International Journal of Mass Spectrometry 307 (2011) 245–252

Contents lists available at ScienceDirect

International Journal of Mass Spectrometry

journal homepage: www.elsevier.com/locate/ijms

Automatic segmentation of tissue sections using the multielement information provided by LA-ICP-MS imaging and *k*-means cluster analysis

A.M. Oros-Peusquens[a,*], A. Matusch[b], J. Sabine Becker[c], N.J. Shah[a,d]

[a] Institute of Neuroscience and Medicine, Physics of Medical Imaging, INM-4, Research Centre Jülich, 52425 Jülich, Germany
[b] Institute of Neuroscience and Medicine, INM-2, Research Centre Jülich, 52425 Jülich, Germany
[c] Central Division of Analytical Chemistry, Research Centre Jülich, 52425 Jülich, Germany
[d] Department of Neurology, School of Medicine, RWTH Aachen University, JARA, 52074 Aachen, Germany

ARTICLE INFO

Article history:
Received 17 January 2011
Received in revised form 29 March 2011
Accepted 30 March 2011
Available online 8 April 2011

Keywords:
Brain imaging
Element bio-distribution
LA-ICP-MS
Clustering
k-means
Segmentation
Brain lesions
Image analysis

ABSTRACT

Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS) is an established and powerful tool to analyse the distribution of elements in tissue sections. Among other applications, the technique is expected to play a central role in the understanding of normal and pathological element distributions in brain tissue.

In order to interpret the distribution of elements such as the bio-metals Cu, Zn, Fe and Mn and proceed to an element-based comparison between groups of samples, it is necessary to anatomically parcel the tissue section into regions-of-interest and to average element signals across these regions. This categorization, also termed segmentation, can be done manually, but the support of automated procedures is highly desirable, especially in order to (1) identify groups of pixels with similar elemental fingerprint, termed clusters, and to determine which degree of discrimination is reasonable; (2) segment anatomical structures known to exhibit substructure but without clearly defined borders, such as the healthy cortex, zones of tumours or ischemic lesions, in an observer-independent way; and (3) to investigate correlation between the distribution of elements in tissue and phenomena which incorporate contributions from several elements in a convoluted way, such as the origin of contrast in magnetic resonance imaging (MRI) experiments.

The multi-parametric information provided by LA-ICP-MS lends itself naturally to multivariate analysis. This study provides a new way to synthesise the information distributed over many element images by demonstrating the possibility to segment tissue sections into biologically meaningful substructures. This data-driven, observer-independent categorization was based on *k*-means clustering. The optimal number of clusters was determined based on the silhouette method.

Segmentation of healthy tissue resulted in a set of substructures in perfect congruence to the anatomical architecture. Segmentation of ischemic lesions identified a number of regions with different fingerprints of C, P, Fe, Cu and Zn deposits. Clustering provides a promising way of combining the information present in several element images and reveals structure which is not entirely present in any isolated image.

As a useful by-product of this study we have found a promising method for investigating the optimal line length within the process of image reconstruction from the continuous stream of raw data points. Images were characterized by their tensor of inertia, in image- as well as in Fourier dual-space (*k*-space) and changes in the ratio of the intrinsic moments of inertia or the orientation of the principal axes were found to closely describe the optimum orientation. The first results look very encouraging, but the method must be extensively tested before it can be used as an automatic procedure.

In conclusion, cluster analysis of mass spectrometric imaging data allows one to define the fingerprint element distribution of different anatomically or functionally distinct regions and opens a new way for the study of correlation between the element distribution and related phenomena.

1. Introduction

Several medical applications can benefit enormously from the capability of laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) to provide spatially resolved quantitative information regarding the element composition of the brain [1–6].

* Corresponding author. Tel.: +49 2461612107.
E-mail address: a.m.oros-peusquens@fz-juelich.de (A.M. Oros-Peusquens).
URL: http://www.brainmet.com/ (J.S. Becker).

1387-3806/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2011.03.014

The questions most accessible to mass spectrometric imaging (MSI), an inherently 2D technique, are those that can be answered by measuring a few slices only. This is due to the long measurement time in the range of 4 h/cm^2. In comparison, Magnetic Resonance Imaging (MRI) is a fast, non-invasive imaging method with excellent and versatile soft tissue contrast and easily obtainable 3D coverage of the entire sample specimen. However, MRI suffers from limited specificity regarding the microscopic source of contrast generation. Consistently combining element analysis on a small number of slices with information obtained from whole brain, *in vivo* or *ex vivo*, could provide valuable insight into the origin of contrast in MRI. In addition, by correlating the contrast of the two methods, it might become possible to extrapolate the information extracted from a few slices by element analysis to the whole brain. To this aim, a pixel-by-pixel comparison between element concentration and the parameters used to describe the MR contrast (magnetisation density, relaxation times) would be highly desirable. This is, however, a difficult task due to the considerable deformations induced in the process of slicing the brain and mounting the slices onto slides. An alternative approach is based on partitioning the tissue section into regions-of-interest (ROI), a process also termed segmentation, and investigating the correlation between element concentrations and MR parameters across and between ROIs.

The partitioning of tissue into ROIs is also an important step in the analysis of LA-ICP-MS data for any application, especially for group comparisons. This categorization can be done manually, but the support of automated and observer-independent procedures is desirable. In general, the ROIs should reflect the anatomical architecture of the sample and be as homogeneous as possible regarding for example MR contrast and element distribution. Depending on the specific question, parcelation can be done on a finer or a coarser scale.

As an example, the dependence between the transversal relaxation rate $R2^*$ of water protons and iron concentration might well be linear in different regions [7], but a different proportionality constant might be expected if – for example – the granularity of iron deposition, which can be different from region to region, plays a role. A sufficient number of points characterising the correlation is required if one is to identify such effects without prior information about their existence and nature. Therefore, in this special application a large number of points inside homogeneous ROIs is required in order to investigate the correlation between the element concentration and MR contrast quantitatively.

When the question is to compare specific diseases or treatments, it is desirable that the parcelling is complete but not redundant resulting in as few ROIs as possible. Here, automated procedures can help to reasonably reduce the plethora of described anatomical microstructures. Especially with respect to layered structures with smooth transitions between layers – such as the neocortex – an automated procedure may outperform an approach based on visual inspection of the contrast. Furthermore, automated procedures can help in the understanding of the degree of similarity of anatomical regions with respect to their elemental signature and thus establish a hierarchy of structures that are more or less cognate.

Cluster analysis, which consists in finding partitions in data [8], has become an increasingly popular part of data analysis. Cluster analysis is a method for finding a reasonable grouping of data points characterized by a set of parameters. In the simple case of two parameters one would plot every data point of the element-image in a coordinate system with e.g., the Zn signal on one and the Cu signal on the other axis. Within the inhomogenous distribution in this *xy*-plot usually clouds of data points can be identified and every data point can be assigned to a given "cloud" – a cluster. A cluster analysis of data points characterized by a large number of parameters can be understood as transposition of the 2D process to an *n*-dimensional space.

Clustering has been applied to areas as different as determining family trees e.g., of monkeys or bacteria based on DNA-sequences in taxonomy, blood cell sorting or establishing specific diagnostic criteria in medicine, in geology, business, engineering systems and image processing, (see for example Refs. [9–11]).

Since mass spectroscopic imaging inherently offers a multiparametric characterization of tissue with a large number of parameters, it is ideally suited for cluster analysis. The application of cluster analysis to MALDI-ToF-IMS data yielding new and anatomically relevant information has been shown very recently [12]. As a first step towards a systematic and quantitative comparison between element distribution and MR contrast, we investigate in this paper the application of clustering analysis to LA-ICP-MS element images. We exemplify the method on three Wistar rat brain slices: one from a formalin fixed brain with a stroke induced by transient mean cerebral artery occlusion (tMCAO), one from a formalin fixed brain with a stroke induced by photo-thrombosis, and one from a fresh brain obtained from a healthy rat.

The proposed clustering analysis should be widely applicable and is expected to become a standard step in the evaluation of element images obtained by LA-ICP-MS for numerous applications.

2. Materials and methods

2.1. Rat brains

All the experiments were performed in accordance with the German law for the protection of animals. Permission for animal experiments was obtained either from the University of Duesseldorf (stroke model) or from the Research Centre Juelich (healthy animal). Two male Wistar rat brains originated from a study of inflammation following stroke reported in [13]. Stroke was induced by either photothrombosis (PT) or transient main cerebral arterial occlusion (tMCAO) and the animals were sacrificed 7 days after stroke. The brains were immediately excised and immersed in 4% formalin solution. After extensive fixation (3.5–4.5 years) the brains were either cryocut in 60 μm thick slices (tMCAO) or embedded in paraffin and then cut in 14 μm thick slices (PT). One male Wistar rat of 250 g body weight was sacrificed under isofluorane anaesthesia, the brain excised and shock frozen in liquid 2-methylbutane at −50 °C and subsequently cryocut in 30 μm thick slices.

The cryocut sections were dried and stored at room temperature for at most 2 weeks before LA-ICP-MS measurements. The slices obtained from the paraffin-embedded brain were deparaffinized with xylene prior to LA-ICP-MS.

2.2. LA-ICP-MS

All the LA-ICP-MS measurements employed a commercial laser ablation (LA) system (UP 266 from NewWave, Freemont, CA, USA) coupled to a quadrupole-based inductively coupled plasma mass spectrometer (XSeries2, Thermo Fisher Scientific, Germany). The laser ablation system was operated with a frequency-quadrupled Nd:YAG laser at a wavelength of 266 nm, a repetition frequency of 20 Hz and a laser spot diameter of 90–120 μm, depending on the sample type. The spot size for the paraffin-embedded PT brain slice, which was substantially shrunken with respect to fresh tissue, was 90 μm; no spaces were left between successive lines. The cryocut slice from the tMCAO brain was measured with 100 μm spot size and 60 μm gap, the cryocut slice of shock-frozen fresh brain with 120 μm spot size and 50 μm gap between the lines.

Brain tissue was ablated line-by-line at scan speeds of 30–60 μm/s and transported by argon to the ICP source. Ioni-

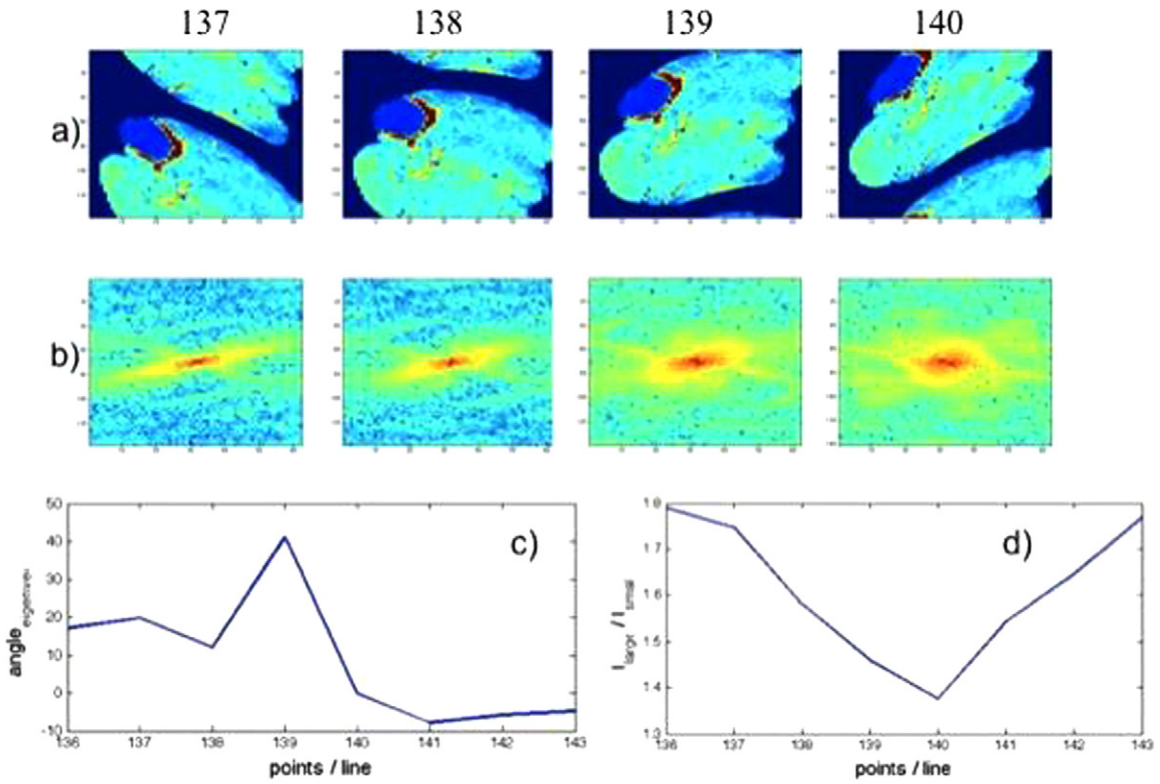

Fig. 1. (a) Images obtained with different reshaping parameters (number of points-per-line P = 137, 138, 139, 140). Severe shearing effects appear in images with erroneous reshaping. (b) The absolute values of the Fourier dual space (k-space) of the images shown in (a). (c and d) Possible indicators of the best approximation of the effective number of points-per-line. Part (c) shows the angle between the orthogonal system defined by the intrinsic axes of inertia and the Cartesian system. The tensor of inertia is calculated in image space. Part (d) shows the ratio between the larger and the smaller moments of inertia when the tensor of inertia is calculated from the k-space distribution of intensity.

sation was obtained in the Ar plasma and the resulting single charged cations were analysed in the quadrupole mass analyzer with respect to their mass/charge ratio. Complete ablation of the entire thickness of the sections was achieved by the laser in all cases. The content of several elements was quantified by comparison of the number of counts/s from tissue with the number of counts/s obtained from homogeneous matrix-matched standards measured in the same run under the same conditions. These aspects are described in more detail in Refs. [1–4,6].

2.3. Data analysis

2.3.1. Reshaping

The raw data output is a continuous list of count rates for each preselected mass/charge value. In order to maximise the sensitivity of the method, the sampling time per element is set to the maximum value of 0.1 s. Each datum point is obtained from a pixel with a width (in μm) given by the product of the sampling time (i.e., 2.5 s for 25 m/z species) and the x-scan speed (in μm/s) of the xyz-stage. The pixel height corresponds to the laser spot diameter if, as in the present setting, the laser spots corresponding to different lines do not overlap. The x-speed is usually chosen to be slightly lower than the quotient of line distance and sampling time, resulting in nearly quadratic pixels.

Concerning reconstruction of the list-mode data, a simple reshape operation should create an image of that isotope's distribution from the string of values for a single mass/charge value. However, as laser repetition frequency, the lateral propagation of the sample stage, and the cycle frequency of the mass spectrometer are not synchronized, the effective number of points per line in the element images is not an integer. Even a slightly erroneous estimate of the "number of points" (for example, 138 instead of 138.65) leads to images with severe shearing. This is easily seen in Figs. 1 and 2, where the reshaping is illustrated using the phosphorus distribution.

We sketch here a simple method to determine the correct number of points per line with a two-digit accuracy. If proven to work in general, the procedure described below could easily be extended to run automatically and also with higher accuracy, if necessary. We have used the property of the Fourier transform to give information about the spatial frequencies in an image, with the aim of finding the image with the simplest pattern. The frequency content and the distribution of intensity in the Fourier as well as image space change with changing reshaping factor. Since the total intensity and the centre of mass (zeroth and first moments of the distribution) do not depend on the number of points-per-line, we have investigated the sensitivity of higher moments of the intensity distribution to reshaping. This was done in Fourier as well as in image space by means of the tensor of inertia, which reflects the intensity distribution in the object with respect to given axes.

The raw data (Excel table) were transformed to ASCII and imported in Matlab (www.mathworks.com). If the number of lines scanned by the laser (L, integer) is known, it is used to estimate the starting value for the number of points/line (P, integer), otherwise an equal number of lines and points/line is taken as a starting value, such that their product gives the closest integer to the original string length. The string is reshaped to a $P \times L$ matrix, which is Fourier transformed. If a significant part of the intensity lies outside the central cross (lines at $x = 0$ and $y = 0$) in the Fourier transformed matrix, the P and L values are modified and the procedure is repeated. The inertia tensor was calculated both in image space and in the Fourier dual space (k-space) using the absolute value of the transformed matrix. The origin was set to the centre of the matrix, (P/2,L/2), and the dimensions were normalized such that the coor-

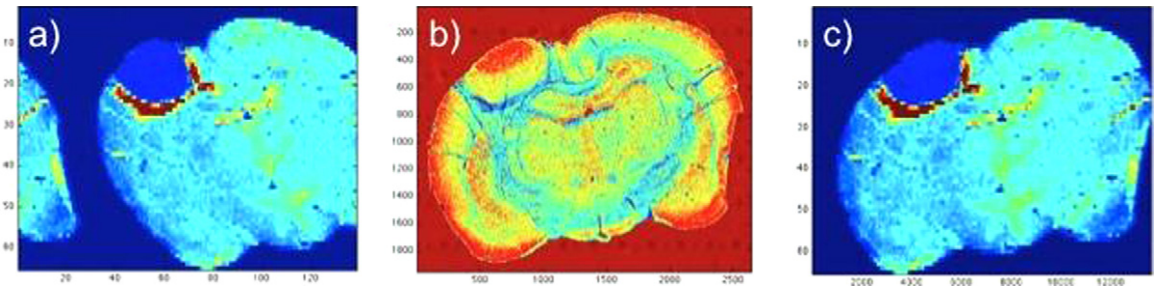

Fig. 2. Influence of reshaping: (a) 138 points-per-line; (b) microphotograph; (c) 138.65 points-per-line. The proper choice of (non-integer) number of points-per-line should correct for shearing and provide an image which is geometrically identical with the microphotograph of the slice mounted on the glass slide before LA-ICP-MS (part b). These conditions are far from being fulfilled by image (a), where an integer number $P = 138$ of points-per-line was chosen, but are well met by the image shown in part (c). The image was obtained with an optimal number of points-per-line determined with two-digit precision, $P = 138.65$.

dinates ran from −0.5 to 0.5 in both x and y. The equivalent of mass for inertia was taken to be the image intensity. The inertia tensor was calculated with respect to the x and y axes, in image as well as in k-space (using the absolute values) and then diagonalized. This was repeated for a number of reshaping factors $P \times L$. The images and corresponding k-spaces are shown in Fig. 1a and b, respectively, for dimensions of $137 \times 62, \ldots, 140 \times 62$. The influence of the reshaping is clearly seen in both image and k-space. Several possible indicators of the correct number of points/line were investigated. In image space the angle of rotation leading from the Cartesian system of coordinates (x, y) to the principal axes of inertia was found to be the best indicator and is shown in Fig. 1c (P-values from 136 to 142). It is seen to change sign between the values of 138 and 140 points per line, as the tilt in the images (Fig. 1a) changes orientation. In k-space, the transition is best reflected in the ratio of the intrinsic moments of inertia, shown in Fig. 1d, again reflecting a change in the tilt of the intensity going through a more circular structure. However, the details of the intensity distribution of the correct k-space seem to play a non-negligible role in this case, and the transition is found at 140 points/line instead of 139. Further investigations of the best indicator are required.

Once the closest integer values, $P1$ and $P2$, of the effective points-per-line value have been found, further fine-tuning is usually required. The string of values acquired for each element is inflated to 10 times its initial size by replicating each value 10 times. Reshaping with numbers of points-per-line between $10 \times P1$ and $10 \times P2$ can be investigated with the method outlined above, and the integers $P1_{10}$ and $P2_{10}$ closest to the effective number of points-per-line are again found. In the next step, the initial string of values is inflated 100 times and the search is repeated. The influence of increasing the precision of the determination of the effective number of points-per-line is illustrated in Fig. 2.

2.3.2. Cluster analysis

Clustering has been performed using the K-means algorithm [14] which has established itself as one of the most popular clustering techniques due to its simplicity and fast convergence. We describe the main ingredients below.

A particular clustering by means of k clusters, each denoted by $\mathbf{C_k}$ represents a partition of the data such that each point belongs to one cluster only. The intraclass inertia of the resulting partition is, for one cluster, the average squared distance from a point $\mathbf{z_j}$ to its cluster centre $\mathbf{c_k}$, and for the whole partition:

$$I_W = \frac{1}{N}\sum_{k=1}^{K}\sum_{j \in \mathbf{C_k}} d^2(\mathbf{z_j}, \mathbf{c_k}) \tag{1}$$

The interclass inertia is the average squared distance from a cluster centre $\mathbf{c_k}$ to the centre of gravity $\bar{c}$ and for the whole partition becomes:

$$I_B = \frac{1}{N}\sum_{k=1}^{K} \left|C_k\right| \cdot d^2(\mathbf{c_k}, \bar{c}) \tag{2}$$

where $d^2(a, b)$ is the squared distance between vectors a and b, $|C_k|$ is the number of elements in cluster C_k, and $\bar{c} = \sum_{k=1}^{K} \left|C_k\right| / N \cdot c_k$ is the weighted average of the cluster centres c_k. Ideally, the intraclass inertia will be minimized, resulting in homogeneous clusters, and the interclass inertia will be maximized, such that the clusters are as well separated as possible.

For a given number K of clusters, the intraclass inertia is iteratively minimized by the steps:

- 0. Choose initial K centres $c_k^{(0)}$. This is usually done in an arbitrary way.
- 1. Assign each data vector x_j to the cluster C_k with the nearest centre $c_k^{(i)}$. The assignment is based on a distance metric $d(x_j, c_k^{(i)})$.
- 2. Determine new cluster centres $c_k^{(i+1)}$ such that the sum of the squared distances from all points in C_k to the new cluster centre is minimized. This is ensured by setting the centre to the mean over all the members of a cluster.
- 3. Increment i and go to step 1 until the partition is stable.

Both steps 1 and 2 decrease the within-class inertia, so that the algorithm converges in a finite number of steps. The final result, however, can depend strongly on the initialisation. In order to reduce or eliminate this dependence, it is common to include heuristics based on a local search, in which centres are swapped in and out of an existing solution, typically at random. Such a swap is accepted if it decreases the between-class inertia, otherwise it is ignored. Alternatively, the procedure (steps 0–3) is repeated several times with different – and random – initialisations.

A shortcoming of the k-means clustering method is that the number of clusters must be defined in advance. A useful tool in deciding which number of clusters describes the data better is provided by the silhouette distribution.

The silhouette value for each point is a measure of how similar that point is to points in its own cluster (how compact the cluster is) compared to points in other clusters (how well separated the clusters are). It is defined as [11]:

$$S(i) = \frac{[\min(b(i,k)) - a(i)]}{\max(a(i), \quad \min(b(i,k)))}$$

where $a(i)$ is the average distance from the ith point to the other points in its cluster, and $b(i,k)$ is the average distance from the ith point to points in another cluster k.

The value of the silhouette coefficient of a point varies between −1 and 1. A value near −1 indicates that the point is clustered badly. A value near 1 indicates that the point is well-clustered. To evaluate

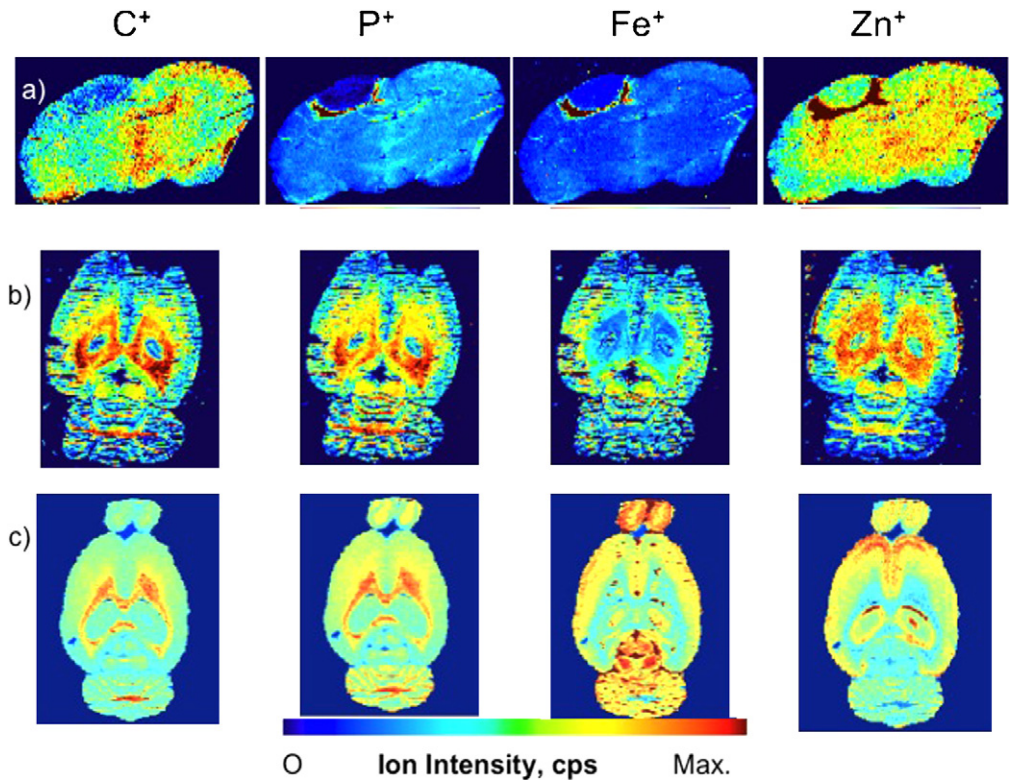

Fig. 3. Distribution of C, P, Fe and Zn in rat brain slices obtained by LA-ICP-MS. The slices describe different physiological conditions and were obtained with different sample preparation schemes: (a) slice obtained from a brain with stroke induced by photothrombosis (PT). After several years of formalin fixation the brain was embedded in paraffin and cut in 40 μm-thick slices which were mounted on glass slides. Before laser ablation, deparaffination was performed; (b) slice obtained from a brain with stroke induced by transient Main Cerebral Arterial Occlusion (tMCAO). After several years of formalin fixation the brain was cryocut and the slices mounted on slides for LA-ICP-MS. (c) Slice obtained from a healthy brain. The brain was excised immediately after death and shock-frozen. After cryocutting, the slices were mounted on glass slides and imaged by LA-ICP-MS. Storage in aqueous solution leads to a wash-out of Zn deposits from grey matter e.g., the hippocampal area CA3 and preservation of Zn in white matter deposits resulting in an artificial pattern with higher white matter concentrations whereas the distribution of Fe, C and P is preserved. The several chemical steps associated with paraffin embedding seem to accentuate the deviation of the element distribution from the native one.

the quality of a clustering one can compute the average silhouette coefficient of all points.

The clustering analysis was performed as follows. In a first step, each element image was normalized to the median value of its intensity distribution, such that all element distributions were compared on a common (and arbitrary) scale. The Euclidian distance was chosen, implying that all elements were given equal weights in determining the clusters.

The *k*-means clustering analysis was performed using Matlab (http://www.mathworks.com), which is a well-established commercial package for technical computing, as well as ImageJ (http://rsbweb.nih.gov/ij/), an open source, Java-based software. The algorithm implemented in ImageJ is based on methods described in Ref. [15]. The same elements were included in the analysis for the cryocut slices (tMCAO and fresh brain) and a reduced number of elements for the paraffin-embedded brain (PT). The distribution of elements was found to be less structured in the latter case, and for the cluster analysis we chose the few distributions with pronounced inhomogeneity either in the lesion or in the normal tissue. The number of clusters was varied from 4 to 12 in each case. For each number of clusters the clustering procedure was repeated 10 times, in order to minimise the dependence of the solution on initialisation. The silhouette was determined for each point and number of clusters and the mean value over the whole slice was calculated.

In the particular case of lesion induced by photothrombosis, since no anatomically based validation of the substructure found by clustering in the rim of the lesion is available, an additional procedure was carried out. We have investigated the stability of the regions found in the rim of the lesion by reducing noise in the element images used for clustering. The filter used was that proposed by Perona and Malik [17,18] based on anisotropic diffusion of information. The implementation was that provided by ImageJ, with the parameter which most pronouncedly influences the noise filtering set to $k = 5$ or 10.

3. Results and discussion

The methods reported here were aimed at applications of LA-ICP-MS to brain research, but are of general validity.

3.1. Image reshaping

Reconstructing the proper shape of the slices analysed by LA-ICP-MS facilitates a comparison with other methods. As an example, block-face photographs before cutting and/or microscope photographs of the brain slices after cutting can be used to identify the anatomy of the slice with high resolution and familiar contrast. Slices adjacent to the ones analysed by LA-ICP-MS can be processed by various histological stains, immuno-histochemistry or microscopy using fluorescent dyes, for example. Furthermore, imaging of the brain using various medical imaging methods before the animal is sacrificed, or on the fixed intact brain, can help investigate a variety of diseases and can be used in longitudinal studies. For a detailed comparison of the element distribution and any of the methods mentioned above, a good coregistration between the slices is essential. These aspects are illustrated in Fig. 2. The phosphorus image obtained with 138.65 points-per-line is compared

Fig. 4. Clustering of brain tissue based on element analysis on a slice of rat brain with stroke lesion, induced by transient Main Cerebral Arterial Occlusion (tMCAO). The tissue was fixed and cryocut. Clustering was performed with the *k*-means algorithm as described in the text and led to an optimal partition with 10 clusters (left-hand side). For better identification of the anatomy we show on the right a similar rat brain slice with Nissl stain taken from [16]. Similar to results on native tissue (Fig. 5), an anatomically meaningful parcellation resulted from clustering. A good separation between the cortex, layers I–IV (violet and red); cortex, layers V–VI (yellow) and white matter (black) was possible. Inside the white matter, the hippocampal structures fascia dentata and cornu ammonis can be seen. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

to the photomicrograph taken before ablation (Fig. 2b). Clearly, the match between the images improves substantially when the number of points per line is varied from 138 (Fig. 2a) to 138.65 (Fig. 2c).

3.2. Segmentation of tissue using cluster analysis and LA-ICP-MS data

Since each mass spectrometric imaging measurement results in a multitude of images reflecting largely independent aspects of tissue, the method is ideally suited for clustering analysis.

Clustering was initially performed using a broad set of 12 elements which yielded rather high signal in the healthy brain tissue

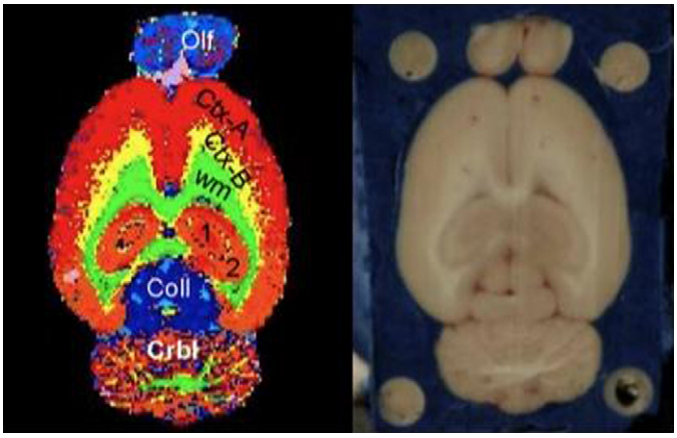

Fig. 5. Clustering of brain tissue based on element analysis on a slice of healthy rat brain, shock frozen immediately after excision and cryocut. Twelve clusters were used for the automatic segmentation. An anatomically meaningful parcellation resulted including the structures bulbus olfactorius (Olf); cortex, layers I–IV (Ctx-A); cortex, layers V–VI (Ctx-B); white matter (WM); plate of four collins (Coll); cerebellar grey matter (Crbl); hippocampus fascia dentata (1) and the cornu ammonis part 1–3 (2). For comparison, the block-face image of the brain and calibration standards embedded in blue stained polyethylenglycole is shown on the right.

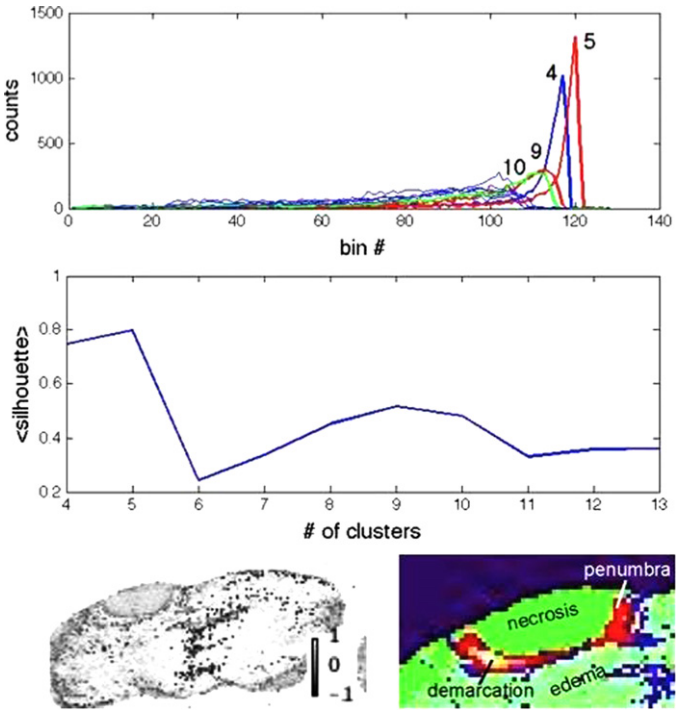

Fig. 6. (a) Overlay of histograms of the silhouette measure for different number of clusters. Thick lines evidence the partitions with a high mean value of the silhouette: partition with 4 clusters (blue), 5 clusters (red), 9 clusters (red), and 10 clusters (green). The range of silhouette values which are histogrammed are between −0.6 and 1 in all cases. The *x* axis represents the bin number (in terms of silhouette values, the *x*-axis unit is 1.6/128/bin). The mean value of the silhouette over the whole slice is plotted in part (b) as a function of the number of clusters. For the optimal partition $N = 9$, the distribution of silhouette values over the slice and the nine colour-coded clusters are shown in parts (c) and (d), respectively. It becomes now possible to discriminate several distinct concentric zones around the central necrosis of the lesion. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

and/or the lesion (C, Na, Mg, P, S, K, Mn, Fe, Cu, Zn, Pb, U). For the PT brain, a restricted set of elements was used (C, P, Fe, Cu, Zn, Pb, U). The latter set was selected using criteria of high SNR and lack of contamination from compounds in at least one isotope. Since the results were found to be qualitatively similar for the tissue samples investigated in our study, we will restrict the discussion to the limited set of elements. Images of C, P, Fe, and Zn are shown in Fig. 3 for all three tissue samples.

We would like to stress the fact that all elements included in the clustering analysis were given equal weights. This should be modified if the transport/accumulation mechanism is found to be the same or strongly correlated for different elements. Ideally the clustering analysis would include independent observations only. An advantage of using this normalization is that quantification of element concentration, which can be difficult, is not essential for clustering analysis. The high quality of the element data, however, is essential. If drifts occur during the rather long measurement time, the image inhomogeneity is reflected in the clustering. In this case, an intensity correction has to be performed.

The result of the *k*-means clustering is shown in Fig. 4 for a slice of fixed brain with tMCAO, in Fig. 5 for a slice of fresh brain and in Fig. 6d for a slice obtained from the fixed brain with photothrombosis.

Without a careful quantitative analysis of the element concentration no effect of the tMCAO-induced stroke is visible in Fig. 4. A left–right asymmetry in tissue properties might be expected, due to the fact that stroke mainly affects one hemisphere. A more detailed analysis is in progress. On this slice, clustering provides very good

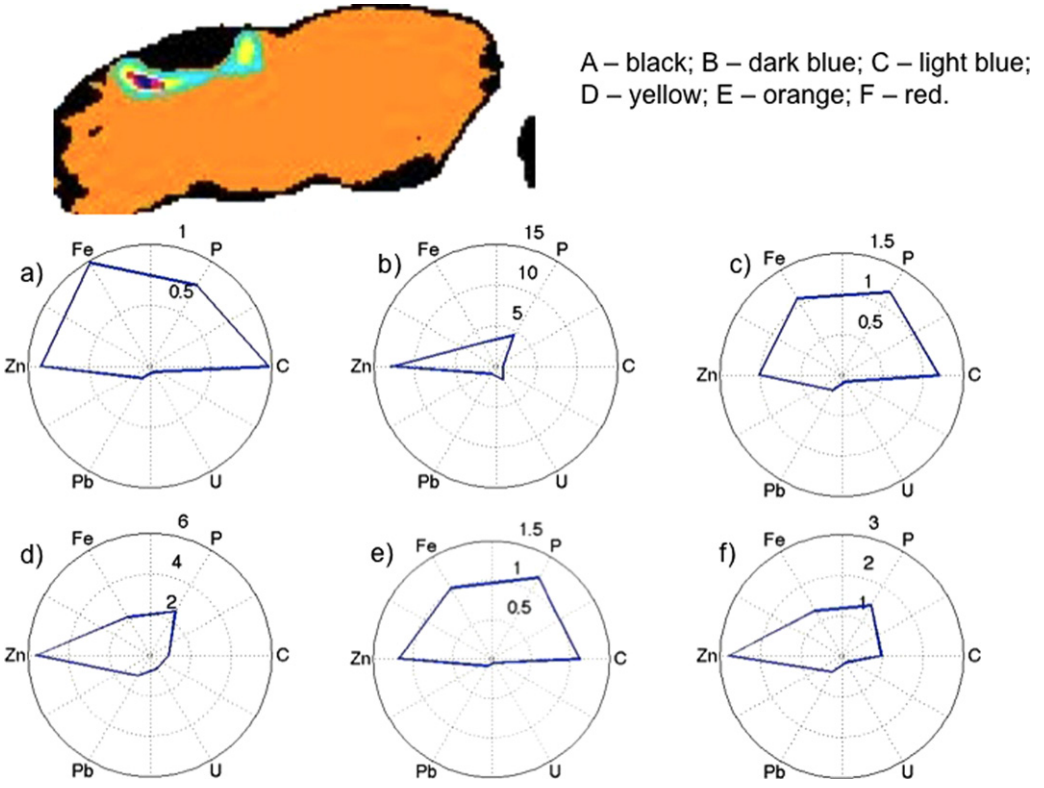

Fig. 7. (a) Result of clustering analysis performed on data filtered with the Perona–Malik method based on anisotropic diffusion. The regions are: cluster 1 (black): necrosis; cluster 2 (dark blue): region inside the rim; cluster 3 (light blue): region inside the rim; cluster 4 (yellow): region inside the rim; cluster 5 (orange): normal tissue; cluster 6 (red): region inside the rim. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

white matter (WM) – grey matter (GM) segmentation and good delineation of hippocampal structures. While the distribution of elements is different from the native one, as seen in Fig. 5, anatomical specificity is still preserved by formalin fixation. A more detailed investigation of the effects of formalin fixation on the distribution of different elements is in place. First results are reported in this special issue [19].

Clustering based on LA-ICP-MS data for native tissue, shown in Fig. 5, leads to excellent anatomical segmentation. Several structures can be easily identified: bulbus olfactorius (Olf); cortex, layers I–IV (Ctx-A); cortex, layers V–VI (Ctx-B); white matter (WM); plate of four collins (Coll); cerebellar grey matter (Crbl); hippocampus fascia dentata (1) and the cornu ammonis part 1–3 (2). The richness of visual detail is at least comparable to that offered by the block-face photograph shown on the right side, and clustering has the great advantage of providing automatic and observer-independent segmentation. Clustering-based tissue segmentation can be expected to have a great impact on e.g., atlasing endeavours using element images.

The results shown in Fig. 6 for tissue lesioned by photothrombosis are significantly different from those discussed above. Interestingly, the segmentation of tissue by *k*-means clustering reveals substructures in the rim region around the lesion. The lesion itself, with microvasculature destroyed by dye coagulation leading to necrosis of tissue, is found to have a very different element signature from that of "normal tissue", as represented by the contralateral hemisphere. It becomes now possible to discriminate several distinct concentric zones around the central necrosis of the lesion. In a systematic study of stroke induced by photothrombosis, the existence of similar partitions in all animals should be investigated. If this proves to be the case, the clustering-based partition of lesions can be used, for example, to investigate changes in element concentration throughout each of the zones at different time points. We stress the point that observer-independent clustering greatly facilitates access to the Terra (nearly) Incognita of lesion substructure. It can be expected to have a huge impact on studies regarding e.g., the evolution of the penumbra zone following stroke or tumour development.

3.3. Optimizing the number of clusters

The dependence of the goodness of clustering, expressed by the silhouette measure, on the number of clusters is exemplified on the slice from the PT brain and shown in Fig. 6a–d. Thus, Fig. 6a shows the histograms of the silhouette value for each point when the number of clusters varies from 4 to 13. The histogram is concentrated at higher values, a sign for good clustering, for a small number of clusters (optimum of 5). In this case, however, the structure is concentrated in the rim of the lesion and no substructure is found in the rest of the brain. A local optimum is found for a moderate number of clusters, $N = 9$, in which the lesion and the brain are well separated, the rim displays substructure, and signs of anatomy (e.g., pixels in the corpus callosum) begin to emerge. The local maximum is apparent in Fig. 6b, where the mean value of the silhouette over the whole slice is plotted as a function of number of clusters. For this partition with $N = 9$, both the distribution of silhouette values (Fig. 6c) and the clusters (Fig. 6d) are shown. In contrast to the rim of the PT lesion, where substructure is well separated, even for a higher number of clusters (for example $N = 13$) the anatomy of the brain is not clearly defined. Among other reasons the washout of Zn and partially of Cu during fixation may have accounted for obliteration of the finer differences in the element distribution, leaving only those which were initially very pronounced (lesion, rim, normal tissue).

For the particular case of lesion induced by photothrombosis, clustering based on element imaging provides an unprecedented characterization. No previous information exists regarding the

number of regions for metal accumulation within the rim. This is different from segmentation of a slice of healthy tissue, where one can make use of very detailed anatomical information, as provided by e.g., histology, in order to validate the results of the clustering procedure. We have therefore repeated the segmentation on filtered images with the aim of reducing the possible influence of noise and investigating the reproducibility of the clustering results with different filter parameters. The reproducibility for different filter parameters was found to be high. Very moderate filtering ($k = 5$ in the ImageJ implementation) separates the lesion into more clear structure than that based on the native data. This is shown in Fig. 7, part a, with a slightly reduced number of clusters than that found optimal for non-filtered images. The structure remains very stable for higher diffusion constants (higher noise suppression; data not shown). We show in Fig. 7b the fingerprint of the element accumulation for selected elements (their mean values within a given cluster) and relevant clusters: necrosis region, "normal" tissue and 4 regions within the rim of yet unknown provenience. Most of the regions found within the rim have very high and varying levels of Zn concentration and could be influenced by any chemical procedure affecting the distribution of Zn between the native one and that in the LA-ICP-MS-imaged slice specimen. However, the segmentation is not based on this single element, since the fingerprint based on the other elements is also characteristic of each cluster.

The multi-element contribution to determining the partition within the lesion increases the confidence in the element-based segmentation as reflecting tissue substructure. Further effort should be directed towards the understanding of this highly specific element accumulation in the rim of the PT lesion.

3.4. Characteristics of the samples assessed and influence of tissue fixation

A careful comparison of the concentration of elements between the tissue samples is beyond the scope of the present paper. However, it is clearly seen from Figs. 4–6 that the distribution of some of the elements in fixed tissue is even qualitatively different from that in fresh tissue and also that the accumulation of elements in the rim of the lesion produced by photothrombosis shows a unique pattern which is not replicated in the tMCAO model of stroke.

In fixed (tMCAO) as well as in fresh tissue the distribution of elements in the white and grey matter shows characteristic patterns, and *k*-means clustering can easily segment anatomically distinct regions (Figs. 4 and 5).

4. Conclusions and outlook

In conclusion, we have presented a chain of operations which is very useful in preparing the data obtained from mass spectrometry imaging for quantitative analysis and neurobiological interpretation.

We have introduced a simple method for automatically determining the proper number of (non-integer) points per line for the reconstruction of geometrically accurate images from the stream of raw data based on the properties of the inertia tensor defined on any high-SNR element image. This procedure was validated by congruence to block-face images. We emphasize the fact that automated methods for data analysis will contribute to further increasing the impact of LA-ICP-MS data on our understanding of the brain. This has already been greatly helped by the development of dedicated software for processing of LA-ICP-MS data such as IMAGENA [20], which for the moment is based on interactive analysis.

Since mass spectrometry imaging produces a multitude of images reflecting largely independent aspects of tissue, the method is ideally suited for clustering analysis irrespective of the object investigated. Clustering based on multi-element distribution using a simple *k*-means algorithm was able to segment the tissue in anatomically meaningful regions. Furthermore, it was possible to identify distinct patterns of element accumulation in a lesion produced by photothrombosis. This latter aspect in particular opens a new avenue to lesion characterization.

Several refinements of the methods presented here should be made in the future. With respect to image reconstruction, the general applicability of the method based on the tensor of inertia to reshaping of LA-ICP-MS images needs further study. To this aim, replacing images with their binary mask and investigating the behaviour of the tensor of inertia with reshaping factor will increase the generality of the method. If available, comparison to the binary mask of a microphotograph of the slice mounted on slide before LA-ICP-MS is expected to increase the accuracy of reshaping. If the method demonstrates broad applicability, both steps should be included in automatic processing of LA-ICP-MS data.

The stability of clustering with respect to the number and nature of elements included in the procedure should also be more systematically investigated.

Consistent application of observer-independent segmentation on LA-ICP-MS data is expected to become a useful tool in the study of the heterogeneity of metal distribution in different applications.

Acknowledgement

The authors thank Thermo Fisher Scientific for instrumental support of the new BrainMet (BrainMet – Bioimaging of Metals and Metallomics) laboratory at Research Centre Juelich (www.brainmet.com).

References

[1] J.S. Becker, Inorganic Mass Spectrometry Principles and Applications, John Wiley and Sons, Chichester, 2007, and references therein.
[2] S. Becker, N. Jakubowski, Chem. Soc. Rev. 38 (2009) 1969–1983.
[3] J.S. Becker, M. Zoriy, B. Wu, A. Matusch, J.Su. Becker, J. Anal. At. Spectrom. 23 (2008) 1275.
[4] J.S. Becker, M.V. Zoriy, C. Pickhardt, N. Palomero-Gallagher, K. Zilles, Anal. Chem. 77 (2005) 3208.
[5] J. Feldmann, A. Kindness, P. Ek, J. Anal. Atom. Spectrom. 17 (2002) 813.
[6] J.S. Becker, A. Matusch, C. Depboylu, J. Dobrowolska, M. Zoriy, Anal. Chem. 79 (2007) 6074.
[7] B. Yao, T.Q. Li, P. Gelderen, K. Shmueli, J.A. de Zwart, J.H. Duyn, Neuroimage 44 (2009) 1259.
[8] M.R. Anderberg, Cluster Analysis for Applications, Academic Press, New York, NY, 1973.
[9] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.
[10] A.K. Jain, R.C. Bubes, Algorithm for Clustering Data, Prentice-Hall, Englewood Cliffs, NJ, 1988.
[11] L. Kaufman, P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, Wiley, New York, 1990.
[12] T. Alexandrov, M. Becker, S.O. Deininger, G. Ernst, L. Wehder, M. Grasmair, F. von Eggeling, H. Thiele, P. Maass, Spatial segmentation of imaging mass spectrometry data with edge-preserving image denoising and clustering, J. Proteome Res. 9 (2010) 6535–6546.
[13] M. Schröter, A.M. Oros-Peusquens, M. Irkens, A. Celik, A. Saleh, N.J. Shah, S. Jander, Spatiotemporal pattern of USPIO enhancement in experimental stroke lesions depends on the type of ischemic injury: a 9. 4T MRI study, in: Proceedings of the 14th Scientific Meeting of the International Society for Magnetic Resonance in Medicine (ISMRM) Seattle, WA, USA, 6–12 May, 2006.
[14] Hartigan, Wong, Algorithm AS136. A K-means algorithm, Appl. Stat. 28 (1979) 100–108.
[15] A.K. Jain, R.C. Dubes, Algorithms for Clustering Data, Prentice Hall, 1988.
[16] G. Paxinos, C. Watson, The Rat Brain in Stereotaxic Coordinates, IVth ed., Academic Press, San Diego, 1998.
[17] P. Perona, J. Malik, Scale space and edge detection using anisotropic diffusion, in: Proc. IEEE Workshop Comput. Vision, Miami, FL, November, 1987.
[18] P. Perona, J. Malik, Scale space and edge detection using anisotropic diffusion, IEEE Pattern Anal. Machin. Intell. 12 (July) (1990) 629–639.
[19] A. Matusch, A. Bauer, J.S. Becker, Element imaging in formalin fixed slices of human mesencephalon, Int. J. Mass Spectrom. 30x (2010), this special issue.
[20] T. Osterholt, D. Salber, A. Matusch, J.S. Becker, C. Palm, IMAGENA: image generation and analysis – an interactive software tool handling LA-ICP-MS Data, Int. J. Mass Spectrom. 30x (2010), this special issue.