IDC-improved direct calibration. Application to wine ethanol quantification in musts and wines

Jean Claude J. C. Boulet, Jean-Michel Roger

IDC-Improved Direct Calibration. Application to ethanol quantification in musts and wines

Jean-Claude BOULET*, Jean-Michel ROGER **
* INRA, UMR-SPO, F-Montpellier
** CEMAGREF, UMR-ITAP, F-Montpellier

Theory

$$\underset{P}{\mathbf{x'}} = y * \underset{P}{\mathbf{k'}} + \underset{Q}{\mathbf{T_\kappa}} * \underset{Q\;\;\;\;P}{\mathbf{K}} + \underset{A}{\mathbf{T_\varphi}} * \underset{A\;\;\;\;P}{\mathbf{P}}$$

Sample spectrum ⎵ Useful part ⎵ Chemical influence factors ⎵ Other influence factors

IDC principle :

to remove the influences from the useful part

- collect $\mathbf{X_G}$ (samples with constant $\mathbf{y}$)
- identify $\mathbf{P}$ with a SVD onto $\mathbf{X_G}$
- project onto $\mathbf{k}$ orthogonally to $\mathbf{K}$ and $\mathbf{P}$

$$\mathbf{R}=\begin{bmatrix}\mathbf{K}\\\mathbf{P}\end{bmatrix} \qquad \Sigma = \mathbf{I} - \mathbf{R'(RR')^{-1}R}$$

$$\mathbf{b} = \Sigma\ \mathbf{k}\ (\mathbf{k'}\,\Sigma\ \mathbf{k})^{-1}$$

Application

Quantification of Ethanol (% vol.) on clarified grape musts during alcoholic fermentation, using NIR spectrometry (500 to 1900 nm / 2nm)
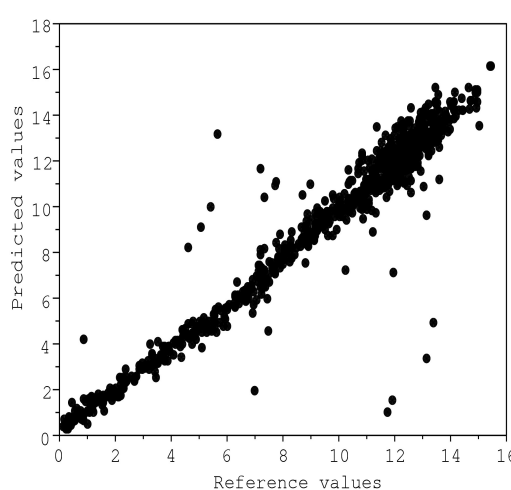
$\mathbf{k}$ = ethanol spectrum

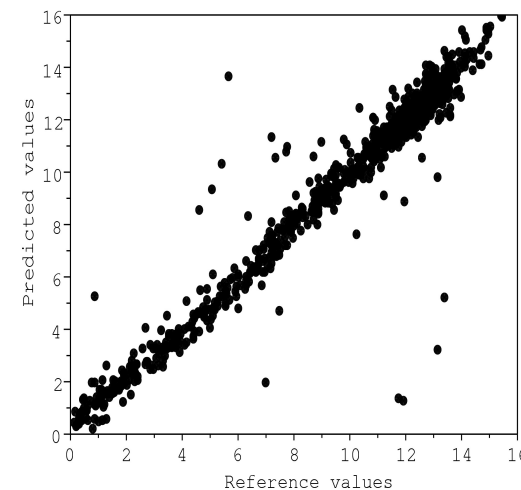$\mathbf{K} = \begin{bmatrix}\text{glycerol}\\\text{lactic acid}\\\text{water}\end{bmatrix}$ } reference = air

$\mathbf{X_G}$ = 165 must spectra
(y = 0 %vol.)

$\mathbf{X_{TEST}}$= 1000 spectra
(0 < y < 16 %vol)

} reference = water

IDC RMSEP=0.96 % vol.

PLSR RMSEP=0.89 %vol.

Discussion and conclusion

PRACTICAL ASPECTS

+ IDC doesn't need any calibration dataset (contrary to PLSR)
  → possible application to hyperspectral images

+ IDC is very flexible: chemical compounds whose pure spectra are unknown can be characterised by $\mathbf{X_G}$

+ Different ways to build $\mathbf{X_G}$ (see [1], [2], [3])

+ IDC predictions can be as accurate as PLSR predictions

− when $\mathbf{X_G}$ is calculated acccording to [2] or [3], constant but not zero influence factors are not taken into account, leading to IDC models with wrong slopes and/or offsets

THEORETICAL ASPECTS

The term $\Sigma\mathbf{k}$ corresponds to Lorber's defintion of the NAS-Net Analyte Signal ([4]):
« the net analyte signal may be computed as the part of its spectrum orthogonal to other coexisting constituants », extended all other -e.g. physical- influence factors.

Despite similar predictions, IDC and PLSR-b-coefficients vectors are different, and don't converge (not shown).

[1] P.W.HANSEN. Pre-processing methods minimizing the need for reference analyses. *Journal of Chemometrics*, 15:123-131, 2001
[2] J.M.ROGER, F.CHAUCHARD, V.BELLON-MAUREL.EPO-PLS external parameter orthogonalisation of PLS, application to temperature-independant measurement of sugar contents in fruits. *Chemometrics and Intelligent Laboratory Systems*, 66:191-204, 2003
[3] R.MARBACH. A new method for multivariate calibration. *Journal of Near Infrared Spectroscopy,* 13:241-254, 2005
[4] A.LORBER, K.FABER, B.R.KOWALSKI. Net analyte signal in multivariate calibration. *Analytical Chemistry,* 69:1620-1626, 1997