

**LETTER**

## Chemometrics Reveals Not-so-Obvious Analytical Information

Fabiola Manhas Verbi Pereira  

*Grupo de Abordagens Analíticas Alternativas (GAAA), Instituto de Pesquisa em Bioenergia (IPBEN), Instituto de Química, Universidade Estadual Paulista (UNESP), 14800-060, Araraquara, SP, Brazil*

The application of chemometric tools in analytical chemistry or other areas of chemistry has become essential. This is mainly due to the large amount and nature of the generated data<sup>1,2</sup> and the need to extract useful information from these and optimize steps throughout a process. It allows the quick decision-making visualization of interactions among variables, such as synergism or antagonism between parameters, during the development of a method,<sup>3</sup> as shown in Figure 1.



**Figure 1.** Pictorial of chemometrics directing analytical information.

Classical chemometric techniques have been disseminated and can be divided according to the study approach, among which exploratory data analysis stands out. Principal component analysis (PCA) is one of the most accessible and well-established ways to perform an initial exploration and extract relevant information from a given dataset and has been used quite successfully in various spectroscopic techniques.<sup>4</sup> Principal component analysis consists of projecting the data in a smaller dimension, enabling the detection of anomalous samples (outliers), the selection of essential variables in a given system, and unsupervised classification.<sup>1,2,4</sup>

Another branch of chemometrics involves the design of experiments (DoE). The primary purpose of the factorial design is to study the influence or effect of a given variable and its interactions in a specific system.<sup>5-9</sup>

Multivariate calibration is another aspect of chemometrics, where several variables are used to calibrate one (or more) property or the concentration of a given chemical analyte.<sup>10,11</sup> Since the first publications of chemometric tools, numerous variations of these techniques, proposals for data fusion strategies, and applications using hyphenated instrumental techniques have been proposed.<sup>12-14</sup>

Industrial quality control and development (R&D) laboratories require an approach addressing adequate quality by design (QbD). The QbD strategies consider four steps that include an analytical target profile (ATP), a risk assessment, a design space (DS), and control strategy and validation based on figures of merit, for instance.<sup>9</sup>

Principal component analysis is the most widely multivariate technique used for data analysis. Jolliffe wrote a review reporting his wonderful experience with PCA in the last 50 years.<sup>15</sup> Indeed, PCA is an invaluable method for data, and I agree with it. PCA is the algorithm of choice for numerous chemometric techniques.<sup>16</sup>

Other computational languages, such as Python, are currently experiencing a rise in popularity in the field of chemistry. The R language has also become more popular than it was ten years ago. The scripts, functions, or codes are easily written with fewer lines and specific commands that minimize steps and help speed up calculations. The dissemination of free software has also become popular, and the sharing of codes through publications, social media, communities, or websites has become relatively easy.

From my point of view, chemometrics is no longer faced as a giant monster or a way to become scientific papers fancier without helpful content. Chemometrics extract information that is not easy to visualize at first through univariate evaluations or using simple plots. Nowadays, thousands of instrumental data can provide important chemical information, and we must use them for significant proposals. Indeed, QbD is proof of that, since Industry 4.0 is a reality.

## REFERENCES

- (1) Olivieri, A. C. *Introduction to multivariate calibration: A practical approach*. Springer International Publishing, Switzerland, 1st ed., 2018, XVII-243.
- (2) Lavine, B. K.; Rayens, W. S. Classification: basic concepts. In: Brown, S. D.; Tauler, R.; Walczak, B. (Eds.). *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*. Elsevier, Amsterdam, The Netherlands, 2009, 507-515.
- (3) Pereira, F. M. V.; Pereira-Filho, E. R. Aplicação de programa computacional livre em planejamento de experimentos: um tutorial. *Quim. Nova* **2018**, 4 (9), 1061-1071. <http://dx.doi.org/10.21577/0100-4042.20170254>
- (4) Geladi, P. Chemometrics in spectroscopy. Part 1. Classical chemometrics. *Spectrochim. Acta Part B* **2003**, 58 (5), 767-782. [https://doi.org/10.1016/S0584-8547\(03\)00037-5](https://doi.org/10.1016/S0584-8547(03)00037-5)
- (5) Myers, R. H.; Montgomery, D. C.; Anderson-Cook, C. M. *Response surface methodology: process and product optimization using designed experiments*. Wiley, Hoboken, 2009.
- (6) Barros Neto, B.; Scarminio, I. S.; Bruns, R. E. *Como fazer experimentos*. Bookman, Porto Alegre, 2010.
- (7) Ferreira, S. L. C. *Introdução às técnicas de planejamento de experimentos*. Vento Leste, Salvador, 2015.
- (8) Pereira Filho, E. R. *Planejamento fatorial em química: maximizando a obtenção de resultados*. Edufscar, São Carlos, 2015.
- (9) Araújo, A. S.; Andrade, D. F.; Babos, D. V.; Castro, J. P.; Garcia, J. A.; Sperança, M. A.; Gamela, R. R.; Machado, R. C.; Costa, V. C.; Guedes, W. N.; Pereira-Filho, E. R.; Pereira, F. M. V. Key Information Related to Quality by Design (QbD) Applications in Analytical Methods Development. *Braz. J. Anal. Chem.* **2021**, 8 (30), 14-28. <http://dx.doi.org/10.30744/brjac.2179-3425.RV-27-2020>
- (10) Martens, H.; Naes, T. *Multivariate Calibration*. John Wiley & Sons, Chichester, 1993.
- (11) Ferreira, M. M. C. *Quimiometria: Conceitos, Métodos e Aplicações*. Editora Unicamp, Campinas, 2015.

- (12) Geladi, P.; Kowalski, B. R. Partial least-squares regression: a tutorial. *Anal. Chim. Acta* **1986**, *185* (1), 1-17.
- (13) Biancolillo, A.; Remo, B.; Magrì, A. L.; Magrì, A. D.; Marini, F. Data-fusion for multiplatform characterization of an Italian craft beer aimed at its authentication, *Anal. Chim. Acta* **2014**, *820*, 23-31. <http://dx.doi.org/10.1016/j.aca.2014.02.024>
- (14) Brownfield, B.; Lemos, T.; Kalivas, J. H. Consensus classification using non-optimized classifiers. *Anal. Chem.* **2018**, *90* (7), 4429-4437. <http://dx.doi.org/10.1021/acs.analchem.7b04399>
- (15) Jolliffe, I. A 50-year personal journey through time with principal component analysis. *J. Multivar. Anal.* **2022**, *188*, 104820. <https://doi.org/10.1016/j.jmva.2021.104820>
- (16) Varmuza, K.; Filzmoser, P. *Multivariate statistical analysis in chemometrics*. CRC Press, Boca Raton, 2009.



**Fabiola Manhas Verbi Pereira** is an Associate Professor at the Chemistry Institute Unesp - campus Araraquara. She received her BS in Chemistry in 1999 from the Federal University of São Carlos (UFSCar, São Carlos, São Paulo State, Brazil) and her Ph.D. in Sciences in 2007 from the Chemistry Institute at Campinas State University (UNICAMP, Campinas, São Paulo State, Brazil). In 2018, she joined the Idaho State University (ISU, Pocatello, Idaho, USA) as a postdoctoral fellow focusing her research on analyzing samples without pre-treatment, thereby improving the speed of sample processing and reducing the generation of residues. She also believes in multidisciplinary interactions to foster collaboration among scientific fields. She is an associate editor in *Food Analytical Methods* (Springer US). [CV](#)