Unsupervised Adaptive Filter for Baseline Thresholding and Elimination in Liquid Chromatography–Mass Spectrometry via Approximation of the Standard Deviation of Baseline Distribution in Retention Time Domain

J. Urban*, J. Vaněk, and D. Štys

University of South Bohemia in Ceske Budejovice, Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, School of Complex Systems (former Institute of Physical Biology), Laboratory of Applied System Biology, Zamek 136, Nove Hrady 373 33, Czech Republic
E-mail: urbanj@frov.jcu.cz

Summary. The omitting presence of a baseline (background, systemic noise, mobile phase) in liquid chromatography–mass spectrometry (LC–MS) measurements impedes objective analysis. Therefore, there is a demand to remove its contribution from the signal response. Elimination of baseline contribution is justified by increasing the data mining output, both qualitatively and quantitatively. Behavior of the baseline content is not perfectly constant on the time axis, and it is often necessary to experiment with gradient changes. However, the behavior could be parametrized using a technique derived from statistical moments. In this article, we propose adaptive thresholding as an unsupervised method for baseline removal from the measurement data. Results of a real analyte measurement are discussed to illustrate its efficiency.

Key Words: LC–MS, mobile phase, systemic noise, baseline removal, adaptive thresholding

Introduction

Liquid chromatography in tandem with mass spectrometry (LC–MS) is widely used globally, and it generates terabytes of measurement data daily [1–3]. Moreover, systemic noise (instrumental and chemical) and random noise complicate the dataset. Thus, LC–MS data are “crowded” and have a non-flat baseline [4]. It is a usual occurrence in the data generated, by both the mobile phase and column bleeding. This systemic noise causes extraneous peaks or a rising baseline during gradient elution [5]. Therefore, it is often advisable for chromatographic signals to consider approaches for denoising, baseline subtraction and correction [1], especially before any chemometric feature extraction and modeling is applied.
Common algorithms based on fixed intensity thresholds [6] or wavelet transformation [7, 8] are not resistant to the loss of information from their principle. For example, mean filter in Mzmine [6] uses only one parameter, the length of the filter window. Advantage of the filter is very quick computation. Disadvantage is that the computation does not change only the maximal intensity value, but may also change the position of the maximum, if not appropriate (but a priori unknown) window length is selected.

The fixed intensity thresholds may (and surely will) filter out signals of low concentration analytes, some isotopologues, adducts, and fragmentation of the molecular ion as well as double charged ions, if their intensity is below the threshold parameter [9]. The wavelet transformations directly change the information content and are sensitive to the window length. Therefore, some of the information cannot be used for further analyzing the process. Noise filtration is a necessary processing step for emphasizing features that are relevant for other steps, especially the segmentation of the measurement into individual eluted compounds.

Sauve and Speed [10] proposed baseline correction based on mathematical morphology (top-hat operator). Their approach is very fast and simple and provides good results; however, the approach requires carefully chosen size of the “morphological window”. Unfortunately, the evaluation of the best window size depends on additional parameters: time sampling rate, mass detector resolution, and quantitation level of the intensity values. In other words, too small window will filter also analyte peaks with large value of FWHM (full width at half maximum). On the other hand, too big window will filter only main composition of the baseline. In literature, especially in image processing and systems theory field [11–17], there are presented other morphological operators with different advantages and disadvantages. The definition of the window size is crucial for each mathematical morphology, and its determination must be supervised.

Malyarenko et al. [18] arise from the similar idea as is proposed in this article: “Most of the linear baseline arises from a constant offset and the accumulation of a slowly decaying charge... In addition to this slow change, the baseline shifts as a step function...” Therefore, Malyarenko et al. created linear model of the incoming signal, and they used the inverted model to approximate the model of baseline. Advantage of Malyarenko's approach is an easy model to create and plot. Unfortunately, the depreciation may not follow a straight line, thus making a linear model may be invalid in some cases, especially for rapid gradient changes. Morris et al. [19] only extended the Malyarenko's approach by characterization of the baseline function as smooth.

Eventually, the segmentation task leads to the peak detection problem. Strong peak candidates allow the alignment to have additional flexibility [8,
Robust peak detectors require advanced analysis, e.g., noise filtration, baseline subtraction, pattern recognition, or curve fitting [20–22]. Noise additions are caused not only by random errors (random noise) but also by an influence of baseline from the liquid chromatography. The sum of the noise and the signal may result in a false interpretation of the signal or even hide it under reasonable level. Therefore, the baseline in LC–MS negatively affects the measurement analysis and represents systematic noise at a nonlinear level on the time axis [9, 23].

However, LC–MS data are complex, and the algorithms based only on filtering noise in the chromatographic domain (total ion current, TIC) or only in the individual mass spectra cannot have a performance as good as algorithms that incorporate information from both domains together. In general, any filtration method that uses fixed threshold values is problematic. Its results are often inconsistent between runs, instrumentation, and methods because the values from the nearest threshold neighborhood may be easily misclassified. The exact time characteristics of the systemic noise vary for each mass. It is necessary to analyze the characteristics independently in every mass. The threshold value is not only an attribute of the measurement but also used as an input parameter.

The contribution of the baseline varied according to the kind of instrument used and the experiment performed. However, the baseline is always present in any analyte measurement, even in a blank measurement. In this study, the blank is considered to be a chromatographic measurement without the addition of a sample. Therefore, it is usually just a mixture of solvents. Hence, the blank is easily obtained for a variety of experiments, and it is often not used any further. The methods used for baseline subtraction are based on direct subtraction of the blank from the measurement data. However, their results are not optimal because of the random influences that add additional noise to the measurement. Moreover, the baseline characteristic in the blank is not chemically affected by the analyzed substances. In other words, the description of the baseline in the analyte has to be a generalized description of the blank baseline.

Therefore, this paper proposes an unsupervised adaptation of the baseline to the analyte measurement based on statistic moments. Behavior of the baseline content is not perfectly constant on the time axis, as it is often necessary for experiments with gradient changes. However, the behavior could be parametrized using a technique derived from statistical moments. Systematic errors become more noticeable as they create border effects, which are systematically over- or underestimated [24]. The basic preprocessing steps are usually baseline removal and data smoothing [4].

Recently, the potential of high-performance liquid chromatography (HPLC)–MS for metabolomic studies has been highlighted due to its capa-
bility of routinely handling large sequences of samples. This instrument provides excellent reproducibility and usefulness for qualitative analysis. However, some questions have been raised about the quantitative abilities of HPLC-MS analysis. Several studies have discussed the fact that ion competition among different analytes exists when they are simultaneously ionized [25–29]. However, the extent of the uncertainty, which highlights the potential impact of ion competition on the analysis of complex biological samples, has not yet been given sufficient attention. Mass spectrometry is not quantitative inherent. Huge amount of physicochemical attributes (hydrophobicity, charge, size, stability, etc.) causes differences in MS detector response. Common mass spectrometers sample just a fraction of all analytes presented in the extract. Because of limitations in data quality, just a fraction of all identified analytes could be reasonably quantified [30].

In quantitative LC-MS bioanalysis, a lot of factors cause instability during sample extraction, collection, storage, LC-MS processing, and analysis [31]. For label-free quantification, which is typical especially in metabolomic screening of a priori unknown complex mixtures, experimental setup affects the quantification of ion intensities. Wenkui et al. [32] described a list of experiment settings avoiding attention from the point of analyzed compounds stability. For example, determination of appropriate matrix is the very first step in the quantitative analysis. To reduce possible degradation, the interval between sample harvesting and storage has to be as quick as possible [33]. Used temperature also plays an important role. However, most of the compounds will show no stability changes, some of them may be more stable using lower temperature [34, 35]. Analysis of biological samples with a priori unknown light sensitivity of the analytes is necessary to be taken into account [36]. Surface of the container and its adsorption of molecules in biological samples have drawn attention in pharmaceutical bioanalysis [37, 38].

Therefore, two separate experiments are usually carried out, identification of the LC fraction of possible bioactive compound, and as second, it is carried out only by direct MS injection of the fraction, to optimize sample signal. Thus, matching of integrated peak intensities is performed by using a combination of accurate mass and retention time [30].

**Experimental Setup and Theory**

The measurements were carried out using an HP 1100 Agilent liquid chromatograph with an HP 100 XD SL-Ion trap. The samples were separated on a reversed phase column (Zorbax XBD C8, 4.6 × 150 mm, 5 μm) at 30°C and eluted by gradient MeOH/H2O + 0.1% HCOOH (30–100% MeOH for 30 min, 100% for 5 min) at a flow rate of 0.6 mL min⁻¹. The settings of the
Unsupervised Adaptive Filter 261

electrospray ionization were as follows: positive mode, ramp range from 1500 to 4500 V, nebulizer at 50 psi, dry gas at a flow rate of 10 L min$^{-1}$ and a dry temperature of 325°C. The ion trap was set to target mass 900 with a range of 100 to 1000 in profile mode, as samples for methodology development were used by routine LC–MS measurements for metabolomic screening of several cyanobacterial species.

In the LC–MS measurement, the compounds eluted from the chromatographic column are separated according to the column-specific chemical properties on the time axis (retention time). Output of the LC column enters the ionization chamber in mass spectrometer. Molecules are then separated according to the mass and charge (2nd axis, m/z) and are detected by the MS detector. The third axis of measurement data represents the intensity, i.e., the amount of molecules detected at a certain position.

Peaks created by compounds separated in time occur only in a specific short time interval. On the other hand, the mobile phase, which carries the compounds through the column, is present in a larger part of the measurement. In the blank measurement, the presence of the baseline as a dominant part of the measurement is expected. However, random spikes (random errors, random noise) and impurity peaks (chemical noise) are also present, as shown in Fig. 1A. The intensity scale range suppresses a contrast between m/z values of the low and high occurrences. It is advisable to

![A](image1.png) ![B](image2.png)

Fig. 1. 3D example of blank measurements in decimal (A) and logarithmic (B) scale

change the scale of the intensity axis to nonlinear (i.e., by logarithms) to increase the information visibility in the measurement. From Fig. 1B, it can be observed that some masses are present during the whole measurement with only small changes on the time axis. The log scale in a 3D graph illustrates the flow as an emerging pattern with a simple structure. The problem, with
classical representation of intensity in decimal scale, is given by the range—the high intensity values are more visible, but the trends are hidden for low intensity values (Fig. 1A). When the logarithmic scale is used for intensity values, the trends (random spikes, peaks, baseline “waves”) even for low intensity values are more visible.

The blank measurement gives us the opportunity to examine the description that separates the baseline from the peaks and random spikes. The baseline presence is similar in both analyte measurements and blank measurements. However, the characteristic is further hidden under analyte influence (presence of the analyte peak influences the systemic noise, or systemic noise influences the analyte; it is just a point of view). The threshold value that separates baseline signals from the analytes is derived from statistical parameters of the whole measurement.

The hypothesis of adaptive baseline thresholding arises from the following knowledge: analyte masses create peaks along the time axis. Therefore, the maximal signal of the analytes increases rapidly above the average signal in the analyte retention time. On the other hand, the baseline signal just shows a slow increase or decrease. One of the standard methods for automated data processing and observation is the estimation of max-to-mean ratio $R_{\text{m}}$:

$$R_{\text{m}}(m_i) = \frac{X_Y(m_i)}{u_Y(m_i)},$$

where $m_i \ [m/z]$ is the $i$-th mass of the measurement, $X_Y(m_i)$ is maximal intensity of the $i$-th mass, and $u_Y(m_i)$ is mean intensity of the $i$-th mass. The max-to-mean ratio is still just a vector of values paired with $m_i$. Every thresholding requires further pre-processing like smoothing or binning to reduce the effects of minor observation errors. The number of binning intervals, $b_i$ should be estimated via Sturges’ rule [41]:

$$b_i = 1.5 + 3.3*\log_{10}(N_R),$$

where $N_R$ is the number of max-to-mean ratio values. Subsequently, the max-to-mean ratio $R$ is binned into a histogram $h$ according to the estimated number $b_i$, with intervals $n$. The binning serves as pre-processing step to construct the histogram, which simply reduces the effects of minor observation errors. The original ratio values which fall in a small interval, a bin, are replaced by a value representative of that interval. It is the basic method to approximate the shape of the distribution. There are several rules how to estimate optimal number of bins as well as length of the bin interval. The most used are Sturges’ rule and the Freedman–Diaconis rule. Because the shape of max-to-mean ratio probability function always diverges to power law
shape from its principle, there cannot be evaluated exactly any central moments. Power law distribution is characterized by having infinite central moments always. However, the most important value from the histogram is then the position \( p \), where the maximal occurrence is in the histogram plot. Thanks to central limit theorem, this position \( p \) is equal to the average value of the max-to-mean ratio in the normal distributions of \( R \). However, the distribution in real samples is more or less shifted, skewed, heavy-tailed, and diverges to power law distribution, when more data are accumulated. Therefore, the position of \( p \) is a priori unknown and identification of the distribution is nontrivial task. In this article, we suggest the approximation of just the standard deviation (positive square root of the second central moment) by a weighted function of interval difference instead of the approximation of the whole distribution function. The standard deviation for a discrete variable \( x \) is defined as

\[
(\Sigma_{i} n \ [X_{k} - E(x)]^{2}p_{k})^{1/2},
\]

where \( E(x) \) is the mean value of variable \( x \) and \( p_{k} \) is the probability that variable \( x \) has the value \( x_{k} \).

The histogram is an estimation of the probability distribution, in our case the probability of the max-to-mean ratios. Because the total area of the histogram is equal to the number of data, it is necessary to normalize the histogram to unity to obtain probability values. Therefore, the histogram \( h \) represents the probability of the max-to-mean ratio \( R \). Then, the position \( n_{k} \) is an average value of the \( k \)-th interval defined by Sturges’ rule. The influence of binning simplification is approximated by the heuristic [42] constant shift:

\[
s = 1/(N_{R} - 1) * (\Sigma_{i}[(n_{k} - p)^{2}h_{k}])^{1/2} + \pi/2.
\]

The threshold value \( Th \) for the max-to-mean ratio \( R \) depends on the relation between the \( p \) value and the approximated standard deviation as shown in Table I. It adapts to the actual position \( p \) in analyte samples via the approximated standard deviation \( s \) as the compensation of the shift in the histogram \( h \) from the ideal normal distribution.

Table I. Description of adaptive threshold \( Th \) selection. Corresponding threshold \( Th \) is set to the \( p \) higher than expressions in the first row. The \( p \) values correspond to the sigma rule distances in normal distribution

<table>
<thead>
<tr>
<th>If ( p )</th>
<th>( &gt; 3s )</th>
<th>( &gt; 2.576s )</th>
<th>( &gt; 1.96s )</th>
<th>( &gt; 1.645s )</th>
<th>( &gt; 0.674s )</th>
<th>( &gt; \text{abs}(p - s) )</th>
<th>( &lt; \text{abs}(p - s) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Th )</td>
<td>( p - 3s )</td>
<td>( p - 2.576s )</td>
<td>( p - 1.96s )</td>
<td>( p - 1.645s )</td>
<td>( p - 0.674s )</td>
<td>( \text{abs}(p - s) )</td>
<td>( p )</td>
</tr>
</tbody>
</table>
The whole process is illustrated on example in Fig. 2, where 2A just plots the maximal intensity values of each mass of \( m/z \) domain, 2B plots the mean intensity values of each mass of \( m/z \) domain, and 2C plots the max-to-mean ratio of each mass of \( m/z \) domain. Finally, the 2D illustrates the histogram of ratios. The \( x \)-axis of 2D is the same as the \( y \)-axis of 2C, and the \( y \)-axis represents the amount of masses in the measurement with the max-to-mean ratio given by \( x \)-axis. The shape of the distribution is clearly power law with a tail of highest ratio. The tail is created by the mass values with narrow high intensive peaks in time domain. There are the analytes, that do not require filtration—they are already clearly visible in the TIC. The filtration is necessary to distinguish low intensity peaks from the baseline contribution. The thresholding is not based directly on the intensity values, but on the max-to-mean ratios. Therefore, even the low intensity peaks may pass the filtration. In other words, the isotopologues, ion fragments, and adducts of the molecular ion rather pass through the filtration process as is shown in Fig. 3C.

![Fig. 2. Example of mass signal max values (A), mass signal mean values (B), mass signal max-to-mean ratio (C), and max-to-mean ratio histogram in continuous representation, (D) evokes heavy tailed distribution](image-url)
Fig. 3. Examples of filtered blank (A) and analyte (B) measurement TICs. Dash-dotted line represents original raw data, dotted line is for removed baseline, and solid line is the remaining peaks with random noise. Overlay of all peaks of analyte measurement after filtration is illustrated on panel C.
Results and Discussion

A computed threshold value is used for thresholding the measurement mass values $m/z$ in the domain of $R$. In other words, the contribution of all $m/z$ values with ratio $R$ below the threshold $Th$ is classified as the baseline. These masses are then removed from the measurement. Adaptive thresholding is a very fast processing method. The total time of computation is about 0.03 s (CPU P8600, 2.4 GHz, 4 GB RAM).

![Fig. 4. At the top: example of adaptive filtration of extract from Nostoc sp. with addition of 0.05 mg mL$^{-1}$ of antifungal drug Nystatin, raw mass spectrum on the upper left, filtered mass spectrum on the upper right. At the bottom: example of top hat filtration of extract from Nostoc sp. with addition of Nystatin (raw mass spectrum is at the top left), top hat filtered mass spectrum is at the bottom left, signal considered as baseline contribution by top hat filtration is at the bottom right.](image)

The proposed adaptive thresholding removes the influence of the baseline from the individual $m/z$ values according to their behavior. This method is independent of the absolute intensity values. In other words, there is no fixed threshold on intensity levels. This nontrivial property
causes two eminent features. First, even the high intensity masses may be evaluated as a baseline contribution, if present. Second, some small peaks hidden in the noise will arise after the filtration process as shown in Fig. 3C. Fig. 3A plots the result of presented adaptive filter applied on the blank measurement, where major contribution is done by mobile phase and minor by random noise and impurities. Fig. 3A illustrates that most of the dataset was filtered as baseline, as expected. In Fig. 3B, the same algorithm is applied to real measurement of cyanobacteria Nostoc sp. Independent peaks of filtered measurement from the Fig. 3B are plotted in Fig. 3C.

Fig. 4 illustrates the results of adaptive filtering applied on measurement of 70% MetOH extract from Nostoc sp. by addition of known amount (0.05 mg mL$^{-1}$) of antifungal drug Nystatin (Formula C47H75NO17, mol. mass 926.09). The left panel of Fig. 4 shows original raw mass spectrum of the Nystatin, and the right panel of Fig. 4 shows filtered mass spectrum of Nystatin via adaptive filtering. The most intensive mass in raw spectrum is most likely an ester-phthalate from preparation plastics. In comparison, top hat baseline correction was applied on the same dataset, with minimal possible size of the structural element (3 datapoints) to remove maximal amount of signal considered as baseline by top hat algorithm. The top hat filtered mass spectrum is illustrated at the bottom left panel of Fig. 4. To emphasize the result of top hat filtration, also the signal removed by top hat filtration is illustrated at the bottom right panel of Fig. 4. The top hat filtration is unable to deal with the dataset of low sampling rate (long intervals in time domain).

Another example of the adaptive thresholding is illustrated in Fig. 5. The dataset origin was experiment of concentration-response dependence in mass spectrometry measurement study on hepatotoxin microcystin-LR analysis in extracts of food additives. The pure MCYST-LR standard (Sigma No. 33893) was diluted in methanol to obtain required concentration. Several food additives were mixed with known concentration of MCYST-LR. The obtained total ion chromatograms (TICs) were evaluated, and protonated molecular ions were detected on the basis of signal intensity, presence of sodium and potassium adducts, and distribution of isotopologues.

Chromatographic peak of MCYST-LR in concentration of 0.01 μg mL$^{-1}$ in Stigeoclonium extract was observed in retention time (rt) of 17.12 min. The most intensive observed peak within the mass spectrum in given retention time was the double charged sodium adduct ion [M + Na]$^{2+}$ (m/z 509) followed by molecular ion [M + H]$^+$ (m/z 995). Within the mass spectrum, the cleavage of Adda moiety (m/z 135) was also visible by formation of ion 861 [M + 2H − 135]$^+$. The upper left panel of Fig. 5 shows raw mass spectrum with marked mass values of MCYST-LR analyte (molecular ion, double charged sodium adduct, Adda moiety, and Adda cleavage).
right panel of Fig. 5 shows top-hatted mass spectrum of MCYST-LR in rt of 17.12 min. The result of adaptive baseline filtration is at the bottom left panel. Both approaches were able to keep MCYST mass values pass the filtration processing. Bottom right panel of Fig. 5 shows three TICs in one plot: raw TIC, filtered TIC, and baseline removed by adaptive filtering.

![Fig. 5. Example of adaptive filtration of MCYST-LR in concentration of 0.01 μg mL in food additive Stigeoclonium extract in retention time (rt) of 17.12 min. Raw mass spectrum at the upper left, filtered mass spectrum at the bottom left. The top right shows example of top hat filtration of the raw spectrum, TICs are in one of the bottom right: raw TIC (solid line), filtered TIC (dash-dotted), and baseline (dotted) removed by adaptive filtering. The molecular ion (m/z 995), double charged sodium adduct (m/z 509), Adda moiety (m/z 135), and Adda cleavage (m/z 861) are marked](image)

An interesting example is illustrated in Fig. 6. Cyanobacteria Nostoc sp. (strain internal ID No. 17) is very promising strain; different bioactive compounds (trypsin inhibitor, cytotoxic compounds) were identified already [23, 43]. Our approach was found as very useful to reveal the minor component of crude extracts. The upper left panel of Fig. 6 shows mass spectrum
of fraction containing compound (m/z 859, sodium adduct 881) with antiinflammation activity in retention time 16.17 min. The baseline filtration itself was a challenge, as is shown at the bottom right panel of Fig. 6. At the bottom left panel of Fig. 6, it is visible that the filtered mass spectrum in the rt 16.17 shows the molecular ion, first isotopologue and sodium adduct as dominant part of the spectrum. The compound was identified later (publication in preparation).

Fig. 6. Example of adaptive filtration of Nostoc sp. 17 extract in retention time (rt) of 16.17 min. Raw mass spectrum at the upper left, filtered mass spectrum at the bottom left. The top right shows example of top hat filtration of the raw spectrum: TICs are in one of the bottom right: raw TIC (solid line), filtered TIC (dash-dotted), and baseline (dotted) removed by adaptive filtering. The molecular ion (m/z 859), isotopologue (860), and sodium adduct (881) are marked

Proposed adaptive filtering is not able to remove all noise contribution in the dataset; however, it removes statistically significant (significant to the standard deviation of max-to-mean ratio) signal. Moreover, the algorithm of adaptive baseline filtration does not change the intensity values of any pair
(time, m/z) in the dataset to prevent available quantitative information. The filter only creates the list of mass values that are considered as blank contribution, and therefore should be removed from the dataset. That way of filtration is therefore primarily qualitative processing of the dataset. On the other hand, a question of an appropriate max-to-mean ratio distribution (approximated via histogram) fitting still remains. Of course, proper identification of the probability density function gives exact values of the relevant central statistical moments instead of any approximation. Quite the contrary, the task of finding a function type specification that fits the distribution is one of the most puzzling issues [44–46]. In fact, it is the searching for data analysis process for constructing mathematical mapping that minimizes displacement of the data points. A common approach is to create a class of possible models, but it is not always obvious which models should be used [47]. Even with the understanding of underlying physical and chemical properties of the problem, it is difficult to choose the right model. Hence, regression analysis is used in both linear and nonlinear modeling [48] to investigate the hypothesis about the relationship between the variables of interest. Specific cases use various iterative methods for value interpolation [49, 50], in which the function must go exactly through the points. The objective of regression analysis is to produce an estimate of the hidden parameters [51]. Unfortunately, any parameter analysis can only help in differentiating between the hypothesis and the model [52]. Very strong results still do not prove that the correct fitting function was chosen [53].

In addition, precise adaptation to analyte signals is conditioned by the shifting and skewing of the max-to-mean ratio distribution in the approximation. This should be estimated from the distribution of the analyte peaks, again by fitting or regression analysis. The problem is that the distribution requires a correct peak description and identification of peaks, which can be done only after baseline filtration. However, the baseline filtration depends on the information about analyte peak distribution. Therefore, an adaptation based on the selection of threshold $Th$ via a relation between the position $p$, which is the maximal occurrence in the histogram, and the approximated standard deviation $s$ was developed. The proposed methodology was developed primary for LC–MS measurements; however, similar baseline drift exists in GC–MS also. The main source of the baseline in the GC–MS is not the mobile phase, but temperature gradient. Therefore, modified evaluation of the max-to-mean ratio $R$ is required. Instead of mean computed as arithmetic average, the mean has to be computed as geometric average to express tempo of baseline growth.
Conclusion

In this paper, we proposed an adaptive thresholding method for baseline removal in LC–MS measurement. This method focused on a statistical approximation of the systemic noise contribution. Estimation of systemic noise produced by the mobile phase separated the baseline contribution from the whole measurement. Further, filtration of the m/z values belonging to the baseline in a real example was illustrated. The proposed method is based on an accepted model of the LC–MS process; this model has issues that have already been discussed; however, they can be solved to make the model more compatible with available datasets.

Acknowledgments

Authors thank Jiří Soukup, Jiří Kopecký, and Pavel Souček for relevant discussion and numerous suggestions; and Pavel Hrouzek and Lucie Marková, as well as our students Pavla Dvořáková and Andrey Timoshenko for assistance with the experiments and software testing.

This work was supported and co-financed by the ERDF and made possible by the INTERREG IVC programme, project Innovation 4 Welfare, sub-project PICKFIBER; by the South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses (CZ.1.05/2.1.00/01.0024); by the Ministry of Education, Youth and Sports of the Czech Republic under the grant MSM 6007665808; and by the South Bohemia University grant GA JU 152/2010/Z.

References

[16] D. Stys, J. Urban, J. Vanek, and P. Cisar, Analysis of biological time-lapse microscopic experiment from the point of view of the information theory, Micron, S0968-4328(10)00026-0 (2010)
[38] P. Kohl and S. Medlar, Occurrence of Manganese in Drinking Water and Manganese Control, American Water Works Association, 2007

Accepted by ISD