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

Paolo Inglese[1,2], Helen Huang[2], Vincen Wu[2], Matthew R Lewis[1,2], Zoltan Takats[2]

[1]National Phenome Centre, Imperial College London, Hammersmith Campus, IRDB Building, London W12 0NN, United Kingdom

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

1 Abstract

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 nominal 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. Here, we present a post-acquisition computational method capable of reducing the observed mass drift in biological samples, exploiting the presence of typical molecules with a known mass-to-charge ratio. The method is tested on time-of-flight and Orbitrap mass spectrometry analyzers interfaced to a desorption electrospray ionization (DESI) source.

2 Introduction

In the last twenty years, mass spectrometry imaging (MSI) has gained increasing attention as a technology capable of capturing the molecular spatial patterns from sample surfaces. Using various ionization techniques, such as desorption electrospray ionization mass spectrometry (DESI-MSI), MSI technologies can collect thousands of ions localized in small areas (ranging from $100\mu m$ to $10\mu m$) of the sample surface. 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 a wide range of application areas, from cancer research (Dória et al., 2016; Guenther et al., 2015; Mascini, Teunissen, Noorlag, Willems, & Heeren, 2018; Schwamborn et al., 2007) to pharmacology (Drexler, Reily, & Shipkova, 2011; Karlsson & Hanrieder, 2017; LaBonia, Lockwood, Heller, Spence, & Hummon, 2016; Nilsson et al., 2015; Prideaux & Stoeckli, 2012). Among the available ionization sources, DESI-MS (Takáts, Wiseman, Gologan, & Cooks, 2004) has gained popularity and been widely applied thanks to the relatively small number of steps required for sample preparation. However, as an emerging technology, MSI still faces implementation challenges for controlling the quality of data produced (Ellis, Bruinen, & Heeren, 2014; Hamm et al., 2012; Swales et al., 2018). The accuracy of mass-to-charge ratio (m/z) measurements under sustained use are chief among these concerns. MSI acquisitions often consist of tens of thousands of spectra (pixels), corresponding to several hours, during which m/z measurements can drift in a manner that correlates with the acquisition (pixel) order. For example, a small image of 10 mm x 10 mm, acquired at a running speed of 1 scan/s, with a spatial resolution of 100 µm, requires about 2.78 h. The situation worsens when multiple samples (thus with larger image sizes) are acquired in a single run. This problem results in confounded data processing and analysis. The mass shift can result in wrong molecular assignments to signal interest and can be

detrimental to the analyzed spatial patterns. Shifted masses can be assigned to two adjacent m/z values, determining a split of the corresponding image in two. Additionally, as MSI relies on only a single dimension of separation for measured ions, errors of a few parts-per-million (*ppm*) can result in hundreds of candidate molecular identities for a given ion species, making the subsequent assignment generation 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., LC-MS) include periodic recalibration or more frequent adjustment of the calibration based on the measurement of one or more standard reference materials during the acquisition. Unfortunately, this approach is unfeasible for the current MSI technology since the acquisition is performed continuously while the probe moves across the sample surface. A way to emulate such an approach is injecting a reference during the acquisition (Olsen et al., 2005) or, in the case of MSI, adding the standard material to the sample (Ràfols, Castillo, Yanes, Brezmes, & Correig, 2018). Although these approaches can achieve a good 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-linearity of the mass drifts 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 therefore map the measured m/z onto their corresponding recalibrated values (Li & Li, 2019). This class of methods is based on the a priori knowledge that specific molecules are likely to be in the analyzed sample.

Thus, contextual information about the fundamental biochemical properties of the sample is of critical importance.

Boskamp et al. (Boskamp et al., 2020) showed that endogenous signal (chemical noise) could improve the mass accuracy in MALDI-MSI of peptide datasets.

Recently, La Rocca et al. (La Rocca et al., 2021) have presented an algorithm to recalibrate MSI datasets using a linear fit on the observed mass errors from endogenous biological peaks. The candidate reference ions were chosen among those most often detected in a given sample type, based on the publicly available MSI datasets on METASPACE (Palmer et al., 2016). Each spectrum was treated as independent, potentially resulting in different reference ions for each of them.

Here, we present a different approach based on a fixed set of reference ions across the entire MSI dataset. The advantages of having global references are: a) global references define a fixed boundary for the m/z extrapolation, b) quality of the matching procedure for the reference ions can be evaluated by visualizing their spatial distribution, c) time-dependent models for each reference can be used to reduce the effect of the random component of the mass error.

We test the performance of the presented approach on DESI-MSI datasets acquired using orthogonal acceleration time-of-flight (oaTOF) and Orbitrap analyzers, using calibration equations specific for the two types of analyzers.

In the presented work, we leverage 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. Since the chosen reference ions may be missing in some of the measured spectra, a polynomial model is employed to infer the mass errors in those spectra, exploiting the assumption that the individual ion masses drift follows a smooth function of the time (or pixel order). The

approach is demonstrated on exemplar biological tissue sections in both the positive and negative ion modes.

3 Samples for DESI-MSI

3.1 TOF MS

DESI-MSI data were acquired from six biological samples were to test the effects of recalibration, mouse (mus musculus) brain, and pig (sus scrofa domesticus) liver, using a TOF MS.
Three brains of a C57BL/6 mouse model were purchased from Charles River Laboratories, while pig liver samples were obtained from a local supermarket.
Methanol and water were purchased from Sigma-Aldrich (St Louis, MO, USA).
All samples were stored at -80°C before data acquisition. A coronal brain sections and one liver section (thickness = 10µm) were cryosectioned from each sample using a cryostat (Thermo Scientific™ Cryotome™ FSE) with the chamber at -20°C and mounted on SuperFrost Plus glass slides (Thermo Fisher Scientific, Waltham, MA, USA).
The six MSI datasets were acquired in positive and negative ion mode, as follows: brain1(+), brain2(-), brain3(+), liver1(-), liver2(-), liver3(+), with the sign between the parenthesis to indicate the ion mode.

3.2 Orbitrap MS

Six public DESI-MSI datasets were downloaded from METASPACE[1]. The set consisted of images captured from a section of one pig liver (negative ion mode), two human ovary (positive ion mode), one human colon (negative ion mode), one rat (rattus norvegicus) kidney (negative ion mode), one rat testis (negative ion mode).
All six datasets were reported to be acquired using a DESI stage interfaced with an Orbitrap MS analyzer.
The datasets consisted of centroided RAW peaks. Thus no peak detection of denoising was applied.
The six MSI datasets were acquired in positive and negative ion mode: liver(-), ovary(+), ovary(+), colon(-), kidney(-), testis(-), with the sign between the parenthesis to indicate the ion mode.

4 DESI-MSI parameters

TOF DESI-MSI was performed on two different Xevo-G2XS mass spectrometers coupled with Acquity UPLC (Waters, Milford, MA, USA), operated under sensitivity mode. The DESI ion source and a two-dimensional DESI stage were from Prosolia (Indianapolis, IN, USA). The custom-built inlet capillary was heated to 270 degrees unless otherwise specified. The MSI experiments were conducted over the mass range of 50 -1000 in the positive ion mode with a spray solvent (MeOH: H2O = 95: 5) at a flow rate of 1.5 µl/min. The DESI source parameters were optimized on a pig liver tissue for best sensitivity and include the following: spray voltage at 4 kV; a sprayer-to-surface distance of 1.5 mm; a sprayer-to-inlet distance of 8 mm; spray impact angle of 75 °; collection angle of 10 °; nebulizing gas (N2) pressure, 4.5 bar at the flow rate of 1.5 µl/min; The MS parameters

[1] https://metaspace2020.eu/

were as following: scan time, 0.186 s; source temperature, 150 °C; sampling cone voltage, -40 V; source offset, -80 V. The sprayer head was grounded throughout the run.
The mouse brain samples were scanned at a rate of 75 µm/s horizontally separated by a 75 µm vertical step, while the pig liver sample was scanned at a rate of 100 µm/s.
The RAW spectra were converted into imzML using the MassLynx SDK (v4.7.0) and Python's pyimzML package[2].

5 Methods

5.1 Signal denoising

A two-step procedure was applied to remove baseline (BN) and signal noise (SN) independently. First, the BN level is identified using a modified version of the method described (Zhurov, Kozhinov, Fornelli, & Tsybin, 2014). As in the original method, we assume that, in a mass spectrum, the signal intensities are marginally distributed as a mixture of two components: a BN component and a sample signal component (with its noise, SN). The BN filtering's main scope identifies the optimal point of separation between two distributions (the "cut level") such that only the sample signal is preserved. To avoid the ambiguity introduced by selecting the number of histogram bins used in Zhurov et al., we use a different approach to identify the cut level. All signal intensities are sorted in decreasing order, and the cut point is chosen at the *elbow point* of the curve. This choice is based on the assumption that the noise level represents the intensity at which the most noticeable intensity distribution change is observed (second derivative of the curve). Since noise can affect the estimated value of the second derivative, we identify the elbow point using a simplified version of the *Kneedle* algorithm (Satopää, Albrecht, Irwin, & Raghavan, 2011) that only looks at the farthest point from the diagonal line connecting the two extremal points of the curve.
All zeros are removed from the signal, and, as described in Zhurov et al., signal intensities are log-transformation before the BN level estimation. The elbow point is used as a cut-off, and all signal intensities below that value are set to zero.
The second component of the noise, SN, is assumed to be signal-dependent (multiplicative noise). To remove SN, we use a smoothing procedure based on a local transformation of the signal through Savitzky-Golay convolution. All negative intensities generated by the application of the Savitzy-Golay filter are set to zero.
The Savitzky-Golay filters were estimated with a degree equal to three and a half-window equal to four points in the presented experiments.
Smoothing followed by peak detection was preferred to one-step approaches, such as wavelet peak-picking (Du, Kibbe, & Lin, 2006), for computational efficiency reasons.

5.2 Peak detection

Denoised signal (as described in the previous section) was used to determine peak mass-to-charge values and intensities.
Peaks and base points (half peak intensity) were detected using the function *find_peaks* available in Python's Scipy library (Virtanen et al., 2020). Minimum width of 3 points and a minimum prominence equal to half of the BN value were used to retain the signal peaks. The centroid m/z value calculated within the base points interval was assigned to the peaks.

[2] https://github.com/alexandrovteam/pyimzML

Since the smoothing modifies the original intensities, we assigned each peak the maximum RAW intensity in the interval between the peak base points. This ensures that the peak intensity is consistent with the original RAW data.

5.3 Region-of-interest

One of the most critical steps of the proposed workflow is the identification of sample-related pixels.
This step was performed immediately after detecting the peaks, as it is necessary to perform m/z recalibration, as described in the following section.
Since only the RAW peaks were available at this stage, we apply a uniform binning with a bin size of 1 m/z to match the peak intensities. This rough matching was accurate enough to preserve the information necessary to discriminate between the sample and off-sample spectra.
A linear support vector machines (SVM) model was fitted on a subset of user-defined labeled pixels. A graphical user interface (GUI) was used to plot the RGB image corresponding to the binned dataset's first three principal components scores (Figure S1). The user then annotated some of the pixels as belonging to the sample or off-sample (or, equivalently, background), drawing the annotated regions' contour. Usually, annotating around 10% of the pixels was enough to achieve a perfect classification.
Subsequently, the SVM model fitted on the annotated pixels was used to predict all pixels' class, providing a binary map of the region-of-interest (ROI). The user could manually refine the segmentation.
The ROI mask was saved in HDF5 format (Koranne, 2011) to be used in the workflow's following steps.

5.4 Mass recalibration

Let us consider an MSI dataset representing a collection of N sequentially acquired spectra (where $N = N_{off} + N_{sam}$ is the sum of the number of sample and off-sample pixels). Given the i-th spectrum, we define the measurement error $\Delta(m_{th}, i)$ of an ion with a *measured mass* m_{obs} and *nominal mass* m_{th}

$$\Delta(m_{th}, i) = m_{obs}(i) - m_{th} \tag{1}$$

Given the values of $\Delta(m_{th}, i)$ for all measured ions, the correct masses can be recalculated by subtracting the error

$$m_{th}(i) = m_{obs}(i) - \Delta(m_{th}, i) \tag{2}$$

However, it may be challenging to know the value of Δ for all detected ions because their nominal masses are mostly unknown, especially in untargeted experiments.
One common strategy is to use a set of reference ions, called *lock masses,* known to be present in the spectra. Therefore, a mathematical model is applied to estimate the error for all observed masses, from the set of lock mass errors, obtaining the recalibrated masses m_{cal}. In our experiments, we used a model specific for each type of MS analyzer.
For TOF data, we used a polynomial in the square root of the observed lock masses (Christian, Arnold, & Really, 2000; Vera, Zubarev, Ehring, Hakansson, & Sunqvist, 1996)

$$\sqrt{m_{cal}} = \alpha_0 + \sum_{p=1}^{P} \alpha_p \sqrt{m_{obs}^p} + \varepsilon_m \tag{3}$$

where P is the polynomial degree (usually $P \leq 5$) and ε_m represents the random error, $\varepsilon_m \sim N(0, \sigma_m^2)$.
In the case of Orbitrap analyzer, instead, we used a linear model in the observed lock masses, based on Wenger et al. (Wenger, McAlister, Xia, & Coon, 2010),

$$m_{cal} = \beta_0 + \beta_1 m_{obs} + \varepsilon_m \tag{4}$$

The parameters were fitted using the least-squares method between the observed and nominal lock masses. Once the parameters of the model were estimated, it was possible to calculate the calibrated mass m_{cal} for all the ions detected in the given spectrum.
An essential aspect of this approach is that, like all regression models, the estimated models become less reliable as we move far from the observed points (extrapolation). Thus, caution must be used when the lock masses cover a small portion of the entire m/z range. Furthermore, this approach requires that the *lock masses* are detected in the spectrum that needs to be recalibrated.
The lock masses can be defined in each spectrum (La Rocca et al., 2021) or globally (same for all spectra). For our models, we used the latter approach. Using a fixed set of lock masses allowed us to control the entire dataset's extrapolation regions and distinguish between modeled and random mass shifts, as discussed in the following.
Under the assumption that the mass drift correlates with the acquisition order, we described the errors of each lock mass as a smooth function of the pixel order:

$$m_{obs}(i) = \beta_0 + \sum_{p=1}^{P} \beta_p i^p + \varepsilon_m, \qquad i = 1, \dots, N \tag{5}$$

where the degree $P \leq 5$ can be estimated by cross-validation.
Nearest peaks, such as shoulder or not well-resolved peaks, are the principal cause of false positives in the reference ion search. To reduce the likelihood of estimating the model using close peaks, a random sample consensus (RANSAC) (Fischler & Bolles, 1981) polynomial was employed. Using a RANSAC regression allowed us to reduce the effect of false positives and outliers on the fitted model. This is especially important in close peaks that are spatially distributed in two complementary regions. They might be wrongly considered as generated by the same ion, using the only percentage coverage method.
Using the fitted model, we estimated the observed lock mass values $\widehat{m}_{obs}$ in all spectra and then used these values to estimate the parameters of Eq. 3 and 4 for the individual spectra recalibration.
Defining a global model for the reference ions mass drift has two advantages: a) the risk of fitting the random component of the mass error (represented by the ε_m term of Eq. 5) in each spectrum is reduced, b) the model defined in Eq. 5 can be used to estimate the mass error in all pixels. The latter is especially useful in biological samples since, due to their molecular heterogeneity, lock masses may be detected in a fraction of the measured spectra.
To achieve an accurate estimation of the reference ions' mass errors, they are required to be detected in a significant fraction of the analyzed spectra. However, due to the expected missing ions in the region outside of the ROI, we only restricted the analysis to the sample-related ROI pixels. The identification of the ROI is described in Section 5.3.
For our experiments, we used as reference a set of endogenous ions with nominal masses $M' = \{m'_1, m'_2, \dots, m'_R\}$ ($R > 5$) that were expected to be found in large part of the ROI spectra.

It is essential to underline that reference ions are sample-dependent. We have identified a set of ions that consistently occur in many biological tissue samples for the presented experiments. These ions mainly represent structural lipids or small molecules necessary for the most fundamental biochemical processes (Table 1).
One crucial requirement for good reference peaks is that they should be the most intense peaks in a reasonably small m/z neighborhood ($\pm 0.1\ m/z$) in almost all spectra. This property reduces the likelihood of mixing two different signals during the search.

5.5 Reference ion matching

Once chosen, the reference ions are searched in all spectra within windows $w \in W$ (measured in *ppm*) around their theoretical m/z value. The reference ions were searched among the peaks with an intensity greater than 10 times the spectrum BN level to improve the mass error estimation's robustness. In the case of Orbitrap data, where the BN level was missing, all peaks were used.
If multiple peaks matched in one spectrum within a search window value, only the most intense peak was retrieved.

For each reference $m'_r \in M'$ and search window value w, we calculate:

- **matched pixels:** set of pixel indices $\mathcal{P}' \subseteq \mathcal{P}$ where the reference peak is detected with mass value $\hat{m}'_r$

$$\mathcal{P}' = \{i \in \mathcal{P} \mid \exists\, \hat{m}'_r \ \ni \hat{m}'_r \ \in [m'_r - \delta(m'_r, w), m'_r + \delta(m'_r, w)]\} \tag{6}$$

  where $\mathcal{P}$ represents the set of ROI pixels and $\delta(m, w)$ is the search window in m/z units:

$$\delta(m, w) = m \cdot w \cdot 10^{-6}\ \ [m/z] \tag{7}$$

- **coverage percentage**: percentage of pixels where the reference peak is found

$$\frac{|\mathcal{P}'|}{|\mathcal{P}|} * 100 \tag{8}$$

- **mass error (in *m/z* units):**

$$\Delta(m'_r, i) = \hat{m}'_r(i) - m'_r, \qquad i \in \mathcal{P}' \tag{9}$$

All search windows corresponding to a coverage percentage smaller than 50% of the ROI pixels were discarded to fit an accurate model.
Afterward, for each reference ion, we defined the *optimal search window* $\hat{w}_r \in W$ as the elbow point (using the same approach used for the BN estimation) of the coverage percentage sorted by w. This choice reflects that if the reference peak is separated enough from adjacent higher peaks, its coverage reaches a plateau as the matched spatial distribution converges to its actual distribution.
After calculating the mass errors from Eq. 9, we fitted the polynomial model (Eq. 5) and subsequently estimated the calibrated masses from Eq. 3 and 4, as described above.

Positive Ion Mode		
***m'*[*m/z*]**	**Formula**	**Adduct**
309.2036	C16H30O4	+Na
104.1070	C5H13NO	+H

184.0733	C5H14NO4P	+H
296.0660	C8H20NO6P	+K
381.0794	C12H22O11	+K
562.3269	C26H54NO7P	+K
577.5190	C37H68O4	+H
782.5670	C42H82NO8P	+Na
820.5253	C44H80NO8P	+K
867.6497	C57H86O6	+H
907.7725	C57H104O6	+Na
923.7464	C57H104O6	+K
Negative Ion Mode		
m'[m/z]	**Formula**	**Adduct**
255.2330	C16H32O2	-H
124.0074	C2H7NO3S	-H
145.0619	C5H10N2O3	-H
146.0459	C5H9NO4	-H
175.0248	C6H8O6	-H
281.2486	C18H34O2	-H
283.2643	C18H36O2	-H
303.2330	C20H32O2	-H
500.2783	C25H44NO7P	-H
599.3202	C27H53O12P	-H
742.5392	C41H78NO8P	-H
885.5499	C47H83O13P	-H
909.5499	C49H83O13P	-H

Table 1 - List of reference ions used for the experiments. The selected molecules are mainly expected to be found in the majority of biological samples

.

6 Results and discussion

To evaluate the proposed recalibration method's efficacy, we looked at two main effects: a) the mass error and the spatial distributions of selected ions, b) the number of putative molecular annotations.
The number of putative annotations was tested similarly to La Rocca et al. (La Rocca et al., 2021). We submitted the original and recalibrated peaks to METASPACE and evaluated the number of annotated formulas in each dataset.
METASPACE molecular assignments are based on a metabolite-signal match score (MSM) calculated from spectral and spatial measures and an FDR-estimation using implausible adducts. Given an MSI dataset, METASPACE assigns molecules to signal with an estimated FDR equal to 5%, 10%, 20%, and 50%.
Additionally, a molecular database must be chosen for the search. Because of our datasets' biological molecular content, we used the following three databases in all our experiments: 1) CoreMetabolome v3, 2) HMDB v4, 3) LipidMaps 2017-12-12.
Since we were interested in estimating the chemical information that METASPACE could retrieve, we compared the number of unique formulas assigned using the three databases. In this way, duplicated assignments from different databases were counted as one.

Furthermore, we computed the number of formulas that were not shared by the original and recalibrated datasets. This allowed us to estimate the amount of information specific to each dataset.
The annotations test was performed without applying any peak matching or spatial filter to preserve the complete information contained in the datasets.
The mass accuracy was tested following a similar approach to (Boskamp et al., 2020). Since the datasets' actual molecular content is unknown, we tested the mass accuracy of a list of ions expected to be detected by DESI-MSI in biological tissue samples (Supplementary Table S1).
Subsequently, we first matched each MSI dataset's peaks using MALDIquant '*relaxed*' method, with a tolerance of 20 ppm (Gibb & Strimmer, 2012). Only the peaks detected in at least 50% of the ROI pixels were considered for the mass accuracy estimation. Then, we compared the mass error calculated from the closest peak to the test ions m/z value. We only kept the results if either the original or the recalibrated m/z differed less than ten ppm from the nominal m/z values of the test ions.
To perform the recalibration, the reference ions were searched around their nominal m/z value within windows varying in $W_{TOF} = \{2, 4, 6, \dots, 100\}\, ppm$ for TOF data and $W_{ORB} = \{1,2,3, \dots, 20\}\, ppm$ for Orbitrap data, respectively.
For each search window value, we calculated the coverage percentage (Eq. 8) and determined the optimal search window $\hat{w}$ from the elbow point, as described in Sec. 5.4.
All tested MSI datasets consisted of peaks with m/z values below 1000 m/z, except for the rat kidney dataset, which had only peaks below 400 m/z. This allowed us to use the provided reference ions (Table 1), as they cover the range between 100 and 1000 m/z. Different reference ions would have been required to cover a different range of m/z values.
The coverage percentage followed the expected trends showing a saturation at the elbow point (Figure 2a).
We always matched more than 5 reference ions for the Orbitrap data, with numbers varying from 6 to 9. In the case of TOF data, the number of matched reference ions varied from 6 to 11. The number of matched ions was sufficient to fit either the linear or polynomial calibration models in both cases.
In all datasets, the images corresponding to the matched peak intensity were consistent with the sample's morphological features (Figure 2c).
In each dataset, RANSAC successfully captured the relationship between the matched reference mass errors and the pixel order, being robust to the presence of strong outliers or close peaks (Figure 1-2b).

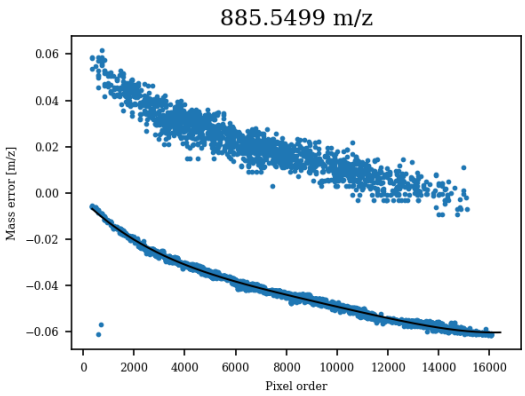

Figure 1 -Example of regression model described in Eq. 5. RANSAC successfully determined the main trend of mass errors of 885.5499 m/z as a smooth function of the pixel order, without being affected by the close false positives. (Dataset: TOF liver2(-)). Blue dots represent the mass errors, and the black line represents the RANSAC fit.

Ten-fold cross-validation determined the degree of the RANSAC models repeated five times. The selected degree, varying between one and five, corresponded to the smallest mean squared error between the predicted and observed independent variables.
Using the estimated mass errors from the RANSAC models, we fitted the calibration models (Eq. 3 and 4) using the nominal masses as independent variables. In TOF data, the optimal polynomial degree for the calibration model was estimated by leave-one-out cross-validation. The optimal degree corresponded to the smallest mean squared error between the independent variable's predicted and observed values.
The models were fitted using the functions available in the Scikit-learn library for Python v3.8 (Pedregosa et al., 2011).
In the case of TOF data, nonlinear fits were often observed among the optimal calibration models, but in all cases, the coefficients assigned to the terms with a degree higher than one were of the order smaller than $10^{-5} m/z$ (Figure S2). These results suggest that the linear terms captured the major correction for the mass errors.
The recalibration reduced the mass error for the reference ions, as shown in Figure 2d.
In all datasets, except for the TOF liver1(-), we observed an increase in the number of putative annotations with an FDR threshold of 5%. The relative increase, calculated as the ratio between the difference of annotations and the number of the original annotations, varied from 0.5% to about 4000%, with an FDR threshold of 5%.
For FDR thresholds greater than 10%, we observed that the number of annotations in the observed data was higher than the recalibrated data in some MSI datasets. Setting instead an FDR threshold of 10% gave similar results to 5% except for Orbitrap ovary1(+) and kidney(-), where the number of recalibrated annotations was smaller than the original dataset. Analogous patterns were observed for the number of dataset-specific annotations (Supplementary Figure S3-S4).
The test ions' mass accuracy generally improved in all datasets, except for TOF liver1(-), where the majority of test m/z had a worse mass accuracy after the recalibration, with the largest increase of mass error equal to 1.25 ppm. We observed the best performance of the TOF data's recalibration method, where we observed a reduction of mass error of about 70 ppm (Figure 2e, Supplementary Figure S5, Sec. S2).
One notable effect of the recalibration was the improved quality of the images corresponding to the assigned annotations by METASCAPE. The reduction of non-linearity in the mass distributions across the spectra resulted in the correction of *band-type* images (Figure 3).

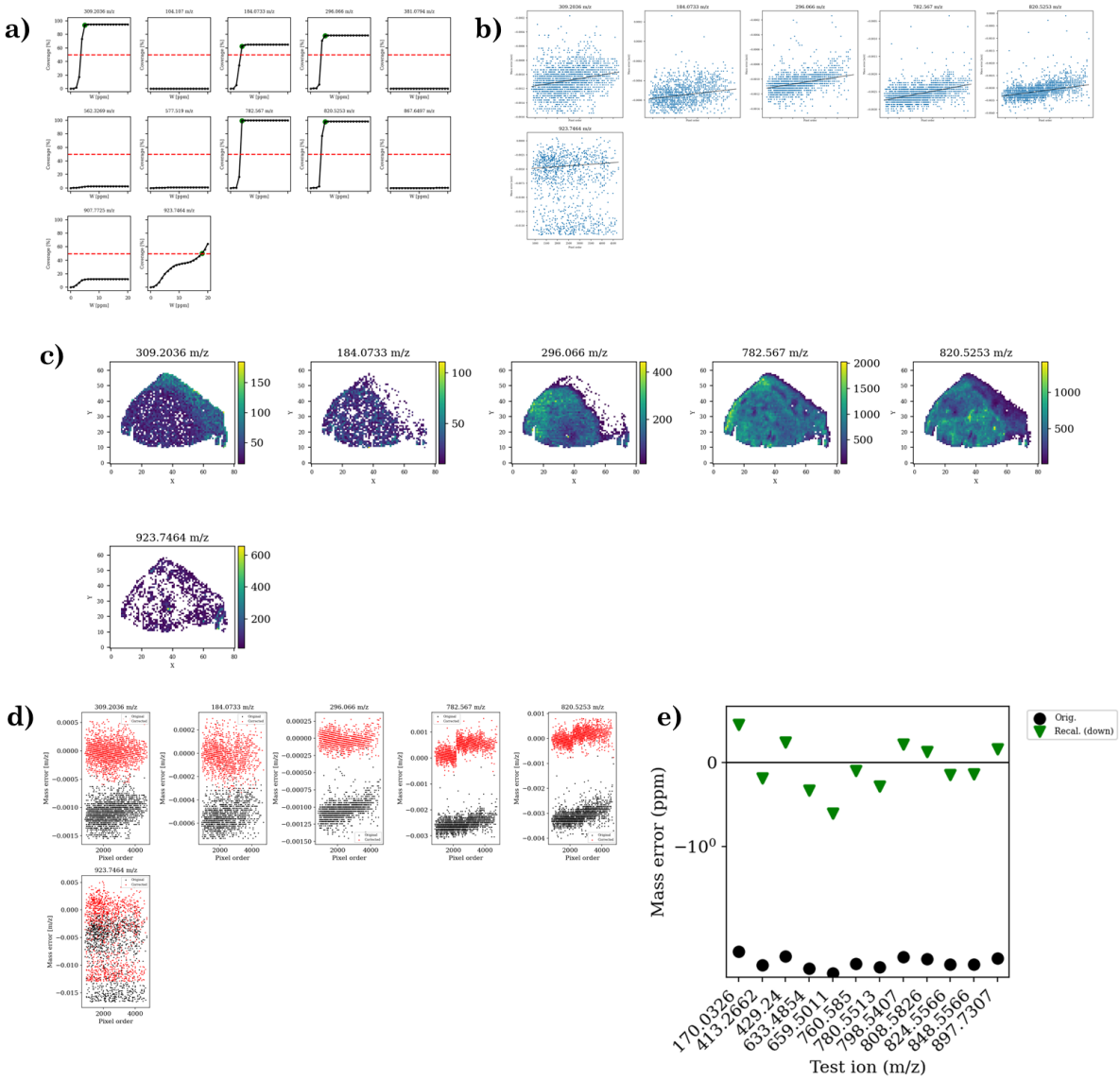

Figure 2 - Results for the Orbitrap ovary2(+) dataset. Six reference ions are detected within the tested search windows and used to estimate the time-dependent mass drift models. The optimal search window corresponds to the elbow point of the coverage percentage curves (a). RANSAC models are not affected by false positives, like for 923.7464 m/z (b). As a qualitative test, the matched peaks' intensity images reveal a spatial distribution consistent with the sample's morphological features (c). The mass errors of the recalibrated reference ions m/z values all show a closer distribution to zero (d). This result is confirmed by the mass errors of the test ions on the matched peaks, with a reduction of about 70 ppm (e).

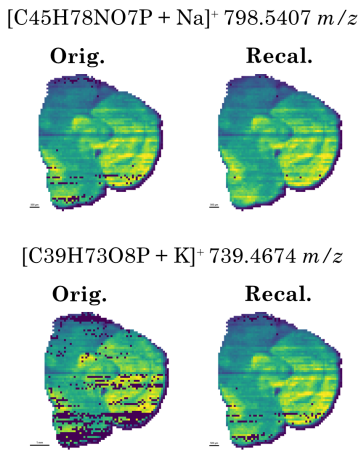

Figure 3 – Example of the effect of recalibration on the quality of the assigned metabolites' spatial distribution by METASPACE for the TOF brain1(+) dataset. The original dataset shows bands of missing intensities caused by the masses' nonlinear distribution across the pixels, significantly reduced in the recalibrated data. The assigned molecules were not part of the reference ions set.

7 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 fixed set of reference ions for the entire MSI dataset allowed us to test the assignment's validity by visualizing the corresponding spatial distributions, together with the estimated optimal coverage percentage. Furthermore, the observation that the mass errors follow the pixel order's smooth function, with no significant discontinuities, represents an additional level of evidence about the assignments' correctness.

RANSAC polynomial models 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 without risking to overfit the random mass errors that do not correlate with the acquisition order. The presented approach can improve the assignment using state-of-the-art molecular annotation methods, such as METASPACE, and their quality in terms of assigned spatial distributions.

Mass accuracy evaluation on a set of test ions after performing peak matching can reveal problems in the recalibrated data. It should always be performed after applying such types of computational models.

The models employed for recalibrating the masses were specific to the MS analyzer. This is an essential aspect because of the unique characteristics of how the signal is generated in each type of analyzer.

It is also important to underline that the recalibration models must also be simple. Thus, they can miss the complexity of the phenomena driving the mass drifts. This is a necessary trade-off to avoid falling into overfitting the noise naturally occurring in this type of data. 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.

The presented work has limitations. A list of reference ions must be known before applying the method. This can be challenging for completely unknown samples. However, building a database of accurately identified molecules in previous studies can generate a likely set of reference ions. In this regard, contextual information is of critical importance. Another limitation of the approach is that it can only be applied to the sample-related ROI pixels. 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 (Inglese, Correia, Takats, Nicholson, & Glen, 2019), 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 by including automatic reference ion selection based on the sample properties and will study the possible integration with the usage of external lock mass reference ions.

8 Acknowledgements

We thank Dr. Nicole Strittmatter (AstraZeneca) and Dr. James McKenzie for providing the public Orbitrap datasets on METASPACE.
PI is supported by the NIHR Imperial Biomedical Research Centre (BRC).

9 References

Boskamp, T., Lachmund, D., Casadonte, R., Hauberg-Lotte, L., Kobarg, J. H., Kriegsmann, J., & Maass, P. (2020). Using the chemical noise background in MALDI mass spectrometry imaging for mass alignment and calibration. *Analytical Chemistry*, *92*(1), 1301–1308. https://doi.org/10.1021/acs.analchem.9b04473

Christian, N. P., Arnold, R. J., & Really, J. P. (2000). Improved calibration of time-of-flight mass spectra by simplex optimization of electrostatic ion calculations. *Analytical Chemistry*, *72*(14), 3327–3337. https://doi.org/10.1021/ac991500h

Dória, M. L., McKenzie, J. S., Mroz, A., Phelps, D. L., Speller, A., Rosini, F., … Takats, Z. (2016). Epithelial ovarian carcinoma diagnosis by desorption electrospray ionization mass spectrometry imaging. *Scientific Reports*, *6*(November), 1–11. https://doi.org/10.1038/srep39219

Drexler, D. M., Reily, M. D., & Shipkova, P. A. (2011). Advances in mass spectrometry applied to pharmaceutical metabolomics. *Analytical and Bioanalytical Chemistry*. https://doi.org/10.1007/s00216-010-4370-8

Du, P., Kibbe, W. A., & Lin, S. M. (2006). Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics*, *22*(17), 2059–2065. https://doi.org/10.1093/bioinformatics/btl355

Ellis, S. R., Bruinen, A. L., & Heeren, R. M. A. (2014). A critical evaluation of the current state-of-the-art in quantitative imaging mass spectrometry. *Analytical and Bioanalytical Chemistry*, *406*(5), 1275–1289. https://doi.org/10.1007/s00216-013-7478-9

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, *24*(6), 381–395. https://doi.org/10.1145/358669.358692

Gibb, S., & Strimmer, K. (2012). Maldiquant: A versatile R package for the analysis of mass spectrometry data. *Bioinformatics*, *28*(17), 2270–2271. https://doi.org/10.1093/bioinformatics/bts447

Guenther, S., Muirhead, L. J., Speller, A. V. M., Golf, O., Strittmatter, N., Ramakrishnan, R., … Takats, Z. (2015). Spatially resolved metabolic phenotyping of breast cancer by desorption electrospray ionization mass spectrometry. *Cancer Research*, *75*(9), 1828–1837. https://doi.org/10.1158/0008-5472.CAN-14-2258

Hamm, G., Bonnel, D., Legouffe, R., Pamelard, F., Delbos, J. M., Bouzom, F., & Stauber, J. (2012). Quantitative mass spectrometry imaging of propranolol and olanzapine using tissue extinction calculation as normalization factor.

Journal of Proteomics, *75*(16), 4952–4961. https://doi.org/10.1016/j.jprot.2012.07.035

Inglese, P., Correia, G., Takats, Z., Nicholson, J. K., & Glen, R. C. (2019). SPUTNIK: An R package for filtering of spatially related peaks in mass spectrometry imaging data. *Bioinformatics*, *35*(1), 178–180. https://doi.org/10.1093/bioinformatics/bty622

Ipsen, A., & Ebbels, T. M. D. (2012). Prospects for a statistical theory of LC/TOFMS data. *Journal of the American Society for Mass Spectrometry*, *23*(5), 779–791. https://doi.org/10.1007/s13361-012-0340-z

Karlsson, O., & Hanrieder, J. (2017). Imaging mass spectrometry in drug development and toxicology. *Archives of Toxicology*. https://doi.org/10.1007/s00204-016-1905-6

Koranne, S. (2011). Hierarchical Data Format 5 : HDF5. In *Handbook of Open Source Tools* (pp. 191–200). https://doi.org/10.1007/978-1-4419-7719-9_10

La Rocca, R., Kune, C., Tiquet, M., Stuart, L., Eppe, G., Alexandrov, T., … Quinton, L. (2021). Adaptive Pixel Mass Recalibration for Mass Spectrometry Imaging Based on Locally Endogenous Biological Signals. *Analytical Chemistry*, *93*(8), 4066–4074. https://doi.org/10.1021/acs.analchem.0c05071

LaBonia, G. J., Lockwood, S. Y., Heller, A. A., Spence, D. M., & Hummon, A. B. (2016). Drug penetration and metabolism in 3D cell cultures treated in a 3D printed fluidic device: assessment of irinotecan via MALDI imaging mass spectrometry. *Proteomics*, *16*(11–12), 1814–1821. https://doi.org/10.1002/pmic.201500524

Li, Y., & Li, L. (2019). Mass Accuracy Check Using Common Background Peaks for Improving Metabolome Data Quality in Chemical Isotope Labeling LC-MS. *Journal of the American Society for Mass Spectrometry*, *30*(9), 1733–1741. https://doi.org/10.1007/s13361-019-02248-w

Mascini, N. E., Teunissen, J., Noorlag, R., Willems, S. M., & Heeren, R. M. A. (2018). Tumor classification with MALDI-MSI data of tissue microarrays: A case study. *Methods*, *151*, 21–27. https://doi.org/10.1016/j.ymeth.2018.04.004

Nilsson, A., Goodwin, R. J. A., Shariatgorji, M., Vallianatou, T., Webborn, P. J. H., & Andrén, P. E. (2015). Mass spectrometry imaging in drug development. *Analytical Chemistry*, *87*(3), 1437–1455. https://doi.org/10.1021/ac504734s

Olsen, J. V., de Godoy, L. M. F., Li, G., Macek, B., Mortensen, P., Pesch, R., … Mann, M. (2005). Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Molecular and Cellular Proteomics*, *4*(12), 2010–2021. https://doi.org/10.1074/mcp.T500030-MCP200

Palmer, A., Phapale, P., Chernyavsky, I., Lavigne, R., Fay, D., Tarasov, A., … Alexandrov, T. (2016). FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry. *Nature Methods*, *14*(1), 57–60. https://doi.org/10.1038/nmeth.4072

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Prideaux, B., & Stoeckli, M. (2012). Mass spectrometry imaging for drug distribution studies. *Journal of Proteomics*.

https://doi.org/10.1016/j.jprot.2012.07.028
Ràfols, P., Castillo, E. del, Yanes, O., Brezmes, J., & Correig, X. (2018). Novel automated workflow for spectral alignment and mass calibration in MS imaging using a sputtered Ag nanolayer. *Analytica Chimica Acta*, *1022*, 61–69. https://doi.org/10.1016/j.aca.2018.03.031
Satopää, V., Albrecht, J., Irwin, D., & Raghavan, B. (2011). Finding a "kneedle" in a haystack: Detecting knee points in system behavior. In *Proceedings - International Conference on Distributed Computing Systems* (pp. 166–171). https://doi.org/10.1109/ICDCSW.2011.20
Schwamborn, K., Krieg, R. C., Reska, M., Jakse, G., Knuechel, R., & Wellmann, A. (2007). Identifying prostate carcinoma by MALDI-imaging. *International Journal of Molecular Medicine*, *20*(2), 155–159. https://doi.org/10.3892/ijmm.20.2.155
Swales, J. G., Dexter, A., Hamm, G., Nilsson, A., Strittmatter, N., Michopoulos, F., … Goodwin, R. J. A. (2018). Quantitation of Endogenous Metabolites in Mouse Tumors Using Mass-Spectrometry Imaging. *Analytical Chemistry*, *90*(10), 6051–6058. https://doi.org/10.1021/acs.analchem.7b05239
Takáts, Z., Wiseman, J. M., Gologan, B., & Cooks, R. G. (2004). Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science*, *306*(5695), 471–473. https://doi.org/10.1126/science.1104404
Vera, C. C., Zubarev, R., Ehring, H., Hakansson, P., & Sunqvist, B. U. R. (1996). A three-point calibration procedure for matrix-assisted laser desorption/ionization mass spectrometry utilizing multiply charged ions and their mean initial velocities. *Rapid Communications in Mass Spectrometry*, *10*(12), 1429–1432. https://doi.org/10.1002/(SICI)1097-0231(199609)10:12<1429::AID-RCM600>3.0.CO;2-W
Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., … Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2
Wenger, C. D., McAlister, G. C., Xia, Q., & Coon, J. J. (2010). Sub-part-per-million precursor and product mass accuracy for high-throughput proteomics on an electron transfer dissociation-enabled orbitrap mass spectrometer. *Molecular and Cellular Proteomics*, *9*(5), 754–763. https://doi.org/10.1074/mcp.M900541-MCP200
Zhurov, K. O., Kozhinov, A. N., Fornelli, L., & Tsybin, Y. O. (2014). Distinguishing Analyte from Noise Components in Mass Spectra of Complex Samples: Where to Cut the Noise? *Analytical Chemistry*, *86*(7), 3308–3316. https://doi.org/10.1021/ac403278t