



**HAL**  
open science

## Un modèle commun à plusieurs méthodes d'étalonnage et de prétraitement

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

► **To cite this version:**

Jean Claude J. C. Boulet, Jean-Michel Roger. Un modèle commun à plusieurs méthodes d'étalonnage et de prétraitement. chimiométrie 2012, Dec 2012, Villeneuve d'Asc, France. pp.81-83. hal-02745386

**HAL Id: hal-02745386**

**<https://hal.inrae.fr/hal-02745386>**

Submitted on 3 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Un modèle commun à plusieurs méthodes d'étalonnage et de prétraitement

M. Boulet<sup>1</sup>, M. Roger<sup>2</sup>

<sup>1</sup>INRA, UMR1083, 2 place Viala, F-34060 Montpellier, bouletjc@supagro.inra.fr

<sup>2</sup>IRSTEA, UMR ITAP, 361 rue J.F. Breton, F-34093 Montpellier, jean-michel.roger@irstea.fr

**Mots-clef:** prétraitement, étalonnage, projection, orthogonale, oblique

## 1 Introduction

Dans le cadre multivarié, comme la spectrométrie, la modélisation linéaire a conduit à deux types de traitements : les prétraitements et les étalonnages. La multiplicité des méthodes est la conséquence de leur efficacité et de leur complémentarité. Nous proposons un modèle commun à plusieurs méthodes d'étalonnage et de prétraitements. Ce modèle est basé sur la gestion des informations présentes dans les spectres.

### 1.1 Espaces vectoriels, métriques

Un spectre acquis avec  $Q$  longueurs d'onde est défini dans  $\mathbb{R}^Q$ , un espace vectoriel de dimension  $Q$ . Supposons que  $\mathbb{R}^Q$  est muni d'une métrique  $M$  afin de mesurer distances et produits scalaires :  $(\mathbb{R}^Q, M)$  constitue un espace métrique.

### 1.2 Projections orthogonales, projections obliques

Une projection  $f$  est une application linéaire de  $\mathbb{R}^Q$  dans  $\mathbb{R}^Q$  qui vérifie:  $f^2 = f$ . La matrice  $\Pi$  de  $f$  vérifie:  $\Pi^2 = \Pi$ . Soit  $E^P$  un sous-espace vectoriel de  $\mathbb{R}^Q$  dont une base est constituée des vecteurs-colonne d'une matrice  $P$  de dimensions  $(Q \times A)$ . Le projecteur oblique  $\Pi$  projetant sur  $P$  en utilisant la métrique  $M$  est défini par:

$$\Pi = P(P'MP)^{-1}P'M \quad (1)$$

Quand  $M$  est la matrice identité, l'équation précédente se simplifie et donne la projection orthogonale classique :

$$\Pi = P(P'P)^{-1}P' \quad (2)$$

## 2 Theory

Soit un sous-espace vectoriel de  $\mathbb{R}^Q$  engendré par une base  $\{p_1, p_2, \dots, p_A\}$  de  $A$  vecteurs formant les colonnes d'une matrice  $P$  de dimensions  $(Q \times A)$ . Ce sous-espace vectoriel peut être utile ou nuisible. Soit  $X$  de dimensions  $(N \times Q)$  une matrice regroupant  $N$  observations. Alors l'information utile ou nuisible contenue dans  $X$  est extraite par projection oblique de  $X$  sur  $P$ , soit :

$$X^{U/N} = XMP(P'MP)^{-1}P' \quad (3)$$

Une fois isolée, l'information utile  $X^U$  ou l'information nuisible  $X^N$  est gérée différemment.

### 2.1 Cas des étalonnages : utilisation de $X^U$

Les coordonnées de l'information utile associée aux  $N$  observations dans la base  $\{p_1, p_2, \dots, p_A\}$  forment une matrice de scores  $T$  vérifiant:  $X^U = TP'$ , d'où :

$$T = XMP(P'MP)^{-1} \quad (4)$$

Supposons qu'un jeu d'étalonnage  $(X,y)$  est disponible. Alors une estimation de  $y$  est obtenue par projection orthogonale de  $y$  sur  $T$  :

$$\hat{y}=T(T'T)^{-1}T'y \quad (5)$$

Sachant que  $\hat{y}=Xb$ , les équations (4) et (5) donnent:

$$b=MP(P'MP)^{-1} [(P'MP)^{-1} P'MX'XMP (P'MP)^{-1}]^{-1} (P'MP)^{-1} P'MX'y \quad (6)$$

## 2.2 Cas des prétraitements: utilisation de $X^D$

L'information nuisible est soustraite de  $X$ , donnant  $X_c$  par projection de  $X$  M-oblique à  $P$  :

$$X_c=X-X^N=X(I_Q-MP (P'MP)^{-1}P') \quad (7)$$

## 3 Applications

### 3.1 Méthodes d'étalonnage inverse : les régressions

→ Méthode des moindres carrés ou *ordinary least squares regression* (OLSR).

Les b-coefficients sont [1]:  $b=(X'X)^{-1} X'y$ . Cette formule est obtenue de l'équation (6) en posant :  $P=M=I_Q$

→ Régression sur composantes principales (PCR).

Soit  $P$  la matrice des  $A$  premiers vecteurs propres de  $X$ . Les b-coefficients sont [1]:  $b=P(P'X'XP)^{-1} P'X'y$ . Cette formule est obtenue de l'équation (6) en posant :  $M=I_Q$ .

→ Régression *partial least squares* (PLSR)

Soit  $M=(X'X)^+$  la pseudo-inverse de  $X'X$  au sens de Moore-Penrose, et  $P$  la matrice des  $A$  premiers *loadings* de PLSR. Nous avons montré [2] que  $b$  de la PLSR standard était obtenu par :  $b=MP(P'MP)^{-1} P'MX'y$ . Cette équation est obtenue à partir de l'équation (6) en simplifiant grâce à une propriété des pseudo-inverses de Moore Penrose selon laquelle :  $MX'XM=M$ .

### 3.2 Méthodes d'étalonnage direct

Les méthodes d'étalonnage direct n'ont pas à proprement parler de jeu d'étalonnage. Toutefois elles nécessitent de connaître au moins le spectre pur  $k$  du composé à quantifier, dont la concentration est 1 par définition. Prenons donc  $(k,1)$  comme jeu d'étalonnage à la place de  $(X,y)$ . De même, la seule base d'information utile est apportée par  $k$ , donc remplaçons également  $P$  par  $k$ . L'équation (6) se simplifie et donne la formule classique des étalonnages directs :

$$b=Mk(k'Mk)^{-1} \quad (8)$$

→ Direct calibration [1] et orthogonal subspace projection [3].

Une matrice  $K$  contient les spectres purs dont l'influence est à éliminer.  $M$  est la projection orthogonale à  $K$  :  $M=I_Q-K(K'K)^{-1}K'$

→ Science based calibration [4].

Une matrice  $X_G$  contient une information nuisible, à éliminer ;  $M=(X'_G X_G)^{-1}$

### 3.3 Méthodes de prétraitement

Elles utilisent l'équation (7). Les différences portent sur l'identification de  $M$  et  $P$ .

→ Detrend [5]:  $M=I_Q$ ;  $P = [(1:Q)^0 (1:Q)^1 (1:Q)^2]$  représente les polynômes de degré 2.

→ Independant interference reduction [6], external parameter orthogonalization [7], transfer by orthogonal projection [8], error removal by orthogonal subtraction [9], dynamic orthogonal projection [10]. Ces méthodes identifient une matrice  $X_G$  d'information nuisible, puis condensent cette information par ACP donnant une matrice  $P$ . La métrique est  $M=I_Q$ .

→ Orthogonal signal correction (OSC). L'OSC de Fearn [11] s'écrit selon l'équation (7) avec  $M=(X'X)^+$  [12].

## 4 Discussion et conclusion

Cette approche montre que les principaux étalonnages et prétraitements partagent des points communs : ils sont définis avec deux matrices  $P$  et  $M$  représentant respectivement les vecteurs d'une base d'un sous-espace vectoriel utile ou nuisible et une métrique. La mise en œuvre est basée sur les projections orthogonales ou obliques. Les relations entre  $P$  et  $M$  sont discutées. L'introduction d'autres informations dans les deux matrices  $P$  et  $M$  ouvre des perspectives vers d'autres méthodes.

## 5 Références

- [1] H.Martens, T.Naes. Multivariate calibration. Wiley, 1989.
- [2] J.C. Boulet, D.Bertrand, G.Mazerolles, R.Sabatier, J.M.Roger. A family of regression methods derived from standard PLSR. Chemom. Intell. Lab. Syst., ressoumis, 2012.
- [3] Q.Du, C.I.Chang. A comparative study for orthogonal subspace projection and constrained energy minimization. IEEE-Trans.Geosci.Remote sens. 41(6),1525-1529, 2003.
- [4] R.Marbach. A new method for multivariate calibration. J. Near Infrared Spectrosc.,13, 241-254, 2005.
- [5] R.J.Barnes, M.S.Dhanoa, S.J.Lister. Standard normal variate and de-trending of near-infrared diffuse reflectance spectra. Appl.Spectrosc.,43, 772-777, 1989.
- [6] P.W.Hansen. Pre-processing method minimizing the need for reference analyses. J.Chemom., 15, 123-131, 2001.
- [7] J.M.Roger, F.Chauchard, V.Bellon-Maurel.EPO-PLS external parameter orthogonalization of PLS, application to temperature-independant measurement of sugar contents in fruits. Chemom. Intell. Lab. Syst., 66, 191-204, 2003.
- [8] A.Andrew, T.Fearn. Transfer by orthogonal projection : making near infra-red calibrations robust to between-instrument variation. Chemom. Intell. Lab. Syst., 72, 51-56, 2004.
- [9] Y.Zhu, T.Fearn, D.Samuel, A.Dhar, O.Hameed, S.G.Brown, L.B.Lovat. Error removal by orthogonal subtraction (EROS) : a customized pre-treatment for spectroscopic data. J. Chemom., 22, 130-134, 2008.
- [10] M.Zeaiter, J.M.Roger, V.Bellon-Maurel. Dynamic orthogonal projection, a new method to maintain the on-line robustness of multivariate calibration, application to NIR-based monitoring of wine fermentations. Chemom. Intell. Lab. Syst., 80, 227-235, 2006.
- [11] T.Fearn. On orthogonal signal correction. Chemom. Intell. Lab. Syst., 50, 47-52, 2000.
- [12] J.C.Boulet, J.M.Roger. Pretreatments by means of orthogonal projections. Chemom. Intell. Lab. Syst., doi 10.1016/j.chemolab.2012.02.002, 2012.