BAYESIAN PEAK DETECTION FOR PRO-TOF MS MALDI DATA

Jianqiu Zhang[1], *Honghui Wang*[3], *Anthony Suffredini*[3], *Denise Gonzales*[3], *Elias Gonzalez*[1], *Yufei Huang*[1,2,4], *Xiaobo Zhou* [5]

1. Dept. of ECE, 2. Dept. of Bio. Eng., University of Texas at San Antonio, San Antonio, TX 78249
3. NIH Clinical Center, Bethesda, MD 20892
4. Greehey Children's Cancer Research Institute, University of Texas Health Science Center at San Antonio, TX 78229
5. Texas Methodist Hospital Research Institute, Houston, TX 77030
Email: Michelle.Zhang@utsa.edu; xzhou@tmhs.org

ABSTRACT

In this paper, a novel Bayesian peak detection algorithm is proposed for peptide peak detection in high resolution prOTOF™ MALDI Mass Spectrometry(MS) data. A nonlinear parametric model is proposed for modeling the peptide signals, chemical noise, and thermal noise. A metropolized Gibbs sampling algorithm is derived for Bayesian peak detection. The proposed algorithm is compared with a popular wavelet-based algorithm and the results show a significant improvement in performance on simulated data. The algorithm is finally tested on real MS MALDI data and the results agree with visual inspection very well.

Index Terms— Mass spectrometry, proteomics, MALDI, Peak detection, Bayesian methods.

I. INTRODUCTION

Mass spectrometers (MS) is an analytical machine used to measure the the mass-to-charge ratio (m/z) of ionized peptides. The measurements across the valid m/z range form mass spectrum of peptides. Mass spectrum can characterize proteins, thus making it an important tool in biomarker discovery. Different MS techniques measure mass spectra of different resolution and mass range. Please refer to [1] for an overview of different techniques and their application. One of the primary methods is Matrix-assisted laser desorption/ionization (MALDI) time-of-flight mass spectrometer (TOF-MS). Especially, prOTOF™ MALDI is capable of holding an accurate mass calibration over a wider mass range and a longer period of time compared to instruments using delayed ion extraction. It has higher resolution for peptides with small m/z value. For m/z < 4000 m/z, the resolution is higher than 1 Dalton, i.e., signals from peptides with 1 Dalton difference in mass can be distinguished.

In a mass spectrum, desired peptide signals are corrupted by unwanted chemical and background high frequency thermal noise. As a first step in protein characterization, signal peaks that corresponds to peptide fragments obtained from a target protein needs to be identified from the interfering noise. This step is referred to as "Peak Detection". This paper focus on peak detection of prOTOF^{TM} MALDI data over low m/z values.

One challenge in peak detection for proTOF MALDI data is to differentiate with confidence the chemical noise from the desired peptide signals, which commonly reside in the same frequency spectrum and form similar bell shaped peaks that are 1 Dalton apart. In most references on peak detection over low m/z range, a threshold is used for distinguishing chemical noise and peptide signals [2], [3]. However, since the height/area of chemical noise peaks are random variables in proTOF MALDI data, thresholding can never be satisfactory.

The second challenge is due to the rather narrow and low intensity peptide peaks. If filtering methods were to be employed to remove high frequency background noise, these peaks would be easily filtered out. As a result, some of the popular nonparametric filtering techniques such as the wavelet based approaches [2] would fail to retain these peaks.

Finally there is the challenge of data fusion. Usually several "replicates" of MALDI data from the same tissue sample are availabe. The task of utilizing data from different replicates for improved peak identification is referred to as "data fusion" here. There are many methods for "fusing" these data together. The simplest and common practice is to average these replicates before performing peak detection[2]. However, such a solution is far from optimal.

In this paper, we propose a nonlinear parametric model that accurately describes the chemical noise, peptide signal and thermal noise within one Dalton. Using such a model, height difference, spread and mean location can all be used for the differentiation of chemical noise and peptide signal. In contrast, methods that only utilize the height for chemical noise/peptide separation will be less effective.

Based on the proposed model, a Bayesian algorithm utilizing metropolized Gibbs sampling is proposed for peak detection. The proposed algorithm performs "Statistical Curve fitting" by estimating the *a posteriori* probabilities (APP) of peptide signals given the observed data. On one hand, the APPs provide a confidence measure about the inference results. Moreover, they are indispensable to the proposed Bayesian data fusion scheme.

The proposed algorithm is tested on both simulated and real dataset. The proposed algorithm is compared to a wavelet based algorithm which has been shown to perform very well for MALDI data processing on machines with lower resolution in the past [2]. For simulated data, it is shown that given the same true positive rates, the false positive rate is reduced dramatically. For the real proTOF data, the results also show better performance of our algorithm.

II. BAYESIAN PEAK DETECTION OF PROTOF DATA OVER LOW M/Z

II-A. Data Model

An example of prOTOP mass spectrum at low m/z range is shown in Figure 1. We can model the proTOF data over a one m/z interval as the superposition of three parts. The first part is generated by charged peptides hitting the receiver plate. Due to numerous random factors, peptides with the same molecular mass do not hit the receiver plate at the same time. Instead, the arrival time of the peptides follows a Gaussian distribution. The second part is the chemical

This work was supported by the National Science Foundation under Award CCF-0546345.

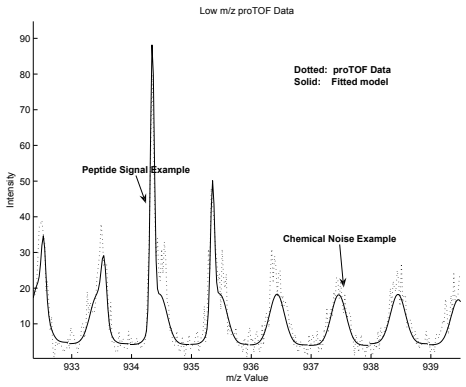

Fig. 1. Example of prOTOFTM MALDI data at Low m/z. figure

noise due to charged chemical impurities. The impurities form similar curves as that of peptides, but the curves are usually wider and with less intensity. Note from Figure 1, both chemical and signal assume bell shape, which inspires us to model them with a Gaussian-like function. The third part is white thermal noise, which can be assumed as white 2D Gaussian. Now, let $y_s[m]$ be the signal generated by ionized peptides observed at m/z value, which is modeled by a Gaussian-like function as

$$y_s[m] = \beta_s exp(-\rho_s^2(m-\mu_s)^2), \tag{1}$$

where β_s models the height of the signal, the inverse of ρ_s models the spread of the signal, and μ_s models the location of the center of the signal on the m/z axis. Chemical noise $y_c[m]$ at m is modeled similarly by a Gaussian-like function as

$$y_c[m] = \beta_c exp(-\rho_c^2(m-\mu_c)^2) + d, \tag{2}$$

where β_c, ρ_c, and μ_c all have the similar meaning as their counterparts in the signal model, and d models the DC element of the chemical noise.

For each data period of 1.0005 m/z, the data is modeled as

$$y[m] = \lambda y_s[m] + y_c[m] + \epsilon[m], \tag{3}$$

where $\lambda \in \{0, 1\}$ is an indicator random variable, which is one if the signal exists and zero otherwise, and $\epsilon[m]$ is additive thermal noise. It is modeled as i.i.d. zero mean Gaussian with standard deviation σ. Note $\boldsymbol{\theta} = [\lambda, \beta_s, \rho_s, \mu_s, \beta_c, \rho_c, \mu_c, d, \sigma]^\top$ is the vector of all the unknowns parameters to be estimated.

The proposed parametric model can fit the data pretty well. As a result, parametric curve fitting can be applied for thermal and chemical removal. Although it is generally believed that non-parametric methods (such as wavelet smoothing) are better suited for thermal noise removal, our proposed parametric model is superior since chemical noise in prOTOF MALDI data assumes distinct structure, whose information (height, mean location, width) can be used for better differentiation between signal and noise. In contrast, non-parametric methods only utilize the height of a signal for chemical noise removal.

II-B. Bayesian Peak Detection

II-B.1. Bayesian Detection Criterion

The peak detection can be done independently on each data period of 1.0005 m/z. The goal of peak detection is to determine if there is a signal in a data period. Given $\mathbf{Y}$, a set of M data samples from on data period, the objective of Bayesian detection is to obtain the APPs $p(\lambda|\mathbf{Y})$ and then decision can be made according to the Maximum *a posteriori* criterion using the APPs. However, the APPs requires high dimensional integration of the joint posterior distribution $p(\boldsymbol{\theta}|\mathbf{Y})$ all the parameters except λ. In addition, the marginal posterior distribution of signal model parameters is required to estimate the shape of signal. Given the highly nonlinear nature of data model 3, none of the desired posterior distributions can be obtained analytically. We resort to a Markov Chain Monte Carlo(MCMC) sampling solution known as metropolized Gibbs sampling.

II-C. Metropolized Gibbs sampling solution

The Metropolized Gibbs (MG) sampling is applied to generate samples from the joint posterior distribution. In the following, we describe the sampling procedure in detail. The joint posterior distribution can be expressed as $p(\boldsymbol{\theta}|\mathbf{Y}) \propto p(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$, where $p(\mathbf{Y}|\theta)$ is the likelihood function and $p(\theta)$ is the prior distribution of the parameters. Since the number of parameters is large, it is difficult to sample all nine parameters at once. Instead we employ a Gibbs sampling to obtain samples from the full conditional posterior densities. However, direct sampling from the full conditional distributions is still infeasible for this problem. To circumvent the problem, a Metropolis-Hastings sampling is then introduced. The resulting scheme is known as the metropolized Gibbs (MG) sampling [4].The general results of the derived MG sampling will be discussed in the following and the details of the derivations are omitted due to page limitations.

The MG sampling is an iterative algorithm. At iteration t, the sample of the noise variance is obtained from the conditional distribution

$$p(\sigma^2|\mathbf{Y}, \boldsymbol{\theta}_{-\sigma^2}) \propto IG(M/2, |\mathbf{Y} - \lambda\mathbf{Y}_s - \mathbf{Y}_c|^2)/2) \tag{4}$$

where $\boldsymbol{\theta}_{-\sigma^2}$ represent the previous samples of all the parameters except σ^2, $\mathbf{Y}_s = \beta_s exp(-(\rho_s)^2(m-\mu_s)^2)$ is estimated signal computed based on previous samples of β_s, ρ_s and μ_s. Similarly, $\mathbf{Y}_c$ is the estimated chemical noise based on previous samples, and $IG(\cdot)$ is the Inverse Gamma distribution. In this case, given $\boldsymbol{\theta}_{-\sigma^2}$, the posterior density on σ^2 has an analytical form and samples can be taken directly from the Inverse Gamma distribution.

Next, samples for $\boldsymbol{\theta}_1 = [\beta_s \beta_c d]^\top$ are drawn from $p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_{-1}, \mathbf{Y})$. These three parameters are special in the sense that given other parameters $\boldsymbol{\theta}_{-\theta_1}$, the observations $\mathbf{Y}$ is linear in these parameters, which implies analytical expressions for the full conditional distribution of $\boldsymbol{\theta}_1$. Now, let $\mathbf{h}_s = exp(-(\rho_s)^2(m-\mu_s)^2)$, and $\mathbf{h}_c = exp(-(\rho_c)^2(m-\mu_c)^2)$. The observations $\mathbf{Y}$ can be then expressed as $\mathbf{Y} = \mathbf{H}\boldsymbol{\theta}_1 + \boldsymbol{\epsilon}$, where $\mathbf{H} = [\mathbf{h}_s\ \mathbf{h}_c\ \mathbf{1}_{M\times 1}]$ is an $M \times 3$ matrix obtained by concatenating three column vectors $\mathbf{h}_s$, $\mathbf{h}_c$ and $\mathbf{1}_{M\times 1}$ together, $\mathbf{1}_{M\times 1}$ is a vector all one elements, and it is evident that the posterior density $p(\boldsymbol{\theta}_1|\boldsymbol{\theta}\_\boldsymbol{\theta}_1, \mathbf{Y})$ is Gaussian, whose mean and covariance matrix are given by

$$\boldsymbol{\mu}_{\boldsymbol{\theta}_1|\mathbf{Y},\boldsymbol{\theta}\_\boldsymbol{\theta}_1} = \mu_{\boldsymbol{\theta}_1} + \Sigma_{\boldsymbol{\theta}_1}\mathbf{H}^\top(\mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{\theta}_1}\mathbf{H}^\top + \sigma^2\mathbf{I}_{M\times 1})^{-1}(\mathbf{Y} - \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{\theta}_1}) \tag{5}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}_1|\mathbf{Y}_i,\boldsymbol{\theta}\_\boldsymbol{\theta}_1} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}_1} - \boldsymbol{\Sigma}_{\boldsymbol{\theta}_1}\mathbf{H}^\top(\mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{\theta}_1}\mathbf{H}^\top + \sigma^2\mathbf{I}_{M\times 1})^{-1}\mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{\theta}_1}. \tag{6}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_1}$ and $\boldsymbol{\mu}_{\boldsymbol{\theta}_1}$ are the prior mean and covariance matrix of $\boldsymbol{\theta}_1$. The prior knowledge can be derived by checking the estimated heights of the first n chemical noise peaks. In this proposed scheme, the prior knowledge is updated after processing every n data periods.

The next group of the unknowns to be sampled are the mean and spread of the chemical noise peak $\boldsymbol{\theta}_2 = [\rho_c \mu_c]^{\mathsf{T}}$. It can be determined that the conditional posterior density of $\boldsymbol{\theta}_2$ can be written as

$$\begin{aligned} p(\boldsymbol{\theta}_2|\mathbf{Y}_i, \boldsymbol{\theta}_{\_\boldsymbol{\theta}_2}) &\propto (\sigma^2)^{-M/2} e^{(-1/(2\sigma^2)|\mathbf{Y}-\lambda\mathbf{Y}_s-\mathbf{Y}_c|^2)} p(\theta_2) \\ &= g(\theta_2) \end{aligned} \tag{7}$$

where $\mathbf{Y}^{cd} = \mathbf{Y} - \lambda\mathbf{Y}_s - d\mathbf{I}_{M\times 1}$ is the estimated chemical noise peak given $\theta_{-\theta_2}$, $p(\boldsymbol{\theta}_2)$ is the prior distribution of $\boldsymbol{\theta}_2$, which is assumed to be a two dimensional uniform distribution. The lower limits of ρ_c and μ_c are set to be zero as these parameters can only be positive. The upper limit on ρ_c is set to be 7 since $6/(\sqrt{2}\rho_c)$ roughly represents the spread of the chemical noise (99% of the area under chemical noise peak). It is observed that the spread of the chemical noise is never smaller than 0.6 Dalton. The upper limit on the center of the chemical noise is set to be 1 above the starting m/z value since at low m/z values, the resolution is greater than one dalton and the center of the peak can never deviate from the starting m/z by one dalton. $g(\boldsymbol{\theta}_2$ denotes the conditional posterior density function. The conditional posterior density of $\boldsymbol{\theta}_2$ in (7) can not be sampled directly. Instead, the Metropolis Hastings (MH) algorithm is applied. In MH sampling schemes, at the tth iteration, we first obtain a new sample θ_2^* from a proposal density function $q(\theta_2)$, then the sample's probability with respect to the proposal density and the targeting density $g(\theta_2)$ are calculated. The probability of accepting the sample is calculated by:

$$r = \frac{g(\boldsymbol{\theta}_2^*)q(\boldsymbol{\theta}_2^{t-1})}{g(\boldsymbol{\theta}_2^{t-1})q(\boldsymbol{\theta}_2^*)} \tag{8}$$

if r is less than one. If r is greater than 1 the sample is always accepted. It can be shown that after an initial "burn-in" period, samples accepted in this process will converge to the true distribution $g(\boldsymbol{\theta}_2)$. The proposal density function is set to be a two dimensional Gaussian distribution. The mean of the proposal density is set to be the value of the last set of samples θ_2^{t-1}, and the covariance matrix is updated every 1000 steps by looking at the sampled covariance on Σ_{θ_2} during previous 1000 steps.

Another group of parameters to be sampled is the mean and the spread of the "signal" peak, $\boldsymbol{\theta}_3 = [\rho_s \mu_s]^{\mathsf{T}}$. The derivation process is essentially the same as that of $g(\boldsymbol{\theta}_2)$.

Lastly, the indicator λ is sampled from the distribution

$$\begin{aligned} & p(\lambda|\mathbf{Y}, \boldsymbol{\theta}_{-\lambda}) \propto p(\mathbf{Y}|\boldsymbol{\theta})p(\lambda) \\ \propto\ & (\sigma^2)^{-M/2} e^{(-1/(2\sigma^2)|\mathbf{Y}-\lambda\mathbf{Y}_s-\mathbf{Y}_c|^2)} p(\lambda) \end{aligned} \tag{9}$$

From this equation, the log-APP ratio(LAR) on λ can be calculated as

$$\begin{aligned} LAR_\lambda &= \ln \frac{p(\lambda=1|\mathbf{Y}, \boldsymbol{\theta}_{-\lambda})}{p(\lambda=0|\mathbf{Y}, \boldsymbol{\theta}_{-\lambda})} \\ &= -\frac{1}{2\sigma^2}(|\mathbf{Y}-\mathbf{Y}_s-\mathbf{Y}_c|^2 - |\mathbf{Y}-\mathbf{Y}_c|^2) \\ &\quad + \ln \frac{p(\lambda=1)}{p(\lambda=0)} \end{aligned} \tag{10}$$

Once the LAR of λ is obtained, then the conditional distribution on λ can be derived as $p(\lambda = 1|\mathbf{Y}, \boldsymbol{\theta}_{-\lambda}) = \frac{1}{1+e^{-LAR_\lambda}}$ and $p(\lambda = 0|\mathbf{Y}, \boldsymbol{\theta}_{-\lambda}) = 1 - p(\lambda = 1|\mathbf{Y}, \boldsymbol{\theta}_{-\lambda})$.

These sampling steps will be repeated and samples taken from these conditional probabilities will converge to the joint posterior distribution of these parameters. In addition, the marginal distribution can be estimated by the samples easily. For instance, the APP $p(\lambda|\mathbf{Y})$ can calculated as

$$p(\lambda|\mathbf{Y}) \approx \sum_{t=1}^{N} \delta(\lambda - \lambda^{(t)})/N \tag{11}$$

where N represent the number of samples collected after "burn-in", or convergence and $\lambda^{(t)}$ denotes the tth sample. A peak can be detected if $p(\lambda = 1|\mathbf{Y})$ is greater than a certain threshold depending on the false alarm rate.

II-C.1. Bayesian Data Fusion

To integrate multiple data replicates, the proposed MG sampling algorithm is first applied to one replicate. At the end of this step, the posterior probability of model parameters $p(\theta|\mathbf{Y}_1)$ will be available based on the first replicate $\mathbf{Y}_1$. When processing the second replicate, we wish to obtain the posterior probability

$$\begin{aligned} p(\theta|\mathbf{Y}_1, \mathbf{Y}_2) &\propto p(\mathbf{Y}_2|\theta, \mathbf{Y}_1)p(\theta|\mathbf{Y}_1) \\ &\propto p(\mathbf{Y}_2|\theta)p(\theta|\mathbf{Y}_1) \end{aligned} \tag{12}$$

where (12) follows since given model parameters, $\mathbf{Y}_1$ and $\mathbf{Y}_2$ become independent. Inspecting (12), $p(\theta|\mathbf{Y}_1)$ can be viewed as a prior density which is multiplied with the likelihood function $p(\mathbf{Y}_2|\theta)$ to calculate the posterior distribution $p(\theta|\mathbf{Y}_1, \mathbf{Y}_2)$. By passing $p(\theta|\mathbf{Y}_1)$ as a prior information, we accomplishe "data fusion" in a statistical way. The same process will be repeated after all replicates are processed.

Comparing to other "Data Fusion" methods such as taking the average of all replicates despite the non-linear data model, the proposed method is less ad hoc.

III. TEST RESULTS

III-A. Test on Simulated System

We first tested the proposed algorithm on several sets of simulated MS data. The simulated data is generated based on the signal model as described in (3). Parameters related to the chemical noise in the model are generated randomly from Gaussian distributions with different means and variances to account for the fact that β_c, μ_c, ρ_c will vary with date period. We modeled β_c as Gaussian with a mean of 20 and standard deviation β_{cd}. We modeled μ_c as a Gaussian random variable with mean 0.50025 m/z and variance 0.01^2, and ρ_c as a Gaussian with mean 5.5 and variance of 0.01^2. Similarly, we have also modeled μ_s and as Gaussian random variables with mean 0.500025 m/z and variance 0.05^2). Other parameters changes significantly and we selected different combinations of $\beta_{cd}, \rho_s, \beta_s, \sigma$ to test the sensitivity of the algorithm with respect to the changes of these parameters. The typical set of values are $\beta_{cd} = 2$, $\rho_s = 20$, $\sigma_n = 2.2$. We let $\beta_s = 0$ or 5. When $\beta_s = 0$, there is no peptide signal exists, and if the algorithm returns a $p(\lambda|\mathbf{Y})$ that is greater than a threshold, then a false alarm occurs. When $\beta_s = 5$, a weak peptide signal exists in the data, and if $p(\lambda = 1|\mathbf{Y})$ is greater than a threshold, then the peak is detected. Higher values of β_s were also tested and peptide peaks are almost always detected. So we did not include those results here.

The proposed algorithm is compared with a wavelet based algorithm [2]. This algorithm first averages several replicates of MS data from the same source, then high frequency noise will be removed from data, and finally a threshold is used to determine whether peptide signal presents. The selection of different threshold will affect

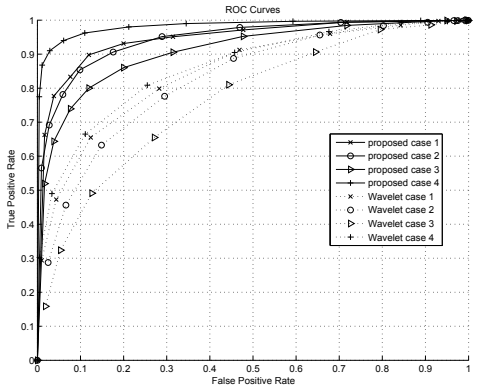

Fig. 2. ROC curve comparisons of the algorithms. figure

the true and false positive rate. If a threshold is set low, then the true positive rate of peptide signals increases at the expense of increased false positive rate. When the threshold is set high, the opposite happens. By varying the threshold, we can obtain the ROC curve. In our simulations, three replicates of simulated MS data is first added together and then averaged. Then we applied the wavelet de-noising function "wden" in Matlab which uses fixed form threshold with a multiple level estimation of noise standard deviation. The applied wavelet is "symlet8". The number of levels used for denoising is 2.

We compared the proposed algorithm and the wavelet-based algorithm in four different cases. In case one, we set all parameters to their typical values except that $\beta_{cd} = 1.5$. In case two, parameters are set to their typical values. In case three, we set $\rho_s = 30$ instead of $\rho_s = 20$ while keeping all other parameters typical. This reduces the spread of peptide signal in effect. In case 4, we reduced σ to be 1.8 and $\beta_{cd} = 1.5$ to estimate the effect of reducing additive noise variance on the tested algorithms. The resulted ROC curve are plotted in Figure 2.

From the ROC curve, we can see a significant improvement of our proposed algorithm over the wavelet based algorithm in all four cases.

III-B. Processing the Real Data

The proposed algorithm was applied to a set of real proTOF MS data. For peak detection, we set the threshold in our algorithm as $p(\lambda|\mathbf{Y}) > 0.8$. For the wavelet based algorithm, we set the threshold on intensity to 34 and 30. The results are shown in Figure 3 and Figure 4. In these figures, the three original replicates of data as well as the "statistical curve-fitting" results are plotted together. Peaks identified by the proposed algorithm are marked by xs. Peaks identified by the wavelet algorithm are indicated by stems. From this small segment, we can see that the proposed algorithm picked out all peptide peaks that are consistent through three replicates of MS data even when the peak intensity is really small such as the peak at m/z 961. On the other hand, the wavelet algorithm failed to pick out the peptide peak at m/z 956, 957 and 961 when the threshold is set as 34. When we lower the threshold to 30, the peptide peak at 956 is identified by the wavelet algorithm. While the peaks at m/z 957 and 961 are still missing, the noise peak at m/z 967 is wrongly classified as peptide signal. This shows the superior performance of the proposed algorithm

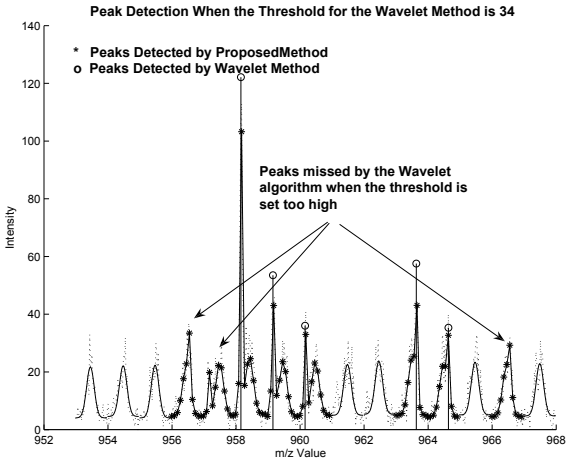

Fig. 3. Comparisons of performance on real data set. figure

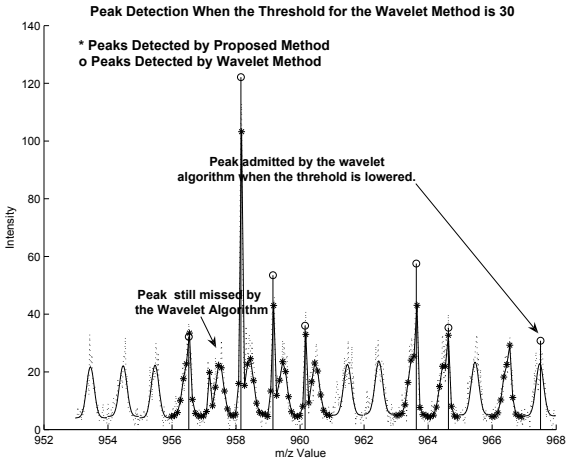

Fig. 4. Comparisons of performance on real data set. figure

IV. REFERENCES

[1] Aebersold R and Mann M., " Mass spectrometry-based proteomics", *Nature* 422:198,07, 2003.

[2] Coombes KR, Tsavachidis S, Morris JS, Baggerly KA, Hung MC, Kuerer HM., " 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 ", *(Technical Report UTMDABTR-001-04). The University of Texas M. D. Anderson Cancer Center.*, 2004.

[3] Yasui Y, Pepe M, Thompson M, Adam B, G Wright J, Qu Y, Potter J,Winget M, Thornquist M, Feng Z., " A data-analytic strategy for protein biomarker discovery: Profiling of high-dimensional proteomic data for cancer detection", *Biostatistics* 4:449463.

[4] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. Springer, 2004.