Using exponential moving average and PLS for UV-Vis spectra analysis

G. González-Aguilar

Centre for Applied Photonics, INESC TEC, Rua do Campo Alegre 687, Porto 4169-007, Portugal

Abstract

Spectroscopic analysis uses several technique to accurately locate the position of the local an general maxima in order to know about the characteristics of a chemical sample. Probably derivatives of the absorption with respect to wavelength is the most used technique. In this work the replacement of derivatives by exponential moving average (EMA) is made. We have found that applying EMAs to UV-Vis spectra are a good replacement form derivative spectroscopy in locating and quantifying compounds. EMAs are easy to calculate and give us similar results when comparing witth derivatives.

Keywords: EMA, UV-Vis spectroscopy, PLS

1. Introduction

Spectroscopy in general and UV-Vis in particular faces the problem of accurately identify and quantify a target of interest. However, one of the most important drawbacks of UV-Vis spectroscopy are the obstacles imposed when work with noising signals or overlapping peaks. These issues make difficult the accurate determination of spectroscopic parameters. For this reason, some sort of data treatment must be done. Over years the spectral processing have been an issue of concern for many researchers ref [1, 2].

The use of derivatives of the signal have been widely used in order to increase the signal-to-noise ratio and to resolve overlapping bands in simultaneous determination of multiple compounds [3, 4] . Derivatives defined as:

$$f'[x] = \lim_{h \to 0} \frac{f[x+h] - f[x]}{h} \tag{1}$$

or for practical purposes:

$$f'[i] = \frac{\Delta y}{\Delta x} = \frac{y[i-1] - y[i]}{x[i] - x[i-1]} \tag{2}$$

however, derivatives in analytical chemistry, derivatives or "derivative-like" are normally calculated by more sophisticated method such as splines, Fast Fourier Transforms (FFTs) or Savitzky-Golay interpolation algorithms [5, 6] . These method are very useful for analysing data *"a posteriori"* being useless for continuous on-line measurements.

The exponential moving average (EMA), defined as [7]:

$$ema_\alpha[i] = (1 - \alpha) * ema_\alpha[i-1] + \alpha * (x[i] - x[i-1]) \tag{3}$$

where the x's are measurements (actual "i" and previous "i-1") and $\alpha \epsilon$ [0,1] "the smoothing factor", is a feature which stresses changes in the trend of a signal. This parameter have been used to detect sensor responses with high accuracy [7]. In the process of getting useful data in uncontrolled environments, it is needed to have the support of more complex mathematical models.

However, derivatives alone does not provide us with quantifications, classifications or trends. That is way some sort of algorithms which perform to the above tasks are needed. classical Regression, Partial Component Analysis/Partial Component Regression [8] , or Partial Least Square/Multiregresion Least Square (among others) ref [8, 9, 10] are then used.

In chemical analysis Partial Least Square regression (PLSr) as a branch of Multiregresion Least Square is used as a two-block predictive model in the determination of the interdependence of the physical response and the sample composition i.e. Y   X. A principal feature of this method is the ability to extract more information about the system than the traditional multiple regression method[9]. Garthwaite have stated that this method "can be better than other methods at forming prediction equations when standard assumptions of regression analysis are satisfied"[11].

In this work we will examine the use of EMAs in the quantification of compounds with similar spectra. For this purpose PLS determination is

2

seem the most adequate method as it allows the simultaneous evaluation of multiple responses at a once.

Amine like compounds, comprising an primary amine (ethylamine) a secondary one (phenylethylmine) and the pyrrolizidine alkaloid monocrotaline having medidas a tertiary nitrogen were the target of this study. The choice of these structural differences in the nitrogen of the target compounds was made to testhow much robust the method is. monocrotaline is an important target because it is a contaminan in goat and dairy cow milk as well as in honey obtained from bees. These compounds were detected by its interactions with Bromothymol Blue. This last compound is a unespecific pH indicator such as, it will only indicates the basicities of the tested solutions, the simple visualization or single wavelength monitoring does not indicates the source of such a change, so its use can conduce to misleading classifications. However, the shape or small displacements of the positions of the bands in resulting spectra contain information about forces acting between molecules. Therefore, the use of mathematical algorithms is a factor helping the improvement of the analysis of the obtained response. However, the preparation of calibration curves is not objective of this work.

2. Materials and Methods

Bromothymol blue, monocrotaline, ethylamine and ethylphenylamine were obtained from Sigma-Aldrich. Several solutions, with different concentrations of the aminocompounds were prepared by mixing the appropriate amount of these compounds with a solution containing 0.3mg/mL of Bromothymol blue (Riedel-de Ha en) in water. The UV-Vis spectra were recorded in a Hitachi U-2010 spectrophotometer operated by Hitachi's UV Solution v2.0 package.

Data were processed in R statistical package by using the RStudio interface. The procedure for calculating EMAs was programmed in R during this work, even when exist an implementation of this method in the package TTR, we make our proper implementation to not charge the PC memory with more another package, and as is formulation is simple, The EMAs where calculated with different values of α: 0.3, 0.5 and 0.7. Derivatives were calculated by using the pspline library and the partial least square (PLS) were calculated using pls library both available in the standard R's CRAN repository [12]. In this case as are both complex procedures I have use the "as provided" tools to obtain the better possible results.

3. Results and Discussions

Figure

Figure

The obtained EMAs and derivatives were used to run PLSr calcuations. The calculations of the Root Mean Square Error of Prediction (RMSEP) for different number of components demonstrate the superior consistence of EMAs to predict the observed values when compared with splines derivatives as can be seen in Figure

The loading scores (Figure

The prediction of the values by PLS have more sucess when EMAs are used as can be seen in Figure

Table

4. Conlusions

Were used to different methods for predicting the concentration of three amine-based compounds. The method consisted in the used of two pre-processing algorithms and the further use of partial least square for calculating and posterior prediction of the concentration. The spline derivatives were more time-consuming than the EMAs. This last method have probed to be useful in searching for transient features in gas sensors. In this work we introduce it as a powerful instrument to analyse UV-Vis spectra when conjugated wit Partial Least Square regression.

5. References

References

[1] I. M. Dubrovkin. Spectrometry by recording derivatives (review). *J Appl Spectrosc*, 39(6):1341–1353, December 1983.

[2] Liudmil Antonov. Fourth derivative spectroscopy - a critical view. *Analytica Chimica Acta*, 349:295–301, August 1997.

[3] M. F. Merrick and H. L. Pardue. Evaluation of absorption and first- and second-derivative spectra for simultaneous quantification of bilirubin and hemoglobin. *Clin. Chem.*, 32(4):598–602, April 1986.

[4] J. J. Berzas Nevado and F. Guiberteau Cabanillas, C.and Salinas López. First Derivative of the Ratio Spectra Method for Resolving 3-Hydroxybenzaldehyde and 4-Hydroxybenzaldehyde in Binary Mixtures. *Bull. Soc. Chim. Belges*, 100(1):79–91, January 1991.

[5] Kritsana Jitmanee, Jaroon Jakmunee, Somchai Lapanantnoppakhun, Sunanta Wangkarn, Norio Teshima, Tadao Sakai, Gary D. Christian, and Kate Grudpan. Enhancing chemical analysis with signal derivatization using simple available software packages. *Microchemical Journal*, 86(2):195–203, August 2007.

[6] R. W. Schafer. What Is a Savitzky-Golay Filter? [Lecture Notes]. *IEEE Signal Processing Magazine*, 28(4):111–117, July 2011.

[7] Mehmet K. Muezzinoglu, Alexander Vergara, Ramón Huerta, and Mikhail I. Rabinovich. A sensor conditioning principle for odor identification. *Sensors and Actuators B: Chemical*, 146(2):472–476, April 2010.

[8] R. A. Viscarra Rossel, R. N. McGlynn, and A. B. McBratney. Determining the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy. *Geoderma*, 137(1-2):70–82, December 2006.

[9] Svante Wold, Michael Sjöström, and Lennart Eriksson. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2):109–130, October 2001.

[10] Sylwester Gawinkowski, Agnieszka Kamińska, Tomasz Roliński, and Jacek Waluk. A new algorithm for identification of components in a mixture: application to Raman spectra of solid amino acids. *Analyst*, 139(22):5755–5764, November 2014.

[11] Paul H. Garthwaite. An Interpretation of Partial Least Squares. *Journal of the American Statistical Association*, 89(425):122–127, 1994.

[12] R: The R Project for Statistical Computing.

List of Tables

Table 1: Exact values and calculate ones using derivatives and EMAs*

	Original			EMAα= 0.7						Derivatives					
	crota	ethylam	phenetam	crota	abs. err	ethylam	abs. err	phenetam	abs. err	crota	abs. err	ethylam	abs. err	phenetam	abs. err
	0.000	0.000	0.000	-0.0007	7.57E-4	0.00	1.66E-4	-0.002	0.002	0.022	0.02	-0.009	9.78E-3	0.028	0.03
	2.970	0.000	0.000	2.971	5.76E-4	0.00	1.27E-4	0.002	0.002	2.910	0.06	0.027	0.02	-0.076	0.07
	5.870	0.000	0.000	5.871	1.33E-3	0.00	2.93E-4	0.004	0.004	5.936	0.06	-0.029	0.03	0.083	0.01
	8.738	0.000	0.000	8.736	1.82E-3	0.00	3.99E-4	-0.005	0.005	8.695	0.04	0.019	0.02	-0.054	0.05
	11.540	0.000	0.000	11.541	1.05E-3	0.00	2.31E-4	0.003	0.003	11.561	0.02	-0.009	9.09E-3	0.026	0.02
	14.286	0.000	0.000	14.286	4.10E-4	0.00	8.92E-5	-0.001	0.001	14.277	9.17E-3	0.004	4.05E-3	-0.012	0.01
	16.980	0.000	0.000	16.980	3.30E-4	0.00	7.20E-4	0.001	0.001	16.974	6.21E-3	0.003	2.75E-3	-0.008	7.81E-3
	19.626	0.000	0.000	19.626	1.40E-4	0.00	8.84E-5	-0.001	0.001	19.646	0.02	-0.009	8.82E-3	0.025	0.02
	20.930	0.000	0.000	20.930	1.40E-4	0.00	0.000	-2.98E-5	2.98E-5	20.922	8.42E-3	0.004	-3.72E-3	-0.011	0.01
	22.220	0.000	0.000	22.220	2.00E-5	0.00	5.08E-6	0.000	0.000	22.217	2.95E-3	0.001	1.31E-3	-0.004	3.72E-3
	0.000	3.429	0.000	0.000	2.67E-4	3.43	5.90E-5	-0.001	0.001	0.013	0.001	3.423	5.56E-3	0.016	0.01
	0.000	6.822	0.000	0.000	4.13E-4	6.82	9.10E-5	0.001	0.001	-0.019	0.02	6.830	8.47E-3	-0.024	0.02
	0.000	10.182	0.000	0.000	1.32E-4	10.18	3.00E-5	0.000	0.000	0.005	5.24E-3	10.180	2.32E-3	0.007	6.60E-3
	0.000	13.510	0.000	0.000	1.46E-4	13.51	3.00E-5	0.000	0.000	0.010	9.54E-3	13.506	4.22E-3	0.012	0.01
	0.000	16.805	0.000	0.000	1.97E-4	16.80	4.00E-5	-0.001	0.001	-0.007	6.99E-3	16.808	-3.09E-3	-0.009	8.79E-3
	0.000	20.068	0.000	0.000	2.36E-4	20.07	5.00E-5	0.001	0.001	0.001	1.40E-3	20.067	6.19E-4	0.002	1.76E-3
	0.000	0.000	13.981	0.000	1.00E-5	0.00	2.21E-5	13.981	0.000	-0.003	3.11E-3	0.001	1.37E-3	13.977	3.91E-3
	0.000	0.000	16.232	0.000	5.97E-5	0.00	1.31E-5	16.232	0.000	0.001	1.49E-3	-0.001	6.58E-4	16.234	1.87E-3
	0.000	0.000	18.462	0.000	2.88E-4	0.00	6.32E-5	18.463	0.001	0.000	3.71E-4	0.000	1.64E-4	18.462	4.67E-4
	0.000	0.000	20.670	0.000	3.99E-4	0.00	8.77E-5	20.669	0.001	-0.004	3.87E-3	0.002	1.71E-3	20.665	4.87E-3
	0.000	0.000	22.857	0.001	5.07E-4	0.00	1.11E-4	22.858	0.001	0.002	1.94E-3	-0.001	8.57E-4	22.859	2.44E-3
	0.000	0.000	25.024	0.000	3.91E-4	0.00	8.60E-5	25.023	0.001	0.001	1.32E-3	-0.001	5.82E-4	25.026	1.66E-3
MAE					4.49E-4		9.84E-5		1.23E-3		1.48E-2		6.54E-3		1.86E-2

*Crota stands for monocrotaline, phenetam for phenylethylamine and ethylam for ethylamine, concentrations are in mg/ml

8

List of Figures

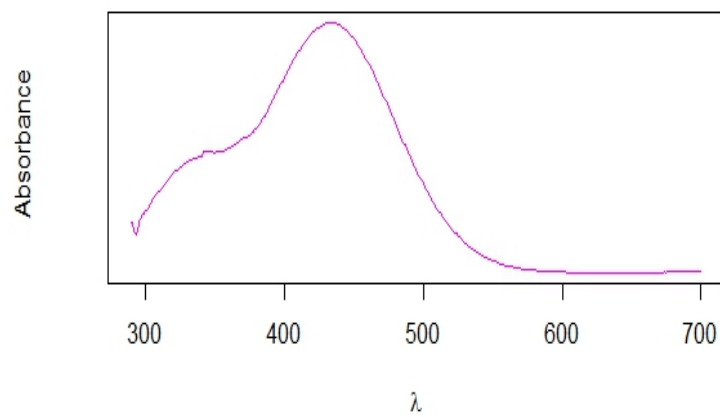Figure 1: UV-Vis spectrum of bromothymol blue c= 0.3mg/mL
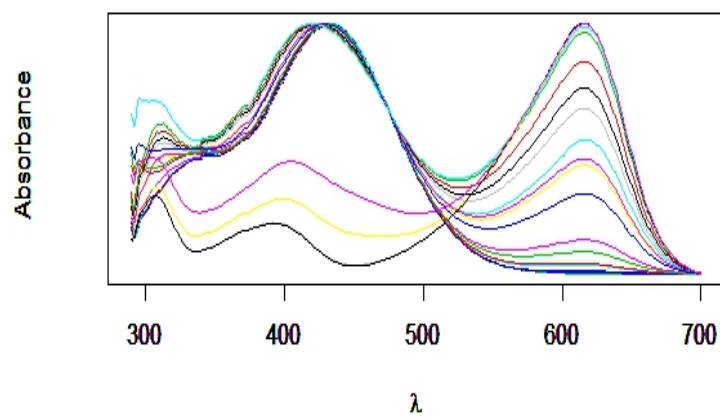

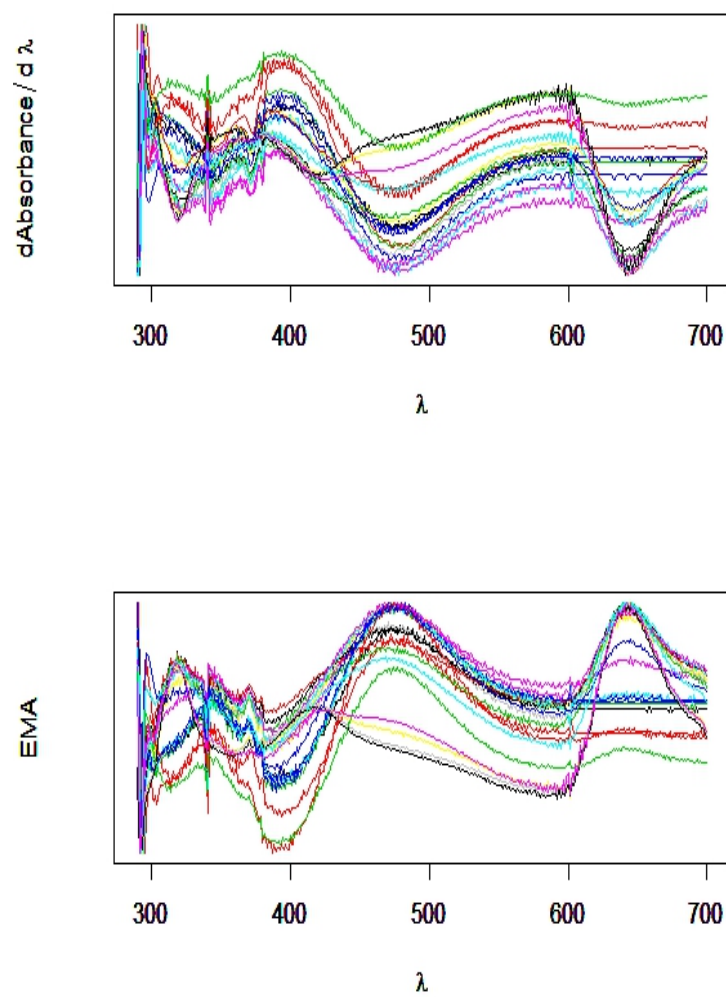Figure 2: UV-Vis of all spectra used in this work

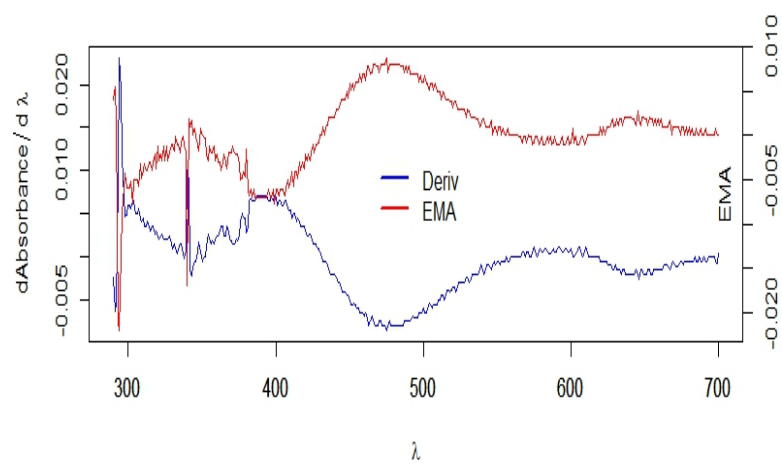Figure 3: Calculated spline derivatives (top) and EMAs (down)

Figure 4: Calculated spline derivative (red) and EMA (blue) for the sixth experiment (monocrotaline concentration was 14.286 mg/ml)
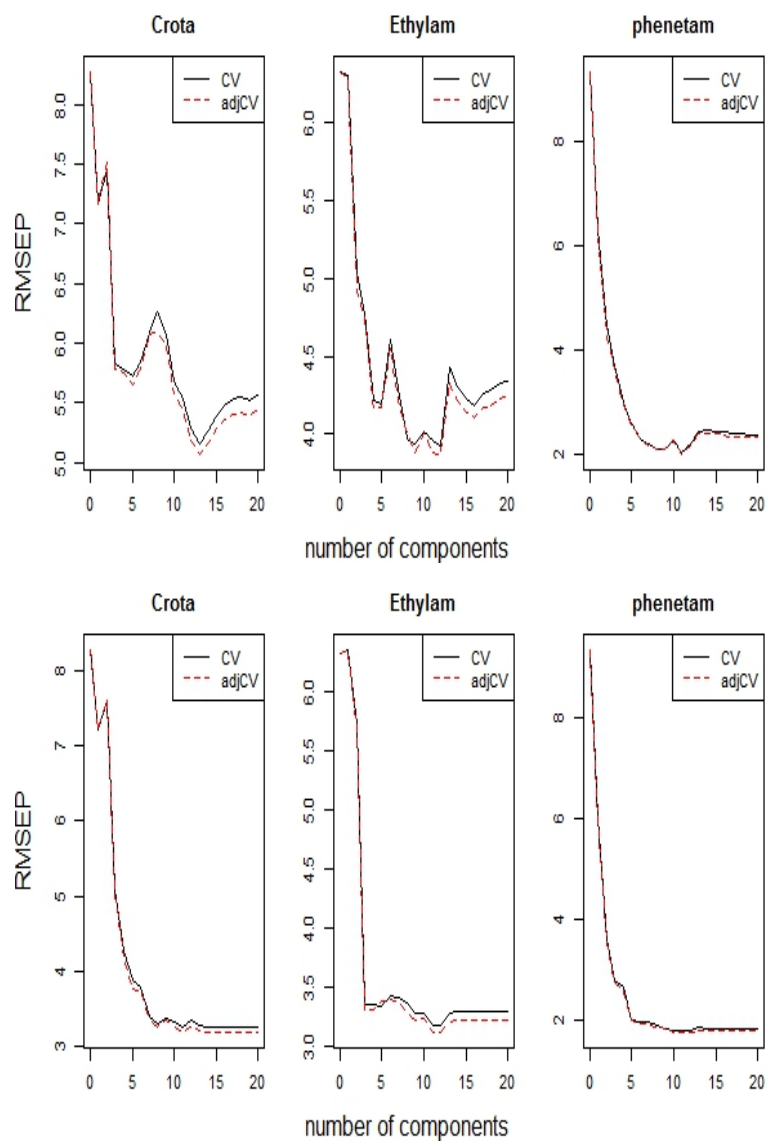
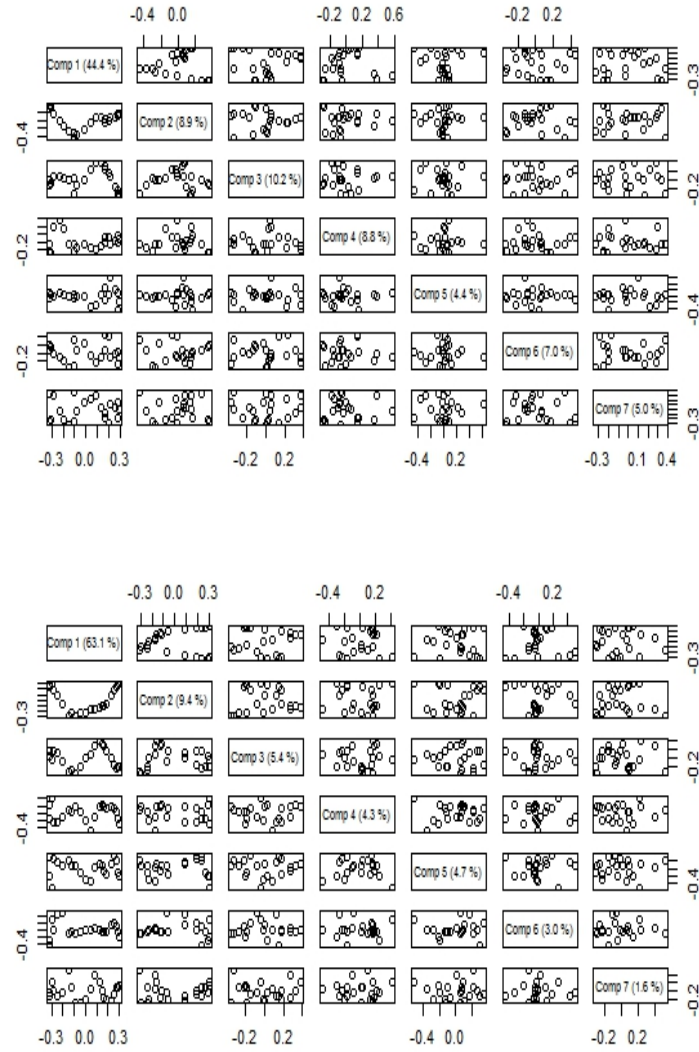Figure 5: RMSEP values for derivatives (top) and EMAs (down)

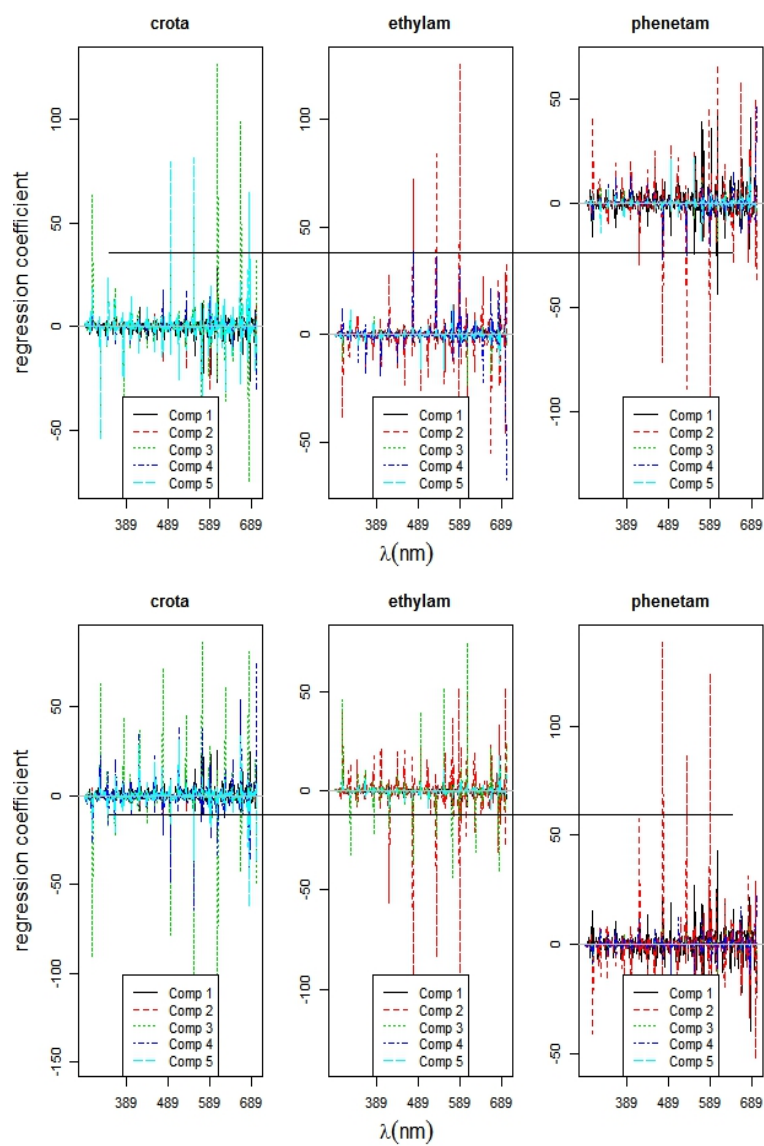Figure 6: Loading scores for derivatives (top) and EMAs (down)

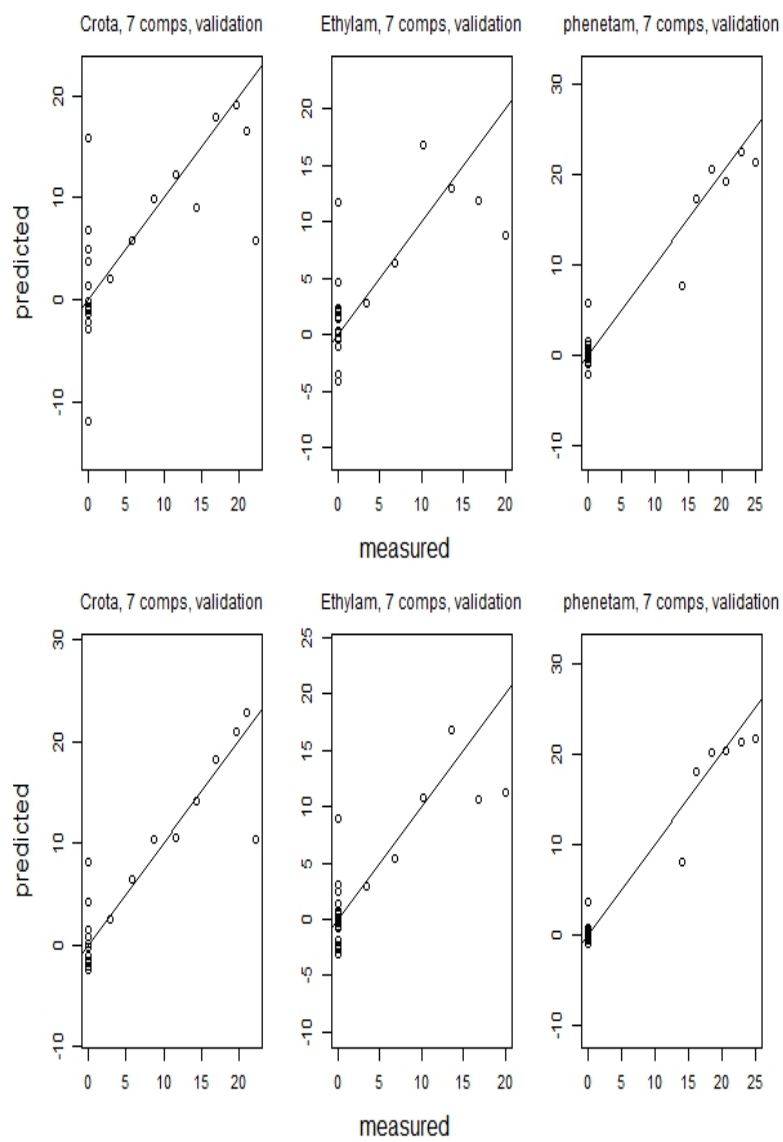Figure 7: Regression coefficients for the first fith variables for derivatives (top) and EMAs (down)

Figure 8: PLS-predicted vs measured values derivatives (top) and EMAs (down)