analytical chemistry

pubs.acs.org/ac
Article

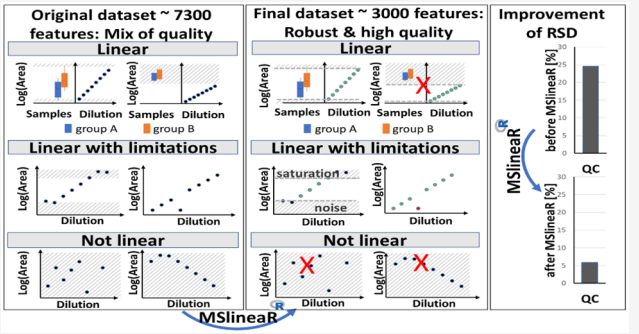
MSlineaR: An R Package Assessing Linear Behavior to Improve Quality Assurance and Statistical Robustness in Untargeted Metabolomics

Janine Wiebach,* Álvaro Fernández-Ochoa, Ulrike Bruning, Jochen Kruppa-Scheetz, Maëlle Bonhomme, Dominique-Marie Votion, and Jennifer A. Kirwan*

Cite This: https://doi.org/10.1021/acs.analchem.5c04480
Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Mass spectrometry-based untargeted metabolomics analyzes complex biological matrices containing thousands of individual features. Linearity is a key analytical parameter in quantitative mass spectrometry, reflecting proportionality between signal intensity and analyte concentration within a defined range. In untargeted metabolomics, linearity cannot be directly assessed due to the absence of reference concentrations. Instead, the range in which features exhibit approximately linear dilution-dependent behavior (ALB) can be evaluated as a practical proxy. Feature selection based on this response enables the early removal of noise and unreliable features, thereby reducing the risk of false-positive findings and improving analytical robustness, while the reduced number of retained features lowers the multiple-testing burden in downstream statistical analyses. We present MSlineaR, an open-source R-based software tool implementing a six-step process to assess dilution-dependent response behavior in metabolomic data sets. MSlineaR evaluates dilution curves to identify nonclassical response patterns, detect outliers, and iteratively trim boundary regions to remove plateau effects. It then reassesses the remaining data to retain features exhibiting ALB. Importantly, it defines boundaries of the approximately linear range (ALR) and applies them for data curation, enabling exclusion of unreliable signals while preserving robust features. Application to three independent data sets demonstrated a ∼10% improvement in the number of features classified as exhibiting ALB compared to classical linear regression-based approaches. MSlineaR was complementary to relative standard deviation (RSD) filtering, improved median RSD values and enhanced the robustness of statistical modeling.



INTRODUCTION

Mass spectrometry-based untargeted metabolomics involves the analysis of complex biological matrices, often containing thousands of individual features. Early curation of data sets to retain highly robust features reduces the risk of false-positive findings. Additionally, reducing the total number of statistical tests lessens the multiple-testing burden, thereby potentially improving the likelihood of detecting true biological differences. It also increases the probability that results are reproducible. Therefore, untargeted metabolomics workflows commonly incorporate data curation steps aimed at removing low-quality molecular features that lack analytical reliability or biological relevance.[1−3]

The use of matrix-specific quality control samples (QC) to assess technical reproducibility and curate data sets is now standard[2] and focuses on the removal of contaminants detected in blank samples, low-abundant features, or those with poor analytical precision. Data quality is often assessed using the relative standard deviation (RSD) of QCs and/or the dispersion-ratio (D-ratio) parameter.[1] However, these metrics primarily reflect technical reproducibility at a fixed concentration and may underestimate overall signal variability as they cannot capture potential heteroscedasticity across the concentration range. In addition to the previously described data curation metrics, the use of a dilution series of QC samples has been proposed as a valuable strategy to assess linearity and the linear range of detected molecular features.[2,4−6]

Linearity is a key analytical parameter in quantitative mass spectrometry, reflecting the proportionality between signal intensity and analyte concentration within a defined range.[7] In targeted analysis, linearity is assessed using calibration curves and is essential for accurate absolute quantification. By contrast,

Received: July 28, 2025
Revised: July 17, 2026
Accepted: August 4, 2026

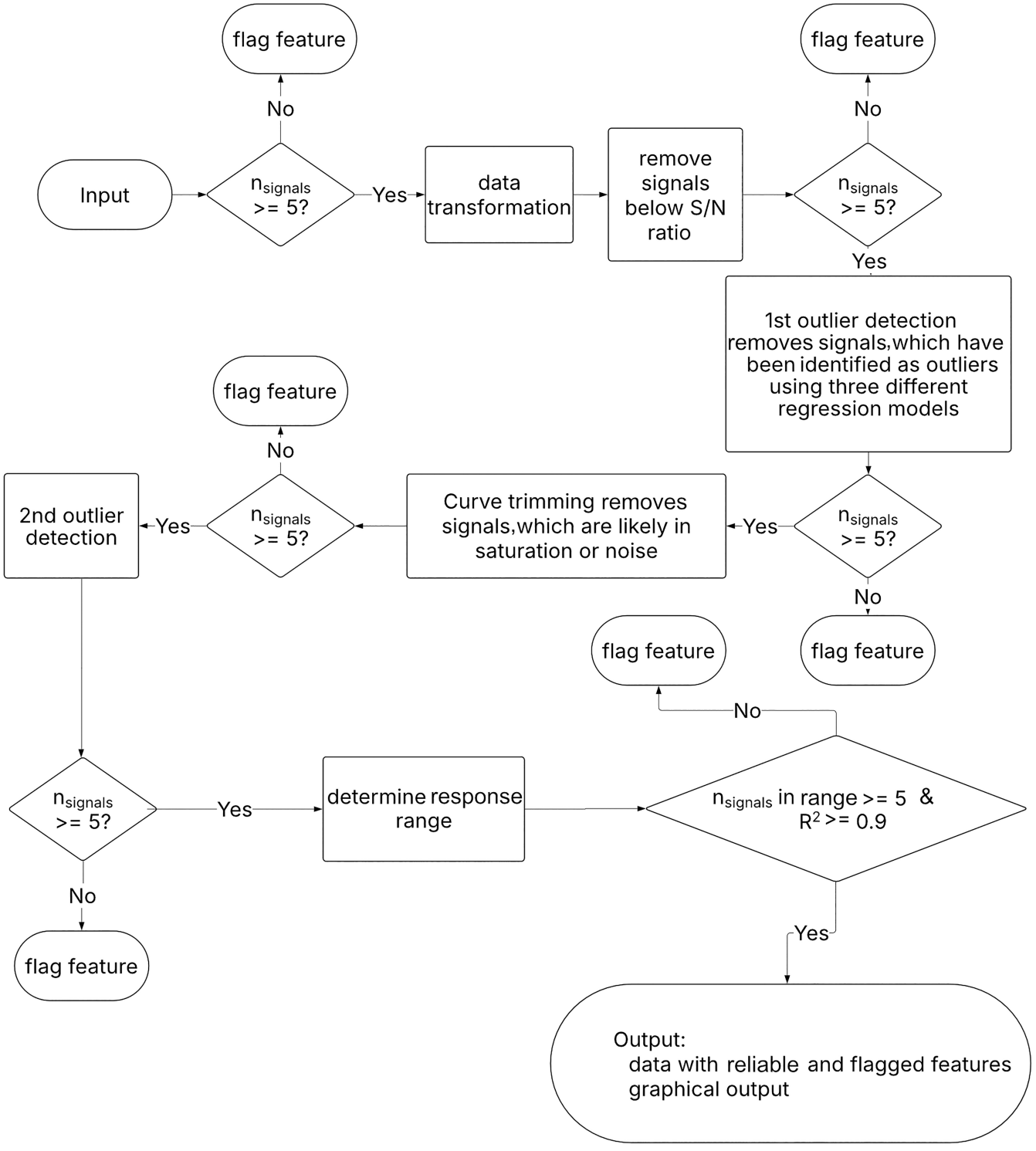


Figure 1. Process diagram of MSlineaR, describing all six functions used for filtering the data according to their dilution-dependent response behavior.

true concentration-signal proportionality is difficult to establish in untargeted analysis due to the absence of known analyte identities and concentrations. Instead, the relationship between signal intensity and dilution factor can act as a proxy to identify features that exhibit reproducible, dilution-dependent behavior suitable for downstream statistical analysis. Accordingly, the terms "linearity" and "linear range" are reserved for targeted analysis in this manuscript. For untargeted analysis, features are instead described in terms of their dilution-dependent signal behavior, which may approximate a linear response within a restricted range. The terms "approximately linear dilution-dependent behavior" (ALB) and "approximately linear range" (ALR) are therefore used to reflect model-based approximations of signal behavior, rather than calibration-defined analytical linearity. Within this context, a feature is defined as a specific RT-m/z pair that can be consistently measured across dilution points. For each dilution point i (i = 1, ..., n) of a sample, this measurement yields a corresponding signal (S_i), collectively forming a dilution curve. This approach is still not widely

https://doi.org/10.1021/acs.analchem.5c04480
Anal. Chem. XXXX, XXX, XXX−XXX

adopted in the untargeted metabolomics community,[2] partly due to limited availability of robust postprocessing tools.

In recent years, feature-based assessment of dilution-dependent behavior and filtering of unreliable features have been automated using bioinformatic tools.[5,8−12] However, to the best of our knowledge, available tools primarily rely on the coefficient of determination (R^2) and/or slope across the full dilution range and often do not explicitly account for missing values or outliers. Additionally, the ability to determine the bounds of dilution-dependency and identify samples that lie outside these bounds may improve the robustness of the statistical results by reducing false positives and thereby enhancing experimental reproducibility.

To address these challenges, MSlineaR was developed to automatically and objectively assess dilution-dependent behavior and define ranges of ALB per feature in untargeted metabolomic data sets. It enables data set curation by retaining analytically reliable features for statistical analysis. Using a feature-by-feature approach, MSlineaR applies a six-step process to assess dilution-dependent behavior and identify "non-classical" dilution curve shapes in QC dilution series. It uses three curve fitting models to determine data structure and removes outliers. Nonresponsive regions (e.g., plateaus) at the beginning and end of concentration ranges are identified to define a consistent response range. A graphical function visualizes sample distribution across the responsive and nonresponsive regions. Finally, features are flagged or excluded when the acquired biological samples lie outside the defined range. To the best of our knowledge, no other tool exists that can automatically handle the observed range of different dilution series shapes, including negative slopes, outliers, and missing values, while establishing dilution-dependency and ranges of ALB. This integrated solution enhances data reliability in untargeted metabolomics and can also support targeted workflows, with the possibility of incorporating quantitative metrics.

EXPERIMENTAL SECTION

Feature Filtering Based on Approximately Linear Dilution-Dependent Behavior Using MSlineaR

The source code of MSlineaR is available on GitHub (https://github.com/bih-metabolomics/MSlineaR).

Data sets were analyzed with a Windows 10 × 64 build 19045 PC (Intel(R) Core(TM) i7−9700 CPU @ 3.00 GHz, 16 GB RAM) using seven cores for parallelization.

The algorithm performs six steps, outlined in Figure 1, to produce a data set containing only features exhibiting dilution-dependent behavior that can be approximated by a linear model, with most samples falling within the corresponding ALR.

MSlineaR requires two data tables (MetaData and FeatureData) as input (templates in Tables S-1 and S-2). The FeatureData table contains information about detected mass spectrometry features, including unique feature names, unique sample names, batch information, and peak areas or intensities. The MetaData table includes information about injection order, unique sample names, sample type, statistical class, and batch information. In addition, dilution factors of the serially diluted samples are provided and normalized to the highest dilution (lowest concentration = 1). For targeted approaches, the expected concentration of the standards can be provided in an additional column. Identical sample names and batches must be used for both data tables. All mandatory and optional parameters are described in detail in Table S-3.

The functions are described below in the order in which they are applied to a single dilution series. For the full mathematical formulation, see Table S-4. The code itself can be executed in parallel. Missing values are retained as missing. After each step, the number of dilution points is checked against the parameter *min_feature,* which describes the minimum number of consecutive signals required per dilution curve for the curve to be considered reliable. Curves with a lower number of dilution points are excluded.

(i) First, data are log−log-transformed to stabilize variance and approximate normality of residuals.

(ii) Next, the lower limit for the range of ALB is calculated as the median of the blank samples multiplied by the parameter *signal_blank_ratio* (default = 5). All dilution points below this threshold are removed. This also allows detection and exclusion of potential contaminants.

(iii) Based on previous observations, we assume that the curve of a biological feature in a dilution series will follow a linear, quadratic or logistic dependency. Therefore, each of the three models is fitted to the dilution series. Model performance is evaluated using the root-mean-square error (RMSE) calculated on the full data set. The model with the lowest RMSE is selected as the best representation of the dilution distribution. Outliers of the selected model are identified using standardized residuals (default $|r| > 2$), which account for residual variance and leverage, providing a scale-independent measure of deviation.

(iv) Metabolites near the upper or lower limits of their detection range may show unusual feature detection behavior due to saturation, ion suppression, or matrix interference and should not be considered within the range of ALB.[13] Plateau-affected regions from dilution curves are removed using an iterative trimming procedure based on local slope and boundary constraints. First, local slopes are computed between adjacent points by dividing the difference in signal intensity by the difference in dilution factor. These slopes capture local changes in signal response along the dilution series.

To distinguish true signal progression from plateau regions, a reference slope is estimated from a subset of points around the midintensity range (corresponding approximately to the half-maximal signal region), where the response is expected to be stable and least affected by saturation or noise.[14]

Plateau regions are identified where local slopes fall substantially below a user-defined fraction of the reference slope (default $\alpha = 0.5$). These regions typically occur at the low-intensity end (noise-dominated) or high-intensity end (saturation).

Boundary constraints ensure that the retained subset represents a physically meaningful response range. The retained range spans a monotonic segment across the observed signal range, preventing inclusion of nonmonotonic edge regions and ensuring that the trimmed curve reflects the central dynamic range. Points at the boundaries of the current range are removed if they violate either the slope-based plateau criterion or the boundary constraints. After each iteration, local slopes at the extremes and the reference slope are compared on the reduced data set. This process continues until no further points are removed, resulting in convergence to a stable set of dilution points.

This combined approach allows MSlineaR to robustly exclude saturation-induced flattening at high concentrations and noise-driven instability at low concentrations, while preserving the central response region suitable for downstream evaluation.

(v) After reducing the potential range of ALB, a second round of outlier detection is performed using the same method as previously described.

(vi) Finally, MSlineaR assesses the range of ALB of the curve by fitting a linear regression model through the remaining dilution points and calculating studentized residuals and Cook's distance. All dilution points with studentized residuals below a user-defined threshold (default = 3) and Cook's distance below $4/(n - p - 1)$, where n is the number of observations and p is the number of model parameters, are assigned to the range of

https://doi.org/10.1021/acs.analchem.5c04480
Anal. Chem. XXXX, XXX, XXX−XXX

ALB. R^2, Spearman correlation, absolute and relative deviation and the slope are calculated for the ALB of all remaining features (Table S-5).

Only curves with sufficient consecutive dilution points (>*min_feature*) and an R^2 greater than or equal to a user-defined threshold (default = 0.9) are flagged as having an ALR.

MSlineaR includes an additional function (MS_filterSamples) for filtering biological sample data based on the upper and lower limits of the range of ALB. Features are retained if at least a user-defined proportion of samples fall within this range (default = 70% per group). The output consists of three data tables, showing information (1) for each dilution curve, (2) for each signal of the dilution curves and (3) for all injected samples. In addition, there is a graphical output to show the dilution curves and linear proportion for all samples and QCs and diagnostic plots for residual behavior and Q-Q plots.

Data Sets

Three data sets were used to demonstrate different aspects of MSlineaR performance. All data sets are liquid chromatography−mass spectrometry based and full details can be found in Supporting Information (Texts S-1, S-2, S-3). Data set 1 (DS1) was used to assess the sensitivity and specificity of MSlineaR. It is based on a previously validated method[15] and consists of a targeted analysis of 33 tryptophan metabolites in 101 charcoal-stripped EDTA plasma samples measured as both a dilution series (10 dilution points) and as a pooled QC every sixth sample. A medium range dilution from the standard curve was analyzed every 10th sample as a second QC check, interspersed with biological samples (data from the biological samples were not used for this study). A second batch was run a day later with the same setup (Table S-6). Data set 2 (DS2) is a method validation data set originally designed to test the robustness of different published untargeted methods[16−18] and was used to evaluate MSlineaR on a complex matrix and to assess its ability to identify true negatives. The test samples consisted of a commercial human plasma sample (GTX73265, GeneTex) extracted using four different solvent compositions (method 1 "MeOH3": plasma to methanol ratio 1:3; method 2 "MeOH1.5": plasma to methanol ratio 1:1.5; method 3 "MeOH9": plasma to methanol ratio 1:9; method 4 "MeOH_EtOH3": plasma to methanol:ethanol mixture (1:1) in a final ratio of 1:3). After protein precipitation, the supernatants were collected and fully dried. Prior to analysis, samples were reconstituted in 50 μL of water containing IROA-^{12}C internal standards (TruQuant Yeast Extract Semitargeted QC Workflow Kit, IROA Technologies). Five replicates of each extraction method were randomly analyzed in each batch (Table S-7). It is important to note that the final injection concentration of MeOH9 samples was approximately 8-fold lower, and MeOH1.5 2-fold lower, than for the other extraction methods. Nine QCs per batch were prepared using the same method as for MeOH3 samples. A dilution series consisting of 9 dilution points was prepared with ratios of 0.05, 0.10, 0.25, 0.50, 0.75, 1, 1.5, 2, 3 to the QC concentration (Table S-8). Three identical analytical batches were analyzed, each with its own dilution series. In a subsequent validation study to assess the usefulness of MSlineaR at very low concentrations, the resuspended QC sample was further diluted 1:3 in the resuspension solvent and a second "pre-diluted" 9-point dilution curve was prepared and analyzed as described above.

The last data set (DS3) was used to illustrate the application of MSlineaR on real-world data. DS3 is derived from a longitudinal field study conducted on horses from three different locations over an 18-month period. Over the study period, venous blood samples were collected every two months. The study aimed to identify metabolomic biomarkers associated with the development of a specific condition (unpublished data, Bonhomme et al., in preparation). Fourteen horses were initially included (n = 7 precondition cases, n = 7 matched controls). However, two horses (one case and one control) originating from a third location were excluded from statistical analysis due to too few samples from that site. All samples were collected in EDTA tubes under field conditions, then processed and analyzed using LC-MS. Samples collected during the 0−2, 2−4, 4−6, and 6−8-month intervals preceding the onset of the condition, and at matched intervals for the controls, were analyzed. These intervals reflect the regular bimonthly sampling visits rather than precise time points relative to the event. Aliquots of all samples were combined to create a pooled QC which was run every 10th sample (Table S-9). A dilution series of the pooled QC was prepared and IROA-^{12}C internal standards were added in equal amounts as described for DS2.

Data Preprocessing without MSlineaR

DS1 was exported into Skyline (v.19.1−64 bit) to identify and quantify peak intensity and area. Data were manually checked to ensure the correct peaks were selected (Text S-1). After internal standard normalization, a cubic spline drift correction using the R notame package[3] (version 0.3.2) was applied per metabolite and to all sample types, with the pooled QC samples as references for fitting the splines.[19] Standards of increasing concentration were background subtracted and used to construct calibration curves for each metabolite using a weighted linear regression model. Compounds with an $R^2 > 0.96$ and at least five dilution points were defined as linear. Concentration values less than or equal to zero were treated as missing values.

For DS2 and DS3, the acquired data were first converted to .mzML format using MSConvertGUI (https://proteowizard.sourceforge.io/). Subsequent data processing was carried out using ClusterFinder (version 4.2.67, IROA Technologies, USA) and with R (version 4.2.0) in the RStudio environment (version 2023.6.2.561).

Peak detection, retention time correction, alignment and isotope and adduct annotation were conducted using XCMS (version 4.6.4) and CAMERA (version 1.64.0) packages[20−23] for DS2 or XCMS Online (https://xcmsonline.scripps.edu, version 3.7.1)[24,25] for DS3. All parameters used in XCMS and XCMS Online can be found in Texts S-2 and S-3.

Features were filtered out if they were below a signal-to-noise ratio of 3:1 compared to the median of blank samples or if they were not detectable in all batches and in at least 80% of samples. Features were excluded where fewer than five pooled QC measurements for that feature were present. Data normalization for DS3 was performed using probabilistic quotient normalization (pqn) with the R package pmp[26] (version 1.20.0). Due to the extreme differences in preanalytical concentrations in DS2, we have omitted a postanalytical normalization step. The notame package[3] (version 0.3.2) was used to correct analytical intrabatch drift deviations using QC samples as a reference for the cubic spline method.[19,26] To correct for interbatch deviations, a groupwise median normalization strategy was used. Technical and biological variabilities were assessed per feature over the pooled QCs by calculating the RSD and D-ratio, respectively. All features were removed if they showed technical variability above 20% (RSD_{QC} > 20%) or a D-ratio[1] above 40%. These filters were applied to ensure that only reproducible features with biological relevance were retained. Metabolite annotation was performed using Compound Discoverer software (Thermo Fisher, 3.3 SP2) and ClusterFinder (version 4.2.67, IROA Technologies, USA).

Filtering Using MSlineaR

For all data sets, MSlineaR was applied before data normalization. DS1 was filtered by MSlineaR with parameters set as follows: *signal_blank_ratio* = 1 (background subtraction and signal-to-noise filtering were applied beforehand), *min_feature* = 5, *min_R2* = 0.96. For DS2 the parameters were set as *signal_blank_ratio* = 3, *min_feature* = 5 and *min_R2* = 0.9 and a threshold of 70% was used per sample group to classify the samples as within ALR or outside. For DS3, the same parameters as for DS2 were used.

Statistical Analysis/Evaluation of MSlineaR

All statistical analyses were performed with R (version 4.2.0) in the RStudio environment (version 2023.6.2.561). To evaluate the performance of MSlineaR, we compared processing pipelines with and without MSlineaR integration using RSD, R^2, number of detected IROA features and principal component analysis (PCA). Prior to PCA, the data were glog-transformed to address heteroscedasticity and skewed distributions.[27] To avoid the influence of missing value imputation, PCA was done solely on features present in all batches and in all samples.

https://doi.org/10.1021/acs.analchem.5c04480
Anal. Chem. XXXX, XXX, XXX−XXX

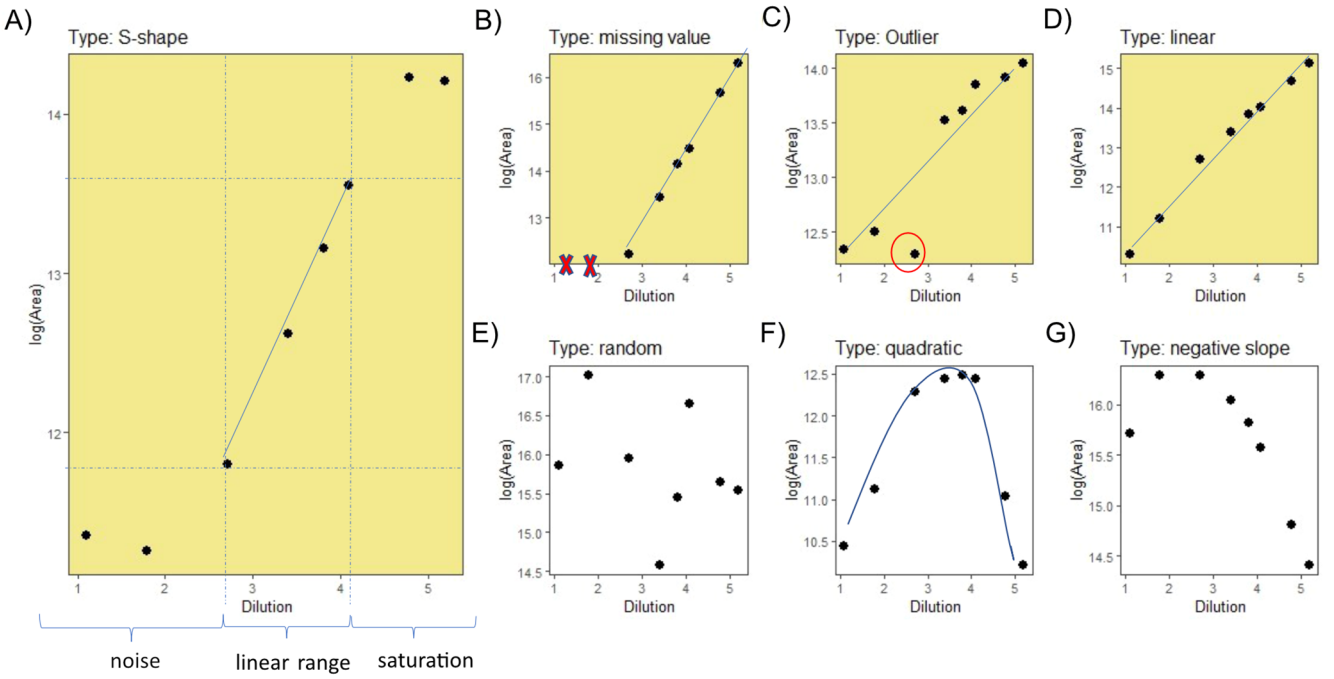


Figure 2. Scatterplots of dilution series from 7 different features showing the variety of behaviors observed for features in a dilution series. A) shows the expected S-shape for a feature on a log−log scale. The consistent response range is highlighted as the central square bounded by the dashed lines. (B−D) show alternative acceptable dilution curve behaviors of, which can be classified as approximately linear, and should be retained (colored background). (E−G) are inconsistent with approximately linear behavior and should be removed.

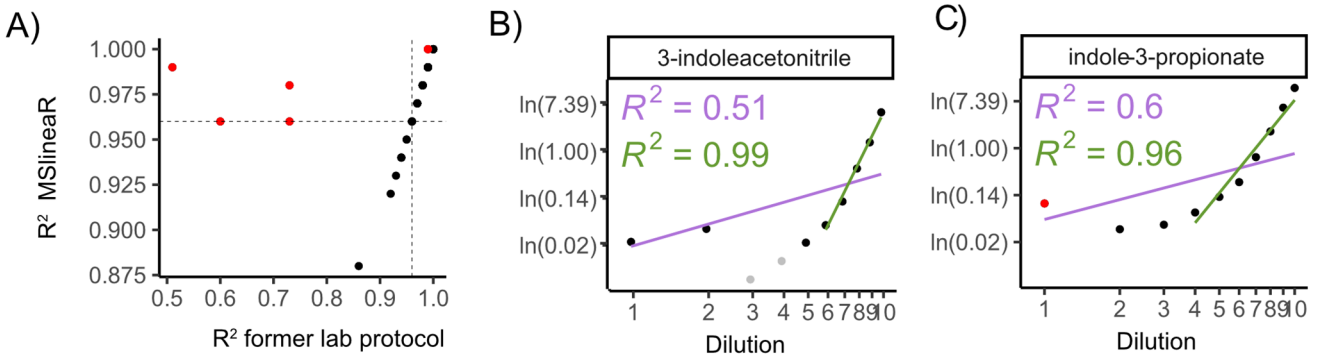


Figure 3. (A) Comparison of R^2 values obtained from DS1 with MSlineaR and the previous method. Compounds rescued by MSlineaR due to improved final R^2 are marked in red. (B and C) Two examples of such features from DS1. The purple line indicates the standard linear regression model, while the green line shows the regression after applying MSlineaR. Gray points indicate concentrations that were removed due to a low signal-to-noise ratio, and red points represent outliers as identified by MSlineaR.

To assess within-group variation, the sum of the squared differences from the group-mean was calculated and compared between data sets analyzed with and without MSlineaR. Between-group variation was calculated as the sum of the group-mean differences compared to the overall mean.

Correct classification was validated by manual inspection of the log−log-plotted graphs of all concentration curves. The curves were checked for visible outliers, shape, and the number of dilution points that were within the ALR.

For DS1, weighted linear regression ($y = \beta_0 + \beta_1 \times x + \epsilon$ with weight for each dilution point i $w_i = \frac{1}{x_i^2}$), representing our standard laboratory workflow, was used for comparative evaluation of non-transformed, untrimmed data and data processed using the MSlineaR trimming workflow. R^2 and linear range as well as the concentration measurement accuracy from the medium range QC samples were compared.

For DS2, data were log−log-transformed and fitted using a nonweighted linear regression. For this comparison the R^2 and ALR were used.

For DS3, to evaluate the impact on downstream statistics, metabolomic differences between the two classes were assessed using a Mann−Whitney U test of all annotated features. FDR-adjusted q-values <0.1 were considered significant.

RESULTS AND DISCUSSION

We benchmarked MSlineaR against both automatic (current method) and manual assessments of dilution-dependent response behavior, using targeted and untargeted data sets.

We started by exploring the dilution curve outputs from real-world data sets. The challenge of using dilution curves to assess dilution-dependent response behavior is as follows: in log−log scale, biological features in a dilution series are expected to follow an S-shaped curve (Figure 2A). In reality, a wide range of

https://doi.org/10.1021/acs.analchem.5c04480
Anal. Chem. XXXX, XXX, XXX−XXX

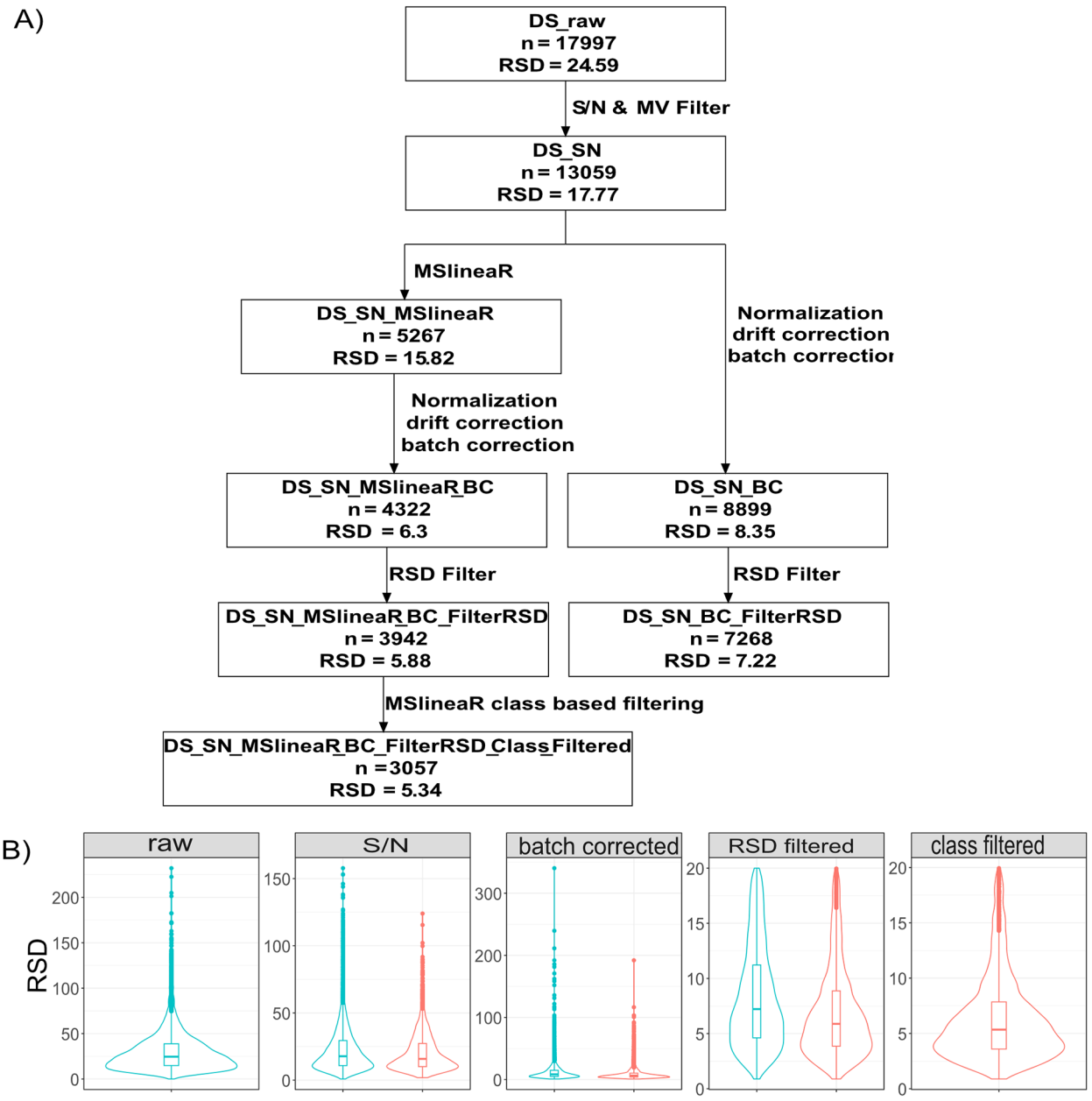


Figure 4. (A) Flowchart for processing an untargeted data set (DS2) with MSlineaR on the left side and without MSlineaR on the right side. The number of features and the median RSD value are shown after each step. (B) RSD improvement after each processing step without (blue) and with MSlineaR (red).

shapes can be observed, the most common of which are displayed in Figure 2B−G. Bioinformatic tools must identify, classify, and filter so that only robust features are kept (Figure 2A−D) and other shapes are flagged for exclusion (Figure 2E−G). Outliers and missing values (Figure 2B and C) are particularly challenging. If not excluded, they will negatively influence the linear regression calculation,[28] which subsequently can lead to the complete exclusion of the feature. Quadratic or parabolic shapes represent a second challenge. Once the signal reaches saturation, counterintuitively, the detected intensity can decrease. In the worst case, this can lead to a full parabolic shape (Figure 2F). However, in most cases, a "partial" parabolic shape is observed.

Sensitivity and Accuracy Testing: MSlineaR Correctly Classified 100% of the Compounds in a Targeted Data Set (DS1)

Using DS1, our first evaluation was an assessment of the sensitivity and specificity of MSlineaR, i.e., its ability to correctly identify peaks that should be linear. DS1 is a targeted data set consisting of a mixture of 33 commercial standards prepared and analyzed as a concentration curve of 10 concentration levels. This was repeated in a second batch yielding 66 concentration curves in total.

For 32 compounds (97%), MSlineaR identified a linear range in both batches. One compound, formylanthranilate, was excluded in both batches, because fewer than five concentration levels were detected. Manual inspection of all concentration curves confirmed the correct classification by MSlineaR for all compounds.

https://doi.org/10.1021/acs.analchem.5c04480
Anal. Chem. XXXX, XXX, XXX−XXX

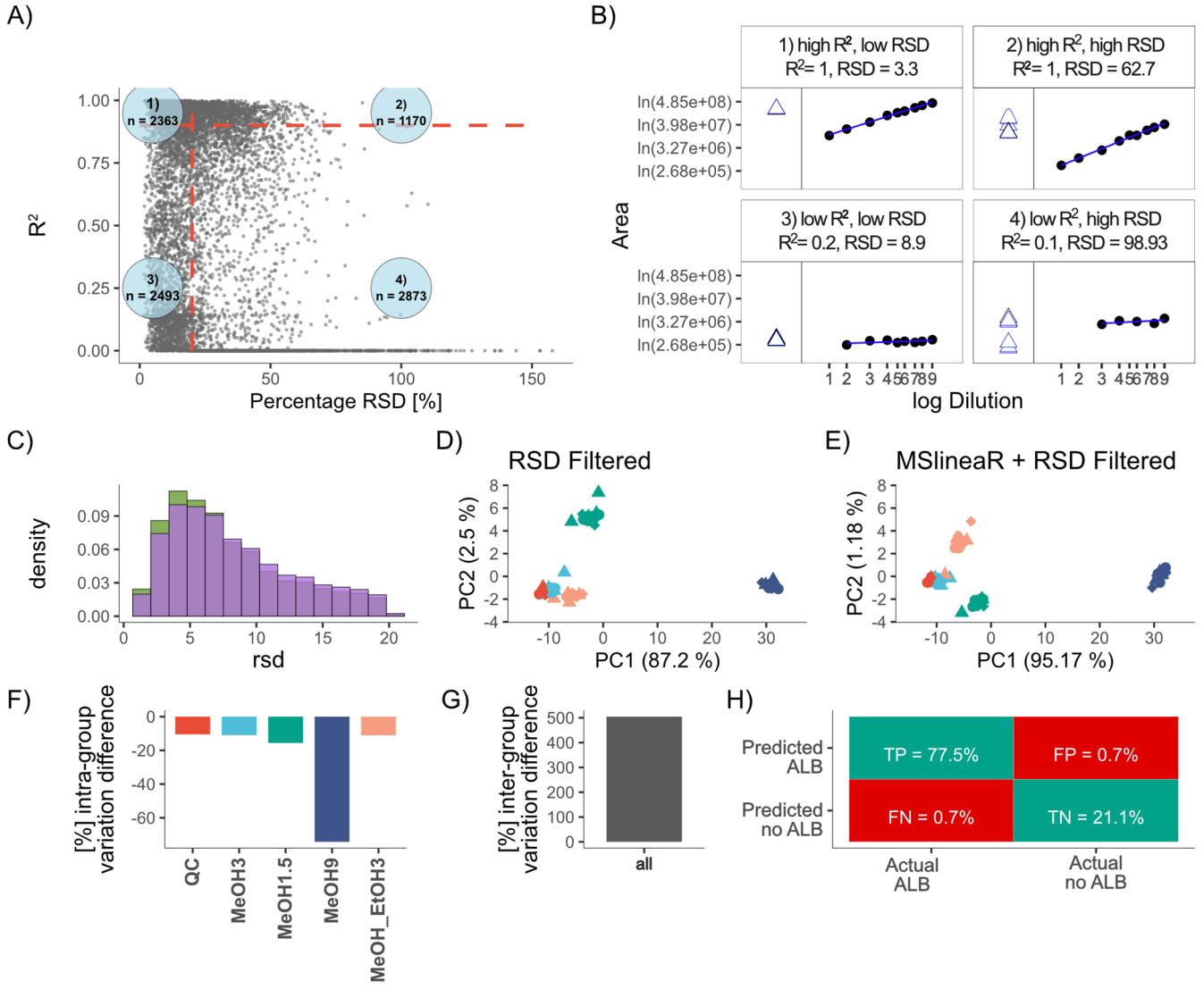


Figure 5. (A) Scatterplot of features from a single batch (batch B3) from data set DS2. The plot is divided into four quadrants based on an RSD_{QC} threshold of 20% and an R^2 threshold of 0.9 (red dashed lines). The number of features in each quadrant is detailed in the numbered circle. Each quadrant reflects a level of data quality: quadrant 1 represents the highest-quality features, while quadrant 4 includes the lowest-quality features. Quadrant 3 denotes features that would pass RSD_{QC} filtering, but which should be excluded based on their poor linear behavior. (B) Examples of peak behavior from each quadrant. The linear fit of the dilution curve is shown on the right, and the measured intensities of the pooled QC samples on the left. (C) Frequency histogram displaying the distribution of RSD_{QC} values from the filtered data set with (green) and without (lilac) MSlineaR filtering. (D-E) PCA score plots of all three batches without (D) and with (E) MSlineaR filtering. Plots are produced to the same scale. The variance explained by PC1 increases with use of MSlineaR and MSlineaR produces a tighter clustering of sample groupings. Analytical batches are indicated by shape, and extraction method by color. (F) Percentage change in intragroup variation without and with MSlineaR filtering, colored according to extraction methods. (G) Overall intergroup variation difference. (H) Percentage classification accuracy of MSlineaR based on 142 IROA standards.

Next, the R^2 and linear range were compared between MSlineaR and our former lab protocol.

With both methods, we observed equivalent R^2 and an equal linear range for 54 (84%) concentration curves. The linear ranges of 10 (16%) concentration curves were trimmed by MSlineaR, resulting in improved overall modeling of the data (two examples shown in Figure 3B and C). This is supported by an increase in R^2 observed in 8 (13%) concentration curves. When our former lab protocol was applied to the same data, 5 of the 8 concentration curves yielded an R^2 below 0.96 and would typically have been discarded. Using MSlineaR, we were able to recover four concentration curves (highlighted as red dots in Figure 3A). The median R^2 of the truncated curves increased from 0.92 to 0.98 with MSlineaR. In Table S-10 we show that regardless of which R^2 cutoff we choose, MSlineaR retained more metabolites than the established standard method.

We detected a change in the mean accuracy for all features where the concentration curve was truncated. For eight out of the ten modified concentration curves, the mean accuracy improved by an average of 6.8% with MSlineaR. For two features, the previous method yielded slightly better results, with an average difference of 1%.

Using this data set, we demonstrate that MSlineaR was able to "rescue" 13% of features which would otherwise have been rejected. Fitrianto et al.[28] have demonstrated that outliers and unusual curve behavior have a significant influence on the R^2 and can lead to inaccurate assessments of the true correlation between signal intensity and concentration. MSlineaR reduces

https://doi.org/10.1021/acs.analchem.5c04480
Anal. Chem. XXXX, XXX, XXX−XXX

these occurrences by applying an automated and objective outlier detection method, followed by truncation of features to consider only the trimmed linear range. These features distinguish MSlineaR from other tools.[5,8−12]

An additional advantage of determining the ALR for each feature is that it can provide an analogue of the targeted analytical methods of analyzing each standard at low, medium and high concentrations to assess behavior, especially when comparing dilution curves over multiple batches.

Specificity Testing: Untargeted Data Set DS2

Next, a technical data set (DS2) was used as an example of an untargeted data set. This was both to evaluate MSlineaR's performance on a more complex data set and to assess its ability to correctly identify true negatives. The samples are technical replicates from a commercial human plasma sample and were extracted with four different methods. In total, XCMS detected 22,767 features, of which ∼18,000 features were detected across all three batches. After preprocessing, the final data set consisted of 8,899 features consistently detected across all three batches (Figure 4A and B).

Evaluation of Approximately Linear Behavior as a Complementary Quality Control Metric

Filtering based on relative standard deviation (RSD) of pooled quality control samples is one of the most common methods employed to curate poorly reproducible peaks. The next step aimed to demonstrate the complementarity of MSlineaR filtering with this approach. The median RSD of QC samples (RSD_{QC}) of every feature from one representative batch (B3) is plotted against the R^2 calculated from MSlineaR in Figure 5A. The plot was subdivided into four quadrants based on a parallel evaluation of both the RSD_{QC} and the R^2 of the dilution curve using thresholds of 20% and 0.9, respectively. A total of 55% (4,856 features) of the features (quadrants 1 and 3, Figure 5B top and bottom left) showed acceptable QC technical reproducibility with RSD_{QC} less than or equal to 20%. Among these, approximately half (2,363) also had an R^2 of greater than or equal to 0.9, indicating a strong linear model fit across the evaluated dilution range and reproducibility. The remaining half (quadrant 3) showed low consistency with ALB across the dilution range, raising concerns about their suitability for quantitative statistics. This is precisely the type of feature that MSlineaR is designed to identify and remove from downstream analyses. The 1,170 features in quadrant 2 (Figure 5B top right) would typically be excluded due to high RSD_{QC} despite showing good linear model fit (i.e., $R^2 > 0.9$). Identifying this subgroup allows for further investigation into the causes of poor reproducibility, such as chemical variability, intrabatch drift, suboptimal dilution coefficients, or the presence of outliers among the QC samples.

MSlineaR Improves General Filter Strategy

To better understand the effects of MSlineaR and its complementary effect on RSD filtering, the data set was processed three times, using standard RSD filtering (RSD_{QC} < 20%) alone, MSlineaR alone, or both filtering approaches combined (Figure 4A). Filtering the data solely on technical and biological variance reduced the data set to 7,268 features (40%). Applying MSlineaR alone as a filter step retained 4,322 features (24%), while combining both filtering approaches yielded 3,942 features (22%). This suggests that nearly 50% of technically reproducible features do not show an adequate fit to an approximately linear model across the dilution range and are therefore suboptimal candidates for quantitative statistical analysis. Conversely, not all features showing good ALB demonstrate good technical reproducibility, highlighting the complementary nature of the two filtering steps. The majority of features (∼50%) were removed during the"curve trimming" step due to an insufficient number of remaining data points after truncation. Additionally, 885 features were excluded by MSlineaR because the sample intensities fell outside the appropriate measured ALR, i.e., 70% of all sample groups were outside this range (one example in Figure S-1), leading to a final filtered data set of 3,057 retained features across all three batches. The original unfiltered data set had a median RSD_{QC} of 24.6%. Standard processing and filtering steps improved this by 70% to a median RSD_{QC} of 7.2%. Applying MSlineaR as an additional filtering step further improved the data quality, reducing the median RSD_{QC} by 26% to a final median RSD_{QC} of 5.3% (Figure 4B). The frequency histogram in Figure 5C demonstrates that applying MSlineaR increased the proportion of features with low RSD values (between 0 and 5), compared to the data set processed without MSlineaR.

Using MSlineaR Improves Group Separation and Removes False Positives

To assess the overall impact of MSlineaR on data interpretation, the additional effects of MSlineaR on standard processing were examined by using principal component analysis (PCA) on features present in all three batches and in all samples. Using RSD filtering alone, the final data set consisted of 3,204 features, whereas after filtering and applying MSlineaR, 1,388 features remained. After applying MSlineaR, the variation explained by the first two principal components increased by ∼6.7% (Figure 5D and E). Prior to MSlineaR, the major source of separation was associated with extraction method MeOH9, which is considerably more dilute than the other methods applied. MSlineaR further enhanced this effect. In addition, MSlineaR also allowed the visualization of the small variation due to the addition of ethanol to the sample extraction mix (MeOH_EtOH3). MeOH1.5 extracts clustered more closely with the MeOH3 and the QC after applying MSlineaR, which is chemically plausible since their only difference lies in the plasma-to-solvent ratio used.

To understand why the behavior of MeOH1.5 samples altered after employing MSlineaR, we examined the 10 strongest contributing features to principal component 2 prior to MSlineaR (4 examples are shown in Figure S-2). All of them exhibited a distinct separation for MeOH1.5, but no consistent dilution-dependent response behavior. This indicates that, without MSlineaR, the PCA-based interpretation of group differences would be driven by inconsistent dilution-dependent signals, potentially leading to misleading and scientifically unreliable conclusions. This suggests that good separation with PCA may be highly misleading in terms of the robustness of results. Interestingly, most of these features displayed a dilution-dependent response with a negative slope, which could imply that they are contaminants from the extraction process, solvents, or signals affected by ion-suppression effects.

To further evaluate the data structure, we calculated intragroup and intergroup variation, as well as the median RSD (Figure 5F and G, Figure S-3). For all groups, the intragroup variation decreased and the intergroup variation increased following MSlineaR filtering, demonstrating MSlineaR's potential to enhance biological interpretation. Interestingly, use of MSlineaR also reduced the median RSD of the

https://doi.org/10.1021/acs.analchem.5c04480
Anal. Chem. XXXX, XXX, XXX−XXX

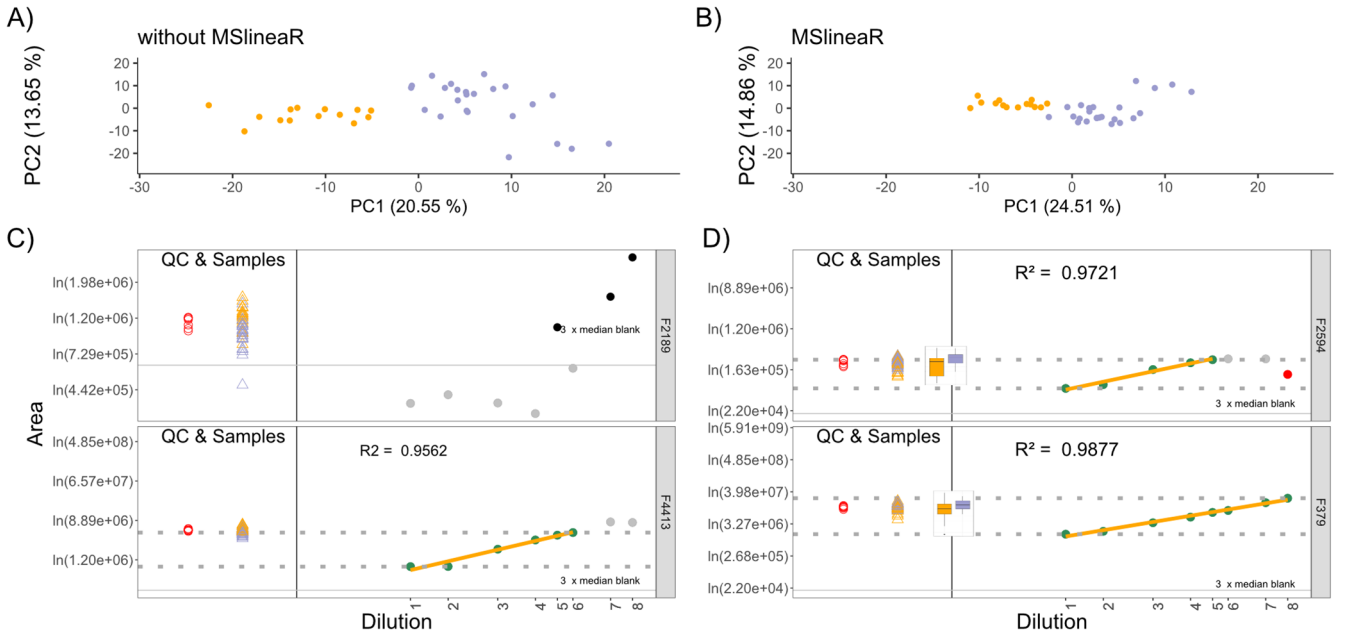


Figure 6. Principal component analysis (PCA) of untargeted metabolomics data from Data set 3 without (A) and with MSlineaR filtering (B), colored according to the different locations. (C) shows two examples of features excluded by MSlineaR: the top panel shows a feature with inconsistent dilution-dependent behavior, while the bottom panel shows a feature where most biological samples fall within the saturation range. (D) shows two examples of features that reached statistical significance (FDR-adjusted q-value ≤ 0.1) subsequent to application of MSlineaR. The dilution curve is displayed at the right side of each panel: pooled QC samples (red circles) and biological samples (triangles, colored according to location) are displayed on the left side with additional boxplots of biological sample intensities overlaid on the plot for better visualization.

samples in all groups except MeOH9. The RSD is influenced by the absolute value of the mean; i.e., for a given standard deviation, the RSD will increase as the mean intensity decreases. Since MeOH9 represents the most diluted samples, removal of noise and high-intensity inconsistent dilution-dependent contaminants by MSlineaR led to a lower average signal intensity, which likely explains the increase in median RSD.

IROA Standards Used as Negative and Positive Controls

To measure the sensitivity and specificity of MSlineaR in an untargeted data set, we required a ground truth. Therefore, each batch was spiked with IROA Internal Standard (IROA-IS), adding uniform concentrations of the U−^{13}C-labeled compounds to each sample.

For the plasma samples, 275 pairs of ^{12}C and U−^{13}C-IS could be found in the dilution curves in at least one batch, of which 210 pairs were consistently found in all three batches. The U−^{13}C-IS, being constant in concentration, are expected to have a uniform signal, whereas the ^{12}C-features are expected to show ALB. MSlineaR accurately identified all U−^{13}C-IS as non-dilution-dependent.

We conducted an additional validation experiment to assess the performance of MSlineaR under highly dilute conditions. In approximately 25% of dilution curves under the most dilute conditions, U−^{13}C-labeled standards exhibited an unexpected response characterized by an apparent signal increase upon further dilution (Figure S-4). The underlying mechanism remains unclear but may reflect moderate ion enhancement effects caused by changes in matrix composition at extremely low concentrations. Notably, this behavior was only observed under highly dilute conditions and was absent in dilution series generated using standard operating concentrations. This phenomenon may be particularly relevant for applications involving highly dilute matrices, such as water quality analysis. Where necessary, additional filtering based on local gradient strength could be applied to exclude such features.

Among the ^{12}C features, 142 (52%) passed all standard quality filters without using MSlineaR but 112 (79%) were classified as exhibiting ALB with MSlineaR. Two researchers manually analyzed all 142 ^{12}C-features by visual inspection of their dilution curves across the three batches to determine MSlineaR's accuracy of classification (Figure 5H). In total, two features (1.4%) from the ^{12}C-isotopes were classified differently by MSlineaR compared to manual classification. Overall, when considering both U−^{13}C and ^{12}C-isotopes, we achieved a classification accuracy of 99.9%. This results in a total of 30 (21%) ^{12}C-features, which were excluded following MSlineaR-based filtering, either because they did not show ALB across the dilution range or because this behavior was limited to a single batch. All ^{12}C-features are displayed in Figure S-5.

R^2 was used as the primary filtering metric. Additional metrics for assessing dilution-dependent behavior were subsequently implemented in the software package and benchmarked against R^2. The ^{12}C-IROA isotopes were used as true positives in this test and the proportion of correctly retained compounds was calculated for each metric. In this comparison, R^2 retained the highest percentage of true positives. Approximately 97% of the retained isotopes additionally exhibited monotonic behavior within the trimmed range. Only 11 curves with high R^2 values were not classified as monotonic due to local deviations within the dilution pattern. Two representative examples are shown in Text S-4, illustrating that these curves nevertheless exhibited clear dilution-dependent signal behavior and were therefore retained when using R^2 as the primary filtering metric. Assessing false positives in this benchmarking context is more challenging, yet potentially less relevant for untargeted methods where results normally require independent validation.

https://doi.org/10.1021/acs.analchem.5c04480
Anal. Chem. XXXX, XXX, XXX−XXX

To determine batch-to-batch variability, the number of features exhibiting ALB included in either (i) all batches, (ii) two batches, or (iii) one batch was determined for the IROA standards (Figure S-6). Of the 275 possible ^{12}C-features present, 126 were found to show ALB with an RSD below 20% in at least one batch. Among them, 94.5% were detected in at least two batches, and 5.5% in only one batch. Four main reasons were identified to explain the variability in feature retention across batches (Figure S-7): (1) there are fewer than five dilution points in a batch, (2) the noise/contamination level was different between batches (Figure S-7A), resulting in more stringent filtering, (3) the feature shows ALB in one batch but this is not reproduced across all batches due to either (i) R^2 close to our cutoff points or (ii) too few retained points, e.g., due to reaching saturation or (iii) some causes remain unclear (Figure S-7B), and (4) the feature ionizes well and has a steep slope, leading to saturation of the detector at relatively low concentrations. Where the samples are also at saturation level, data are not reliable for statistical analysis (Figure S-7C).

Benchmarking Using a Standard Linear Algorithm Protocol

The unprocessed data set (22,767 features) was further evaluated by comparing the results of fitting a linear regression model to the untrimmed and the trimmed data.

Both approaches agreed on the ALB classification of approximately 94% (21,314) of features, with 52% classified as exhibiting ALB. For 46% of the features, both methods reported the same R^2 values and identified comparable ranges of ALB. However, MSlineaR shortened the range of ALB for 11% of the features (2,364), resulting in an average R^2 of 0.97 compared to 0.93 for the standard model. In total, 1,453 (6%) dilution series were classified differently by MSlineaR throughout all three batches. Among them, 7% (111 features) were classified as exhibiting ALB by the conventional approach but inconsistent with ALB by MSlineaR. The vast majority of this discordance was due to insufficient dilution points remaining after MSlineaR truncated unstable ends of the curve. In contrast, 29% (421 features) were discarded by the standard method but were retained by MSlineaR, illustrating its ability to recover valid features that would otherwise be lost.

Using MSlineaR on a Biological Untargeted Data Set: DS3

Finally, to demonstrate the application of MSlineaR in a real-world biological data set, we used a subset of unpublished data from a longitudinal study involving horses sampled across different locations. In this test case, we assessed the impact of MSlineaR on the detection of metabolomic differences between groups of horses sampled from two different locations. Using conventional quality-based filtering alone, the data set was reduced from 10,900 to 2,717 features (25%). Applying MSlineaR as an additional filtering step further reduced the data set to 1,846 retained features (17%) showing consistent ALB (871 features removed). Principal component analysis (Figure 6A and B) showed improved clustering structure and increased variance explained by PC1 and PC2 following MSlineaR filtering. Intragroup variance was reduced by 1% and 5% per group, while intergroup variance increased by ∼65%, indicating improved data structure after removal of signals outside of the range of ALB.

Using Compound Discoverer, 139 out of 2,717 features (5.1%) in the data set without MSlineaR could be annotated, compared to 118 out of 1,846 features (6.4%) with MSlineaR, indicating a slightly higher percentage identification rate. To evaluate the impact on downstream statistics, metabolomic differences between the two classes were assessed using a Mann−Whitney U test of all annotated features. A total of 39% of annotated features without MSlineaR had FDR-adjusted q-values below 0.1, compared to 44% with MSlineaR. Figure 6C illustrates two features that showed apparent group separation but were excluded by MSlineaR: one due to inconsistent dilution-dependent behavior (F2189), and the other (F4413) displays ALB but the majority of the sample values were beyond the range of ALB (saturation). Conversely, Figure 6D shows two features that previously did not survive false discovery rate testing (FDR-adjusted $q > 0.1$), but improved accuracy in filtering with MSlineaR. Together with a reduction in the overall number of statistical tests this resulted in lower q-values meeting the threshold for statistical significance.

Based on three data sets, we have demonstrated that MSlineaR enhances data quality and reproducibility, resulting in more robust and reliable outcomes. The benchmark data sets comprised between 66 and ∼60,000 dilution curves and were processed within 5 to 40 min on a standard desktop PC.

We have chosen to use linear modeling primarily because nonlinear modeling requires more extensive and complex validation. For simplicity, R^2 was adopted as our main filtering metric and MSlineaR was benchmarked on this metric, although the software has since been developed to include other filtering parameters that can be used for more comprehensive data filtering. Although a high R^2 alone does not guarantee linearity, a low R^2 suggests that a linear model does not fully explain the variation in the data. This metric is thus employed to measure how precisely a linear model describes the curated dilution range, i.e., after removal of nonlinear regions (e.g., plateau effects, negative gradients, and outliers), and provides a first-pass filter.

Compared to higher-order polynomial or nonlinear models, linear models facilitate the calculation of uncertainty and goodness of fit[29] and require fewer data points.[30] They also simplify calculation of key parameters including limit of detection and quantification boundaries. In theory, a linear calibration model is also more reproducible, enabling better comparison of results across different instruments or laboratories.

By contrast, many existing calibration models rely on polynomial equations. While these may extend the working range and will typically have a better fit than linear models, there is a high risk of overfitting the model and such models cannot distinguish between signals within the range of ALB and those already in saturation or noise. Thus, they typically require more validation to prove accuracy and robustness.[31] Future developments will explore the integration of polynomial models, provided this can be achieved without compromising reproducibility or quantitative reliability.

MSlineaR effectively removes unreliable and nonrobust features from the data set, and it complements other existing filtering methods. We demonstrated that following a conventional/standard RSD-based filter, about 30% of the remaining features still lacked consistency with ALB across the dilution range, resulting in a large number of statistically unreliable features. Applying MSlineaR as an additional filtering step resulted in the reduction of noninformative and biologically irrelevant features, indicating that the two filtering methods are complementary.

Although this was not evaluated in the present study, MSlineaR also offers potential for optimizing sample dilution strategies. A dilution series of pooled QCs can be analyzed (with

https://doi.org/10.1021/acs.analchem.5c04480
Anal. Chem. XXXX, XXX, XXX−XXX

replicates) and then processed with MSlineaR to identify the maximum number of reliably measurable signals and thereby establish the optimal dilution prior to running the real samples. A modified approach can also be used to objectively compare different methods or instrument performances based on the number of features detected in the linear range. Combining such an approach with adduct, isotope and modification identification software is predicted to enable a realistic estimate of the number of unique metabolites quantifiable for any individual method or instrument.

Although we demonstrate only plasma as a sample matrix here, other matrices can be used if the samples are at a comparable overall concentration level when chemically analyzed, if necessary, using preacquisition dilution to achieve this.[32] This approach helps to avoid or compensate for ion suppression effects between samples arising from concentration differences. We would recommend a pilot run incorporating different dilutions of the pooled QC to determine the optimal concentration range to use for all samples, including the pooled QCs.

Cross-checking the graphical output allows identification of cases where features were rejected for not showing ALB across the dilution range because their steep slopes did not meet the required data-point threshold before they reached saturation. In such cases, MSlineaR can highlight the need for additional sample dilution or a different method to ensure reliable quantification of certain features. The graphical output can also be used to flag systematic issues, such as anomalies with a single dilution sample. Any challenges presented by this can be mitigated by ensuring biological results are robustly validated and internal standards are evaluated for unexpected dilution-dependent behavior as part of the standard pipeline.

In the present study, MSlineaR was applied in a stringent manner, which may result in the exclusion of a small number of real metabolites. However, this conservative approach ensures retention of only high-quality, quantitative features, likely to result in improved experimental reproducibility. User-defined parameters allow relaxation of stringency for individual experimental needs, albeit with a higher risk of false positive results.

How to Implement MSlineaR

Pooled QC and Number of Dilution Points. In conventional studies, dilution curves are typically generated from a range of concentrations of a pooled QC sample, with the highest concentration set at three times that of the undiluted pooled QC (Table S-8). We make an initial assumption that no sample is likely to have a concentration of over three times the mean value and thus deal with these cases by defining the ALR below this subjective cutoff line. This applies equally to signals within the ALR and those outside it.

MSlineaR requires a user-defined minimum number of data points to assess dilution-dependency. In our studies, we used at least five dilution points, which is in accordance with the ICH and USP guidelines[33,34] and the minimum we recommend. More dilution points decrease the probability of features being excluded due to not reaching the threshold for minimum number of linear data points. Nine data points are a typical starting point in our workflows. The optimal number, spacing of points, and maximum concentration will ultimately be dependent on the dynamic range of the specific matrix under analysis. Accordingly, the recommended dilution strategy should be considered context-specific and may need to be adapted for different experimental systems.

Sample Design. Experimental design is key to the success of any metabolomics experiment, and the success of MSlineaR is reliant on appropriate selection of concentration or dilution ranges for the initial dilution curve. This can only be determined empirically or based on the experience of working with a particular matrix. Ideally, the dilution curves would be analyzed throughout the run in both random order and spacing, but the reality of working with the higher concentrations necessitates keeping the dilution curves separate from the samples while still running in the same sequence. Typically, they are analyzed just once at the beginning of a sequence, but for more robust analysis, analyzing them at the beginning, middle and end of the sequence allows the intrabatch measurement drift to be measured. Analytical factors, including carryover, column equilibration and total batch length, become important considerations in such studies. In general, we tend to evaluate degradation over the batch length and attempt to solve this issue by restricting batches to a defined number of samples or analysis times based on objective measured evidence of intrabatch drift for a given method and matrix.

It is matrix-dependent, but at least two QC injections should be analyzed after the highest 3-fold concentrated pooled QC to evaluate potential carryover effects and equilibrate the instrument for the real samples. Where specific matrices or methodologies require more than one QC to account for carryover, adjusting the dilutions and washing solvents to reduce the initial carryover is preferable to more stabilization samples where possible.

Implementing into Preprocessing Pipeline. We have thus far only tested MSlineaR prior to major data correction processes. Using it prior to any drift correction enables it to assess ALB more reliably based on the true detector performance limitations. Additionally, there is a subsequent reduction in computational cost and processing time due to a smaller data set. Further exploration and validation are required to see how it performs postcorrection when strong drift and batch effects are present.

Choice of Metrics for Consistent Dilution-Dependent Response. Different metrics reflect complementary aspects of the dilution response and have inherent limitations.[35] For example, Spearman's ρ is sensitive to local violations of monotonicity and may decrease substantially due to a single deviating point, even if the overall trend is well behaved. In contrast, such isolated deviations typically have a limited impact on R^2, which reflects the overall consistency of the signal.

To provide a balanced and robust assessment, we therefore use R^2 as the suggested initial metric. R^2 cutoff values for filtering are selected by the user. These are best selected after examining the distribution of the R^2 values across the data set; R^2 of 0.99 is conventional but can be too strict to use in complex matrices. In particular, we have observed matrix enhancement at particularly low dilution levels (Figure S-4) which may justify relaxing R^2 values if working across a wide dynamic range. Based on our evaluation, R^2 of at least 0.9 maintained a good balance between retaining robust genuine signals and removing most noise peaks.

In addition, we have complemented this metric with additional measures such as Spearman's ρ, slope, and residual-based measures. Importantly, these metrics are implemented within a flexible framework, allowing users to adjust the stringency of filtering depending on the data set and analytical objective.

https://doi.org/10.1021/acs.analchem.5c04480
Anal. Chem. XXXX, XXX, XXX−XXX

While graphical diagnostics such as residual and Q−Q plots are provided, manual inspection of every peak is not required, ensuring scalability to large untargeted data sets.

For targeted analyses, where absolute quantification is required, we apply stricter criteria, including assessment of proportionality (slope) and deviation from the fitted line, consistent with CLSI EP06 recommendations.[7]

CONCLUSION AND OUTLOOK

In this paper, we introduce MSlineaR, an easy-to-use R software tool that improves statistical robustness and quality assurance in data analysis. It filters out features that show poor consistency with ALB across the dilution range, including noise, at an early stage in the data cleanup. It reduces the risk of false positive statistical findings, "rescues" features with outliers or restricted ranges of ALB and reduces computational demands for downstream analysis while improving the percentage of statistically relevant features. It can be integrated into an existing preprocessing pipeline or be used as a standalone approach. It can be used to improve data set robustness and reproducibility, support the selection of appropriate dilution levels, and facilitate objective comparison of different methods or instruments in untargeted data analysis.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.5c04480.

Texts S-1−S-4, Tables S-1−S-10, and Figures S-1−S-4 and S-6−S-7: detailed analytical workflows for DS1−DS3; evaluation of linearity metrics; MSlineaR templates, parameters, mathematical formulation, and assessed metrics; analytical sequences and dilution-series preparation; additional results on feature filtering, RSDs, ion behavior, and IROA standards (PDF)

Figure S-5: Assessment of the dilution-dependent signal behavior of 142 IROA^{12}C features from DS2 across three batches, showing dilution curves and signal variation across sample types and statistical classes for each feature and batch (PDF)

AUTHOR INFORMATION

Corresponding Authors

Jennifer A. Kirwan − *Berlin Institute of Health at Charité Universitätsmedizin, BIH Metabolomics, Berlin 10117, Germany; Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin 13125, Germany; Metabolomics Unit, University of Veterinary Medicine Vienna, Vienna 1210, Austria;* orcid.org/0000-0002-5423-1651; Email: Jennifer.kirwan@vetmeduni.ac.at

Janine Wiebach − *Berlin Institute of Health at Charité Universitätsmedizin, BIH Metabolomics, Berlin 10117, Germany; Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin 13125, Germany;* orcid.org/0000-0002-7416-9859; Email: Janine.wiebach@bih-charite.de

Authors

Álvaro Fernández-Ochoa − *Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin 13125, Germany; Department of Analytical Chemistry, University of Granada, Granada 18071, Spain;* orcid.org/0000-0001-5547-3209

Ulrike Bruning − *Berlin Institute of Health at Charité Universitätsmedizin, BIH Metabolomics, Berlin 10117, Germany; Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin 13125, Germany;* orcid.org/0000-0003-0112-4032

Jochen Kruppa-Scheetz − *Hochschule Osnabrück, University of Applied Sciences, Osnabrück 49076, Germany*

Maëlle Bonhomme − *University of Liège, Faculty of Veterinary Medicine, Department of Functional Sciences, Pharmacology and Toxicology, Comparative Veterinary Medicine, FARAH, Liège 4000, Belgium;* orcid.org/0009-0003-9884-848X

Dominique-Marie Votion − *University of Liège, Faculty of Veterinary Medicine, Department of Functional Sciences, Pharmacology and Toxicology, Comparative Veterinary Medicine, FARAH, Liège 4000, Belgium;* orcid.org/0000-0003-2899-8207

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.5c04480

Author Contributions

J.W., A.F.-O., and J.A.K. conceived and planned the study. A.F.-O., U.B., and M.B. performed analytical measurements. J.W. processed and analyzed the data and performed all calculations. J.W. and J.A.K. wrote the manuscript with contributions from all authors. All authors have given approval to the final version of the manuscript.

Funding

Work conducted by the BIH Metabolomics Lab was partially funded by the German Federal Ministry of Education and Research (BMBF) , grant number 03LW0239K, through the project titled "Multimodal Clinical Mass Spectrometry to Target Treatment Resistance (MSTARS). Open access funding provided by Veterinarmedizinische Universitat Wien.

Notes

ChatGPT version 4.0 was used for minor editing of individual paragraphs to improve wording.
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge the assistance of Maureen Kirwan in proofreading the manuscript.

ABBREVIATION

QC, quality control sample; R^2, coefficient of determination; RSD, relative standard deviation; D-ratio, dispersion-ratio; PCA, principal component analysis; ALB, approximately linear behavior; ALR, approximately linear range

REFERENCES

(1) Broadhurst, D.; Goodacre, R.; Reinke, S. N.; Kuligowski, J.; Wilson, I. D.; Lewis, M. R.; Dunn, W. B. Guidelines and Considerations for the Use of System Suitability and Quality Control Samples in Mass Spectrometry Assays Applied in Untargeted Clinical Metabolomic Studies. *Metabolomics* **2018**, *14* (6), 72.

(2) Broeckling, C. D.; Beger, R. D.; Cheng, L. L.; Cumeras, R.; Cuthbertson, D. J.; Dasari, S.; Davis, W. C.; Dunn, W. B.; Evans, A. M.; Fernández-Ochoa, A.; Gika, H.; Goodacre, R.; Goodman, K. D.; Gouveia, G. J.; Hsu, P.-C.; Kirwan, J. A.; Kodra, D.; Kuligowski, J.; Lan, R. S.-L.; Monge, M. E.; Moussa, L. W.; Nair, S. G.; Reisdorph, N.; Sherrod, S. D.; Ulmer Holland, C.; Vuckovic, D.; Yu, L.-R.; Zhang, B.;

Theodoridis, G.; Mosley, J. D. Current Practices in LC-MS Untargeted Metabolomics: A Scoping Review on the Use of Pooled Quality Control Samples. *Anal. Chem.* **2023**, *95* (51), 18645−18654.
(3) Klåvus, A.; Kokla, M.; Noerman, S.; Koistinen, V. M.; Tuomainen, M.; Zarei, I.; Meuronen, T.; Häkkinen, M. R.; Rummukainen, S.; Farizah Babu, A.; Sallinen, T.; Kärkkäinen, O.; Paananen, J.; Broadhurst, D.; Brunius, C.; Hanhineva, K. "Notame": Workflow for Non-Targeted LC−MS Metabolic Profiling. *Metabolites* **2020**, *10* (4), 135.
(4) Croixmarie, V.; Umbdenstock, T.; Cloarec, O.; Moreau, A.; Pascussi, J.-M.; Boursier-Neyret, C.; Walther, B. Integrated Comparison of Drug-Related and Drug-Induced Ultra Performance Liquid Chromatography/Mass Spectrometry Metabonomic Profiles Using Human Hepatocyte Cultures. *Anal. Chem.* **2009**, *81* (15), 6061−6069.
(5) Sands, C. J.; Gómez-Romero, M.; Correia, G.; Chekmeneva, E.; Camuzeaux, S.; Izzi-Engbeaya, C.; Dhillo, W. S.; Takats, Z.; Lewis, M. R. Representing the Metabolome with High Fidelity: Range and Response as Quality Control Factors in LC-MS-Based Global Profiling. *Anal. Chem.* **2021**, *93* (4), 1924−1933.
(6) Yu, H.; Xing, S.; Nierves, L.; Lange, P. F.; Huan, T. Fold-Change Compression: An Unexplored But Correctable Quantitative Bias Caused by Nonlinear Electrospray Ionization Responses in Untargeted Metabolomics. *Anal. Chem.* **2020**, *92* (10), 7011−7019.
(7) CLSI. *Evaluation of Linearity of Quantitative Measurement Procedures; CLSI guideline EP06*, 2nd ed.; Clinical and Laboratory Standards Institute: 2020.
(8) Sands, C. J.; Wolfer, A. M.; Correia, G. D. S.; Sadawi, N.; Ahmed, A.; Jiménez, B.; Lewis, M. R.; Glen, R. C.; Nicholson, J. K.; Pearce, J. T. M. The NPYc-Toolbox, a Python Module for the Pre-Processing, Quality-Control and Analysis of Metabolic Profiling Datasets. *Bioinformatics* **2019**, *35* (24), 5359−5360.
(9) Guo, J.; Yu, H.; Xing, S.; Huan, T. Addressing Big Data Challenges in Mass Spectrometry-Based Metabolomics. *Chem. Commun.* **2022**, *58* (72), 9979−9990.
(10) Duan, L.; Molnár, I.; Snyder, J. H.; Shen, G.; Qi, X. Discrimination and Quantification of True Biological Signals in Metabolomics Analysis Based on Liquid Chromatography-Mass Spectrometry. *Mol. Plant* **2016**, *9* (8), 1217−1220.
(11) Yu, H.; Huan, T. Patterned Signal Ratio Biases in Mass Spectrometry-Based Quantitative Metabolomics. *Anal. Chem.* **2021**, *93* (4), 2254−2262.
(12) Jankevics, A.; Merlo, M. E.; de Vries, M.; Vonk, R. J.; Takano, E.; Breitling, R. Separating the wheat from the chaff: a prioritisation pipeline for the analysis of metabolomics datasets. *Metabolomics* **2012**, *8* (S1), 29−36.
(13) Crawford Scientific *Hidden Problems in your LCMS data?*; Crawford Scientific, 2026. https://www.crawfordscientific.com/chromatography-blog/post/hidden-problems-in-your-lcms-data. accessed 26 September 2024.
(14) Findlay, J. W. A.; Dillard, R. F. Appropriate Calibration Curve Fitting in Ligand Binding Assays. *AAPS J.* **2007**, *9* (2), 29.
(15) Holle, J.; Bartolomaeus, H.; Löber, U.; Behrens, F.; Bartolomaeus, T. U. P.; Anandakumar, H.; Wimmer, M. I.; Vu, D. L.; Kuhring, M.; Brüning, U.; Maifeld, A.; Geisberger, S.; Kempa, S.; Schumacher, F.; Kleuser, B.; Bufler, P.; Querfeld, U.; Kitschke, S.; Engler, D.; Kuhrt, L. D.; Drechsel, O.; Eckardt, K.-U.; Forslund, S. K.; Thürmer, A.; McParland, V.; Kirwan, J. A.; Wilck, N.; Müller, D. Inflammation in Children with CKD Linked to Gut Dysbiosis and Metabolite Imbalance. *J. Am. Soc. Nephrol.* **2022**, *33* (12), 2259−2275.
(16) De Paepe, E.; Van Meulebroek, L.; Rombouts, C.; Huysman, S.; Verplanken, K.; Lapauw, B.; Wauters, J.; Hemeryck, L. Y.; Vanhaecke, L. A Validated Multi-Matrix Platform for Metabolomic Fingerprinting of Human Urine, Feces and Plasma Using Ultra-High Performance Liquid-Chromatography Coupled to Hybrid Orbitrap High-Resolution Mass Spectrometry. *Anal. Chim. Acta* **2018**, *1033*, 108−118.
(17) Kouassi Nzoughet, J.; Bocca, C.; Simard, G.; Prunier-Mirebeau, D.; Chao de la Barca, J. M.; Bonneau, D.; Procaccio, V.; Prunier, F.; Lenaers, G.; Reynier, P. A Nontargeted UHPLC-HRMS Metabolomics Pipeline for Metabolite Identification: Application to Cardiac Remote Ischemic Preconditioning. *Anal. Chem.* **2017**, *89* (3), 2138−2146.
(18) Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R. Procedures for Large-Scale Metabolic Profiling of Serum and Plasma Using Gas Chromatography and Liquid Chromatography Coupled to Mass Spectrometry. *Nat. Protoc.* **2011**, *6* (7), 1060−1083.
(19) Kirwan, J. A.; Broadhurst, D. I.; Davidson, R. L.; Viant, M. R. Characterising and Correcting Batch Variation in an Automated Direct Infusion Mass Spectrometry (DIMS) Metabolomics Workflow. *Anal. Bioanal. Chem.* **2013**, *405* (15), 5147−5157.
(20) Tautenhahn, R.; Boettcher, C.; Neumann, S. Highly Sensitive Feature Detection for High Resolution LC/MS. *BMC Bioinf.* **2008**, *9* (1), 504.
(21) Benton, H. P.; Want, E. J.; Ebbels, T. M. D. Correction of Mass Calibration Gaps in Liquid Chromatography-Mass Spectrometry Metabolomics Data. *Bioinformatics* **2010**, *26*, 2488.
(22) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **2006**, *78* (3), 779−787.
(23) Kuhl, C.; Tautenhahn, R.; Böttcher, C.; Larson, T. R.; Neumann, S. CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of LC/MS Data Sets. *Anal. Chem.* **2012**, *84* (1), 283−289.
(24) Gowda, H.; Ivanisevic, J.; Johnson, C. H.; Kurczy, M. E.; Benton, H. P.; Rinehart, D.; Nguyen, T.; Ray, J.; Kuehl, J.; Arevalo, B.; Westenskow, P. D.; Wang, J.; Arkin, A. P.; Deutschbauer, A. M.; Patti, G. J.; Siuzdak, G. Interactive XCMS Online: Simplifying Advanced Metabolomic Data Processing and Subsequent Statistical Analyses. *Anal. Chem.* **2014**, *86* (14), 6931−6939.
(25) Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Anal. Chem.* **2012**, *84* (11), 5035−5039.
(26) Jankevics, A.; Lloyd, G. R.; Weber, R. J. M. *Pmp: peak Matrix Processing and Signal Batch Correction for Metabolomics Datasets*; R package: Birmingham, England, 2021.
(27) Parsons, H. M.; Ludwig, C.; Günther, U. L.; Viant, M. R. Improved Classification Accuracy in 1- and 2-Dimensional NMR Metabolomics Data Using the Variance Stabilising Generalised Logarithm Transformation. *BMC Bioinf.* **2007**, *8* (1), 234.
(28) Fitrianto, A.; Rana, S.; Midi, H.; Hydara, K. Effects of a Single Outlier on the Coefficient of Determination: An Empirical Study. *AIP Conf Proc.* **2015**, *1643* (1), 409−413.
(29) Cheng, W. L.; Markus, C.; Lim, C. Y.; Tan, R. Z.; Sethi, S. K.; Loh, T. P. Calibration Practices in Clinical Mass Spectrometry: Review and Recommendations. *Ann. Lab. Med.* **2023**, *43* (1), 5−18.
(30) Guthrie, W. F. *NIST/SEMATECH e-Handbook of Statistical Methods (NIST Handbook 151)*; NIST, 2020. DOI: .
(31) Frisbie, S. H.; Mitchell, E. J.; Sikora, K. R.; Abualrub, M. S.; Abosalem, Y. Using Polynomial Regression to Objectively Test the Fit of Calibration Curves in Analytical Chemistry. *Int. J. Appl. Math. Theor. Phys.* **2016**, *1* (2), 14−18.
(32) Edmands, W. M. B.; Ferrari, P.; Scalbert, A. Normalization to Specific Gravity Prior to Analysis Improves Information Recovery from High Resolution Mass Spectrometry Metabolomic Profiles of Human Urine. *Anal. Chem.* **2014**, *86* (21), 10925−10931.
(33) ICH *Validation of Analytical Procedures Q2(R2); Q2(R2)*; International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, 2023. https://www.ich.org/page/quality-guidelines.
(34) U. S. Pharmacopeia *Validation of Compendial Methods*; United States Pharmacopeial Convention: Rockville, MD, 2024. http://ftp.uspbpep.com. accessed 12 November 2025.
(35) Overdahl, K. E.; Collier, J. B.; Jetten, A. M.; Jarmusch, A. K. Signal Response Evaluation Applied to Untargeted Mass Spectrometry Data to Improve Data Interpretability. *J. Am. Soc. Mass Spectrom.* **2023**, *34* (9), 1941−1948.

https://doi.org/10.1021/acs.analchem.5c04480
Anal. Chem. XXXX, XXX, XXX−XXX