Multi order correction algorithms to remove image distortions from mass spectrometry imaging datasets

Florian Gerber[1]‡, Florian Marty[2,3] ‡ , Gert B.Eijkel[3]‡ , Konrad Basler[2], Erich Brunner[2], Reinhard Furrer[1], Ron M.A. Heeren[3]**

[1)] Institute of Mathematics, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich; [2)] Institute of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich; [3)] FOM-AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands

Mass spectrometry imaging (MSI), image distortions, statistical correction, LDA, GLM

Time-of-flight secondary ion mass spectrometry imaging is a rapidly evolving technology. Its main application is the study of the distribution of small molecules on biological tissues. The sequential image acquisition process remains susceptible to measurement distortions that can render imaging data less analytically useful. Most of these artifacts show a repetitive nature from tile to tile. Here we statistically describe these distortions and derive two different algorithms to correct them. Both, a generalized linear model approach and the linear discriminant analysis approach are able to increase image quality for negative and positive ion mode datasets. Additionally, performing simulation studies with repetitive and non-repetitive tiling error we show that both algorithms are only removingrepetitivedistortions. It is further shown that the spectral component of the dataset is not altered by the use of these correction methods. Both algorithms presented in this work greatly increase the image quality and improve the analytical usefulness of distorted images dramatically.

Introduction

Imaging mass spectrometry is a technology with increasing use in the bio analytical field. The capability to obtain molecular images and at the same time chemical identity makes it a great tool to analyze a wide variety of organic and inorganic samples[1]. Secondary ion mass spectrometry imaging (SIMS) is one of the most used techniques to generate ion images of different molecular species. In a typical SIMS imaging experiment a primary ion beam is rapidly rastered over the sample of interest in a predefined track. At each position the primary ion beam is rastered over a squared surfaces area of a defined size that contains a fixed number of pixels (256×256). This is referred to as a tile. By decreasing the tile size the researcher can increase the spatial resolution down to the sub micrometer level. A larger area is analyzed by performing a mosaic of a large number of adjoining tiles (e.g.8×8)[2].

In a typical SIMS experiment the surface is bombarded with primary ions with energies in the keV range. As a consequence secondary ions are generated from the analyzed surface. Generally the ionization efficiency during this process is very low and decreases towards higher m/z values[2]. The secondary ions are analyzed in a mass spectrometer to generate position specific mass spectra. After completing the measurement at every (x, y)-position (tile) the ion images can be reconstructed. These images facilitate the investigation of the distributions of elements and small molecules such as lipids and other metabolites[3–5].

For data interpretation it is crucial that the images have a very high contrast, as well as sharp borders between sample and sample holder. Yet, concealed measurement-based artifacts may vitiate the dataset and subsequently impair the analysis. These distortions predominantly occur in every tile and show a high reproducibility from tile to tile. We therefore call these reproducible image errors tiling errors.

The cause of these errors is not completely understood. One of the possible reasons arises from the physical nature of the operating principle of SIMS and the surfaces analyzed. Most biological

surfaces are considered electrical insulators[6]. During the process of bombarding the insulator with primary ions the sample will start to charge up. This charge up will influence or even prevent the emission of secondary ions.

Different methods have been developed to avoid these charge up effects. For example coating the surface of interest with conductive materials (e.g. gold nanolayer) or charge compensation by electron bombardment. These applications tremendously improved image quality in SIMS imaging mass spectrometry[7].

But, charge build up is not the only source of tiling errors. Misalignment of the primary ion gun or local detector sensitivity issues can also lead to similar distorted images. Unfortunately, most of these errors are only visible after the entire image has been acquired. Therefore, it is of great importance to generate tools that can correct for these tiling artifacts after thedata acquisition.

Because of the repetitive nature of the tiling errors it is likely that these errors can be stochastically described. Such a stochastic model of tiling error would possibly allow the construction of methods to remove these artifacts. As a direct benefit the image quality would be enhanced and therefore multivariate analysis methods such, as principle component analysis (PCA) may gain significance. In this paper we show that the tiling errorscan indeed be statistically explained based on spectral features andwe have developed two algorithms that can directly remove these artifacts efficiently on the spectral level of the data. The algorithms and analytical approach described in this paper allows the user to get the most out of data acquired with tiling errors and thereby improve the data interpretation.

Experimental Section

Data acquisition:

Data used in this study was from atherosclerotic segments of mice arteria that were analyzed for lipid distribution. Additional information about the datasets can be found in[8]. Briefly, the time-of-light (ToF) SIMS data were acquired on a Physical Electronics (Eden Prairie, MN) TRIFTII secondary ion mass spectrometer. Datasets were acquired in positive and negative ion mode ensuring that the total ion dose was kept below 10^{13} ions/cm^2, the so-called static limit of SIMS. Data acquisition was performed using the software WinCadence, version 4.4.0.17 (Physical Electronics, Chanhassan, MN). Data analysis was performed using the ChemomeTricks software package[9].

Generalized linear model (GLM) approach:

In this approach the measurement region of the detector is divided into two sub regions, namely the *region of the sample* (RS) as the region containing the measured tissue and the *background* (BG)as the region where we observe sample holder (glass) only. First, one single m/z-peak is corrected. To correct an entire dataset we repeat the procedure for each peak (Fig.1). We assume a parametric description of the distortion and estimate its parameters on the BG where the tilingerrorpattern is not disturbed by biological information. The tilingerroris then extrapolated to the RS and the detector signal is separated into a tiling error and biological information component. The resulting corrected data is used as input for further multivariate data analysis (Fig.1).

The approach is based on the following two assumptions on the data structure:

(A1) The tiling error in each peak can be described by a common model formulation.

(A2) The tiling error structure on the RS is the same as on the BG.

The first assumption allows us to perform the correction from peak to peak without altering the model.

We denote the observed intensity count at a pixel/location (x, y) with z(x,y). For each m/z count we assume a Poisson distribution with rate λ(x, y) and model the log rate log (λ(x,y)) additively through a *biological information* part bioInfo(x,y), a *tiling error*part tiling(x,y), a *constant intercept* on the RS μ_{RS} and an unstructured noise component ε. On the RS we have

$$\log (\lambda_{RS}(x,y)) = \mu_{RS} + \text{tiling}(x,y) + \text{bioInfo}(x,y) + \varepsilon \quad (1)$$

Assumption (A2) implies that the tilingerroris common over the entire slide and for the BG the log rate is thus

$$\log(\lambda_{BG}(x,y)) = \mu_{BG} + \text{tiling}(x,y) + \varepsilon \quad (2)$$

as there is no biological informationbut a potentially different constant intercept μ_{BG}. To model the tilingerroron the BG a GLM with log-link function is used[10]. Its linear predictors have to capture the tiling error. More specifically, we use fourth order polynomials in terms of (x mod k) and (y mod k), with k the size of an individual tile. The modulus construction imposes a repeated structure on the model for the tiling error component.

Finally the estimated tiling error is removed from the data accordingly. The corrected rate linked to the biological information can be estimated as

$$\exp(\text{bioInfo}(x,y)) = z_{RS}(x,y) / \exp(\mu_{RS} + \text{tiling}(x,y)) \quad (3)$$

where$z_{RS}(x,y)$ is the observed intensity count at a pixel/location (x,y) on the RS. exp(bioInfo(x,y))does not have a tilingerrorcomponent anymore.

The correction routine is implemented in the statistical software R[11]. The R package R.matlab is used to access the data from ChemomeTricks software package[9, 12] and later to return data

structures to the same software to perform principal component analysis. Parallel computing is exploited with the R package multicore[13], resulting in a virtually linear speed up. The first principal component of the dataset is used to select the RS and obvious outliers from the BG are excluded. In case the peak has a low overall intensity, the estimated tilingerrormay be very small for some (x, y), leading to unstable results. The fitting procedure is further improved by transforming the ratio in (3) to

$$(z_{RS}(x,y) + q) / (\mu_{RS} + \exp(\text{tiling}(x,y)) + q), \quad (4)$$

where $q = \max(0, 1 - \exp(\min(\mu_{RS} + \text{tiling}(x,y))))$. The corrected intensities are rescaled to have the same overall intensity as the original ones, allowing comparisons of total ion images between different methods.

Linear discriminant analysis (LDA) approach:

The second correction method assumes the correlation of the structure of the data with its position of the pixel in each tile. We used a LDA approach to model and remove the tiling error from the data[14].

The first discriminant function (DF_1) is defined as

$$\frac{D_1^T B D_1}{D_1^T W D_1} = \text{maximal}$$

Where D_1^T is the transposed DF, B and W are the between and within group covariance matrices. The same definition holds for the second and higher DFs with the precondition that all DFs are orthogonal.

This approach first assigned a category (i.e. class) to each pixel according to their position within each tile. Then LDA was used to create a model based on these classes. The resulting first discriminant function (DF_1) described the common behavior of the spectra at each tile position over all tiles and therefore represented the tiling error. DF_1 scores of all pixels were back transformed to the original (mass) feature space by multiplication of the DF1 score with DF1 loadings, and auto scaling and normalization were reversed. The resulting dataset was subtracted from the original dataset and the LDA procedure was repeated until no more tilingerrorrelated variance remained in the data.The procedure is summarized in the following equation:

$$S_{corr} = \left((S - ScoreDF_1 \times LoadDF_1) \cdot* Var + Mean\right) \times Norm_s$$

Where S_{corr} is the corrected spectrum, S is the auto scaled and normalized spectrum to be corrected, $ScoDF_1$is the score of S, on $DF_1 LoadDF_1$ is the loading vector of DF_1, Var is the vector that contains the variances of the mass channels, $Mean$ is the vector that contains the mean values of the mass channels and $Norm_s$ is the normalization factor for S. Please note that operator $\cdot*$ means the component wise multiplication of the vectors.

After all the tiling error has been removed, the DF_1scores would be random and is assessed with the Wald-Wolfowitz runs test on the DF_1 scores. If the Wald-Wolfowitz runs test is positivethe cycling would be stopped[15]. The resulting corrected data were used as input for further multivariate data analysis (Fig.1).

Both presented algorithms in this study were tested on samples containing large amount of tiling distortions.

Simulation study:

To further support that the correction routines only remove tilingerrorsand do not alter the biological information, i.e. to correctly separate signal and repetitive noise, we have performed multiple simulation studies. Two representative examples are presented and discussed here. We took a corrected dataset after the GLM correction was applied. This corrected dataset represents the `ground truth' signal and was subsequently altered by adding different artificial tiling noise. The tiling pattern was regular or alternated by inverting every second tile (checkerboard like) to simulate non-repetitive noise. In the latter case, the correction routines should not detect a pattern and leavethe data unaltered.

Results and discussion

Correction of biological datasets

We developed two algorithms, GLM and LDA to remove artificial repetitive distortions (tiling errors) (Fig. 2a). The datasets used in here were arterial cross sections of atherosclerotic origin[8]. First, we corrected the data usingGLM that is a parametric approach. The GLM approach uses the background (in our case, the sample holder) to estimate the tiling error structure and corrects it for one peak at the time. The procedure is then repeated for every peak (details see Experimental Section). Using appropriate model functions for the background, the biological information and the tiling error part - the tiled data can be corrected using a multiplicative correction approach. Subjectively, the algorithm performs well in negative ion mode and removed all the tiling error(Fig. 2b). In positive ion modecompared to the non-corrected datasetthe algorithm removed most of the tiling error and facilitates image interpretation (Fig. 2b).

Second, we applied a different correction algorithm based on LDA to correct the same datasets and comparefor the suitability of the different algorithms. Comparable to the GLM approach the LDA approach removes the tiling in the negative ion mode (Fig. 2c) and performs equally well

in positive ion mode (Fig. 2c). This indicates that the choice of an iteratively applied additive model seem to be as appropriate as the use of a multiplicative model. Both algorithms are not able to entirely remove the tiling errorartifacts in the positive mode dataset. Possibly, the remaining tiling error may not be as repetitive as assumedfor the algorithms to be corrected. Further investigation of the remaining tiling errorsare necessary to improve the correction efficiency for non-repetitive artifacts.

Correction of simulated tiling error on de-tiled real life sample

As both algorithms lead to a subjectiveimprovement of the data quality we also examined their effect on the spectral components (principal component loadings). In a first step we added artificially generated regular tiling error to a corrected dataset. The tiling error was constructed to resemble the typical tiling error patterns (Fig. 3a). Clearly, the addition of tiling error heavily influences the loadings. This is indicated by the positive and negative differences between the loadings of the initial dataset and the artificially tiled dataset (red and green bars below and above the spectra Fig. 3). Butthe GLM, as well as the LDA procedure, removed these artifacts completely (Fig. 3b). Comparing the initial and corrected dataset the loadings only vary slightly indicating that also the spectral distributions were restored to the initial state (again red and green bars below and above the spectra). This demonstrates that the algorithms do remove tiling error and thusincrease the data quality. In an additional simulation we added artificially generated alternating tiling error which was not observed previously on real data (Fig. 4a). We call this tiling error ‘checkerboard tiling’.Again the loadings of the checkerboard-tiled dataset are different from the initial one (red and green bars below and above the loadings spectra Fig.4). Neither the GLM nor the LDA procedure removed this pattern (Fig. 4b). Also the loadings were kept unchanged if the corrected dataset was compared with the dataset containing the distortions. (Note the small scales of the red and green bars.) This illustrates that the proposed algorithms

only remove repetitive tiling error structures and no other potentially relevant information. Taken together these findings clearly demonstrate the use of such algorithms to increase image quality.

Conclusion:

We have investigated a long-standing problem in mass spectrometry imaging, the occurrence of repetitive image artifacts called tiling error through the use of two different algorithms. We demonstrated that a GLM and LDA approach are both equally well suited to correct tiled datasets. Simulation studies of different tiling error patterns showed that the structure of the tiling erroris only corrected if it is repetitive in every tile. Moreover, we demonstrated that both algorithms perform the correction without corrupting the spectral components of the data. Both algorithms have been applied to several dozen different samples. The versions described here represent a stable and comprehensive implementation. Both methods are built on different components that are almost orthogonal to each other (see Suppl. Table 1). Naturally all these features can be interpreted as a disadvantage. For example, the GLM approach does not use the sample area to estimate the tiling error but subsequently uses a non-negligible area of background.

It is straightforward to build new and different correction approaches by exchanging or extending individual model components, e.g. using generalized additive models instead of GLM, using quadratic discrimination instead of a linear one. The utilizationof the software R illustrates that there is no limitation to useproprietary software and that it is possible to built new or more complex correction routines. Finally, we believe that other fields with similar problems (e.g. MALDI MSI, microscopy) may apply the algorithms presented herein, provided a repetitive error pattern is present.

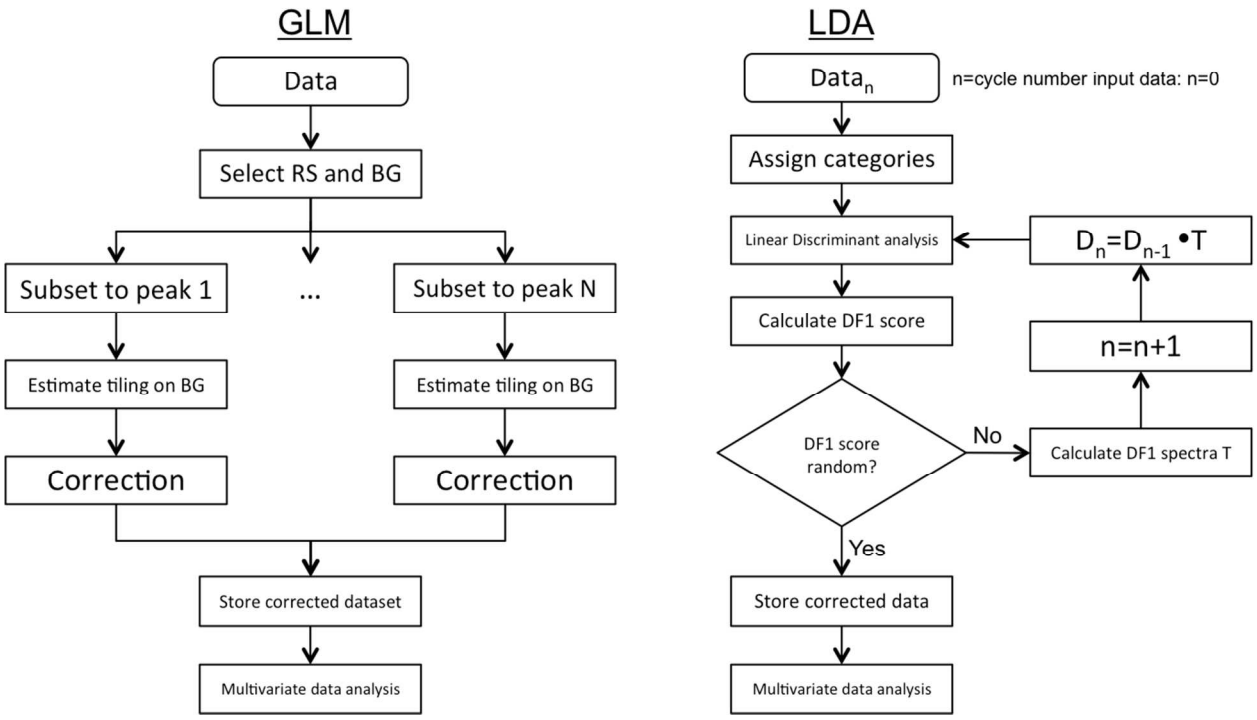


Figure 1. Workflow of the GLM and LDA approach used in this work. The GLM approach corrects the tiling errorin a peak-by-peak manner, whereas the LDA performs linear discriminant analysis on the entire spectra of an assigned category. Due to its non-iterative nature the GLM approach does not require any statistical test to stop the correction procedure. In contrast, LDA is iterative and hence has to stop when the DF1 score calculated is random. The randomness is tested using a Wald-Wolfowitz runs test[15].

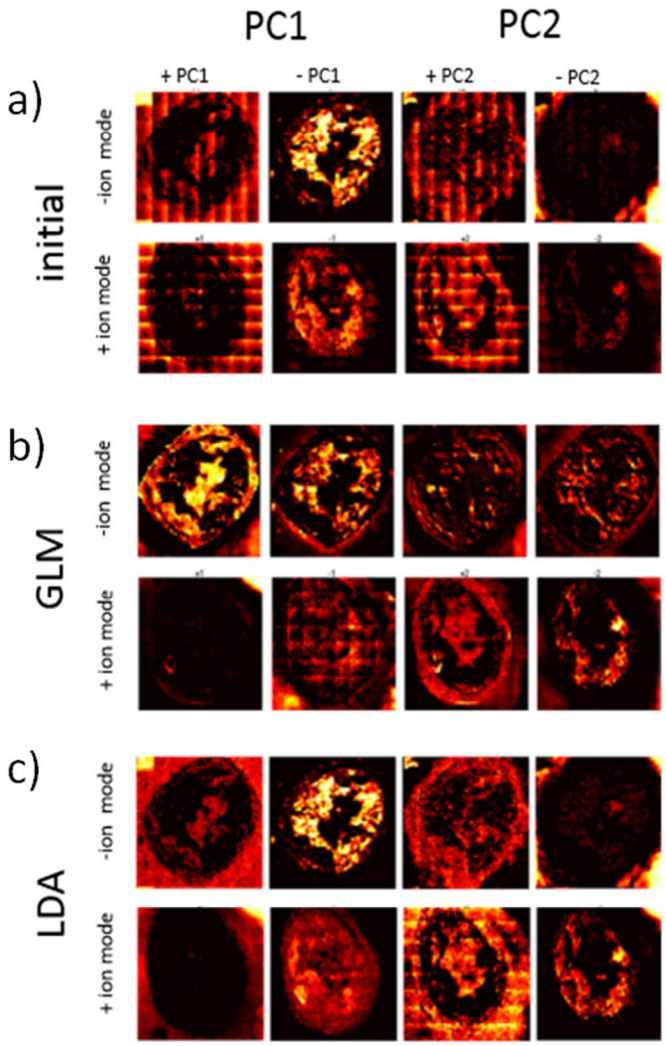


Figure 2. First and second principal components (PC1, PC2) of uncorrected data shows strong image distortions both in positive and negative ion mode **(a)**.Correction of the distorted data called initial with the GLM and LDA approach removed the distortions completely in negative ion mode **(b, c)**. In positive ion mode some distortions are not removed. The remaining distortion needs to be further examined to improve the correction of it**(b, c)**.

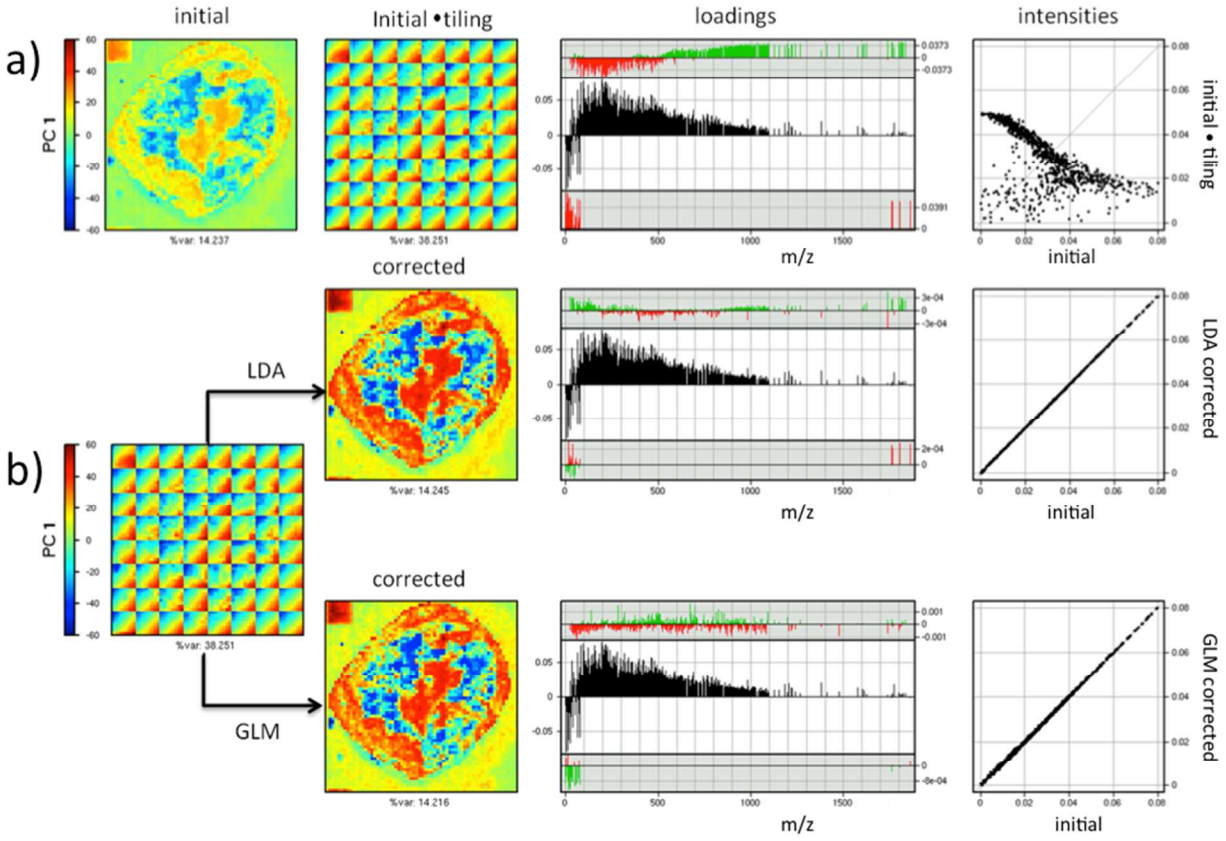


Figure 3. The corrected dataset (initial) was distorted with a simulated repetitive tiling error pattern (initial • tiling)**(a)**. This changed the loadings spectra and lead to a significant shift of the peak intensities. As indicated above the loadings spectra, most of the loadings increase (green) in the higher mass range whereas the lower masses decrease in loadings valueby a factor of ±0.0379(red). For the negative part of the principle component only minor changesof a factor of 0.0391are observed as shown under the spectra indicated with green and red bars. The distorted simulated data set was then corrected with GLM and LDA **(b)**. A direct comparison of the initial and the corrected dataset does not reveal any changes in the loadings(±3e-04 for positive and ±2e-04 for negative loading components) for the GLM. For theLDA approach the change is also negligibly small (±0.001 for the positive and 8e-04 for the negative loadings).In addition for both algorithms the peak intensities remain the same after correction compared to the initial.

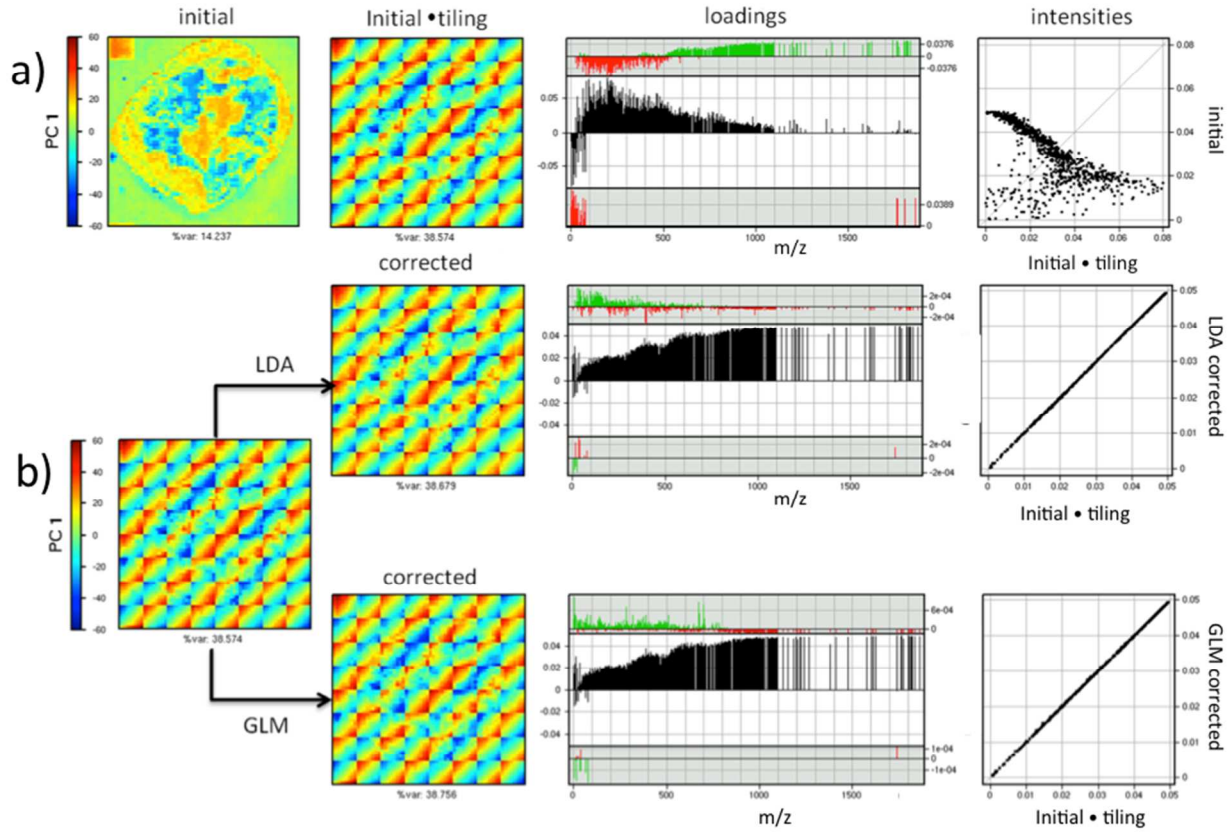


Figure 4. The corrected dataset (initial) was distorted with a simulated non-repetitive tiling error pattern (checker board tiling)(initial • tiling)**(a)**. Comparable to the repetitive tiling error the loading spectra and the peak intensities are highly changed. As shown above the loadings spectra, most of the loadings increase (green) in the higher mass range whereas the lower masses the loadings value decreases (red) by a factor of 0.0376. The negative part of the principal component does only change little (shown below, ±0.0389). The distorted simulated data set was then corrected with GLM and LDA **(b)**. As expected, both methods do not remove this distortion since it is not based on tile-by-tile repetition. Comparing the corrected data set to the

initial•tiling dataset the loadings and peak intensities remain unchanged (±2e-04 for the GLM and 6e-04 for the LDA positive loadings and 1e-04 for the negative loadings, respectively)

Corresponding Author

* Ron M.A.Heeren, FOM-AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands and ReinhardFurrer, Institute of Mathematics, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‡These authors contributed equally to the work.

ACKNOWLEDGMENT

This work is part of the research program of the StichtingvoorFundamenteelOnderzoek der Materie (FOM), which is financially supported by the NederlandseOrganisatievoorWetenschappelijkOnderzoek (NWO). This project was in part financed with a grant from the Swiss SystemsX.ch initiative, grant IPP-2011/"Identification of Small Molecules Relevant for Patterning and Growth of the Wing Imaginal Disc Using High Resolution Mass Spectrometry" to EB, KB and RMAH and by the University Research Priority Program (URPP) Systems Biology/Functional Genomics of the University of Zurich to RF.

ABBREVIATIONS

BG, Background; DF, discriminant function; GLM, generalized linear model; LDA, linear discriminant analysis; MALDI MSI, matrix assisted laser desorption/ionization mass spectrometry imaging; PC, principle component; PCA, principal component analysis; RS, region of sample; SIMS, secondary ion mass spectrometry, ToF, time-of-flight

REFERENCES

(1) McDonnell, L.; Heeren, R. *Mass spectrometry reviews26*, 606–643.

(2) Bertrand, P.; Weng, L. *MikrochimicaActaSuppl***1996**, *13*, 167–182.

(3) Debois, D.; Bralet, M.-P.; Le Naour, F.; Brunelle, A.; Laprevote, O. *Analytical chemistry***2009**, *81*, 2823–2831.

(4) Cillero-Pastor, B.; Eijkel, G.; Kiss, A.; Blanco, F. J.; Heeren, R. M. A. *Analytical chemistry***2012**, *84*, 8909–8916.

(5) Xu, K.; Proctor, A.; Hercules, D. M. *MikrochimicaActa***1996**, *122*, 1–15.

(6) Müller, G. *Applied Physics***1976**, *10*, 317–324.

(7) Altelaar, A. F.; Klinkert, I.; Jalink, K.; De Lange, R. P.; Adan, R. A.; Heeren, R. M.; Piersma, S. R. *Anal Chem***2006**, *78*, 734–742.

(8) Bot, M.; De Jager, S. C. A.; MacAleese, L.; Lagraauw, H. M.; Van Berkel, T. J. C.; Quax, P. H. A.; Kuiper, J.; Heeren, R. M. A.; Biessen, E. A. L.; Bot, I. *Journal of lipid research***2013**, *54*, 1265–74.

(9) Eijkel, G.; Kokrer Kaleta, B.; Van Der Wiel, I.; Kros, J.; Luider, T.; Heeren, R. *Surface and Interface Analysis***2009**, *41*, 675–1360.

(10) Nelder, J. A.; Wedderburn, R. W. M. *Journal of the Royal Statistical Society. Series A (General)***1972**, *135*, 370–384.

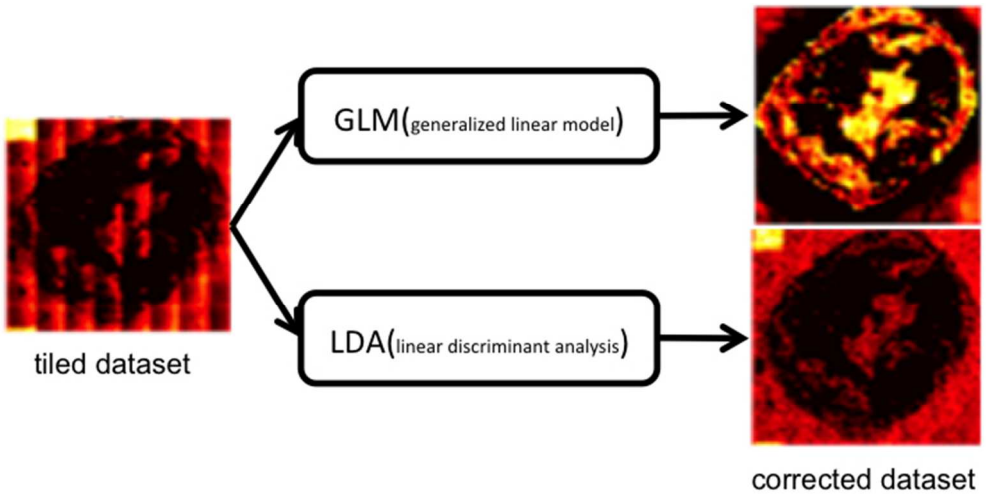
(11) R.Core Team R: A language and environment for statistical computing http://www.r-project.org.

(12) Bengtsson, H.;Riedy, J. R.matlab: Read and write of MAT files together with R-to-Matlab connectivity **2012**.

(13) Urbanek, S. Multicore: Parallel processing of R code on machines with multiple cores or CPUs **2011**.

(14) Hoogerbrugge, R.; Willig, S. J.; Kistemaker, P. G. *Analytical Chemistry***1983**, *55*, 1710–1712.

(15) Bradley, J. V. *Distribution-Free Statistical Tests*; Prentice-Hall, **1968**; 388.



121x61mm (150 x 150 DPI)