

Evaluation Strategies for Coupled GC-IMS Measurement including the Systematic Use of Parametrized ANN Training Data

Artur Scheinemann¹, Stefanie Sielemann², Jörg Walter², Theodor Doll¹

¹Institute of Physics, Johannes Gutenberg-University Mainz, Mainz, Germany

²Gesellschaft für Analytische Sensorysysteme mbH G.A.S., Dortmund, Germany

Email: artur.scheinemann@gmail.com

Received August 24, 2012; revised September 26, 2012; accepted October 8, 2012

ABSTRACT

Data evaluation strategies for the novel coupled MCC-IMS sensory system are developed. Major attention to the plausibility of applied procedures and the feasibility of automation was paid. Three stages of extraction levels with increasing data reduction are presented for several fields of application. According to suitable extraction levels, real data were tested on various structures of artificial neural networks (ANN) with the result, that the computational levels must still be chosen by expertise, but subsequent processing and training can be fully automated. For the training of larger networks a method of automated generation of secondary training data is presented which exceeds the quality of previous noise models by far. It is concluded that the combination of MCC-IMS as measuring instrument and ANNs as evaluation technique have high potential for industrial use in process monitoring.

Keywords: Gas Chromatography; Ion Mobility Spectrometry; GC-IMS; MCC-IMS; Artificial Neural Network; Measurement Evaluation

1. Introduction

Ion Mobility Spectrometry (IMS [1]) and Gas Chromatography (GC [2]) have been well established measuring technologies for several decades. However their coupling into a combined measuring technology (GC-IMS) is relatively new [3,4]. Though this method is very promising in terms of sensitivity and accuracy, all evaluation tools have to be developed from the very beginning. Several attempts have been made to find an analytical approach for the description of GC-IMS spectra in order to automate the evaluation of these measurements, only little progress was gained [5,6]. Due to their 2-dimensional nature GC-IMS measurements contain great quantities of data, which, depending on the measurement setup, may contain up to 10^6 or even 10^7 data points.

2. Experimental Details

2.1. Measuring Principle and Resulting Properties of the Measurements

The GC-IMS is a system that measures 2 different properties independent from each other [7-9]. While the GC column separates analytes depending on their ability to adsorb and desorb on the inner column surface (see **Figure 1**), the IMS separates charged particles under the influence of an electrical field depending on their drift behavior in a carrier gas atmosphere. This can be seen in **Figure 2**. Beginning at the moment of the analyte injection into the GC column the output of the GC column is permanently analyzed by the IMS. At a given rate per second the IMS is recording 1-dimensional spectra until all fractions of the analyte have been eluted from the column. The taken 1-dimensional spectra are combined into one 2-dimensional spectrum with certain characteristics as it can be seen in **Figure 3**.

The carrier gas is always present in the measurement process and therefore is seen in all spectra. The according peak in the spectrum is called RIP (Reactant Ion Peak). The RIP is a constant feature of IMS spectra. The ions which are analyzed by the IMS are ionized by a radiation source in the reaction region of the IMS and then pulled by an electrical field toward the detector. The time from the opening of the gate, when the electric field starts to pull the ions till they hit the detector and cause an electric current is called the drift time of this particular ion sort. The ionized carrier gas (reactant ions) reacts with the analyte sample when latter is eluted from the GC into the IMS. Due to the reaction mechanism the formation of analyte ions competes with the amount of

reactant ions present in the reaction region. Hence the detection of a peak at a given retention time t_R reduces the intensity of the RIP at the same t_R (see **Figure 3**). Since the radiation in the ionization region is constant, the amount of produced ions is assumed to be nearly constant and hence the amount of detected charged particles at the Faraday detector.

Thus a decreasing or even vanishing RIP intensity is always the result of detected analytes forming a peak at $t_D \neq t_{RIP}$ and preserving the overall ion amount detected at a given retention time t_R .

$$\frac{d}{dt_R} Q(t_R) = \frac{d}{dt_R} \int_1^{t_D, \max} I(t_D, t_R) dt_D \approx 0 \quad (1)$$

2.2. Available Measurement Data

Depending on the processing approach, the demanded

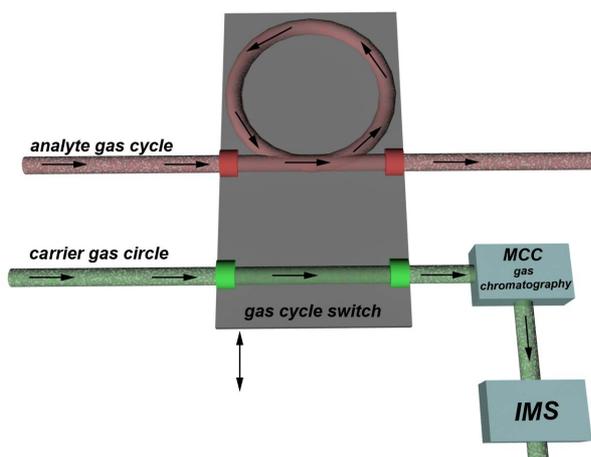


Figure 1. Schematic draft of the measurement technique and the gas flow inside of a GC-IMS detector.

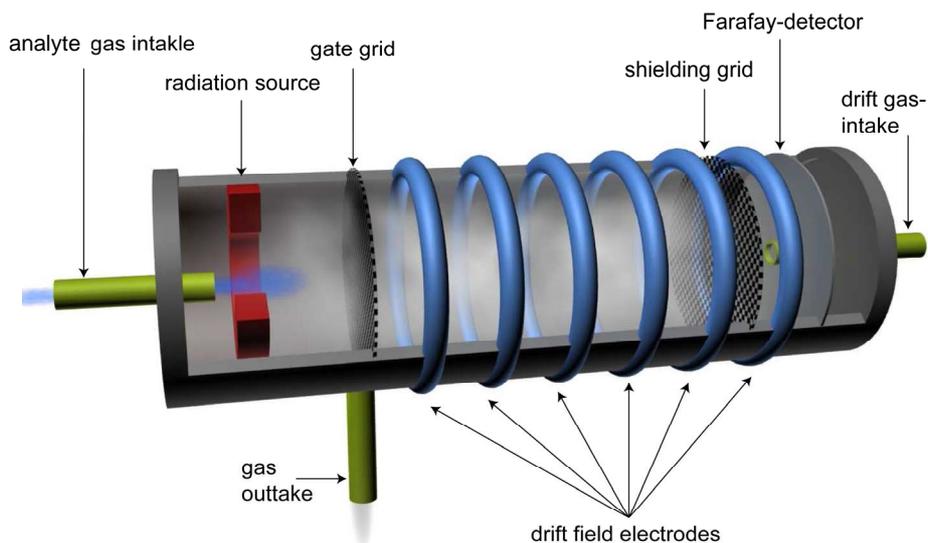


Figure 2. Schematic picture of the IMS and its parts.

amount of measurement data to initialize a certain processing algorithm can differ by orders of magnitude. Here initialization means a first and one-time execution of the algorithm code which sets the processing program into the state where it can simply read an unknown measurement sample and produce an appropriate output to classify the unknown sample. Initialization can be the storage of reference samples, the evaluation of reference samples in order to adopt certain threshold parameters which determine the classification process or the training of an ANN. Furthermore any processing algorithm, once initialized, needs to be evaluated with new data sets that were not used before during the initialization.

The most simple case is the attempt to distinguish between two possible categories for example the breath of people with and without some specific disease or the detection of some specific contaminant in any food or beverage product.

Since every processing approach has to be evaluated with really measured data it is important to have a dataset of measurements big enough. Though the motto is, the bigger the dataset the better for the evaluation, the analysis in this work had to content itself to the dataset given in **Table 1**. The measurements were carried out using the FlavourSpec and GC-IMS by G.A.S. Both systems consist of a combination of an IMS with a chromatographic column for a better pre separation of volatile organic compounds (VOCs) in complex mixtures. The essential technical and physical data of the measuring devices are assembled in **Listing 1**.

The involved interaction potentials between drift gas and analyze particles are of enormous complexity. This is especially the case with complex analyze particles as MVOCs. There are no analytical models to predict the appearance of a molecule or atom in the spectra just from

Table 1. Data sets for GC-IMS analysis tests.

Measurement class	Categories	Measurements	Measurements total
Cola	13	≈10	130
Juice	18	≈10	180
Rice	4	≈5	20
Olive oil	3	≈50	147
Breath diff. candy flavors	3	≈16	50
Meat	14	≈5	70

-IMS Parameters

- Drift length: 50 mm
- Drift voltage: 2 kV
- Electrical field strength: 400 V/cm

-Radioactive Ionisation source

- Tritium ^3H (β -radiation) 300 MBq
- Radioactive half-life: 12.5 years

-Multi Capillary Column (MCC) Parameters

- Film thickness: 0.2 μm
- Column Length: 20 cm
- Capillaries I.D. μm : 40
- Number of capillaries: 1200

Listing 1. Technical Specs of used GC-IMS.

fundamental modeling using first principle physics. Another way to distinguish between different spectra would therefore be of interest. In the following alternative approaches for evaluation and processing of GC-IMS spectra are presented. The benefits of using them in Neural Networks are discussed and results of classification tests produced by them are presented.

One problem is to find a suitable and general evaluation strategy which helps to determine from which category a given measurement is. A successful strategy not only has to yield stable classification rates on unseen spectra/measurements, it furthermore needs to be general enough to be applied to new classification problems without cumbersome modifications.

3. Simple Metrics Evaluation

After inspecting the spectra of the breath measurements (two measurements from two different flavors can be seen in **Figure 4**) a quite simple and straight forward approach gets obvious. It is to define a metrics on the two dimensional spectra in order to determine a distance between two spectra. One possible definition of this distance Δ between two arbitrary spectra $R_i(t_R, t_D)$ and $M(t_R, t_D)$ is:

$$d_i = \Delta(R_i, M) = \sum_{t_D=1}^{t_{D,\max}} \sum_{t_R=1}^{t_{R,\max}} |R_i(t_R, t_D) - M(t_R, t_D)| \quad (2)$$

This method needs only one reference measurement

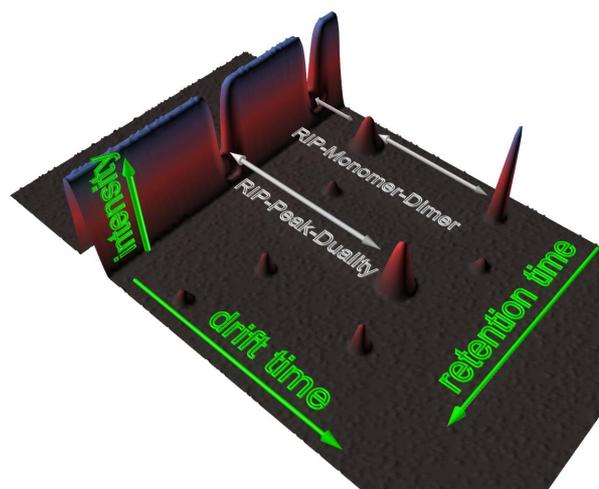


Figure 3. GC-IMS spectrum shown in perspective view with the duality of RIP and Peak. This is an artificial illustration intended to visualize basic principles of GC-IMS measurements. It is therefore exaggerated and not completely consistent since the smaller peaks should as well diminish the RIP intensity. The properties of this spectrum are discussed in the text.

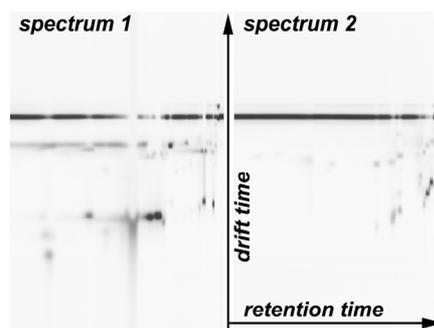


Figure 4. Two different candy flavors taken from the breath measurement. The spectra show obvious differences. Left: Candy flavor 1; Right: Candy flavor 2 (see Table 2 for the complete list of measurements).

R_i for every substance i that needs to be classified. Any uncategorized measurement M would then be evaluated against all references and the minimum distance d_{\min} of all distances $d_i = \Delta(R_i, M)$ would be given as the most likely classification result for the unknown measurement. Meaning that if $d_{\min} = d_k$ then k is the category where M most likely belongs to.

We find that only measurement categories with visible, obvious differences like those in **Figure 4** (like the breath samples with different candy flavors as seen in **Figure 4**) can be separated by this algorithm.

Since the simplest straight forward approach doesn't work and several attempts to align measurements according to the RIP position failed to gain any improvement, it is mandatory to get more insight into the measuring principle.

4. Profile Evaluation

4.1. Evaluation of RIP Profiles

Given the measuring characteristics explained in Section 2.1, one could use the RIP-shape for further analysis and neglect the information given by the points with $t_D > t_{RIP}$. By doing so the magnitude of data points is reduced from millions down to several hundreds. **Figure 5** shows an exemplary extraction of a RIP-Profile along the retention time axis. At fixed drift time (which is the drift time of the RIP peak) the intensity values along the retention time axis are extracted and combined to one profile. This process can be automated because the RIP can be easily found in every GC-IMS measurement. This approach, though being simplifying, allows the use of Artificial Neural Networks (ANNs) for the evaluation and classification of measured spectra. The use of ANNs seems even imperative. As can be seen in **Figure 6** from the overlaid different spectra, it is not possible to define discrimination levels for the signals to distinguish between different measurements without multi-feature analysis, for which ANN are well suited if training data are sufficient.

4.2. Evaluation of Drift Profiles

The extraction of the RIP profiles loses a lot of information that is enclosed in the drift axis. A similar approach would be to extract a profile along the drift time axis and to lose information that is enclosed in the retention time axis. Since there is no exceptional point along the retention time axis like the RIP is on the drift time axis one has to find another extraction formalism differing from the RIP profile extraction. Two possible ways to obtain a drift profile $D(t_D)$ are:

$$D(t_D) = \sum_{t_R=1}^{t_{R,max}} I(t_D, t_R) \quad (3)$$

OR

$$D(t_{D,RIP}) = \max(I(t_D, t_R) | t_R \in [0, t_{R,max}]) \quad (4)$$

Equation (3) is nothing else then just a simple IMS measurement, since all information that was gained during the separation process of the GC column was now again summed up as if it was never separated. Since Equation (4) is a projection and doesn't integrate all information along the retention time axis for every point in our drift spectrum this evaluation formalism was used.

4.3. Virtual Measurements

Depending on the setup, the RIP-Profiles contain from 10^2 up to 10^3 data points. Combined with the drift profile

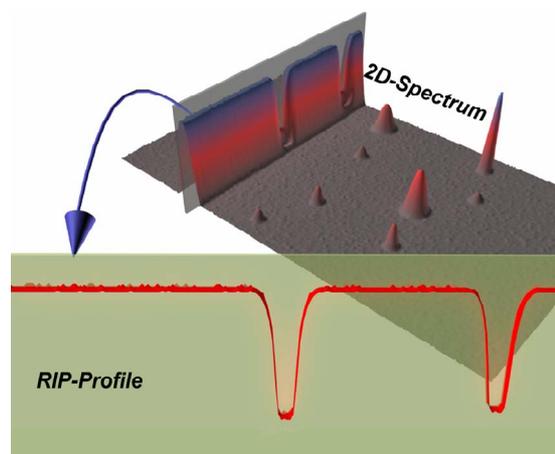


Figure 5. Extraction of a RIP-Profile out of a measured GC-IMS spectrum.

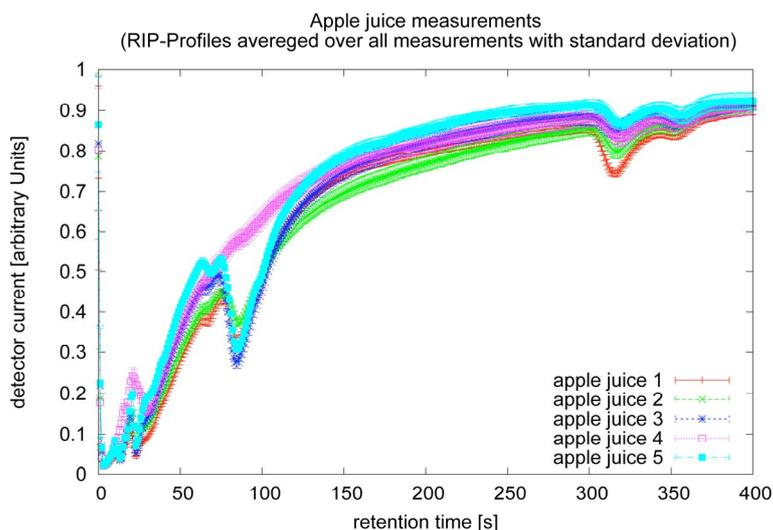


Figure 6. Averaged RIP spectra of 5 different substances in the same measurement class with the according standard deviation.

that contains up to 3000 data points due to higher sampling rate, a total data volume of 4×10^3 points per single measurement is obtained. This still is a powerful reduction of the original spectra by several orders of magnitude. Nonetheless the result is a vector of high dimensionality. Let the dimension of this vector be defined as n and the overall amount of weights in the Network be N . It is obvious that $n \leq N$ must apply. Furthermore it is well known that the amount of examples T for the training of the Network must be greater than the amount of network weights N (with a rule of thumb $T \approx N^2$ [10,11]). This fact implicates that one needs sets of 10^2 - 10^6 measured examples to train a network. With a measuring duration between 3 to 10 minutes this is hardly to be accomplished. Therefore one must consider a way to produce new “virtual” measurements in parametrized manner to simulate the naturally measured scattering.

Generating new datasets by just superimposing white noise over the existing data sets yields only poor results in the training process of networks. In contrast we find that virtual generation by functional approximation and subsequent parameter variation is more promising. As deduced earlier the information about detected substances is enclosed in the RIP profile as dip. Every dip represents a different detected substance. Position, amplitude and general form of every dip are considered to be the most decisive and relevant properties of the RIP profile. Hence one can use exactly these features and distort them slightly to generate new “virtual” measurements for the training set of the ANN. Since slight variations in the detection time (drift or retention time respectively) don't cause gaussian noise on the peaks but shift these peaks and change their shape. In order to produce

variations, the RIP profile was cut in separate dips surrounded by two peaks. These partial profiles

$$F_i(t_R) | (t_{\min,i} \leq t_R \leq t_{\max,i})$$

were superposed with the distortion profiles, where $F_i(t_R)$ is the profile of the i^{th} dip and $t_{\min,i}, t_{\max,i}$ are the retention times of the surrounding peak maxima. It is important to understand that the connecting points do not contain important information since at these points the signal of the RIP is relaxed back to its zero-level. Thus it is ensured that relevant data are varied and methodology is kept free from artefacts. As distortion functions Bézier curves [12,13] were used. A general Bézier curve of order n is defined as

$$B(t) = \sum_{j=0}^n \binom{n}{j} (1-t)^{n-j} t^j P_j \quad (5)$$

As control points $t_{\min,i}, t_{\max,i}$ and $t_{DIP,i}$ are used. Where $t_{DIP,i}$ is the retention time of the minimum in the partial profile. With only 3 control points P_j the Bézier curve simplifies to quadratic form and one gets a parabolic segment as distortion function.

$$B_3(t_R) = (1-t_R)^2 P_0 + 2(1-t_R)t_R P_1 + t_R^2 P_2, \quad t \in [0,1] \quad (6)$$

These partial distortion profiles for the individual dips are joined to one distortion profile $D(t_R)$ to be superposed on the original measurement. In **Figure 7** a RIP profile and the distortion functions in different distortion strengths can be seen. The final “virtual” measurement is just the sum of the measured original profile $R(t_R)$ and the distortion function $D(t_R)$ which is weighted by a random parameter, which is generated randomly for

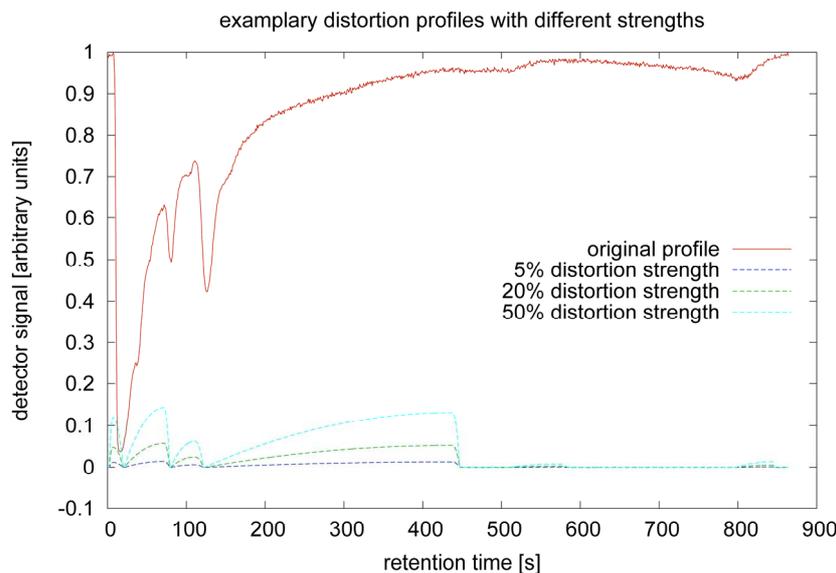


Figure 7. Extracted RIP profile and overlay distortions for the generation of “virtual” measurements.

every “virtual” measurement.

$$V(t_R) = R(t_R) + a \cdot D(t_R) \quad (7)$$

4.4. The Classifying ANN

The first tests of the classification performance of this approach showed poor results (as shown in **Figure 8**).

Further analysis of the problem revealed that the used datasets had many categories which the ANN should discriminate. Modifications in the design structure of the used ANN brought a major improvement to the classification performance. The modified network structure is shown in **Figure 9**. Instead of presenting the example measurement to one single network and training this network to distinguish all categories from each other, the problem set with N categories was divided. Now N Networks were trained to distinguish the category i from all other categories. This approach doesn't reduce the overall ratio of data sets selected for training per network weight. But for every binary network ANN_i which has to decide whether a given measurement is from the category i or not, the ratio of training examples per weight improved since the network needs less weights. An unknown measurement is subsequently presented to all Networks. The ideal case is, that only one of the networks produces the output “yes”. This Network and the RIP profile based method yielded up to 100% classification rate on nearly all presented problem sets. This redesigned architecture (Binary Decision Tree) has another advantage. All datasets were pure measurements of one substance at a time thus every measurement is

definitely assignable to one category. So only one of the n ANNs can have the output “yes” after evaluating the profile vector with all ANNs. The result is a more redundant classification.

To substantiate the found conclusions another test run was made. **Figure 10** shows the comparison between

- the standard training approach with an ANN to distinguish all categories from each other (yellow plot);
- the training of a binary decision tree with virtual training data generated as described in Section 4.3 (blue plot);
- the training of a binary decision tree with virtual data generated just by superposing white noise over the real measurements (green plot).

One must keep in mind that in the binary training mode the evaluation files of $n-1$ categories are combined and to one category so that the ANN can classify these $n-1$ categories against one other category k . So the relation of profiles in category “ k ” and category “not k ” is about 90%. This means that only classification rates that are greater than this “mixRate” are actually classifying something. As can be seen in **Figure 10** only the binary decision tree method reaches classification rates above the mix rate, and therefore shows the best performance. Since the training sets for the binary decision tree were generated according to Section 4.3 this indicates that the principle of generating virtual measurements is correct.

Yet worth mentioning is the fact that the values shown in this plot are averaged values. Actually there were training runs, where the ANN in the binary decision tree architecture yielded classification rates of 100%.

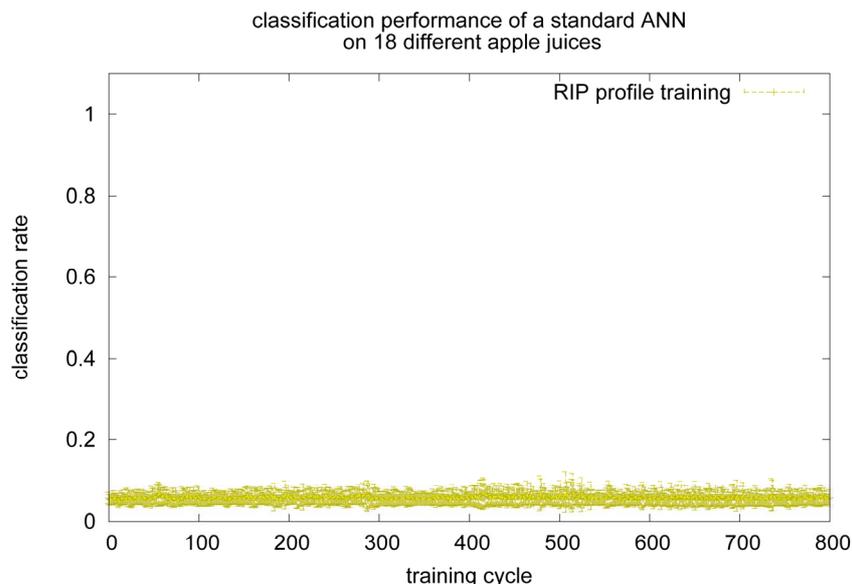


Figure 8. Classification performance of the standard ANN approach on the basis of RIP profiles. Results are poor. A rate of about 5% - 10% is just as good as guessing when trying to classify between 18 possible categories.

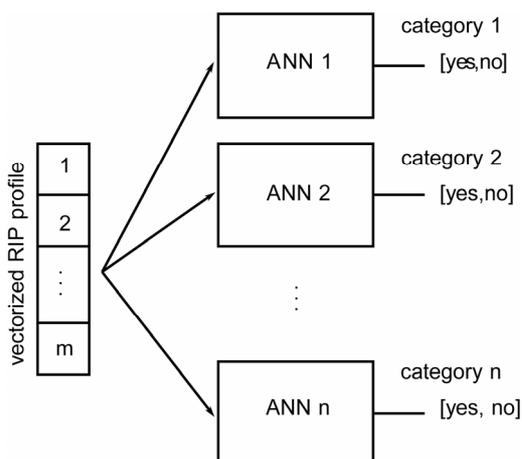


Figure 9. Split ANN architecture to find a redundant decision mechanism as explained in Section 4.4.

5. Refinement of Data Extraction

Though the usage of the RIP Profile seems to be a promising approach and furthermore reduces the data by several orders of magnitude. It skips many useful data though, which is only justified as long as the classification performance is high enough. This is the case for many of the measurements but not for all of them. During the evaluation tests with olive oil measurements this method failed to classify as can be seen in **Figure 11**. The broad bands of standard deviation indicate that the training process doesn't converge in every training attempt to a high classification rate. This is another indication of insufficient discernible training data. The extracted RIP profiles do not carry enough information to train an ANN successfully in the case of olive oil. Since

the RIP extraction loses all information about the drift time, the next logical step to add information from projections onto the drift time axis of the spectra (drift profile). These drift profiles can resolve peaks which are distinct on the drift time axis but have the same retention time and therefore can not be resolved in a RIP profile. Strictly spoken, substances which are not separated by the GC column can't be seen as disjoint features on the RIP profile but they have a chance to be separated during the drift process.

The test of different extraction strategies from the two dimensional spectra reveals that in some cases the RIP Extraction is not the ideal method to obtain training vectors and one should use another extraction technique to obtain less data intensive training sets. **Figure 12** shows the comparison of the average of 10 training runs with

- RIP profiles
- Drift profiles
- RIP + Drift profiles

as training sets. As seen before the training with RIP profiles only shows insufficient classification performance with the olive oil measurements. The training with combined RIP and Drift profiles shows better classification in the average but is very unstable. This is very probably due to the higher dimensionality of the training vector, which reduces the probability to reach a global minimum in the error function of the ANN. In the case of olive oil we find that the training with Drift profiles shows the best classification results. Beside the good classification performance one can see the very narrow band from the standard deviation indicating that this training approach yields very stable results and reaches good classification results in every training run.

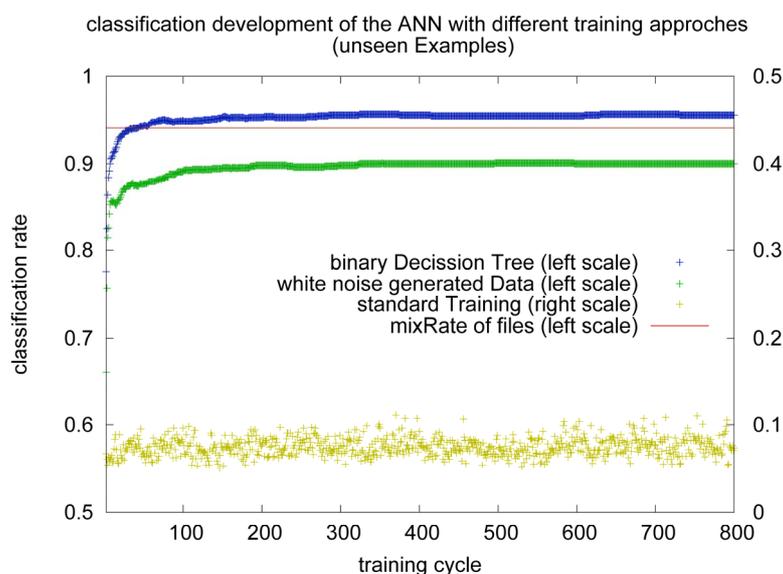


Figure 10. Comparison between different data generation modes and different ANN structures. Green, red and blue curves are plotted on the left axis, the yellow curve is plotted on the right axis.

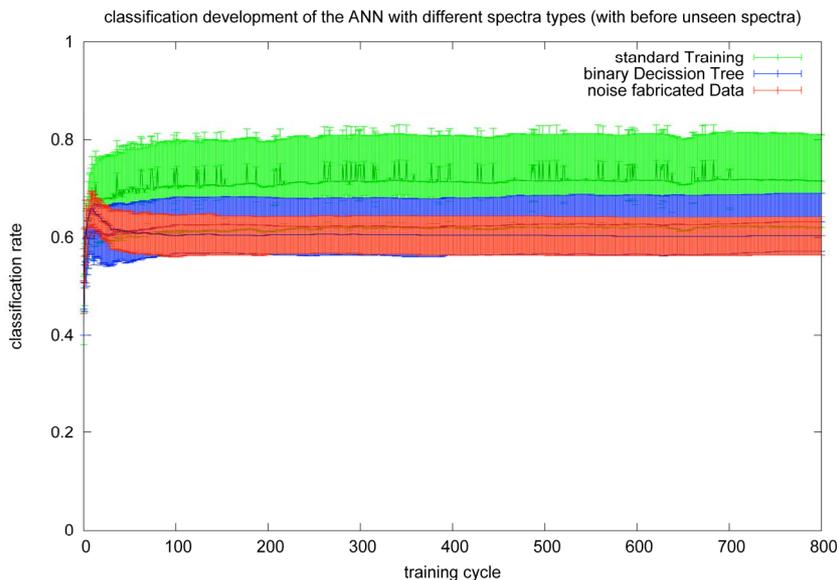


Figure 11. Classification performance of different training and data generation approaches on the basis of the RIP profiles.

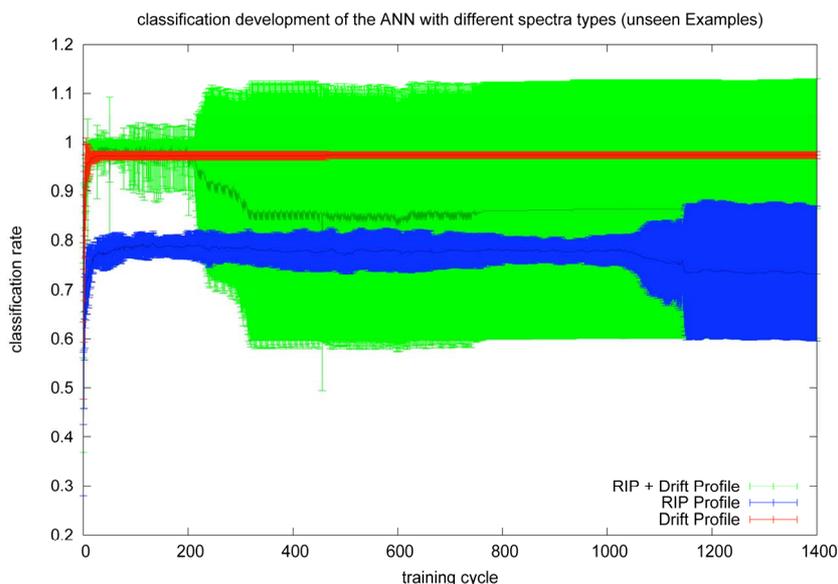


Figure 12. Classification performance of different trainings sets.

6. Conclusions

Various methods have been tested on several datasets. Different approaches of data reduction/extraction have been developed and evaluated as well as different structures of ANNs for the classification of data samples. To reach automated evaluation a peak detection algorithm was developed on the basis of the existing “growing islands” algorithm. This algorithm though developed to generate virtual measurements from RIP profiles is generic enough to detect peaks on any continuous profile where detected information is enclosed in the formation of peaks and dips.

The achieved classification performance of the ANN indicated that the data compression via RIP profile extraction is a powerful method. Out of 6 data sets with about 10 categories to be distinguished in each data set, the network classified 5 of the data sets with a rate of over 97%. A variety of measured substances like different sorts of juices to be distinguished from each other, several soft drinks, various oil sorts and meat in several aging stadiums have been used to evaluate the aforementioned methods. The discussed draw backs of this extraction method with RIP profiles and with drift profiles were observed during the classification tests. The two presented extraction methods showed different results on

the available data sets. In the end at least one strategy for every available classification problem was found, which managed to reach high classification rates. The combined classification results of RIP profiles and drift profiles are encouraging. Since they lay ground for automated evaluation of measurements and possible monitoring applications.

An overview of the reached classification results with the different evaluation strategies is given in **Table 2**. The classification rates given in this table are maximum rates that were reached during several training attempts.

One disadvantage was already mentioned in the introduction. Though the RIP profile contains information about peaks being measured in the GC-IMS spectrum it doesn't show double peaks appearing at the same retention time, which arise from monomer and dimer formation. Another drawback is the loss of information on the drift time of the peaks. This can lead to the creation of RIP profiles without any discriminating information. This is the reason why in some cases the drift profiles achieve better classification results. Since the principle is the same, information appears in form of the peaks, the profiles are interchangeable.

Further Improvements: Although this paper shows great potential in evaluation of GC-IMS measurements with ANNs, there are still improvements possible which are to be investigated. Other data extraction and data reduction methods should be considered.

Furthermore tests with bigger data sets should be implemented to investigate the convergence and the classification behavior of the ANNs and the according data extraction strategies. Only statistically more relevant data sets are able to determine the stableness and usability of this method.

Table 2. Classification rates for all measured data with different training profiles used to train the ANN.

Measurement	Classification rate	
	Drift profiles	RIP profiles
Cola	91.0%	96.3%
Juice	98.6%	88.8%
Rice	100%	100%
Olive oil	98.9%	76.4%
Candy flavors	100%	100%
Aging meat	96.6%	100%

7. Acknowledgements

This work has been supported by the German BMBF, contract 01IS09046A.

REFERENCES

- [1] G. A. Eiceman, "Ion Mobility Spectrometry," Taylor and Francis Group, London, 2005.
[doi:10.1201/9781420038972](https://doi.org/10.1201/9781420038972)
- [2] M. Tsvet, "Physical Chemical Studies on Chlorophyll Adsorptions," *Berichte der Deutschen Botanischen Gesellschaft*, Vol. 24, 1906, pp. 316-323.
- [3] F. W. Cohen and M. J. Karasek, "Plasma Chromatography—A New Dimension for Gas Chromatography and Mass Spectrometry," *Journal of Chromatographic Science*, Vol. 8, No. 6, 1970, pp. 330-337.
- [4] C. P. Cram and S. N. Chesler, "Coupling of High Speed Plasma Chromatography with Gas Chromatography," *Journal of Chromatographic Science*, Vol. 11, No. 8, 1973, pp. 391-401.
- [5] B. Bodeker and J. I. Baumbach, "Analytical Description of IMS-Signals," *International Journal for Ion Mobility Spectrometry*, Vol. 12, No. 3, 2009, pp. 103-108.
- [6] J. Xu and B. W. William, "Monte Carlo Simulation of Ion Transport in Ion Mobility Spectrometry," *International Journal for Ion Mobility Spectrometry*, Vol. 11, No. 1-4, 2008, pp. 13-17.
- [7] J. I. Baumbach, S. Sielemann and P. Pilzecker, "Coupling of Multi-Capillary Columns with Two Different Types of Ion Mobility Spectrometer," *International Journal for Ion Mobility Spectrometry*, Vol. 402, No. 1, 2012, pp. 489-498.
- [8] C. L. P. Thomas, "Sensitivity and Resolution in Gas Chromatography-Ion Mobility Spectrometry," *International Journal for Ion Mobility Spectrometry*, Vol. 4, No. 2, 2001, pp. 62-68.
- [9] Z. Xie, S. Sielemann, H. Schmidt and J. I. Baumbach, "A Novel Method for the Detection of MTBE: Ion Mobility Spectrometry Coupled to Multi Capillary Column," *International Journal for Ion Mobility Spectrometry*, Vol. 4, No. 1, 2000, pp. 77-86.
- [10] R. Rojas, "Neural Networks," Springer, Berlin, Heidelberg, 1996.
- [11] R. L. Harvey, "Neural Network Principles," Prentice-Hall International, Upper Saddle River, 1994.
- [12] F. Gerald, "Curves and Surfaces for Computer-Aided Geometric Design," 4th Edition, Elsevier Science and Technology Books, Amsterdam, 1997.
- [13] D. Lasser, "Bernstein-Bezier Representation of Volumes," *Computer Aided Geometric Design*, Vol. 2, No. 1-3, 1985, pp. 145-149.

Nomenclature

Symbol	Description
t_R	Retention time
t_D	Drift time
$t_{R,max}, t_{D,max}$	Measurement specific maximal values of retention or drift time
$I(t_R, t_D), R_i(t_R, t_D), M(t_R, t_D)$	2-dimensional matrix representation of a MCC-IMS measurement, in some cases this discrete data representation is regarded as continuous, which is justified by the high resolution due to a high sampling rate in the available measurements
$Q(t_R)$	Complete charge accumulated at one certain retention time by summing up charge at all drift times
$\Delta(R_i(t_R, t_D), R_j(t_R, t_D))$	Function mapping 2 matrices R_i and R_j of identical dimension to a scalar value $\Delta: R_i(t_R, t_D) \times R_j(t_R, t_D) \rightarrow \mathbb{R}$
$D(t_D)$	1-dimensional extraction of a profile along the drift time axis
$F_i(t_R)$	Partition of a 1-dimensional profile into piece wise profiles that contain only one extremum which is not located on the edge points of this profile
$B_n(t)$	Bezier curve of order n defined by n + 1 control points