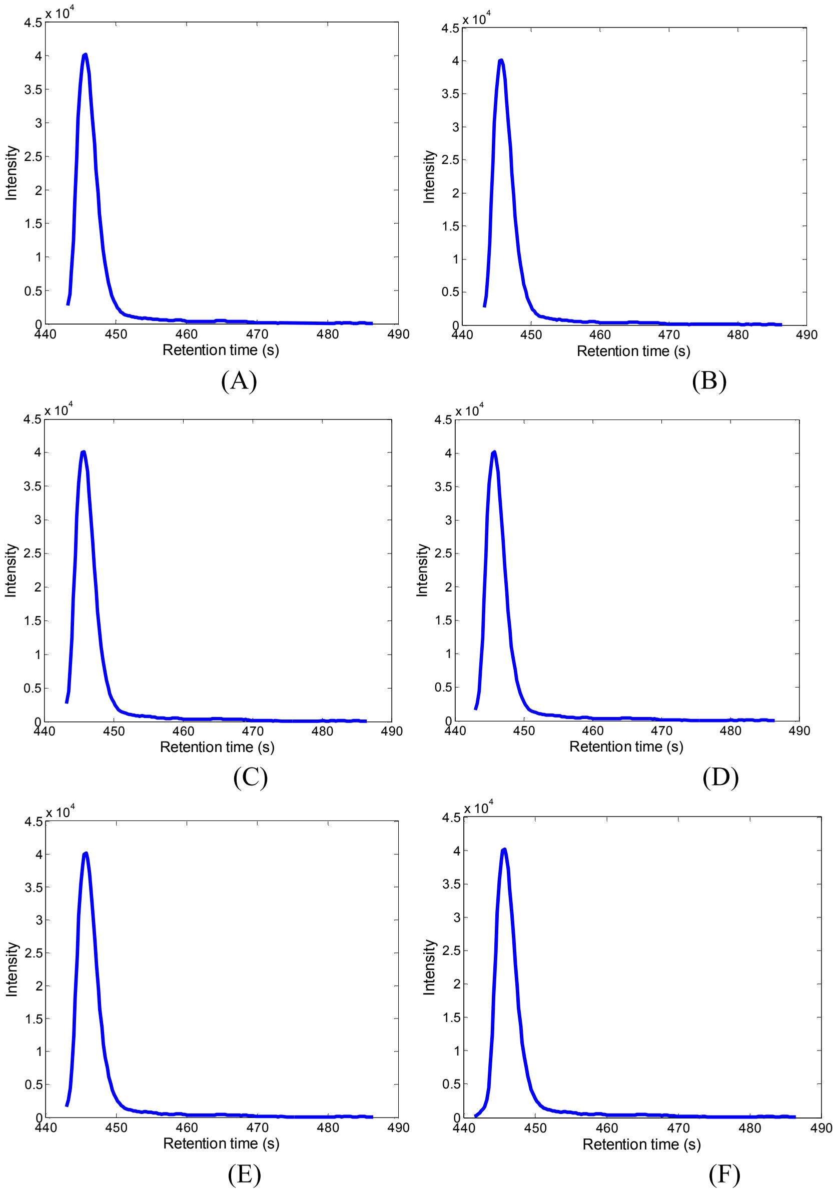
Data Dependent Peak Model-based Spectrum Deconvolution for Analysis of High Resolution LC-MS Data

Xiaoli Wei, Xue Shi, Seongho Kim, Jeffrey S. Patrick, Joe Binkley, Maiying Kong, Craig McClain, Xiang Zhang



(A)

(B)

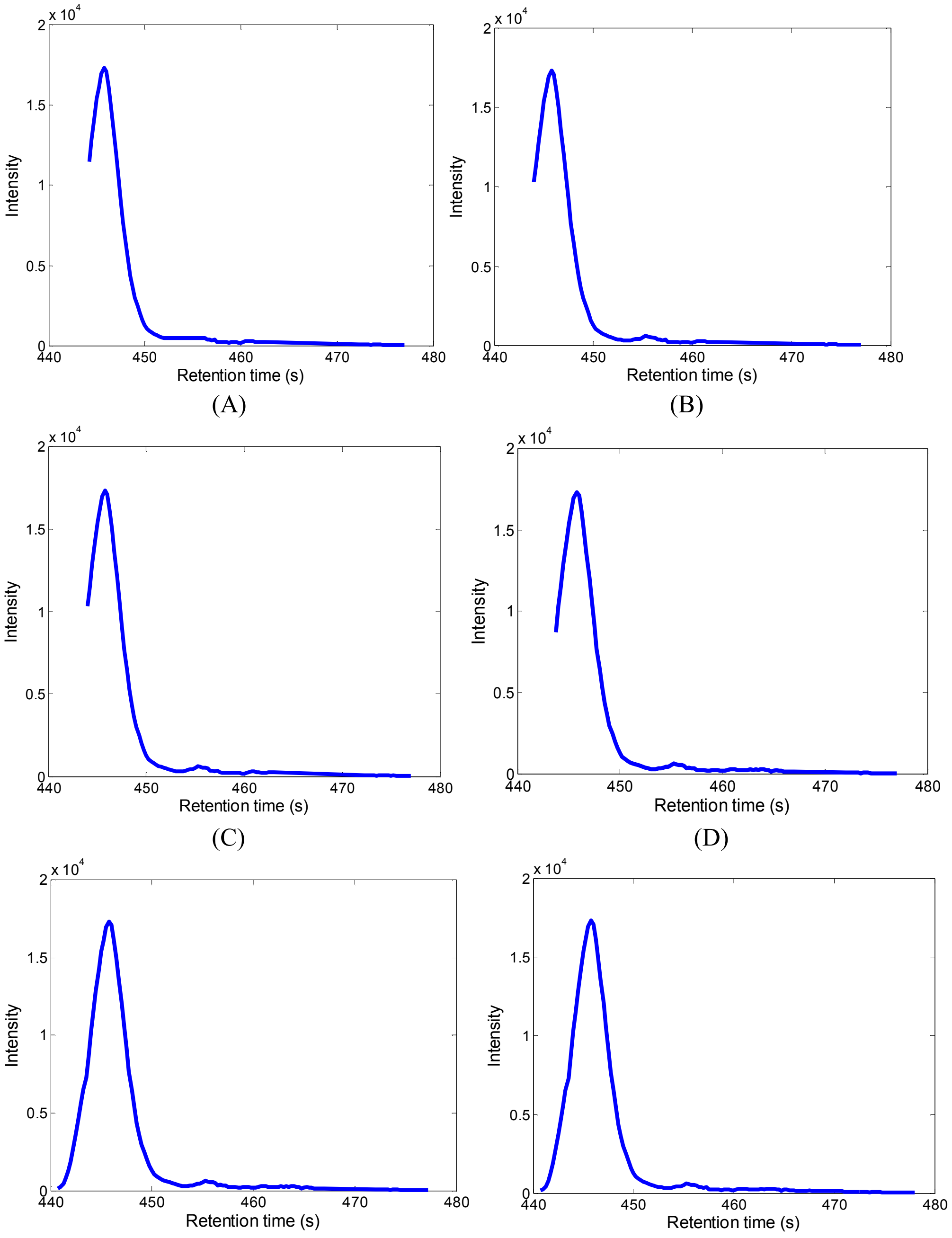
(C)

(D)

(E)

(F)

Figure S-1. An example of XIC with maximum peak height equals to 40128 generated by using different user defined *m/z* variation windows (A) $\Delta m/z \leq$ 4 ppm, (B) $\Delta m/z \leq$ 5 ppm, (C) $\Delta m/z \leq$ 6 ppm, (D) $\Delta m/z \leq$ 7 ppm, (E) $\Delta m/z \leq$ 8 ppm, and (F) DBSCAN approach.

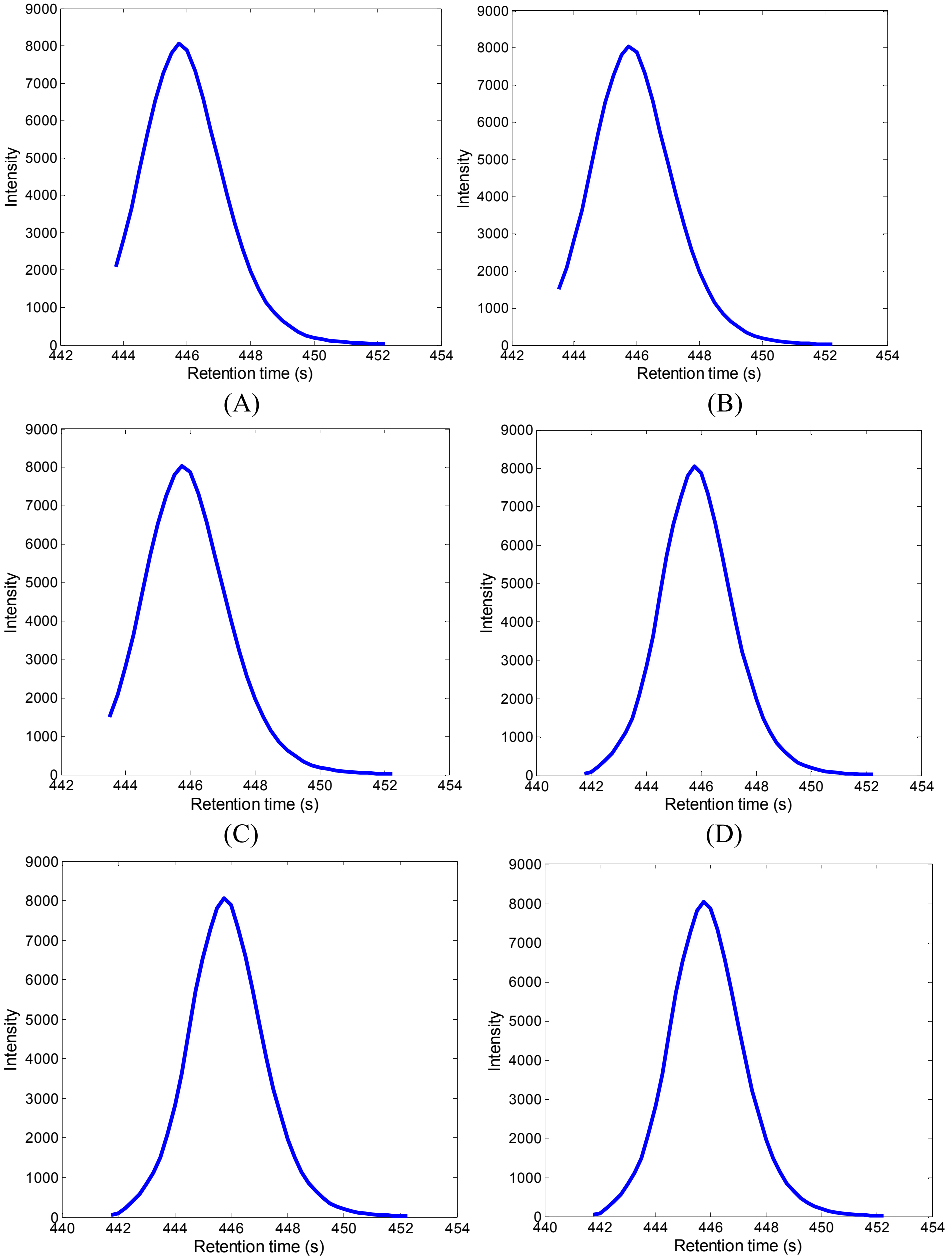


(A) (B)

(C) (D)

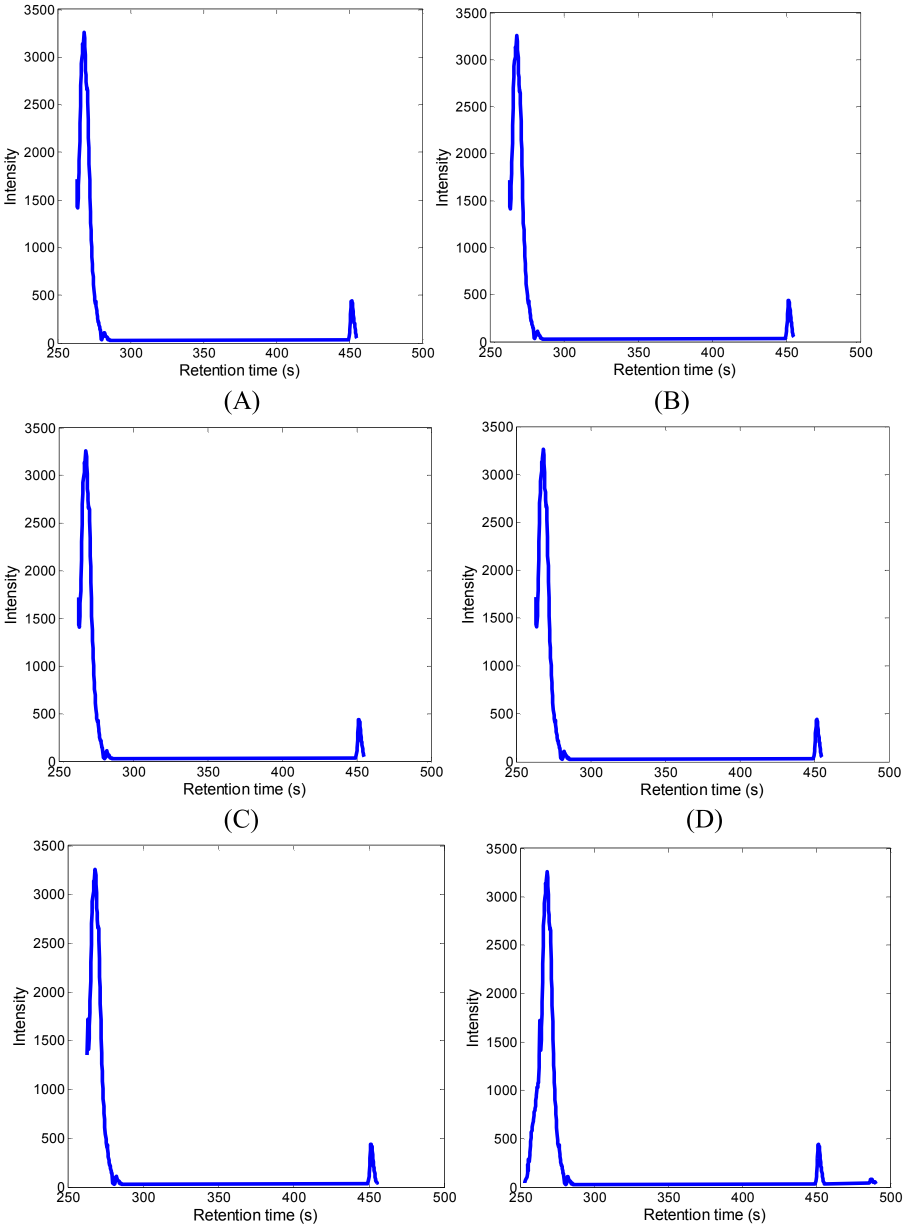
(E) (F)

Figure S-2. An example of XIC with maximum peak height equals to 17344 generated by using different user defined m/z variation windows (A) $\Delta m/z \leq 4$ ppm, (B) $\Delta m/z \leq 5$ ppm, (C) $\Delta m/z \leq 6$ ppm, (D) $\Delta m/z \leq 7$ ppm, (E) $\Delta m/z \leq 8$ ppm, and (F) by DBSCAN approach.



(A) (B) (C) (D) (E) (F)

Figure S-3. An example of XIC with maximum peak height equals to 8056 generated by using different user defined m/z variation windows (A) *Δm/z* ≤ 4 ppm, (B) *Δm/z* ≤ 5 ppm, (C) *Δm/z* ≤ 6 ppm, (D) *Δm/z* ≤ 7 ppm, (E) *Δm/z* ≤ 8 ppm, and (F) by DBSCAN approach.



(A) (B)

(C) (D)

(E) (F)

Figure S-4. An example of XIC with maximum peak height equals to 3260 generated by using different user defined m/z variation windows (A) $\Delta m/z \leq$ 4 ppm, (B) $\Delta m/z \leq$ 5 ppm, (C) $\Delta m/z \leq$ 6 ppm, (D) $\Delta m/z \leq$ 7 ppm, (E) $\Delta m/z \leq$ 8 ppm, and (F) by DBSCAN approach.

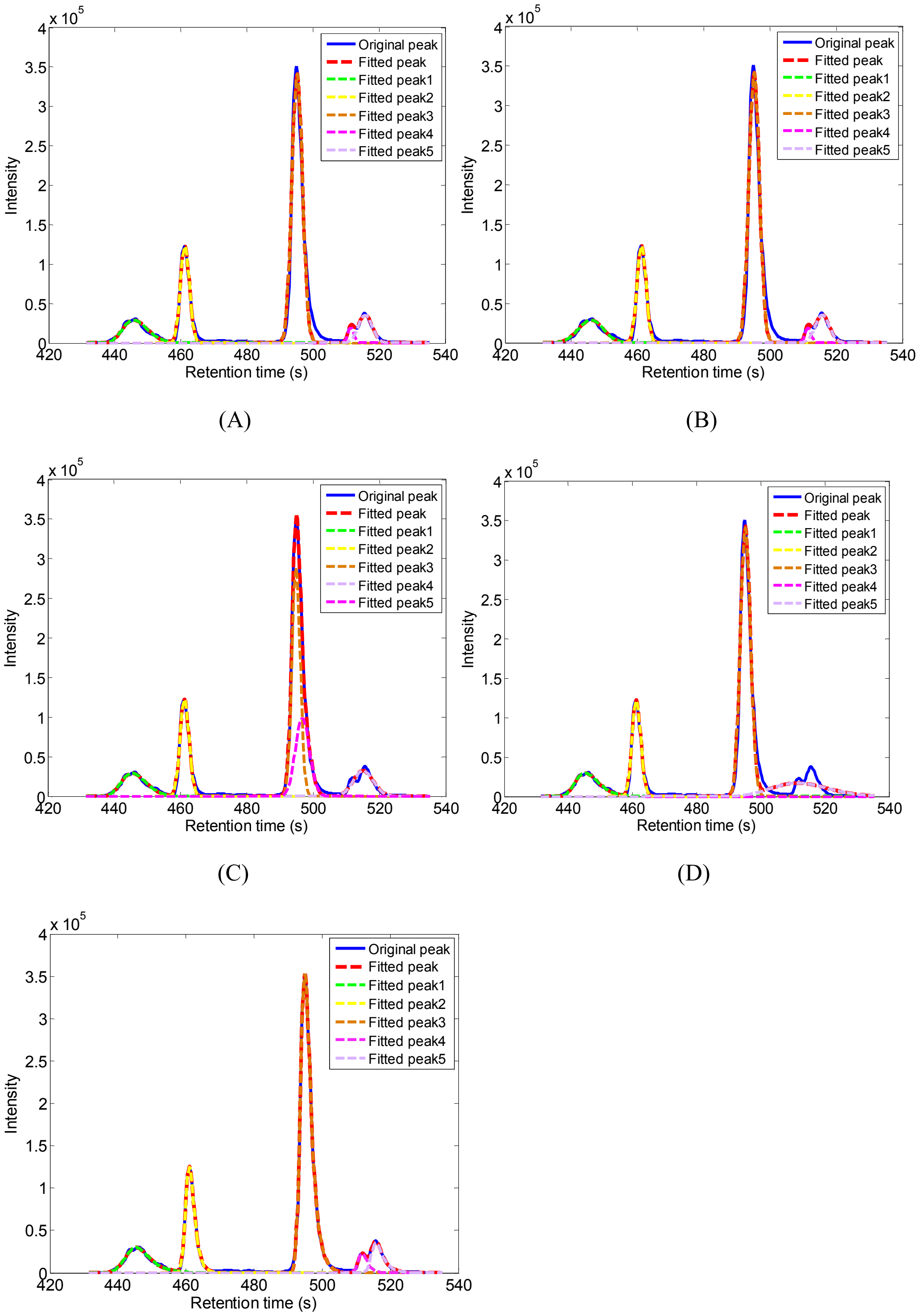
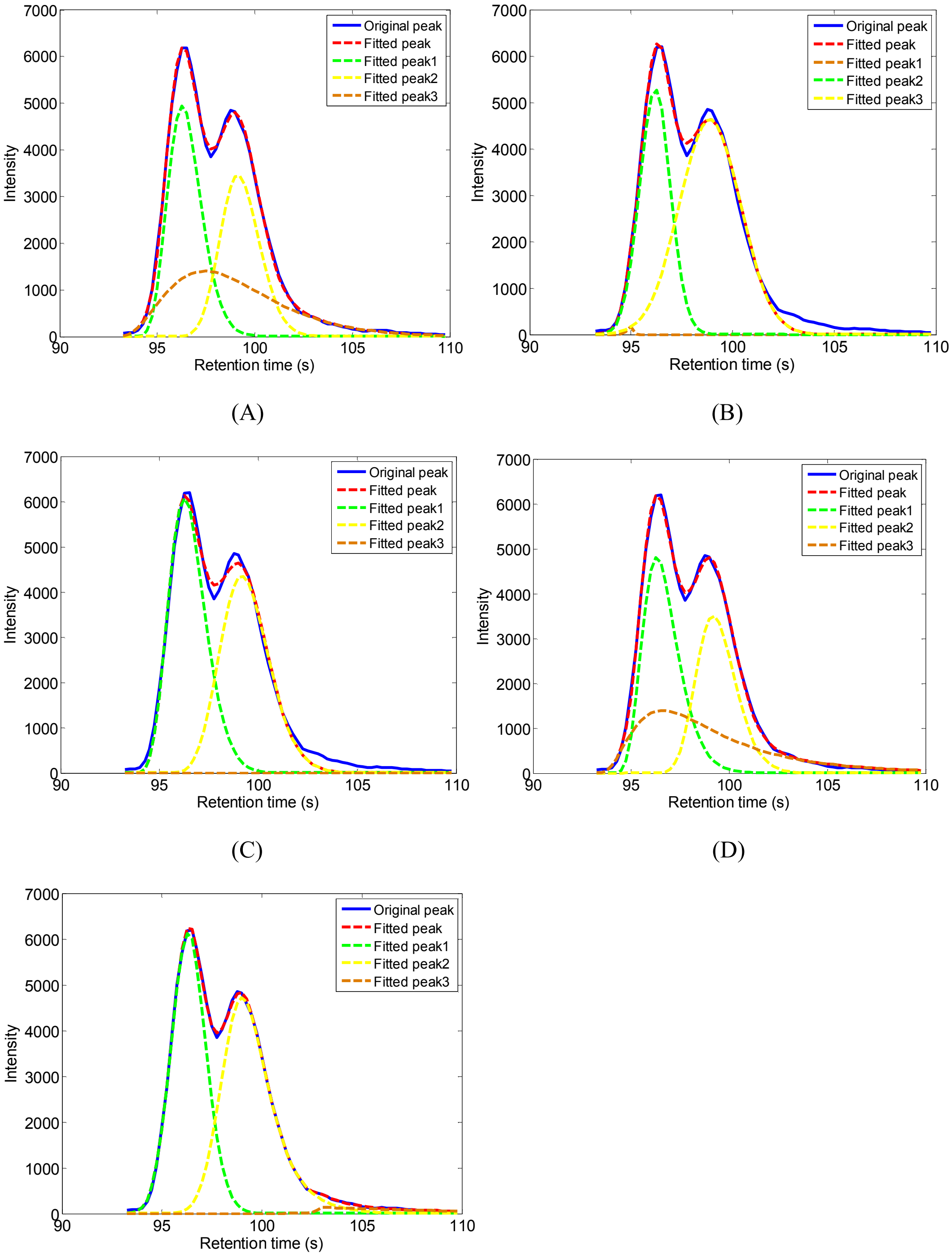


Figure S-5. A sample of effect of five chromatographic peak models in fitting a region of an XIC contains most abundant chromatographic peak. (A) is PMM model, (B) is GMM model, (C) is GaMM model, (D) is LNMM model, and (E) is EMGM model. The fitting scores of these five peak models in ascending order are LNMM (R^2 = 0.9815) < GMM (R^2 = 0.9820) < PMM (R^2 = 0.9844) < GaMM (R^2 = 0.9970) < EMGM (R^2 = 0.9992).



(A)

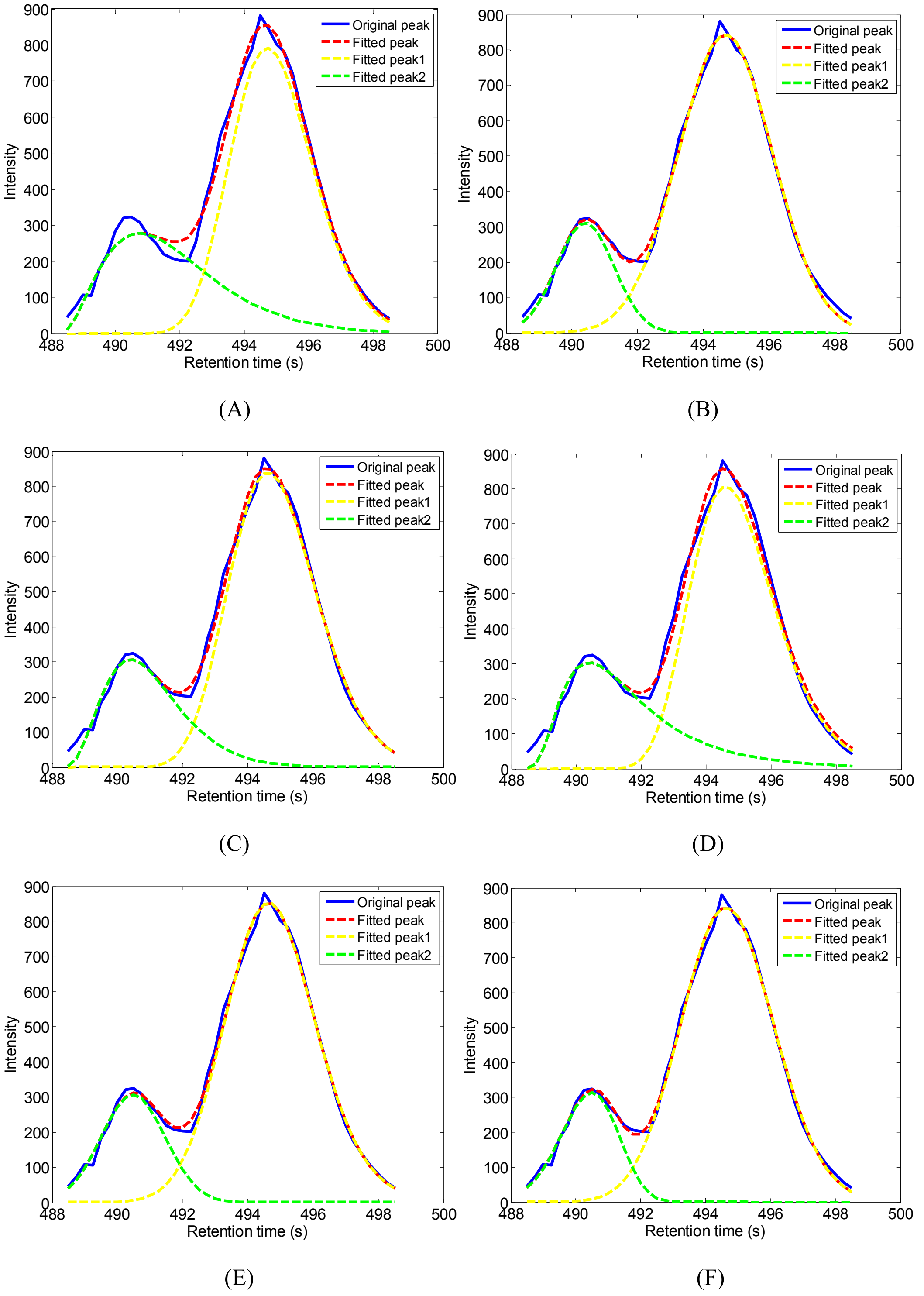
(B)

(C)

(D)

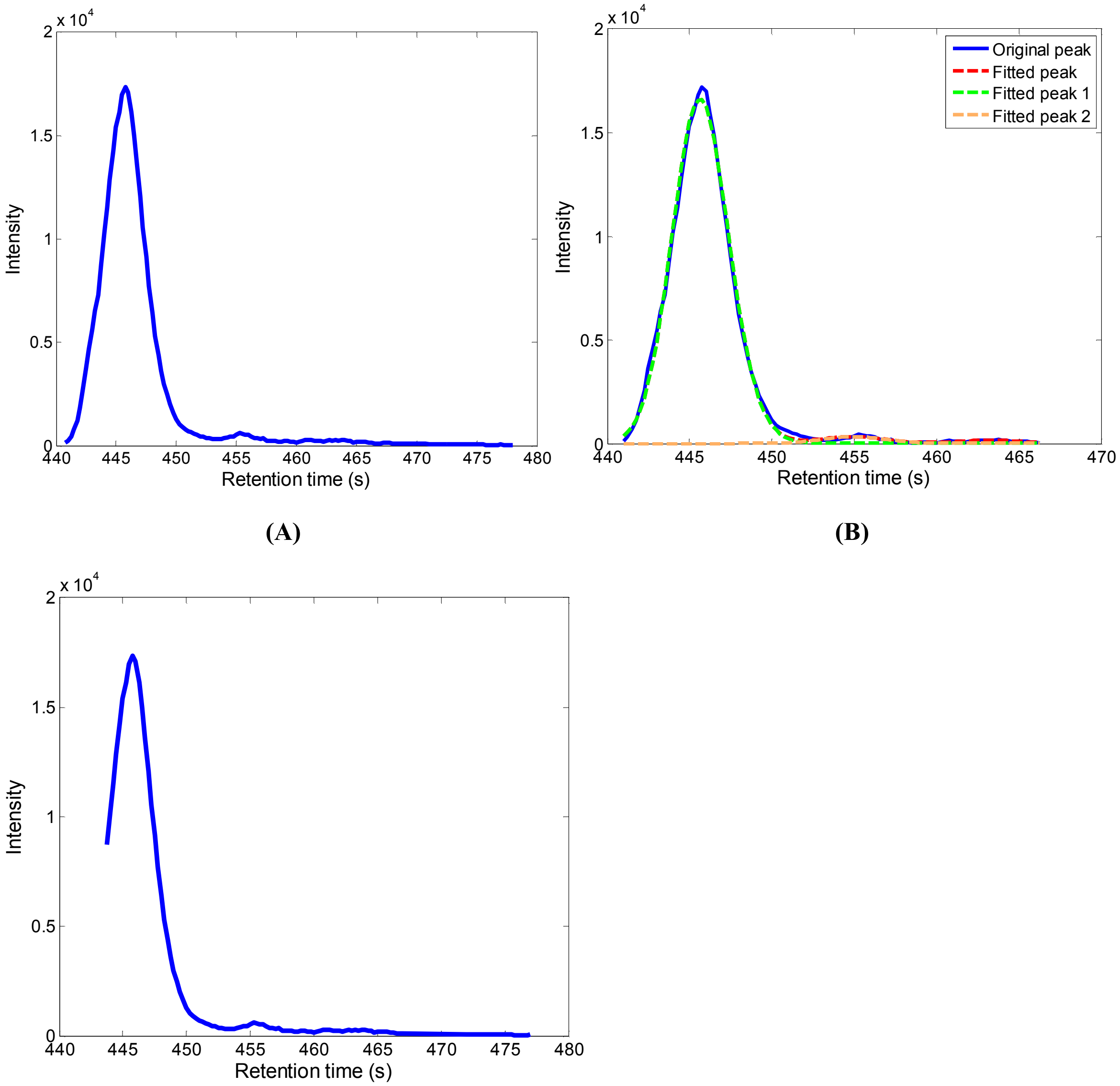
(E)

Figure S-6. A sample of effect of five chromatographic peak models in fitting a region of an XIC contains less abundant chromatographic peaks. (A) is PMM model, (B) is GMM model, (C) is GaMM model, (D) is LNMM model, and (E) is EMGM model. The fitting scores of these five peak models in ascending order are GaMM (R^2 = 0.9944) < GMM (R^2 = 0.9952) < LNMM (R^2 = 0.9991) < PMM (R^2 = 0.9992) < EMGM (R^2 = 0.9997).



(A) (B) (C) (D) (E) (F)

Figure S-7. A sample of effect of six chromatographic peak models in fitting a region of an XIC contains low abundant chromatographic peaks. (A) is PMM model, (B) is GMM model, (C) is GaMM model, (D) is LNMM model, (E) is EGH model, and (F) is EMGM model. The fitting scores of these six peak models in ascending order are PMM ($R^2 = 0.9857$) < LNMM ($R^2 = 0.9887$) < GaMM ($R^2 = 0.9927$) < GMM ($R^2 = 0.9950$) < EGHM ($R^2 = 0.9958$) < EMGM ($R^2 = 0.9965$).



(A)

(B)

(C)

Figure S-8. A sample of analysis difference of spectrum deconvolution by DBSCAN approach and *m/z* variation window approach. (A) is a XIC peak constructed by DBSCAN approach. (B) depicts the fitting results of the detected peaks. by the DBSCAN approach and the fitting parameter $R^2 = 0.9978$. (C) is the XIC of the same ion constructed by *m/z* variation window approach ($\Delta m/z \leq 7$ ppm).