

RESEARCH

Open Access



Mass recalibration for desorption electrospray ionization mass spectrometry imaging using endogenous reference ions

Paolo Inglese^{1,2*}, Helen Xuexia Huang², Vincen Wu², Matthew R. Lewis^{1,2} and Zoltan Takats^{2*}

*Correspondence:

p.inglese14@imperial.ac.uk;

z.takats@imperial.ac.uk

¹ National Phenome Centre, Imperial College London, Hammersmith Campus, IRDB Building, London W12 0NN, UK

² Section of Bioanalytical Chemistry, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

Abstract

Background: Mass spectrometry imaging (MSI) data often consist of tens of thousands of mass spectra collected from a sample surface. During the time necessary to perform a single acquisition, it is likely that uncontrollable factors alter the validity of the initial mass calibration of the instrument, resulting in mass errors of magnitude significantly larger than their theoretical values. This phenomenon has a two-fold detrimental effect: (a) it reduces the ability to interpret the results based on the observed signals, (b) it can affect the quality of the observed signal spatial distributions.

Results: We present a post-acquisition computational method capable of reducing the observed mass drift by up to 60 ppm in biological samples, exploiting the presence of typical molecules with a known mass-to-charge ratio. The procedure, tested on time-of-flight and Orbitrap mass spectrometry analyzers interfaced to a desorption electrospray ionization (DESI) source, improves the molecular annotation quality and the spatial distributions of the detected ions.

Conclusion: The presented method represents a robust and accurate tool for performing post-acquisition mass recalibration of DESI-MSI datasets and can help to increase the reliability of the molecular assignment and the data quality.

Introduction

In the last 20 years, mass spectrometry imaging (MSI) has attracted increasing attention as a technology capable of capturing molecular spatial patterns from sample surfaces. Among the available ionization sources, such as matrix-assisted laser desorption [1] (MALDI) or secondary ion mass spectrometry [2] (SIMS), desorption electrospray ionization [3] (DESI) has gained popularity and been widely applied thanks to its relatively simpler preparation process. A map of the detected ions' spatial distribution is usually represented as an image and analyzed for identifying informative spatial patterns. MSI has been successfully deployed in various application areas, from cancer research [4–7] to pharmacology [8–12]. However, as an emerging technology, MSI still faces implementation challenges for controlling the quality of data produced [13–15]. The accuracy of mass-to-charge ratio (m/z) measurements under



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

sustained use are chief among these concerns. MSI acquisitions often consist of tens of thousands of spectra (pixels), corresponding to several hours of operation, during which mass measurements can be subject to substantial drift. For example, a small image of 10 mm × 10 mm, acquired at a running speed of 1 scan/s, with a spatial resolution of 100 μm, requires about 2.8 h. Shifting mass measurements are detrimental to the analyzed spatial patterns and can result in erroneous chemical annotations. Additionally, as MSI relies on only a single dimension of separation for detected ions, errors of a few parts per million (ppm) can result in hundreds of candidate molecular identities, interpreting in terms of biochemical hypotheses a practical challenge.

For these reasons, it is crucial to develop quality control approaches for MSI that mitigate this issue and yield accurate m/z measurements with high precision during extended instrument use. Solutions successfully applied in more routine hyphenated techniques (e.g., liquid chromatography-mass spectrometry LC-MS) include periodic recalibration or more frequent calibration based on the measurement of one or more standard reference materials during the acquisition. Such an approach is infeasible for the current MSI technology since the acquisition is performed continuously while the probe moves across the sample surface. However, this approach may be emulated by adding the standard material to the sample [16]. Although this approach can be used to facilitate post-acquisition m/z correction of mass accuracy, concerns may arise about the unknown effect on the ionization efficiency caused by these exogenous molecules' presence. In the specific case of DESI-MSI, usually, only one reference molecule (lock mass) is added to the solvent, which can be insufficient to capture the non-linear drift across the measured m/z range.

Another class of approaches relies on using endogenous or background peaks as candidate reference ions to estimate the mass shift occurring in each spectrum and map the measured m/z onto their corresponding recalibrated values [17]. This class of methods requires prior knowledge about the molecular content of the sample.

Boskamp et al. showed that endogenous signal (chemical noise) could improve the mass accuracy in MALDI-MSI of peptide datasets [18]. Recently, La Rocca et al. have presented an algorithm to recalibrate MSI datasets using a linear fit on the observed mass errors from endogenous biological peaks [19]. In their method, they treated each spectrum as independent.

Here, we present a different approach based on a fixed set of reference ions across the entire MSI dataset, in which the individual mass shifts are modeled as a time series. The advantages of global references are that they (1) define a fixed boundary for the m/z extrapolation, (2) improve evaluation of the quality of the matching procedure by visualizing their spatial distribution, and (3) increase the robustness of the matching process as a time-dependent model.

The approach leverages the ubiquity of common structural components of tissues as dependable sources of reference values, using them as multiple points to estimate error and correct all observed masses in each spectrum. The approach is demonstrated on biological tissue datasets acquired using a DESI-MSI source interfaced with time-of-flight (TOF) or Orbitrap ion analyzers, in both positive (ES+) and negative (ES-) ion mode.

For simplicity, we will refer to m/z values as “masses” throughout the manuscript.

DESI-MSI datasets

In-house DESI-MSI datasets

DESI-MSI data of six tissue sections from mouse (*mus musculus*) brain and pig (*sus domesticus*) liver were acquired using a Waters Xevo G2-XS QToF mass spectrometer.

Three brains of a C57BL/6 mouse model were purchased from Charles River Laboratories, while pig liver samples were obtained from a local supermarket.

We used the sample preparation procedure and DESI parameters reported in Tillner et al. [20]. The mouse brain samples were scanned at a rate of 75 $\mu\text{m/s}$ horizontally, while the pig liver sample was scanned at a rate of 100 $\mu\text{m/s}$.

The RAW spectra from the six acquired TOF DESI-MSI were first converted into imzML using the MassLynx SDK (v4.7.0) and Python's pyimzML package.¹ Then, they were filtered from the baseline noise using a modified version of the *kneedle* algorithm [21] (Additional file 1: Section S1) and smoothed using a Savitzky-Golay kernel convolution. Centroided peaks were detected and used as input for the recalibration procedure (Additional file 1: Section S2). Throughout the manuscript, we will refer to spectrum and list of centroid peaks as synonyms.

Public DESI-MSI datasets

Twenty-four DESI-MSI datasets were downloaded from the public service METASPACE.² We considered various tissue types analyzed with either Orbitrap or TOF ion analyzers. Details about the datasets can be found in Additional file 1: Table S1.

The datasets consisted of centroided RAW peaks. No peak detection or denoising was applied.

Methods

Selection of sample pixels

The first step of the presented workflow consists of determining the pixels associated with the biological sample. Since our method uses endogenous molecules that are expected to be detected by DESI-MS in tissue, it is crucial to remove all pixels that do not correspond to tissue-related signals.

The procedure aims at discriminating the sample-related from off-sample pixels, using a supervised classifier trained on a set of user-defined labeled pixels.

We first define the set of features from the RAW peaks, applying a uniform binning with a bin size of 1 m/z . Subsequently, through a graphical user interface (GUI), the user manually labels a set of pixels as either 'sample' or 'background' (Additional file 1: Fig. S2). Then, a linear support vector machines (SVM) model is fitted on this set of features and labels and used to predict all pixels' labels, providing a binary map of the region-of-interest (ROI). The user can manually refine the segmentation through the GUI. Finally, connected regions (8-neighborhood) smaller than a given size are assigned to the background class. In all the experiments, we set the size threshold equal to 50 pixels.

The ROI mask is saved in a comma-separated values (CSV) file to be used in the workflow's following steps.

¹ <https://github.com/alexandrovteam/pyimzML>.

² <https://metaspace2020.eu/>.

General recalibration workflow

In this section, a general description of the recalibration method is given, with details reported in the following sections.

Let us consider a DESI-MSI dataset representing a collection of N sequentially acquired spectra (where $N = N^{(ROI)} + N^{(off)}$) is the sum of the number of sample ROI and off-sample pixels).

Given an observed mass spectrum, the calibration procedure aims to estimate the theoretical (“calibrated”) masses from the observed masses of the detected ions.

One common strategy is based on reference points, called *lock masses* [22]. Given a mass spectrum from a pixel p , this method assumes a statistical model g between the theoretical $\mathbf{M}_p = (M_{k,p})$ and observed $\mathbf{M}_p^\# = (M_{k,p}^\#)$ masses:

$$M_p = g_p(\mathbf{M}_p^\#, \boldsymbol{\theta}_p) \tag{1}$$

where the index p indicates that the observed and theoretical masses, the model g and its parameters $\boldsymbol{\theta}$ are spectrum (or pixel)-specific.

Given a set of reference masses and their observed values, the parameters are fitted from data, and the model g_p is used to predict the calibrated values of all detected masses.

When using endogenous signals as a reference, lock masses can be spectrum-specific [19] ($\mathbf{M} = \mathbf{M}_p$) or be defined globally. Our method uses the latter approach.

When using a global set of reference masses, the models g_p can only be fitted if the reference masses are observed in all pixels p . However, due to the molecular heterogeneity of biological samples, this property is not guaranteed. To overcome this challenge, for all reference masses \mathbf{M} , we model their observed values in the pixels p as a smooth function of the acquisition time (or, equivalently, the pixel order). This is equivalent to assuming that the observed masses $\mathbf{M}^\#$ depend on the actual conditions of the instrument, and that these smoothly vary in a controlled environment. In practice, we model each mass in $\mathbf{M}^\#$ as a time series, where the pixel order is a proxy variable for the acquisition time.

The temporal trends of the masses’ shifts are fitted as follows. Given a the theoretical value of a reference mass $M_k \in \mathbf{M}$, let $\mathbf{P}_M = (p_1, p_2, \dots) = (p_i)$ be the subset of ROI pixels where it is observed with mass $\mathbf{M}_k^\# = (M_{k,p_1}^\#, M_{k,p_2}^\#, \dots) = (M_{k,p_i}^\#)$. We model the time series trend using a generalized additive model (GAM) [23]:

$$M_{k,p_i}^\# | \eta_{k,p_i}, \sigma^2 \sim \text{Normal}(\eta_{k,p_i}, \sigma_k^2)$$

$$\eta_{k,p_i} = \beta_{0,k} + \sum_{j=1}^J s_{j,k}(p_i) \tag{2}$$

where $s_{j,k}$ are penalized cubic spline functions and J is equal to 20.

The procedure, repeated for all $M_k \in \mathbf{M}$, separately, provides the trend models for all reference masses. As for any regression models, if \mathbf{P}_M corresponds to ‘a large portion’ of the ROI, the fitted model can accurately predict the values of each database mass,

$\widehat{M}_k = \eta_k = (\eta_{k,p_i})^3$,³ that would have been observed (up to a random error term) in all ROI pixels p_i , given the instrumental condition at the acquisition time.

In the following section, we describe the procedure to detect and select the reference masses for the shift trends estimation.

Selection of reference masses and shift trend modelling

Since the molecular composition of the sample is largely unknown, it is practically impossible to define a general set of lock masses that works for all datasets. Before describing the procedure aimed at identifying them, we need to introduce properties that define ‘good’ endogenous reference masses.

The candidate reference mass is optimal if: (a) it is detected in a ‘large’ portion of the sample, (b) its peak is ‘isolated’ in the m/z space. Property (a) allows us to use the model described in Eq. 2, while property (b) reduces the uncertainty of the matching procedure.

The procedure aimed at establishing if an observed mass matches a database mass is described in the following. As the reference ions consist of molecules that are expected to be found in tissue samples and be detected by DESI-MS, we perform a database search to determine the list of candidates. The database Ω includes phospholipids, fatty acids, mono/di/triglycerides, cholesterol, and ceramides from Lipidmaps [24] and Human Metabolite DataBase [25] (HMDB) databases. Deprotonated and chlorine adducts ($[M-H]^-$, $[M+Cl]^-$) are considered for negative polarity mode, while protonated, sodium and potassium adducts ($[M+H]^+$, $[M+Na]^+$, $[M+K]^+$) are considered for positive polarity mode.

Given a database mass $M \in \Omega$, a mass m_i among observed masses in pixel p_i is considered a candidate match if $m_i \in [M - \lambda, M + \lambda]$, where $\lambda = W \times M \times 10^{-6}$ is the mass distance from M corresponding to the relative error W in ppm units. All masses satisfying the condition are considered possible matches, meaning that more matches per pixel can be found. For each candidate match, we retrieve its peak index π_i , observed mass $M_i^\# = m_i$, and intensity $I_i^\#$.

To satisfy property (a), for all $M \in \Omega$, we remove the candidate matched peaks that are detected in less than 75% of the sample ROI pixels.

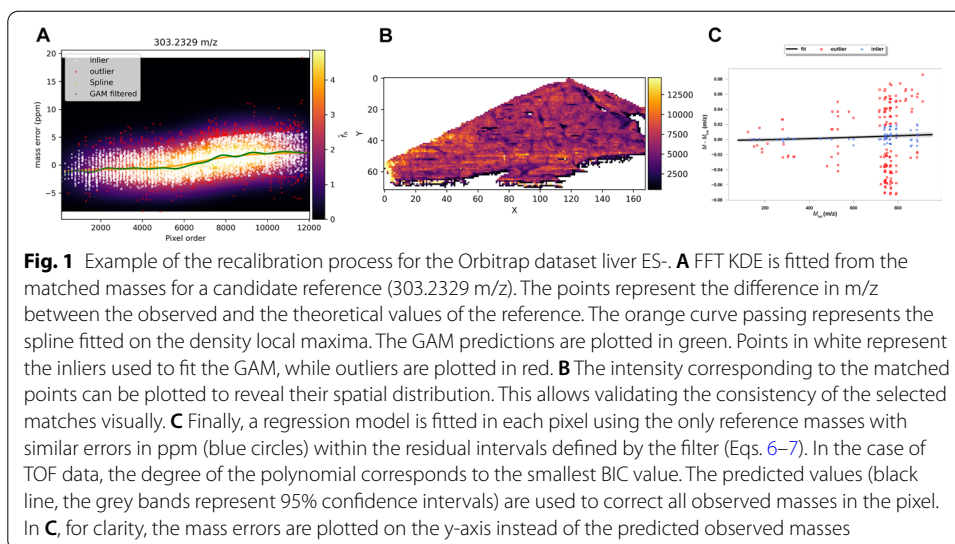
Subsequently, we further filter the list of candidate reference ions using a method based on a kernel density estimator (KDE) following a similar procedure to that described in Smirnov et al. [26] If candidate reference peaks for M are represented as points with coordinates $\mathbf{x} = (x, y) = (p_i, M_i^\#)$, we assume that highly dense connected regions represent the shift trends of M .

Given a candidate reference mass $M \in \mu \subseteq \Omega$ and its observed values $\mathbf{M}^\# = (M_i^\#)$, we estimate the density $f_h = \frac{\sum_i K(\frac{x-x_i}{h})}{nh}$ of the (x, y) -points using a 2D Fast Fourier Transform (FFT) KDE [27], with a triangular kernel K , defined as

$$K(\mathbf{x}) = 1 - |\mathbf{x}|, |\mathbf{x}| \leq 1 \quad (3)$$

The kernel is fitted on the coordinates of the points scaled to $[0, 1]$ interval.

³ To simplify the notation, we denote the predicted observed masses with the symbol \widehat{M}_k instead of $\widehat{M}_k^\#$.



The triangular kernel is chosen because of its computational efficiency. The kernel bandwidth is set to $h = 2.576 \times \sigma \times N^{-1/5}$, where σ represents the standard deviation of the whole set of points and N represents the number of points \mathbf{x} [28]. The kernel is fitted on a regular grid of size $G \times G$, with $G = 1024$.

Given the estimated 2D density $\hat{f}_h = \hat{f}_h(x, y)$, we identify the curve passing through its local maxima as follows. For each $x' \in \{1, \dots, G\}$, we determine the local density maximum $y' = \arg \max_{x=x'} \hat{f}_h(x, y)$. If $\hat{f}_h(x', y) = 0$ for all y , no maximum is considered for that value of x' . Subsequently, a cubic spline $S(x)$ is fitted on the vector of local maxima (\hat{x}, \hat{y}) after transforming them back to the original (pixel, mass)-space. This step allows to determine a smooth model along the detected local KDE maxima. Using the predicted spline values $S(p_i)$ for all p_i , we calculate the absolute residuals r_i and the dispersion d_i :

$$r_i = |M_i^\# - S(p_i)| \tag{4}$$

$$d_i = \frac{2 \times \text{mad}(\mathbf{r})}{S(p_i)} \times 10^6 \tag{5}$$

Points with $r_i \geq 2 \times \text{mad}(\mathbf{r})$, $\mathbf{r} = (r_i)$, where “*mad*” is the median absolute deviation multiplied by the inverse of cumulative Normal distribution $1/\Phi^{-1}(3/4) \approx 1.4826$, are considered outliers and removed from the list of matched peaks (Fig. 1A).

Finally, the candidate reference M is kept if: the number of distinct inlier pixels is greater or equal than 75% of the ROI size and $\max_i(d_i) \leq 10$ ppm. In this way, we select isolated signals (property (b)). The choice of a threshold equal to 10 ppm for the dispersion around the fitted spline is based on the expected scattering of the mass shifts due to intrinsic measurement noise, and it can be customized by the user.

The procedure is repeated for all candidate reference masses.

The final set of reference masses is denoted as \mathbf{M}^* . The mass shift GAM trends (Eq. 2) for each $M^* \in \mathbf{M}^*$ are fitted using the corresponding inlier points, and the

models are used to predict the observed values $\widehat{\mathbf{M}} = (\widehat{M}_i)$ in all ROI pixels (Fig. 1A). The use of GAM with penalized cubic splines aims at capturing the details of the shift trends better than the single spline regression. The intensities $I_i^\#$ of the matched peaks can be plotted to reveal their spatial distribution (Fig. 1B). Visual inspection provides additional validation of the consistency of the matched peaks. When multiple matches are available per pixel, the intensities of the peaks with the smallest r_i are plotted.

Pixel-wise mass recalibration

In this section, we will describe the procedure used to finally recalibrate the masses detected in all ROI pixels.

Although the selected reference ions have passed the filtering procedure, mismatched peaks may still be present if: (a) same peaks fall within two search windows so that they are assigned to two reference ions, (b) shifted peaks fall within the search window by chance (especially if the mass shift is large).

To reduce the chance of using mismatched reference ions masses, we apply a pixel-specific mass filter, based on the assumption that most of the true matches share a similar relative mass error. The details of the filtering procedure are reported in Additional file 1: Section S4.

Once the set of reference masses $\mathbf{M}_p^* = (M_{1,p}^*, M_{2,p}^* \dots) = (M_{k,p}^*)^4$ is determined for pixel p together with the values predicted by GAMs, we fit a model g (Eq. 1) specific for each type of ion analyzer (Additional file 1: Section S3). For Orbitrap analyzers, we fit a linear regression model [29]

$$\begin{aligned}
 M_{k,p}^* &\sim \text{Normal}(\eta_{k,p}, \sigma_p^2) \\
 \eta_{k,p} &= \beta_{0,p} + \beta_{1,p} \widehat{M}_{k,p}^\#
 \end{aligned}
 \tag{6}$$

Instead, for the TOF analyzer, we use a polynomial regression model [30, 31]

$$\begin{aligned}
 \sqrt{M_{k,p}^*} &\sim \text{Normal}(\eta_{k,p}, \sigma_p^2) \\
 \eta_{k,p} &= \beta_{0,p} + \sum_{d=1}^D \beta_{d,p} \sqrt{(\widehat{M}_{k,p}^\#)^d}
 \end{aligned}
 \tag{7}$$

where the optimal polynomial degree D , $D \leq 5$, corresponds to the smallest Bayesian Information Criterion (BIC) [32].

The fitted model is then used to predict the calibrated mass for all the detected ions in the specific pixel (Fig. 1C).

Schematically, the mass recalibration method follows these steps:

1. ROI pixels and reference mass database are given.
2. Search reference masses within a user-defined tolerance (in ppm units).
3. Keep masses with hits in at least 75% of ROI pixels.

⁴ For simplicity, we use the index p instead of p_i to denote the pixel dependency.

4. For each kept mass, fit a 2D FFT KDE of the pixels and observed mass values:
 - Fit a spline regression on the KDE local maxima.
 - Identify the inliers and outliers from the spline residuals.
 - Exclude the matched mass if the inliers points are less than 75% of ROI pixels or their residuals have a too large “mad” (Eq. 5).
 - If the reference is kept, fit a GAM using the inlier points.
 - Predict the observed masses in all ROI pixels using the fitted GAM.
5. For all ROI pixels:
 - Select the masses predicted by GAM in the pixel that have a common relative mass error.
 - Fit a linear or polynomial regression using the selected predicted masses and the corresponding theoretical values.
 - Use the regression model to predict the recalibrated masses in the pixel.

Recalibration method from La Rocca et al.

We compared the performance of the presented method with that described in La Rocca et al. (La Rocca) [19].

The ROI pixels masses of the DESI-MSI datasets were recalibrated using the default parameters, as described in their original work: bandwidth for the density estimation function (“st”) equal to 0.0005, Da tolerance for the identifications (“tl”) equal to 0.01, limit in Da for hits selection (“lm”) equal to 0.002. To perform a fair comparison with our method, we used the same database described in "[Selection of reference masses and shift trend modelling](#)" section

Results and discussion

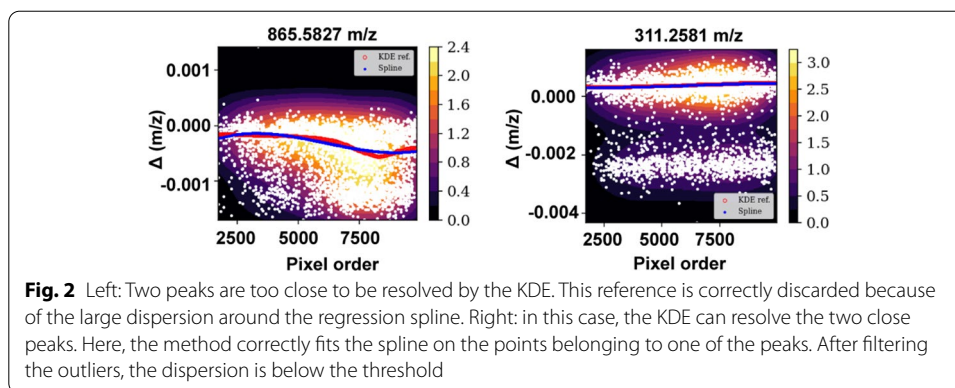
As a first experiment, we tested the robustness of the pixel-wise recalibration model and the reference mass filtering described in Additional file 1: Section S4. We simulated a series of pixel-wise reference masses and their observed values following the method described in Additional file 1: Section S5 (Additional file 1: Fig. S1). In all simulations, we observed a significant reduction of the median relative error, in both Orbitrap and TOF models. Also, the simulations confirmed the reduction of the heteroskedasticity of the residuals (Additional file 1: Fig. S3).

We then tested the DESI-MSI datasets after removing the pixels outside of the ROI as described in "[Selection of sample pixels](#)".

All recalibration parameters were kept fixed within the Orbitrap or TOF datasets, as described in the "[Methods](#)" section. For Orbitrap datasets, we used a search window $W = 20$ ppm, while for TOF datasets we used $W = 100$ ppm.

FFT KDEs for matched points were fitted using the “KDEpy” package for Python⁵ (v. 1.1.0). The cubic splines are fitted using the function *UnivariateSpline* available

⁵ <https://github.com/tommyod/KDEpy>.



in “SciPy” package for Python (v 1.7.1) [33]. The smoothing parameter s of the function was determined by fivefold cross-validation, among 30 values from 0.00001 to 0.1 evenly spaced in the \log_{10} scale, corresponding to the smallest mean squared error.

The KDE-based model of the reference mass shift showed its robustness in the case of close peaks, correctly discarding wrongly matched peaks (Fig. 2).

The GAMs were fitted using the pyGAM package (v. 0.8.0) for Python [34]. The spline penalization parameter was chosen among eleven values varying between 0.001 and 1000 evenly spaced in the \log_{10} scale, corresponding to the smallest generalized cross-validation value (GCV) [35].

In all tested DESI-MSI datasets, except for Orbitrap pancreas ES-, the selected reference masses for the pixel-wise recalibration well covered their acquisition m/z range (Additional file 1: Figs. S4–S5). The number of reference masses used for the recalibration varied between 15 and 174 (Additional file 1: Tables S2–S3).

In general, the reference masses were more concentrated in the m/z ranges below 400 and above 600, corresponding to small molecules and phospholipids, respectively (annotations are available online, see “Code and data availability” section).

In nine TOF datasets, there were pixels in which the optimal recalibration model was polynomial with a degree greater or equal to two. In particular, the distribution of the regression models coefficients revealed that the TOF mass shifts were greater than Orbitrap (Orbitrap error within 5 ppm, while TOF error up to 65 ppm) (Additional file 1: Table S4–S5). This observation may indicate a higher tendency of one class of analyzers to be subject to the changes of the environmental conditions, although these may depend on the unknown actual instrumental conditions for each dataset.

To evaluate the recalibration efficacy, we looked at two main effects: (1) the number of putative molecular annotations, (2) mass error of test ions.

The number of putative annotations was generated using the METASPACE website. METASPACE assigns molecule identities to MSI datasets based on a metabolite-signal match score (MSM) calculated from spectral and spatial measures and an FDR-estimation using a decoy strategy. Each assignment is characterized by an FDR equal to 5%, 10%, 20%, and 50%. The annotation is performed by database search. Because of the biological nature of the analyzed samples, we selected the following

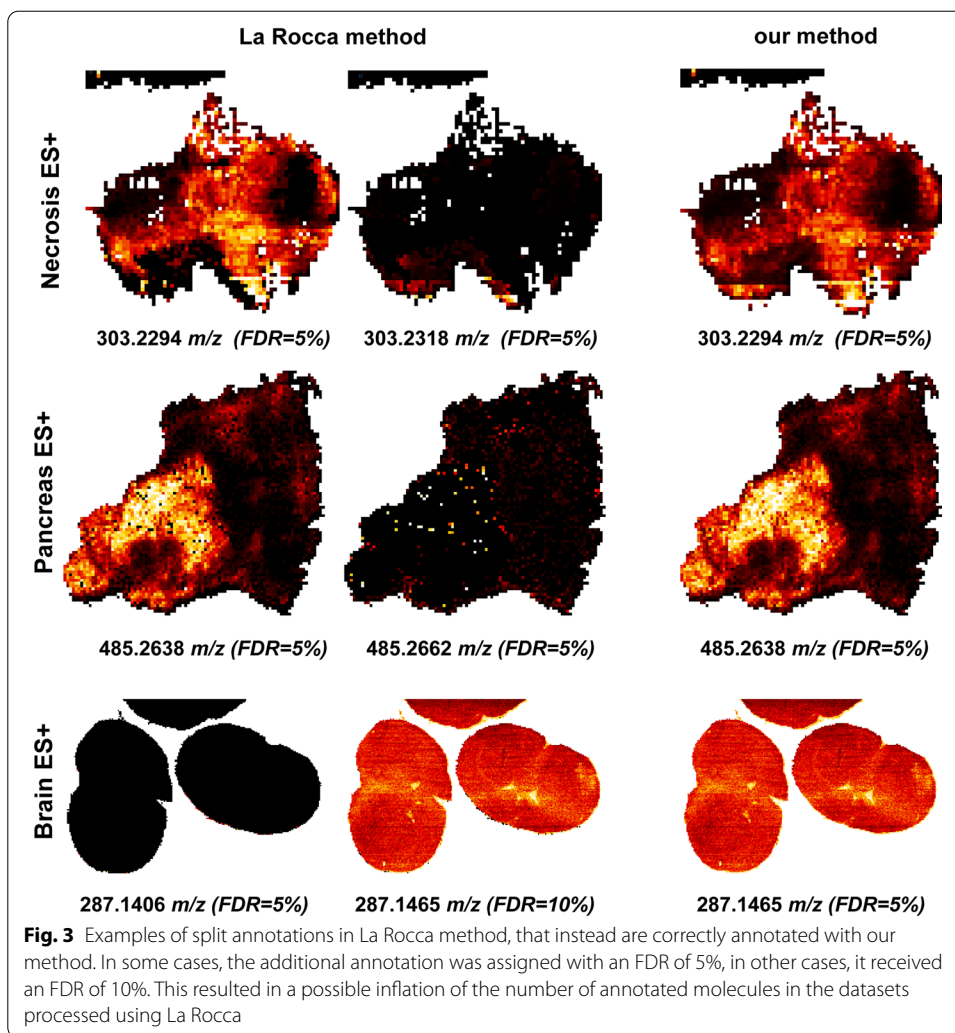
Table 1 Number of annotated molecules by METASPACE using the combined set of databases (FDR=5%)

Sample	Orbitrap					
	# Orig.	# LR	# Recal.	# Recal - Orig.	#LR - Orig	# Recal. - LR
adrenal gland ES-	589	627	630	41	38	3
breast ES- 1	280	276	279	-1	-4	3
breast ES- 2	186	184	187	1	-2	3
breast ES- 3	258	262	265	7	4	3
colon ES- 1	193	255	246	53	62	-9
colon ES- 2	29	266	277	248	237	11
Colon ES- 3	299	320	304	5	21	-16
liver ES-	319	350	335	16	31	-15
necrosis ES+	542	402	474	-68	-140	72
ovary ES-	146	153	147	1	7	-6
ovary ES+ 1	376	402	394	18	26	-8
ovary ES+ 2	14	163	148	134	149	-15
ovary ES+ 3	17	283	298	281	266	15
pancreas ES-	103	101	101	-2	-2	0
pancreas ES+	536	587	484	-52	51	-103
brain ES+	1150	27	1206	56	-1123	1179
kidney ES-	647	709	721	74	62	12
testis ES-	218	230	218	0	12	-12
thyroid ES-	149	162	153	4	13	-9
tumour ES-	893	911	910	17	18	-1
Sample	TOF					
	# Orig.	# LR	# Recal.	# Recal - Orig.	# LR - Orig	# Recal. - LR
brain ES- 1	3	2	87	84	-1	85
brain ES- 2	62	87	78	16	25	-9
brain ES+ 1	69	38	70	1	-31	32
brain ES+ 2	9	10	48	39	1	38
kidney ES- 1	76	49	68	-8	-27	19
kidney ES- 2	115	107	101	-14	-8	-6
liver ES- 1	74	74	70	-4	0	-4
liver ES- 2	3	25	83	80	22	58
liver ES- 3	83	102	97	14	19	-5
liver ES+	13	12	53	40	-1	41

The number of annotations for the original (non-recalibrated), processed with La Rocca (LR) and processed with our method datasets are reported in the first three columns respectively. The last three columns represent the difference between the number of annotations in the recalibrated datasets and the original dataset and between the two recalibration methods. Cases where our method results in a greater number of annotations are shaded in green, otherwise in red. In orange, the cases where the number of annotations is equal. The three datasets where we observed split annotations in LR are coloured in light blue

four databases among those available: (a) “ChEBI 2018-01”, (b) “CoreMetabolome v3”, (c) “HMDB v4”, (d) “LipidMaps 2017-12-12”.

We only considered the number of unique annotated *m/z* values corresponding to the stringent criterion of FDR=5% (column “*mz*” of the annotation tables). We considered



the unique annotated *m/z* values corresponding to the four databases combined. The annotations were performed using the ROI filtered raw and recalibrated peaks without applying any peak binning or spatial filter.

We checked whether the largest number of unique annotations was found in the original or recalibrated dataset.

In twenty-two cases ($\approx 73\%$), the recalibrated dataset received a larger number of assignments compared to the raw dataset when considering the combined databases. The recalibration increased the number of annotations by up to 281. In contrast, the largest decrease was observed in Orbitrap necrosis ES+ with 68 fewer annotations, equivalent to about 12.5% of the original annotations (Table 1). In this and pancreas ES+, where we observed the largest decrease of annotated molecules, the missing annotations received an FDR = 10% in the recalibrated datasets, due to a small number of missing pixels. Because no specific pattern was associated to these empty pixels, we hypothesize that they are due to statistical noise (Additional file 1: Fig. S6).

Using La Rocca method, in three Orbitrap datasets: “necrosis ES+”, “pancreas ES+” and “brain ES+”, we observed annotations corresponding to the same image split in two.

These were instead correctly annotated as a single molecule when processed with our method (Fig. 3). Evidently, this issue affected the final number of annotated molecules in the datasets processed using La Rocca method. In the case of pancreas ES+, it resulted in an inflation of the number of annotated molecules, since both the split parts received an FDR=0.05, while in the other two affected datasets, one of the split parts received an FDR=0.1. This result confirmed that having a global model (time-dependent shift trend) increases the robustness of the final recalibration models, compared to considering each pixel as an independent sample.

In all other datasets, we observed a balanced outcome, with a slightly better performance of La Rocca in the Orbitrap datasets, and significantly better performance of our method in the TOF datasets. Specifically, in four TOF datasets, we observed that La Rocca was unable to improve the number of annotated molecules as our method (Table 1).

Subsequently, we tested the mass accuracy of a set of test masses to evaluate the effect of the recalibration. We followed the same idea described in Boskamp et al. [18]. Since the actual molecular content of a sample was unknown, we compared the mass accuracy of a list of ions expected to be detected by DESI-MSI in biological tissue samples.

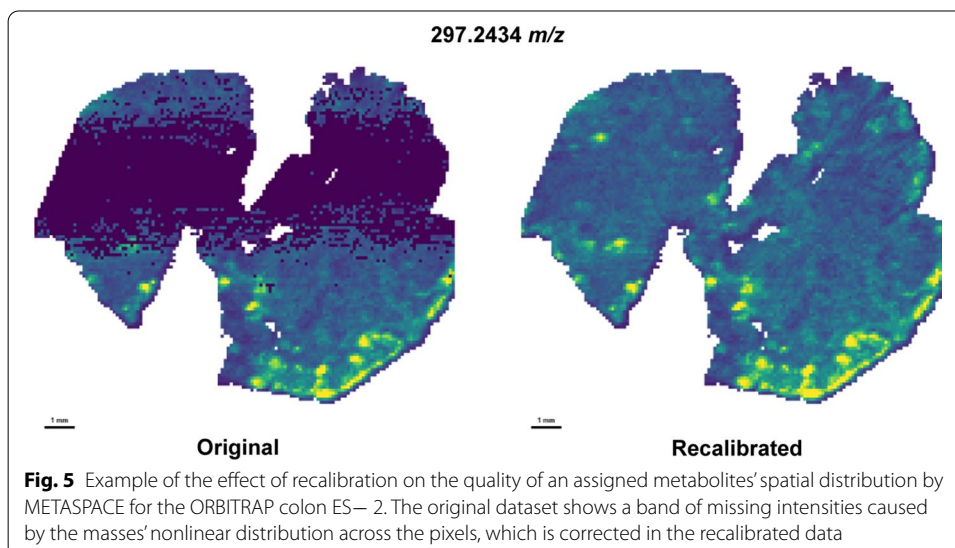
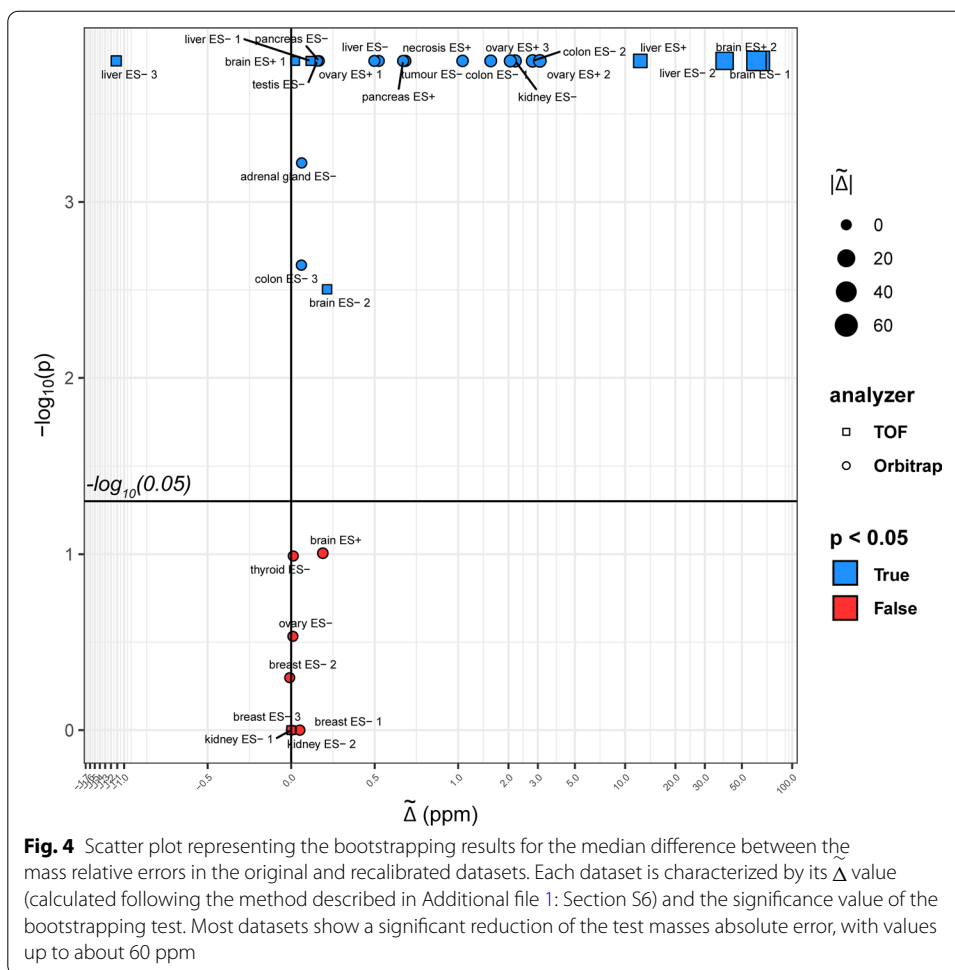
The test ions list was generated from the available annotated molecules in the public datasets of METASPACE. The list consisted of the monoisotopic form plus the first three identified isotopes detected in more than 10% of the datasets from the same tissue type and ion mode of the DESI-MSI dataset. We used a Python script from LaRocca et al.⁶ (version available in June 2021) to generate the list of candidate test masses. The mass values common to the set used for fitting the pixel-wise recalibration models were excluded. These masses could either belong to different databases from those used for the mass recalibration or could have been skipped during the recalibration due to the more stringent filtering (limit of mass error dispersion).

In all datasets, the test masses covered the acquisition m/z range, confirming that they represented a good set of probes for evaluating the recalibration effects (Additional file 1: Figs. S3–S4).

Using the procedure described in Additional file 1: Section S6, we selected the test masses within ± 2.5 ppm from the most abundant relative errors and calculated the median of the pixel-wise median difference between the absolute mass errors $\tilde{\Delta}$. We tested if we could reject the null hypothesis $H_0 : \tilde{\Delta} = 0$, using bootstrapping (number of repetitions equal to 10,000). We observed in twenty datasets a significant (Benjamini–Hochberg corrected p value < 0.05) decrease of the median relative error after the recalibration, with values varying between 0.02 and 64 ppm. In all these datasets, the final median relative error was below 3.7 ppm for Orbitrap, and below 6 ppm for TOF. In one case (TOF liver ES– 3), the median relative error increased by about 1.8 ppm, with a final median relative error equal to 4 ppm. In nine datasets, $\tilde{\Delta}$ was not significantly different from zero (Fig. 4, Additional file 1: Tables S6–S7).

Consistently, the improvement of mass accuracy observed in the TOF datasets corresponded to a substantial increase of annotated molecules.

⁶ https://github.com/LaRoccaRaphael/MSI_recalibration/blob/master/recalibration/internal_calibrants_generation.ipynb.



Furthermore, as already discussed before, the recalibration improved the ion image quality. After removing the mass shifts, all pixels correctly displayed the expected molecular spatial distributions (Fig. 5).

When compared with La Rocca method, our method performed slightly worse in the Orbitrap datasets, with a difference below 1 ppm in all cases (Additional file 1: Table S8). However, in the TOF datasets, our method outperformed La Rocca, which failed to improve the mass accuracies when the original relative errors were above 15 ppm (Additional file 1: Table S9). Since in these cases, the original error was large, we tested La Rocca with an optional set of parameters provided by the authors on GitHub for “low precision analyzers”: “st”=0.03, “tl”=0.8, “lm”=0.08. Using these parameters, we observed an improvement in the datasets “brain ES– 1” and “liver ES– 2”, passing from 69 to 6 ppm and from 46 to 21 ppm, respectively. However, in all datasets, La Rocca still performed worse than our method, with also a reduction of the mass accuracy in some datasets (Additional file 1: Table S10).

We tested the effect of using polynomial models for the mass recalibration of the TOF datasets, forcing a linear model. The results of the relative mass errors for the test masses were close to those obtained in the original models (Additional file 1: Table S11) suggesting that also linear models would be feasible for TOF datasets.

We found that among the Orbitrap datasets, three masses were always used to fit the recalibration models: 253.2173 m/z, 269.2486 m/z, 279.2329 m/z for ES–, and 441.2975 m/z, 457.2715 m/z, 469.3288 m/z for ES+. Among the TOF datasets, we observed four masses always used as a reference for ES–: 255.2329 m/z, 281.2486 m/z, 283.2642 m/z and 303.2329 m/z. For ES+, we observed 14 masses always used: 309.2036 m/z, 621.4855 m/z, 621.4878 m/z, 649.5168 m/z, 649.5191 m/z, 713.4518 m/z, 723.4935 m/z, 739.4675 m/z, 782.5670 m/z, 798.54077 m/z, 798.5410 m/z, 820.5253 m/z, 824.5566 m/z, 826.5721 m/z. Although these could be used as the only reference masses, using a database is a more robust solution for unknown MSI datasets.

Conclusion

Despite its enormous potential to capture the spatial characteristics of the metabolic and proteomic mechanisms in a wide range of biological samples, MSI remains a relatively young technology. Advancements in fundamental data quality control are necessary to transform it into a more reliable approach.

Accuracy of measured masses is particularly challenging in imaging approaches since usually a run consists of tens of thousands of pixels and may require several hours to complete. Thus, the changes of the instrumental condition during this time may invalidate its initial configuration, resulting in mass drifts that correlate with time or pixel order.

Here, we have presented a computational workflow that exploits the presence of typical biological molecules in most of the sample spectra and uses them as a set of reference ions for applying a spectra-wise lock mass correction.

Using a global set of reference ions for the entire MSI dataset allowed us to test the assignment’s validity by visualizing the corresponding spatial distributions. Furthermore, we modeled the mass drifts as a smooth function of the acquisition time, as

expected in acquisitions occurring in a controlled environment. Thus, these models represent an additional level of evidence about the assignments' correctness.

FFT KDE-based reference match filtering, together with GAMs, proved robust against outliers and false-positive reference ions, for instance, in the presence of close peaks with a complementary spatial distribution to that of the actual reference ions peaks.

We showed that the presented approach could improve the mass accuracy, removing the nonlinear fluctuations of the measured masses. Additionally, our approach improved the molecular assignment to the detected peaks, using state-of-the-art molecular annotation methods, such as METASPACE, and their quality in terms of assigned spatial distributions.

In datasets with already accurate masses, the recalibration can introduce errors due to its statistical nature. For this reason, it is crucial to evaluate the test masses accuracies before and after performing the recalibration, using the methods described here.

We employed pixel-wise recalibration models specific to the physics of the MS analyzer. This is an essential aspect of the procedure since these characteristics influence the statistical properties of the signal generated from the detected ions.

It is also important to underline that we designed the mass shift models to be simple, considering only the time-dependency of the acquisition. Unfortunately, this means that several complex aspects of the phenomena driving the mass drifts are not captured. This is a necessary trade-off between generality and reduction of the risk of overfitting. More realistic models may be designed if additional variables are tracked, such as temperature, humidity, just to name a few.

We showed that computational models could effectively reduce the mass drifts that affect MSI datasets. However, it is crucial to stress that they can represent a temporary solution to this problem until technological solutions capable of performing an accurate online calibration will not be available.

When compared with a similar method, the results confirmed that considering the non-independence of the ion mass shifts increased the accuracy of the results, avoiding split annotations and being able to correct substantial relative mass errors.

The presented work has limitations. Although we used an extensive list of publicly available molecular masses as references, some peculiar molecules for unknown samples may be missing. For this reason, building a database of accurately identified molecules in previous studies is crucial.

Another limitation of the approach is that it can only be applied to the sample-related ROI pixels. Thus, when selecting these pixels, the spatial information necessary to identify possible non-sample-related signals is lost. However, this difficulty can be easily overcome by applying spatial-aware filters, such as SPUTNIK [36], before the recalibration. In this way, most of the signal correlated with regions outside of the ROI is removed before performing the pixels selection. This aspect is crucial for the correct biological interpretation of the observed molecular spatial patterns.

In the future, we will work to extend the method to other types of ion analyzers and will study the possible integration with the usage of external lock mass reference ions.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-022-04671-5>.

Additional file 1. Details about the MSI pre-processing and mass recalibration algorithm. Results of the comparison between original and recalibrated mass accuracies.

Acknowledgements

We thank Clara L. Feider, Meghan P. Friis, Kyana Garza, Emrys Jones, James McKenzie, Nicole Strittmatter, Jialing Zhang, for sharing their datasets on METASPACE.

Author contributions

PI. designed the model and the computational framework and analyzed the data. H.X.H. and V.W. performed the DESI-MSI experiments. M.R.L. and Z.T. conceived the study and oversaw overall direction and planning. All authors discussed the results and contributed to the final manuscript. All authors read and approved the final manuscript.

Funding

PI is supported by the NIHR Imperial Biomedical Research Centre (BRC). HH is supported by the UK Dementia Research Institute which receives its funding from UK DRI Ltd, funded by the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK. This work was supported by the Medical Research Council UK Consortium for MetAbolic Phenotyping (MAP UK) [grant number MR/S010483/1], infrastructure support was provided by the NIHR Imperial Biomedical Research Centre (BRC).

Availability of data and materials

The Python code is freely distributed under the MIT license and can be downloaded from https://github.com/paoloinglese/DESI_MSI_recal, together with the reference masses databases. Additional results consisting of all annotations from METASPACE, annotated reference masses, number of reference masses per pixel and model coefficients are available on Open Science Framework at <https://osf.io/q4h76/>.

Declarations

Ethics approval and consent to participate

Three snap-frozen brains of a C57BL/6 mouse model were purchased from Charles River Laboratories. No experiment was carried out on live animals.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 25 August 2021 Accepted: 30 March 2022

Published online: 15 April 2022

References

1. Hillenkamp F, Karas M, Beavis RC, Chait BT. Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal Chem*. 1991;63:1193A-1203A.
2. Herzog RFK, Viehböck FP. Ion source for mass spectrography. *Phys Rev*. 1949;76:855-6.
3. Takáts Z, Wiseman JM, Gologan B, Cooks RG. Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science*. 2004;306:471-3.
4. Mascini NE, Teunissen J, Noorlag R, Willems SM, Heeren RMA. Tumor classification with MALDI-MSI data of tissue microarrays: a case study. *Methods*. 2018;151:21-7.
5. Guenther S, et al. Spatially resolved metabolic phenotyping of breast cancer by desorption electrospray ionization mass spectrometry. *Cancer Res*. 2015;75:1828-37.
6. Schwamborn K, et al. Identifying prostate carcinoma by MALDI-imaging. *Int J Mol Med*. 2007;20:155-9.
7. Dória ML, et al. Epithelial ovarian carcinoma diagnosis by desorption electrospray ionization mass spectrometry imaging. *Sci Rep*. 2016;6:1-11.
8. Karlsson O, Hanrieder J. Imaging mass spectrometry in drug development and toxicology. *Arch Toxicol*. 2017;91:2283-94.
9. Drexler DM, Reily MD, Shipkova PA. Advances in mass spectrometry applied to pharmaceutical metabolomics. *Anal Bioanal Chem*. 2011;399:2645-53.
10. Nilsson A, et al. Mass spectrometry imaging in drug development. *Anal Chem*. 2015;87:1437-55.
11. Prideaux B, Stoeckli M. Mass spectrometry imaging for drug distribution studies. *J Proteomics*. 2012;75:4999-5013.
12. LaBonia GJ, Lockwood SY, Heller AA, Spence DM, Hummon AB. Drug penetration and metabolism in 3D cell cultures treated in a 3D printed fluidic device: assessment of irinotecan via MALDI imaging mass spectrometry. *Proteomics*. 2016;16:1814-21.
13. Ellis SR, Bruinen AL, Heeren RMA. A critical evaluation of the current state-of-the-art in quantitative imaging mass spectrometry. *Anal Bioanal Chem*. 2014;406:1275-89.

14. Swales JG, et al. Quantitation of endogenous metabolites in mouse tumors using mass-spectrometry imaging. *Anal Chem*. 2018;90:6051–8.
15. Hamm G, et al. Quantitative mass spectrometry imaging of propranolol and olanzapine using tissue extinction calculation as normalization factor. *J Proteomics*. 2012;75:4952–61.
16. Ràfols P, del Castillo E, Yanes O, Brezmes J, Correig X. Novel automated workflow for spectral alignment and mass calibration in MS imaging using a sputtered Ag nanolayer. *Anal Chim Acta*. 2018;1022:61–9.
17. Li Y, Li L. Mass accuracy check using common background peaks for improving metabolome data quality in chemical isotope labeling LC–MS. *J Am Soc Mass Spectrom*. 2019;30:1733–41.
18. Boskamp T, et al. Using the chemical noise background in MALDI mass spectrometry imaging for mass alignment and calibration. *Anal Chem*. 2020;92:1301–8.
19. La Rocca R, et al. Adaptive pixel mass recalibration for mass spectrometry imaging based on locally endogenous biological signals. *Anal Chem*. 2021;93:4066–74.
20. Tillner J, et al. Faster, more reproducible DESI-MS for biological tissue imaging. *J Am Soc Mass Spectrom*. 2017;28:2090–8.
21. Satopää V, Albrecht J, Irwin D, Raghavan B. Finding a 'needle' in a haystack: detecting knee points in system behavior. In: Proceedings—international conference on distributed computing systems. 2011. p. 166–71. <https://doi.org/10.1109/ICDCSW.2011.20>.
22. Michelsen P, Karlsson AÅ. Accurate mass determination of a metabolite of a potential diagnostic imaging drug candidate by high performance liquid chromatography with time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom*. 1999;13:2146–50.
23. Hastie T, Tibshirani R. Generalized additive models. *Stat Sci*. 1986;1:297–310.
24. Sud M, et al. LMSD: LIPID MAPS structure database. *Nucleic Acids Res*. 2007;35:D527–32.
25. Wishart DS, et al. HMDB: the human metabolome database. *Nucleic Acids Res*. 2007;35:D521–6.
26. Smirnov KS, Forcisi S, Moritz F, Lucio M, Schmitt-Kopplin P. Mass difference maps and their application for the recalibration of mass spectrometric data in nontargeted metabolomics. *Anal Chem*. 2019;91:3350–8.
27. Silverman BW. Algorithm AS 176: kernel density estimation using the fast Fourier transform. *Appl Stat*. 1982;31:93.
28. Silverman BW. Density estimation: for statistics and data analysis. CRC Press; 2018.
29. Wenger CD, McAlister GC, Xia Q, Coon JJ. Sub-part-per-million precursor and product mass accuracy for high-throughput proteomics on an electron transfer dissociation-enabled orbitrap mass spectrometer. *Mol Cell Proteomics*. 2010;9:754–63.
30. Vera CC, Zubarev R, Ehring H, Hakansson P, Sunqvist BUR. A three-point calibration procedure for matrix-assisted laser desorption/ionization mass spectrometry utilizing multiply charged ions and their mean initial velocities. *Rapid Commun Mass Spectrom*. 1996;10:1429–32.
31. Christian NP, Arnold RJ, Really JP. Improved calibration of time-of-flight mass spectra by simplex optimization of electrostatic ion calculations. *Anal Chem*. 2000;72:3327–37.
32. Schwarz G. Estimating the dimension of a model. *Ann Stat*. 1978;6:461–4.
33. Virtanen P, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17:261–72.
34. Servén D, Brummitt C, Abedi H. dswah/pyGAM: v0.8.0 (2018). <https://doi.org/10.5281/ZENODO.1476122>.
35. Wood SN. Generalized additive models: an introduction with R. 2nd ed. Chapman and Hall/CRC; 2017. <https://doi.org/10.1201/9781315370279>.
36. Inglese P, Correia G, Takats Z, Nicholson JK, Glen RC. SPUTNIK: an R package for filtering of spatially related peaks in mass spectrometry imaging data. *Bioinformatics*. 2019;35:178–80.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

