# Wine chemical characterization by Near Infrared Spectroscopy and chemometric analysis

Amaral, AL[1,2], Genisheva, Z[2], Quintelas, C[2], Mesquita, DP[2], Ferreira, EC[2] and Oliveira, JM[2]

[1]Instituto Politécnico de Coimbra, Instituto Superior de Engenharia de Coimbra, Rua Pedro Nunes, Quinta da Nora, 3030-199 Coimbra, Portugal
[2]CEB – Centre of Biological Engineering, University of Minho, 4710-057 Braga, Portugal

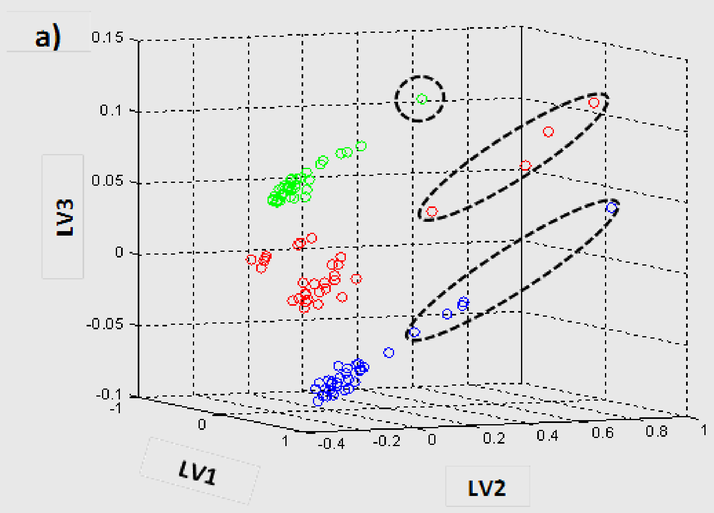
EUROANALYSIS 2017
Europe's Analytical Chemistry Meeting
28 August - 1 September 2017 • Stockholm Sweden

## SCOPE

- To examine the potential of near infrared spectroscopy (NIR) spectroscopy, a rapid and non-destructive technique with minimal sample processing prior to analysis, to determine the concentration of 10 different compounds in white wines. The collected NIR spectra ranged from 5435 $cm^{-1}$ to 6357 $cm^{-1}$.
- Initially a boxplot analysis, regarding the dependent variables (*Y*), was performed resulting in Y outliers identification and removal. Next, a PCA–X analysis was carried out, regarding the independent variables (*X*) for the identification of distinct clusters and possible *X* outliers.
- This led to 4 different datasets fed to the PLS analysis: [1] – ensemble with no *X* outliers removed; [2] – ensemble with *X* outliers removed; [3] – dataset divided in 3 clusters (1, 2 and 3) with no *X* outliers removed; and [4] – dataset divided in 3 clusters with *X* outliers removed.
- An iterative PLS method was then applied, first determining the weights of each wavelength, next grouping the wavelength values together according to weight similarity and, finally, recalculating the PLS with the averaged wavelength values.

## RESULTS AND CONCLUSIONS

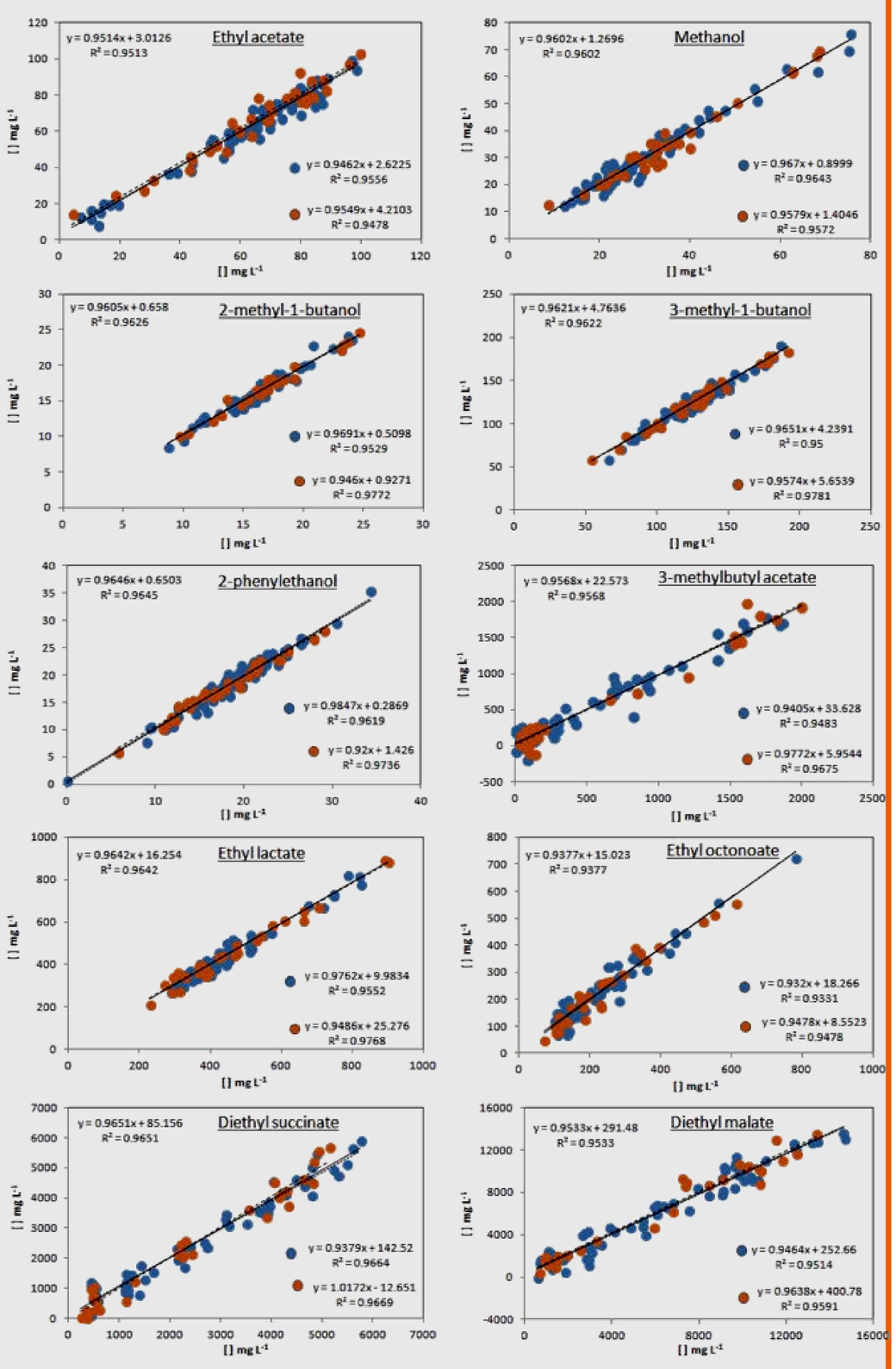
- The PCA analysis allowed to identify 3 different clusters
- The PLS analysis allowed to model all compounds with $R^2$ above 0.94 and *RMSE* values bellow 15% of the sample average values for 9 out of 10 compounds




PCA analysis of the *X* dataset

| Compound | [1] | | | [2] | | | [3] | | | [4] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $R^2$ | *RMSE* | *n* | $R^2$ | *RMSE* | *n* | $R^2$ | *RMSE* | $n_1, n_2, n_3$ | $R^2$ | *RMSE* | $n_1, n_2, n_3$ |
| ethyl acetate | 0.88 | 7.2 | 25 | 0.79 | 8.7 | 15 | **0.95** | **4.8** | **10, 10, 9** | 0.91 | 6.1 | 13, 9, 7 |
| methanol | 0.90 | 4.3 | 18 | 0.92 | 3.7 | 24 | **0.96** | **2.7** | **8, 10, 9** | 0.92 | 3.7 | 11, 7, 7 |
| 2-methyl-1-butanol | 0.96 | 0.7 | 23 | **0.96** | **0.6** | **19** | 0.92 | 0.9 | 9, 9, 9 | 0.91 | 0.9 | 11, 8, 7 |
| 3-methyl-1-butanol | **0.96** | **5.0** | **20** | 0.95 | 5.8 | 18 | 0.92 | 7.2 | 9, 9, 8 | 0.91 | 7.3 | 9, 8, 7 |
| 2-phenylethanol | **0.97** | **1.0** | **19** | 0.94 | 1.2 | 17 | 0.92 | 1.4 | 10, 7, 8 | 0.91 | 1.5 | 11, 9, 7 |
| 3-methylbutyl acetate | **0.96** | **120.1** | **21** | 0.96 | 120.1 | 27 | 0.95 | 134.4 | 10, 9, 8 | 0.83 | 221.8 | 11, 9, 6 |
| ethyl lactate | 0.90 | 44.4 | 18 | 0.94 | 35.0 | 21 | 0.95 | 32.2 | 9, 10, 7 | **0.96** | **27.0** | **12, 8, 7** |
| ethyl octanoate | 0.91 | 37.7 | 18 | **0.94** | **32.3** | **19** | 0.92 | 37.1 | 8, 8, 9 | 0.93 | 34.2 | 11, 9, 7 |
| diethyl succinate | **0.97** | **310.8** | **20** | 0.94 | 417.4 | 22 | 0.92 | 470.1 | 8, 8, 10 | 0.92 | 461.0 | 8, 9, 7 |
| diethyl malate | **0.95** | **916.1** | **16** | 0.95 | 859.6 | 20 | 0.93 | 1081.1 | 7, 8, 9 | 0.95 | 873.4 | 9, 7, 7 |

$R^2$, *RMSE* and number of PLS components (*n*) of the PLS analysis




Best model results for the studied compounds

This study was supported by FCT under the scope of the strategic funding of UID/BIO/04469/2013 unit and COMPETE 2020 (POCI-01-0145-FEDER-006684) and BioTecNorte operation (NORTE-01-0145-FEDER-000004) funded by the European Regional Development Fund under the scope of Norte2020 - Programa Operacional Regional do Norte. The authors also acknowledge the financial support to Cristina Quintelas through the postdoctoral grant (SFRH/BPD/101338/2014) and Zlatina Genisheva trough the postdoctoral grant (SFRH/BPD/108868/2015) provided by FCT .