



HAL
open science

Pretreatments by means of orthogonal projections

J.C. Boulet, J.M. Roger

► **To cite this version:**

J.C. Boulet, J.M. Roger. Pretreatments by means of orthogonal projections. *Chemometrics and Intelligent Laboratory Systems*, 2012, 117, p. 61 - p. 69. 10.1016/j.chemolab.2012.02.002 . hal-00842452

HAL Id: hal-00842452

<https://hal.science/hal-00842452>

Submitted on 8 Jul 2013

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.

Pretreatments by means of orthogonal projections

Jean-Claude Boulet¹

INRA, UMR1083 Sciences Pour l'Oenologie, 2 place Viala, F-34060 Montpellier, France

Jean-Michel Roger²

IRSTEA, UMR ITAP Information Technologies Analyse environnementale Procédés agricoles, F-34196 Montpellier, France

Abstract

This article describes several linear pretreatments based on orthogonal projections. The main differences of these pretreatments lie in the way the information to be removed is identified, using calibration dataset, pure spectra, experimental designs or mathematical models. Removing all the undesired spectral information yields spectra proportional to the net analyte signal, so it is important to collect the most complete information possible, using the complementarities of different approaches. The correction should then be processed with a single Euclidian orthogonal projection that gathers all the information, rather than with successive operations. By embedding Euclidian orthogonal projections into the calibration, it is not necessary to reapply them to new datasets.

Keywords: orthogonal projection, pretreatment, preprocessing, subspace, linear model

1. Introduction

2 Spectroscopy has spread throughout many industries as an on-line process
3 control tool because calibration models are able to extract quantitative infor-
4 mation about a compound of interest from the spectra. Among the models
5 proposed, regressions or inverse calibrations such as partial least square re-
6 gression (PLSR) (1), extract the relevant spectral information by means of a
7 calibration dataset. The term "partial" recalls that just a few dimensions or
8 latent variables are used, and thus the information related to the other variables
9 is dropped. Pretreatments or preprocessings are positioned prior to calibration.
10 Their purpose is to identify and to remove spectral information that interferes

Email addresses: bouletjc@supagro.inra.fr (Jean-Claude Boulet),
jean-michel.roger@irstea.fr (Jean-Michel Roger)

¹Tel.(33)499613148 Fax (33)499612857 (corresponding author)

²Tel.(33)467046383 Fax (33)467166440

11 with the desired prediction. Pretreatments and regressions share a same ob-
12 jective, so pretreatments allow regressions to perform better. Many different
13 pretreatments are available. This paper focuses on describing the ones based
14 upon orthogonal projections and complements a recent review by Rinnan et al
15 (2). After introducing the notations, we describe several pretreatments focused
16 on orthogonal projections, and then discuss their properties. We propose a clar-
17 ified view of several pretreatments by putting forward their resemblances and
18 complementarities and suggesting the best methods for their use. The detailed
19 relationships of these pretreatments with other pretreatments (e.g. Savitsky-
20 Golay (SG), standard normal variate) and with other calibration methods are
21 outside the scope of this article.

22 1.1. Notations

23 Vectors are noted in bold lowercase, matrices in bold uppercase, and scalars
24 in uppercase characters. Vectors are arranged in columns, except in matrices \mathbf{X}
25 and \mathbf{X}_G where the lines represent the spectra. The transposed forms of vector
26 \mathbf{a} and matrix \mathbf{A} are noted \mathbf{a}' and \mathbf{A}' , respectively. The main notations are
27 gathered in Table 1, and a glossary is also available in Table 2.

28 2. Pretreatments based on orthogonal projections

29 Pretreatments based on orthogonal projections deal with the correction of
30 additive effects. For example, suppose that for sample i the observed spectrum
31 $\mathbf{x}_{i,obs}$ is the sum of the expected spectrum \mathbf{x}_i plus an unwanted contribution \mathbf{h}_i :

$$\mathbf{x}_{i,obs} = \mathbf{x}_i + \mathbf{h}_i \quad (1)$$

32 If a good estimation of \mathbf{h}_i is available, the first possibility would be to perform
33 a subtraction, and so \mathbf{x}_i is estimated as:

$$\hat{\mathbf{x}}_i = \mathbf{x}_{i,obs} - \hat{\mathbf{h}}_i$$

34 Unfortunately, because \mathbf{h}_i is not well estimated for each spectrum i , this
35 configuration is very uncommon in spectrometry. Nevertheless, it is possible to
36 obtain a good estimation of the subspace \mathcal{E}^D spanned by the different vectors
37 $\{\mathbf{h}_i\}$. Thus it becomes possible to build a projector orthogonal to this subspace.
38 Let \mathbf{P} be a matrix of dimensions $P \times A$ whose column-vectors $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_A\}$
39 form a basis of \mathcal{E}^D . Let \mathbf{I}_P be the identity matrix of dimensions $P \times P$. The
40 Euclidian orthogonal projector to \mathbf{P} is:

$$\mathcal{P}_P^\perp = \mathbf{I}_P - \mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'$$

41 A spectrum $\mathbf{x}_{i,corr}$ corrected from the information due to any \mathbf{h}_i is obtained
42 after a projection of $\mathbf{x}_{i,obs}$ orthogonally to \mathbf{P} :

$$\mathbf{x}_{i,corr} = \mathcal{P}_P^\perp \mathbf{x}_{i,obs} = \mathcal{P}_P^\perp \mathbf{x}_i \quad (2)$$

43 Note that the vector $\mathbf{x}_{i,corr}$ that is obtained after an orthogonal projection
44 is very different from $\hat{\mathbf{x}}_i$ obtained when a subtraction is possible. However in
45 both cases, the influence of \mathbf{h}_i has been reduced to nought. The orthogonal
46 projector \mathcal{P}_P^\perp is symmetrical: $\mathcal{P}'_P^\perp = \mathcal{P}_P^\perp$, so for N spectra forming the matrix
47 \mathbf{X} of dimensions $N \times P$:

$$\mathbf{X}_{corr} = \mathbf{X}\mathcal{P}_P^\perp = \mathbf{X}(\mathbf{I}_P - \mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}') \quad (3)$$

48 The performances of the different pretreatments are directly explained by
49 their ability to obtain a good approximation of a basis of \mathcal{E}^D . Different ap-
50 proaches are possible: using pure spectra, information extracted from experi-
51 mental design, models, and calibration datasets. For each method, matrix \mathbf{X}
52 represents centered or uncentered data, depending on the centering option cho-
53 sen. In order to simplify the presentation, all pretreatments presented here are
54 for correcting spectra in which just one compound of interest is to be quantified.
55 However, some pretreatments can also be written for the correction of several
56 compounds of interest.

57 *2.1. Pretreatment using pure spectra*

58 A basis of the space spanned by chemical components is given by their pure
59 spectra. A method derived from hyperspectral imaging uses this property.

60 *2.1.1. Orthogonal subspace projection*

61 The orthogonal subspace projection (OSP) use pure spectra, called undesired
62 signatures, which are associated with all the chemical influences present except
63 the one of the compound of interest. These undesired signatures form matrix \mathbf{K} .
64 They can be determined after a clustering process (3; 4), in which homogeneous
65 groups of spectra are obtained, followed by selection of a spectrum representative
66 of each group. However in Harsanyi et al (5) they were chosen within the image.
67 The OSP correction is a projection that is orthogonal to \mathbf{K} , in accordance with
68 equation 3:

$$\mathbf{X}_{OSP} = \mathbf{X}(\mathbf{I}_P - \mathbf{K}(\mathbf{K}'\mathbf{K})^{-1}\mathbf{K}')$$

69 In Harsanyi et al (5), the OSP method was applied to an hyperspectral im-
70 age from an airborne VIS-IR spectrometer using the radiance spectra directly.
71 Several end members were identified and alternatively chosen as the compound
72 of interest. An OSP was performed for each endmember. For each OSP, the
73 corrected spectra were used to classify the pixels, and the results were in accor-
74 dance with the measured values, or ground truth. However, the limits of this
75 method are such that all the pure spectra must be known in advance and they
76 cannot be collinear. In addition, influences such as temperature are not taken
77 into account.

78 *2.2. Pretreatments using spectra issued from an experimental design*

79 If pure spectra are not available due to the chemical complexity of the sam-
80 ples or because the influence to be removed is physical and no pure spectrum
81 exists, OSP cannot be applied. It is possible to construct experimental designs
82 to obtain a matrix \mathbf{X}_G whose spectra contain targeted spectral perturbations
83 without any useful information. A singular value decomposition (SVD) or a
84 principal component analysis (PCA) applied to \mathbf{X}_G gives a matrix of eigenvectors
85 \mathbf{P} of dimensions $(P \times A)$ whose columns represent an orthonormal basis of
86 the subspace to be removed. The matrix \mathbf{X} is corrected to \mathbf{X}_{corr} by a projection
87 orthogonal to \mathbf{P} . The following formula is the same as equation 3 and can be
88 simplified to:

$$\mathbf{X}_{corr} = \mathbf{X}(\mathbf{I}_P - \mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}') = \mathbf{X}(\mathbf{I}_P - \mathbf{P}\mathbf{P}')$$

89 Several methods have been based on this principle, but they differ in the
90 way \mathbf{X}_G is obtained, and in how the dimension of the SVD or the PCA applied
91 to \mathbf{X}_G is determined.

92 *2.2.1. Independent interference reduction*

93 The independent interference reduction method (IIR), Hansen (6) uses spectra
94 from samples where the compound of interest is null and these spectra are
95 gathered into the matrix \mathbf{X}_G . The IIR method implies that such samples should
96 be easy to collect in large numbers. This was the case in the reported applica-
97 tion, due to the absence of acetone for the milk of healthy cows. IIR yielded
98 models that were more interpretable and predictions that were more stable. IIR
99 removes more interferences with higher numbers of PCA factors and thus the
100 PLSR needs fewer latent variables.

101 *2.2.2. External parameter orthogonalization and transfer orthogonal projection*

102 External parameter orthogonalization (EPO), Roger et al (7) and transfer
103 orthogonal projection (TOP), Andrew et al (8) are two closely related methods.
104 For the same set of M samples, M spectra are acquired at R levels of one phys-
105 ical influence. Thus, each sample is associated with R spectra. Centering each
106 set of R spectra removes chemical information to leave only the information
107 from the physical influence remaining. Matrix \mathbf{X}_G of dimensions $(MR \times P)$
108 is obtained by merging the R centered spectra from each of the M samples.
109 EPO was first applied to correct temperature effects when predicting the sugar
110 concentration in apples in the near infrared. Ten apples were set at different
111 temperatures between 5°C and 40°C. The spectra of the ten apples were centered
112 for each temperature level, and then merged, to yield \mathbf{X}_G . The PCA and PLSR
113 dimensions were determined using either cross validation or Wilks Λ . The EPO
114 method produced a dramatic drop in the root mean square error of prediction
115 ($RMSEP$) value compared to that obtained without any pretreatment. TOP
116 aims at transferring calibrations from one instrument to another, and was used
117 in two applications: to determine the protein content in barley and moisture

118 in corn. Near infrared (NIR) spectra of the same five samples of barley were
119 acquired onto seven spectrometers, and the NIR spectra of five samples of corn
120 were acquired onto three spectrometers. Spectra were centered for each instru-
121 ment, and then merged into two matrices \mathbf{X}_G associated with barley and corn
122 respectively. The PCA dimensions were determined such that 98 – 99 % of the
123 variance was captured. For barley and corn, the TOP-PCR and TOP-PLSR
124 models were determined on one instrument and then applied to the others. The
125 low *RMSEP* values produced by the analysis demonstrated the value of using
126 TOP for calibration transfer. Calibration was also successfully transferred to an
127 unseen instrument (not used for the calculation of TOP).

128 2.2.3. Dynamic orthogonal projection

129 Dynamic orthogonal projection (DOP), Zeaiter et al (9) was inspired by
130 EPO. The aim of DOP is the on-line correction of unexpected disruptive influ-
131 ences. The spectra and reference values of a few disturbed samples are assumed
132 to be known. Let (\mathbf{X}, \mathbf{y}) be the calibration dataset, \mathbf{x}_1 a spectrum acquired after
133 the appearance of a disruptive influence, and y_1 the reference value associated
134 with \mathbf{x}_1 . The ideal spectrum $\hat{\mathbf{x}}_1$ which would have been obtained without the
135 disturbance is estimated by a kernel function or weighted mean of spectra from
136 \mathbf{X} chosen for their proximity to y_1 . The difference between the spectra $(\hat{\mathbf{x}}_1 -$
137 $\mathbf{x}_1)$ characterizes the disturbance. The same operation is repeated with other
138 measurement points $\mathbf{x}_2, \mathbf{x}_3$, etc., and spectral differences are gathered into \mathbf{X}_G .
139 The method was applied to the prediction of ethanol in wine fermentations, and
140 aimed at correcting the temperature effects. A calibration database was first
141 obtained during an isothermal fermentation. Then, temperature was monitored
142 during a second fermentation. Five reference points were acquired during the
143 increase of temperature and the NIR predictions of ethanol were corrected using
144 DOP. A Gaussian kernel function was used to build virtual spectra of the avail-
145 able reference points, and the database was successively corrected from one to
146 five points. Ethanol predictions were well-corrected by DOP, even at the end of
147 the fermentation when the temperature had decreased to its initial value. It was
148 also shown that DOP was able to capture the vertical (baseline) and horizontal
149 (wavelength) shifts due to the temperature.

150 2.2.4. Error removal by orthogonal subtraction

151 Error removal by orthogonal subtraction (EROS), Zhu et al (10) is a method
152 derived from TOP. EROS takes into account and then corrects variations due
153 to repetitions. Several spectra acquired from a same sample are centered. The
154 centered spectra associated with the different samples are gathered into a ma-
155 trix \mathbf{X}_G which contains only information due to the repetitions and in which the
156 row mean is null. EROS was applied to the diagnosis of precancerous polyps
157 using visible-NIR spectroscopy. The repeatability of the spectra was poor due
158 to different angles, pressures or locations at which the optical probe was used
159 during the *in vivo* data acquisitions. EROS was able to correct these influences
160 using five repetitions for each analysis. Normal and precancerous polyps were
161 classified by principal component discriminant analysis (PCDA), the dimensions

162 were tuned using cross-validation. EROS enabled the accuracy of the classifica-
163 tion to be slightly increased. However, the models generated were expected to
164 be more robust because they had been simplified and because EROS+PCDA
165 needed fewer components than PCDA alone.

166 2.3. Pretreatments using a polynomial model

167 Let λ be the vector of dimensions $P \times 1$ composed of values $\{v_1, v_2, \dots, v_P\}$.
168 The i° column of the matrix Λ_R of dimensions $(P \times R + 1)$ is obtained by raising
169 each term of λ to the power i . A vector \mathbf{z} can be written in a polynomial form
170 if it exists as vector \mathbf{a} such that:

$$\mathbf{z} = \Lambda_R \mathbf{a} \quad (4)$$

171 If an observed spectrum \mathbf{x}_{obs} is the sum of the expected spectrum \mathbf{x} plus the
172 contribution of an unexpected influence (e.g. scattering), which is represented
173 by a polynomial vector \mathbf{z} , then:

$$\mathbf{x}_{obs} = \mathbf{x} + \mathbf{z} = \mathbf{x} + \Lambda_R \mathbf{a}$$

174 An estimation of \mathbf{x} , corrected from the polynomial, is obtained after \mathbf{x} is
175 projected orthogonally to Λ_R :

$$\hat{\mathbf{x}}_{corr} = (\mathbf{I}_P - \Lambda_R (\Lambda_R' \Lambda_R)^{-1} \Lambda_R) \mathbf{x}_{obs}$$

176 The following pretreatments: detrend, constrained principal spectra analysis
177 (CPSA) and iterative polynomial fitting, are based on polynomial corrections.

178 2.3.1. Detrend

179 Detrend was proposed by Barnes et al(11; 12) to correct baselin distortions
180 due to scattering effects able to be modeled by a polynomial. These situations
181 are very common in spectroscopy and are difficult to model due to the variabil-
182 ity of particle sizes (2). Detrend calculates a matrix Λ_2 of dimensions $(P \times 3)$
183 as described above, to model second-order polynomials. A detrend correction
184 consists of an orthogonal projection of the raw spectra to Λ_2 . Detrend was
185 applied to spectra from crystalline and powder forms of sucrose (11), and the
186 corrected spectra of the same component were nearly identical. However, as
187 expected for an orthogonal projection, they had lost the classical shape of the
188 sucrose spectra. Other authors have proposed modeling baselines with polyno-
189 mials having orders up to two (13; 14) and with solutions for the choice of the
190 polynomial order. However the corrections were done using subtraction and an
191 orthogonal projection was also possible, at least for the spectra.

192 *2.3.2. Constrained principal spectra analysis*

193 CPSA addresses the correction of influences unrelated to the chemical com-
 194 position of the sample (15). As for detrend, scattering is modeled by a matrix
 195 Λ_2 . Spectra from external influences such as water and carbon dioxide gas,
 196 form matrix \mathbf{K} . Matrices Λ_2 and \mathbf{K} are merged to yield matrix \mathbf{R} . The CPSA
 197 correction consists of projecting raw spectra \mathbf{x}_i orthogonally to \mathbf{R} .

198 *2.4. Pretreatments using data compression methods*

199 The main goal of data compression methods is to obtain fewer numbers of
 200 new variables A such that $A \ll P$, where P is the dimension of a spectrum.
 201 These variables form the best basis (BB) (16), in that they span the subspace
 202 of \mathbb{R}^P containing the most relevant information. However a consequence of this
 203 pretreatment is that information is dropped while projecting onto the BB.

204 Let \mathbf{W} be a $P \times P$ matrix in which column-vectors constitute an orthonormal
 205 basis of \mathbb{R}^P . Thus $\mathbf{W}'\mathbf{W} = \mathbf{I}_P$. The projection of any vector \mathbf{x}_i of dimension
 206 P onto \mathbf{W} is invariant and gives \mathbf{x}_i :

$$\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{x}_i = \mathbf{W}\mathbf{W}'\mathbf{x}_i = \mathbf{x}_i \quad (5)$$

207 The scores of \mathbf{x}_i in the basis represented by the column-vectors of \mathbf{W} are
 208 represented by a unique vector \mathbf{t}_i , which by definition verifies:

$$\mathbf{x}_i = \mathbf{W}\mathbf{t}_i \quad (6)$$

209 Thus, from equations 5 and 6:

$$\mathbf{t}_i = \mathbf{W}'\mathbf{x}_i$$

210 \mathbf{W} is split into two matrices: $\mathbf{W}_{1:A}$ containing A selected vectors of \mathbf{W} (the
 211 BB); and $\mathbf{W}_{A+1:P}$ containing the last $P - A$ vectors of \mathbf{W} . The spectrum \mathbf{x}_i is
 212 rebuilt using $\mathbf{W}_{1:A}$ and the A first scores of \mathbf{t}_i , noted $\mathbf{t}_{i,1:A}$, into equation 6:

$$\mathbf{x}_{i,corr} = \mathbf{W}_{1:A}\mathbf{t}_{i,1:A} = \mathbf{W}_{1:A}\mathbf{W}'_{1:A}\mathbf{x}_i$$

213 It can easily be deduced that $\mathbf{x}_{i,corr}$ is also the orthogonal projection of \mathbf{x}_i
 214 to $\mathbf{W}_{A+1:P}$:

$$\mathbf{x}_{i,corr} = (\mathbf{I}_P - \mathbf{W}_{A+1:P}\mathbf{W}'_{A+1:P})\mathbf{x}_i$$

215 The equation obtained is equivalent to equation 2 with $\mathbf{W}_{A+1:P}$ replacing
 216 \mathbf{P} . The $\mathbf{W}_{1:A}$ and $\mathbf{W}_{A+1:P}$ matrices can be obtained using several compression
 217 methods. The simplest way is to perform a PCA, then select respectively the
 218 A first and $P - A$ last components. Discrete Fourier transform (DFT) is used
 219 in signal processing. However in chemometrics more attention has been paid to
 220 wavelet transforms. Thus a discrete wavelet transform (DWT) and a wavelet
 221 packet transform (WPT) offer alternative solutions.

222 *2.5. The wavelet packet transform and the discrete wavelet transform*

223 All of the vectors of \mathbf{W} are calculated using a WPT, whereas only a few of
224 them are obtained with DWT (16). Thus DWT is a particular case of WPT.

225 Two parameters must be determined for a WPT: (1) an orthonormal wavelet
226 basis of \mathbb{R}^P , represented by the columns of \mathbf{W} ; and (2) the A vectors from \mathbf{W}
227 which form the BB (16). Examples of orthonormal wavelet families of bases
228 are Haar-Daubechie, Symmlet, and Coiflet (16; 17; 18). The vectors forming
229 the BB are determined using a threshold. Several methods are available to
230 obtain the value of the threshold: predetermined (16); minimum description
231 length (MDL), which is a compromise between A and the percent of spectral
232 reconstruction (16; 19); and universal thresholding (ThU) (20). Cross-validation
233 on a calibration dataset (\mathbf{X}, \mathbf{y}) can help to determine these parameters (19; 21).
234 A WPT-DWT can be used for data compression and for smoothing, denoising
235 and baseline corrections. Smoothing consists of removing high frequency signals
236 (16; 19; 22), regardless of their amplitudes (23). Denoising consists in removing
237 the lower coefficients in the frequency domain, regardless of their frequencies
238 (23). Smoothing and denoising are different, but high frequencies are often
239 associated with low coefficients, so in practice both denoising and smoothing
240 are performed in the same operation. Baseline can also be removed. Hu et al
241 (20) applied DWT to Raman spectra for baseline and noise corrections.

242 DWT-WPT performance can be very good in terms of signal reconstruction
243 and data compression. Barclay et al obtained corrections which overwhelmed
244 the DFT and the SG method (23). Trygg et al obtained quite the same in-
245 formation after a compression rate of 30 (17). However, when associated with
246 PLSR, DWT-WPT did not significantly improve the quality of prediction (17).

247 *2.6. Pretreatments using a calibration dataset*

248 Orthogonal signal correction (OSC) methods are based on a calibration
249 dataset (\mathbf{X}, \mathbf{y}) and do not need any additional information. Several methods
250 have been proposed for removing the spectral information orthogonal to \mathbf{y} from
251 \mathbf{X} : Wold et al OSC (24); Sjoblom et al OSC (25); Wise et al OSC (cited in (26));
252 Andersson direct orthogonalization (DO) (27); Fearn's OSC (28); Westerhuis et
253 al direct orthogonal signal correction (DOSC) (26); Goicoechea net analyte pre-
254 processing (NAP) (29); and Trygg orthogonal projection to latent structures
255 (OPLS) (30). These methods have been reviewed previously (26; 31) and have
256 been cited extensively in the litterature. OSC methods are not reviewed here
257 but will be discussed briefly to clarify their relationships with orthogonal pro-
258 jections.

259 *2.6.1. Overview of the OSC methods*

260 All OSC methods aim at determining scores \mathbf{T} containing information or-
261 thogonal to \mathbf{y} . The correction can be written as an orthogonal projection into
262 \mathbb{R}^N :

$$\mathbf{X}_{OSC} = (\mathbf{I}_N - \mathbf{T}(\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}')\mathbf{X} \quad (7)$$

263 The methods differ in the way \mathbf{T} is obtained.

264 • Wold's OSC

$$\begin{aligned}
 \mathbf{X}_{1:0} &= \mathbf{X}; \text{ for step } i : \\
 \mathbf{t} &= \text{first score of a PCA onto } \mathbf{X}_{1:i-1} \\
 \mathbf{t}_o &= (\mathbf{I}_N - \mathbf{y}(\mathbf{y}'\mathbf{y})^{-1}\mathbf{y}')\mathbf{t} \\
 \mathbf{T}_{PLSR} &= \text{first scores of a PLSR calculated with } (\mathbf{X}_{1:i-1}, \mathbf{t}_o) \\
 \mathbf{t}_{o,new} &= \mathbf{T}_{PLSR}(\mathbf{T}'_{PLSR}\mathbf{T}_{PLSR})^{-1}\mathbf{T}'_{PLSR}\mathbf{t}_o \\
 \mathbf{t} &= \mathbf{t}_{o,new} \text{ then repeat until convergence of } \mathbf{t} \rightarrow \mathbf{t}_i \\
 \mathbf{p}_i &= \mathbf{X}'_{1:i-1}\mathbf{t}_i(\mathbf{t}'_i\mathbf{t}_i)^{-1} \\
 \mathbf{X}_{1:i} &= (\mathbf{I}_N - \mathbf{t}_i(\mathbf{t}'_i\mathbf{t}_i)^{-1}\mathbf{t}'_i)\mathbf{X}_{1:i-1}
 \end{aligned}$$

265 At each loop, the projection of \mathbf{t}_o onto \mathbf{T}_{PLSR} means that $\mathbf{t}_{o,new}$ lies in
 266 the subspace spanned by the columns of $\mathbf{X}_{1:i-1}$. Thus the \mathbf{t}_i are orthog-
 267 onal. After A iterations, they form matrix \mathbf{T} and the OSC correction is
 268 performed according to equation 7.

269 • Sjoblom and Wise OSCs

$$\begin{aligned}
 \mathbf{X}_{1:0} &= \mathbf{X}; \text{ for step } i : \\
 \mathbf{t} &= \text{first score vector of a PCA onto } \mathbf{X}_{1:i-1} \\
 \mathbf{t}_o &= (\mathbf{I}_N - \mathbf{y}(\mathbf{y}'\mathbf{y})^{-1}\mathbf{y}')\mathbf{t} \\
 \mathbf{w}_i &= \mathbf{X}'_{1:i-1}\mathbf{t}_o \text{ then } \mathbf{w}_i \text{ is normed} \\
 \mathbf{t}_1 &= \mathbf{X}_{1:i-1}\mathbf{w}_i \\
 \mathbf{t} &= \mathbf{t}_1 \text{ then repeat until convergence of } \mathbf{t}_1 \\
 \mathbf{T}_{PLSR} &= \text{first scores of a PLSR calculated with } (\mathbf{X}_{1:i-1}, \mathbf{t}_1) \\
 \mathbf{t}_i &= \mathbf{T}_{PLSR}(\mathbf{T}'_{PLSR}\mathbf{T}_{PLSR})^{-1}\mathbf{T}'_{PLSR}\mathbf{t}_1 \\
 \mathbf{p}_i &= \mathbf{X}'_{1:i-1}\mathbf{t}_i(\mathbf{t}'_i\mathbf{t}_i)^{-1} \\
 \mathbf{X}_{1:i} &= (\mathbf{I}_N - \mathbf{t}_i(\mathbf{t}'_i\mathbf{t}_i)^{-1}\mathbf{t}'_i)\mathbf{X}_{1:i-1}
 \end{aligned}$$

270 This is Sjoblom's OSC. Wise' OSC is similar, with an additional step
 271 consisting of orthogonalizing \mathbf{t}_i to \mathbf{y} (26). By construction each score-
 272 vector \mathbf{t}_i lies in the $\mathbf{X}_{1:i-1}$ subspace, so the \mathbf{t}_i vectors are orthogonal. After
 273 A iterations, they form matrix \mathbf{T} and the OSC correction is performed into
 274 \mathbb{R}^N according to equation 7.

275 • Andersson's DO and Goicoechea's NAP

$$\begin{aligned}\mathbf{Z} &= (\mathbf{I}_N - \mathbf{y}(\mathbf{y}'\mathbf{y})^{-1}\mathbf{y}')\mathbf{X} \\ \mathbf{P} &= A \text{ first loadings of a PCR calculated with } \mathbf{Z} \\ \mathbf{T} &= A \text{ first scores of a PCR calculated with } \mathbf{Z}\end{aligned}$$

276 DO and NAP are the same method. The correction is be performed into
 277 \mathbb{R}^N according to equation 7, but also into \mathbb{R}^P using \mathbf{P} :

$$\mathbf{X}_{DO-NAP} = \mathbf{X}(\mathbf{I}_P - \mathbf{P}\mathbf{P}')$$

278 • Fearn's OSC

$$\begin{aligned}\mathbf{M} &= \mathbf{I}_P - \mathbf{X}'\mathbf{y}(\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{y})^{-1}\mathbf{y}'\mathbf{X} \\ \mathbf{W} &= A \text{ first eigenvectors of a PCA calculated with } \mathbf{M}\mathbf{X}'\mathbf{X} \\ \mathbf{T} &= \mathbf{X}\mathbf{W}\end{aligned}$$

279 As the \mathbf{t}_i are orthogonal (28), the correction is performed into \mathbb{R}^N accord-
 280 ing to equation 7.

281 • Westerhuis' DOSC

$$\begin{aligned}\mathbf{y}_1 &= \mathbf{X}(\mathbf{X}'\mathbf{X})^+\mathbf{X}'\mathbf{y} \\ \mathbf{X}_1 &= (\mathbf{I}_N - \mathbf{y}_1(\mathbf{y}'_1\mathbf{y}_1)^{-1}\mathbf{y}'_1)\mathbf{X} \\ \mathbf{T} &= A \text{ first scores of a PCA calculated with } \mathbf{X}_1\end{aligned}$$

282 The scores of a PCA are orthogonal, so the correction is performed into
 283 \mathbb{R}^N according to equation 7.

284 • Trygg's OPLS

$$\begin{aligned}\mathbf{X}_{1:0} &= \mathbf{X}; \text{ for step } i : \\ \mathbf{w}_{PLS}, \mathbf{p}_{PLS} &= \text{first weight and loading vectors of a PLSR calculated with } (\mathbf{X}_{1:i-1}, \mathbf{y}) \\ \mathbf{w}_i &= \mathbf{p}_{PLS} - (\mathbf{w}'_{PLS}\mathbf{p}_{PLS})(\mathbf{w}'_{PLS}\mathbf{w}_{PLS})^{-1}\mathbf{w}_{PLS} \text{ then } \mathbf{w}_i \text{ is normed} \\ \mathbf{t}_i &= \mathbf{X}_{1:i-1}\mathbf{w}_i \\ \mathbf{p}_i &= \mathbf{X}'_{1:i-1}\mathbf{t}_i(\mathbf{t}'_i\mathbf{t}_i)^{-1} \\ \mathbf{X}_{1:i} &= (\mathbf{I}_N - \mathbf{t}_i(\mathbf{t}'_i\mathbf{t}_i)^{-1}\mathbf{t}'_i)\mathbf{X}_{1:i-1}\end{aligned}$$

285 Each vector \mathbf{t}_i lies in the column space of $\mathbf{X}_{1:i-1}$. Thus the \mathbf{t}_i vectors are
 286 orthogonal and can be gathered into a matrix \mathbf{T} for the OPLS correction
 287 into \mathbb{R}^N according to equation 7.

288 *2.6.2. Properties of the OSC methods*

289 OSC methods were created to improve inner properties (26), such as: no (or
290 nearly no) correlation between the OSC scores \mathbf{t}_i and the reference values \mathbf{y} ;
291 removal of the largest variability in \mathbf{X} ; belonging of the \mathbf{t}_i to the column space;
292 and more direct calculation. In order to prevent overfit, just a few dimensions
293 are usually removed. The usefulness of OSC has been widely discussed, and
294 many authors agree that OSC does not improve accuracy (26; 27; 28; 29; 30; 31).
295 Several OSC methods are connected with PLSR. For example, Wold's OSC is
296 derived from PLSR, and OPLS can be obtained from PLSR (32). Often the same
297 calibrations obtained after OSC would be obtained without OSC, but the latter
298 calibrations contain more PLSR latent variables. Nevertheless OSC improves
299 the interpretability of the calibrations by identifying outliers (27) and explaining
300 the regression model obtained after OSC, by analyzing the information removed
301 by OSC (30; 31).

302 OSC models can always be presented as a subtraction: $\mathbf{X}_{OSC} = \mathbf{X} - \sum \mathbf{t}_i \mathbf{p}'_i$
303 (31). The calculations of these models are similar to orthogonal projection into
304 the individual space \mathbb{R}^N . It is less obvious to link OSC to orthogonal projections
305 into the variable space \mathbb{R}^P , but it is possible in some instances. For example, DO
306 and NAP are Euclidian orthogonal projections into \mathbb{R}^P , whereas Fearn's OSC is
307 an oblique orthogonal projection into \mathbb{R}^P , see Appendix A. These mathematical
308 considerations are of importance for the embedded or not embedded properties
309 of OSCs, and also for their compatibilities with other orthogonal projection
310 methods (see below).

311 **3. Discussion**

312 Orthogonal projections are common to the main linear pretreatment meth-
313 ods. We discuss several of their properties, including the relationship to the
314 net analyte signal, the complementarity and the association of different pre-
315 treatments, and the correction of additive and multiplicative effects. We also
316 propose for practical rules using orthogonal projections.

317 *3.1. Net analyte signal and pretreatments in the calibration process*

318 Calibration consists of two steps: an optional pretreatment or preprocessing
319 step, and a mandatory calibration step. The aims of pretreatments and linear
320 calibrations are linked.

321 A linear multivariate calibration extracts the spectral information produced
322 by the compound of interest, and uses this information to predict the concen-
323 tration of the compound. From a mathematical point of view, it is a function
324 from \mathbb{R}^P to \mathbb{R} , represented by a vector \mathbf{b} , called the regression or b-coefficients
325 vector. Let (\mathbf{X}, \mathbf{y}) be a calibration dataset, $(\mathbf{X}_c, \mathbf{y}_c)$ be the calibration dataset
326 after mean centering, and let \mathbf{b} and \mathbf{b}_c be the models obtained from raw or
327 centered data. For a spectrum \mathbf{x}_i from \mathbf{X} acquired on sample i , an estimation
328 \hat{y}_i for the concentration of the compound of interest is given by the following
329 equations, depending on the centering option:

$$\begin{aligned}\widehat{y}_i &= \mathbf{x}'_i \mathbf{b} \text{ (no centering)} \\ \widehat{y}_i &= \mathbf{x}'_{i,c} \mathbf{b}_c + y_0 \text{ (centering)}\end{aligned}$$

330 The intercept y_0 is induced by the centering. A good calibration model gives
 331 a prediction close to the true values of y_i , so $|y_i - \widehat{y}_i|$ is minimized under different
 332 constraints. For that purpose, it is able to extract only the relevant information
 333 from \mathbf{x}_i (or $\mathbf{x}_{i,c}$ if centered). The most condensed relevant spectral information
 334 is called the net analyte signal (NAS). Lorber et al (33) introduced the NAS (a
 335 vector of the same dimension as a spectrum) in multivariate calibration. For
 336 one component of interest, two definitions were proposed:

- 337 • First definition
- 338 "NAS is the part of the spectrum of the component of interest which, is
 339 orthogonal to the spectra of the other components" (33; 34)
- 340 • Second definition
- 341 "NAS is the part of the raw signal that is useful for prediction of the
 342 component of interest" (34)

343 Suppose that all of the contributions to the spectra, except the contribution
 344 of the compound to be predicted, span a subspace \mathcal{E}^D . Let the column-vectors
 345 of a matrix \mathbf{P} be a basis of \mathcal{E}^D . Thus the NAS part of a spectrum \mathbf{x}_i , called
 346 $\mathbf{x}_{i,nas}$, is obtained as an orthogonal projection of \mathbf{x}_i to \mathbf{P} (35), so that the NAS
 347 is in the null space of \mathcal{E}^D :

$$\mathbf{x}_{i,nas} = (\mathbf{I}_P - \mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}')\mathbf{x}_i \quad (8)$$

348 The net sensitivity vector (NSV) \mathbf{s}_{nsv} is the NAS for a compound of interest
 349 whose concentration is one. The vectors \mathbf{s}_{nsv} and \mathbf{b} are collinear and they verify,
 350 for direct and inverse calibrations (34):

$$\mathbf{s}_{nsv} = \mathbf{b}(\mathbf{b}'\mathbf{b})^{-1} \quad (9)$$

351 For the same compound of interest, the NAS is strongly dependent on the
 352 experiment. For instance, glucose presents different NAS, when measured in
 353 fruit juices or blood. If a new compound whose spectra is close to that of glucose
 354 is added either to juice or blood, the glucose NAS is immediately modified.
 355 Thus, the NAS is conceptual rather than actual and is difficult or impossible
 356 to measure. However, the NAS can be estimated. From equation 9, the best
 357 calibration models are expected to be the best NAS estimates.

358 Equations 8 and 9 also clarify the relationships among orthogonal pretreat-
 359 ments, NAS and calibrations. The definitions of the NAS and a pretreatment
 360 based on orthogonal projections are very similar. The differences lie in the in-
 361 formation represented by the matrices \mathbf{P} . For the NAS, \mathbf{P} contains the vectors

362 spanning *all* the spectral contributions to be removed. In general pretreatment
363 cases, \mathbf{P} contains the vectors spanning *only a part* of the space containing the
364 contributions to be removed. The extent to which pretreatments are able to
365 gather the comprehensive information to be removed determines how close cor-
366 rected spectra will be to their NAS. And from equation 9, determining regression
367 vector \mathbf{b} is straightforward. The relationship between the NAS and \mathbf{b} , along
368 with the calculation of the NAS using Euclidian orthogonal projections, has at
369 least two consequences.

370 A first consequence is the embedded / not embedded property of pretreat-
371 ments. A pretreatment is embedded into the calibration if it is not necessary to
372 apply the pretreatment to a new spectrum $\mathbf{x}_{i, val}$ before using it for prediction.
373 If Euclidian orthogonal projection is applied, e.g. to correct the spectra from
374 \mathbf{P} , the regression vector \mathbf{b} of the subsequent calibration is built in a subspace
375 already orthogonal to \mathbf{P} : $\mathbf{b} = \mathcal{P}_{\mathbf{P}}^{\perp} \mathbf{b}$. Thus, it is not necessary to apply the same
376 pretreatment to $\mathbf{x}_{i, val}$. However this property is not verified for non-Euclidian
377 orthogonal projections. For example, OSC methods have different behaviors.
378 Fearn's OSC is not embedded because it is an *oblique* rather than an Euclidian
379 projection. In contrast, the DO and NAP OSCs, which are written as Euclidian
380 orthogonal projections, are embedded. No conclusions can be drawn for the
381 other OSC methods because they have not yet been rewritten as projections
382 into \mathbb{R}^P .

383 A second consequence is an identification of the useful and detrimental spaces
384 within the variables space \mathbb{R}^P (Fig.2). The useful space \mathcal{E}^U contains informa-
385 tions for the prediction of the compound of interest. Its dimension is one ac-
386 cording to the first definition, but can be greater than one according to the
387 second definition. The subspace related to other compounds or chemical influ-
388 ences is detrimental (\mathcal{E}^D) if it shares common information with the useful space.
389 The information from \mathcal{E}^U which is not shared by \mathcal{E}^D is the NAS, i.e. the NAS
390 is orthogonal to \mathcal{E}^D . All the information from \mathbb{R}^P which does not belong to
391 \mathcal{E}^U or \mathcal{E}^D is unuseful information because it is already orthogonal to \mathcal{E}^U . Pre-
392 treatments have different strategies to remove \mathcal{E}^D . OSC methods rely on the
393 information provided by \mathbf{y} . However, as PLSR uses the same approach, there
394 is no added value, at least for prediction. Nevertheless OSCs calculated using
395 other data (e.g. pure spectra or experimental design) and then applied to the
396 calibration data \mathbf{X} can increase the accuracy of measurement (36; 37) because
397 the two sets of spectral informations are complementary. On the other hand,
398 the non OSC orthogonal projections (OSP, IIR, EPO, TOP, DOP, EROS, de-
399 trend, CPSA, DWT-WPT) rely on additional information, and various sources
400 of information can be withdrawn. The disadvantages of the methods are: (1)
401 except for DWT-WPT, a prior knowledge is necessary, such as a model, pure
402 spectra or information that can be used to build an experimental design; and
403 (2) non OSC orthogonal pretreatments *may* remove information from the NAS
404 subspace, contributing to a drop in the accuracy of prediction. Therefore, choos-
405 ing the correct pretreatments and tuning and applying them properly, are very
406 important for optimizing their predictive value.

407 *3.2. Pretreatment complementarity*

408 A spectrum can be viewed as the sum of signals of different frequencies: low,
 409 medium, and high (38).

410 Baseline shifts correspond to low frequencies. They can be modeled and
 411 removed either by detrend and related methods or by data compression methods
 412 such as WPT. Detrend is used more often than WPT. One reason for this may
 413 be that the polynomials used are based on models of low rank (e.g. 2). These
 414 models may already fit well with the observed deviations of the baseline, and
 415 thus detrend is more parsimonious.

416 Noise corresponds to high frequencies, usually removed using the SG algo-
 417 rithm. However WPT also removes noise, smooths the spectra and is more
 418 compatible with the other orthogonal projections.

419 Most of the relevant spectral information corresponds to medium frequencies,
 420 such as the information from the component of interest and the information to
 421 be removed. Orthogonal projections perform best in these situations.

422 Therefore different pretreatments should be selected according to their comple-
 423 mentarity; that is, their ability to represent the whole space to be removed
 424 (Fig. 2, hatched part). Pretreatment choice is guided by the data to be pro-
 425 cessed and also by the available information. However, disrupting information
 426 can be expected in all frequency ranges. Thus, a relevant association of several
 427 methods, such as detrend + EPO + WPT, would respectively be able to remove
 428 the low, medium and high frequency spectral perturbations. How can several
 429 projection methods be used to produce the best results?

430 *3.3. Association of pretreatments*

431 Using successively two or more orthogonal projections is not equivalent to
 432 performing a single orthogonal projection containing all the information to be
 433 removed. An example is given in Fig. 1. Suppose that the spectral information
 434 from water and ethanol are to be removed from the spectrum of a sample that
 435 is half ethanol and half water. The expected result is the null vector. It is
 436 obtained if the orthogonal projection to ethanol and water is done in one step,
 437 but not if two orthogonal projections, one for ethanol and one for water, are done
 438 separately. Moreover, successive orthogonal projections are not commutative,
 439 so different orders lead to different results.

440 Let \mathbf{P} and \mathbf{Q} be two matrices whose column-vectors span the subspaces \mathcal{E}_P
 441 and \mathcal{E}_Q . Let \mathcal{P}_P and \mathcal{P}_Q be the orthogonal projectors for \mathcal{E}_P and \mathcal{E}_Q , respectively.
 442 The combined orthogonal projector for the \mathcal{E}_P and \mathcal{E}_Q subspaces is called \mathcal{P}_{P+Q} .
 443 According to Piziak et al (39):

$$\mathcal{P}_{P+Q} = (\mathcal{P}_P + \mathcal{P}_Q)(\mathcal{P}_P + \mathcal{P}_Q)^+$$

444 The orthogonal projector to \mathcal{E}_P and \mathcal{E}_Q is \mathcal{P}_{P+Q}^\perp defined as:

$$\mathcal{P}_{P+Q}^\perp = \mathbf{I}_P - (\mathcal{P}_P + \mathcal{P}_Q)(\mathcal{P}_P + \mathcal{P}_Q)^+ \quad (10)$$

445 If \mathcal{E}_P and \mathcal{E}_Q are orthogonal, equation 10 is simplified:

$$\mathcal{P}_{P+Q}^\perp = \mathbf{I}_P - \mathcal{P}_P - \mathcal{P}_Q = (\mathbf{I}_P - \mathcal{P}_P)(\mathbf{I}_P - \mathcal{P}_Q) = \mathcal{P}_P^\perp \mathcal{P}_Q^\perp$$

446 However, in the other cases where \mathcal{E}_P and \mathcal{E}_Q are not orthogonal, $\mathcal{P}_{P+Q}^\perp \neq$
447 $\mathcal{P}_P^\perp \mathcal{P}_Q^\perp$. So, if several Euclidian orthogonal pretreatments must be applied, it
448 is strongly recommended to merge their respective \mathbf{P} matrices into a matrix \mathbf{R}
449 and then performing a single orthogonal projection to \mathbf{R} . For example, suppose
450 that spectra are to be corrected from scattering and from chemical components
451 whose pure spectra are known. The best choice is to merge the \mathbf{A}_2 to the
452 pure spectra, yielding \mathbf{R} , and to orthogonalize to \mathbf{R} . This is also the option
453 recommended by Brown (15). In the case of successive orthogonal projections,
454 calculations will be erroneous if the bases of the different subspaces are not
455 orthogonal.

456 3.4. Limits of orthogonal projections

457 Orthogonal projections have at least two limits: the number of dimensions
458 of the spectra to be processed and the correction of multiplicative effects.

459 3.4.1. Reduction of dimensions and number of variables

460 Orthogonal projection to a subspace of dimension A retains the number of
461 variables but reduces the mathematical dimension of the spectra by A , which
462 cannot exceed the dimension P of the spectra. This is not a problem if spectra
463 contain hundreds of variables, because A remains lower than P . However, if
464 the spectra contain only a few variables (e.g. between 10 and 20), then only a
465 few dimensions can be removed by orthogonal projections, and often orthogonal
466 projection cannot be applied.

467 3.4.2. Additive and multiplicative effects

468 Orthogonal projection methods correct additive effects according to equation
469 1. Suppose that the spectral perturbations also induce a multiplicative effect
470 α_i , such that Eq. 1 can be written:

$$\mathbf{x}_{i,obs} = \alpha_i(\mathbf{x}_i + \mathbf{h}_i) = \alpha_i \mathbf{x}_i + \alpha_i \mathbf{h}_i \quad (11)$$

471 An orthogonal projector \mathcal{P}_R^\perp can be obtained such that:

$$\mathbf{x}_{i,corr} = \mathbf{x}_{i,obs} \mathcal{P}_R^\perp = \alpha_i \mathbf{x}_i \mathcal{P}_R^\perp \quad (12)$$

472 Orthogonal projections are able to remove the term $\alpha_i \mathbf{h}_i$, but are not able to
473 correct the value α_i associated with \mathbf{x}_i ; they cannot handle multiplicative effects.
474 Thus, methods such as standard normal variate (SNV) proposed by Barnes (11),
475 or (extended) multiplicative signal correction proposed by Martens (40; 41) do
476 not use orthogonal projections to correct multiplicative effects.

477 4. Conclusion

478 All the pretreatments presented here are based on orthogonal projections.
479 Some pretreatments that also use the principle of orthogonal projection were
480 not addressed because they do not fulfill all the conditions. The MSC and EMSC
481 methods perform a corrections using subtraction and division. SG is a piecewise
482 orthogonal projection. Even variable selection is an orthogonal projection into
483 the variables space, but in practice, variables are dropped by reshaping the
484 spectra rather than by setting them to 0. Thus, the orthogonal projection
485 principle, which is directly associated with the least squares approximation, is
486 used extensively and its prevalence is a confirmation of its power. Projections
487 are performed usually in a *Euclidian* space but at least one OSC is an oblique
488 projection performed in a *non-Euclidian* space.

489 The information used by calibrations is actually an estimation of the NAS.
490 A very important issue for pretreatments is identifying the detrimental infor-
491 mation to be removed by orthogonal projection. Two opposing strategies are
492 used: supervised and unsupervised. On one hand, OSC methods are supervised,
493 the information to be removed is chosen orthogonal to the reference values \mathbf{y} .
494 These methods obey the definition of the NAS, and thus OSC methods should
495 perform very well. However, due to OSC's redundancy with PLSR, which is
496 based on the same principles its accuracy has not improved substantially. On
497 the other hand, information removed by unsupervised orthogonal projection
498 methods is chosen based on prior knowledge of the experimental conditions and
499 the analyzed samples. All the available information, pure spectra, spectra from
500 an experimental design, and models (such as polynomials), should be consid-
501 ered to correct the detrimental information as comprehensively as possible. It is
502 important to avoid removing any information about the compound of interest,
503 which would lead to worse calibration models. Thus, proper use of orthogonal
504 projections requires being very careful about the quality and the completeness
505 of the removed information.

506 Orthogonal projections have practical properties. Information from different
507 origins should be merged in order to process only one orthogonal projection. Eu-
508 clidian orthogonal projections are embedded; they are applied only once when
509 building the calibration model. There is no need to apply them to new spec-
510 tra. Orthogonal projections also have limits. They cannot be applied in good
511 conditions for spectra containing just a few variables, and they do not correct
512 multiplicative effects. Nevertheless, orthogonal projections have proven to be
513 valuable tools when building the calibration models of many applications.

514 5. Acknowledgements

515 The authors would like to thank the reviewers whose relevant remarks were
516 very helpful for the improvement of this paper.

\mathbf{X}	matrix $N \times P$, N samples and P spectral variables
\mathbf{y}	vector $N \times 1$, the reference values
$\mathbf{X}_{1:i}$	projection of \mathbf{X} orthogonally to $\{\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_i\}$
\mathbf{T}	matrix $N \times A$, scores for \mathbf{X}
\mathbf{P}	matrix $P \times A$, loadings for \mathbf{X}
\mathbf{W}	matrix $P \times A$, weights for \mathbf{X}
Σ	Moore-Penrose pseudo-inverse of $(\mathbf{X}'\mathbf{X})$; $\Sigma = (\mathbf{X}'\mathbf{X})^+$
$\mathbf{I}_N, \mathbf{I}_P$	identity matrices for \mathbb{R}^N and \mathbb{R}^P spaces
\mathcal{P}_P	Euclidian orthogonal projector $P \times P$ onto \mathbf{P} ; $\mathcal{P}_P = \mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'$
\mathcal{P}_P^\perp	Euclidian projector $P \times P$ orthogonally to \mathbf{P} ; $\mathcal{P}_P^\perp = \mathbf{I} - \mathcal{P}_P$
\mathbf{t}_i	i^{eme} column vector of \mathbf{T}
\mathbf{p}_i	i^{eme} column vector of \mathbf{P}
\mathbf{w}_i	i^{eme} column vector of \mathbf{W}
\mathcal{E}_X	subspace of \mathbb{R}^P spanned by the line vectors of \mathbf{X}
$\mathcal{E}^U, \mathcal{E}^D$	useful / detrimental subspaces of \mathbb{R}^P

Table 1: Main notations

BB	best basis
CPSA	constrained principal spectra analysis
DFT	discrete Fourier transform
DO	direct orthogonalization
DOP	dynamic orthogonal projection
DOSC	direct orthogonal signal correction
DWT	discrete wavelet transform
EMSC	extended multiplicative signal correction
EPO	external parameter orthogonalization
EROS	error removal by orthogonal subtraction
IIR	independent interference reduction
MSC	multiplicative signal correction
MDL	minimum description length
NAP	net analyte preprocessing
NAS	net analyte signal
NSV	net sensitivity vector
OPLS	orthogonal projection to latent structures
OSC	orthogonal signal correction
OSP	orthogonal subspace projection
PCA	principal component analysis
PCDA	principal component discriminant analysis
PLSR	projection to latent structures regression
RMSEP	root mean square error of prediction
SG	Savitsky-Golay
SNV	standard normal variate
SVD	singular value decomposition
TOP	transfer by orthogonal projection
WPT	wavelet packet transform

Table 2: Glossary

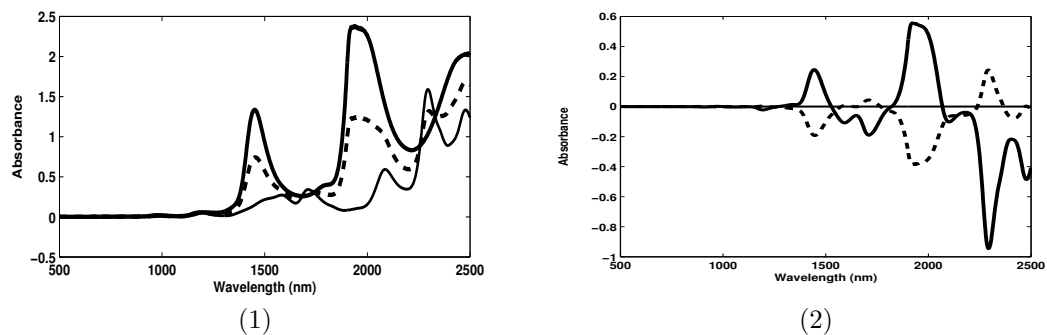


Figure 1: Water and ethanol spectra, and their orthogonal projection. (1) From top to bottom, spectra of water, water+ethanol, ethanol. (2) Projection of the (water+ethanol) spectrum orthogonally to: ethanol then water (thick line), water then ethanol (dashed line), and a matrix containing water and ethanol spectra (horizontal light line)

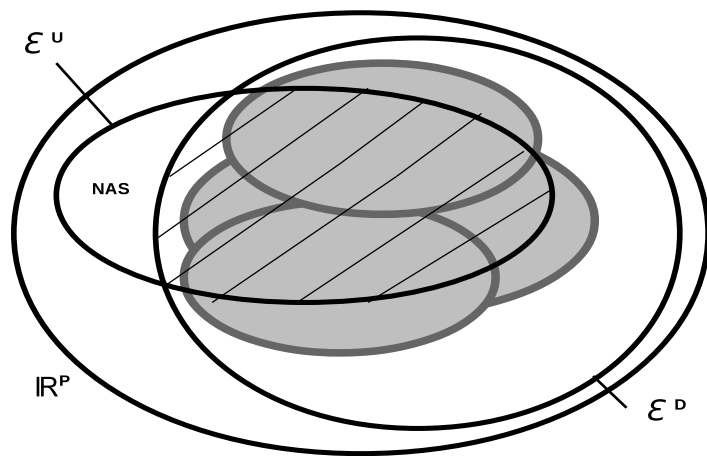


Figure 2: Different subspaces of \mathbb{R}^P . The useful subspace \mathcal{E}^U ; the detrimental subspace \mathcal{E}^D ; the NAS; the common part of \mathcal{E}^U and \mathcal{E}^D to be removed (hatched); and the information represented by three orthogonal pretreatments (grey) are depicted in the figure. Information inside of and outside of each subspace are orthogonal

517 **References**

- 518 [1] S.Wold, A.Ruhe, H.Wold, W. D. III, The collinearity problem in linear
519 regression, the partial least square (pls) approach to generalized inverses, J.
520 Sci Stat. Comput. 5 (1984) 735–743.
- 521 [2] A.Rinnan, F.VanDenBerg, S. Engelsen, Review of the most common pre-
522 processing techniques for near infrared spectra, TrAC 28(10) (2009) 1201–
523 1222.
- 524 [3] C.I.Chang, M.L.G.Althouse, Unsupervised interference rejection approach
525 to target detection and classification for hyperspectral imagery, Opt. Eng.
526 37(3) (1998) 735–743.
- 527 [4] C.I.Chang, Hyperspectral imaging, Kluwer academic / Plenum, N.Y., 2003.
- 528 [5] J.C.Harsanyi, C.I.Chang, Hyperspectral image classification and dimen-
529 sionality reduction: an orthogonal subspace projection approach, IEEE
530 Trans. Geosci. Rem. Sens. 32(4) (1994) 779–786.
- 531 [6] P.W.Hansen, Pre-processing method minimizing the need for reference
532 analyses, J. Chemom. 15 (2001) 123–131.
- 533 [7] J.M.Roger, F.Chauchard, V.Bellon-Maurel, Epo-pls external parameter or-
534 thogonalisation of pls, application to temperature-independant measure-
535 ment of sugar contents in fruits, Chemom. Intell. Lab. Syst. 66 (2003)
536 191–204.
- 537 [8] A.Andrew, T.Fearn, Transfer by orthogonal projection: making near infra-
538 red calibrations robust to between-instrument variation, Chemom. Intell.
539 Lab. Syst. 72 (2004) 51–56.
- 540 [9] M.Zeaiter, J.M.Roger, V.Bellon-Maurel, Dynamic orthogonal projection, a
541 new method to maintain the on-line robustness of multivariate calibration,
542 application to nir-based monitoring of wine fermentations., Chemom. Intell.
543 Lab. Syst. 80 (2006) 227–235.
- 544 [10] Y.Zhu, T.Fearn, D.Samuel, A.Dhar, O.Hameed, S.G.Brown, L.B.Lovat, Er-
545 ror removal by orthogonal subtraction (eros): a customised pre-treatment
546 for spectroscopic data, J. Chemom. 22 (2008) 130–134.
- 547 [11] R.J.Barnes, M.S.Dhanao, S.J.Lister, Standard normal variate transforma-
548 tion and de-trending of near-infrared diffuse reflectance spectra, Appl. Spec-
549 trosc. 43 (1989) 772–777.
- 550 [12] R.J.Barnes, M.S.Dhanao, S.J.Lister, Correction to the description of stan-
551 dard normal variate (snv) and de-trend (dt) transformations in practical
552 spectroscopy with applications in food and beverage analysis, 2nd edition,
553 J. Near Infrared Spectrosc. 1 (1993) 185–186.

- 554 [13] V.Mazet, C.Carteret, D.Brie, J.Idier, B.Humbert, Background removal
555 from spectra by designing and minimizing a non-quadratic cost function,
556 Chemom. Intell. Lab. Syst. 76 (2005) 121–133.
- 557 [14] F.Gan, G.Ruan, J.Mo, Baseline correction by improved iterative polyno-
558 mial fitting with automatic threshold, Chemom. Intell. Lab. Syst. 82 (2006)
559 59–65.
- 560 [15] J.M.Brown, Method for correcting spectral data for data due to the spec-
561 tral measurement process itself and estimating unknown property and/or
562 composition data of a sample using such method, U.S. patent 5.121.337
563 (1992).
- 564 [16] B.Walczak, D.L.Massart, Noise suppression and signal compression using
565 the wavelet packet transform, Chemom. Intell. Lab. Syst. 36 (1997) 81–94.
- 566 [17] J.Trygg, S.Wold, Pls regression on wavelet compressed nir spectra,
567 Chemom. Intell. Lab. Syst. 42 (1998) 209–220.
- 568 [18] L.Eriksson, J.Trygg, E.Johansson, R.Bro, S.Wold, Orthogonal signal cor-
569 rection, wavelet analysis and multivariate calibration of complicated pro-
570 cess fluorescence data, Anal.Chim. Acta 420 (2000) 181–195.
- 571 [19] L.Pasti, B.Walczak, D.L.Massart, P.Reschiglian, Optimization of signal de-
572 noising in discrete wavelet transform, Chemom. Intell. Lab. Syst. 48 (1999)
573 21–34.
- 574 [20] Y.Hu, T.Jiang, A.Shen, W.Li, X.Wang, J.Hu, A background elimination
575 method based on wavelet transform for raman spectra, Chemom. Intell.
576 Lab. Syst. 85 (2007) 94–101.
- 577 [21] R.K.H.Galvao, G.E.Jose, H.A.D.Filho, M.C.V.Aranjo, E.C.DaSilva,
578 H.M.Paiva, T.C.B.Saldanha, E.S.O.N.DeSouza, Optimal wavelet filter con-
579 struction using x and y data, Chemom. Intell. Lab. Syst. 70 (2004) 1–10.
- 580 [22] B.K.Alsberg, A.M.Woodward, D.B.Bell, An introduction to wavelet trans-
581 form for chemometricians: a time-frequency approach, Chemom. Intell.
582 Lab. Syst. 37 (1997) 215–239.
- 583 [23] V.J.Barclay, R.F.Bonner, I.P.Hamilton, Application of wavelet transforms
584 to experimental spectra: smoothing, denoising and data compression, Anal.
585 Chem. 69 (1997) 78–90.
- 586 [24] S. Wold, H. Antti, F. Lindgren, J. Ohman, Orthogonal signal correction of
587 near infra-red spectra., Chemom. Intell. Lab. Syst. 44 (1998) 175–185.
- 588 [25] J.Sjoblom, O.Svensson, M.Josefson, H.Kullberg, S.Wold, An evaluation of
589 orthogonal signal correction applied to calibration transfer of near infrared
590 spectra, Chemom. Intell. Lab. Syst. 44 (1998) 229–244.

- 591 [26] J.A.Westerhuis, J.DeJong, A.K.Smilde, Direct orthogonalization,
592 Chemom. Intell. Lab. Syst. 56 (2001) 13–25.
- 593 [27] C.A.Andersson, Direct orthogonalization, Chemom. Intell. Lab. Syst. 47
594 (1999) 51–63.
- 595 [28] T.Fearn, On orthogonal signal correction., Chemom. Intell. Lab. Syst. 50
596 (2000) 47–52.
- 597 [29] H.C.Goicoechea, A.C.Olivieri, A comparison of orthogonal signal correction
598 and net analyte preprocessing methods, theoretical and experimental study,
599 Chemom. Intell. Lab. Syst. 56 (2001) 73–81.
- 600 [30] J.Trygg, S.Wold, Orthogonal projection to latent structures (o-pls), J. of
601 Chemom. 16(3) (2002) 119–128.
- 602 [31] O.Svensson, T.Kourti, J.F.MacGregor, An investigation of orthogonal sig-
603 nal correction algorithms and their characteristics, J. of Chemom. 16(4)
604 (2002) 176–188.
- 605 [32] E.K.Kemsley, H.S.Tapp, Opls filtered data can be obtained directly from
606 non-orthogonalized pls1, J. Chemom. 23 (2009) 518–529.
- 607 [33] A.Lorber, N.K.M.Faber, B.R.Kowalski, Net analyte signal calculation in
608 multivariate calibration, Anal. Chem. 69(8) (1997) 1620–1626.
- 609 [34] J.Ferre, N.K.M.Faber, Net analyte signal calculation for multivariate cali-
610 bration, Chemom. Intell. Lab. Syst. 69 (2003) 123–136.
- 611 [35] R.Bro, C.M.Andersen, Theory of net analyze signal vectors in inverse re-
612 gression, J. of Chemom. 17 (2003) 646–652.
- 613 [36] J.C.Boulet, T.Dococ, J.M.Roger, Improvement of calibration models using
614 two successive orthogonal projection methods, application to quantification
615 of wine mannoproteins., Chemom. Intell. Lab. Syst. 87 (2007) 295–302.
- 616 [37] S.Preys, J.M.Roger, J.C.Boulet, Robust calibration using orthogonal pro-
617 jection and experimental design, application to the correction of the light
618 scattering effect on turbid nir spectra., Chemom. Intell. Lab. Syst. 91 (2006)
619 28–33.
- 620 [38] K.H.Liland, T.Almoy, B.H.Mevik, Optimal choice of baseline correction
621 for multivariate calibration of spectra, Appl. Spectrosc. 64(9) (2010) 1007–
622 1016.
- 623 [39] R.Piziak, P.L.Odell, R.Hahn, Constructing projections on sums and inter-
624 sections, Comput. Math. Appl. 37 (1999) 67–74.
- 625 [40] P.Geladi, D.MacDougall, H.Martens, Linearization and scatter-correction
626 for near infrared reflectance spectra of meat, Appl. Spectrosc. 39(3) (1985)
627 491–499.

628 [41] H.Martens, E.Stark, Extended multiplicative signal correction and spec-
 629 tral interference subtraction: new preprocessing methods for near infrared
 630 spectroscopy, J. Pharm. Biomed. Anal. 9 (1991) 625–635.

631 [42] R.DeMaesschlack, D.JouanRimbaud, D.L.Massart, The mahalanobis dis-
 632 tance, Chemom. Intell. Lab. Syst. 50 (2000) 1–18.

633 **Appendix A. Writing Fearn’s OSC in the form of an orthogonal pro-**
 634 **jection into \mathbb{R}^P**

635 Two equations are extracted from Fearn’s text:

$$\mathbf{t}_i = \mathbf{X}\mathbf{w}_i \quad (\text{A.1})$$

$$\mathbf{p}_i = \mathbf{X}'\mathbf{t}_i(\mathbf{t}'_i\mathbf{t}_i)^{-1} \quad (\text{A.2})$$

636 Let Σ be $(\mathbf{X}'\mathbf{X})^+$, the Moore-Penrose pseudo-inverse of $(\mathbf{X}'\mathbf{X})$. Our first
 637 goal is to write \mathbf{w}_i using \mathbf{p}_i and Σ . The calculation of the product $\mathbf{p}'_i\Sigma\mathbf{p}_j$ is
 638 performed using equations A.1 and A.2 :

$$\begin{aligned} \mathbf{p}'_i\Sigma\mathbf{p}_j &= (\mathbf{t}'_i\mathbf{t}_i)^{-1}\mathbf{w}'_i\mathbf{X}'\mathbf{X}\Sigma\mathbf{X}'\mathbf{X}\mathbf{w}_j(\mathbf{t}'_j\mathbf{t}_j)^{-1} \\ &= (\mathbf{t}'_i\mathbf{t}_i)^{-1}\mathbf{w}'_i\mathbf{X}'\mathbf{X}\mathbf{w}_j(\mathbf{t}'_j\mathbf{t}_j)^{-1} \\ &= (\mathbf{t}'_i\mathbf{t}_i)^{-1}\mathbf{t}'_i\mathbf{t}_j(\mathbf{t}'_j\mathbf{t}_j)^{-1} \end{aligned}$$

639 If $i \neq j$, because of the orthogonality of the \mathbf{t}_i :

$$\mathbf{p}'_i\Sigma\mathbf{p}_i = (\mathbf{t}'_i\mathbf{t}_i)^{-1} \quad (\text{A.3})$$

$$\mathbf{p}'_i\Sigma\mathbf{p}_j = 0 \quad (\text{A.4})$$

640 After multiplying on the left by $\Sigma\mathbf{X}'$, a different combination of equations
 641 A.1 and A.2, completed with equation A.3, yields:

$$\Sigma\mathbf{X}'\mathbf{X}\mathbf{w}_i = \Sigma\mathbf{X}'\mathbf{t}_i = \Sigma\mathbf{p}_i(\mathbf{t}'_i\mathbf{t}_i) = \Sigma\mathbf{p}_i(\mathbf{p}'_i\Sigma\mathbf{p}_i)^{-1} \quad (\text{A.5})$$

642 The left term is the projection of \mathbf{w}_i onto the space spanned by $\mathbf{X}'\mathbf{X}$, which
 643 is also the space spanned by the lines of \mathbf{X} (39), called \mathcal{E}_X . On the other hand,
 644 the \mathbf{w}_i are the eigenvectors of the matrix $\mathbf{M}\mathbf{X}'\mathbf{X}$, where \mathbf{M} is the projector
 645 orthogonal to $\mathbf{X}'\mathbf{y}$, which belongs to \mathcal{E}_X . Thus, \mathbf{w}_i also belongs to \mathcal{E}_X , it is not
 646 modified after projection onto $\mathbf{X}'\mathbf{X}$ and equation A.5 is simplified:

$$\mathbf{w}_i = \Sigma\mathbf{p}_i(\mathbf{p}'_i\Sigma\mathbf{p}_i)^{-1} \quad (\text{A.6})$$

647 The \mathbf{p}_i are Σ - orthogonal according to equation A.4, and therefore $(\mathbf{P}'\Sigma\mathbf{P})$
648 is a diagonal matrix of full rank. Thus, equation A.6 can be written in a matrix
649 form:

$$\mathbf{W} = \Sigma\mathbf{P}(\mathbf{P}'\Sigma\mathbf{P})^{-1}$$

650 The OSC correction is:

$$\begin{aligned}\mathbf{X}_{OSC} &= \mathbf{X} - \mathbf{X} \sum_{i=1}^A \mathbf{w}_i \mathbf{p}'_i \\ &= \mathbf{X} - \mathbf{X}\mathbf{W}\mathbf{P}' \\ &= \mathbf{X}(\mathbf{I}_P - \Sigma\mathbf{P}(\mathbf{P}'\Sigma\mathbf{P})^{-1}\mathbf{P}')\end{aligned}$$

651 To conclude, Fearn's OSC can be also expressed into \mathbb{R}^P as an oblique (not
652 Euclidian) projection to the space spanned by the column-vectors of \mathbf{P} . If \mathbf{X}
653 has been centered, Σ is a metric associated to a Mahalanobis distance (42).