Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 16-21
doi: 10.30744/brjac.2179-3425.letter-dnrutledge

BrJAC

LETTER

Multiblock Methods in Analytical Chemistry

Douglas N. Rutledge[1,2]

[1]*Université Paris-Saclay, INRAe, AgroParisTech, UMR SayFood, 75005, Paris, France*
[2]*National Wine and Grape Industry Centre, Charles Sturt University, Wagga Wagga, Australia*

Chemometrics, and multivariate data analysis in particular, has become a significant component of Analytical Chemistry as can be seen from the NGrams plot in Figure 1 for the word “chemometrics”. This is because of the need to have mathematical methods capable of extracting the pertinent information from the ever-increasing amounts of data generated by modern instruments. Usually, these multivariate data analysis methods are concerned with the exploratory analysis of a single data matrix, as in PCA, or with relating one explanatory matrix to another descriptive matrix, as in regression methods such as PCR and PLS, or discriminant methods, such as FDA and PLS-DA.

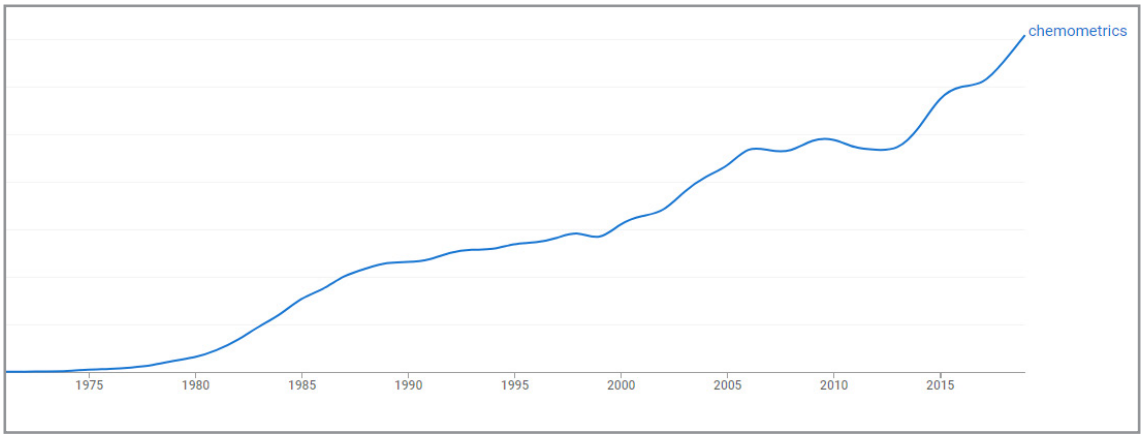

Figure 1. Evolution of the usage of the word “chemometrics” as given by NGrams (https://books.google.com/ngrams).

Recently however, there has been a trend towards analyzing many matrices simultaneously, the data being in the form of blocks of variables describing the same individuals. This trend of multiblock analysis (Figure 2) is the result of two forces: the availability of a wide range of very different instrumental techniques, and a paradigm shift towards a holistic study of complex systems. This is the case, for example, for -omics data, where combining or fusing data from different instruments can result in a better characterization of the individuals under study than that which is possible using each source of information separately.

Cite: Rutledge, D. N. Multiblock methods in Analytical Chemistry. *Braz. J. Anal. Chem.,* 2021, *8* (32), pp 16-21. doi: http://dx.doi.org/10.30744/brjac.2179-3425.letter-dnrutledge

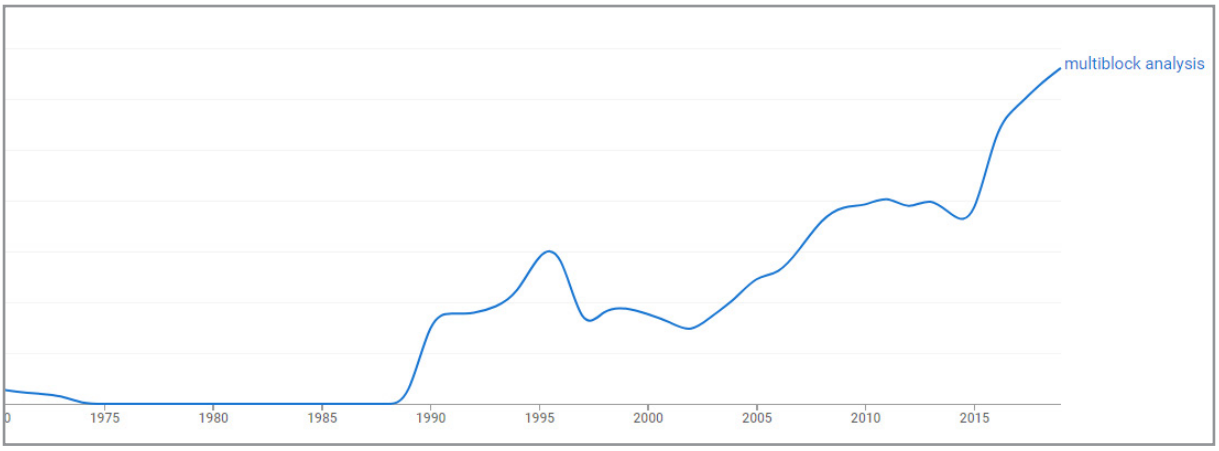

Figure 2. Evolution of the usage of the term “multiblock analysis” as given by NGrams (https://books.google.com/ngrams).

Much work has been done recently on the development of multiblock data analysis methods in order to treat this sort of data. Multiblock methods were initially used in sensometrics for the analysis of sensory data. For example, when a set of samples are evaluated by a group of judges, each using several (possibly different) descriptors, it is interesting to know more than just the relations between the samples but also which judges consider the samples as similar and which evaluate them differently. In chemometrics, the aim is to highlight the relationships between various blocks of variables. Early discussions of multiblock methods can be found in Smilde, Westerhuis and de Jong [1], Westerhuis et al. [2], and Qin et al. [3]. A very recent overview of the many different multiblock methods and their applications can be found in the book “Data fusion methodology and applications” [4].

The ComDim multiblock data analysis method

The main characteristic of multiblock methods is that they extract global components corresponding to the directions of greatest dispersion of the individuals, common to the multidimensional spaces defined by each of the data blocks. Although, as can be seen in [4], there are many interesting methods available to do multiblock, multivariate data analysis, I will concentrate in the following on a particular algorithm simply because I know it well and I think it works very well.

“Common Components and Specific Weights” (CCSWA or ComDim) is an unsupervised, multi-block (or multi-table) data analysis method developed by Qannari et al. [5-7], in the context of sensory profiling for the simultaneous analysis of several data tables describing the same individuals. It has since been widely applied in chemometrics, for example: simultaneous analysis of Mid Infrared (MIR) and Fluorescence spectra of cheeses [8], chemical and sensory characteristics of wines [9], the fatty acid composition of edible oils evaluated by combining Near Infrared (NIR) and Ultraviolet-Visible (UV-Vis) spectroscopy and Gas Chromatographic (GC-FID) data [10], interpretation of NIR and NMR spectral data, quality parameters and sensory properties of Brazilian coffees [11], monitoring surface water quality using physico-chemical, microbiological and 3D Fluorescence data [12], characterizing structural changes in a semisolid pharmaceutical formulation by NIR spectroscopy and Raman imaging [13], discrimination of commercial Yerba mates by combining HPLC, phytochemical composition, antioxidant activity, Visible and NIR spectroscopy, colorimetry and electronic nose data [14], coupling data from 3 laser-induced breakdown spectroscopy (LIBS) detectors to sort geological materials from caves [15], combining NMR and MIR spectra with stable isotope data to differentiate organically- and conventionally-produced tomatoes [16], combining 1H NMR and 13C NMR spectra with stable isotope data to differentiate organically- and conventionally-produced milk [17], combining NMR, MIR and Isotope Ratio MS data to discriminate tomato varieties [18].

ComDim and two of its extensions have also been applied in metabolomics, for example in a large-scale, multi-instrument inter-laboratory study [19], using ComDim and OPLSDA to combine positive- and negative-mode ElectroSpray Ionisation data from an UHPLC-TOF/MS system [20], using ComDim and OPLSDA to evaluate the therapeutic potential of a series of 83 flavonoid derivatives by relating five blocks of physicochemical properties to their affinity toward P-glycoprotein, and to differentiate a series of 60 human cancer cell lines by combining transcriptomic, metabolomic and proteomic data [21].

The ComDim method consists in determining a common space for all the data tables, with each matrix having a specific contribution ("salience") to the definition of each orthogonal direction of this common space. The components are iteratively extracted so as to correspond to the maximum amount of variance that is common to the largest number of tables. Each table is first normalized so that larger tables do not automatically have more influence in the calculation.

An iterative process is used to estimate the contribution ("salience") of each block to each CC. A significant difference in the saliences of the blocks for a given CC reflects their different contributions to the construction of that common dimension.

The coordinates of the observations on the ComDim directions are the 'Global Scores' and the contributions of the variables within each of the normalized tables are the 'Scaled Loadings'.

In the original algorithm, each CC is the first normed scores vector of a weighted sum of scalar matrices calculated from all the data tables as shown in Figure 3, in the simplest case of two centered and normalized data blocks, $\mathbf{X}_1$ and $\mathbf{X}_2$. A weighted sum $\mathbf{W}_G$ of the samples-based variance–covariance matrices, $\mathbf{W}_i = \mathbf{X}_i \times \mathbf{X}_i^T$, is calculated using an initial weighting, or salience, of $\lambda_i = 1$ for all tables. The vector of scores of the first normed Principal Component is extracted from $\mathbf{W}_G$ as an initial estimate of the first Common Component (CC). The salience, λ_i, of each block **W**i is then recalculated from these scores. The estimations of the Global Scores and saliences are optimized by iterative recalculation until convergence. Each original matrix $\mathbf{X}_i$ is then deflated, and the procedure is repeated for the calculation of the second CC, and so on.

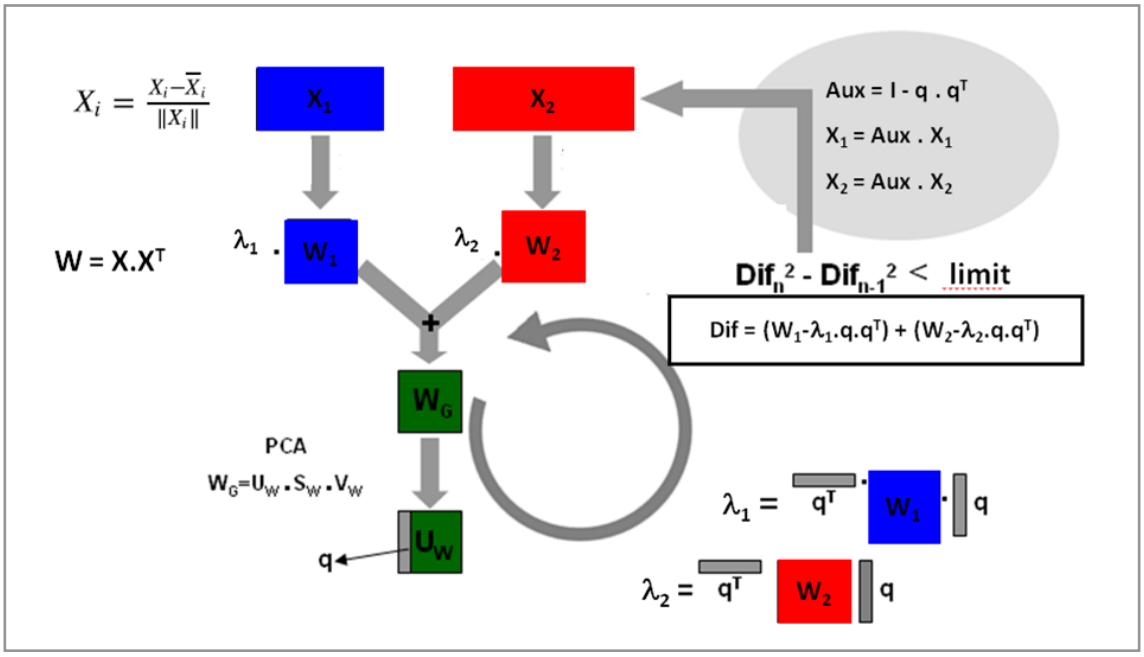

Figure 3. Schema of the original ComDim algorithm in the case of two data blocks.

Extensions of ComDim

As mentioned above, there have been a number of extensions and adaptations of the ComDim algorithm, such as replacing the PCA step by a PLS regression, or a discriminant analysis using PLSDA or OPLSDA [20,21].

Another entirely different predictive method, P-ComDim, has been developed by Qannari et al. [22] and even extended as a Path-Modelling method [23] which is useful if all the blocks are assumed to have a specific pattern of directed relations among them reflecting, for instance, a chain of influence.

In a way similar to the ANOVA–PCA method proposed by Harrington et al. [24], AComDim is an adaptation of ComDim to identify significant factors and interactions in an experimental design [25,26].

Software

A Graphical User Interface for multiblock data analysis (MB-GUI) [27] has been developed to make the implementation of multi-block data analysis easier, so that it can also be done by practitioners with no programming skills. The GUI can be downloaded from (https://github.com/puneetmishra2/Multi-block.git) and can be either installed to run in the MATLAB environment or as a standalone executable program.

The program covers a range of tasks such as multi-block data pre-processing, visualization, exploration, predictive modelling, variable selection and multi-block analysis for data fusion. The article also includes a list of other free software resources available for multi-block data analysis.

CONCLUSION

The acquisition of multi-modal data in order to have a more complete understanding of the characteristics of complex systems is becoming widespread. This has necessitated the development of new algorithms to perform multiblock data analyses. The existence of these new tools is now having the interesting effect of producing a positive feedback, leading to even more multi-modal analyses. We are only at the beginning of this revolution.

To conclude, I would like to point out that much of the progress in this field is a result of the heritage of Ronei Jesus Poppi, through the inspiration that he has given to so many young chemometricians.

REFERENCES

1. Smilde, A. K.; Westerhuis, J. A.; de Jong, S. *J. Chemom*., **2003**, *17* (6), pp 323–337, (https://doi.org/10.1002/cem.811).
2. Westerhuis, J. A.; Kourti, T.; MacGregor, J. F. *J. Chemom*., **1998**,*12*, pp 301–321 (https://doi.org/10.1002/(SICI)1099-128X(199809/10)12:5<301::AID-CEM515>3.0.CO;2-S).
3. Qin, S. J.; Vallee, S.; Piovoso, M. J. *J. Chemom*., **2001**, *15* (9) pp 715–742 (https://doi.org/10.1002/cem.667).
4. Cocchi, M. (Ed.) *Data Fusion Methodology and Applications*, ISBN: 978-0-444-63984-4, Elsevier, Amsterdam, **2019**.
5. Qannari, E.; Wakeling, I.; Courcoux, P.; MacFie, H. J. H. *Food Quality and Preference*, **2000**,*11*, pp 151–154 (https://doi.org/10.1016/S0950-3293(99)00069-5).
6. Hanafi, M.; Qannari, E. M. Nouvelles propriétés de l'analyse en composantes communes et poids spécifiques, *Journal de la Société Française de Statistique*, **2008**,*149*, pp 75–97.
7. Cariou, V.; Jouan-Rimbaud-Bouveresse, D.; Qannari, E. M.; Rutledge, D. N. ComDim Methods for the Analysis of Multiblock Data. In: Cocchi, M. (Ed.), *Data Fusion Methodology and Applications*, ISBN: 978-0-444-63984-4, **2019**, pp 179-204.
8. Mazerolles, G.; Devaux, M. F.; Dufour, E.; Qannari, E. M.; Courcoux, P. *Chemom. Intell. Lab. Syst.*, **2002**, *63*, pp 57-68 (https://doi.org/10.1016/S0169-7439(02)00036-9).
9. Blackman, J.; Rutledge, D. N.; Tesic, D.; Saliba, A.; Scollary, G. R. *Anal. Chim. Acta*, **2010**, *660*, pp 2-7 (https://doi.org/10.1016/j.aca.2009.10.062).

10. Rosa, L. N.; de Figueiredo, L. C.; Bonafé, E. G.; Coqueiro, A.; Visentainer, J. V.; Março, P. H.; Rutledge, D. N.; Valderrama, P. *Anal. Chim. Acta*, **2017**, *961*, pp 42-48. (https://doi.org/10.1016/j.aca.2017.01.019).
11. Baqueta, M. R.; Coqueiro, A.; Março, P. H.; Mandrone, M.; Poli, F.; Valderrama, P. ComDim for multi-block data analysis of Brazilian coffees: Identification of metabolites and application to the interpretation of the near infrared and 1H NMR spectral data, quality parameters and sensory properties, SLACA 2019, Anais do 13º Simpósio Latino Americano de Ciências de Alimentos, ISSN: 2447-2840, **2019**, *4*, 114673.
12. Daou, C.; El Hoz, M.; Kassouf, A.; Legube, B. *Water*, **2020**, *12* (6), 1673 (https://doi.org/10.3390/w12061673).
13. Mitsutake, H.; de Géa Neves, M.; Rutledge, D. N.; Poppi, R. J.; Breitkreitz, M. C. *J. Chemom*., **2020** (https://doi.org/10.1002/cem.3288).
14. Vieira, T. F.; Makimori, G. Y. F.; dos Santos Scholz, M. B.; Zielinski, A. A. F.; Bona, E. *Food Analytical Methods*, **2020**, *13*, pp 97–107 (https://doi.org/10.1007/s12161-019-01520-9).
15. Hohmann, M.; Monakhova, Y.; Erich, S.; Christoph, N.; Wachter, H.; Holzgrabe, U. *J. Agric. Food Chem.*, **2015**, *63*, pp 9666–9675 (https://doi.org/10.1021/acs.jafc.5b03853).
16. Hohmann, M.; Monakhova, Y.; Erich, S.; Christoph, N.; Wachter, H.; Holzgrabe, U. *J. Agric. Food Chem*., **2015**, *63* (43), pp 9666–9675 (https://doi.org/10.1021/acs.jafc.5b03853).
17. Erich, S.; Schill, S.; Annweiler, E.; Waiblinger, H.-U.; Kuballa, T.; Lachenmeier, D. W.; Monakhova, Y. B. *Food Chem*., **2015**, *188*, pp 1–7 (https://doi.org/10.1016/j.foodchem.2015.04.118).
18. Monakhova, Y. B.; Hohmann, M.; Christoph, N.; Wachter, H.; Rutledge, D. N. *Chemom. Intell. Lab. Syst*., **2016**, *156*, pp 1–6 (https://doi.org/10.1016/j.chemolab.2016.05.006).
19. Martin, J. C.; Maillot, M.; Mazerolles, G.; Verdu, A.; Lyan, B.; Migne, C.; Defoort, C.; Canlet, C.; Junot, C.; Guillou, C.; et al. *Metabolomics*, **2015**, *11*, pp 807-821 (https://doi.org/10.1007/s11306-014-0740-0).
20. Boccard, J.; Rutledge, D. N. *Anal. Chim. Acta*, **2013**, *769*, pp 30–39 (https://doi.org/10.1016/j.aca.2013.01.022).
21. Boccard, J.; Rutledge, D. N. *Anal. Chim. Acta*, **2014**, *813*, pp 25–34 (https://doi.org/10.1016/j.aca.2014.01.025).
22. El Ghaziri, A.; Cariou, V.; Rutledge, D. N.; Qannari, E. M. *J. Chemom*., **2016**, *30*, pp 420–429 (https://doi.org/10.1002/cem.2810).
23. Cariou, V.; Qannari, E. M.; Rutledge, D. N.; Vigneau, E. *Food Quality and Preference,* **2018**, *67*, pp 27–34 (https://doi.org/10.1016/j.foodqual.2017.02.012).
24. Harrington, P. d. B.; Vieira, N. E.; Espinoza, J.; Nien, J. K.; Romero, R.; Yergey, A. L. *Anal. Chim. Acta*, **2005**, *544* (1–2), pp 118-127 (https://doi.org/10.1016/j.aca.2005.02.042).
25. Jouan-Rimbaud-Bouveresse, D.; Pinto, R. C.; Schmidtke, L. M.; Locquet, N.; Rutledge, D. N. *Chemom. Intell. Lab. Syst.*, **2011**, *106* (2), pp 173-182 (https://doi.org/10.1016/j.chemolab.2010.05.005).
26. Guisset, S.; Martin, M.; Govaerts, B. *Chemom. Intell. Lab. Syst.*, **2019**, *184*, pp 44-63 (https://doi.org/10.1016/j.chemolab.2018.11.006).
27. Mishra, P.; Roger, J-M.; Rutledge, D. N.; Biancolillo, A.; Marini, F.; Nordon, A.; Jouan-Rimbaud-Bouveresse, D. *Chemom. Intell. Lab. Syst.*, **2020**, *205*, 104139 (https://doi.org/10.1016/j.chemolab.2020.104139).

Douglas Neil Rutledge is Emeritus Professor at AgroParisTech in Paris, where he was until 2017 Director of the Analytical Chemistry Laboratory. His research interests cover chemometrics, spectroscopy, chromatography and their application to the characterization of a wide range of materials, including agricultural, food, forensics, and pharmaceuticals. His work has resulted in the publication of over 200 refereed journal papers, 420 conference communications and 20 book chapters mainly on the development and application of novel chemometrics methods. He received the prestigious prize “Celestino da Costa-Jean Perrin” from the Portuguese Foundation for Science and Technology (FCT).

Prof. Rutledge’s collaboration with Brazil started in 2008 when two former PhD students of Prof. Ronei Poppi, Patricia Valderrama and Paulo Henrique Março, came to do a Postdoc in his laboratory. This visit led to a very fruitful long-distance collaboration with them after their return to Brazil. In 2013, thanks to a CAPES project that they set up, Prof. Rutledge was invited to spend a month in Brazil, participating in a series of conferences and giving lectures in universities and research centres in Rio de Janeiro, Curitiba, Campo Mourão, Dois Vizinhos, Campinas and Maringá. This visit gave him the opportunity to meet Prof. Ronei Poppi, as well as two others of his former PhD students, Prof. Dr. Márcia Cristina Breitkreitz and Dr. Andre Marcelo Souza, who are both now important members of the Brazilian Chemometrics community.

Prof. Poppi, Prof. Breitkreitz and Dr. Souza prepared a research and teaching project within the program of “*Cátedras Franco-Brasileiras*” in the State of São Paulo in partnership with the General Consulate of France. This project was selected in 2016, and Pr. Rutledge was invited to spend a month teaching and developing research projects at the Chemistry Institute of the University of Campinas – UNICAMP.

Since then, he has been invited back to Brazil several times and has also had the opportunity to welcome PhD students, Postdocs and Professors to visit his Laboratory in Paris.

These collaborations have been extremely fruitful. Much of it would not have been possible without the role played by Prof. Ronei Poppi as a teacher, a mentor and a major force in Chemometrics.