Received: 12 December 2019 | Accepted: 16 December 2019

DOI: 10.1002/cem.3214

SPECIAL ISSUE - EDITORIAL

Journal of CHEMOMETRICS WILEY

Orthogonal PLS (O-PLS) and related algorithms

Federico Marini

Department of Chemistry, University of Rome La Sapienza, Rome, Italy

Correspondence
Federico Marini, Department of Chemistry, University of Rome La Sapienza, P. le Aldo Moro 5, Rome I-00185, Italy.
Email: federico.marini@uniroma1.it

The concept of orthogonalized partial least squares regression or, better, as it was originally named, orthogonalized projection to latent structures (O-PLS) was first introduced in 2001 [1,2] by Johann Trygg and Svante Wold, as a way to deal with the large amount of variation in predictor matrices for multivariate calibration (and classification), not correlated to the responses. In this context, O-PLS operates by partitioning the systematic variance in the X block into a Y relevant and an orthogonal data sets, both having a bilinear structure. As soon as the algorithm has been published, it has immediately gained a lot of interest and rapid popularity, as shown in Figure 1, where it is evident that the trend of published papers using algorithms from the O-PLS family or referring to them is growing exponentially.

In particular, as shown in Figure 2, where the main field of application of O-PLS and O-PLS-related algorithm is displayed in the form of a pie chart, it is unequivocal that, apart from the "natural" application to purely chemistry-related problems, as expected for an algorithm born within the chemometric community, a large share of the researches fall under the wide umbrella of -omic sciences. It is undubious that a part of the O-PLS popularity in the -omic field can be related to its availability as the main tool in one of the most widely adopted software in the area, but it is also true that the -omic sciences are one of the disciplines where the impact of irrelevant variation on the data is high and where the quest for straightforward identification of interpretable marker is the stake.

On the other hand, probably also because of its popularity, during the years, there has been a large debate on the O-PLS algorithm itself, to highlight its real peculiarities, whether it really had unique features or, on the other hand, to understand whether it should just be seen as one out of the many possibilities of identifying sources of common and distinctive variation among two or more blocks [3-7].

Starting from these considerations, in this special section of the issue, four papers representing the different aspects of how the research and the debate around O-PLS and related issues have evolved so far, are collected.

Sjögren et al confirm the multidisciplinary interest around O-PLS, by showing how multivariate data analysis can be applied in the context of text analysis in a combination with common machine learning preprocessing methodologies. In particular, their attention is focused on demonstrating how PCA, O-PLS, and its multiblock analog On-PLS on bag-of-words representations of abstracts, claims, and detailed descriptions could provide an accurate classification of patents from big pharma companies [8].

On the other hand, starting from the above-mentioned multiblock analog of O-PLS, On-PLS, but extending the applicability also to other multiblock models, Skotare et al. present different tools for data visualization. In particular, they propose a correlation matrix plot, to highlight the relationships between blocks found by multiblock models. Moreover, they also introduce additional plots, a multiblock scatter plot, a metadata correlation plot, and a variation distribution plot, which are meant to simplify the interpretation [9].

Indahl's paper [10] is an example of the theoretical debate which has developed around O-PLS and similar algorithms sharing the characteristics of being all derived from the orthogonal signal correction (OSC) principle [11]. Here, the mathematics behind O-PLS and related algorithm is thoroughly examined to verify whether all the claims associated with these methods are really met or not.

Lastly, Stocchero [12] introduces a new posttransformation method, called posttransformation of the latent variable space (ptLV), to separate the nonpredictive data from the predictive one. Its peculiarities are that it works on the score

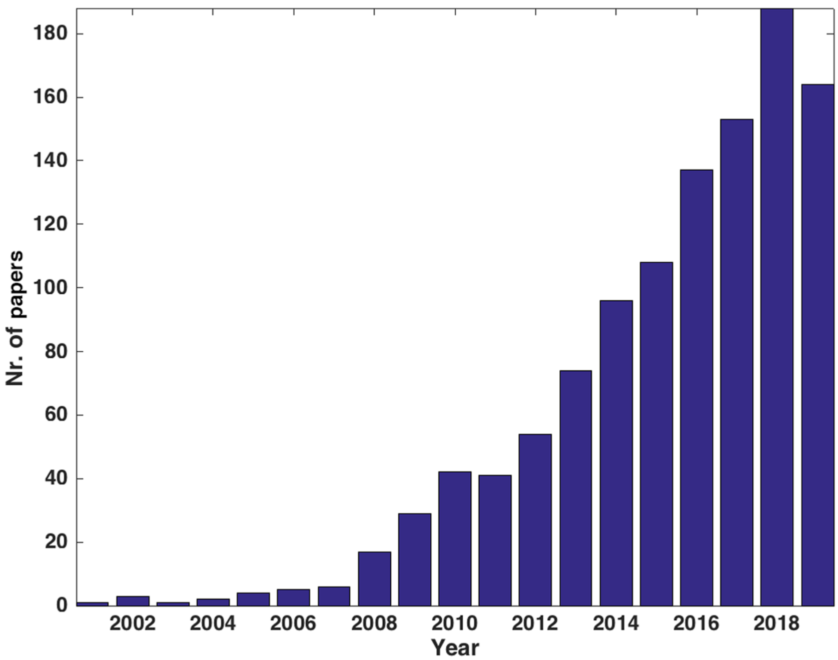

FIGURE 1 Number of studies based on O-PLS, O-PLS-DA, On-PLS, and related methods (source: Scopus, December 2019)

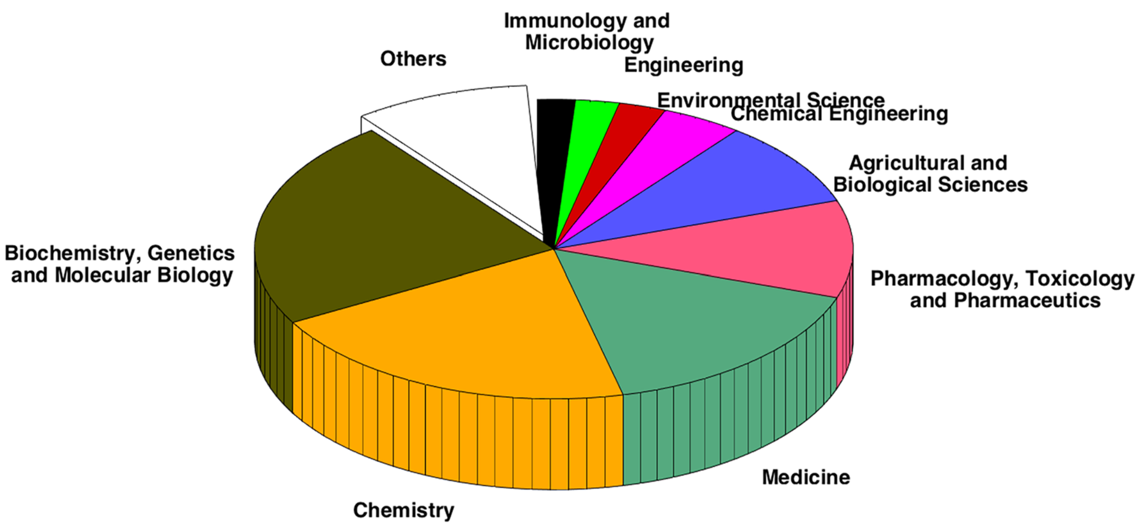

FIGURE 2 Distribution of the studies shown in Figure 1 based on the field of application (source: Scopus, December 2019)

space and can be applied also to kernel-PLS2 (KPLS2) and that, differently from other posttransformation approaches, does not require the explicit calculation of the weight matrix.

ORCID

Federico Marini https://orcid.org/0000-0001-8266-1117

REFERENCES

1. Wold S, Trygg J, Berglund A, Antti H. Some recent development in PLS modeling. *Chemometr. Intell. Lab. Syst.* 2001;58:131-150.
2. Trygg J, Wold S. Orthogonal projection to latent structures. *J Chemometr*. 2002;16:119-128.
3. Ergon R. PLS post-processing by similarity transformation (PLS+ST): a simple alternative to OPLS. *Chemometr. Intell. Lab. Syst.* 2005;96:1-4.
4. Kemsley EK, Tapp HS. OPLS filtered data can be obtained directly from non-orthogonalized PLS1. *J Chemometr*. 2009;23:263-264.
5. Kvalheim OM. Interpretation of partial least squares regression models by means of target projection and selectivity ratio plots. *J Chemometr*. 2010;24:496-504.

6. Indahl UG. Towards a complete identification of orthogonal variation in multiple regression from a PLS1 modeling point of view: including OPLS by a change of orthogonal basis. *J Chemometr*. 2014;28:508-517.
7. Smilde AK, Måge I, Næs T, et al. Common and distinct components in data fusion. *J Chemometr*. 2017;31:e2900.
8. Sjögren R, Stridh K, Skotare T, Trygg J. Multivariate patent analysis—using chemometrics to analyze collections of chemical and pharmaceutical patents. *J Chemometrics*. 2020;34:e3041.
9. Skotare T, Sjögren R, Surowiec I, Nilsson D, Trygg J. Visualization of descriptive multi-block analysis. *J Chemometr*. 2020;34:e3071.
10. Indahl UG. The O-PLS methodology for orthogonal signal correction—is it correcting or confusing? *J Chemometr*. 2020;34:e2884.
11. Fearn T. On orthogonal signal correction. *Chemometr Intell Lab Syst*. 2000;50:47-52.
12. Stocchero M. Exploring the latent variable space of PLS2 by post-transformation of the score matrix (ptLV). *J Chemometr*. 2020;34:e3079.

How to cite this article: Marini F. Orthogonal PLS (O-PLS) and related algorithms. *Journal of Chemometrics*. 2020;34:e3214. https://doi.org/10.1002/cem.3214