

NECTAR: A New Algorithm for Characterizing and Correcting Noise in QToF-Mass Spectrometry Imaging Data

Ariadna González-Fernández,* Alex Dexter, Chelsea J. Nikula, and Josephine Bunch*



Cite This: *J. Am. Soc. Mass Spectrom.* 2023, 34, 2443–2453



Read Online

ACCESS |



Metrics & More

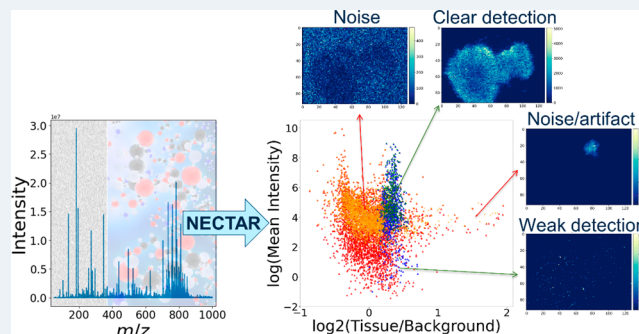


Article Recommendations



Supporting Information

ABSTRACT: A typical mass spectrometry imaging experiment yields a very high number of detected peaks, many of which are noise and thus unwanted. To select only peaks of interest, data preprocessing tasks are applied to raw data. A statistical study to characterize three types of noise in MSI QToF data (random, chemical, and background noise) is presented through NECTAR, a new Noise Correction Algorithm. Random noise is confirmed to be dominant at lower m/z values (~ 50 – 400 Da) while systematic chemical noise dominates at higher m/z values (>400 Da). A statistical approach is presented to demonstrate that chemical noise can be corrected to reduce its presence by a factor of ~ 3 . Reducing this effect helps to determine a more reliable baseline in the spectrum and therefore a more reliable noise level. Peaks are classified according to their spatial S/N on the single ion images, and background noise is thus removed from the list of peaks of interest. This new algorithm was applied to MALDI and DESI QToF data generated from the analysis of a mouse pancreatic tissue section to demonstrate its applicability and ability to filter out these types of noise in a relevant data set. PCA and t-SNE multivariate analysis reviews of the top 4000 peaks and the final 744 and 299 denoised peak list for MALDI and DESI, respectively, suggests an effective removal of uninformative peaks and proper selection of relevant peaks.



1. INTRODUCTION

One of the main challenges in the preprocessing of mass spectrometry (MS) data is robust separation of noise from signal. Noise can decrease the mass accuracy of MS peaks due to centroid shifting and lead to incorrect identification of detected ions. The term “noise” is not clearly defined in the MS field, and different authors refer to several types of noise with slightly varied approaches. The definition of noise given by the IUPAC¹ is the random fluctuations occurring in a signal that are inherent in the combination of instrument and method’.

A number of mass spectrometry articles refer to “nonrandom noise” (i.e., sinusoidal noise,² low-frequency noise,³ chemical noise^{4,5} or background noise^{6,7}). This type of “nonrandom noise” appears to contradict the definition provided by IUPAC but is a widely accepted concept in the MS community. However, the definition and origin of these noise contributions are not always consistent or well-known in the literature.

Several methods have been introduced for filtering noise in MS, which can be grouped into two main types: noise threshold-based algorithms^{8–10} and wavelet techniques.^{11–13} Some examples of noise threshold algorithms are (1) Autopicker,⁸ which assumes that real peaks should display regular spacing while noise will not. This method uses the isotopic structure to optimize the noise threshold along the m/z regime. (2) LIMPIC (linear MALDI-TOF-MS peak

indication and classification)⁹ was developed for the detection of consistent protein peaks from a set of calibrated mass spectra. For each, mass-spectral smoothing and baseline correction are applied, followed by noise estimation and peak picking. From the combination of the multiple spectra analysis, LIMPIC provides a list of “true” molecular signal peaks. The noise threshold in this case is determined using a smoothing procedure based on a Kaiser filter.

Other approaches use wavelet based algorithms, for example: (1) MassSpecWavelet¹⁴ is based on the continuous wavelet transform (CWT) algorithm. The shape and distribution of the peaks are considered when fitting the model. In this case, the local noise around one peak is defined as the 95-percentage quantile of the absolute CWT coefficient values within a local window surrounding the peak. (2) The undecimated discrete wavelet transform (UDWT) method consists of three steps: (i) compute the wavelet coefficients from the data, (ii) set small wavelet coefficients to zero, and

Received: March 23, 2023

Revised: September 8, 2023

Accepted: September 28, 2023

Published: October 11, 2023



Table 1. Collection of Some of the Methods for Noise Determination and Peak Picking Present in the Literature

Software	Technique	Availability	Ref	Summary
Autopiquer	MALDI-ToF	No	8	Threshold estimation by selecting as much (isotopic) structure as possible
	LC-Q-ToF			
	FT-ICR			
LIMPIC	MALDI-ToF	Matlab	9	Noise threshold by smoothing technique
MassSpecWavelet	SELDI-ToF	R - Open source	14	Continuous wavelets transform
UDWT	MALDI-ToF	Matlab	15	Undecimated discrete wavelet transform
N-sigma data dependent thresholding for FTMS	Orbitrap-FTMS	No	10	Noise threshold that involves analysis of the distribution of logarithmic intensity of all peaks, including noise and analyte
	FT-ICR MS			
Subspectral analysis	SELDI	R - Available under request	16	Continuous wavelets transform to detect peaks in individual subspectra, retaining information to construct noise distribution
Bioconductor PROcess	MALDI-ToF	R - Open source	15	Undecimated discrete wavelet transform-based peak detection on the mean spectrum
Hilbert–Huang Transformation	MALDI-ToF	No	17	Decomposition of the spectrum in very high and very low frequencies signal. The protein frequency domain should be in the middle part of frequency domain
	SELDI-ToF			

(iii) compute the inverse wavelet transform to recover the denoised spectrum. When using wavelet methods, the user needs to select a basic wavelet function on which to base the transform, the kind of transform, and the thresholding method.

Some of these techniques are summarized in Table 1, where the main principle of the method alongside information about coding language and availability are shown.

Mass spectrometry imaging enables the mapping of thousands of molecules in tissue. Uninformative peaks in MSI are often those that represent ions relating to background. Two types of background noise can be distinguished: (i) Sample background, which may be defined as peaks only detected in nonsample regions and/or those clearly showing a lack of spatial organization or structure. (ii) Spectral interference, referring to those peaks that are not resolved with the instrument used. These unresolved peaks, if only present off-tissue, are noise, while if in-tissue they are potentially peaks of interest. The distribution on tissue and background of the detected ions can be used to make a judgment on signal vs noise.

Identification of uninformative peaks in MSI is normally performed by visual examination of each of the individual single ion images. This manual analysis is highly time-consuming (typical MALDI experiments consist of thousands of peaks) and dependent on the experience of the user. Some methods to automatically remove these peaks can be found in the literature. Alexandrov and Bartels¹⁸ developed a method based on spatial structure analysis on single-ion-images. The level of spatial structure is ranked based on the original measure of spatial chaos (lack of spatial pattern in the pixels intensities). Inglese et al.¹⁹ created a reference image and studied the similarity between peak intensity images and the reference image. The similarity is determined using Pearson's correlation, Spearman's correlation, structural similarity index measure,²⁰ and normalized mutual information. Inverse maximum signal factors (MSF) denoising is presented in Tyler et al.²¹ to remove correlated noise as well as to improve contrast in MSI images with low signal-to-noise detections.

Removal of peaks relating to the use of a MALDI matrix and associated solvent, or sampling solvents used in DESI, can be

eliminated when cross-matching with databases of known compounds (i.e., HMDB,²² lipid maps,²³ etc.) such as is performed in METASPACE²⁴.

Noise is a particular problem in MSI workflows because further data analysis will typically consist of multivariate analysis,^{25–27} where the number and the lists of peaks have an important effect in the sample classification. The implication of peak selection for subsequent multivariate analysis cannot be neglected.²⁸

In this paper we present NECTAR (NoiseE CorreCtion AlgoRithm), a new method to characterize and remove noise in MSI-QToF instruments (a python package is available at https://github.com/NiCE-MSI/nectar_msi for transparency purposes⁴). Although several algorithms are available in the literature to remove noise, none of these methods analyze spectral and spatial noise together. This novel algorithm statistically determines and characterizes noise considering three factors: (i) a random noise level threshold, (ii) a chemical noise wavelet correction, and (iii) a spatial signal-to-noise ratio (S/N) analysis of tissue and background. The definitions given to the different types of noise in this work are instrumental, random, chemical and background noise (Figure 1). Instrumental noise does not seem to affect mass spectra significantly, affecting mostly the true shape of peaks.³ Smoothing algorithms are normally used to correct this effect (e.g., Savitzky–Golay, Gaussian smoothing, moving average²⁹). We do not investigate instrumental noise as it does not affect the detection limit of the spectrum in MS. We refer to “random noise” as the intrinsic random noise which follows the Poisson distribution due to the “ion counting” detectors of MS instruments, to “chemical noise” as the systematic noise present at higher *m/z* range values that is sinusoidal and mostly constant in amplitude, and “background noise” to those uninformative peaks associated with matrix, sampling solvents, or other contaminants which are not found to provide sample related spatial structure.

2. METHODS

2.1. Mass Spectrometry Imaging. Two 10 μm thick mouse pancreatic tissue sections were analyzed in two different

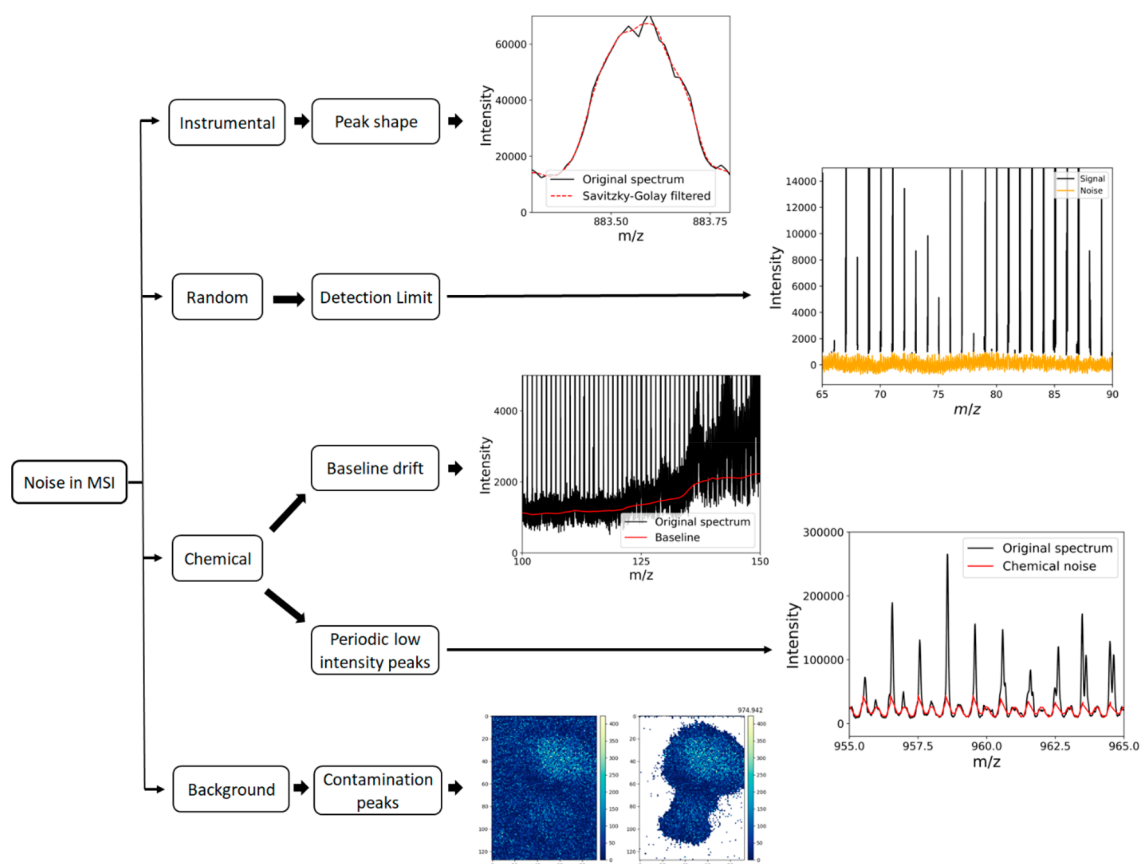


Figure 1. Diagram of the different noise contributions found in the literature affecting the MSI data. Instrumental noise mostly affects the shape of the peaks, while random and chemical noise are direct noise contributions affecting the detection limit of the spectrum. Background noise refers to uninformative peaks, most likely related to the presence of matrix or solvent.

modalities. First, tissue was acquired in positive polarity using a uMALDI Synapt G2-Si (Waters, UK) with a mass range of 50–1000 m/z in sensitivity mode. A scan time of 0.05 s/pixel and a pixel size of 50 $\mu\text{m} \times 50 \mu\text{m}$ were used, and 2,5-dihydroxybenzoic acid was used as a matrix. Second, tissue was acquired in negative polarity using a DESI XEVO G2-XS (Waters, UK) with a mass range of 80–1000 m/z in sensitivity mode. A scan time of 0.5 s/pixel and a pixel size of 100 $\mu\text{m} \times 100 \mu\text{m}$ were used, and 95:5 methanol/water was used as solvent.

2.2. Data Conversion and Handling. Proprietary raw data files were converted to mzML using msconvert from ProteoWizard³⁰ and converted to the imzML format using imzMLConverter.³¹ The imzML was read in Python with pyimzML package (<https://github.com/alexandrovteam/pyimzML>). To calculate the mean spectrum, first we created a common x -axis for all pixels. This axis was created from the m/z values of the individual pixels, assuming the m/z sampling is proportional to $(m/z)^{0.5}$. Using this scale, every pixel-spectrum was binned to this common x -axis. SpectralAnalysis software³² was used to apply K-means clustering algorithm to separate tissue from background, so the resulting spectrum analyzed included peaks detected from tissue regions only. Multivariate analysis was performed with Matlab2018b (The Mathworks Inc., Natick, MA, USA).

2.3. Determination of Noise. Three types of noise are studied in this work, i.e., random, chemical, and background noise. Examination of random and chemical noise is performed on discrete mass spectra. Our review of background noise

involved examination of a spatially resolved MS imaging data set and the distribution of selected ions of interest.

2.3.1. Random Noise Determination: Sigma-Clipping Function. The “sigma-clipping function” has been widely developed and used in astronomy to define the continuum of a spectrum, to remove bad pixels from astronomical images (i.e., cosmic rays, hot pixels), as well as to separate signal from noise.^{33–35} In astronomy, a CCD detector is commonly used to count photons arriving at the telescope. The random noise produced in this “photon counting” has a Poisson distribution, similar to the noise found in the ion-counting detectors of the MS-ToF instruments.^{36,37}

The main concept behind the sigma-clipping is the removal of outliers from a distribution. Outliers present in the data can bias the noise level. Signal peaks count as outliers for the noise determination; therefore, these outliers are removed, which reduce the noise and allow new outliers to be identified. This process was repeated until a certain tolerance is reached. The main steps of this function are as follow: (1) The standard deviation (σ) and the median m of the distribution are calculated. The median is used instead of the mean because the mean is much more affected by outliers than the median. (2) Every point that is smaller or larger than $m \pm \alpha\sigma$ is removed, where σ is the “sigma threshold” and α a scaling parameter. Values beyond this threshold are considered outliers and therefore rejected. Higher α implies less data being removed and vice versa. (3) We go back to step 1 and iterate the same steps until a certain tolerance value (θ) is achieved,

$$\theta = \frac{\sigma_{\text{old}} - \sigma_{\text{new}}}{\sigma_{\text{new}}} \quad (1)$$

where σ_{old} is the standard deviation before removing the new outliers and σ_{new} the standard deviation after.

In this work, we use this function to define the noise level from the mean spectrum of the MS data. Noise in MS is not constant along the m/z range; hence, the sigma-clipping function is applied within a sliding window which determines the noise level locally. Two different window sizes are evaluated (4 and 20 Da) to ensure that enough values are used to robustly identify outliers while still revealing local variation in the noise level on these mass-charge scales. This process is applied in every single mass-charge bin; therefore, no gaps between windows are present.

The tolerance value was set at 0.001. This figure was chosen after testing several threshold values on different data sets and finding this value the most appropriate when inspecting the random noise level in each data set.

As well as estimating the noise level using the sigma-clipping function, it is possible to correct the baseline drift sometimes present in MS data. The baseline of a spectrum should be flat at zero; however, MS spectra normally present an uneven baseline often observed as a monotonically decreasing bias. This baseline affects the peak intensity values of a mass spectrum, so it is essential to correct for this effect to obtain accurate information (Figure 1, baseline drift). To apply the baseline correction with sigma-clipping, the median intensity of peaks determined to be noise is calculated and subtracted from the original signal. The baseline is calculated within the same sliding window that moves along the m/z range. To apply the baseline correction, the threshold is set at 2σ . A lower threshold was chosen here to avoid the bias due to the presence of chemical noise in the sample (more bins were identified as signals and thus not included in the noise). This high intensity systematic noise can bias the baseline level, resulting in a higher number of negative values when performing a standard 3σ baseline subtraction. After correcting the baseline and separating the signal from the noise, the random noise is evenly distributed around zero (Figure 1, random noise).

2.3.2. Chemical Noise Correction. Chemical noise is characterized by a sinusoidal signal of almost constant amplitude and local maxima with a periodicity of ~ 1 Da.^{38,39} The amplitude of this signal is smaller than that of the true peaks in a local region. This type of sinusoidal noise is characteristic of MALDI and DESI QTOF instruments (Figure 1, periodic low intensity peaks). Its origin and nature are not clearly understood, but it is generally accepted that this type of noise is due to cluster ions arising from the matrix or solvent.⁴⁰ This nonrandom noise contributes to the systematic noise of the mean spectrum, and therefore, the identification of a baseline noise cannot be obtained easily, which makes it difficult to separate noise from signal.

To characterize and correct chemical noise we use a combination of the sigma-clipping function and the “Adaptive Background Subtraction” (ABS) algorithm.³⁹ The local noise level is determined and masked via sigma-clipping, and a new “noise spectrum” is used to estimate the chemical noise pattern with ABS. This intensity distribution pattern is then subtracted from the original signal, and thus, the chemical noise corrected.

The main steps of ABS are as follows: (1) A sliding window of 21 Da mass units is selected. (2) This window is divided in

units of 1 Da each and subdivided into 10 bins. (3) The 21 bins of 1 Da each are overlaid, so that for each 0.1 Da sub-bin, there are now 21 values available. (4) The local intensity distribution for each nominal mass region is calculated, where the chemical noise is defined as the 45th percentile of the 21 values in each sub-bin. (5) An interpolation is applied between the different types of sub-bins to obtain the correction at each mass-charge. (6) This intensity distribution pattern is then subtracted from the original signal (for more details see ref 39).

2.3.3. Peak Picking. To identify individual peaks, we make use of the first derivative properties. All of the identified peaks with a distance less than 30 ppm between them are considered one single peak. This value was selected due to the intrinsic resolution of the QTOF instruments that have been used. When multiple peaks fall within the 30 ppm window, the highest intensity peak is selected as a potential peak while the second one is considered an artifact or an unresolved peak due to insufficient resolution.

Once the chemical noise has been removed and the noise level has been determined, all peaks above 3σ of their local noise are considered potential peaks. However, it often happens that poorly resolved peaks are present in the spectrum.

From this first list of peaks, a subset with the 200 most intense peaks is created to determine the resolution of the peaks in the data set under study. A Gaussian model is fitted to each of these 200 peaks by minimizing the χ^2 , which is defined as

$$\chi^2 = \frac{(\text{obs} - \text{model})^2}{\text{noise}^2} \quad (2)$$

Three free parameters are considered in the fitting, i.e., the centroid, width, and intensity of the peak. The χ^2 minimization provides an optimized result, σ_{res} , for the fit. To obtain a better correlation, outliers are removed with the use of the sigma-clipping function, eliminating those peaks that are further than $\pm 3\sigma$ from the correlation ($\theta \sim 0.001$). The intercept of the correlation is forced to be located at zero. This resolving power is used as a reference to estimate the expected width of the peaks along the m/z range of the spectrum (Supporting Information Figure S1).

The process is then repeated, and the second Gaussian fitting is applied to the full list of peaks by minimizing χ^2 , but this time constraints are put on the fits. These constraints are applied to the centroid, intensity, and width of the peak. The centroid of the peak is constrained to be within $\pm 1 \sigma_{\text{res}}$ of the mass charge where the maximum intensity was found. The intensity is constrained to the intensity of the peak $\pm 0.5 \times \text{intensity}_{\text{peak}}$, and the width is constrained to be within $3\sigma_{\text{res}}$. If peaks are closer to each other than 15 sample bins, the two or three peaks are fitted together, as the contribution of close peaks could affect the shape and intensity of each other peaks (Supporting Information Figure S1.2).

2.3.4. Background Noise Removal. It is expected that some of the selected peaks are related to the presence of a matrix or solvent. These peaks are valid signals from noninformative features, i.e., background noise. NECTAR performs a S/N analysis on tissue and background and classifies these potentially noninteresting peaks and flags them as noise.

To identify peaks that are present only in the background but not in the tissue or signal that has no spatial structure, a

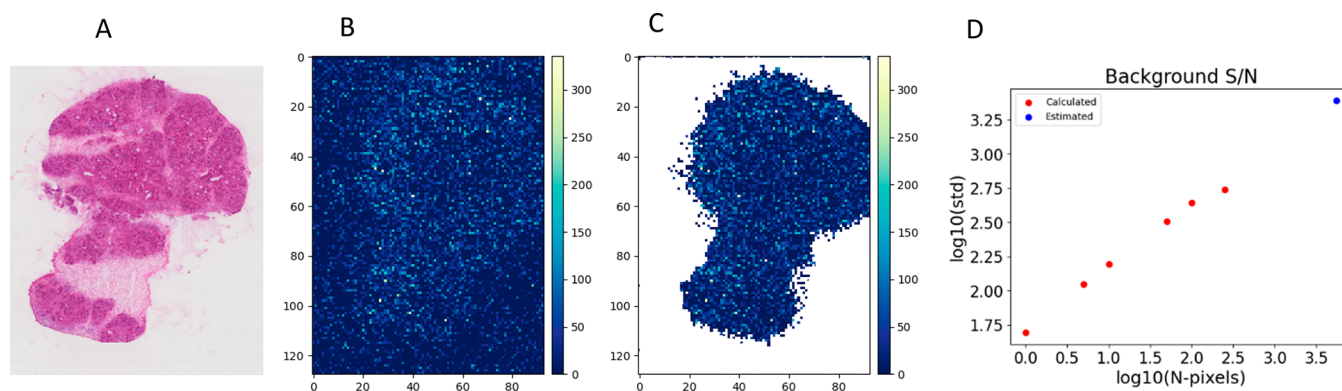


Figure 2. (A) H&E staining was performed post MALDI MSI analysis on same tissue section. The stained section was imaged at 40 $\times$ magnification (0.226 $\mu\text{m}/\text{pixel}$) with the Aperio CS2 digital pathology scanner (Leica Biosystems) and visualized ImageScope software (Leica Biosystems version 12.3.2.8013). (B) Example of a compound that is related to the matrix and thus not relevant for the study of the biological tissue sample. (C) If no background is considered when selecting the peak, it could be wrongly selected at first. (D) An estimation of the S/N on background for the same number of pixels in tissue is performed to obtain the S/N ratio between tissue and background.

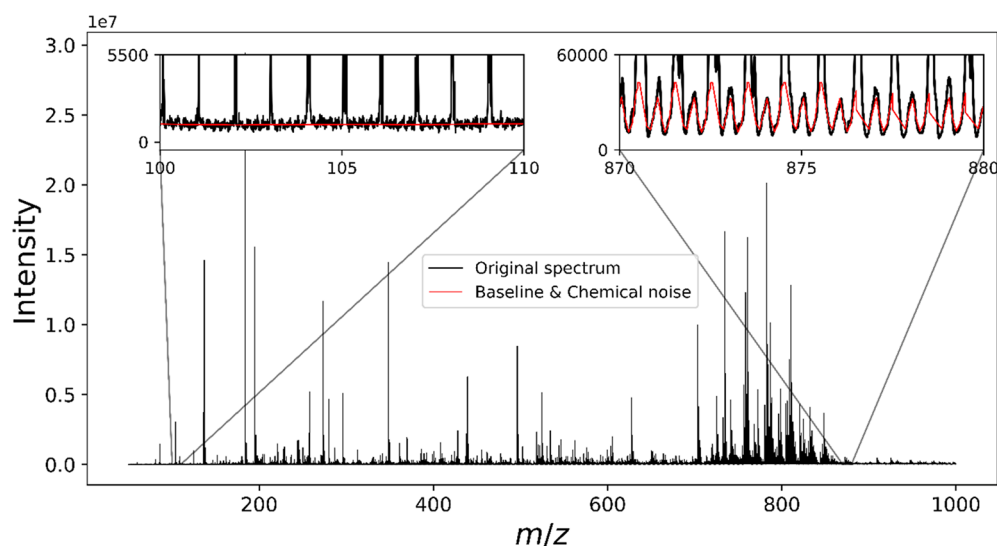


Figure 3. Characterization of the different types of noise in QTOF-MSI instruments. Noise level is determined (red), and baseline correction is applied to the spectrum. Random noise is identified, and chemical noise is modeled (red) and subtracted from the original spectrum.

comparison of the signal-to-noise on tissue (S/N_T) and signal-to-noise on background (S/N_B) is evaluated in each single ion image. First, masks for the tissue and background were created by using a k-means approach ($k = 2$). To obtain a statistically meaningful S/N ratio between tissue and background, the number of pixels in both masks needs to be the same. This is because the mean signal will remain mostly constant with increasing number of pixels, yet the noise in the mean spectrum will decrease with the root of the number of pixels. In our sample under study, the number of pixels for the background is smaller than for the tissue, so we estimate what the background noise would be for the same number of pixels that are present in the tissue. To do so, an estimation of the background noise to the number of pixels in tissue is performed by extrapolation. First, the sum intensity of 1, 5, 10, 50, 100, and 250 random pixels are measured on the background mask. Because single ion images of low intensity peaks might have many pixels whose intensities are zero, we take the standard deviation of the sum intensity of those random pixels by iterating these measurements 5000, 1000, 500, 100, 50, and 20 times respectively. Second, we obtain the linear relationship in logarithmic scale of the background signal

versus the number of pixels and extrapolate the background signal to the desired number of pixels, i.e., number of pixels in the tissue. An example of this is shown in Figure 2, where the red dots are the measured background signal for the different number of pixels and the blue dot represents the estimated background signal for the number of pixels in the tissue. The definitions of the S/N for tissue and background are defined in the Supporting Information, Section S2.

3. RESULTS AND DISCUSSION

Random and chemical noise are known to be dominant in different parts of the spectrum (random noise is dominant in lower m/z values, while chemical noise is dominant in high m/z values⁴¹), as well as presenting different patterns (Figure 3).

The characteristics of these two types of noise were studied in more detail by creating different mean spectra of randomly selected pixels with an increasing number of pixels. When averaging pixels, the random uncertainties will cancel out to some extent (which is why the random noise is inversely proportional to the root of the number of pixels, i.e., random noise $\propto \text{std} \sqrt{N}$, where std is the standard deviation

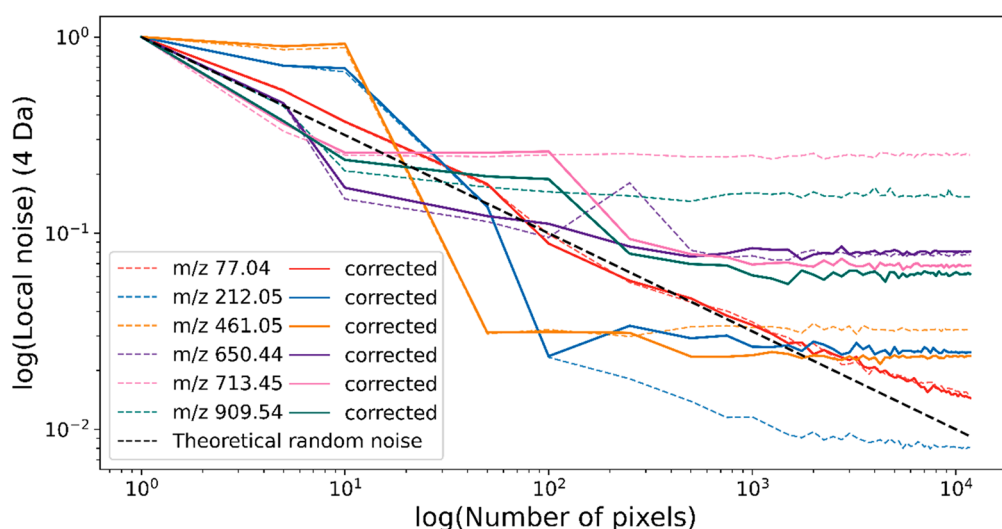


Figure 4. Noise contribution for 6 selected peaks in the MALDI spectrum. Random noise is dominant for peaks at lower m/z values, while systematic chemical noise is dominant for peaks at higher m/z values. Dashed lines correspond to data before correction, and solid lines correspond to data after correction (baseline and chemical correction).

or local noise around the peak of interest and N is the number of pixels used to create the mean spectrum), and the random noise will thus get relatively smaller with increasing number of pixels.

The chemical noise is always positive and does not cancel out, thus remaining constant with an increasing number of pixels. Therefore, it is possible to identify where and how the systematic chemical noise dominates the random noise. Six peaks along the m/z range were selected (m/z 77.04, m/z 212.05, m/z 461.05, m/z 650.44, m/z 713.45, m/z 909.54) and the local noise measured around these peaks of interest. To estimate the random noise in the spectrum, mean spectra with different numbers of pixels were created (1, 5, 10, 50, 100, 250, 500, 750, (+250)..., "total number of pixels"), and the local noise for the six peaks of interest was obtained for each of the spectra. Due to variabilities among pixels, to obtain a better statistical result for those mean spectra created from a number of pixels lower than 100, 50 different mean spectra were created, and the local noise averaged. For instance, for the five-pixel mean spectrum, 50 subsets of 5 random pixels on the sample were selected, their mean spectrum was created, and their local noise averaged, obtaining the final standard deviation value.

3.1. Random Noise. The local noise around the selected peaks under study is plotted on a logarithmic scale in Figure 4. For comparison purposes, the local noise of the different peaks is normalized in the plot by dividing the local noise by the corresponding averaged local noise of the mean spectrum for one pixel. The theoretical relationship, represented as a black dashed line, is linear (the square root becomes a linear slope of 0.5). The local noise around the peak detected at m/z 77.04 correlates almost perfectly to the estimated theoretical random noise, indicating that at this m/z value the random noise is dominating the spectrum. In addition, this indicates that the estimation of the local noise done via sigma-clipping performs well. For m/z 212.05 and m/z 461.05, although the correlation between theoretical and measured noise is not as precise as for the previous peak, random noise is still dominant for these two peaks. Thus, low m/z peaks within a local window can be selected with the certainty that the noise level is properly determined.

3.2. Chemical Noise. At higher m/z values, systematic noise is dominant as shown in Figure 4, (m/z 650.44, m/z 713.45, m/z 909.54). The standard deviation around the peaks of interest becomes constant after a certain number of pixels; this indicates a systematic noise present in the spectrum, most likely due to chemical noise. For very small numbers of pixels, there will be a contribution from both random and chemical noise, yet when averaging a higher number of pixels, the random uncertainties will minimize, and the uncertainties will be dominated by the chemical noise. Therefore, it is difficult to separate the noise from the signal with only the use of the sigma-clipping function.

If all of the chemical noise were successfully subtracted, the random noise should dominate the spectrum. For those peaks that were more affected by chemical noise (m/z 650.44, m/z 713.45, m/z 909.54), applying NECTAR on the MALDI data set decreased the chemical noise $\sim 63\%$, and the noise of the mean spectra (solid - lines) by a factor of ~ 3 . Therefore, the local noise was closer to the theoretical random noise estimation when higher numbers of pixels were averaged.

The benefit of using NECTAR over ABS is shown in Figure S3 in the Supporting Information. In Figure S3, left, the sinusoidal signal (dashed-red in the plot) has been removed from the original spectrum (blue line) and a baseline noise level is better identified (green line). However, the ABS sometimes does not perform ideally when multiple high intensity peaks are next to low intensity peaks (as shown in this same figure). When there are multiple peaks that are separated by 1 Da within the 21 Da window, this can affect the 45th percentile of the bin corresponding to these peaks. As a result, the modeled chemical noise might end up higher than the signal itself. When subtracting the chemical noise, these low intensity peaks become negative and therefore are removed from the signal, potentially missing peaks of interest (shown in the same figure). To overcome this effect, we created a novel algorithm that is a combination of sigma-clipping and ABS. Because all potential peaks above 3σ noise level have been already removed at this stage, they will not be considered when determining the 45th percentile, and the chemical noise model will not be affected by the intense peaks in the spectrum. Thus,

no peaks will be accidentally removed (Supporting Information Figure S3, right).

The relative contribution of the noise across the full m/z range is shown in Figure 5 for MALDI data set, where a clear

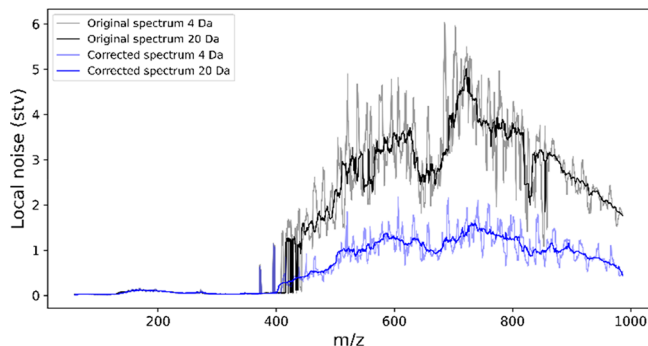


Figure 5. Relative noise level in the MALDI data spectrum obtained with two different window sizes (4 and 20 Da) before and after applying the noise correction.

division of the noise level is found (around 400 Da), which is a direct consequence of the noise being dominated by random or chemical noise. The local noise for each detected peak along the mean spectrum was measured with the sigma-clipping function before (black) and after (blue) the correction. There is a significant decrease in the local noise for m/z values for which the chemical noise is dominant, indicating an improvement on the detection limit of the ions in that range. Two different windows (4 and 20 Da) were used to study the potential effects of different window sizes when determining the local noise. While the trend of the noise is similar for both windows, a smaller window results in a more scattered distribution of the noise along the spectrum.

We next examined noise distribution histograms. After baseline correction and noise determination, the noise was rescaled to be randomly distributed around zero; thus, a Gaussian distribution around zero was expected. Several histogram plots are shown in Figure 6, where the noise for the MALDI data set is represented before baseline correction (green), after baseline correction (blue), and after chemical noise correction (red). Different areas of the spectrum are represented. Figure 6A shows the noise along the whole mean spectrum. In this case, most of the noise is located around zero

after the correction, while before the correction the distribution is skewed to positive values. Figure 6B shows the noise up to 350 Da. Due to the random noise being dominant in this part of the spectrum, a mostly Gaussian distribution is already obtained only after baseline correction, which is improved after the chemical noise correction. Figure 6C shows the noise distribution from 400 to 1000 Da, where the chemical noise is more dominant. In this case, the noise correction greatly improves the noise distribution in this area of the spectrum, although the histogram is a bit skewed to negative values, which might indicate that part of the signal could have been removed during subtraction of the chemical noise (Figure S3). Similar plots for the DESI data set are presented in Figure S4.

3.3. Background Noise. To evaluate the variability of S/N between tissue and background, the logarithmic mean intensity of the single ion image versus the fold change (FC) of the mean intensity of tissue and background was reviewed (Figure 7, center). In addition, the difference and ratio between tissue/background are calculated for each single ion image, which indicates the presence of background peaks when this difference is negative and the ratio is greater than 5. The threshold value for the ratio was selected after visual inspection of some of the single ion images in this work. Four categories of different peaks were identified (Table 2) according to their S/N values as well as fold change information (classification of peaks for the DESI data is shown in Table S1). Here we provide an example for each of the four categories.

Characteristics of the categories (Table 2 and Figure 7): (i) Robust detection: Peaks with very high S/N on tissue and lower S/N on background, whose ratios are on the order of tens (Figure 7D). (ii) Low intensity detections: Peaks with ratios between S/N_T and $S/N_B > 5$ and positive $S/N_T - S/N_B$. These peaks are the most difficult to classify as it is not easy to determine if the detection on tissue is statistically significant or not versus the background (Figure 7A). (iii) Artifact: Peaks with negative value for the difference between tissue and background but with positive $\text{Log}_2(\text{FC})$. Figure 7E is an example of a clear detection outside tissue that is most likely an artifact originated during the sample preparation. (iv) Background noise: Peaks with S/N higher on background than on tissue. These ion images present a negative difference and are easily identified as noninformative peaks (Figure 7B). Another example of background noise is shown in Figure 7C.

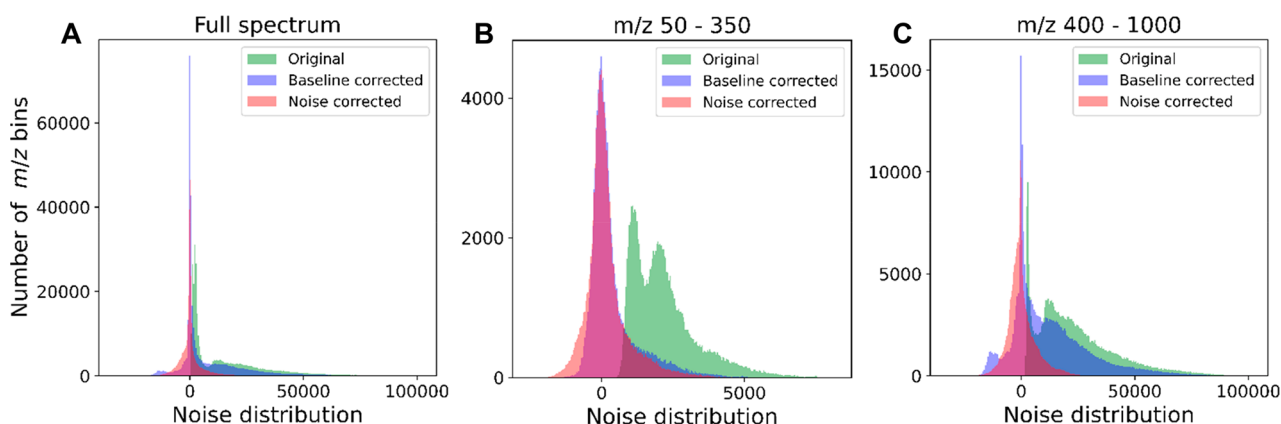


Figure 6. Histogram of the MALDI noise distribution before and after baseline and chemical noise correction at different areas of the spectrum: (A) full spectrum, (B) m/z 50–350, (C) m/z 400–1000. After final correction, a Gaussian distribution of noise is obtained (in red).

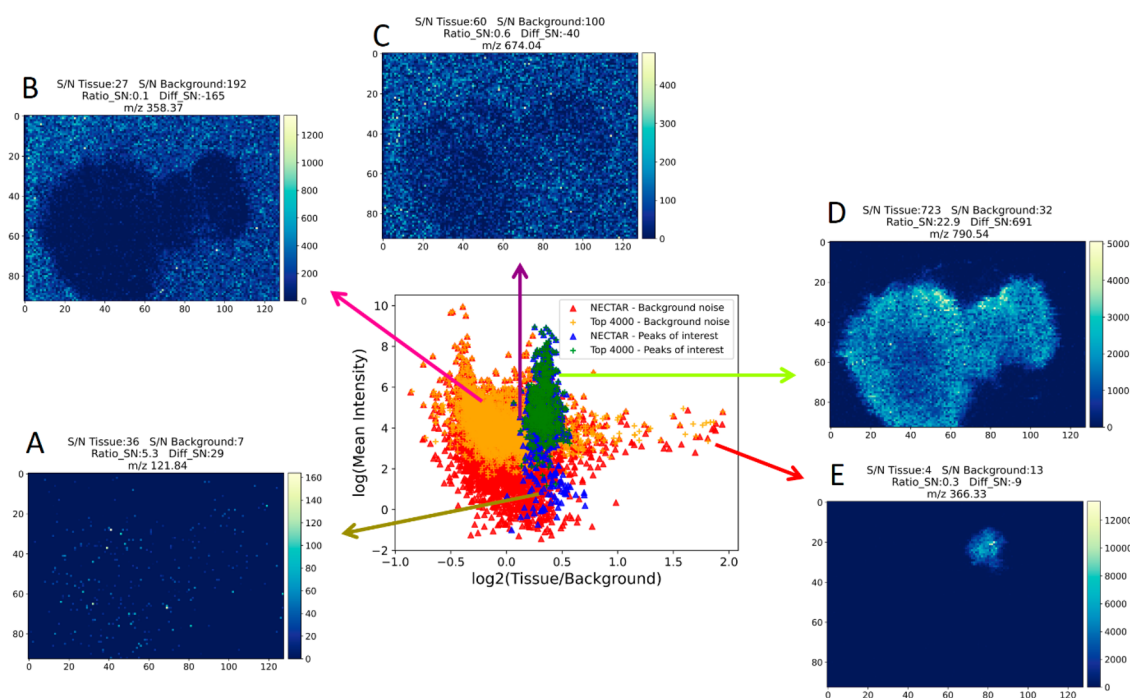


Figure 7. Classification of peaks according to their signal spatial distribution. The central figure represents the logarithmic mean intensity versus the fold change of the mean intensity of the tissue background on the single ion images. Triangles correspond to the NECTAR selected peaks, while crosses correspond to the top 4000 most intense peaks. Blue and orange symbols represent noise peaks ($S/N_T - S/N_B < 0$ and $S/N_T/S/N_B < 5$), while blue and green are peaks of interest ($S/N_T - S/N_B > 0$ and $S/N_T/S/N_B > 5$). Information related to the single ion images is given in Table 2. (A) Example of a detected very low intense peak. (B,C) Examples of background noise originated in the matrix. (D) Clear detection. (E) Artifact produced most likely during the sample preparation.

Table 2. Information and Classification of the Final List of Peaks Made by NECTAR

Measured m/z	Peak Intensity	Width	S/N_T	S/N_B	$S/N_T - S/N_B$	$S/N_T/S/N_B$	$\text{Log}_2(\text{FC})$	Classification
121.838 [A]	440	0.015	36	7	29	5.3	0.25	Low intensity peak
358.370 [B]	81,004	0.012	27	192	-165	0.1	-0.45	Background noise
674.038 [C]	41,829	0.019	60	100	-40	0.6	-0.13	Background noise
790.542 [D]	1,806,152	0.033	723	32	691	22.9	0.39	Highly abundant ion
366.326 [E]	73,178	0.013	4	13	-9	0.3	1.94	Artifact

In this scenario, the difference between the S/N for tissue and background is still negative, but the single ion image presents a uniform intensity across the image.

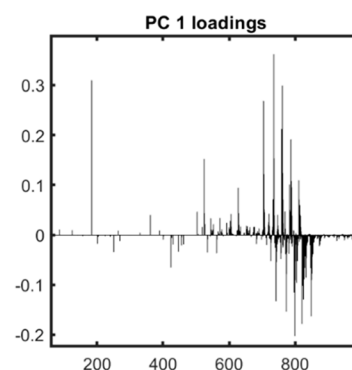
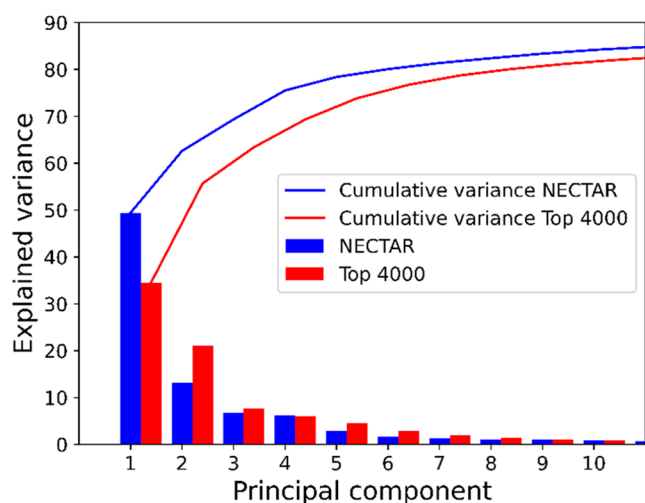
Comparison with Top 4000 Most Intense Peaks. Several methods select the list of peaks by defining a threshold value^{8,9} or, for instance, by selecting the top-N number of most intense peaks.^{28,42} This criterion might cause the loss of some potentially interesting peaks, while uninformative peaks might be selected as well (i.e., background noise or chemical noise peaks). A comparison of the selected peaks using NECTAR and the top 4000 most intense peaks in the mean spectrum was performed.

The central plot in Figure 7 represents the mean intensity of single ion images versus the fold change of the mean intensity of tissue/background for both lists of peaks used in the comparison. The main difference between both examples is the lower limit on the ordinate axis of the plot. NECTAR detects lower values (red and blue symbols) than the top 4000 (yellow and green symbols). Some peaks are detected in both lists, overlapped in this plot, which correspond to the very intense peaks in the spectrum. NECTAR also identifies more peaks relating to the artifact present in the sample (Figure 7E), although these peaks are removed by the S/N criteria. The

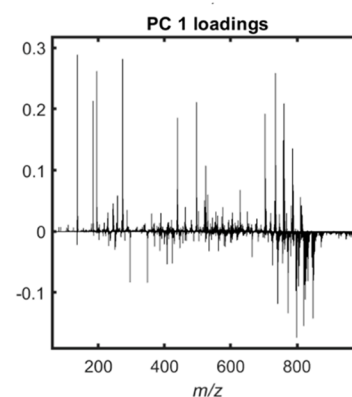
number of peaks found in the m/z low regime is much higher using NECTAR, while for the higher m/z regime more are identified from the top 4000 list. This is a consequence of the presence of chemical noise, which has periodic peaks and increases the average intensity of the spectrum, thus selecting chemical noise peaks among the top 4000 most intense peaks in the spectrum. Applying NECTAR to the original mean spectrum resulted in a list of 4215 peaks after noise determination and chemical noise correction. A final list of 740 peaks is obtained after eliminating uninformative peaks according to the spatial S/N criteria for the MALDI data set. NECTAR selects a final list of 299 peaks of interest when performed on the DESI data set (Figure S5).

Multivariate Analysis. To study the effects of this much shorter list in subsequent post processing tasks, we performed principal component analysis (PCA) using both peak lists. Two normalization methods (root-mean-square and total-ion-count) as well as no-normalization were tested, obtaining similar results in all cases. The explained variance as well as cumulative variance with PCs is shown for the total-ion-count normalization in Figure 8 (Figure S6 for DESI). For MALDI, an explained variance of 34.53% in PC1 for the top 4000 is obtained, while for NECTAR we obtain a 49.39% explained

PCA



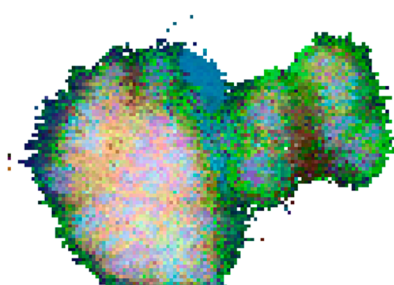
NECTAR (744 peaks)



Top 4000

t-SNE

NECTAR (744 peaks)



Top 4000

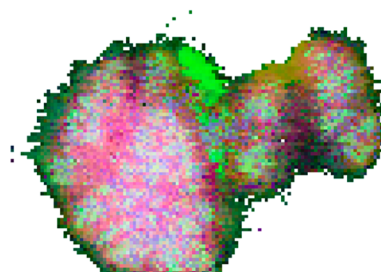


Figure 8. Multivariate analysis comparison of the top 4000 peaks and the NECTAR denoised list for the MALDI data set. A greater cumulative variance is obtained with the first 10 PCs for the denoised list than with the top 4000. T-SNE clustering demonstrates a proper removal of uninformative peaks and selection of relevant peaks.

variance. When calculating the cumulative variance, NECTAR results in a greater explained variability with the first 10 PCs than the top 4000, indicating a good performance in the removal of noise peaks. The first 4 PCs scores from both lists under study are compiled in Figure S7 (4 PCs for the DESI data set are shown in Figure S8). T-SNE was performed for both lists of peaks, resulting in similar segmentation results in both scenarios (note that the color scheme is different between figures, but the clustering results enable similar segmentation), which shows an effective removal of uninformative peaks and proper selection of relevant peaks, making the interpretation of biological features simpler for further analysis. The computational time required to process the multivariate analysis was shortened by ~70%.

4. Conclusions. The removal of noise in MSI can simplify the interpretation of complex biological samples as well as significantly reduce the computational time needed to process the data. A proper noise characterization results in a reliable

noise level determination and thus a more reliable peak picking selection. Obtaining a local noise threshold along the spectrum avoids the removal of peaks of interest due to the presence of different types of noise, and it is the first step toward quantitative analysis of the abundances of the compounds of interest. Selecting only informative peaks simplifies the interpretation of further multivariate analysis, i.e., PCA, t-SNE.

In the present study, the determination of the S/N threshold of tissue/background to classify background noise was determined after visual inspection of many single ion images. This value might not always be the same for different data sets, as well as is dependent on the experience of the user. The threshold for both MALDI and DESI data set was set at the same value for consistency ($S/N_T/S/N_B < 5$). Although after visual examination, this value could have been set to a lower figure for the DESI data set, obtaining a more complete list of low intensity peaks of interest. Subsequent studies will focus into an in-depth study of different modalities obtaining a

statistically robust method to classify these peaks without the intervention of the user.

The resolving power obtained for the specific data set during the peak picking step could be used to estimate an appropriate threshold to classify resolve/unresolved peaks, instead of using a generic ppm as it is currently developed.

This method has been developed for QToF MSI data and preliminary tested in Ion-trap-MS (Orbitrap) data. In order to extend the applicability of the proposed method, future work will focus on refining NECTAR for different instruments and spectrometers to facilitate comparison among multimodal MSI studies.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jasms.3c00116>.

Resolving power determination of the MALDI and DESI data sets, example of χ^2 Gaussian fitting for three overlapping peaks in the mean spectrum of MALDI, equations used to obtain the S/N in tissue and background, corrected spectrum of the *adaptive background subtraction* method versus NECTAR for the MALDI data set, noise distribution on the DESI data before and after noise and baseline correction, peak classification for the DESI data, first 4 PCs of PCA for the top 4000 peaks vs denoised list obtained with NECTAR for MALDI and DESI data set, cumulative variance and t-SNE for the DESI data set (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Ariadna González-Fernández – National Physical Laboratory, Teddington, Middlesex TW11 0LW, U.K.;

orcid.org/0009-0003-3225-4718;

Email: ariadna.gonzalez@npl.co.uk

Josephine Bunch – National Physical Laboratory, Teddington, Middlesex TW11 0LW, U.K.;

Email: josephine.bunch@npl.co.uk

Authors

Alex Dexter – National Physical Laboratory, Teddington, Middlesex TW11 0LW, U.K.; orcid.org/0000-0001-8536-4417

Chelsea J. Nikula – National Physical Laboratory, Teddington, Middlesex TW11 0LW, U.K.; orcid.org/0000-0002-6817-2592

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/jasms.3c00116>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by Cancer Research UK [C58393/A24034] and the UK's National Measurement System programs.

■ ADDITIONAL NOTE

^aThe python package contains a jupyter notebook as well as a main.py file to run the pipeline on a imzML file, and obtain a

denoised list of peaks. The different steps on how to use the code are explained in the repository.

■ REFERENCES

- (1) Guibault, G. G.; Hjelm, M. Nomenclature for automated and mechanised analysis. *Pure and Applied Chemistry* **1989**, 61 (9), 1657–1664.
- (2) Baggerly, K. A.; Morris, J. S.; Wang, J.; Gold, D.; Xiao, L. C.; Coombes, K. R. A comprehensive approach to the analysis of matrix-assisted laser desorption/ionization-time of flight proteomics spectra from serum samples. *Proteomics* **2003**, 3 (9), 1667–1672.
- (3) Shin, H.; Mutlu, M.; Koomen, J. M.; Markey, M. K. Parametric power spectral density analysis of noise from instrumentation in MALDI TOF mass spectrometry. *Cancer Informatics* **2007**, 3, 219–230.
- (4) Busch, K. L. Chemical noise in mass spectrometry: Part I. *Spectroscopy (Santa Monica)* **2002**, 17 (10), 32–37.
- (5) Hilario, M.; Kalousis, A.; Pellegrini, C.; Müller, M. Processing and classification of protein mass spectra. *Mass Spectrom. Rev.* **2006**, 25 (3), 409–449.
- (6) Lieb, F.; Boskamp, T.; Stark, H.-G. Peak detection for MALDI mass spectrometry imaging data using sparse frame multipliers. *Proteomics* **2020**, 103852.
- (7) Malyarenko, D. I.; Cooke, W. E.; Adam, B. L.; Malik, G.; Chen, H.; Tracy, E. R.; Trosset, M. W.; Sasinowski, M.; Semmes, O. J.; Manos, D. M. Enhancement of sensitivity and resolution of surface-enhanced laser desorption/ionization time-of-flight mass spectrometric records serum peptides using time-series analysis techniques. *Clinical Chemistry* **2005**, 51 (1), 65–74.
- (8) Kilgour, D. P. A.; Hughes, S.; Kilgour, S. L.; Mackay, C. L.; Palmblad, M.; Tran, B. Q.; Goo, Y. A.; Ernst, R. K.; Clarke, D. J.; Goodlett, D. R. Autopicker - a Robust and Reliable Peak Detection Algorithm for Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2017**, 28 (2), 253–262.
- (9) Mantini, D.; Petrucci, F.; Pieragostino, D.; Del Boccio, P.; Di Nicola, M.; Di Ilio, C.; Federici, G.; Sacchetta, P.; Comani, S.; Urbani, A. LIMPIC: A computational method for the separation of protein MALDI-TOF-MS signals from noise. *BMC Bioinformatics* **2007**, 8, 1–17.
- (10) Zhurov, K. O.; Kozhinov, A.N.; Fornelli, L.; Tsybin, Y.O. Distinguishing Analyte from Noise Components in Mass Spectra of Complex Samples: Where to Cut the Noise? *Anal. Chem.* **2014**, 86, 3308.
- (11) Cappadona, S.; Levander, F.; Jansson, M.; James, P.; Cerutti, S.; Pattini, L. Wavelet-based method for noise characterization and rejection in high-performance liquid chromatography coupled to mass spectrometry. *Anal. Chem.* **2008**, 80 (13), 4960–4968.
- (12) Kwon, D.; Vannucci, M.; Joon, J. S.; Jeong, J.; Pfeiffer, R. M. A novel wavelet-based thresholding method for the pre-processing of mass spectrometry data that accounts for heterogeneous noise. *Proteomics* **2008**, 8 (15), 3019–3029.
- (13) Zhang, J.; Du, Q.; Song, X.; Gao, S.; Pang, X.; Li, Y.; Zhang, R.; Abliz, Z.; He, J. Evaluation of the tumor-targeting efficiency and intratumor heterogeneity of anticancer drugs using quantitative mass spectrometry imaging. *Theranostics* **2020**, 10 (6), 2621–2630.
- (14) Du, P.; Kibbe, W. A.; Lin, S. M. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics* **2006**, 22 (17), 2059–2065.
- (15) Coombes, K. R.; Tsavachidis, S.; Morris, J. S.; Baggerly, K. A.; Hung, M. C.; Kuerer, H. M. Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform. *Proteomics* **2005**, 5 (16), 4107–4117.
- (16) Meuleman, W.; Engwegen, J. Y. M. N.; Gast, M. C. W.; Wessels, L.F. A.; Reinders, M. J. T. Analysis of mass spectrometry data using sub-spectra. *BMC Bioinformatics* **2009**, 10 (SUPPL. 1), 1–9.
- (17) Wu, L.-c.; Chen, H.-h.; Horng, J.-t.; Lin, C.; Huang, N. E. A Novel Preprocessing Method Using Hilbert Huang Transform for

MALDI-TOF and SELDI-TOF Mass Spectrometry Data. *PLoS ONE* **2010**, *5* (8), e12493.

(18) Alexandrov, T.; Bartels, A. Testing for presence of known and unknown molecules in imaging mass spectrometry. *Bioinformatics* **2013**, *29* (18), 2335–2342.

(19) Inglese, P.; Correia, G.; Takats, Z.; Nicholson, J. K.; Glen, R. C. SPUTNIK: An R package for filtering of spatially related peaks in mass spectrometry imaging data. *Bioinformatics* **2019**, *35* (1), 178–180.

(20) Wang, W.; Li, Y. Bayesian denoising for remote sensing image based on undecimated discrete wavelet transform, *Proceedings - 2009 International Conference on Information Engineering and Computer Science*; ICIECS, 2009; p 1–4.

(21) Tyler, B. J.; Kassenböhmer, R.; Peterson, R. E.; Nguyen, D. T.; Freitag, M.; Glorius, F.; Ravoo, B. J.; Arlinghaus, H. F. Denoising of Mass Spectrometry Images via Inverse Maximum Signal Factors Analysis. *Anal. Chem.* **2022**, *94* (6), 2835–2843.

(22) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M. A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncio, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; MacInnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. HMDB: The human metabolome database. *Nucleic Acids Res.* **2007**, *35* (Database), D521–D526.

(23) Fahy, E.; Sud, M.; Cotter, D.; Subramaniam, S. LIPID MAPS online tools for lipid research. *Nucleic Acids Res.* **2007**, *35* (Web Server), W606–W612.

(24) Alexandrov, T.; Ovchinnikova, K.; Palmer, A.; Kovalev, V.; Tarasov, A.; Stuart, L.; Nigmatzianov, R.; Fay, D.; Gaudin, M.; Lopez, C. G.; Vetter, M.; Swales, J.; Bokhart, M.; Kompauer, M.; McKenzie, J.; Rappez, L.; Velickovic, D.; Lavigne, R.; Zhang, G.; Thinagaran, D.; Ruhlmann, E.; Sans, M.; Triana, S.; Sammour, D. A.; Aboulmagd, S.; Bagger, C.; Strittmatter, N.; Rigopoulos, A.; Gemperline, E.; Joensen, A. M.; Geier, B.; Quason, C.; Weaver, E.; Prasad, M.; Balluff, B.; Nagornov, K.; Li, L.; Linscheid, M.; Hopf, C.; Heintz, D.; Liebeke, M.; Spengler, B.; Boughton, B.; Janfelt, C.; Sharma, K.; Pineau, C.; Anderton, C.; Ellis, S.; Becker, M.; Pánczél, J.; Violante, G. D.; Muddiman, D.; Goodwin, R.; Eberlin, L.; Takats, Z.; Shahidi-Latham, S. METASPACE: A community-populated knowledge base of spatial metabolomes in health and disease. *bioRxiv* **2019**, 1–22.

(25) Alexandrov, T.; Kobarg, J. H. Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering. *Bioinformatics* **2011**, *27*, 230–238.

(26) Jones, E. A.; Deininger, S. O.; Hogendoorn, P. C. W.; Deelder, A. M.; McDonnell, L. A. Imaging mass spectrometry statistical analysis. *J. Proteomics* **2012**, *75* (16), 4962–4989.

(27) Wehrli, P. M.; Michno, W.; Blennow, K.; Zetterberg, H.; Hanrieder, J. Chemometric Strategies for Sensitive Annotation and Validation of Anatomical Regions of Interest in Complex Imaging Mass Spectrometry Data. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (11), 2278–2288.

(28) Murta, T.; Steven, R. T.; Nikula, C. J.; Thomas, S. A.; Zeiger, L. B.; Dexter, A.; Elia, E. A.; Yan, B.; Campbell, A. D.; Goodwin, R. J. A.; Takats, Z.; Sansom, O. J.; Bunch, J. Implications of Peak Selection in the Interpretation of Unsupervised Mass Spectrometry Imaging Data Analyses. *Anal. Chem.* **2021**, *93* (4), 2309–2316.

(29) Yang, C.; He, Z.; Yu, W. Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis. *BMC Bioinformatics* **2009**, *10*, 1–13.

(30) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M. Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.

Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **2012**, *30* (10), 918–20.

(31) Race, A. M.; Styles, I. B.; Bunch, J. Inclusive sharing of mass spectrometry imaging data requires a converter for all. *Journal of Proteomics* **2012**, *75* (16), 5111–5112.

(32) Race, A. M. Investigation and Interpretation of Large Mass Spectrometry Imaging. Thesis. 2016 (May).

(33) Hassan, A. H.; Fluke, C. J.; Barnes, D. G.; Kilborn, V. A. Tera-scale astronomical data analysis and visualization. *Mon. Not. R. Astron. Soc.* **2013**, *429* (3), 2442–2455.

(34) Maples, M. P.; Reichart, D. E.; Konz, N. C.; Berger, T. A.; Trotter, A. S.; Martin, J. R.; Dutton, D. A.; Paggen, M. L.; Joyner, R. E.; Salemi, C. P. Robust Chauvenet outlier rejection. *arXiv* **2018**, DOI: 10.48550/arXiv.1807.05276.

(35) Sánchez-Monge, A.; Schilke, P.; Ginsburg, A.; Cesaroni, R.; Schmiedeke, A. STATCONT: A statistical continuum level determination method for line-rich sources. *Astron. Astrophys.* **2018**, *609*, 1–15.

(36) Coombes, K. R.; Koomen, J. M.; Baggerly, K. A.; Morris, J. S.; Kobayashi, R. Understanding the characteristics of mass spectrometry data through the use of simulation. *Cancer Informatics* **2005**, *1* (1), 41–52.

(37) Ipsen, A. Derivation from first principles of the statistical distribution of the mass peak intensities of MS data. *Anal. Chem.* **2015**, *87* (3), 1726–1734.

(38) Shuqin, Z. A Novel peak detection approach with chemical noise removal using short-time FFT for PrOTOF MS data. *Proteomics* **2009**, 1–7.

(39) Green, M. R.; Richardson, K.; Palmer, M.; Denny, R.; Campuzano, I.; Skilling, J. A NOVEL ADAPTIVE BACKGROUND SUBTRACTION ALGORITHM FOR REDUCTION OF CHEMICAL NOISE IN MASS SPECTRA. Number of Peptides. *Proceedings of the 54th ASMS Conference on Mass Spectrometry and Allied Topics* **2006**, 2006–2006.

(40) Krutchinsky, A. N.; Chait, B. T. On the nature of the chemical noise in MALDI mass spectra. *J. Am. Soc. Mass Spectrom.* **2002**, *13* (2), 129–134.

(41) Andreev, V. P.; Rejtar, T.; Chen, H. S.; Moskovets, E. V.; Ivanov, A. R.; Karger, B. L. A Universal Denoising and Peak Picking Algorithm for LC-MS Based on Matched Filtration in the Chromatographic Time Domain. *Anal. Chem.* **2003**, *75* (22), 6314–6326.

(42) Taylor, A. J.; Dexter, A.; Bunch, J. Exploring Ion Suppression in Mass Spectrometry Imaging of a Heterogeneous Tissue. *Anal. Chem.* **2018**, *90* (9), 5637–5645.