An alternative for the robust assessment of the repeatability and reproducibility of analytical measurements using bivariate dispersion

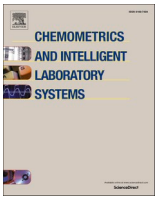
Elfried Salanon, Blandine Comte, Delphine Centeno, Stéphanie Durand, Estelle Pujos-Guillot, Julien Boccard

Chemometrics and Intelligent Laboratory Systems 250 (2024) 105148

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics



An alternative for the robust assessment of the repeatability and reproducibility of analytical measurements using bivariate dispersion

Elfried Salanon [a,*], Blandine Comte [a], Delphine Centeno [a], Stéphanie Durand [a], Estelle Pujos-Guillot [a], Julien Boccard [b]

[a] Université Clermont Auvergne, INRAE, UNH, Plateforme d'Exploration du Métabolisme, MetaboHUB Clermont, Clermont-Ferrand, France
[b] School of Pharmaceutical Sciences, University of Geneva, Geneva, Switzerland

ARTICLE INFO

Keywords:
Reproducibility
Intermediate precision
LC/MS
Dispersion
Drift

ABSTRACT

Introduction: Assessing repeatability and reproducibility in analytical chemistry is commonly based on parametric dispersion indicators, such as relative standard deviation and standard deviation, calculated for each detected variable using repeated measurements of Quality Control (QC) samples collected throughout the data acquisition sequence. However, their reliability strongly relies on the assumption of normality distribution. Knowing that analytical variability is conditional to many sources, the use of such parametric estimators is not always suitable. There is therefore a need for robust indicators of data quality independent of central values and any parametric assumption.

Methods: Three specific indicators were developed: (i) intra-group dispersion, based on the median area of the convex hull of QC samples within an analytical batch; (ii) inter-group dispersion, defined as the gradient of the deviation between analytical batches; and (iii) dispersion index. Mathematical properties of these indicators, including positivity, stability, and translation invariance, were then evaluated using synthetic data under normal and non-normal distributions. Finally, the relevance of these indicators and the associated visualization methods were highlighted based on a metabolomics case study involving liquid chromatography coupled to mass spectrometry measurements of the NIST SRM1950 reference material analyzed over more than one year within different projects.

Results: The proposed indicators were shown to be translation invariant and always positive, while first investigations performed on synthetic data revealed a high stability for multiplication. Moreover, their application to experimental data revealed specific behaviors depending on the characteristics of the signal associated with the different detected analytes, showing their ability to capture the variability observed either in parametric or non-parametric conditions. Moreover, this investigation showed different structures of sensitivity to analytical variability all along the data processing steps. The proposed indicators also allowed a visualization of the analytical drift in two dimensions, to facilitate result interpretation.

Conclusion: These indicators open the way to a better and more robust assessment of repeatability and reproducibility but also to improvements of long-term data comparability involving suitability testing.

1. Introduction

In analytical chemistry, the measurement error of any analyte present in a sample can be divided into two components, namely random and systematic error [1]. In many application fields, the multiple sources of random error can be reduced through good laboratory practices, to the point that they are considered negligible in comparison to the biological variability. Systematic errors, generally caused by drifting analytical systems, are usually evaluated using multiple measurements of samples, but cannot be determined with certainty without internal standards or certified reference materials. Analytical methods that do not involve absolute quantification, such as mass spectrometry-based (MS) untargeted metabolomics, may therefore suffer from this limitation [2].

Evaluating the variability of a measurement method is one of the key performance criteria required in analytical chemistry. In untargeted

* Corresponding author. INRAE, UNH, Plateforme d'Exploration du Métabolisme, MetaboHUB Clermont, F63122, Saint-Genès Champanelle, France.
E-mail address: elfried.salanon@inrae.fr (E. Salanon).

https://doi.org/10.1016/j.chemolab.2024.105148
Received 7 March 2024; Received in revised form 2 May 2024; Accepted 16 May 2024
Available online 18 May 2024

metabolomics, recent guidelines were published to provide researchers with a framework to describe quality assessment and quality control procedures [3,4]. In routine practice, pool quality control (QC) samples are generally analyzed along the data acquisition sequence (*also known as batch in some domains*) to evaluate the system performance and assess within-study data quality.

Many different approaches have been developed to evaluate the performance of liquid chromatography coupled to mass spectrometry (LC/MS) methods, but they have not yet been standardized [2,5]. From a formal point of view, any measurement procedure can be described as having either a constant absolute or a relative variation in the measuring interval. When the assessment is related to the measurement done over a short time period in the same experimental conditions (defined as within one analytical batch), the variation is related to the method's 'repeatability'. The 'intermediate precision' reflects the variability observed between different days of analysis using the same setting, while 'reproducibility' is related to differences observed in different measurement conditions. The latter can be defined in the present case as occurring between analytical batches in the present case. The definitions of these concepts used in the present work, are based on the ones proposed by *Kirwan* et al. [4].

Assessing repeatability, intermediate precision and reproducibility is commonly based on standard deviation (SD) or on relative standard deviation (RSD) following the principle of the variance components. Although these dispersion indicators are well-known and widely used, they rely on a strong assumption of Gaussian distribution which is not always fulfilled in experimental conditions. Indeed, in untargeted metabolomics, linearity is evaluated using diluted QCs to filter out variables that do not follow dilution trend, using linear correlations. However, it should be noted that linear correlation cannot be used to evaluate some important characteristics of an analyte response function, such as the slope of the concentration-signal relationship. This is a further argument for the use of dispersion estimators that do not depend on the assumption of normal distribution. Consequently, a Gaussian error in analyte concentrations does not result in a Gaussian error in measured values.

In practice, RSD is often used as the variation parameter to assess repeatability and reproducibility. Whether intra-assay or inter-assay, the RSD is calculated by considering the values across multiple samples within one or several batch (es) [6]. However, this intuitive approach does not account for the structure of the variance components in the case of inter-batches RSD estimates. It has also been noted that RSD is not a relevant index of measurement variability when the number of replicates varies across samples [7]. Both SD and RSD are based on the use of the mean, which is known to lack robustness. Moreover, knowing that overall measurement variability is built from the combined variations of many sources, RSD based on repeated measurements of QC samples cannot constitute reliable estimates in every experimental situation. To overcome the limitations of the RSD, some authors proposed alternative based on central value [2]. However, knowing that the central value is also affected by the variability, this work proposed a method independent of these parameters, bringing additional information on the variability of the instrument response. In this context, the use of a bivariate dispersion indicator constitutes a relevant approach that could help to overcome the limitations linked to classical parameters. Bivariate dispersion, defined as the measure of spread or variability in two dimensions, is a fundamental concept in statistics for analyzing relationships between two variables. It quantifies the degree to which data points deviate from the central tendency along both the horizontal and vertical axes on a scatter plot. Techniques like covariance and correlation coefficients aid in quantifying bivariate dispersion, providing insights into the strength and direction of the relationship between variables. In the present study, we developed three original indicators for a robust assessment of the repeatability and reproducibility of analytical measurements.

2. Methods

2.1. Notion of convex hull

All the methods proposed here are based on the convex hull. We consider a Euclidean space defined by a plane made of the "*measurement order*" and the "*intensity*" of the analyte signal. Given a finite set $S = \{S_1, \ldots..,S_n\}$ where the coordinates of S_i are (x_i,y_i), there is a unique convex hull *CH(S) of S* in the plane defined by different repeated measurements of a QC/reference material. A subset S of the plane is called *convex* if and only if for any pair of points p, q ∈ S the line segment $\overline{pq}$ is completely contained in S. The *convex hull* CH(S) of a set S is the smallest convex set that contains S. To be more precise, it is the intersection of all convex sets that contain S [8].

2.2. Proposed indicators

For the set of *S* in the affine Euclidean plane, we assume that the measurements are done in groups and the indicative groups are represented by the variable $Z = \{Z_1 \ldots Z_h\}$ with $h > 1$.

2.2.1. Intra-group dispersion (IntraD)

The Intra-group dispersion (*IntraD*) is meant to assess the dispersion of *S*, considering that the part of dispersion carried by the *"Measurement order"* is known and linear. The observed dispersion is then linked to the "*intensity*" of the measurement. The bivariate dispersion is estimated, knowing that the variation is mostly linked to the intensity of the analytes in three steps.

2.2.1.1. Step 1: Identification of the convex hull. The convex hull of each subgroup is identified; this step is completed by the identification of *h* (with $h > 1$) convex hulls.

2.2.1.2. Step 2: Estimation of the area of each group convex hull. We assume that each convex hull is a simple planar polygon with a *positively oriented* (counter clock wise) sequence of points $S_i=(x_i,y_i)$, $i = 1, \ldots,n$, in a Cartesian coordinate system. It is convenient to set $S_0=S_n$, $S_{n+1}=S_1$ for the simplicity of the formula *(Eq. (1))*. Then, we estimate the signed area of the convex hull of each group by using the following ***shoelace formula*** **(**Eq. (1)**)** [9]:

$$A_t = \frac{1}{2}\left(\sum_{i=0}^{n-1}\left(x_i y_{i+1} - x_{i+1} y_i\right)\right), \text{where } x_n = x_0 \text{ and } y_n = y_0 \qquad \text{(Eq.1)}$$

Remarks. If the vertices of the convex hull are ordered counterclockwise, $A_i > 0$. Otherwise, $A_i<0$.

2.2.1.3. Step 3: estimation of the median of the areas. Once the area of each convex hull is estimated, the *IntraD* is calculated as the median of the different convex hull's areas weighted by the quotient of the number of measurement points in each group and the total number of measurement points.

$$IntraD = median\left(\frac{A_t \bullet n_t}{n}\right), \text{with } t = 1, \ldots, h \text{ and } h > 1 \qquad \text{(Eq.2a)}$$

Rationale: If the deviation in intensity increases, the covered area will also increase.

A compact form of the formula can be calculated by:

$$IntraD = median\left(\frac{n_t}{2n}\left(\sum_{i=0}^{n-1}\left(x_i y_{i+1} - x_{i+1} y_i\right)\right)\right) \qquad \text{(Eq.2b)}$$

with t = 1, …, h; h > 1 and $\sum_{i=0}^{n-1}\left(x_i y_{i+1} - x_{i+1} y_i\right)$ being the area of the convex hull of each batch,

2.2.2. Inter-group dispersion (InterD)

The assessment of the InterD is meant to evaluate the gradient of dispersion from one group to another, and is assessed by (Eq. (3)):

$$\text{InterD} = \frac{\Delta_{injection\ order} * \Delta_{intensity}}{2} \text{ or InterD} = \frac{\sqrt{\left(x_{c_i} - x_{c_i'}\right)^2} * \sqrt{\left(y_{c_1} - y_{c_i'}\right)^2}}{2} \quad \text{(Eq.3)}$$

where X_c and Y_c refer to the centroid in the 'sequence' and 'intensity' dimension, respectively.

The establishment of this equation is done in three steps.

2.2.2.1. Step 1: The identification of the convex hull centroid of each batch. The convex hull centroid of the set of point $S=\{S_0, \ldots, S_n\}$ in the group j, is the point C_j (C_{xj}, C_{yj}) [10], where:

$$C_{x_j} = \frac{1}{6A} \sum_{i=0}^{n-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i), \quad \text{(Eq.4)}$$

$$\text{and } C_{y_j} = \frac{1}{6A} \sum_{i=0}^{n-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i), \quad \text{(Eq.5)}$$

With $x_n = x_0$ and $y_n = y_0$.

2.2.2.2. Step 2: Translation and projection. To estimate the gradient, we use (Eq. (6))

$$\text{InterD} = \left(\frac{\partial f(x,y)}{\partial y} + \frac{\partial f(x,y)}{\partial x}\right) \quad \text{(Eq.6)}$$

Considering the centroid of the first measurement group as the reference, the gradient of the others is estimated relatively to this group. Henceforth, the origin of the Euclidean space is the centroid of the first group. By projecting the centroid of each group on the new affine space, a triangle (c_1 c_i' c_i) is obtained in each case.

2.2.2.3. Step 3: estimation of the gradient. The estimation of the gradient is performed by evaluating the area of the triangle c_1 c_i' c_i [9].:

Applied to the triangle of interest, we have (Eq. (7)),

$$\text{InterD} = \frac{c_i' c_i * c_i' c_1}{2} \quad \text{(Eq.7)}$$

Using the Euclidean distance, we have:

$$\text{InterD} = \frac{\sqrt{\left(x_{c_i} - x_{c_i'}\right)^2 + \left(y_{c_i} - y_{c_i'}\right)^2} * \sqrt{\left(x_{c_1} - x_{c_i'}\right)^2 + \left(y_{c_1} - y_{c_i'}\right)^2}}{2} \quad \text{(Eq.8)}$$

with $p_{(D)}(c_i) = c_i'$, we have $y_{c_i} = y_{c_i'}$ and $x_{c_1} = x_{c_i'}$. Then, $\left(y_{c_i} - y_{c_i'}\right)^2 = 0$ and $\left(x_{c_1} - x_{c_i'}\right)^2 = 0$.

Henceforth,

$$\text{InterD} = \frac{\sqrt{\left(x_{c_i} - x_{c_i'}\right)^2} * \sqrt{\left(y_{c_1} - y_{c_i'}\right)^2}}{2} \quad \text{(Eq.9)}$$

$$\text{with : } \sqrt{\left(x_{c_i} - x_{c_i'}\right)^2} = \Delta_{measurement\ order} \text{ and } \sqrt{\left(y_{c_1} - y_{c_i'}\right)^2} = \Delta_{intensity}$$

we have,

$$InterD = \frac{\Delta_{measurement\ order(j)} * \Delta_{intensity(j)}}{2} \quad \text{(Eq.10)}$$

With *j the number of batch; j= 2, …, h,* and h being the number of groups.

If h =2:

$$InterD = \frac{\Delta_{measurement\ order} * \Delta_{intensity}}{2} \quad \text{(Eq.11)}$$

If h>2:

$$InterD = median\left(\frac{\Delta_{measurement\ order(j)} * \Delta_{intensity(j)}}{2}\right) \quad \text{(Eq.12)}$$

2.2.3. Dispersion index (D − Index)

The dispersion index (D-Index) aims at comparing the dispersion of both intra-group (within) and inter-group (between) variabilities. The rationale behind the computation of this parameter, is that two measurements could show different inter-group variability, while having the same intra-group variability. This indicator then captures this structure in a single value, as calculated using equation (Eq. (13)):

$$D - Index = \frac{IntraD}{1 + InterD} \quad \text{(Eq.13)}$$

2.3. Data

2.3.1. Simulated data

The assessment of the theoretical characteristics of these indicators (invariance by translation, positivity, and nullity) was first investigated based on simulated data. Two variables were simulated: one following a normal (variable Y) and another (variable X) following a dummy non-normal distribution. This was simulated for n = 30 samples within 3 groups (10/group). The histograms of the two simulated variables are presented in supplementary material 1; X = 11892.5 ± 3939.4, Y = 1955.4 ± 740.8. These variables were used to evaluate the theoretical characteristics of the developed indicators and the associated visualization method. From variables X and Y, five variables *(T, J, G, F and N)* were generated to assess these characteristics. A random value (340) was subtracted from variables X and Y to obtain variables T and J, respectively, thus mimicking constant loss of sensitivity. Constant multiplicative noise was simulated by multiplying variables X and Y by 1.5 to produce variables G and F, while variable N was obtained by multiplying variable X by 4. It should be noted that the proposed synthetic data do not reflect the complexity that can be encountered in the analysis of real-world data, but they are aiming to demonstrate some mathematical properties (invariance by translation and the stability by multiplication) of the newly proposed estimators.

2.3.2. Experimental data: NIST 1950 and experimental conditions

The different developed indicators were assessed using experimental data, as a proof-of-concept study. Metabolomics data obtained from the analysis of a human plasma reference material (SRM1950) made available to the research community by the National Institute of Standards and Technology (NIST) as Standard Reference Material® [11], were used. This material was specifically developed for the metabolomics research field, and can be used for calibration, measurement accuracy assessment and analytical method development. This material can also be used for comparison of measurement technologies used in metabolomic studies, interlaboratory analyses, and for quality assurance when assigning experimental values to in-house reference materials [11].

Metabolomic data were collected from the repeated analysis of the NIST1950 plasma sample over one year, using a mass spectrometry-based untargeted approach [12]. Deproteinized plasma extracts were injected in triplicates at the beginning of each analytical sequence of various metabolomics studies. The experimental analysis plan of the full NIST dataset is presented in Fig. 1. Raw data were processed with the XCMS R-package [13] using a Galaxy web-based platform [14], to yield a data matrix containing retention times, accurate masses and processed peak intensities. This pre-processing step included noise filtering, automatic peak detection and chromatographic alignment allowing the appropriate comparison of multiple samples by further processing

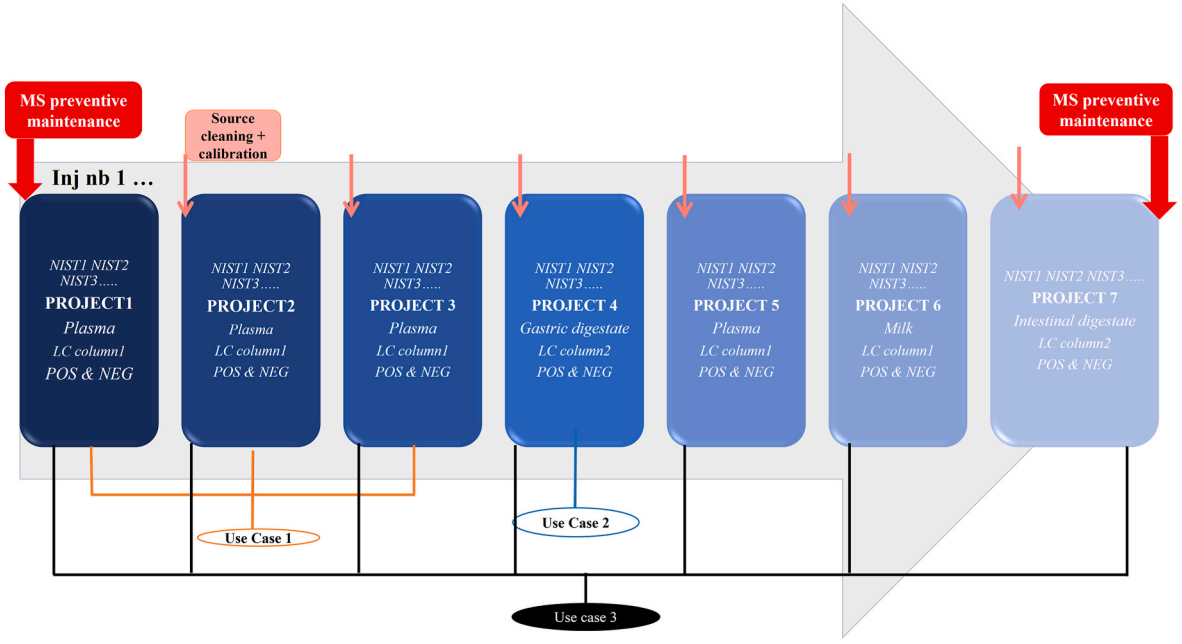


Fig. 1. Analytical sequence of the NIST in one year.

methods. Following these steps, a missing data filter was applied (threshold percentage of missing data <20 %), leading to a dataset of 930 features. From this, 43 known chemical compounds were selected to get a relevant coverage of the physicochemical diversity in terms of retention time and *m/z* values. The subset of 43 compounds was selected to include metabolites that are commonly detected and formally identified in human blood samples. Additionally, the aim was to reflect the physicochemical diversity of metabolites in term of polarity (consequently retention time) and *m/z* by including members of several families with different chemical natures, and detectable using the present LC/HRMS method. Finally, the selected compounds cover a wide range of biological absolute concentrations that can be considered reliable in the context of metabolomics analysis.

These data were used to assess three experimental setups involving different dimensions of the measurement uncertainty, namely ***repeatability, intermediate precision and within laboratory reproducibility*** [4]. The experimental conditions are presented in Fig. 1: the first use case focusses on repeatability assessment (i.e., the typical measurement variability associated with repeated measurements of the same sample within a single batch) [4] and involved NIST samples from *Projects 1, 2, and 3*. The second use case relates to intermediate precision assessment (i.e. between batches variability) and involves NIST samples from *Project 4*. The last use case addresses within-laboratory or intra-laboratory reproducibility assessment, with NIST samples from *Projects 1 to 7*. The NIST samples were analyzed in the beginning of the analytical sequence as they are used to check system suitability prior to project analysis being performed. Therefore, they reflect potential consequences on the analytical system of previous projects (LC column damage, source clogging, etc …). To illustrate the batch effects visualization method, we consequently only used the NIST samples following a plasma project, excluding then samples from *Project 2 and 5*.

2.4. Simulations and analyses

Simulations and analyses were performed with the R software using a fixed seed of "123". A heatmap was used to assess the correlations between the proposed indicators and other dispersion and central indicators: *CV or RSD, range, median, interquartile range, variance, SD, and the Fisher variance ratio*. Pearson correlation coefficients were also calculated to quantify the agreement between indicators and presented as heatmap. Principal component analysis (PCA) was then applied to visualize the distribution of the different groups of measurements.

3. Results

3.1. Properties of the indicators

3.1.1. Invariance by translation

The InterD *(Eq. (14))*, the IntraD *(Eq. (15))* and the D – Index *(Eq. (16))* remained unchanged when a constant value *b* was added to the random variable as shown in Table 1.

$$X : InterD\{S(x+b,y)\} = InterD\{S(x,y)\} \quad \text{(Eq.14)}$$

$$X : IntraD\{S(x+b,y)\} = IntraD\{S(x,y)\} \quad \text{(Eq.15)}$$

$$X : D-Index\{S(x+b,y)\} = D-Index\{S(x,y)\} \quad \text{(Eq.16)}$$

The invariance by translation is important to assess the sensitivity of measurement to a constant noise. When the same additive measurement errors affect the different measurements, the results will remain the same for the different indicators.

Some examples on the invariance by translation based on simulated data are presented in Table 1. In the following examples, a case of instrument sensitivity loss was considered with a constant/non-differential signal decrease. As expected, the behavior of variables X, Y, T and J revealed that translation did not produce any effect on the estimated indicators.

3.1.2. Positivity and nullity

The IntraD, InterD and the D-Index are all positive values. They are nulls only when the difference between the measurements is null.

Table 1
Estimation of the indicators in X and Y and their translation.

	X	T = X-340	Y	J = Y-340
Intra-batch dispersion	2667.87	2667.87	471.03	471.03
Inter-group dispersion	24451.20	24451.20	19746.56	19746.56
Dispersion Index	0.11	0.11	0.02	0.02

Table 2
Exploration of the stability by multiplication of *X* and *Y*.

	X	G = X*1.5	*F=X*4*	*Y*	*N=Y*1.5*
Mean	11892.54	17838.80	47570.15	1955.44	2933.17
Mean quotient		1.5	4		1.5
Standard deviation	3939.44	5909.15	15757.72	740.80	1111.21
Standard deviation quotient		1.5	4		1.5
Intra-group dispersion	2667.87	3999.79	10659.42	471.03	704.54
Intra-group dispersion quotient		1.5	4.0		1.5
Inter-group dispersion	24451.20	55015.20	391219.16	19746.56	44429.77
Inter-group dispersion quotient		2.25	16.00		2.25
Dispersion Index	0.11	0.07	0.03	0.02	0.02
Quotient of the Dispersion Indexs		0.67 (1/1.5)	0.25 (1/4)		0.67 (1/1.5)

3.1.2.1. Demonstration for

IntraD. $IntraD = median\left(\frac{n_t}{2n}\left(\sum_{i=0}^{n-1}(x_i y_{i+1} - x_{i+1}y_i)\right)\right)$; *with* t = 1, …,

$\forall \left(x_{c_i}; y_{c_i}\right)$; and $C_1\left(x_{c_1}; y_{c_1}\right)$; *with* x & y $\in \mathbb{R}$ $\sqrt{\left(x_{c_1} - x_{c_i'}\right)^2} \geq 0$ & $\sqrt{\left(y_{c_1} - y_{c_i'}\right)^2} \geq 0$

h; h > 1 and $\sum_{i=0}^{n-1}(x_i y_{i+1} - x_{i+1}y_i)$ being the area of the convex hull of each batch, knowing that, ∀ X ∈ [0; + ∞ [; M(X) ≥ 0.

$$\forall\, n, h;\ \sum_{i=0}^{n-1}(x_i y_{i+1} - x_{i+1}y_i) > 0 \text{ and } \frac{n_t}{2n} > 0 \Longleftrightarrow \forall\, n, h \in \mathbb{R}\ M\left(\frac{n_t}{2n}\left(\sum_{i=0}^{n-1}(x_i y_{i+1} - x_{i+1}y_i)\right)\right) > 0 \quad \text{(Eq.17)}$$

Let *S (x, y)* being the vertex of the convex hull and a series of measurements, with *x: being the injection order* and *y: the intensity of the measurement*,

$$\forall\, S, x_i = x_{i+1} = x_{i+2},\ \text{when } y_i = y_{i+1} = y_{i+2} = \ldots = y_n,\ \left(\sum_{i=0}^{n-1}(x_i y_{i+1} - x_{i+1}y_i)\right) = 0;\ \text{if } \forall\, h, \left(\sum_{i=0}^{n-1}(x_i y_{i+1} - x_{i+1}y_i)\right) = 0, \textit{then } M\left(\frac{n_t}{2n}\left(\sum_{i=0}^{n-1}(x_i y_{i+1} - x_{i+1}y_i)\right)\right) = 0 \quad \text{(Eq.18)}$$

From *(Eq. (17))* and *(Eq. (18))*, ∀*S (x, y)*, ∀ *h, n* ∈ ℝ. $M\left(\frac{n_t}{2n}\left(\sum_{i=0}^{n-1}(x_i y_{i+1} - x_{i+1}y_i)\right)\right) \geq 0$

That means that if the measured intensity (y) and the measurement order (x) are the same for all measurements, the area of the convex hull of each group is then null. This behavior is the only condition leading to a null median, knowing that the IntraD is a positive parameter.

3.1.2.2. InterD

$$InterD = \frac{\Delta_{measurement\ order} \bullet \Delta_{intensity}}{2} \text{ or } InterD = \frac{\sqrt{\left(x_{c_i} - x_{c_i'}\right)^2} \bullet \sqrt{\left(y_{c_1} - y_{c_i'}\right)^2}}{2}$$

Let *S (x, y)* being the vertex of the convex hull and series of measurements, with *x: being the injection order* and *y: the intensity of the measurement*,

$$\sqrt{\left(x_{c_1} - x_{c_i'}\right)^2} \geq 0\ \&\ \sqrt{\left(y_{c_1} - y_{c_i'}\right)^2} \geq 0 \Leftrightarrow \sqrt{\left(x_{c_1} - x_{c_i'}\right)^2} \cdot \sqrt{\left(y_{c_1} - y_{c_i'}\right)^2} \geq 0$$

$$\Leftrightarrow \frac{\sqrt{\left(x_{c_1} - x_{c_i'}\right)^2} \cdot \sqrt{\left(y_{c_1} - y_{c_i'}\right)^2}}{2} \geq 0$$

$$\Leftrightarrow InterD \geq 0$$

Moreover,

$$InterD = 0 \Longleftrightarrow \sqrt{\left(x_{c_1} - x_{c_i'}\right)^2} \cdot \sqrt{\left(y_{c_1} - y_{c_i'}\right)^2} = 0, for\ 2 \neq 0.$$

Knowing that *x* is the measurement order, ∀ *y; x* ≠ 0. We have then,

$$InterD = 0 \Longleftrightarrow \sqrt{\left(y_{c_1} - y_{c_i'}\right)^2} = 0\ ;\ for\ \sqrt{\left(x_{c_1} - x_{c_i'}\right)^2} \neq 0$$

$$\Longleftrightarrow y_{c_1} = y_{c_i'}$$

The InterD is strictly positive, and its value is null if, and only if, the barycenter of both group measurements has the same intensity.

3.1.2.3. Dispersion index (D − Index)

$$D - Index = \frac{IntraD}{1 + InterD} \quad \text{(Eq.13)}$$

Knowing that *IntraD* ≥ 0 & ≥ 0 ; and 1 + *InterD* ≠ 0, we have *D − Index* ≥ 0.

3.1.3. Stability by multiplication

The InterD *(Eq. (19))*, the IntraD *(Eq. (20))* and the *D − Index (Eq. (21))* showed different behaviors when a random variable X is multiplied by a constant *b*.

$$X : IntraD\{S(y \bullet b, x)\} \approx b \bullet (IntraD\{S(y, x)\}) \quad \text{(Eq.19)}$$

$$X : InterD\{S(y \bullet b, x)\} = b^2 \bullet (InterD\{S(y, x)\}) \quad \text{(Eq.20)}$$

$$X : \{D - Index\{S(y \bullet b, x)\} = \frac{1}{b} \bullet D - Index\{S(y, x)\}, \quad \text{(Eq.21)}$$

for large values of InterD and IntraD.

The application of the equations *(Eq. (19))*, *(Eq. (20))*, and *(Eq. (21))* to the simulated variables X, Y, G, F and N are presented in Table 2. The results were consistent with the expected theoretical behavior, being that the quotient of the intra-group dispersion of two variables, with an affine multiplicative relation is equal to the multiplicative factor of the two variables. For the InterD, the quotient is equal to the square of the multiplicative factor. Concerning the $D-$ *Index*, the quotient is approximately equal to the inverse of the multiplicative factor.

3.2. Investigation of the indicators in different experimental setups

3.2.1. Use case 1: Repeatability assessment

Three different examples were considered for repeatability assessment, namely the NIST samples of Project 1, 2 and 3 as presented in Fig. 1. Each example relates to a group of 3 NIST samples analyzed in one batch.

In these cases, the indicator under investigation was therefore the IntraD. As presented in Fig. 2, a strong correlation (>0.8) between the coefficient of variation and the IntraD could be highlighted in all cases. This shows the ability of this indicator to capture the variation within batch, while not submitted to the same theoretical constraints as the coefficient of variation. It should be noted that the intra-group dispersion behaves similarly to the coefficient of variation in presence of a gaussian distribution but remains also relevant in the case of a non-gaussian distribution.

3.2.2. Use case 2: intermediate precision assessment: estimation of the variability between batches

This use case was based on the two analytical batches (within the same project separated by a mass spectrometer source cleaning) of the NIST samples analyzed in Project 4 (Fig. 1). As presented in Fig. 3, a strong linear correlation between the IntraD and several parameters, namely the interquartile range, the median absolute deviation, the variance, and the SD ($r \geq 0.8$) was observed. Moreover, a strong linear correlation was also found with the SD ($r \geq 0.8$) but also with the range, the Fisher score, and the RSD ($r = -1$) for the InterD. Finally, the D − Index was negatively correlated with the interquartile range ($r = -1$), while a positive association with the coefficient of variation, range, and Fisher score ($r \geq 0.8$) was observed. These results highlighted the ability of the proposed indicators to capture different aspects of variability structure related to intermediate precision.

3.2.3. Use case 3: Within laboratory reproducibility

All groups of NIST samples analyzed in the various projects over one year (Fig. 1) were then included in this analysis to assess within laboratory reproducibility. As presented in Fig. 4, PCA was first applied to explore the major trends in the data, and the score plot of the first factorial plane highlighted the overall homogeneity of dispersion between the different groups analyzed. However, even if the structure of variability between the different projects may seem similar, the situation is much more complex, and specific patterns could be revealed using additional bivariate dispersion indicators and corresponding visualization tools.

As presented in Fig. 5, the correlation between the different indicators was consistent with the observed behaviors for intermediate precision (Fig. 3). Moreover, a strong linear correlation was observed between the IntraD, the range, the inter-quartile range, the median absolute deviation, and the standard deviation ($r \geq 0.8$). In addition, a strong positive correlation was found with the variance and the inter-quartile range ($r \geq 0.8$), while a strong negative correlation with the coefficient of variation ($r = -1$) could be highlighted for InterD. Moreover, the D − Index was negatively correlated with the Fisher score ($r = -1$).

3.2.4. Other uses of the indicators

Rather than using the IntraD and the InterD as indicators to evaluate dispersion in quality control, the D − Index intends to capture the structure between the inter-batch and the intra-batch dispersion. Alternatively, this can be used to quantify the dispersion between the QCs and samples as presented by Broadhurst et al. for the D-ratio. In fact, the D-ratio can take any indicator of variability into account, whether parametric or not. We could therefore include the new indicators in the calculation of the D-ratio by estimating these parameters separately on the QCs and the study samples, to apply a threshold for features that are subject to too much technical variability, following the D-Ratio rationale.

3.3. Visualization of the bivariate interstudy dispersion of selected reference compounds

3.3.1. Application on simulated data

The distribution of the dispersion computed from simulated data, with different types of bias (additive, multiplicative) is presented in Fig. 6. Results showed that the structure of the bivariate dispersion remained unchanged as far as the bias was constant.

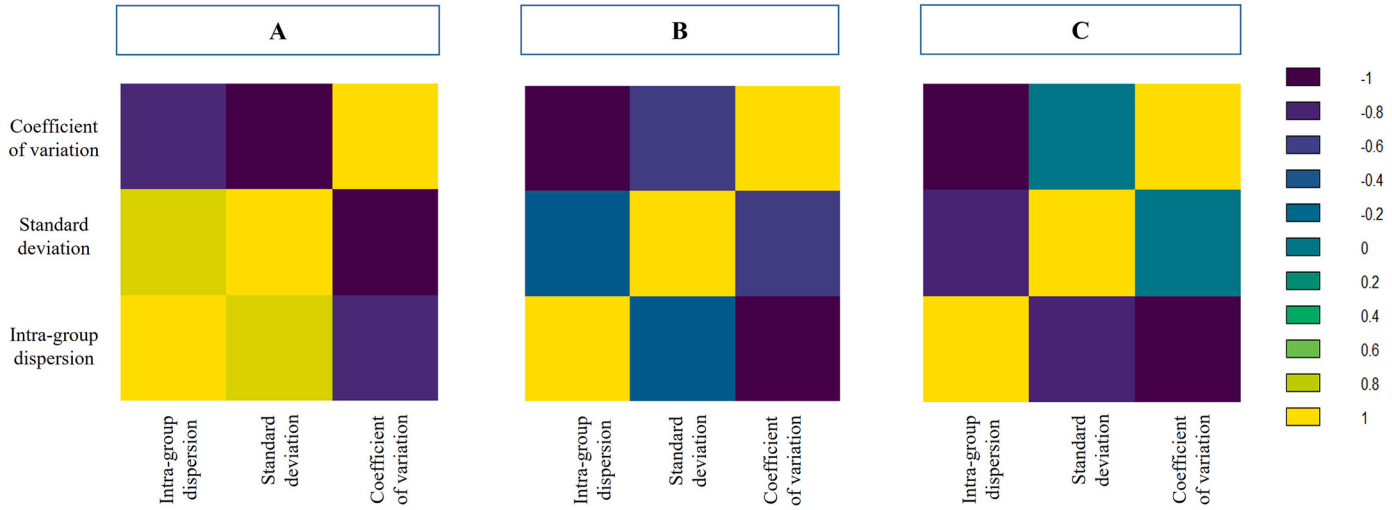


Fig. 2. Heatmaps of the intra-batch dispersion and the RSD and the standard deviation on Project 1 (A), Project 2 (B) and Project 3 (C).

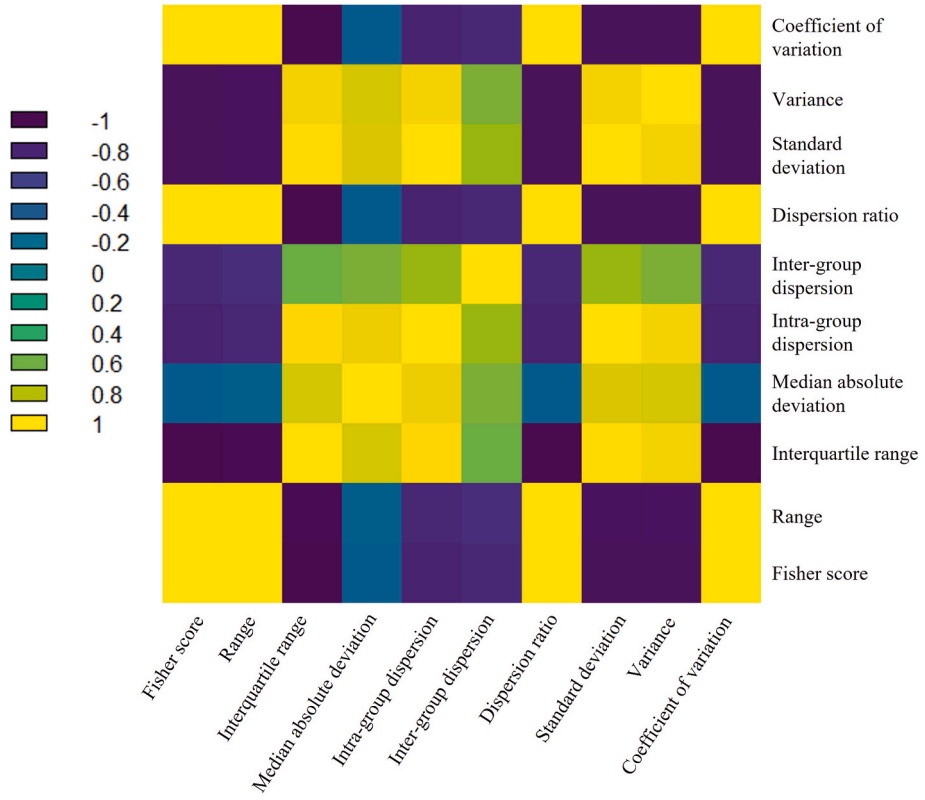


Fig. 3. Heatmaps of the correlation between the indicators and the others dispersion indicators.

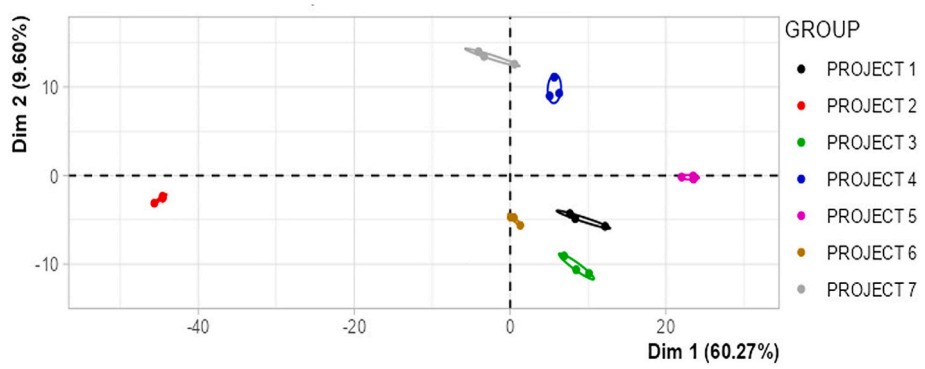


Fig. 4. Projection of NIST samples collected from the different projects in the first factorial plane using the identified metabolites with RSD <50 % and percentage of missing values < 20 %.

3.3.2. Application to experimental data

The visualization of the experimental data dispersion revealed different types of structures with various shapes. On the Y-axis, a long shape reflected an important variation in the signal intensities within replicates, while a flattened structure revealed good repeatability of measured values. For the different analytes presented in Fig. 7, a majority of compounds were associated with small areas and flat shapes illustrating a good repeatability within the Project 1. However, there is still an important heterogeneity revealing the complexity of the comparisons between different projects and different behavior among the referent analytes. This difference could be explained by the fact that NIST samples from Project 1 were analyzed just after a global maintenance of the mass spectrometer, thus reflecting the basic sensitivity of the instrument in a clean condition. By considering taurine, creatine and hypoxanthine in Fig. 7, the dispersion within Project 3 was highlighted as extremely high, while the data within the other projects presented smaller dispersion. In practice, observing this difference could lead to detect problems occurring during the data acquisition process. Interestingly, this clearly showed that the proposed visualization method can be useful to display the structure of the dispersion in the different measurement batches and guide decisions for a potential correction method. Moreover, this approach allowed the dispersion in the system at different time points to be investigated, which is of particular interest in the assessment of the system suitability, independently of a central parameter. This visualization tool can also be used to identify efficiently, similar behaviors (dispersion, sensitivity to batch effect) within analytes as presented in Fig. 7. Its combination with the proposed indicators, brings more information for a robust and easily interpretable quantification of the dispersion, than classical analytical visualization using extracted ion chromatograms and signal areas of replicates along sequences and batches (supplementary material 2).

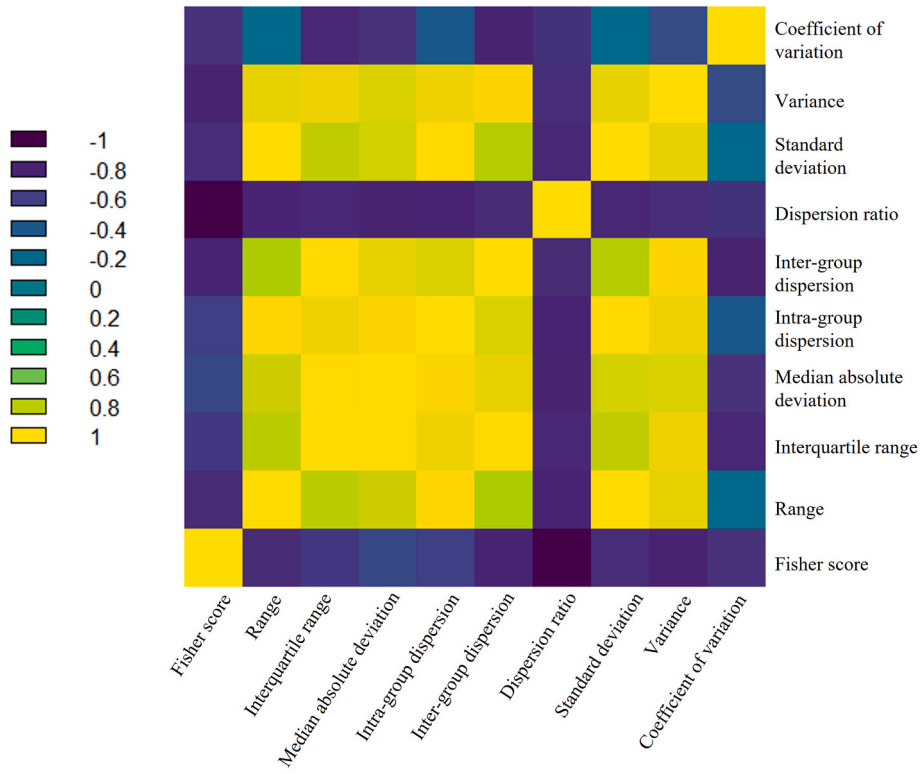


Fig. 5. Heatmap of the correlation between the proposed indicators and the other dispersion indicators.

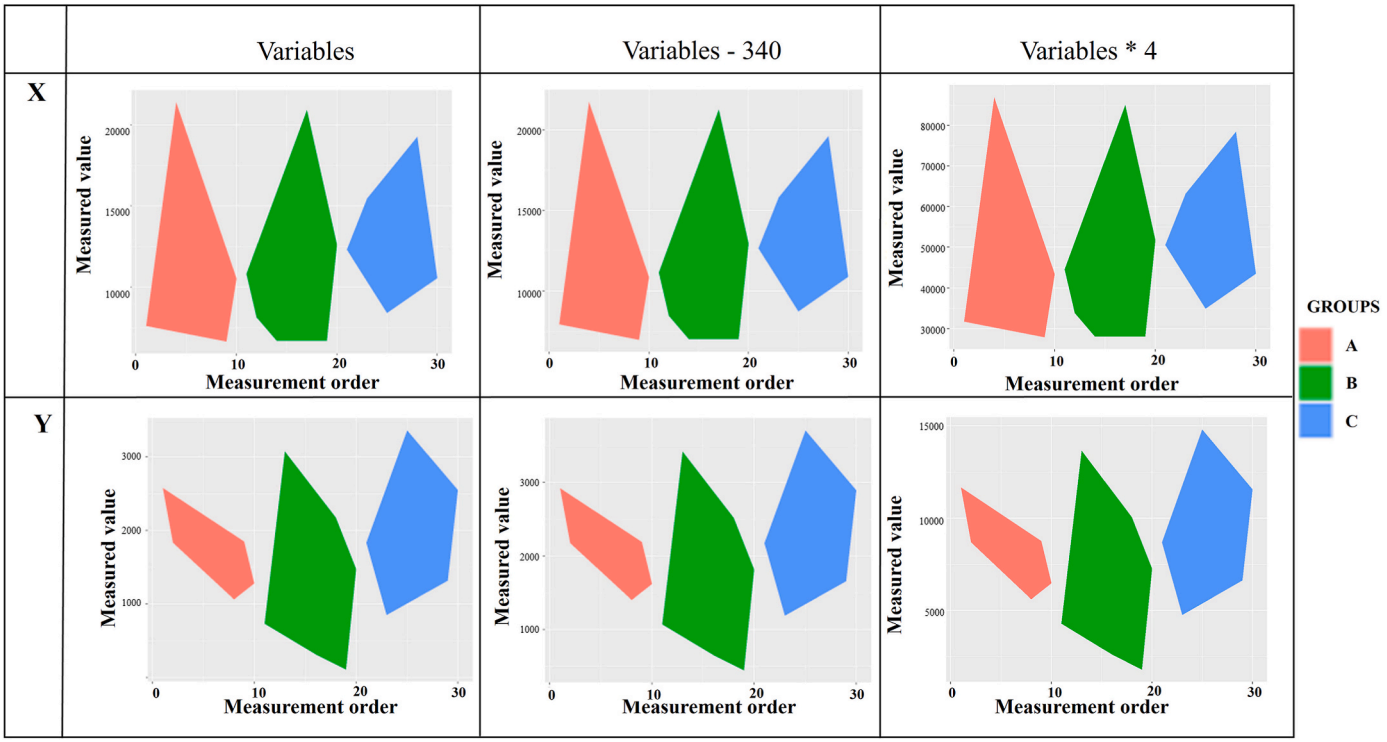


Fig. 6. Visualization of the dispersion on simulated variables, using the proposed visualization method.

4. Conclusion

In the present work, three new robust indicators were developed for assessing the analytical dispersion. Their combination with a visualization method, allowed an efficient evaluation of the structure of variability related to repeatability, intermediate precision, and reproducibility. Datasets collected in experimental setups can show a large dispersion, but this variability could originate from a specific analytical batch. This is particularly the case in untargeted metabolomics, when a large number of signals are measured in a single analysis. In that case,

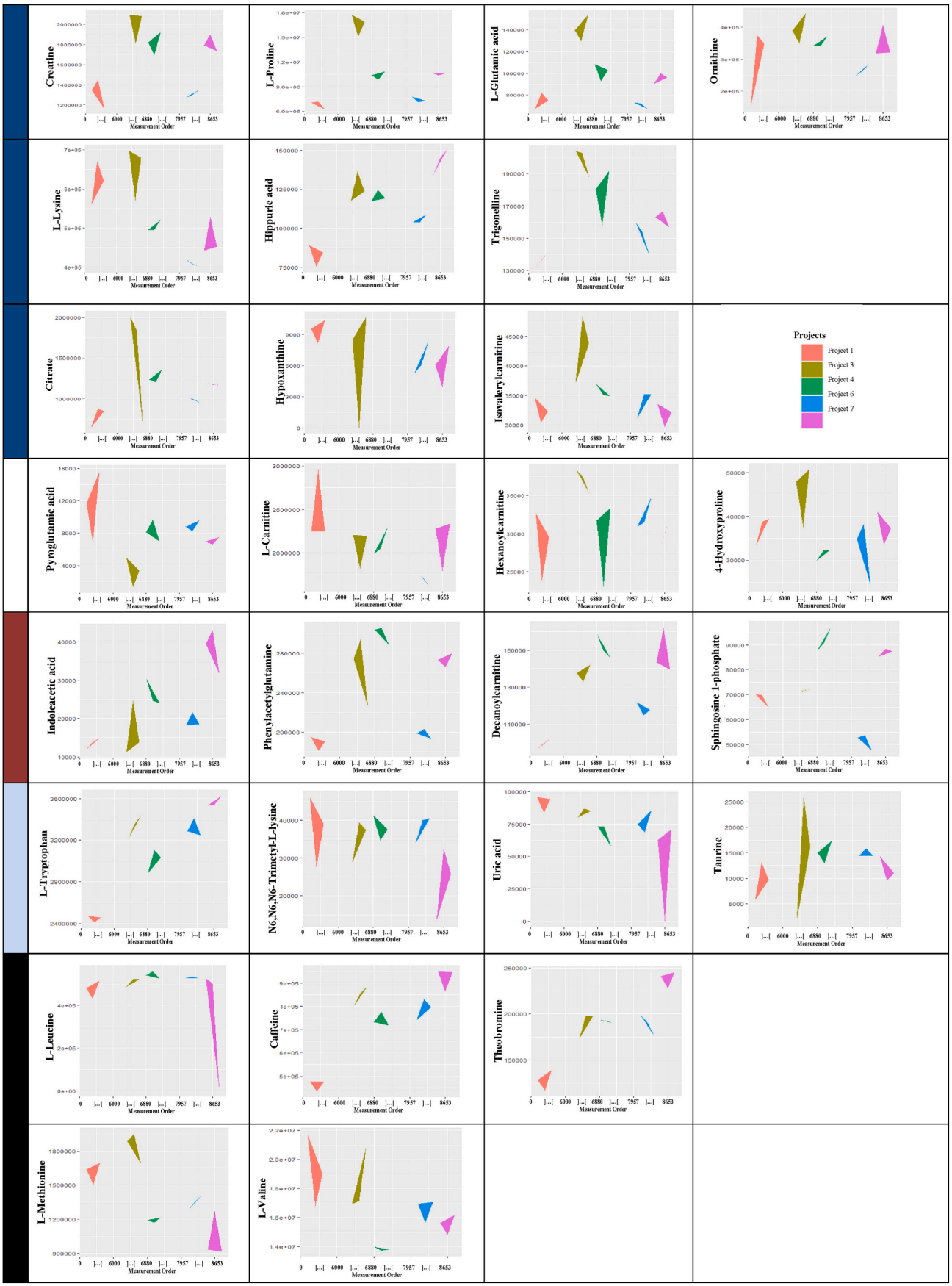


Fig. 7. Application of the visualization methods on the reference compounds in some groups NIST.

combining the proposed indicators with the visualization method contributed to reveal different types of structures and make diagnostic for further choice of an adapted correction method. Moreover, these indicators open the way for a robust assessment of the intra and inter assays variability assessment and for long-term suitability testing.

CRediT authorship contribution statement

Elfried Salanon: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Blandine Comte:** Writing – review & editing, Validation, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization. **Delphine Centeno:** Writing – review & editing, Validation, Data curation. **Stéphanie Durand:** Writing – review & editing, Data curation, Conceptualization. **Estelle Pujos-Guillot:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation. **Julien Boccard:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

All metabolomics analyses were funded and performed within the MetaboHUB French infrastructure (ANR-11-INBS-0010). E. Salanon is recipient of a doctoral fellowship from the INRAE DIGIT BIO metaprogramme and Geneva University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2024.105148.

References

[1] D.P. Green, J. Citrin, Measurement error and the structure of attitudes: are positive and negative judgments opposites? Am. J. Polit. Sci. 38 (1994) 256–281.
[2] D. Broadhurst, R. Goodacre, S.N. Reinke, et al., Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies, Metabolomics 14 (2018) 72.
[3] D. Dudzik, C. Barbas-Bernardos, A. García, et al., Quality assurance procedures for mass spectrometry untargeted metabolomics. a review, J. Pharm. Biomed. Anal. 147 (2018) 149–173.
[4] J.A. Kirwan, H. Gika, R.D. Beger, et al., Quality assurance and quality control reporting in untargeted metabolic phenotyping: mQACC recommendations for analytical quality management, Metabolomics 18 (2022) 70.
[5] B.A. Rappold, Review of the use of liquid chromatography-tandem mass spectrometry in clinical laboratories: Part II–operations, Ann. Lab. Med. 42 (2022) 531–557.
[6] A.M. Evans, C. O'Donovan, M. Playdon, et al., Dissemination and analysis of the quality assurance (QA) and quality control (QC) practices of LC–MS based untargeted metabolomics practitioners, Metabolomics 16 (2020) 113.
[7]] .
[8] De Berg M, Cheong O, Van Kreveld M, et al. Computational Geometry: Algorithms and Applications. Third. Berlin Heidelberg: Springer..
[9] B. Braden, The surveyor's area formula, Coll. Math. J. 17 (1986) 326–337.
[10] Polygons, Meshes, https://paulbourke.net/geometry/polygonmesh/(accessed 24 October 2023).
[11] K.A. Lippa, J.J. Aristizabal-Henao, R.D. Beger, et al., Reference materials for MS-based untargeted metabolomics and lipidomics: a review by the metabolomics quality assurance and quality control consortium (mQACC), Metabolomics 18 (2022) 24.
[12] E. Pujos-Guillot, M. Brandolini, M. Pétéra, et al., Systems metabolomics for prediction of metabolic syndrome, J. Proteome Res. 16 (2017) 2262–2272.
[13] R. Tautenhahn, C. Böttcher, S. Neumann, Highly sensitive feature detection for high resolution LC/MS, BMC Bioinf. 9 (2008) 504.
[14] F. Giacomoni, G. Le Corguillé, M. Monsoor, et al., Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics, Bioinformatics 31 (2015) 1493–1495.