



HAL
open science

IDC-Iproved direct calibration: a new direct calibration method applied to hyperspectral image analysis

Jc Boulet, N. Gorretta, J.M. Roger

► **To cite this version:**

Jc Boulet, N. Gorretta, J.M. Roger. IDC-Iproved direct calibration: a new direct calibration method applied to hyperspectral image analysis. *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009 (IEEE, First Workshop on)*, Aug 2009, Grenoble, France. 4 p., 10.1109/WHISPERS.2009.5289094 . hal-00468860

HAL Id: hal-00468860

<https://hal.science/hal-00468860>

Submitted on 31 Mar 2010

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.

IDC-IMPROVED DIRECT CALIBRATION: A NEW DIRECT CALIBRATION METHOD APPLIED TO HYPERSPECTRAL IMAGE ANALYSIS

Jean-Claude BOULET*

Nathalie GORRETTA, Jean-Michel ROGER

INRA
UMR-SPO

2 Place Viala, Montpellier, France

CEMAGREF
UMR-ITAP

361 Rue JF Breton, Montpellier, France

ABSTRACT

Calibrating consists in predicting \mathcal{Y} , a quantitative variable of interest, using P explaining variables. PLSR-Projection to Latent Structures Regression (Wold,[1]) is the most popular method, very powerful when a calibration dataset is available. Other methods don't need a calibration dataset, they are called direct calibrations. Two of them have been proposed previously: DC-Direct Calibration (as described in Martens and Naes, [2]) and SBC-Science Based Calibration (Marbach([3])). New method called IDC-Improved Direct Calibration is proposed. As for DC, this approach is based on an orthogonal projection. IDC projector is obtained by merging DC projector (consisting only in pure spectra of chemical compounds), and vectors characterising physical influence factors (consisting in PCA loadings onto a design dataset). Indeed, hyperspectral image analysis is a case where calibration data are not available. Thus, it's interesting to use direct calibration methods instead of PLSR. With the prior knowledge of a few reference spectra, and modelling some noise from the hyperspectral image itself, it's possible to identify objects of interest from the background. This method is also simple and understandable, very quick and easy to compute.

Index Terms— direct calibration, DC, SBC, IDC.

1. INTRODUCTION

Fast analysis methods to control or analyse processes are of great importance. Spectrometry is often chosen. Indeed, it has many advantages: very quick, sample preservation, low cost. Most of predictions are made using PLSR-Projection to Latent Structures Regression, a supervised method which needs a calibration dataset. However, the building of the calibration dataset can be a real problem. Sometimes the cost of the reference analysis is very expensive, other times it's not possible to obtain a calibration dataset. In these cases, it can be interesting to use alternative unsupervised methods, as direct calibration approach based on pure spectra knowledge.

*Thanks to Ralph MARBACH (VTT Electronics, Finland) for his useful discussions and advices.

In the litterature, two methods have been proposed. The first one is called DC-Direct Calibration. Recently, Marbach proposed a new calibration method called SBC-Science based Calibration. This paper proposes a third method, based on the two previous ones, called IDC-Improved Direct Calibration. Firstly a theoretical part is devoted to the presentation of these three methods. Then experimental results obtained on hyperspectral image, identification of wheat leaves using the IDC method, are presented and discussed.

2. THEORY.

Matrices will be noted with uppercase boldface characters, vectors with lowercase boldface characters, scalars with lightface characters. The following notations will be also used:

\mathcal{Y} : interest factor

\mathbf{X} : matrix (N, P) containing the values of P explaining variables measured onto N samples

$\hat{\mathbf{X}}$: good estimation of \mathbf{X} : $\hat{\mathbf{X}} - \mathbf{X} \approx \mathbf{O}$ the null matrix

\mathbf{y} : column vector $(N, 1)$ containing the values of \mathcal{Y}

\mathbf{X}_G : matrix (N_G, P) obtained onto a dataset different from \mathbf{X} , and containing only information from the influence factors.

\mathbf{I} : identity matrix, (P, P)

$\Sigma_{DC}, \Sigma_{SBC}, \Sigma_{IDC}$: symmetrical matrices (P, P) associated respectively to DC, SBC and IDC.

\mathbf{K} : matrix $(Q - 1, P)$ where the lines are \mathbf{k}_i , the $(Q - 1)$ pure influence factor spectra. $\{\mathbf{k}_1 \dots \mathbf{k}_{Q-1}\}$ is a base of \mathcal{E}_X

\mathbf{K}^* : matrix obtained by merging \mathbf{K} and \mathbf{k}

\mathbf{P} : matrix (A, P) ; the lines $\{\mathbf{p}_1 \dots \mathbf{p}_A\}$ are a base of \mathcal{E}_ϕ

\mathbf{R} : matrix $(A + Q - 1, P)$ obtained by merging \mathbf{K} and \mathbf{P}

\mathbf{T}_X et \mathbf{T}_ϕ : matrices $(N, Q - 1)$ and (N, A) of the concentrations or scores onto N samples associated to the vectors defined respectively by \mathbf{K} and \mathbf{P}

\mathbf{k} : vector $(1, P)$, pure spectra of the interest factor

Let \mathbf{X} , \mathbf{y} and $\hat{\mathbf{y}}$ the prediction of \mathbf{y} using \mathbf{X} . $\hat{\mathbf{y}}$ is calculated with a calibration model using a b-coefficients vector $(P, 1)$ named $\hat{\mathbf{b}}$ which verifies:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}}$$

Calibration methods differ in the way $\widehat{\mathbf{b}}$ is estimated.

2.1. Theory of DC.

In order to avoid confusions, terms 'DC' and 'Direct Calibration' will denote the method presented by Martens and Naes ([2]); otherwise the term 'direct calibration' will refer to one of the three existing methods. The assumption of DC is that all spectra of all chemical compounds in the analysed samples are known. In accordance with the mixing law derived from Beer-Lambert, the final spectra is the contribution of all spectra of the chemical compounds, weighted by their concentrations. Assuming that \mathbf{X} is a (N, P) matrix of N spectra, \mathbf{Y} a matrix (N, Q) containing the concentrations of the Q compounds for the N samples, and \mathbf{K}^* a matrix (Q, P) of the pure compound spectra, a good estimation of \mathbf{X} , noted $\widehat{\mathbf{X}}$, is given by Equation (1).

$$\widehat{\mathbf{X}} = \mathbf{Y}\mathbf{K}^* \quad (1)$$

We suppose now that all of the pure spectra are different, and that we have less compounds than explaining variables, i.e. $Q < P$. Thus $(\mathbf{K}^*\mathbf{K}^{*\prime})$ is invertible. Then multiplying on the right each member of (1) by $\mathbf{K}^{*\prime}(\mathbf{K}^*\mathbf{K}^{*\prime})^{-1}$ leads to the classical solution:

$$\widehat{\mathbf{Y}} = \mathbf{X}\mathbf{K}^{*\prime}(\mathbf{K}^*\mathbf{K}^{*\prime})^{-1}$$

In the case of quantification of one interest factor \mathcal{Y} , the solution has been demonstrated by Marbach ([3]). From \mathbf{K}^* , the line corresponding to the spectrum of the interest factor, i.e. \mathbf{k} , is deleted. Thus \mathbf{K} is obtained. Σ_{DC} is the (P, P) orthogonal projector to \mathbf{K} :

$$\Sigma_{\text{DC}} = \mathbf{I} - \mathbf{K}'(\mathbf{K}\mathbf{K}')^{-1}\mathbf{K}$$

The solution can then be written:

$$\widehat{\mathbf{y}}_{\text{DC}} = \mathbf{X}\Sigma_{\text{DC}}\mathbf{k}'(\mathbf{k}\Sigma_{\text{DC}}\mathbf{k}')^{-1}$$

and:

$$\widehat{\mathbf{b}}_{\text{DC}} = \Sigma_{\text{DC}}\mathbf{k}'(\mathbf{k}\Sigma_{\text{DC}}\mathbf{k}')^{-1}$$

2.2. Theory of SBC.

The equation (1) is very theoretical. In practice there is always noise added to the signal of the interest factor. If we note \mathbf{E} this error, then equation (1) is not true and should be replaced by:

$$\widehat{\mathbf{X}} = \mathbf{Y}\mathbf{K} + \mathbf{E} \quad (2)$$

Marbach supposes that noise is random-distributed around he mean, i.e; \mathbf{E} is centered. He proposes a solution to minimise its influence, increasing signal-to-noise ratio. First, an experimental design is dedicated to the obtention of spectra where variations are due only to the influence factors -the interest

factor has no effect onto this spectra- The obtained centered-matrix is called \mathbf{X}_{G} . Marbach's main idea lies in the fact that variables that have a great variability in \mathbf{X}_{G} are the same that have a great variability in \mathbf{E} . The influence of these variables can be minimized multiplying by \mathbf{X}_{G} variance-covariance inverse matrix. Let note:

$$\Sigma_{\text{SBC}} = (\mathbf{X}'_{\text{G}}\mathbf{X}_{\text{G}})^{-1}$$

If \mathbf{X}_{G} is not invertible, we can also replace \mathbf{X}_{G} in previous equation by $\mathbf{X}_{\text{G}}\mathbf{Q}'\mathbf{Q}$, its projection onto \mathbf{Q} the first loadings of a PCA onto \mathbf{X}_{G} . After Σ_{SBC} is determined, multiplication of each term of equation (2) by Σ_{SBC} is performed. Assuming that $\mathbf{E}\Sigma_{\text{SBC}}$ can be approximated to the null matrix leads to the Marbach solution:

$$\widehat{\mathbf{y}}_{\text{SBC}} = \mathbf{X}\Sigma_{\text{SBC}}\mathbf{k}'(\mathbf{k}\Sigma_{\text{SBC}}\mathbf{k}')^{-1}$$

and:

$$\widehat{\mathbf{b}}_{\text{SBC}} = \Sigma_{\text{SBC}}\mathbf{k}'(\mathbf{k}\Sigma_{\text{SBC}}\mathbf{k}')^{-1}$$

2.3. Theory of IDC.

Now, we suppose that we are able to describe correctly both chemical and physical influence factors. Chemical influence factors are described by the corresponding pure spectra, in a \mathbf{K} matrix of dimension $(Q-1, P)$. Physical influence factors are described by a \mathbf{P} matrix, dimension (A, P) , determined in the same way as \mathbf{Q} in the SBC. Then, it exists unknown matrices \mathbf{T}_{χ} and \mathbf{T}_{ϕ} that verify:

$$\widehat{\mathbf{X}} = \mathbf{y}\mathbf{k} + \mathbf{T}_{\chi}\mathbf{K} + \mathbf{T}_{\phi}\mathbf{P} \quad (3)$$

In order to minimise \mathbf{K} and \mathbf{P} , an orthogonal projection is performed, instead of using SBC variance-covariance inverse. First \mathbf{K} and \mathbf{P} are merged to obtain \mathbf{R} , matrix $(Q-1+A, P)$. Then the orthogonal projector Σ_{IDC} is calculated:

$$\Sigma_{\text{IDC}} = \mathbf{I} - \mathbf{R}'(\mathbf{R}\mathbf{R}')^{-1}\mathbf{R}$$

Thus, terms of equation (3) are multiplied on the right by Σ_{IDC} . After simplification it gives next equations:

$$\widehat{\mathbf{y}}_{\text{IDC}} = \mathbf{X}\Sigma_{\text{IDC}}\mathbf{k}'(\mathbf{k}\Sigma_{\text{IDC}}\mathbf{k}')^{-1}$$

and:

$$\widehat{\mathbf{b}}_{\text{IDC}} = \Sigma_{\text{IDC}}\mathbf{k}'(\mathbf{k}\Sigma_{\text{IDC}}\mathbf{k}')^{-1}$$

3. MATERIAL AND METHODS.

A Hypslex-VNIR 1600 (Norsk Elektro Optikk, Norway) hyperspectral system was used to acquire an image in 160 narrow wavebands (400-980 nm, with a 3.6 nm resolution). Pure spectra were acquired later, using a Zeiss MMS1 spectrophotometer with a 310-1150 nm spectral range and a 3.3 nm spectral resolution. A pre-processing was applied on each

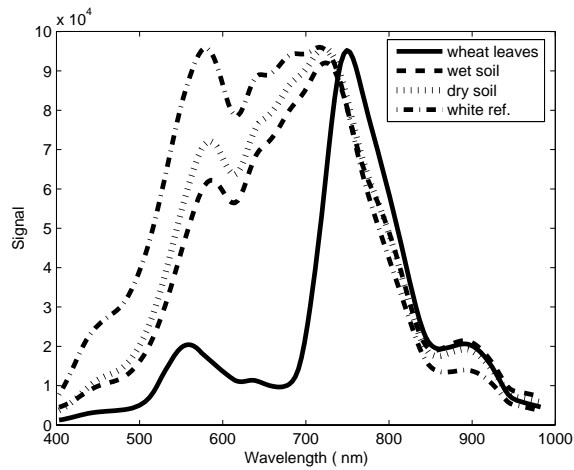


Fig. 1. Pure spectra acquired using a spectrophotometer.

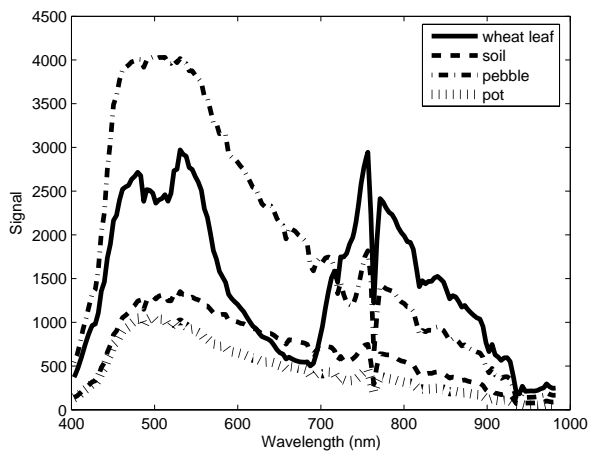


Fig. 2. Spectra extracted from the hyperspectral image.

data. The image spectra were normalized: each spectrum was divided by its raw signal at 800 nm. Zeiss raw spectra were replaced by a linear extrapolation of their corresponding values in the camera range (160 wavebands). No other pre-processing was achieved. So, signal intensities remain spectral units for respectively the camera and the spectrophotometer. All data were processed with Matlab software.

To determine pure spectra, 12 spectra were acquired onto 4 samples using the Zeiss spectrophotometer: fresh wheat leaves, wet and dry soil, and a white ceramic. The mean of fresh leaves spectra is considered as the interest compound spectrum k , whereas the means of white ceramic, wet and dry soil, are considered as pure spectra of the previously-known influence factors, giving matrix K . Other influence factors spectra were extracted from the image itself. They concern variabilities in the background composition that hadn't been taken into account by the pure spectra: i.e. pebbles and pot.

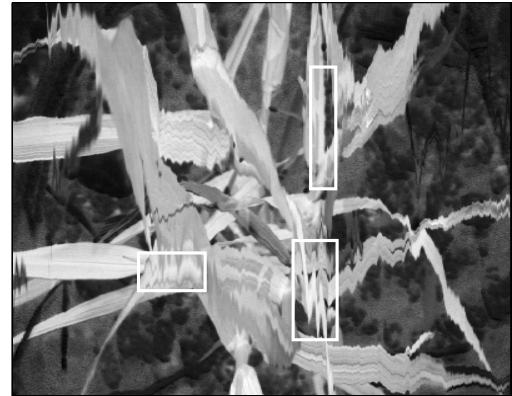


Fig. 3. NDVI image.

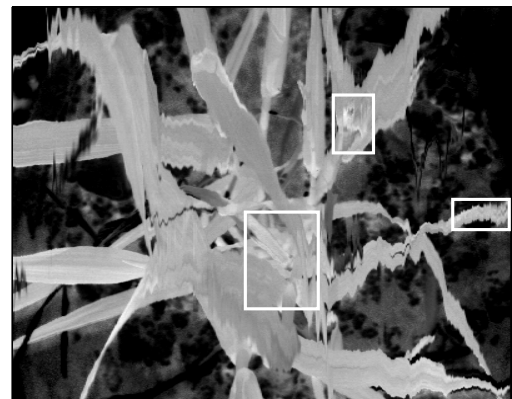


Fig. 4. IDC image.

Then the 7 first PCA loadings onto the matrix containing these spectra were kept, yielding P . A grey-scale image was then computed, the maximum and minimum values are respectively affected to white and black levels.

To compare the results obtained with IDC to existing methods, the NDVI- Normalised Difference Vegetation Index (Rouse,([4])) was calculated. This index used with pre-determined channels on satellite sensors is well-known for its ability to separate vegetation from background. In our application, the red channel was estimated by the signal at 650 nm, the NIR one by the signal at 850 nm. As for IDC, NDVI grey scale image was computed.

4. RESULTS AND DISCUSSION.

Figures (1) and (2) show the raw spectra shapes in both acquisitions. We can note strong differences among them, de-

pending mainly from the light source, the sun for the image, an artificial light for the pure spectra. To correct this source of variability, an usual approach is to transform raw signal into reflectance. In our example, both white reference and pebble spectra could have been used to perform this correction. However, as these spectra were introduced into IDC model, a correction is automatically performed by orthogonal projection.

Obtained grey scale images (NDVI and IDC results) are given respectively in Figures (3) and (4). In both case, a good discrimination between leaves and background is obtained. NDVI appears very powerful. Indeed, leaves and soil spectra are greatly different one from the other. However, some differences between NDVI and IDC predictions can be observed. Thus, on one hand, leaf prediction using IDC appears more homogeneous in some regions, see white-edge windows in Figure (4). In these regions, we suppose that leaves have moved during data acquisition, due to the wind, and the resulting spectra are a mix of leaf and background spectra. On the other hand, NDVI highest values are obtained by spectral perturbations on leaves spectra, see white-edge windows in Figure (3), whereas these perturbations have totally disappeared in IDC image. To conclude, IDC appears to be more robust to varying data acquisition conditions than NDVI.

5. CONCLUSION.

IDC is a direct calibration method suitable for quantification of an interest compound in complex matrices. Contrary to PLSR, it takes advantage of external chemical and physical information. Pure spectra are known information which is directly computed into the model. More unknown informations

as structured noise are first modelled, then condensed using a PCA. The corresponding loadings are computed into the model. IDC and SBC theories are close one to the other, but matrix inversion isn't a problem with IDC, contrary to SBC where choices have to be done. It's easier to take into account a new influence factor with IDC than with SBC. Also IDC is very easy to understand and to compute, it needs only a PCA routine.

IDC was applied first in classical spectrometry, but its properties and the previous example show that this method has potential applications in hyperspectral image analysis. Its interest should increase with the complexity of the image, in particular when each pixel spectra is a mix of interest and influence factor spectra.

6. REFERENCES

- [1] S.Wold, A.Ruhe, H.Wold, and W.J. Dunn III, "The collinearity problem in linear regression. the partial least square (pls) approach to generalized inverses.," *Journal of Science and Statistical Computations*, vol. 5, pp. 735–743, 1984.
- [2] H.Martens and T.Naes, *Multivariate Calibration*, Wiley, 1989.
- [3] R.Marbach, "A new method for multivariate calibration.," *Journal of Near Infrared Spectroscopy*, vol. 13, pp. 241–254, 2005.
- [4] J.Rouse, R.Haas, J.Schell, and D.Deering, "Monitoring vegetation systems in the great plains with erts," *Third ERTS Symposium, NASA*, vol. 1, pp. 309–317, 1973.