PEAK MODELING FOR ION MOBILITY SPECTROMETRY MEASUREMENTS

Dominik Kopczynski[†] *Jörg Ingo Baumbach*[*] *Sven Rahmann*[‡]

[†]Bioinformatics for High-Throughput Technologies, Computer Science XI, TU Dortmund, Germany
dominik.kopczynski@tu-dortmund.de
[*]Department Microfluidics and Clinical Diagnostics, KIST Europe, Saarbrücken, Germany
baumbach@kist-europe.de
[‡]Genome Informatics, Institute of Human Genetics,
Faculty of Medicine, University of Duisburg-Essen, Germany
sven.rahmann@uni-due.de

ABSTRACT

Ion mobility spectrometry (IMS), coupled with multicapillary columns (MCCs), is a technology for analyzing the concentration of volatile organic compounds (VOCs) in the air or in exhaled breath. We introduce a new model-based method for peak description and thus for manifold data reduction. Depending on the number of peaks, we reduce the raw data by five orders of magnitude (about 250 000-fold). Each peak is described with seven parameters, which are estimated by an expectation maximization (EM) algorithm that copes well with overlapping peaks. We transform the parameters into robust and interpretable peak descriptors, which are suitable for downstream analysis steps.

Index Terms— Ion mobility spectrometry (IMS), data reduction, parameter estimation, mixture model, EM algorithm, three-parameter inverse Gaussian distribution

1. INTRODUCTION

Ion mobility spectrometry (IMS) technology has recently gained importance because it can measure volatile organic compounds in the air or in exhaled breath under normal air pressure, in contrast to mass spectrometry (MS), which requires a vacuum. Therefore IMS instruments are easier to build and less expensive. So IMS technology, coupled with multi-capillary columns (MCCs) for pre-separating complex mixtures, is now being explored for exhaled breath analysis in medical applications; several diseases like lung cancer [1] can potentially be diagnosed early.

Main results. The size of a full MCC/IMS measurement can be several tens or hundreds of megabytes depending on the instrument's resolution, e.g. 12500x6000 data points from 10 minutes of measurement. Comparing measurements with an annotated reference database of metabolic compounds at this level of detail is not only time-consuming, but also complicated due to inherent noise and small differences in location and shape of the peaks. Therefore we reduce the original data to few peak descriptors. Our approach is to describe the peaks with appropriate parameterized model functions which reduce the amount of data down to the parameters of the functions. Our model has seven parameters per peak, reducing the data by five orders of magnitude or about 250 000-fold. We introduce peak descriptors, which have the advantage that they can be naturally interpreted as location and shape parameters of the peaks and are in a one-to-one correspondence with the (technical) model parameters. Since peaks may overlap, we interpret the measured data as a sample from a mixture of several peak models plus background, and we use an expectation maximization (EM) algorithm for parameter estimation.

Previous work. Bader [2] used hard assignments to distinguish the peaks, with the disadvantage that such methods do not cope well with overlapping peaks. In Vogtland's approach [3], the peak description in drift time is a combination of two semi-distributions. The fronting of a peak is described by a normal distribution while after the mode a Breit-Wigner distribution is used for a better tailing approximation. Bödecker [4] used a least-square method to estimate the parameters for Vogtland's function to describe peaks. The approach [3, 4] complicates parameter estimation, since inconsistent relative weights for both semi-distributions are used to ensure continuity. Rossoni and Feng [5] previously used the EM algorithm to describe interspike interval data in the neurosciences with mixtures of (two-parameter) inverse Gaussian distributions (among others).

Outline. In Section 2 we summarize the background of MCC/IMS technology and mention standard preprocessing steps. In Section 3 we introduce the peak model and explain how a measurement is interpreted as a sample from a mixture distribution of several peaks and the background. Section 4

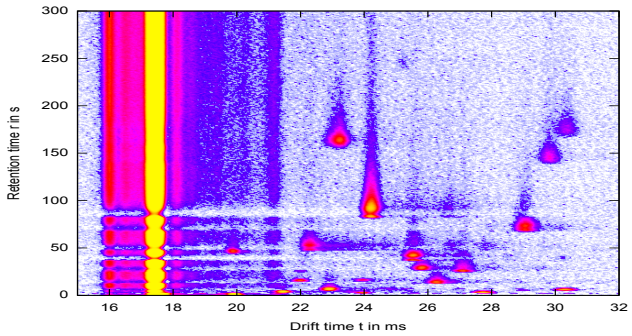

Fig. 1. Heat map of an MCC/IMS measurement. X-axis: drift time t in ms; Y-axis: retention time r in seconds; intensity: white (lowest) $<$ blue $<$ purple $<$ red $<$ yellow (highest), *reaction ion peak* (RIP) at $t = 17.5$ms

describes the EM algorithm used to estimate peak parameters, and in Section 5 we introduce robust interpretable peak descriptors. An evaluation of our approach is presented in Section 6, while Section 7 concludes.

2. BACKGROUND

MCC/IMS experiments. The function of IMS devices is well documented [6] and is only summarized here briefly. For this work, a BioScout IMS (B&S Analytik, Dortmund, Germany) with a ^{63}Ni ß-ionization source is used. Neutral analyte compounds are ionized by reaction ions into product ions before they are moved though the IMS tube by an electric field. Molecule properties like mass, polarizability and structure lead to a characteristic drift time for every single compound. To obtain comparable drift times independent of external conditions (temperature, pressure), they are converted into reduced inverse mobilities [6]. The measurement process takes about 100 milliseconds. A signal intensity (voltage change) is captured at each time point with 250 kHz.

It is impossible to distinguish different compounds with identical mean drift time. Therefore the IMS device is coupled with an MCC, which separates the compounds before they enter the IMS device. According to the strength of interacting molecules with the matrix of the MCC, different compounds are retained for different times in the MCC. After a specific retention time r, several molecules of a compound reach the IMS device. Thus an IMS measurement is executed many times, about ten times each second. Let R be the finite set of retention time points where an IMS measurement is taken, and let T be the finite set of measured drift time points. Because of equidistant time points, we may assume that $R = \{1, \ldots, |R|\}$ and $T = \{1, \ldots, |T|\}$. We obtain a two-dimensional spectrum $S = (s_{r,t})$ indexed by retention time $r \in R$ and drift time $t \in T$. Here $(s_{r,t})$ can be interpreted as a sum of events counted at index (r, t). It can be visualized as a two-dimensional heat map (Fig. 1). A column $S_{\cdot,t}$ with fixed drift time t is called a *chromatogram*.

Preprocessing. The data is noisy and contains a dominant reaction ion peak (RIP, visible in Fig. 1 around $t = 17.5$ ms). To compensate for such signals that are constantly present at specific drift times t, we subtract the chromatogram's median from each chromatogram. We next use a low-pass filter to remove high frequencies via a two-dimensional discrete Fourier transform (DFT). We smooth the data in each spectrum with a Savitzky-Golay filter, which computes weighted averages across small drift time windows (9 data points). These are standard procedures and already used in commercial software.

3. PEAK MODEL

Regions with a high signal intensity are called peaks. Since peaks may overlap, we cannot assign each coordinate (r, t) to a unique peak (or to the background). So a probabilistic assignment method is desirable, and we describe the whole measurement S as a mixture of several *peak components* and a *background component* accounting for remaining noise.

The main feature of our proposed model is that the measured values $s_{r,t}$ are interpreted as a (finite) sample from a two-dimensional probability density $f = f(r, t)$, the said mixture model. Detecting (overlapping) peaks is challenging, and we assume that the number of peaks c and their mode positions are known. The problem is to first choose a parametric family for the components of f and then, from the data, determine the parameters of each component and probabilistically assign coordinates (r, t) to components.

To choose appropriate functions for the peak model, we examined several hundred measured peaks and noted that (a) horizontal cross sections (intensity over drift time) are almost symmetric, but sometimes slightly skewed (when the ionized molecules moved through the drift tube, the drift gas flows in opposite direction through the tube to blow the deionized molecules out of the tube, slowing the ions), (b) vertical cross sections (intensity over retention time) are almost always skewed (long tailing due to MCC interaction).

The inverse Gaussian distribution can represent both symmetric and tail-skewed shapes. It phenomenologically fits well to many of the observed peaks; therefore it was chosen as the parametric model in what follows. Note that there is no principled physical theory yet available to describe the peak shape on either axis that fits with the observations.

While the original inverse Gaussian distribution has two parameters $\mu > 0, \lambda > 0$ with a fixed origin at zero, we use an additional offset parameter $o \in \mathbb{R}$. It is known as the *shifted* or *three-parameter inverse Gaussian distribution* [7]. The density is

$$g_{\mu,\lambda,o}(x) = [x > o] \cdot \sqrt{\frac{\lambda}{2\pi(x-o)^3}} \cdot \exp\Big(-\frac{\lambda\big((x-o)-\mu\big)^2}{2\mu^2(x-o)}\Big).$$

Here $\mu > 0$ is the mean of the distribution (relative to the offset o; thus the true mean is $\mu + o$), and $\lambda > 0$ controls the

shape and skewness of the distribution.

The two-dimensional peak model density function p is a product of shifted inverse Gaussians in both retention time r and drift time t and defined as

$$p_{\mu_R,\lambda_R,o_R,\mu_T,\lambda_T,o_T}(r,t) := g_{\mu_R,\lambda_R,o_R}(r) \cdot g_{\mu_T,\lambda_T,o_T}(t).$$

The model is a probability density, so $\int_{\mathbb{R}^2} p(r,t)\,\mathrm{d}r\,\mathrm{d}t = 1$. To describe peaks of different (relative) volume, the model function can be scaled by an arbitrary positive factor. The data is not on $\mathbb{R}^2$, but on a finite lattice $R \times T$, so using the density values $p(r,t)$ directly yields a discretization error that is corrected by introducing a normalization factor $Z_p = \sum_{r,t} p(r,t)$ and replacing p by $P := p/Z_p$.

A measurement is modeled as a sample from a weighted mixture of peak models. Let c be the number of peaks and

$$\theta_j := (\mu_{R,j}, \lambda_{R,j}, o_{R,j}, \mu_{T,j}, \lambda_{T,j}, o_{T,j})$$

be the parameter vector of peak j ($1 \le j \le c$), and let ω_j be the normalized weight of peak j. To explain the remaining noise and regions without peaks, we assume a background component with uniform density $1/(|R||T|)$ across all retention and drift times and normalized weight ω_0; for convenience, we set θ_0 to an empty parameter vector. Thus the full mixture model has probability mass function

$$f_{\omega,\theta}(r,t) = \frac{\omega_0}{|R||T|} + \sum_{j=1}^{c} \omega_j \cdot P_{\theta_j}(r,t). \tag{1}$$

4. PEAK PARAMETER ESTIMATION

The goal is to estimate parameters ω_j and θ_j for each component, such that the log-likelihood

$$\sum_{r\in R,\, t\in T} s_{r,t} \cdot \log f_{\omega,\theta}(r,t) \tag{2}$$

of the measurement S is maximized, where $f_{\omega,\theta}$ is the mixture from (1). However, this is difficult because (2) is not concave in ω, θ and has several local maxima.

The expectation maximization (EM) algorithm [8] is an iterative method to optimize parameters in mixture models, starting with reasonable estimates, and increasing the log-likelihood in each step until convergence to a local optimum. It has been successfully applied in many contexts, among others to mixtures of (non-shifted) one-dimensional inverse Gaussian distributions [5].

EM works by introducing hidden assignment variables $W_{(r,t),j} \in \{0,1\}$ indicating whether data point $(r,t) \in R \times T$ belongs to model component $j \in \{0,1,\dots,c\}$. The joint log-likelihood of W, ω, θ for data $S = (s_{r,t})$ is

$$\sum_{r,t} s_{r,t} \sum_{j=0}^{c} W_{(r,t),j}\, \log(\omega_j P_{\theta_j}(r,t)). \tag{3}$$

For EM, the assignment is soft (probabilistic) instead of hard, and $W_{(r,t),j}$ is replaced by the expected value

$$\overline{W}_{(r,t),j} := \mathbb{E}[W_{(r,t),j} \mid S] = \mathbb{P}(W_{(r,t),j} = 1 \mid S),$$

conditional on the data S. The EM algorithm alternates between two phases: In the expectation (E) phase, the expectations of the hidden variables are estimated. In the maximization (M) phase, the parameters of each components and the relative weights are optimized with maximum-likelihood-estimators (MLE), using the fixed hidden variables. Each of the phases results in convex optimization problems if the model functions P_{θ_j} are appropriately chosen, and one can prove that the log-likelihood increases in every step until convergence to a local optimum [9].

Expectation Step. We estimate $\mathbb{P}(W_{(r,t),j} = 1 \mid S)$, where $\mathbb{P}$ denotes the probability measure of the mixture model with current parameters (θ^0, ω^0). Using Bayes Theorem, we arrive at the intuitively appealing result that $\overline{W}_{(r,t),j}$ is proportional to $\omega_j^0\, P_{\theta_j^0}(r,t)$, that is, ensuring proper normalization,

$$\overline{W}_{(r,t),j} = \frac{\omega_j^0\, P_{\theta_j^0}(r,t)}{\sum_k \omega_k^0\, P_{\theta_k^0}(r,t)}. \tag{4}$$

Maximization step. Using $\overline{W}_{(r,t),j}$, the data $s_{r,t}$ decomposes into independent components $s_{r,t}^{(j)}$, each of which represents a single peak (or the background), via

$$s_{r,t}^{(j)} := \overline{W}_{(r,t),j} \cdot s_{r,t}. \tag{5}$$

We first estimate the new model weights ω_j^* by maximum likelihood; a standard calculation shows that

$$\omega_j^* = \frac{1}{Z_S} \sum_{r,t} s_{r,t}^{(j)},$$

where $Z_S = \sum_{r,t} s_{r,t}$ is the total signal intensity.

For the background component ($j = 0$), there are no further parameters, and we are done.

For the peak component, we estimate each parameter of θ_j. MLEs for the one-dimensional shifted inverse Gaussian are known [7, 10]. Since our two-dimensional model factors into two one-dimensional inverse Gaussians, these estimators can be used by marginalizing over the respective other dimension. We here provide the resulting estimators $(\mu_{R,j}^*, \lambda_{R,j}^*, o_{R,j}^*)$ for the retention time axis. Let $\tilde{s}_r^{(j)} := \sum_{t\in T} s_{r,t}^{(j)}$ be the marginalized signal.

The relative mean is naturally estimated as

$$\mu_{R,j}^* = \frac{\sum_r r \cdot \tilde{s}_r^{(j)}}{\omega_j^*\, Z_S} - o_{R,j}^*.$$

Thus to estimate $\mu_{R,j}^*$ the offset estimate $o_{R,j}^*$ has to be known. We solve this problem by first using the previous value $o_{R,j}^0$,

then estimate $o^*_{\mathrm{R},j}$ (see below) using $\mu^*_{\mathrm{R},j}$, and finally update $\mu^*_{\mathrm{R},j}$ by using $o^*_{\mathrm{R},j}$. This can be repeated until convergence, but we found that one iteration is sufficient in practice.

The shape parameter $\lambda^*_{\mathrm{R},j}$ is ML-estimated by

$$\lambda^*_{\mathrm{R},j} = \frac{\omega^*_j \, Z_S}{\sum_r \tilde{s}^{(j)}_r \cdot \left(1/(r - o^*_{\mathrm{R},j}) - 1/\mu^*_{\mathrm{R},j}\right)}.$$

We circumvent the difficulty of requiring $o^*_{\mathrm{R},j}$ as above.

The ML-estimator for the offset $o^*_{\mathrm{R},j}$ is determined by

$$0 = \sum_r \tilde{s}^{(j)}_r \cdot \left(\frac{3}{\lambda^*_{\mathrm{R},j}(r - o^*_{\mathrm{R},j})} + \frac{1}{(\mu^*_{\mathrm{R},j})^2} - \frac{1}{(r - o^*_{\mathrm{R},j})^2}\right).$$

There is no closed formula for $o^*_{\mathrm{R},j}$, but its value can be efficiently determined using Newton's method.

Starting values. EM requires starting parameters ω^0, θ^0, from which it will converge towards a local optimum. Choosing reasonable starting values is essential for quick convergence and good results. The initial component weights are chosen uniformly as $\omega_j := 1/(c+1)$, but they are immediately re-estimated. The key issue is choosing reasonable peak parameters θ_j. Since the shifted inverse Gaussian distribution is zero to the left and below of $(o_{\mathrm{R}}, o_{\mathrm{T}})$, the origin parameters should be set to the left and below the main peak volume. This is at present done visually and interactively with a heatmap visualization of the IMS spectrum as in Fig. 1. The other parameters can be initialized with any small positive value.

Convergence criteria. There are two alternatives for stopping EM: no significant log-likelihood improvement, or no significant change in the model parameters. We check the model parameters. Let $\theta_{j,k}$ be the k-th parameter in θ_j; let $\theta^0_{j,k}$ denote the old value and $\theta^*_{j,k}$ the new value. We compute the *regularized relative change* $C_{j,k} = \left|\theta^*_{j,k} - \theta^0_{j,k}\right| / (|\theta^*_{j,k}| + 1)$. Iteration stops when $C := \max_{j,k} C_{j,k} < \varepsilon$, a given threshold. In practice $\varepsilon = 10^{-3}$ works well, and we usually reach convergence after 10–20 iterations, depending on the number of peaks and their overlap. On an Intel i5 Quad-Core 2.8GHz, this takes about one second per peak on a standard-resolution IMS (3×10^6 data points) and 20 seconds per peak on a high-resolution IMS (75×10^6 data points).

5. PEAK DESCRIPTORS

Substantially distinct parameter values of the shifted inverse Gaussian may result in similar distributions (Fig. 2). For human interpretation and for comparing different experiments, it help if similar distributions were described with similar *peak descriptors*, replacing the technical parameters (μ, λ, o). Here we advocate the mean μ', the standard deviation σ, and the mode m (position of the maximum). Note that, because of

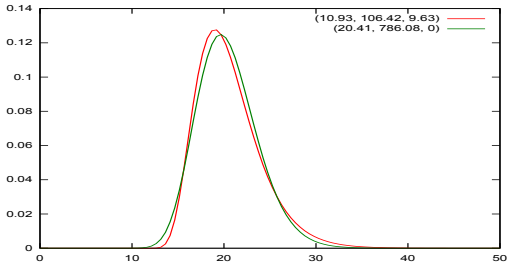

Fig. 2. Significantly different shifted inverse Gaussian parameter combinations (μ, λ, o), here $(10.93, 106.42, 9.63)$ and $(20.41, 786.08, 0)$, result in similar distributions.

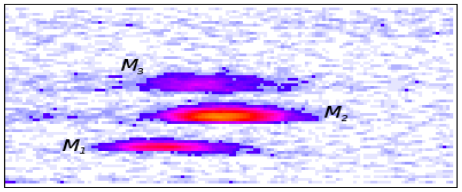

Fig. 3. Section of an IMS measurement with three peaks

the skewness, mean and mode generally do not agree for an inverse Gaussian. The three descriptors are in a one-to-one-correspondence to the technical parameters. In particular,

$$\mu' = \mu + o, \qquad \sigma = \sqrt{\mu^3/\lambda},$$
$$m = \mu\left(\sqrt{1 + (9\mu^2)/(4\lambda^2)} - (3\mu)/(2\lambda)\right) + o.$$

We obtain the original parameters back from the descriptors; define

$$p := \big(-m(2\mu' + m) + 3 \cdot (\mu'^2 - \sigma^2)\big)/\big(2(m - \mu')\big),$$
$$q := \big(m(3\sigma^2 + \mu' \cdot m) - \mu'^3\big)/\big(2(m - \mu')\big).$$

Then $o = -p/2 - \sqrt{p^2/4 - q}$, $\mu = \mu' - o$, $\lambda = \mu^3/\sigma^2$.

Example. We independently ran EM 1000 times on a section of a measurement (Fig. 3) with dimensions $|R| = 54$, $|T| = 141$. Starting modes $(m_{\mathrm{R}}, m_{\mathrm{T}})$ for the three peaks were visually chosen as $(9, 55)_{M_1}$, $(20, 72)_{M_2}$, $(32, 65)_{M_3}$. To obtain randomized EM start parameters (μ, λ, o) in each dimension, we draw them uniformly from the intervals $\mu \in [m + 40, m + 60]$ and $\lambda \in [\mu^3/4, \mu^3]$. We set $o := m - \mu + 10^{-3}$. These ranges ensure that the initial model is smaller in its shape than the original peak. Table 1 shows the resulting mean and standard deviation for model parameters and descriptors. The descriptors are much more robust than the technical model parameters.

6. EVALUATION

We evaluate the quality of the obtained peak models by computing the log-likelihood of the observed normalized signal

	M_1	M_2	M_3
μ_R	59.3 ± 181	70.8 ± 180	82 ± 179
λ_R	$129000 \pm 1.16 \cdot 10^6$	133000 ± 10^6	120000 ± 787000
o_R	-50.2 ± 181	-50 ± 180	-49.6 ± 179
μ_T	96.7 ± 181	113 ± 181	108 ± 176
λ_T	8010 ± 45500	14100 ± 67100	6470 ± 30600
o_T	-49.6 ± 182	-50 ± 182	-50.3 ± 178
μ'_R	9.09 ± 0.032	20.76 ± 0.026	32.37 ± 0.055
σ_R	1.28 ± 0.013	1.64 ± 0.006	2.15 ± 0.037
m_R	9.05 ± 0.099	20.7 ± 0.122	32.28 ± 0.134
μ'_T	47.05 ± 0.739	63.58 ± 0.409	57.82 ± 1.49
σ_T	10.68 ± 0.45	10.25 ± 0.051	14.03 ± 0.997
m_T	45.29 ± 2.65	62.2 ± 1.79	55.12 ± 2.48

Table 1. Comparison between model parameters and descriptors, resulting from 1000 independent EM executions with different start parameters, mean $\pm$ standard deviation.

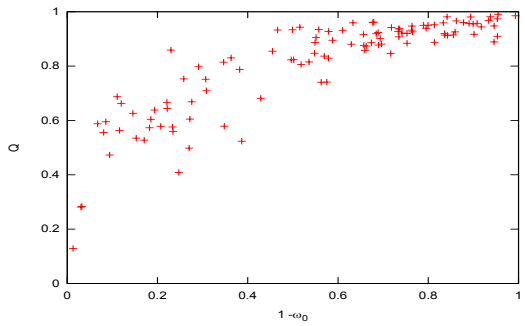

Fig. 4. Q in relation to the signal weight $1 - \omega_0$ for 110 datasets.

$s^o_{r,t} := s_{r,t}/Z_S$ under three different distributions: (1) the empirical distribution $s^o_{r,t}$ itself, the most accurate description of the data requiring $|R|\,|T|$ parameters; (2) the estimated mixture model $f_{\omega,\theta}$ requiring seven parameters per peak; (3) the uniform distribution on $R \times T$, unaware of the data and requiring no parameters. Thus let

$$L_D := \sum_{r,t} s^o_{r,t} \log(s^o_{r,t}),$$
$$L_M := \sum_{r,t} s^o_{r,t} \log(f_{\omega,\theta}(r,t)),$$
$$L_U := \sum_{r,t} s^o_{r,t} \log(1/(|R|\,|T|)) = -\log(|R|\,|T|).$$

Of course $L_D \geq L_M \geq L_U$; we claim that for IMS measurements with high peak density, $L_D \approx L_M \gg L_U$, even though our model requires only seven parameters per peak instead of $|R|\,|T|$ data points. Thus we compute the relative size of L_M in the interval $[L_U, L_D]$, that is $Q := (L_M - L_U)/(L_D - L_U) \in [0,1]$; values closer to 1 mean more accurate models.

Sparse measurements with few peaks (and much background noise) are (globally) already well described by the uniform model; thus we plot Q in relation to the weight $1-\omega_0$ of the signal components in our model (Fig. 4). For data sets that predominantly consist of peaks ($1 - \omega_0 > 0.4$), almost all Q-values are above 0.8, with few exceptions.

7. CONCLUSION

We have introduced two-dimensional shifted inverse Gaussian distributions as models for peaks in IMS measurements; they phenomenologically fit well to the observed symmetrical and skewed peak shapes. Each peak is described with seven (technical) parameters and alternatively with seven descriptors that allow better interpretation and comparison of peak position and shape. Our evaluation shows that these seven descriptors per peak suffice to describe IMS measurements accurately. Our approach results both in a data reduction by at least five orders of magnitude and simplifies further processing of peaks, such as comparisons between measurements and to reference databases. Data and additional material can be found at `www.rahmannlab.de/research/ims`.

Acknowledgments. The authors are supported by the Collaborative Research Center (Sonderforschungsbereich, SFB) 876 "Providing Information by Resource-Constrained Data Analysis" within project TB1, see `http://sfb876.tu-dortmund.de`.

8. REFERENCES

[1] M. Westhoff, P. Litterst, L. Freitag, W. Urfer, S. Bader, and J.I. Baumbach, "Ion mobility spectrometry for the detection of volatile organic compounds in exhaled breath of lung cancer patients," *Thorax*, vol. 64, pp. 744–748, 2009.

[2] S. Bader, *Identification and Quantification of Peaks in Spectrometric Data*, Ph.D. thesis, TU Dortmund, 2008.

[3] D. Vogtland and J.I. Baumbach, "Breit-Wigner-Function and IMS-signals," *International Journal for Ion Mobility Spectrometry*, vol. 12, pp. 109–114, 2009.

[4] B. Bödeker and J.I. Baumbach, "Analytical description of IMS-signals," *International Journal for Ion Mobility Spectrometry*, vol. 12, no. 3, pp. 103–108, 2009.

[5] E. Rossoni and J. Feng, "A nonparametric approach to extract information from interspike interval data," *Journal of Neuroscience Methods*, vol. 150, pp. 30–40, 2006.

[6] J.I. Baumbach and G.A. Eiceman, "Ion mobility spectrometry: Arriving on site and moving beyond a low profile," *Appl. Spectrosc.*, vol. 53, no. 9, pp. 338A–355A, 1999.

[7] I.A. Koutrouvelis, G.C. Canavos, and S.G. Meintanis, "Estimation in the three-parameter inverse gaussian distribution," *Computational statistics & data analysis*, vol. 49, no. 4, pp. 1132–1147, 2005.

[8] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[9] R. A. Boyles, "On the convergence of the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 45, no. 1, pp. 47–50, 1983.

[10] R. C. H. Cheng and N. Amin, "Maximum likelihood estimation of parameters in the Inverse Gaussian distribution, with unknown origin," *Technometrics*, vol. 23, no. 3, pp. 257–264, 1981.