]=SFAMS(reftic,refmz,xtic,xmz,phi,max_shift,valve,AMWFA_wz,l)

7. All the peaks are moved through linear interpolation in the non-peak parts, so that the aligned chromatograms remain unchanged.