Article

A New Approach for Multivariate Data Analysis in Interlaboratory Comparisons Based on Multidimensional Scaling and Robust Confidence Ellipse

Luiz Henrique C. Leal [a] *and Werickson F. C. Rocha* *,[a]

aInstituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO), 25250-020 Duque de Caxias-RJ, Brazil

Interlaboratory comparisons (IC) present a challenge related to multivariate data analysis. ISO 13528:2015 is a reference document for interlaboratory comparisons. This standard does not provide descriptions of statistical methods for multivariate analysis and, according to our best knowledge, there is no practical guidance for the organizing and evaluation of multivariate data analysis for interlaboratory comparisons available. Due to this reason, some researchers have made efforts to develop methodologies that make it possible to analyze multivariate data in IC. Generally, these approaches are based on dimensionality-reduction methods like principal component analysis. This paper proposes a new approach to reduce the dimensionality of large data set and check the performance of laboratories based on multidimensional scaling (MDS) and robust confidence ellipse/ellipsoid (RCE). MDS is a multivariate analysis technique that allows grouping laboratories according to their similarity in a Euclidean space. On the other hand, RCE is a statistical method for outlier detection in a multivariate data set. In this work, it is proposed to combine MDS and RCE to evaluate laboratory proficiency in interlaboratory comparison. This methodology was compared with the multivariate z-score and both methodologies identified the same outlying laboratories. This preliminary result indicates that MDS/RCE is promising for classifying IC results.

Keywords: multidimensional scaling, robust confidence ellipse, interlaboratory comparisons, multivariate analysis

Introduction

Interlaboratory comparison (IC) seeks to assess the performance of a given measurement method. To organize an IC, the same or similar items test are sent to each participating under predetermined conditions.[1] In some cases, the reported results by participants are multivariate data and univariate techniques are not suitable to analyze this kind of data.[2] For this situation, multivariate analysis techniques can be used for investigating the inherent structure of data without losing valuable information of measurement.

Multivariate analysis, in a general way, refers to all statistical methods that simultaneously analyze multiple measures in each individual or object (in this case laboratory).[3] The official document for interlaboratory comparison, ISO 13528:2015,[1] does not provide descriptions for multivariate analysis. Due to this gap left by the official document, some researchers have been proposed many ways to realize multivariate data analysis in interlaboratory

comparison. Sheen *et al*.[2] have proposed multivariate z-score to identify outlying laboratory results.

In the proposed methodology by Sheen *et al*.,[2] the spectra should be grouped so that the cluster consists of multiple nuclear magnetic resonance (NMR) spectra of the k-th sample provided by the participants. From these clusters, the interspectral distance matrix $\mathbf{D}_k$ is calculated, whose elements are the distances $D_{ij,k} = d(s_{i,k}, s_{j,k})$ where $s_{i,k}$ is the spectrum of laboratory i, $s_{j,k}$ is the spectrum of laboratory j (both belonging to the conglomerate S_k) and d(.) is a multivariate distance measure. The authors suggest Euclidean, Mahalanobis, Hellinger, Kullback-Leibler, Jensen-Shannon and Jeffreys distances.[2]

Based on the values $D_{ij,k}$, the average distance $\hat{D}_{i,k} = (1/n)\sum_j D_{ij,k}$ is computed. The $\hat{D}_{i,k}$ values need to be fitted to a given probability distribution for each laboratory i. After this step, the matrix $\mathbf{Z}$ is obtained, where Z_i is the z-score vector of the i-th laboratory. The vector Z_i is obtained by $Z_{i,k} = C^{*-1}(C_k(\hat{D}_{i,k}))$ where C_k is the cumulative distribution function after being fitted to the conglomerate k and C* is the corresponding standard distribution function.[2]

In this approach, the next step is to perform the principal component analysis (PCA) on the $\mathbf{Z}$ matrix. In the PCA

*e-mail: wfrocha@inmetro.gov.br
Editor handled this article: Maria Cristina Canela (Associate)

model, $\mathbf{T} = \mathbf{ZP}^t$ where $\mathbf{T}$ is the matrix of the scores of the principal components and $\mathbf{P}$ is the matrix of the loadings. The most significant L principal components are identified by obtaining $\mathbf{T}_L = \mathbf{ZP}_L{}^t$. For each participant, the Euclidean norm $\|\mathbf{T}_{i,L}\|$ is calculated and these statistical distances are adjusted to a new probability distribution with distribution function $\hat{C}$.[2]

This new distribution has a z-score associated with it. The authors called the projected statistical score ($\hat{Z}_i$). This score is calculated using $\hat{Z}_k = \hat{C}^{*-1}(\hat{C}(\|T_{i,L}\|))$. If any $\hat{Z}_i$ value falls outside the 95% confidence interval, the corresponding laboratory is considered an outlier and removed from the data set. The process is repeated until no dataset falls outside the 95% confidence interval.[2] In this study, 10 participants provide a one-dimensional ^{1}H NMR spectrum and three of them were identified as outlying by the proposed methodology.

Another approach to deal with multivariate data was proposed by Viant et al.[4] It was proposed principal component analysis (PCA) to clustering participating results. The items test sent to laboratories were synthetic metabolite mixtures and European flounder liver extracts (biological) from clean and contaminated sites. The goal of the intercomparison exercise was to evaluate the effectiveness of ^{1}H NMR metabolomics to generate comparable data sets from environmentally derived samples and each participant provides a one-dimensional spectrum. The associated PCA scores plots were used to visually assess if individual laboratories could reveal which metabolites discriminated the synthetic mixtures and biological samples. In both cases, the PCA approach allows concluding that there was a high degree of similarity across all laboratories. The comparability and precision of the laboratories participating were good, reflecting good results from NMR spectra.[4]

Gallo et al.[5] have proposed a new performance index, named Q_p-score, to assess the laboratory performance in multi-component analyses. Eight nuclear magnetic resonance signals (3 for aldicarb, 1 for methamidophos, 2 for oxadixyl, and 2 for pirimicarb) were obtained by 36 participants in interlaboratory comparison. It was proposed the following parameter: $Q_p = (a_i \cdot \bar{a})/\sigma_{slope}$ where a_i is the slope of the calibration line determined by the i-th participant, $\bar{a}$ is the consensus slope value, and σ_{slope} is the inter-laboratory standard deviation on slopes. The Q_p-score is considered satisfactory when $|Q_p| \leq 2$, questionable when $2 < |Q_p| < 3$ and unsatisfactory when $|Q_p| \geq 3$. The proposed methodology allows classifying the reported results by participants in each NMR signal. In general, 9 laboratories were classified as unsatisfactory, 2 were classified as questionable and 25 were classified as satisfactory.[5]

Other authors reported multivariate data analysis by PCA in interlaboratory comparisons. Danzer et al.[6], Henrion[7] and, Škrbić et al.[8] suggested PCA to identify outlying laboratories in interlaboratory comparison. Minkkinen[9] proposes principal components score plot to visualize the variation of the results between and within laboratories. Heininger et al.[10] have proposed PCA to group the laboratories and identify the type of method used to analyze the samples by each participants. Aoki et al.[11] proposed multiple hypothesis testing to assess the equivalence among the laboratories measurements with respect to the reference laboratory. The authors suggest build a confidence regions between each participating and reference laboratory based on Wald statistic.[11]

This paper proposes a new procedure to assess the performance of interlaboratory comparisons that are different from the methodologies mentioned above. A new methodology is derived from the concepts of outlying detection in two- and three-dimensional Euclidean space. The methodology is based on following steps.

Initially the data (Figure 1a) is arranged as matrix $\mathbf{X}_{n \times p}$ where n represents the number of variables, that is, the chemical shift of the spectrum provided by i-th laboratory and p represents the number of participants in the interlaboratory comparison (Figure 1b).

The second step refers to performing a multidimensional scaling (MDS) on the multivariate data to transform them into a dimension that can be used to visualize the results of the laboratories in two or three dimensions. After that, a confidence limit based on a robust confidence ellipse/ellipsoid (RCE) is plotted to identify outlying laboratory result (Figures 1c and 1d). Points outside of robust confidence ellipse/ellipsoid provide evidence that laboratory result is an outlier with a specific confidence level (95%, for example). On the other hand, points inside of RCE are not considered outlying result with same confidence level.

Multidimensional scaling is a multivariate technique that allows revealing "hidden" structures in a multivariate data set.[12] In other words, it is a method that allows visualizing the similarity/dissimilarity among laboratories participating in an interlaboratory comparison which are represented as points in a two- or three-dimensional space.[13] The proximity/distance between the points represents the similarity/dissimilarity among laboratories.[13]

Robust confidence ellipse/ellipsoid is a multivariate analysis technique for outlier detection. A confidence region is built from the variance-covariance matrix of original data set, which allows identifying if there are laboratories that differ statistically from the others at some specific significance level.[14]
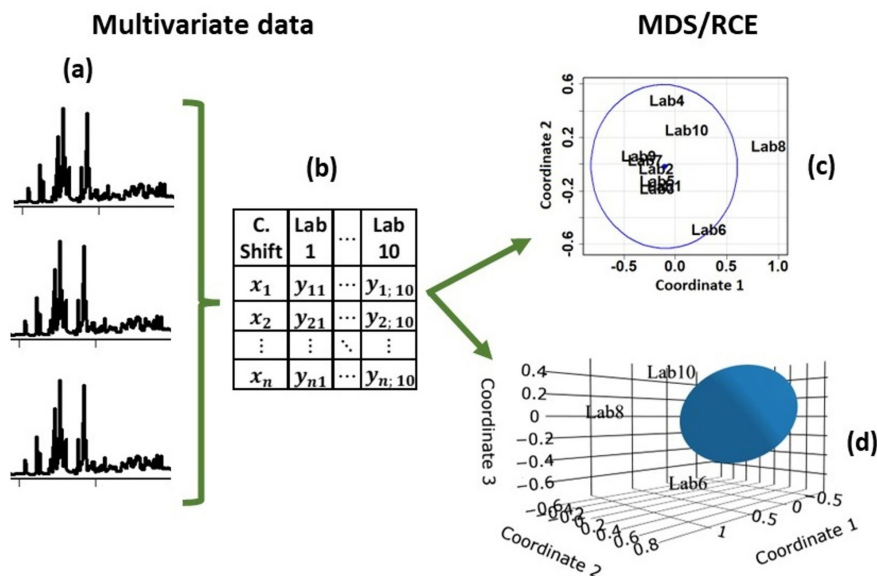
Figure 1. The proposed methodology: (a) multivariate reported spectrum by each participating; (b) multivariate data set organized in matrix where the first column contains the chemical shift and the remaining columns contain the spectra reported by each participant; (c) multidimensional scaling combined with robust confidence ellipse (MDS/RCE 2D); (d) multidimensional scaling combined with robust confidence ellipsoid (MDS/RCE 3D).

The combination of these two techniques (Figure 1) constitutes the proposal for evaluating the performance of the results reported by the participants.

Methodology

Data set

The data used in this study can be obtained from the available data set described by Sheen *et al.*[2] In this data set, an interlaboratory comparison was carried out to assess the effectiveness of ^{1}H NMR metabolomics. Seven laboratories from the United States, one from Canada, one from the United Kingdom, and one from Australia participated in the intercomparison. Mixtures of synthetic metabolites and samples of biological origin from liver extracts of European flounder from clean and contaminated sites were analyzed.[4]

The data described by Sheen *et al.*[2] refers to adult female European flounder collected from Tyne in the United Kingdom. This is a biological sample selected from a polluted site. The fish were sacrificed, liver tissues were dissected, snap-frozen in liquid nitrogen, and stored at −80 °C until extraction by IC participating.[4] The samples were extracted using methanol:chloroform:water method and Precellys-24 bead-based homogenizer (Stretton Scientific Ltd., U.K.).[4]

Each participating laboratory obtained a one-dimensional ^{1}H NMR spectrum. The spectra are reported as chemical shift frequencies with a range from 10.0 to 0.2 ppm. The region from 4.7 to 5.2 ppm was excluded

due to water solvent suppression artifacts and the NMR spectra were renormalized. The spectra were binned with a bin width of 0.005 ppm, for a total of 1860 variables in each spectrum.[2]

In this work, the multivariate techniques of multidimensional scaling and robust confidence ellipse/ellipsoid (Figures 1c and 1d) were explored to assess the performance of laboratories participating in a laboratory intercomparison classifying them as outlier or not. The basis for the development of this methodology is described below.

Multidimensional scaling

Let p be the number of different laboratories (Figure 1b) and δ_{ij} the dissimilarity between laboratories i and j. The coordinates are gathered in the matrix $\mathbf{X}_{n \times p}$ where n is the dimensionality of the solution to be specified. Thus, the column i of $\mathbf{X}_{n \times p}$ provides the coordinates of the laboratory i (Figure 1b). Let $d_{ij}(\mathbf{X})$ be the Euclidean distance (most used)[15,16] between columns i and j defined as:

$$d_{ij}(\mathbf{X}) = \left(\sum_{s=1}^{n} \left(x_{is} - x_{js} \right)^2 \right)^{\frac{1}{2}} \tag{1}$$

which is the shortest distance between laboratories i and j. In the equation 1, x_{is} is the spectrum of laboratory i and x_{js} is the spectrum of laboratory j.

The Euclidean distance is more favorable in visual representations because a more isotropic display is obtained using it.[17] The purpose of multidimensional scaling is to find an $\mathbf{X}_{n \times p}$ matrix that $d_{ij}(\mathbf{X})$ is equal to δ_{ij} as much as possible.[13]

To obtain this matrix, the least squares MDS model is used, which consists of minimizing the equation:

$$\sigma^2(\mathbf{X}) = \sum_{i=2}^{p}\sum_{j=1}^{i-1} w_{ij}\left(\delta_{ij} - d_{ij}(\mathbf{X})\right)^2 \tag{2}$$

where w_{ij} is a user defined weight which must be nonnegative. The minimization problem of $\sigma^2(\mathbf{X})$ is quite complex and it is necessary to use interactive algorithms to find the matrix $\mathbf{X}_{n\times p}$ that minimizes $\sigma^2(\mathbf{X})$. The most used for the solution of this is the SMACOF algorithm.[12,13,18]

Robust confidence ellipse

After applied MDS technique, the dimensionality data was reduced for two or three dimensions. The following step is identifying the outlying laboratories by robust confidence ellipse (Figure 1c).

The robust confidence ellipse is built from the matrix equation:

$$\hat{\boldsymbol{\mu}}_{rob} + \left(\sqrt{2 \cdot F_{2;(n-1);(1-\alpha)}}\right) \cdot \mathbf{U} \cdot \mathbf{Q} \tag{3}$$

where $\hat{\boldsymbol{\mu}}_{rob} = \begin{bmatrix}\overline{x}_{rob} & \overline{y}_{rob}\end{bmatrix}$ is the vector of robust means; $F_{2;(n-1);(1-\alpha)}$ is the quantile of the Fisher-Snedecor distribution with 2 and $(n-1)$ degrees of freedom and confidence level of $(1-\alpha)\%$; $\mathbf{Q}$ is the Cholesky decomposition of the robust variance-covariance matrix $\mathbf{S}_{rob}$.[18,19] Both $\hat{\boldsymbol{\mu}}_{rob}$ and $\mathbf{S}_{rob}$ are estimated by the iterative process described in the next section.

The matrix $\mathbf{U}$ is the unit circle defined by

$$\mathbf{U} = \begin{bmatrix}\cos(\mathbf{a}) & \sin(\mathbf{a})\end{bmatrix} \tag{4}$$

where $\mathbf{a} = [a_1 \ldots a_m]$ is a vector of size m $(0 \leq \mathbf{a} \leq 2\pi)$.

Robust means ($\hat{\boldsymbol{\mu}}_{rob}$) and variance-covariance matrix ($\mathbf{S}_{rob}$)

The vector of robust means $\hat{\boldsymbol{\mu}}_{rob}$ and the variance-covariance matrix $\mathbf{S}_{rob}$, for bi-dimensional data set, mentioned in the previous sub-section, are fitted from the following iterative process.[19] Let $(\mathbf{x}; \mathbf{y})$ a bi-dimensional data set.

$$(\mathbf{x};\mathbf{y}) = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix} \tag{5}$$

Step (i): $\hat{\mu} = \begin{bmatrix}\overline{x} & \overline{y}\end{bmatrix}$ and $wi = 1 + p/\upsilon \ \forall \ i = 1, \ldots, n$ where p is the number of variables and υ is degree of freedom of multivariate t distribution.

Step (ii): compute the matrix:

$$\mathbf{X} = \begin{bmatrix} (x_1 - \overline{x}) & (y_1 - \overline{y}) \\ \vdots & \vdots \\ (x_n - \overline{x}) & (y_n - \overline{y}) \end{bmatrix} \tag{6}$$

Step (iii): compute the singular value decomposition of matrix $\mathbf{A}$, svd($\mathbf{A}$) = $\mathbf{USV}^T$, where:

$$\mathbf{A} = \begin{bmatrix} \sqrt{\dfrac{w_1}{\sum_{i=1}^{n}w_i}}(x_1 - \overline{x}) & \sqrt{\dfrac{w_1}{\sum_{i=1}^{n}w_i}}(y_1 - \overline{y}) \\ \vdots & \vdots \\ \sqrt{\dfrac{w_n}{\sum_{i=1}^{n}w_i}}(x_n - \overline{x}) & \sqrt{\dfrac{w_n}{\sum_{i=1}^{n}w_i}}(y_n - \overline{y}) \end{bmatrix} \tag{7}$$

Step (iv): compute the matrix:

$$\mathbf{w}' = \mathbf{XVS}' = \begin{bmatrix} w'_{11} & w'_{12} \\ \vdots & \vdots \\ w'_{n1} & w'_{n2} \end{bmatrix} \tag{8}$$

where,

$$\mathbf{S}' = \begin{bmatrix} 1/s_1 & \\ & 1/s_2 \end{bmatrix} \tag{9}$$

Step (v): compute the vector $\mathbf{Q} = [Q_1 \ldots Q_n] = [(w'_{11})^2 + (w'_{12})^2 \ldots (w'_{n1})^2 + w'_{n2})]$

Step (vi): compute the new weights:

$$w_i^* = \frac{(\upsilon + p)}{(\upsilon + Q_i)} \quad \forall \ i = 1, \ldots, n \tag{10}$$

Step (vii): compute the robust vector of means:

$$\hat{\boldsymbol{\mu}}_{rob} = \begin{bmatrix}\overline{x}_{rob} & \overline{y}_{rob}\end{bmatrix} = \begin{bmatrix} \dfrac{\sum_{i=1}^{n}w_i^* x_i}{\sum_{i=1}^{n}w_i^*} & \dfrac{\sum_{i=1}^{n}w_i^* y_i}{\sum_{i=1}^{n}w_i^*} \end{bmatrix} \tag{11}$$

Step (viii): the new fitted vector $\hat{\boldsymbol{\mu}}_{rob}$ and the weights w_i^* are used to obtain new values in step (i) ($\hat{\mu} = \hat{\boldsymbol{\mu}}_{rob}$ and $w_i = w_i^*$). This procedure is repeated until the values of w_i^* converge, that is, $|w_i - w_i^*| < \varepsilon$.

Step (ix): compute the robust variance-covariance matrix $\mathbf{S}_{rob}$.[19]

$$\mathbf{S}_{rob} = \begin{bmatrix} \dfrac{\sum_{i=1}^{n}w_i^*(x_i - \overline{x}_{rob})^2}{n} & \dfrac{\sum_{i=1}^{n}w_i^*(x_i - \overline{x}_{rob})(y_i - \overline{y}_{rob})}{n} \\ \dfrac{\sum_{i=1}^{n}w_i^*(x_i - \overline{x}_{rob})(y_i - \overline{y}_{rob})}{n} & \dfrac{\sum_{i=1}^{n}w_i^*(y_i - \overline{y}_{rob})^2}{n} \end{bmatrix} \tag{12}$$

Robust confidence ellipsoid

It is recommended to investigate through the robust confidence ellipsoid (Figure 1d) the possible presence of outliers that were not identified in the 2D analysis mentioned in previous sub-section.

The robust confidence ellipsoid is built from an equation analogous to the robust confidence ellipse:

$$\hat{\boldsymbol{\mu}}_{rob} + \left(\sqrt{3 \cdot F_{3;(n-1);(1-\alpha)}} \right) \cdot U \cdot Q \qquad (13)$$

where $\hat{\boldsymbol{\mu}}_{rob} = \begin{bmatrix} \overline{x}_{rob} & \overline{y}_{rob} & \overline{z}_{rob} \end{bmatrix}$ is the vector of robust means; $F_{3;(n-1);(1-\alpha)}$ is the quantile of the Fisher-Snedecor distribution with 3 and $(n-1)$ degrees of freedom and confidence level of $(1-\alpha)\%$; Q is the Cholesky decomposition of the robust variance-covariance matrix. Furthermore, $\hat{\boldsymbol{\mu}}_{rob}$ and S_{rob} are estimated analogous to iterative process described in previous sub-section.

The $m \times 3$ matrix U is the sphere of radius 1 defined by:

$$U = \begin{bmatrix} \cos(\theta)\sin(\varphi) & \sin(\theta)\sin(\varphi) & \cos(\varphi) \end{bmatrix} \qquad (14)$$

where $\boldsymbol{\theta} = [\theta_1 \ \dots \ \theta_m]$ is a vector of size m $(0 \leq \boldsymbol{\theta} \leq 2\pi)$ and $\boldsymbol{\varphi} = [\varphi_1 \ \dots \ \varphi_m]$ is a vector of size m $(0 \leq \boldsymbol{\varphi} \leq \pi)$.[18]

Software

All statistical analyses were performed using the R statistical software, an open-source free environment for statistical computing and graph creation.[18] The 2D plot analysis (Figure 1c) was built using the R packages (CAR and STATS).[18] In addition, an R code was built to define which points obtained in the multidimensional scaling technique were outside the robust confidence ellipse. The MASS R package[18] was necessary for this purpose.

Regarding the 3D analysis (Figure 1d), the plot was built using PLOTLY R package[18] which is a graphing library that makes interactive plots. It is worth pointing out that to build the robust confidence ellipsoid it was necessary to develop an R code based on the dataEllipse R function from CAR R package.[18] Moreover, an R code was built to identify outlying results. This code also depends on the MASS R package.[18]

Results and Discussion

The results of the participants (Figure 1a) were arranged in a matrix with 1860 rows and 10 columns (Figure 1b). The columns contain the results of the participants in which each column is the spectrum reported by the i-th participant. The lines represent the variables (chemical shift in ppm).

It should be noted that this is the first study combining multidimensional scaling and robust confidence ellipse/ellipsoid (MDS/RCE) to evaluate laboratory proficiency in interlaboratory comparison. In order to validate the methodology suggested in this article, the MDS/RCE results will be compared with multivariate z-score results obtained by Sheen *et al.*[2]

In the proposed methodology, the two-dimensional multidimensional scaling provides a location in Euclidean space of each laboratory participating according to similarity/dissimilarity between ^{1}H NMR spectra. Additionally, the two-dimensional robust confidence ellipse allows identifying if there is an outlier spectrum (Figure 1c). The Figure 2 presents the results obtained by two-dimensional approach.
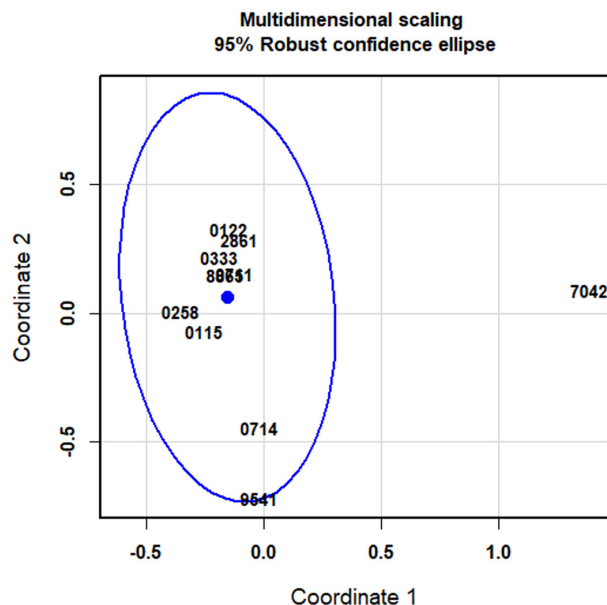

Figure 2. Multidimensional scaling from ^{1}H NMR spectra reported by participating and 95% robust confidence ellipse (2D plot).

It can be seen in Figure 2 that the laboratory 7042 is located out of ellipse. This provides evidence that its result differs statistically from other participants. At 5% significance level (Figure 2) the spectrum reported by laboratory 7042 is classified as an outlier when compared to other participants. On the other hand, laboratories that are located within the robust confidence ellipse do not differ statistically from each other. In this case, there is no evidence that the results reported by these participants are outliers.

The following step is analyzing the three-dimensional limit according to Figure 1d. The 3D plot allows seeing an outlier result that is "hidden" in the 2D plot. In other words, the three-dimensional approach provides more information about participating performances.

It can be seen from Figure 3 that more laboratories show results that differ from the others (at the 5% significance level) when compared to the bi-dimensional plot showed in Figure 2. Altogether there are three laboratories (0714, 7042, and 9541) out of ellipsoid which means that their results may be considered outliers, that is, statistically different from the others. The rest, inside the ellipsoid, do not differ statistically from each other. By an analogous reasoning followed for the univariate z-score suggested by ISO 13528:2015,[1] performance assessment by the MDS/RCE is considered acceptable when results are inside of ellipse/ellipsoid and unacceptable when results are outside.
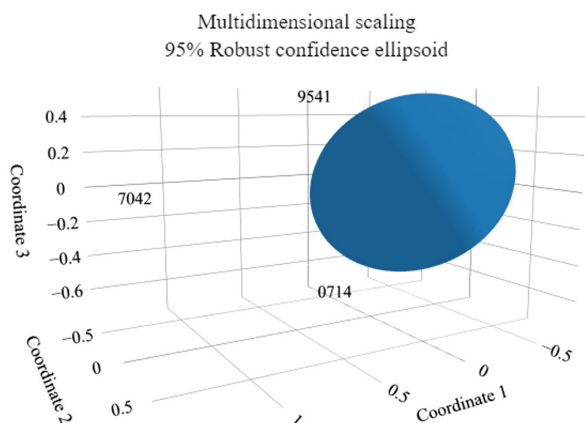

Figure 3. Multidimensional scaling from ^1H NMR spectra reported by participating and 95% robust confidence ellipsoid (3D plot).

The confidence level adopted to consider the reported result as an outlier was 95% in 2D (Figure 2) and 3D (Figure 3) plot analysis.

In some cases, two (Figure 1c) and three (Figure 1d) dimensional analysis can provide the same conclusions, however, there are situations where this does not occur (such as Figures 2 and 3). In this situation, the two-dimensional analysis differs from the three-dimensional analysis, thus the latter must be adopted because it offers a more comprehensive (and reliable) analysis of the participants' results.

In the multivariate z-score approach suggested by Sheen *et al.*,[2] the interspectral distance matrix $\mathbf{D}_k$ was obtained from the Kullback-Leibler, Mahalanobis, Hellinger and Jensen-Shannon distances. The $\hat{D}_{i,k}$ values were fitted to a lognormal distribution. The choice of this probability distribution was based on the Q-Q plot. Additionally, according to the authors, the lognormal is the maximum entropy distribution for a specified mean and standard deviation.[2]

Each $Z_{i,k}$ value is an indication of where the $s_{i,k}$ spectrum is in relation to the others in the S_k cluster. In the case of the lognormal distribution, $Z_{i,k}(1/2) = 1$ and $Z_{i,k}(0.95)$ ca. 5. In

this context, $Z_{i,k} = 1$ indicates that $s_{i,k}$ is closer to the center of S_k while $Z_{i,k}$ greater than 5 indicates that $s_{i,k}$ is outside the range 95% confidence in the conglomerate S_k.[2]

The $\|T_{i,L}\|$ values were fitted to a lognormal distribution and the projected statistical score $\hat{Z}_i$ was calculated for each data set. The scores calculated from the Kullback-Leibler, Mahalanobis, Hellinger, and Jensen-Shannon distances showed evidence that participants 0714, 7042 and 9541 reported results (spectra) are outliers.[2]

The methodology based on multidimensional scaling combined with the robust confidence ellipse/ellipsoid (Figure 1) identified the same outlying participant results as the multivariate z-score method proposed by Sheen *et al.*[2] The main advantage of the methodology proposed in this article is that it does not depend on choosing a probability distribution. This provides evidence that MDS/RCE has potential to be considered as performance evaluation method in interlaboratory comparisons.

Conclusions

In this paper, it was proposed a new approach to analyze multivariate data from interlaboratory comparisons. The methodology combines multidimensional scaling and robust confidence ellipse/ellipsoid to identify an outlying laboratory result. The results obtained by the methodology suggested in this work were compared with the results obtained by the multivariate z-score method described by Sheen *et al.*[2] The MDS/RCE method found the same three outlying laboratories (0714, 7042, and 9541) identified in the multivariate z-score method. The proposed methodology has the advantage in relation to the latter of not depending on the choice, sometimes subjective, of a probability distribution.

This approach proved to be promising as a performance evaluation method for multivariate data analysis in interlaboratory comparison. It should be noted that this methodology can be used, in a similar way, to evaluate the performance of laboratories participating in proficiency testing scheme. In this approach, it is suggested classifying the participating laboratory results located outside of robust confidence ellipse/ellipsoid as unacceptable and inside as acceptable like univariate z-score showed in ISO 13528:2015.

The proposed method therefore constitutes a valuable tool that contribute to filling a gap in the literature regarding the multivariate data analysis in interlaboratory comparisons and proficiency trials. Other multivariate data analysis techniques like factor analysis and Kohonen's self-organizing map, for example, are multivariate techniques that can be considered in future works for performance assessment tool in interlaboratory comparisons.

References

1. ISO 13528: *Statistical Methods for Use in Proficiency Testing by Interlaboratory Comparisons*, ISO: Geneva, 2015.

2. Sheen, D. A.; Rocha, W. F. C.; Lippa, K. A.; Bearden, D. W.; *Chemom. Intell. Lab. Syst.* **2017**, *162*, 10. [Crossref]

3. Hair Jr., J. F.; Black, W. C.; Babin, B. J.; Anderson, R. E.; *Multivariate Data Analysis*, 7th ed.; Prentice Hall: New Jersey, USA, 2009.

4. Viant, M. R.; Bearden, D. W.; Bundy, J. G.; Burton, I. W.; Collette, T. W.; Ekman, D. R.; Ezernieks, V.; Karakach, T. K.; Lin, C. Y.; Rochfort, S.; De Ropp, J. S.; Teng, Q.; Tjeerdema, R.S.; Walter, J. A.; Wu, H.; *Environ. Sci. Technol.* **2009**, *43*, 219. [Crossref]

5. Gallo, V.; Intini, N.; Mastrorilli, P.; Latronico, M.; Scapicchio, P.; Triggiani, M.; Bevilacqua, V.; Fanizzi, P.; Acquotti, D.; Airoldi, C.; Arnesano, F.; Assfalg, M.; Benevelli, F.; Bertelli, D.; Cagliani, L. R.; Casadei, L.; Marincola, F. C.; Colafemmina, G.; Consonni, R.; Cosentino, C.; Davalli, S.; De Pascali, S. A.; D'aiuto, V.; Faccini, A.; Gobetto, R.; Lamanna, R.; Liguori, F.; Longobardi, F.; Mallamace, D.; Mazzei, P.; Menegazzo, I.; Milone, S.; Mucci, A.; Napoli, C.; Pertinhez, T.; Rizzuti, A.; Rocchigiani, L.; Schievano, E.; Sciubba, F.; Sobolev, A.; Tenori, L.; Valerio, M.; *Anal. Chem.* **2015**, *87*, 6709. [Crossref]

6. Danzer, K.; Wank, U.; Wienke, D.; *Chemom. Intell. Lab. Syst.* **1991**, *12*, 69. [Crossref]

7. Henrion, R.; *Chemom. Intell. Lab. Syst.* **1994**, *25*, 1. [Crossref]

8. Škrbić, B.; Héberger, K.; Đurišić-Mladenović, N.; *Anal. Bioanal. Chem.* **2013**, *405*, 8363. [Crossref]

9. Minkkinen, P.; *Chemom. Intell. Lab. Syst.* **1995**, *29*, 263. [Crossref]

10. Heininger, P.; Pelzer, J.; Henrion, R.; Henrion, G.; *Fresenius' J. Anal. Chem.* **1998**, *360*, 344. [Crossref]

11. Aoki, R.; Leão, D.; Bustamante, J. P. M.; Vilca, F.; *J. Appl. Stat.* **2021**, 1. [Crossref]

12. Leeuw, J.; Mair, P.; *J. Stat. Software* **2009**, *31*, 1. [Crossref]

13. Groenen, P. J. F.; van de Velden, M.; *J. Stat. Software* **2016**, *73*, 1. [Crossref]

14. ISO 13528: *Statistical Methods for Use in Proficiency Testing by Interlaboratory Comparisons*, ISO: Geneva, 2005.

15. Agrafiotis, D. K.; Rassokhin D. N.; Lobanov, V. S.; *J. Comput. Chem.* **2001**, *22*, 488. [Crossref]

16. Morrison, A.; Ross, G.; Chalmers, M.; *Inf. Vis.* **2003**, *2*, 68. [Crossref]

17. Asan, U.; Ercan, S. In *Computational Intelligence Systems in Industrial Engineering*; Kahraman, C., ed.; Atlantis Press: Paris, France, 2012, ch. 14. [Link]

18. Bates, D.; Chambers, J.; Dalgaard, P.; Gentleman, R.; Hornik, K.; Ihaka, R.; Kalibera, T.; Lawrence, M.; Leisch, F.; Ligges, U.; Lumley, T.; Maechler, M.; Meyer, S.; Murrell, P.; Plummer, M.; Ripley, B.; Sarkar, D.; Lang, D. T.; Tierney, L.; Urbanek, S.; *R Project*, version 4.2.0; R Foundation, Austria, 2022.

19. Kent, J. T.; Tyler, D. E.; Vard, Y.; *Commun. Stat. Simul. Comput.* **1994**, *23*, 441. [Crossref]