Received: 6 June 2017 | Revised: 14 February 2018 | Accepted: 13 April 2018

DOI: 10.1002/cem.3045

REVIEW

WILEY Journal of CHEMOMETRICS

A review of orthogonal projections for calibration

Jean-Michel Roger[1] | Jean-Claude Boulet[2]

[1]ITAP, Irstea, Montpellier SupAgro, University of Montpellier, Montpellier, France

[2]INRA, UMR1083, Montpellier SupAgro, University of Montpellier, Montpellier, France

Correspondence
Jean-Michel Roger, ITAP, Irstea, Montpellier SupAgro, University of Montpellier, Montpellier, France.
Email: jean-michel.roger@irstea.fr

Abstract

Effective methods often rely on simple mathematical operators. Among these operators, orthogonal projections have been widely used because of their simplicity in compensating for detrimental factors. This efficiency depends largely on the way these tools are prepared. This article links the mathematical basics of orthogonal projections to the notion of vectoral subspaces, highlighting which information should be removed in the process and the important practical properties concerned with optimizing this technique. This review covers several methods involving orthogonal projections and focuses specifically on their practical use. This concerns the identification of detrimental information and its removal together with adjusting the dimension of the projection. The methodology discussed in this review will enable the reader to optimize orthogonal projections for any given situation.

KEYWORDS

dimension, detrimental, orthogonal, projection, subspace

1 | INTRODUCTION

For decades, analytical chemistry equipment has provided laboratories with increasingly multivariate signals. Among these, near-infrared spectroscopy (NIRS) has emerged as a technique applicable to online monitoring (eg, in food processing, agriculture, pharmaceuticals, or petrochemicals). The NIRS will be used in this document as a case study. Chemometrics has developed methods to process these signals and perform spectrometer calibrations. Factorial methods such as principal component regression[1] or partial least squares regression (PLSR)[2] are renown calibration methods. Many studies advocate that NIRS can replace analytical measures. However, few online applications have been developed. A reason for this partial success may be the calibration model's lack of robustness (the model is often developed with small databases and with a *blackbox* calibration software). It is known that calibration methods based on statistical optimization such as PLSR are suboptimal, producing models based on the optimization of statistical criteria, for example, covariance for PLSR. There is no guarantee that the relationship captured by the model is based on the composition of the product.

Imagine that we can measure the spectra of a set of samples encompassing all sources of variations that impact the spectra. A calibration model calculated on this set would extract the part of the signal that concerns the compound of interest. This signal, called the net analytical signal (NAS), is what all calibration methods look for. In practice, the calibration bases mainly contain a large variation of the compound of interest and a few small variations of the interfering factors. It is generally easy to calibrate a model on this type of set, but the model can be sensitive to any interfering factor that does not vary in the calibration set. These factors can affect the programs robustness.[3-5] The NIRS is confronted with the change of the spectrometers,[6] the variations of temperature,[7] the variations of external light,[8] the variations of humidity,[9] the year, or the conditions of growth.[10] In response to these problems, a recent trend in chemometrics has been to produce methods coupling multivariate regression with better management of the spectral measurement space. Among these methods are orthogonal projection (OP) methods. The OP methods aim to make the calibration insensitive to interfering factors

that are not present in the calibration set. The principle is to identify the subspace affected by interfering factors and to eliminate it before calibration. These methods have been used successfully to improve the robustness of calibrations with respect to temperature,[11] humidity,[12] crop year,[13] spectrometer change,[14,15] and even for some combinatory effects.[16]

This paper proposes a review of the OP methods. A theoretical recall introduces both the decomposition into subspaces and the concepts of OPs. Then, the main methods are reviewed. The practical implementation of OPs is then presented in the context of real examples. Finally, the advantages, disavantages, and limitations of the OP methods are discussed.

2 | THEORY

2.1 | Notations

Table 1 presents the main symbols used in this paper. Capital bold characters are used for matrices, eg, $\mathbf{A}$; small bold characters are used for column vectors, eg, $\mathbf{a}_j$ denotes the jth column of $\mathbf{A}$; row vectors are denoted with the transpose notation. Nonbold lowercase characters are used for scalars, eg, matrix element a_{ij} or index i. Index i will be used for rows and j for columns. Constant scalars are denoted with uppercase nonbold characters, eg, the number of rows N of a matrix. Chemical or physical factors are denoted by italic uppercase characters, eg, $\mathcal{Y}$, the variable to be predicted, which will be called the *response*.

The column and row vectors of a given $(N \times Q)$ matrix $\mathbf{X}$ span 2 subspaces of $\mathbb{R}^N$ and $\mathbb{R}^Q$, respectively.[17] They are called the column space and the row space of $\mathbf{X}$. In the present paper, because we will address chemometrics data, the column space of $\mathbf{X}$ will be named the *sample space*, and the row space of $\mathbf{X}$ will be named the *signal space*. A vector of the sample space contains N values. It is, for example, a wavelength recorded on N samples. A vector of the signal space contains Q values. It is for example a spectrum recorded on Q wavelengths. Most of the concepts presented in this paper refer to the signal space and are thus related to linear combinations of spectra.

2.2 | Orthogonal projections

A theoretical presentation of OPs can be found in Meyer.[17] Let $\mathcal{E}$ be a vector space, and $\mathcal{F}$ be a subspace of $\mathcal{E}$. Let $\mathbf{P}$ be a basis of $\mathcal{F}$. The projection onto $\mathcal{F}$ ($\mathcal{P}_{\mathcal{F}}$) keeps in any vector $\mathbf{x}$ of $\mathcal{E}$ its part contained into $\mathcal{F}$. The projection orthogonal to $\mathcal{F}$ ($\mathcal{P}_{\mathcal{F}}^{\perp}$) removes from any vector $\mathbf{x}$ its part contained in $\mathcal{F}$. From a matrix $\mathbf{X}$, one can determine and use OPs in the sample space or in the signal space. An OP in the sample space is a $N \times N$ matrix, which acts on the rows of $\mathbf{X}$ and consequently left multiplies $\mathbf{X}$. An OP in the signal space is a $Q \times Q$ matrix, which acts on the columns of $\mathbf{X}$ and consequently right multiplies $\mathbf{X}$ (see Table 1).

Sample space OPs are used to separate sources of information related to the collection of the N samples. For example, let $\mathbf{1}_N$ be a vector made up of N ones. Orthogonal projection of $\mathbf{X}$ onto $\mathbf{1}_N$, noted $\mathcal{P}_{\mathbf{1}_N}(\mathbf{X})$, produces a matrix containing N times the mean spectrum of $\mathbf{X}$. This projection only keeps the information related to the mean. In contrast, projecting $\mathbf{X}$ orthogonally to $\mathbf{1}_N$, noted $\mathcal{P}_{\mathbf{1}_N}^{\perp}(\mathbf{X})$, removes from $\mathbf{X}$ its row of means (it mean centers the $\mathbf{X}$ matrix). This projection removes the information related to the mean. Another example is the so called deflation process, used in iterative algorithms as nonlinear iterative partial least squares.[18] At each step of the iteration a vector of N scores is produced, and the data are projected orthogonally to these scores. Thus, the scores are orthogonal, and the variances they carry are additive.

Signal space OPs are used to separate signal features. They are the basis of the methods discussed in this paper and so will be presented in detail. Figure 1 presents 2 types of analytical signals. On the left part of the figure, a chromatogram separates at best all the different chemical compounds of a sample. As a result, the information about one compound lies in a limited location of the chromatogram: compound contributions can easily be extracted by addressing corresponding

TABLE 1 Glossary

N, Q	Integers; numbers of observations, of variables
$\mathbb{R}^N$, $\mathbb{R}^Q$	Vector spaces of dimensions N and Q respectively
$\mathbf{X}$	N observations of Q variables; a matrix $(N \times Q)$
$\mathbf{P}$	$(Q \times A)$ matrix of loadings defining a
	A dimensional subspace $\mathcal{E}$ of $\mathbb{R}^Q$
$\mathcal{P}_{\mathcal{E}}$ or $\mathcal{P}_{\mathbf{P}} = \mathbf{P}(\mathbf{P}^{\mathrm{T}}\mathbf{P})^{-1}\mathbf{P}^{\mathrm{T}}$	Projection onto $\mathcal{E}$[17]
$\mathcal{P}_{\mathcal{E}}^{\perp}$ or $\mathcal{P}_{\mathbf{P}}^{\perp} = \mathbf{I}_Q - \mathbf{P}(\mathbf{P}^{\mathrm{T}}\mathbf{P})^{-1}\mathbf{P}^{\mathrm{T}}$	Projection orthogonal to $\mathcal{E}$[17]

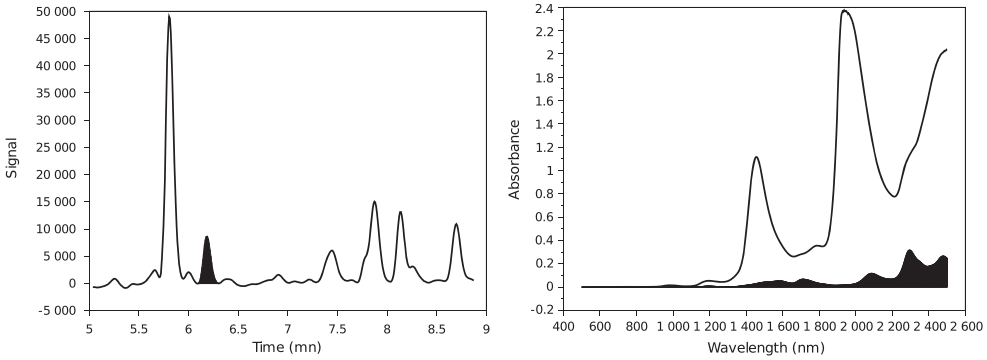

FIGURE 1 Comparison of a chromatogram (left) vs an near-infrared spectrum (right); the black area corresponds to the contribution of one compound

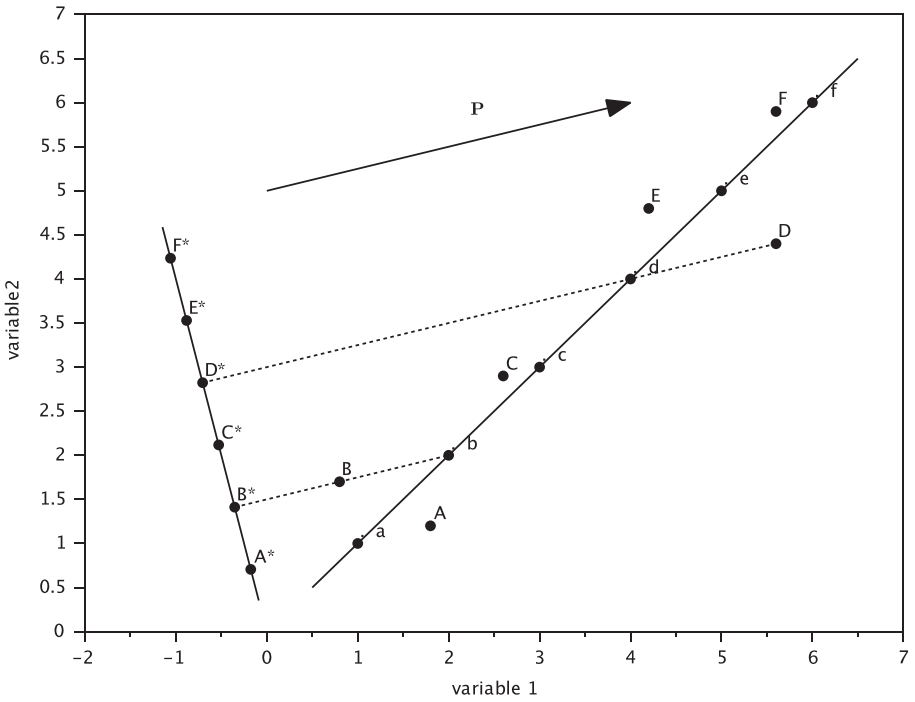

FIGURE 2 Illustration of a projection. The original values of observations {a...f} are aligned. An interfering factor represented by the direction of the arrow shifts these points to {A...F}. The points {A*...F*} are the result of the projection of {A...F} on a subspace orthogonal to the direction represented by the arrow. This projection is illustrated by a dotted line for B and D

variables. On the right part of the figure, an near-infrared (NIR) spectrum does not clearly separate the compounds. The information about one compound appears as a latent structure spread all over the spectrum: the contributions of a compound lie in a subspace of the signal space. Orthogonal projection methods particularly address the latter case.

Figure 2 shows an illustration of OP for a 2-dimensional space. Consider 6 individuals $\{a...f\}$ characterized by 2 variables. In the absence of interfering factors, these points are aligned. Imagine that an interfering factor represented by a space spanned by a vector $\mathbf{p}$ shifts the measured points from $\{a...f\}$ to $\{A...F\}$. The effect of the interference is different for each point, but all share the same direction. Points $\{A...F\}$ are no longer aligned. A projection orthogonal to $\mathbf{p}$ yields the 6 corrected points {A*...F*}. Some properties of $\{a...f\}$ have been lost: the initial direction of the point alignment, the original distances between the points. However, other properties of $\{a...f\}$ that were lost in $\{A...F\}$ due to the interfering factor are restored in {A*...F*}: the alignment along a straight line and the relative position of the 6 points.

Figure 3 shows another example of OP in a higher dimension space. This example uses NIR reflectance spectra of corn, available in Blackburn.[19] Figure 3A shows these spectra put in $\mathbf{X}$. A linear baseline clearly affects the spectra. The general expression of a linear baseline is $a + bw$, where w is the wavelength and a and b are 2 coefficients. Thus, in the signal space, linear baselines are linear combinations of 2 vectors: $\mathbf{1}_Q$, a column of Q ones and $\mathbf{w}$ containing the Q wavelengths.

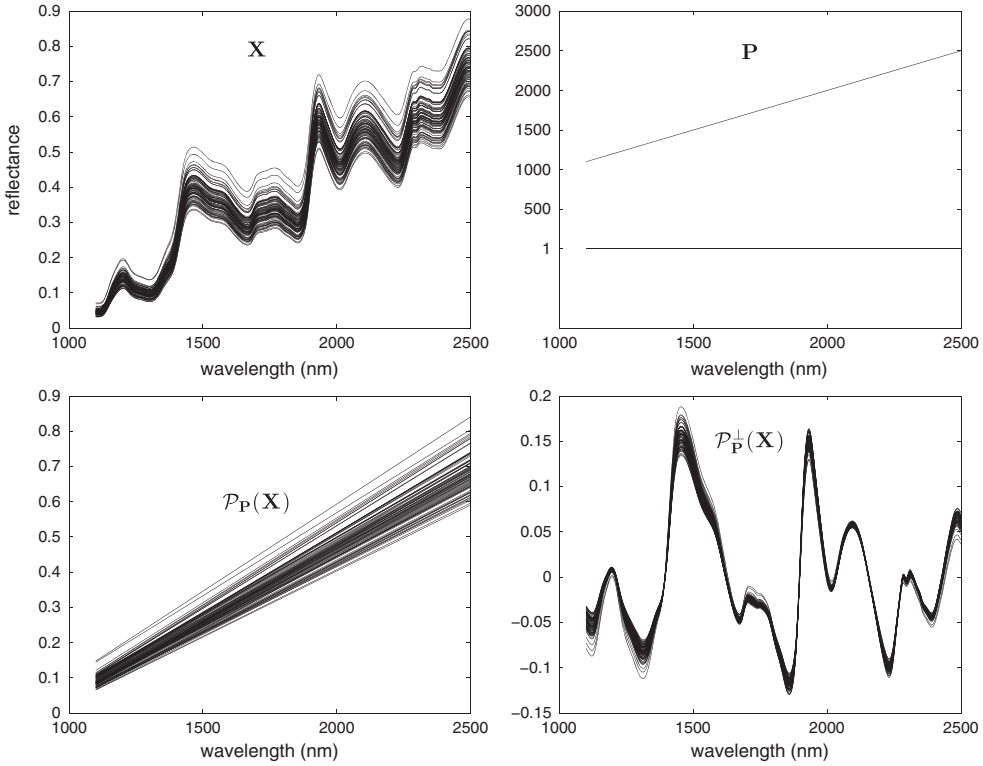

FIGURE 3 Example of orthogonal projection. A, Near-infrared reflectance spectra of corn. B, Basis of the linear baseline subspace. C, Projection of the spectra on the linear baseline subspace. D, Projection of the spectra orthogonal to linear baseline subspace

Let $\mathbf{P}$ be the matrix containing $(\mathbf{1}, \mathbf{w})$. $\mathbf{P}$ is illustrated in Figure 3B. Figure 3C shows $\mathcal{P}_{\mathbf{P}}(\mathbf{X})$, the projection of the spectra onto $\mathbf{P}$ that exhibits the baselines of the spectra, as Figure 3D shows $\mathcal{P}^{\perp}_{\mathbf{P}}(\mathbf{X})$, the projection of the spectra orthogonal to $\mathbf{P}$ that exhibits the baseline corrected spectra. This example shows that OPs can remove from the signal space any linear combination of the columns of a matrix $\mathbf{P}$.

2.3 | Useful and detrimental subspaces

The measured spectra are rarely formed only by the interference of the estimated response. Other chemicals and some physical factors such as product temperature or light scattering participate in the formation of the measured spectra. The signal space is the union of vector subspaces, each one being generated by the variations of a factor. In an ideal situation (Beer's law conditions), the spectral contribution of a chemical is included in the subspace spanned by its pure spectrum, which is of unitary dimension. Consequently, chemical factors ideally produce unitary subspaces. Physical factors generate more complex interferences and thus span larger dimensions. This fact was verified for the temperature,[7,20] and also for the turbidity.[21] Some other effects cannot easily be linked to a unique factor. For example, the vintage effect produces complex deformations in the signal space.[13,22] Thus, the spectral contribution of a response $\mathcal{Y}$ lies in a low dimensional subspace $\mathcal{E}^U$, which cohabits with a set of other subspaces having variable dimensions. The part of $\mathcal{E}^U$ orthogonal to all these subspaces is the part of the signal space that is sensitive to the response only, ie, the NAS.[23,24]

A linear calibration aims at calculating a regression vector $\mathbf{b}$ so that $\mathbf{x}^{\mathrm{T}}\mathbf{b}$ would be a good estimate of y for any couple $(\mathbf{x}, y)$ of spectrum and response. Two main types of linear calibrations have been reported[25]: the direct-classical and the indirect-inverse calibrations. Let $\mathbf{k}$ be the pure spectrum of the compound $\mathcal{Y}$ to be quantified. The direct-classical calibrations build the regression vector $\mathbf{b}$ according to the NAS definition. Thus, $\mathbf{b}$ is the projection of $\mathbf{k}$ orthogonal to the identified interfering factors, eg, the pure spectra of the other chemicals. Indirect-inverse calibrations aim at finding a linear relationship between $\mathbf{X}$ and $\mathbf{y}$ by other means. Nevertheless, the regression vector of each method is a good estimation of the NAS, as verified for PLSR.[26]

The concepts of useful space, detrimental space and regression vector are illustrated in Figure 4. The useful space $\mathcal{E}^U$ is the part of the signal space that contains the useful information. For direct-classical calibration methods, $\mathcal{E}^U$ is spanned by $\mathbf{k}$. For indirect-inverse methods such as PLSR, $\mathcal{E}^U$ is spanned by the latent variables. Every factor spanning a subspace

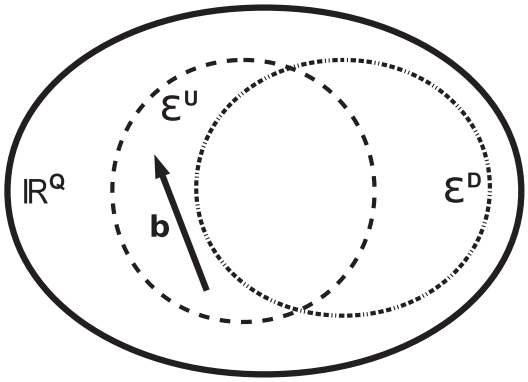

FIGURE 4 The useful and detrimental spaces $\mathcal{E}^U$ and $\mathcal{E}^D$; the regression vector $\mathbf{b}$

that does not intersect $\mathcal{E}^U$ has no detrimental effect on the model. The union of these subspaces is the useless space. Every factor spanning a subspace that does intersect $\mathcal{E}^U$ potentially has a detrimental effect. The union of these subspaces is the detrimental space $\mathcal{E}^D$. Typically, the regression vector $\mathbf{b}$ should lie in $\mathcal{E}^U$ but not in $\mathcal{E}^D$ (see Figure 4).

3 | THE MAIN PRINCIPLES

When a linear model is applied to a spectrum, the estimated response is given by $\hat{y} = \mathbf{x}^T\mathbf{b}$. Let $\delta\mathbf{x}$ be an alteration added to the spectrum, due to an interfering factor variation. The error induced on the estimation is given by $\delta\hat{y} = \delta\mathbf{x}^T\mathbf{b}$. Then,

$$|\delta\hat{y}| = \|\delta\mathbf{x}\| \times \|\mathbf{b}\| \times |\cos(\delta\mathbf{x}, \mathbf{b})|. \quad (1)$$

The Equation 1 shows that the absolute error caused by the interfering factor variation is the product of 3 positive terms. Lowering any of these terms should lower the error. Reducing $\|\delta\mathbf{x}\|$ is the aim of generic preprocessing as derivative or normalization. Reducing $\|\mathbf{b}\|$ belongs to the good modeling practices.[27]

The principle of the OP methods is to lower the term $|\cos(\delta\mathbf{x}, \mathbf{b})|$, to constrain the model to be as orthogonal as possible to the spectral perturbation. The parameter $\delta\mathbf{x}$ is not constant; $\delta\mathbf{x}$ varies inside the space $\mathcal{E}^D$. The parameter $\mathbf{b}$ belongs to $\mathcal{E}^U$. The orthogonality between $\mathbf{b}$ and $\delta\mathbf{x}$ is obtained by forcing $\mathcal{E}^U$ to be orthogonal to $\mathcal{E}^D$. This is done by removing from $\mathcal{E}^U$ the part that is colinear to $\mathcal{E}^D$, using OP in the signal space. Let $\mathcal{G}$ be a detrimental interference, and $\mathbf{G}$ a matrix containing a basis of the subspace spanned by $\mathcal{G}$. The following methods implement OP processing. They involve preprocessing and direct-classical calibration.

3.1 | Preprocessings using OPs

External parameter orthogonalisation (EPO[11]) and transfer by OP (TOP[14]) aim at making an existing calibration database robust against a specific interfering factor. External parameter orthogonalisation addresses physical interfering factors as temperature while TOP addresses discrete factors as spectrometer changes. The detrimental subspace is identified thanks to a specific experimental design, which exhibits examples of perturbations.

Dynamic OP (DOP[28]) aims at making an existing calibration database robust against changes in the measuring conditions that induce variations of unknown interfering factors. Dynamic OP assumes that several values of the response are known in correspondence with altered spectra. Originally addressing process monitoring, DOP has been successfully applied to remove block effects due to calibration transfer[15] or vintage changes.[13]

Error removal by orthogonal subtraction (EROS[29]) addresses the repeatability problem. In cases where many spectra are acquired for the same sample, a correspondence between a set of spectra and one value of the response is obtained. The variability contained in these spectra represents the repeatability error. This error is due to the reproduction of the measurement that can be considered an interfering factor $\mathcal{G}$. If this error was independently and identically distributed over the variables, the best processing would be to replace the repetitions by their mean. In spectrometry this error is often structured and yields a subspace of low dimension. Error removal by orthogonal subtraction proposes to remove this subspace by means of an OP.

Although it is not sensu stricto an OP, orthogonal signal correction (OSC) is briefly described because its concept relies on subspace separation. Orthogonal signal correction[30] aims at removing from **X** dimensions that carry a large part of the variance and are orthogonal to **y**. Orthogonal signal correction leads to computation problems because the orthogonality to **y** is defined in the sample space, and the correction must be realized in signal space to be applied on a single new spectrum. Several solutions to this problem can be found in other studies.[30-32] A review of the differences between these solutions and a discussion on their projection properties can be found in Boulet and Roger.[33] Orthogonal signal correction effectiveness is controverted. Indeed, its objective is strictly the dual of partial least squares (PLS), which aims at finding directions that both carry variance and are correlated to **y**. Thus, when associated with PLS, OSC lowers the number of latent variables, but the predictive ability of the model is not improved[31,34,35] because OSC removes the useless space, not the detrimental space. The integration of OSC processing inside the PLS algorithm led to orthogonal projections to latent structures.[36] This method inherits all the OSC properties.[35]

3.2 | Direct calibration methods using OPs

Methods related to direct-classical calibration are all based on the knowledge of the useful space and of the detrimental space. They estimate the NAS by modifying the useful space according to the detrimental space and then projecting the spectrum in the resulting space. The useful space is always the space spanned by the pure spectrum of $\mathcal{Y}$. Direct unmixing[25] and orthogonal subspace projector (OSP[37-39]) have been proved to be equivalent methods[40] based on the following hypotheses: (1) the detrimental space is exclusively due to chemical compounds; (2) all the pure spectra of these compounds are known. Direct unmixing and OSP consist of projecting $\mathcal{E}^U$ orthogonally to $\mathcal{E}^D$, to project the pure spec-

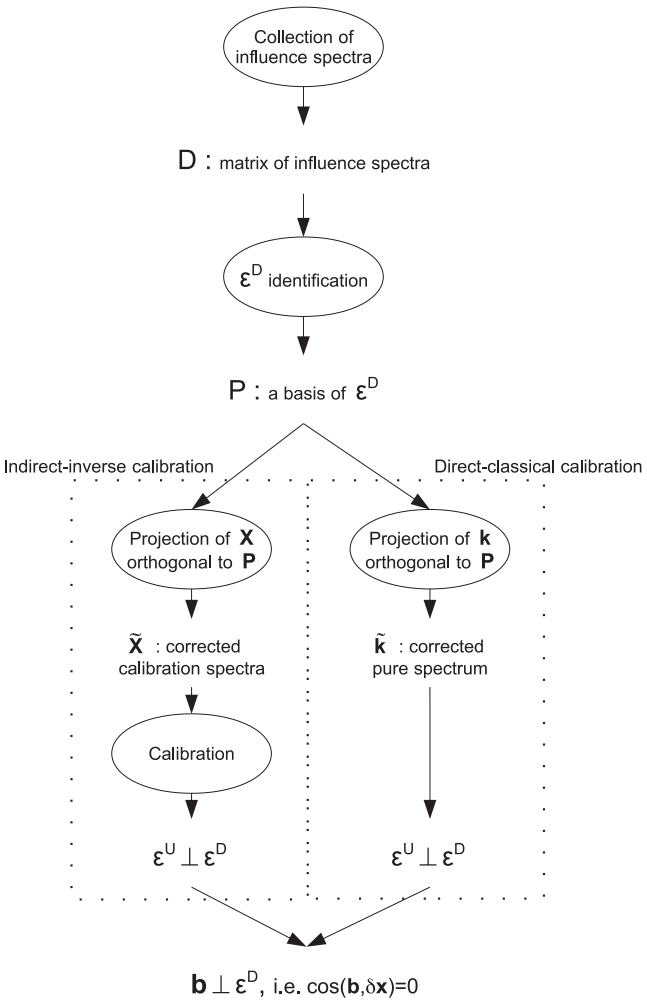

FIGURE 5 Main principle of orthogonal projection methods

trum of $\mathcal{Y}$ orthogonally to the other pure spectra. Improved direct calibration[41] operates the same projection, but performs a better estimation of $\mathcal{E}^D$. The information given by the available chemical spectra is completed with other effects.

All the above methods obey the principle described in Figure 5. A set of interference spectra belonging to the detrimental space $\mathcal{E}^D$ is collected and stored into an interference matrix **D**. Then, a basis of the space spanned by interferences is extracted and yields the matrix **P**. For indirect-inverse calibrations (left path of Figure 5), EPO, TOP, DOP, and EROS act as preprocessing. Spectra are projected orthogonally to **P**, yielding $\widetilde{\mathbf{X}}$. A new calibration is then computed on these preprocessed spectra. Because the useful space computed by this calibration is included in the signal space of $\widetilde{\mathbf{X}}$ and because this space is orthogonal to $\mathcal{E}^D$, the orthogonality between $\mathcal{E}^U$ and $\mathcal{E}^D$ is obtained. For direct calibrations (right path of Figure 5), direct unmixing, OSP, and improved direct calibration modify the pure spectrum of the response (**k**). Its projection orthogonally to **P** provides the orthogonality between $\mathcal{E}^U$ and $\mathcal{E}^D$.

Hence, the 2 most important steps for using OP methods are the detrimental information collection (matrix **D**) and the detrimental subspace estimation (matrix **P**).

4 | DETRIMENTAL INFORMATION COLLECTION (MATRIX D)

The matrix **D** is obtained by merging as much detrimental information as possible. The sources are very diverse and are summarized in Table 2. Consequently, the first step is to identify the sources of detrimental information in relation to the different situations that are encountered. The knowledge of the samples, the conditions of data acquisition and the variable to be predicted are of great importance when using an OP.

The first source of detrimental information is provided by the *expert knowledge*: eg, pure spectra or geometrical modelization.

- Pure spectra
  Detrimental interference from a chemical compound is easily characterized by its pure spectrum. Direct calibration (or spectral unmixing) and OSP[37,44] are based on this principle. Pure spectra can be obtained from databases and put into **D**.
- Geometrical modelization
  Some detrimental interferences can be modeled as geometrical curves. For example, light scattering affects NIR spectra with baselines. These baselines are classically estimated by polynomial functions of the column index j.[42,43,46] The set of all the kth order polynomial baselines is a subspace of the signal space. Let **V** be the Vandermonde matrix, rows of which contain $[1; 2; 3; ...Q]$ to powers 0, 1, 2, $\cdots$, k. Thus, the detrimental information carried by kth order polynomial baselines is obtained by putting **V** into **D**.

A second source of information is provided by the *experimental knowledge*: spectra acquired under specific conditions.

- Response-free spectra
  Samples may not contain the compound of interest but contain at least detrimental information. Independent interference reduction uses their spectra directly to build **D**.[45]
- Reference values allowing the estimation of ideal spectra

TABLE 2 Different methods for building **D**, according to the type of sources of information

Source of information		Methods for filling D	Refs
Expert		Geometrical modelization	42,43
	$y = 0$	Pure spectra of interference compounds	37,44
		Response-free spectra	45
Experimental		Estimation of ideal spectra	28
	$y \neq 0$	Repeating acquisitions	29
	$\triangle y = 0$	Experimental design without reference values	11,14
	$y \neq 0$	Experimental design with reference values	21
	$\triangle y \neq 0$		

This method addresses the case where a few altered spectra have been acquired, and the corresponding response values have been measured. Let $(\mathbf{X}_c, \mathbf{y}_c)$ be this dataset. The method relies on the estimation of $\hat{\mathbf{X}}_c$, the ideal spectra that would have been obtained without the detrimental interferences. This estimation can be performed on the basis of the calibration database. For example, in Zeaiter et al,[28] Gaussian kernels are used on the $\mathbf{y}$. Then $\hat{\mathbf{X}}_c - \mathbf{X}_c$ contains the detrimental information and is put into $\mathbf{D}$.

- A detrimental interference varying for a set of samples

  Spectral acquisitions are performed on the same set of samples at different levels of a detrimental interference. The reference values are not necessary and not processed. Two approaches have been proposed, differing on the way the useful information is removed:

  - In Figure 6, an experimental design is realized at several levels of the detrimental interference. At each level, spectra of the same set of samples are acquired. The mean spectrum of each level is calculated and put in a matrix $\mathbf{M}$. This matrix is then centered column-wise. This centering removes the useful information brought by the samples while preserving the detrimental information. Because of a loss of one dimension due to the centering, one of the resulting spectra is dropped. The others are put into $\mathbf{D}$. For example, in Roger et al,[11] the set of samples contained 10 apples, and the detrimental interference was the temperature with 8 levels. Seven spectra were put into $\mathbf{D}$.
  - In Figure 7, some samples are measured at several levels of an interfering factor. The difference with the previous case is that the interfering factor is not controlled. The spectra are gathered by sample. The corresponding submatrices are centered, thus removing the useful information. For the sample k, a matrix $\mathbf{D}_k$ is obtained. In Zhu et al,[29] scatter matrices $\mathbf{D'}_k\mathbf{D}_k$ are calculated and added. Merging of all the $\mathbf{D}_k$ to yield $\mathbf{D}$ is equivalent.

  The 2 approaches present some differences. Whereas the first needs an experimental design, the second can be applied in most situations. Usually, the second approach yields bigger $\mathbf{D}$ matrix and may catch more detrimental information.

- Experimental design with reference values

  Orthogonal signal correction was originally developed for processing calibration datasets. However, its algorithm has been applied to spectra acquired according to an experimental design merging the response and the interfering factors[21,47] to separate their respective contributions. Consider a collection of spectra acquired while $\mathcal{Y}$ and $\mathcal{G}$ vary. An OSC of these data yields loadings that represent directions containing $\mathcal{G}$ spectral effects but orthogonal to $\mathcal{Y}$ spectral effects. These vectors are put into $\mathbf{D}$.

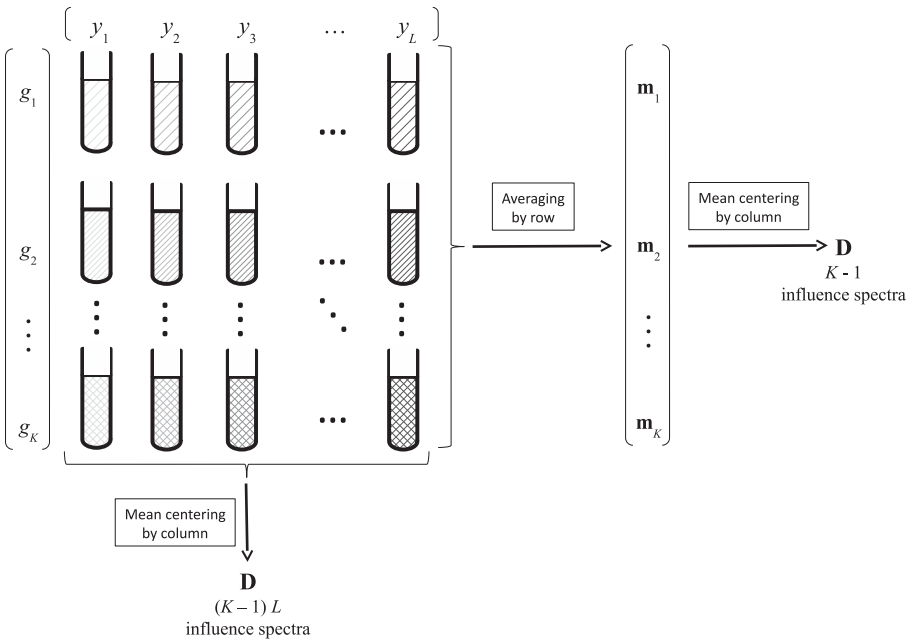

FIGURE 6 Illustration of collecting detrimental information by means of a complete experimental design. Response is figured by gray level while interfering factors are figured by patterns

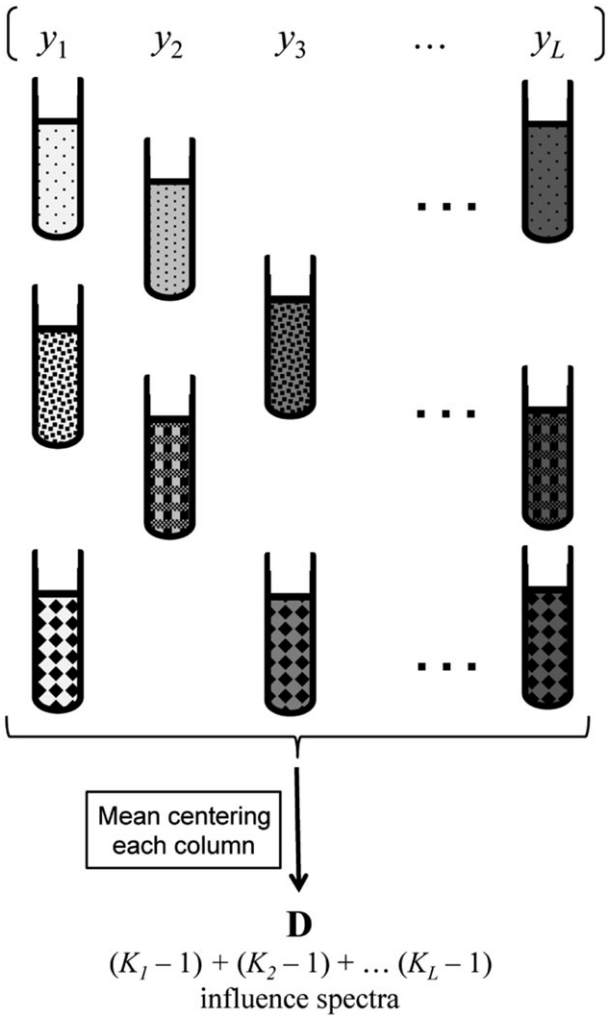

FIGURE 7 Illustration of collecting detrimental information without experimental design. Response is figured by gray level while interfering factors are figured by patterns

5 | DETRIMENTAL SUBSPACE ESTIMATION (MATRIX P)

Once the matrix **D** is obtained, a basis **P** of the detrimental subspace must be calculated. The matrix **P** contains R vectors of $\mathbb{R}^Q$, which best represent the detrimental information of **D**.

In Roger et al[11] and Andrew and Fearn,[14] the proposed method to identify **P** is principal components analysis (PCA). The first loadings of a PCA on **D** correspond to the main directions of **D** observed in the signal space, ie of $\mathcal{E}^D$. To avoid losing the first main direction, the **D** matrix must not be centered before PCA calculation. Special attention must be paid when using PCA because centering is a default preprocessing step performed by many commercial software. Alternately, **P** can be filled with the eigenvectors of $\mathbf{D}^T\mathbf{D}$, eg, by means of a singular value decomposition. Using the eigenvectors of $\mathbf{D}^T\mathbf{D}$ (from PCA or singular value decomposition) produces orthonormal vectors for **P**. This property makes the expression of the OP simpler: $\mathcal{P}_{\mathbf{P}}^{\perp} = \mathbf{I}_Q - \mathbf{P}\mathbf{P}^T$.

The value R corresponds to the dimension of the subspace removed by the OP. If R is too small, not all interferences are removed, and the correction is not complete. If R is too big, some useful dimensions can be removed, and consequently, the model becomes poorer. Thus, finding the optimal value R_{opt} is a key point in the OP setup. That issue is discussed below, using the dataset of the Chimiométrie 2007 contest[16] referred to as dataset 1, except for the last point dedicated to image analysis, which uses the data of Jaillais et al[48] referred to as dataset 2.

Dataset 1 is made up of 5 subsets. A calibration set contains 10 spectra of ground wheat and 10 protein contents. An experimental design dataset contains the spectra of 11 wheat samples prepared according to 2 moisture levels by 2 particle sizes and acquired under 3 temperatures by 2 repetitions. Thus, this experimental design dataset contains 11×2×2×3×2 =

264 spectra. Moreover, a first standardization dataset contains spectra of 10 sealed cells of ground wheat acquired on 31 spectrometers. A second standardization dataset contains spectra of 10 other cells acquired on 17 spectrometers. A test set contains 2000 spectra. In the challenge reported in Pierna et al,[16] the $\mathcal{Y}$ value of these samples were unknown. But in the following, we will use them. The $\mathbf{D}$ matrix was calculated according to the scheme of Figure 7.

Dataset 2 is built from multispectral images of wheat grain cuts. Each pixel of these images is described by $Q = 24$ variables, corresponding to the crossing of 3 RGB signals of a classical color camera by 8 monochromatic illuminations, as described in Chevallier et al.[49] Three classes of pixels were considered: external layers of the grain (walls, 1), endosperm of normal grains (normal, 2), and endosperm of waxy grains (waxy, 3). Area corresponding to these 3 classes were manually extracted, yielding 3 matrices of pure spectra $\mathbf{K}^j, j = 1, 2, 3$. The $\mathbf{D}$ matrix was obtained by merging $\mathbf{K}^2$ and $\mathbf{K}^3$ to reveal the compound 1, $\mathbf{K}^1$ and $\mathbf{K}^3$ to reveal the compound 2, and $\mathbf{K}^1$ and $\mathbf{K}^2$ to reveal the compound 3.

If enough data are available, performing the process with increasing values of R can efficiently guide the user to the optimal value. Applying the models to a separate validation set, the optimal value can be detected by observing the prediction error as a function of R. If the calibration is performed by means of a factorial method such as PLSR, the optimization of R and of the latent variable number must be realized simultaneously. As for any optimization process, once chosen the good value for R, the final model must be tested on another independent test set. Table 3 resumes the evolution of root mean square error of prediction (RMSEP) as a function of R and of the number of PLSR latent variables for dataset 1. Without any interference, considering the root mean square error of cross-validation (RMSECV) reported in the first row of Table 4, the best model uses 7 to 9 latent variables. When applied to the test set (Table 3, row 1, columns 7-9), we can notice that the factors involved in the dataset are very influential: the RMSEP is more than 7 times the RMSECV. The effect of OP corrections (Table 3, row 2 and higher) begins for $R = 2$ and is optimal for $R \approx 7$. Bigger values for R lead to a poorer model. For the optimal zone ($R \approx 7$), the model is very simple because it requires only 1 latent variable. The OPs have produced a 1-dimensional useful subspace yielding a good estimation of the NAS.

In the case where no test set is available or as a complement to the method presented above, some other methods can be used to choose R_{opt}.

5.1 | Adjusting R_{opt} on the eigenvalue evolution

As explained in Section 2.3, the spectral interferences to be removed are structured, and they, therefore, lie in a small subspace. It is, thus, natural to expect that only a few eigenvectors of $\mathbf{D}^T\mathbf{D}$ are necessary to represent $\mathcal{E}^D$. A simple approach for choosing R_{opt} is to search for a breakpoint in the evolution of the eigenvalues of $\mathbf{D}^T\mathbf{D}$, sorted in decreasing order. As this curve is always decreasing, it is difficult to make an objective decision. Figure 8 shows this curve. A visual inspection of the detail (right part of the figure) reveals 2 breaks at $R = 4$ and $R = 8$. In Zhu et al,[29] the authors suggest using a

TABLE 3 Evolution of the root mean square error of prediction of a partial least squares regression model calculated on dataset 1, as a function of the dimension R of the orthogonal projection and of the number LV of latent variables[a]

$R \setminus LV$	1	2	3	4	5	6	7	8	9
0	20.7	21.4	16.6	21.5	9.7	25.7	30.8	29.8	29.8
1	28.3	8.8	11.2	16.3	23.3	19.7	28.5	27.5	26.5
2	16.7	8.2	7.2	8.7	10.6	11.6	12.0	11.8	11.7
3	7.9	7.6	7.1	9.3	13.7	18.1	16.9	14.5	14.5
4	7.9	7.8	6.2	6.9	7.7	7.1	7.5	8.0	8.3
5	8.5	5.8	5.7	7.2	8.1	8.5	8.6	8.2	8.8
6	6.4	6.1	8.2	8.9	10.1	10.4	10.3	9.9	10.2
7	5.0	6.6	6.8	6.2	7.9	8.3	8.8	9.6	9.6
8	6.3	6.8	6.9	6.3	6.8	7.7	8.9	9.3	9.2
9	5.8	6.7	6.6	6.6	7.2	7.9	9.2	9.2	9.3
10	7.9	8.3	9.2	9.9	8.2	8.1	8.3	8.3	8.3
11	8.8	9.9	10.9	12.4	10.9	11.7	10.1	10.2	10.2
12	13.3	13.0	12.7	12.8	13.2	12.7	12.1	12.2	12.1

[a] $R = 0$ means "no projection."

TABLE 4 Evolution of the leave one out root mean square error of cross-validation of a partial least squares regression model calculated on dataset 1, as a function of the orthogonal projection dimension R and of the number of latent variables LV[a]

$R \setminus LV$	1	2	3	4	5	6	7	8	9
0	31.6	8.6	5.3	4.8	5.9	5.7	4.2	3.8	3.8
1	28.6	4.9	4.3	4.9	4.3	3.6	3.6	3.5	3.5
2	17.2	4.8	3.8	3.8	3.6	4.3	3.9	3.8	3.8
3	4.5	3.4	3.0	3.5	4.0	5.1	4.7	4.6	4.6
4	4.4	3.6	3.4	3.7	4.9	6.3	5.7	5.7	5.7
5	5.8	4.9	4.9	4.9	5.5	5.6	5.4	5.2	5.2
6	5.1	4.3	5.1	6.3	6.2	5.7	5.4	4.8	4.8
7	6.1	5.2	5.7	6.7	6.0	4.7	3.9	3.7	3.7
8	4.8	5.7	6.0	6.1	5.8	5.5	5.0	5.1	5.1
9	4.9	4.3	4.6	4.9	4.6	4.1	4.1	4.1	4.1
10	4.5	5.0	6.2	7.0	6.9	8.2	8.5	8.2	8.2
11	4.8	5.0	5.9	7.0	7.4	8.3	8.2	8.1	8.1
12	4.3	6.3	7.1	6.3	8.2	8.6	8.5	8.4	8.4
13	6.3	7.7	5.8	7.2	7.7	7.6	7.7	7.7	7.7
14	9.3	7.0	7.8	8.5	8.0	7.8	7.8	7.9	7.9

[a] $R = 0$ means "no projection."

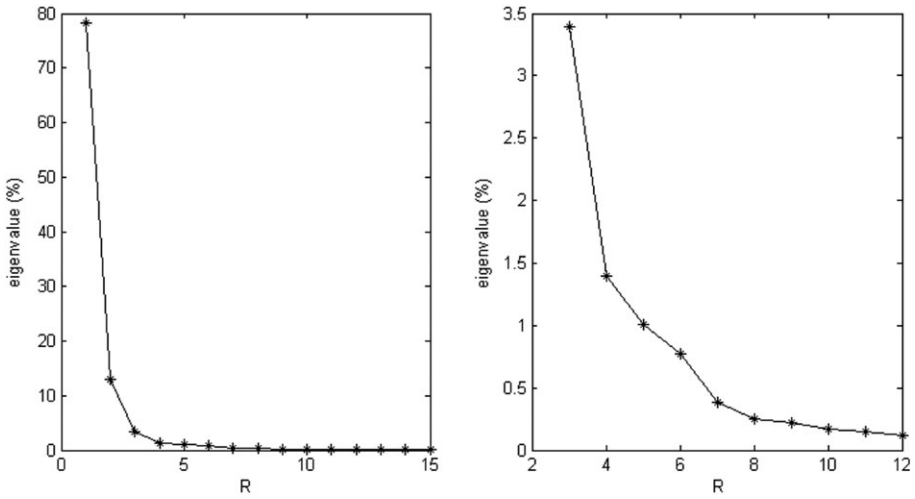

FIGURE 8 Evolution of $\mathbf{D}^T\mathbf{D}$ eigenvalues for dataset 1, as a function of R (left). Zoom on $3 \leq R \leq 12$ part (right)

threshold of 1% to cut this curve and to decide about the R_{opt} value; in our case, that method should produce a value of $R = 5$ or $R = 6$. Some statistical tests have been proposed to detect the breakpoints for these kind of curves.[50] Finally, we can conclude that this eigenvalue curve can give some clues about R_{opt}—in our case one breakpoint is very close to the true optimum—although an objective decision cannot be achieved.

5.2 | Adjusting R_{opt} with the cross-validation

For each value of R between 0—no projection—and R_{max}, a correction using R dimensions is performed on the calibration database and a regression is validated for 1 to A_{max} latent variables. The RMSECVs of all these models provide a matrix of dimensions $(A_{max} \times (R_{max} + 1))$. Because the calibration database is not supposed to contain any examples of interferences to be corrected, the positive effect of the projection cannot be observed plainly. A detrimental effect can, however, be observed. A visual inspection of this matrix can give a clue about the maximal value for R. Table 4 displays the results of this procedure for dataset 1. Surprisingly enough, this table would indicate a maximal value for R of about 3 or 4, which

is very far from the optimal value of 7. This difference is probably due to the very low number of calibration samples (10), which leads to a very stochastic cross-validation. Nonetheless, from our experience, this procedure yields interesting clues about the maximal value of R when more calibration samples are used.

5.3 | Adjusting R_{opt} with the Wilks' lambda evolution

When a set of observations is structured in groups, the Wilks' lambda (Λ) is defined as the ratio of the between group variance and the total variance. This criterion varies between 0 and 1; the better the group separation, the greater is the Λ.[51] Let consider the data regrouped by samples, ie, forming 11 groups of 24 spectra. The between group variance is carried by $\mathcal{E}^U$, while the within group variance is carried by $\mathcal{E}^D$. If there were no external interference and if the spectra were perfectly acquired, the spectra of each group would match, yielding a Λ of 1. Thus, as far as the OP removes detrimental information, Λ increases with R. However, the dimensions that are removed can also carry some information related to $\mathcal{Y}$. So removing too many dimensions can make Λ decreasing. Hence, plotting Λ as a function of R produces a curve that should present an initial increase, followed by a plateau or a decrease. R_{opt} corresponds to the maximum or to the breakpoint of this curve. Figure 9 shows the evolution of Λ as a function of R for dataset 1. From this curve, a breakpoint appears at $R_{opt} = 6$ and the maximum at $R_{opt} = 9$.

5.4 | Adjusting R_{opt} for spectral image analysis

In spectral images, each spectrum corresponds to a pixel, the exact composition of which is often inaccessible. In this case, the methods based on cross-validation or test are not feasible but the extra information yielded by the spatial structure of the image can be used as in.[48] Let assume that an area $\mathcal{A}$ of the image can be considered as homogeneous regarding the response $\mathcal{Y}$. The spectra of $\mathcal{A}$ are gathered in a matrix $\mathbf{K}^a$. Let $\mathbf{k}$ be the pure spectrum of $\mathcal{Y}$. Let $\mathbf{x}$ be a spectrum of $\mathbf{K}^a$ and $\mathcal{P}_R$ the OP using R dimensions. If the detrimental subspaces were correctly removed, then $\mathcal{P}_R(\mathbf{x})$ and $\mathcal{P}_R(\mathbf{k})$ would be almost collinear and their squared cosine would be close to 1. The average value $C(R)$ of the squared cosines is computed on $\mathcal{A}$ and plotted as a function of R. The optimum value R_{opt} is chosen according to a breakpoint or a maximum of this curve, corresponding to the correction that produces an area $\mathcal{A}$ the most homogeneous regarding $\mathcal{Y}$. Figure 10 presents the evolution of $C(R)$ for dataset 2. Each curve shows the evolution of $C(R)$ on an area of the grain, among "1: walls," "2: normal," and "3: waxy." As the pure spectrum $\mathbf{k}$ of each compound was unavailable, it has been replaced by the mean spectrum of the area, as proposed in Jaillais et al.[48] The higher $C(R)$, the more the pixels of the area globally resemble the mean pixel of the area. In this case, R_{opt} corresponds to the dimension of the subspace removed by the projection that makes the area homogeneous, regardless $\mathcal{Y}$. One can observe on Figure 10 a clear break in the curves for the areas *walls* and *waxy*, for $R_{opt} = 2$ and $R_{opt} = 8$, respectively. That would mean that *walls* area is naturally homogeneous and different from the other areas , also that there is no need to remove a lot of dimensions to make it appear clearly. On the contrary, *waxy* pixels contain a mix of spectral signals that are common to the other areas. For the *normal* area, no maximum appears on the curve, meaning that the OPs do not improve the identification of the area. The images resulting from this process can be found in Jaillais et al.[48]

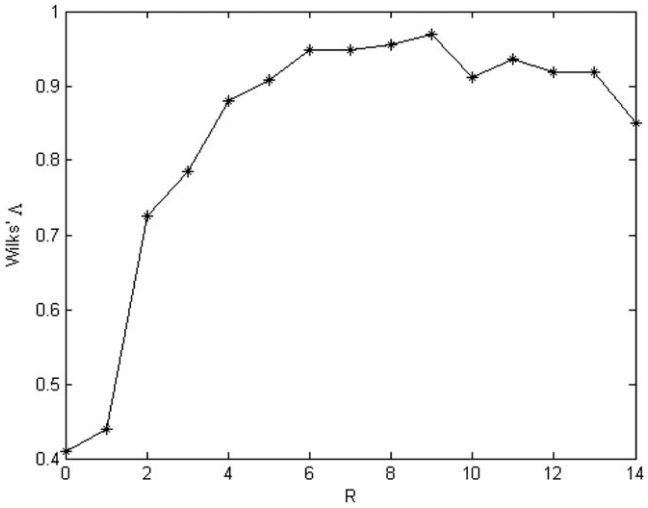

FIGURE 9 Evolution of the Wilks' Λ for dataset 1, as a function of R

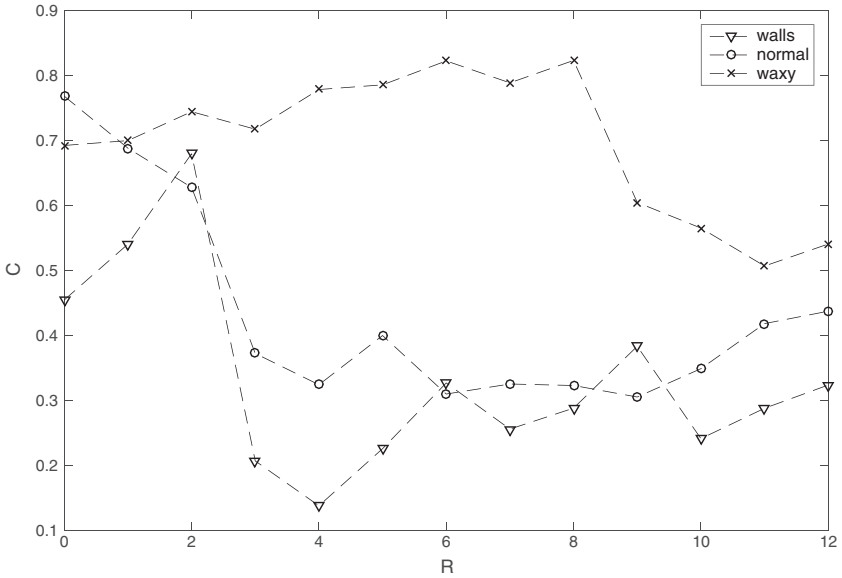

FIGURE 10 Evolution of the average of the squared cosines between the mean spectrum of the area $\mathcal{A}$ and the spectra of the area $\mathcal{A}$, as a function of the number of dimension removed by the projection (R)

6 | MANAGEMENT OF OPS

Three features must be managed in practical use: the embedding of the OP into the model, the management of several detrimental interferences, and the maintenance of OPs over time.

Orthogonal projections are embedded into the calibration models. The OPs remove $\mathcal{E}^D$ from the signal space, and $\mathbf{b}$ lies in the subspace obtained. So $\mathbf{b}$ is orthogonal to any vector belonging to $\mathcal{E}^D$. Consequently, there is no need to apply the OP to the new samples to be predicted. Nonetheless, this property applies only if no other preprocessing is performed on the spectra after the OP. Consequently, the combination of OP with other preprocessing should be examined carefully.

Several detrimental interferences can be compensated by OP. It was shown[33,52] that successive OPs cannot correct for all the interferences if they are not orthogonal to each other. Consequently, several OPs must be performed in paralell, not sequentially. The solution consists in merging all of the data into $\mathbf{D}$ before performing a single OP. This configuration is proposed in Brown[53] and was used to solve the Chimiométrie 2007 challenge.[16] Thus, special attention must be paid when combining some preprocessing with OPs. For example, Detrend[42,43] is formally equivalent to a projection orthogonal to a Vandermonde matrix (cf Section 4). Consequently, Detrend should neither be preceded nor followed by an OP. If we want to apply an EPO against temperature effects and a Detrend, it is mandatory to realize an OP with a $\mathbf{D}$ matrix including interferences due to temperature and a Vandermonde matrix.

The maintenance of OPs over time addresses many practical situations, eg, calibration of spectrometer networks or online sensor maintenance. Orthogonal projections can take a new detrimental interference that is not present in the original calibration dataset quickly into account. According to the previous point, it is not advised to perform successive OPs over time. One solution consists of retaining the original dataset and the different matrices $\mathbf{D}_k$ of detrimental information. When a new detrimental interference appears, a new $\mathbf{D}_{k+1}$ matrix is determined, then merged with the others, yielding $\mathbf{D}$. The OP is processed with $\mathbf{D}$. Stacking all these interference matrices should theoretically finally produce a complete detrimental subspace and thus a useful space close to the NAS. Nevertheless, if some $\mathbf{D}_i$ matrices are not well built, this stacking could eventually erode the useful space and make the model poorer.

Nevertheless, OPs also present some limits that must be known to users to prevent any misuse. Four issues are described hereafter.

The first issue is the added value of OPs within the global calibration process. Models using OPs are supposed to perform better than the same models without the OP. It is the case for direct calibration because OPs yield a better definition of the NAS. It is also the case for indirect calibration such as PLSR, if OPs use any detrimental information that is not included in the calibration dataset. When OPs do not use extra information, as OSC, we should not expect any improvement of the model accuracy. An alternative to OPs consists of augmenting the calibration dataset with examples altered by the interference. In Gujral et al,[54] the 2 solutions were compared in the framework of DOP, and it was concluded that OPs

outperformed model updates when the noise measurement rises. According to Hansen,[45] OPs minimize the need for reference values.

The second issue is the number of variables. Orthogonal projections are designed for data containing a large number of highly correlated variables, as spectra. Formally, the signal space must be large enough to contain 2 supplementary spaces related to the response and to the interfering factors, plus the measurement noise. For a small number of variables, it may be preferable to replace the removal of dimensions by shrinking the signal space. This shrinkage does not consume dimensions. It is performed according to the detrimental information, as suggested in science-based calibration[55] or general weighted least squares.[56]

The third issue concerns the multiplicative effects. Orthogonal projections handle additive effects only. If multiplicative effects are expected, then other methods should be used. The most famous methods are standard normal variate,[42,43] and the multiplicative signal correction family, ie, multiplicative signal correction[57] and extended multiplicative signal correction (EMSC).[58,59] It should be noticed that these methods partly use OPs. Standard normal variate removes the horizontal baseline of spectra by subtracting their mean value, which is a projection orthogonal to **1** in the signal space. Multiplicative signal correction removes linear baselines; EMSC removes additive interferents of any type: geometric baselines, theorically-based scattering effects, and chemical interferents. In Martens and Stark,[60] EMSC is described upon a model where the measured spectra contain a linear combination of the interferents. The contibutions of these interferents are calculated, and the interfering part is subtracted. In Chen et al,[61] it is proposed to perform this subtraction by projecting the spectra orthogonally to the subspace spanned by the interferents.

The fourth issue is about the possible erosion of the useful subspace that can occur when removing the detrimental subspace. Indeed, there is no guarantee that the spectra put into the **D** matrix do not contain information about the $\mathcal{Y}$. If that occurs, the directions calculated in **P** contain a part of the useful subspace. Consequently, the OP erodes a part of the useful subspace, and the model becomes poorer. This situation may particularly arise with the 2 methods using the response value, ie, with DOP and with OSC performed on experimental design. In this case, the uncertainty about the $\mathcal{Y}$ can affect the quality of the OP.

7 | CONCLUSION

This paper reviews OP methods that can be used to preprocess multivariate signals, as spectra. Orthogonal projections are based on very simple and long-known mathematical equations that are briefly revisited in this review. The main methods using OPs are presented. The setup of OPs relies on the identification of the detrimental subspace to be removed. This process is divided into 2 steps: the detrimental data collection and the detrimental subspace estimation. These 2 points are extensively developed and provide the users with useful tactical knowledge. Some particular points of caution are also listed to avoid a misuse of the OPs. It is worth noting that, because the OP methods remove detrimental information, the only risk they present is that they may erode the data and the model. In no case, they can overfit a solution.

Orthogonal projection softwares are available into the Free-Access Chemometrics Toolbox for Scilab (https://atoms.scilab.org/toolboxes/FACT) and into the free ChemFlow software (www.chemproject.org).

ACKNOWLEDGEMENTS

The authors are grateful to Pierre Dardenne and all his team at the Centre Wallon de Recherches Agronomiques, Gembloux, Belgium, who kindly provided the Chimiométrie 2007 challenge data and allowed their use in this paper.

ORCID

Jean-Michel Roger http://orcid.org/0000-0003-2123-5266

REFERENCES

1. Jolliffe IT. A note on the use of principal components in regression. *J R Stat Soc C-APP*. 1982;31:300-303.
2. Wold S, Martens H, Wold H. The multivariate calibration method in chemistry solved by the PLS model. In: Ruhe A, Kagstrom B, eds. *Proc. Conf. Matrix Pencils*, Lecture Notes in Mathematics. Heildelberg: Springler-Velay; 1983:286-293.
3. Gemperline PJ. Rugged spectroscopic calibration for process control. *Chemom Intell Lab Syst*. 1997;39(1):29-40.
4. Wortel VAL, Hansen WG, Wiedemann SCC. Optimising multivariate calibration by robustness criteria. *J Near Infrared Spectrosc*. 2001;9(2):141-151.

5. Zeaiter M, Roger JM, Bellon-Maurel V, Rutledge DN. Robustness of models developed by multivariate calibration. Part I the assessment of robustness. *TrAC Trends Anal Chem*. 2004;23(2):157-170.
6. Wang Y, Kowalski BR. Calibration transfer and measurement stability of near-infrared spectrometers. *Appl Spectrosc*. 1992;46(5): 764-771.
7. Wülfert F, Kok WT, Smilde AK. Influence of temperature on vibrationnal spectra and consequences for the predictive ability of multivariate models. *Anal Chem*. 1998;70:1761-1767.
8. Sánchez NH, Lurol S, Roger JM, Bellon-Maurel V. Robustness of models based on NIR spectra for sugar content prediction in apples. *J Near Infrared Spectrosc*. 2003;11:97-102.
9. Bogrekci I, Lee WS. *Trans ASABE*. 2006;49(4):1175-1180.
10. Peirs A, Tirry J, Verlinden B, Darius P, Nicolaı BM. Effect of biological variability on the robustness of NIR models for soluble solids content of apples. *Postharvest Biol Technol*. 2003;28(2):269-280.
11. Roger J. M., Chauchard F., Bellon-Maurel V. EPO-PLS external parameter orthogonalisation of PLS, application to temperature-independant measurement of sugar contents in fruits. *Chemom Intell Lab Syst*. 2003;66:191-204.
12. Minasny B, McBratney AB, Bellon-Maurel V, et al. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. *Geoderma*. 2011;167:118-124.
13. Roger JM, Chauchard F, Williams R. Removing the block effects in calibration by means of dynamic orthogonal projection. Application to the year effect correction for wheat protein prediction. *J Near Infrared Spectrosc*. 2008;16:311-315.
14. Andrew A, Fearn T. Transfer by orthogonal projection: making near-infrared calibrations robust to between-instrument variation. *Chemom Intell Lab Syst*. 2004;72:51-56.
15. Igne B, Roger JM, Roussel S, Bellon-Maurel V, Hurburgh CR. Improving the transfer of near infrared prediction models by orthogonal methods. *Chemom Intell Lab Syst*. 2009;99:57-65.
16. Pierna JAF, Chauchard F, Preys S, et al. How to build a robust model against perturbation factors with only a few reference value: a chemometric challenge at 'Chimiométrie 2007'. *Chemom Intell Lab Syst*. 2011;106:152-159.
17. Meyer KD. *Matrix Analysis and Applied Linear Algebra*. New York: SIAM; 2000.
18. Wold Sv, Esbensen K, Geladi P. Principal component analysis. *Chemometr Intell Lab Syst*. 1987;2(1-3):37-52.
19. Blackburn M. Corn dataset; 2005.
20. Hansen WG, Wiedemann SCC, Snieder M, Wortel VAL. Tolerance of near infrared calibrations to temperature variations: a practical evaluation. In: Davis AMC, Giangiacomo R, eds. *Near Infrared Spectroscopy: Proceedings of the 9th International Conference*. Verona, Italy: NIR Publications; 2000:307-311.
21. Preys S, Roger JM, Boulet JC. Robust calibration using orthogonal projection and experimental design, application to the correction of the light scattering effect on turbid NIR spectra. *Chemom Intell Lab Syst*. 2006;91:28-33.
22. Peirs A, Tirry J, Verlinden B, Darius P, Nicolaï B. Effect of biological variability on the robustness of NIR models for soluble solids content of apples. *Postharvest Biol Technol*. 2003;28:269-280.
23. Lorber A, Faber K, Kowalski BR. Net analyte analyte signal calculation in multivariate calibration. *Anal Chem*. 1997;69:1620-1626.
24. Lorber A. Error propagation and figures of merit for quantification by solving matrix equations. *Anal Chem*. 1986;58:1167-1172.
25. Martens H, Naes T. *Multivariate Calibration*. New York: Wiley; 1989.
26. Nadler B, Coifman RR. Partial least squares, Beer's law and the net analyte signal: statistical modeling and analysis. *J Chemom*. 2005;19:45-54.
27. Seasholtz MB, Kowalski BR. The parsimony principle applied to multivariate calibration. *Anal Chim Acta*. 1993;277:165-177.
28. Zeaiter M, Roger JM, Bellon-Maurel V. 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*. 2006;80:227-235.
29. Zhu Y, Fearn T, Samuel D, et al. Error removal by orthogonal substraction (eros): a customised pre-treatment for spectroscopic data. *J Chemom*. 2008;22:130-134.
30. Wold S, Antti H, Lindgren F, Ohman J. Orthogonal signal correction of near-infrared spectra. *Chemom Intell Lab Syst*. 1998;44:175-185.
31. Fearn T. On orthogonal signal correction. *Chemom Intell Lab Syst*. 2000;50:47-52.
32. Sjoblom J, Svensson O, Josefson M, Kullberg H, Wold S. An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra. *Chemom Intell Lab Syst*. 1998;44:229-244.
33. Boulet JC, Roger JM. Pretreatments by means of orthogonal projections. *Chemom Intell Lab Syst*. 2012;117:61-69.
34. Kemsley EK, Tapp HS. OPLS filtered data can be obtained from non-orthogonalizer PLS1. *J Chemom*. 2009;23:263-264.
35. Verron T, Sabatier R, Joffre R. Some theoretical properties of the O-PLS method. *J Chemom*. 2004;18:62-68.
36. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). *J Chemom*. 2002;16:119-128.
37. Harsanyi JC, Chang CI. Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. *IEEE-Trans Geosci Remote Sens*. 1994;32(4):779-785.
38. Chang CI, Althouse MLG. Unsupervised interference rejection approach to target detection and classification for hyperspectral imagery. *Opt Eng*. 1998;37(3):735-743.
39. Du Q, Chang CI. A comparative study for orthogonal subspace projection and constrained energy minimization. *IEEE-Trans Geosci Remote Sens*. 2003;41(6):1525-1529.
40. Settle JJ. On the relationship between spectral unmixing and subspace projection. *IEEE-Trans Geosci Remote Sens*. 1996;34(4):1045-1046.

41. Boulet JC, Roger JM. Improvement of direct calibration in spectroscopy. *Anal Chim Acta*. 2010;668:130-136.
42. Barnes RJ, Dhanoa MS, Lister SJ. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl Spectrosc*. 1989;43(5):772-777.
43. Barnes RJ, Dhanoa MS, Lister SJ. Correction to the description of standard normal variate (SNV) and de-trend (DT) transformations in Practical Spectroscopy with Applications in Food and Beverage Analysis—2nd edition. *J Near Infrared Spectrosc*. 1993;1:185-186.
44. Chang CI. *Hyperspectral Imaging, Techniques for Spectral Detection and Classification*. New York: Kluver Academic / Plenum; 2003.
45. Hansen PW. Pre-processing method minimizing the need for reference analysis. *J Chemom*. 2001;15:123-131.
46. Martens H, Nielsen JP, Engelsen SB. Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures. *Anal Chem*. 2003;75:394-404.
47. Boulet JC, Doco T, Roger JM. Improvement of calibration models using two successive orthogonal projection methods, application to quantification of wine mannoproteins. *Chemom Intell Lab Syst*. 2007;87:295-302.
48. Jaillais B, Boulet JC, Roger JM, Balfourier F, Berbezy P, Bertrand D. Application of direct calibration in multivariate image analysis of heterogeneous materials. *Anal Chim Acta*. 2012;734:45-53.
49. Chevallier S, Bertrand D, Kohler A, Courcoux P. Application of PLS-DA in multivariate image analysis. *J Chemom*. 2006;20(5):221-229.
50. Jackson JE. *A User's Guide to Principal Components*. New York: Wiley and Sons; 1991.
51. Lebart L, Morineau A, Warwick KM. *Multivariate Descriptive Statistical Analysis*. New York: Wiley and Sons; 1984.
52. Piziak R, Odell PL, Hahn R. Constructing projectors on sums and intersections. *Comput Math Appl*. 1999;37:67-74.
53. Brown JM. Method for correcting spectral data for data due to the spectral measurement process itself and estimating unknown property and/or composition data of a sample using such method. U.S. Patent; 1992.
54. Gujral P, Amrhein M, Bonvin D. Drift correction in multivariate calibration models using on-line reference measurements. *Anal Chim Acta*. 2009;642(1):27-36.
55. Marbach R. A new method for multivariate calibration. *J Near Infrared Spectrosc*. 2005;13:241-254.
56. Martens H, Høy M, Wise BM, Bro R, Brockhoff PB. Pre-whitening of data by covariance-weighted pre-processing. *J Chemom*. 2003;17(3):153-165.
57. Geladi P, MacDougall D, Martens H. Linearization and scatter-correction for near infrared reflectance spectra of meat. *Appl Spectrosc*. 1985;39(3):491-499.
58. Martens H, Bruun SW, Adt I, Sockalingum GD, Kohler A. Pre-processing in biochemometrics: correction for path-length and temperature effects on water in FTIR bio-spectroscopy by emsc. *J Chemom*. 2006;20:402-417.
59. Kohler A, Bocker U, Warringer J, Blomberger A, Omholt SW, Martens H. Reducing inter-replicate variation in Fourier-transform infrared spectroscopy by extended multiplicative signal correction. *Appl Spectrosc*. 2009;63(3):296-305.
60. Martens H, Stark E. Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. *J Pharm Biomed Anal*. 1991;9(8):625-635.
61. Chen Z-P, Morris J, Martin E. Extracting chemical information from spectral data with multiplicative light scattering effects by optical path-length estimation and correction. *Anal Chem*. 2006;78(22):7674-7681.

How to cite this article: Roger J-M, Boulet J-C. A review of orthogonal projections for calibration. *Journal of Chemometrics*. 2018;32:e3045. https://doi.org/10.1002/cem.3045