XI COLLOQUIUM CHEMOMETRICUM MEDITERRANEUM
27 - 30 JUNE 2023
PADOVA - ITALY

expertPLS: a MATLAB toolbox for spectral preprocessing selection

J. Ezenarro, D. Schorn-García, L. Aceña, M. Mestres, O. Busto and R. Boqué

Universitat Rovira i Virgili. CHEMOSENS group, Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, Edifici N4, C/Marcel·lí Domingo 1, Tarragona, 43007, Spain

jokin.ezenarro@urv.cat

www.chemosens.recerca.urv.cat

Introduction

Spectroscopy plays a crucial role in various scientific fields, enabling researchers to analyze the composition and properties of materials. However, the success of spectroscopic data analysis heavily depends on the preprocessing applied to the data, which is a critical and often time-consuming step. By automating the selection of preprocessing algorithms and providing a comprehensive assessment of model performance, expertPLS (ePLS) offers researchers a more efficient and accurate approach, leading to more accurate and reliable results [1, 2].

Aim of study

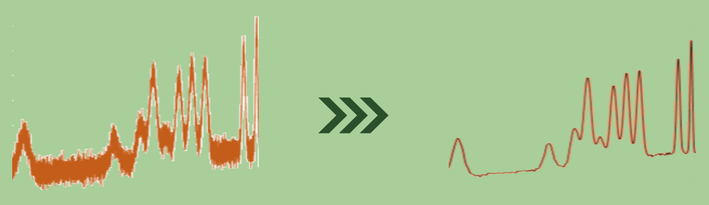
- To propose a toolbox, the expertPLS, that automates the search of the optimal preprocessing for spectroscopic data.
- To demonstrate the effectiveness and efficiency of expertPLS for analyzing spectroscopic data.
- To emphasize the advantages of using a reliable tool for reducing trial-and-error labor and improving the objectivity of the process.

Materials and Methods: Spectral preprocessing algorithms [3]

Figures: Keogan et al., *Translational Biophotonics,* 2021

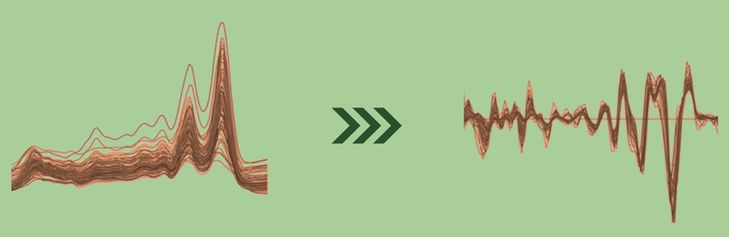
Smoothing

- Gaussian smoothing
- Savitzky-Golay smoothing
- Wavelet denoising



Derivative

- Differentiation
- Savtitzky-Golay derivative

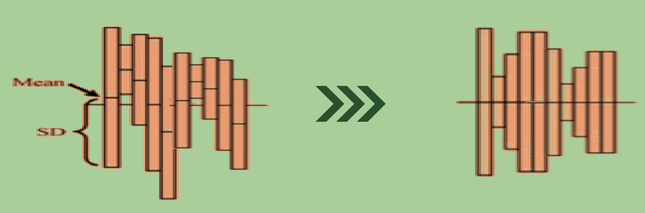


Sample-wise normalization

- Detrending
- Baseline correction
- Standard normal variate

Variable-wise normalization

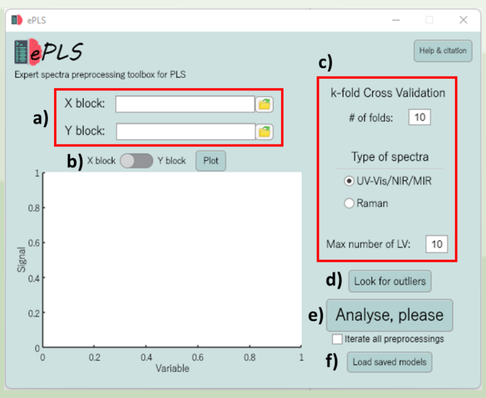
- Mean centering
- Autoscaling



Workflow

ePLS accompanies the analyst in the optimal spectra preprocessing selection process: after the data are loaded and analysed, the best PLS models are proposed to the analyst, from which they must choose based on expertise or relying on the choice of the toolbox.

1. Main window



a) Spectra and reference values are loaded.

b) Both can be visualized for a first inspection.

c) A model validation method is selected, also the type of spectra for an optimal smoothing (if needed).

d) Outliers can be removed. A model is calculated with raw data in order to visualize Q residuals vs. Hotelling's T^2 values of the samples.

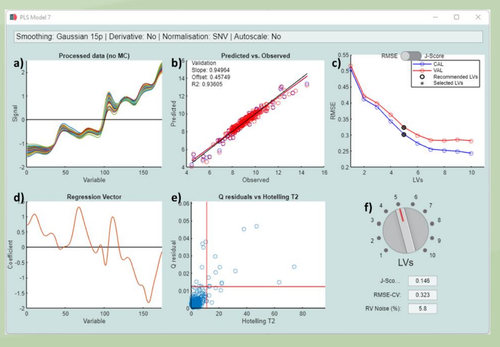
2. Results window

By following the decision chart (on the right) the ePLS creates all the datasets with a sensible preprocessing. Then a PLS model is created for each one and the figures of merit are shown in a table.

Preprocessed PLS models

Model	Smoothing	S. Window	Pol. Order	Derivative	Normalisation	N. Pol. Order	Limit	Autoscale	Sugg. LVs (J-Score)	J-Score	RMSE-CV	R2-CV	RV Noise
1	No			0	No			No	3	0.2380	0.3718	0.9155	5.4223
2	No			0	SNV			No	4	0.2402	0.2950	0.9474	14.6721
3	Gaussian	7		0	SNV			No	4	0.1967	0.2986	0.9463	10.5416
4	Gaussian	9		0	SNV			No	4	0.1834	0.2962	0.9455	9.0242
5	Gaussian	11		0	SNV			No	5	0.1647	0.2827	0.9510	9.6240
6	Gaussian	13		0	SNV			No	5	0.1535	0.2999	0.9451	6.6864
7	Gaussian	15		0	SNV			No	5	0.1452	0.2983	0.9453	5.8349
8	S-G	7	2	0	SNV			No	4	0.2195	0.2797	0.9518	15.2428
9	S-G	9	2	0	SNV			No	5	0.2027	0.2812	0.9517	13.8527
10	S-G	11	2	0	SNV			No	4	0.1899	0.2816	0.9519	12.1329
11	S-G	13	2	0	SNV			No	5	0.1762	0.2886	0.9505	9.2872
12	S-G	15	2	0	SNV			No	4	0.1621	0.2902	0.9492	7.9696
13	S-G	7	3	0	SNV			No	4	0.2229	0.2932	0.9480	13.3115
14	S-G	9	3	0	SNV			No	4	0.2059	0.2799	0.9521	13.8527
15	S-G	11	3	0	SNV			No	4	0.1875	0.2824	0.9521	12.1329
16	S-G	13	3	0	SNV			No	5	0.1757	0.2855	0.9505	9.2872

2. Model visualisation window



a) Spectra after preprocessing.

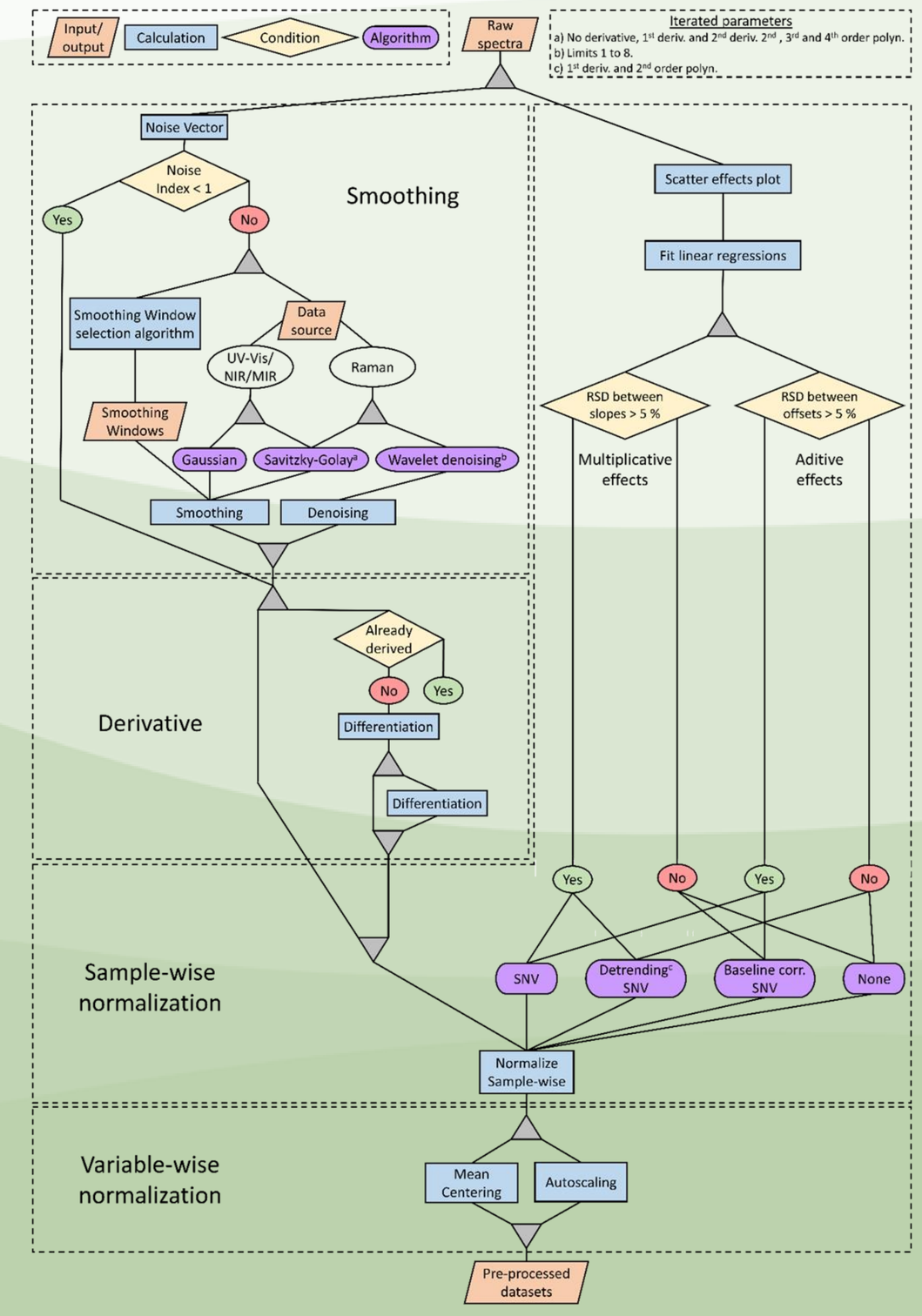
b) Predicted vs. measured **Y**-values.

c) Model performance vs. number of LVs. ($RMSE_{Cal}$ and $RMSE_{Val}$ or J-Score).

d) Coefficients of the regression vector.

e) Samples Q residuals vs Hotelling T^2 values with 95% confidence limits, to help detect outliers.

f) Wheel to change the number of LVs.



Conclusions

- The expertPLS toolbox automates the selection of preprocessing methods in spectroscopic data analysis, improving efficiency and objectivity.
- It incorporates various preprocessing techniques that include smoothing, derivatives, scatter correction, and normalization.
- The toolbox utilizes the J-Score to evaluate PLSR model performance, providing a comprehensive assessment to a non-expert analyst.

The expertPLS toolbox provides analysts with a user-friendly interface to streamline the preprocessing steps and improve the overall performance of PLSR models.

References

[1] H. Jonsson, J. Gabrielsson, Evaluation of Preprocessing Methods, in: S.D. Brown, B. Walczak, R. Tauler (Eds.), Comprehensive Chemometrics, Elsevier, 2009: pp. 199–206.

[2] J. Gerretzen, E. Szymańska, J.J. Jansen, J. Bart, H.J. Van Manen, E.R. Van Den Heuvel, L.M.C. Buydens, Simple and Effective Way for Data Preprocessing Selection Based on Design of Experiments, Anal Chem. 87 (2015) 12096–12103.

[3] A. Rinnan, F. Van Den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, Trends in Analytical Chemistry. 28 (2009) 1201–1222.

Acknowledgements

Grant PID2019-104269RR-C33 funded by MCI/AEI/10.13039/501100011033.

Grant URV Martí i Franqués – Banco Santander (2021PMF-BS-12; Ezenarro, J.).

This publication has been possible with the support of the Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya (2020 FISDU 00221; Schorn-Garcia, D.).

Poster Download