International Journal of Pharmaceutics 373 (2009) 179–182

Contents lists available at ScienceDirect

International Journal of Pharmaceutics

journal homepage: www.elsevier.com/locate/ijpharm

Note

A comparison of a common approach to partial least squares-discriminant analysis and classical least squares in hyperspectral imaging

José Manuel Amigo [a], Carsten Ravn [b,c], Neal B. Gallagher [d], Rasmus Bro [a,*]

[a] Department of Food Science, Quality and Technology, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark
[b] Faculty of Pharmaceutical Sciences, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark
[c] Novo Nordisk A/S, Novo Nordisk Park, DK-2760 Måløv, Denmark
[d] Eigenvector Research Inc., Manson, WA 98831, USA

ARTICLE INFO

Article history:
Received 23 October 2008
Received in revised form 14 February 2009
Accepted 17 February 2009
Available online 3 March 2009

Keywords:
Hyperspectral imaging
CLS
PLS-DA
PLS-Classification
PLS-Class

ABSTRACT

In hyperspectral analysis, PLS-discriminant analysis (PLS-DA) is being increasingly used in conjunction with pure spectra where it is often referred to as PLS-Classification (PLS-Class). PLS-Class has been presented as a novel approach making it possible to obtain qualitative information about the distribution of the compounds in each pixel using little a priori knowledge about the image (only the pure spectrum of each compound is needed). In this short note it is shown that the PLS-Class model is the same as a straightforward classical least squares (CLS) model and it is highlighted that it is more appropriate to view this approach as CLS rather than PLS-DA. A real example illustrates the results of applying both PLS-Class and CLS.

1. Introduction

The powerful combination of hyperspectral images and multivariate analysis has not been overlooked by the pharmaceutical industry and the number of pharmaceutical applications has increased significantly the past few years (Gowen et al., 2007). One of the major challenges is selecting the appropriate data analysis methodology for extracting desired information from hyperspectral images. In hyperspectral images, a high-resolution spectrum is obtained for each pixel and for example, principal component analysis (PCA) (Amigo et al., 2008; Kohler et al., 2007) has been used for unsupervised pixel classification, multivariate curve resolution—alternating least squares (MCR-ALS) (Amigo and Ravn, 2009; de Juan et al., 2004) has been used for chemical mapping and interpretation of images, and partial least squares (PLS) regression has been used for quantification (Ravn et al., 2008).

PLS-discriminant analysis (PLS-DA) (Chevallier et al., 2006) is being increasingly used in conjunction with pure spectra where it is often referred to as PLS-Classification (PLS-Class) or 'PLS2 using a library of pure component spectra' (see, for example references Clarke, 2004; Furukawa et al., 2007; Henson and Zhang, 2006; Lyon et al., 2002; Ma and Anderson, 2008; Weiyong et al., 2008; Westenberger et al., 2005). This method is based on the development of a PLS2 calibration model with a calibration matrix with one pure spectrum for each chemical component and a dummy matrix as target matrix. It has been advocated as attractive because it merges the easiness of obtaining a calibration set (just the pure spectra of the components) and the exploratory and visualizing properties of the PLS2 model. In this paper, it is demonstrated that the PLS-Class approach is equivalent to performing classical least squares (CLS) (Amigo et al., 2008; Gallagher, 2007; Martens and Naes, 1984; Naes and Martens, 1984). This is shown both mathematically and by means a practical example.

2. PLS-Classification in hyperspectral imaging and its similarity with CLS

PLS-DA is a PLS2-based model, and therefore, the general model form can be written as Eq. (1):

$$\boldsymbol{X}\hat{\boldsymbol{B}} = \boldsymbol{Y} \tag{1}$$

where $\mathbf{X}$ is an $M \times J$ matrix of calibration spectra with a corresponding $M \times I$ class membership matrix $\mathbf{Y}$ and where residuals are avoided for simplicity. The matrix $\hat{\mathbf{B}}$ holds the estimated regression coefficients. The number of spectra used for calibration is M and the number of spectral channels is J. The I columns of $\mathbf{Y}$ (a dummy matrix) correspond to class memberships in classes $i = 1, \ldots, I$. Rows with a 1 in column i indicates membership in class i. The model

* Corresponding author.
E-mail address: rb@life.ku.dk (R. Bro).

0378-5173/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpharm.2009.02.014

is applied to new data $\mathbf{X_{new}}$ (e.g., all the spectra in a hyperspectral image) by using the estimated $\hat{\mathbf{B}}$ (Martens and Naes, 1989) (Eq. (2)):

$$\hat{\mathbf{Y}}_{\mathbf{new}} = \mathbf{X_{new}}\hat{\mathbf{B}} \quad (2)$$

where $\hat{\mathbf{Y}}_{\mathbf{new}}$ is the estimated class membership matrix for the new data.

Because it can be difficult in hyperspectral imaging to identify classes a priori, a modification of the PLS-DA approach that uses measured pure component spectra for $\mathbf{X}$ can be employed. To avoid confusion with PLS-DA, this approach is called PLS-Class. In PLS-Class, the calibration matrix now contains the pure component spectra $\mathbf{S}$ ($I \times J$) and $\mathbf{Y}$ is replaced with a corresponding dummy matrix $\mathbf{I}$ ($I \times I$) which is an identity matrix. The PLS-DA model then becomes

$$\mathbf{S}\hat{\mathbf{B}}_{PLS\text{-}Class} = \mathbf{I} \quad (3)$$

The regression matrix $\hat{\mathbf{B}}_{PLS\text{-}Class}$ can then be estimated using a PLS algorithm. As there are typically only I samples in the data, it is not possible to calculate a PLS-Class model with more components than the number of classes. Even when there are more samples than the number of classes, the chemical rank will usually be I because additional samples are just replicates of the pure spectra. Analogously, it is not meaningful to calculate the PLS-Class model with fewer components than the number of classes as it would imply that the I analytes do not have distinct spectral features. Hence, the PLS-Class model is really not a low-rank regression model typical of PLS. Note that the use of this PLS-Class model is only one specific application of PLS-DA. Mostly, PLS-DA is used in applications where low-rank models are expected to outperform full rank models.

Eq. (3) is in the form of an inverse least squares (ILS) model and $\hat{\mathbf{B}}_{PLS\text{-}Class}$ can be estimated using any ILS regression algorithm that handles rank deficient systems of equations. This is because the least squares estimate of $\hat{\mathbf{B}}_{PLS\text{-}Class}$ is

$$\hat{\mathbf{B}}_{PLS\text{-}Class} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{I} \quad (4)$$

However, for spectroscopic applications, it is typical that $J > I$ (and often $J \gg I$). Therefore the $J \times J$ matrix $\mathbf{S}^T\mathbf{S}$ is at most rank I and is not invertible. PLS is a common method for estimating $\hat{\mathbf{B}}_{PLS\text{-}Class}$, but principal component regression (PCR) could easily be employed as well. To see this, define the singular value decomposition of the pure component spectra as

$$\mathbf{S} = \mathbf{U}\Lambda\mathbf{V}^T \quad (5)$$

where $\mathbf{U}(I \times I)$ and $\mathbf{V}(J \times J)$ are orthogonal matrices of left- and right-singular vectors respectively, and $\mathbf{\Lambda}$ ($I \times J$) is a diagonal matrix. Because $J > I$, only the first I singular values are non-zero. Keeping only the factors that correspond to non-zero singular values (and considering that $\mathbf{U}^T\mathbf{U}$ is $\mathbf{I}$) gives

$$\mathbf{S}^T\mathbf{S} = \mathbf{V}\Lambda\mathbf{U}^T\mathbf{U}\Lambda\mathbf{V}^T = \mathbf{V}\Lambda^2\mathbf{V}^T \quad (6)$$

Therefore, the pseudo-inverse is given by Eq. (7):

$$(\mathbf{S}^T\mathbf{S})^{\dagger} = \mathbf{V}\Lambda^{-2}\mathbf{V}^T \quad (7)$$

and the least squares estimate is

$$\hat{\mathbf{B}}_{PLS\text{-}Class} = (\mathbf{S}^T\mathbf{S})^{\dagger}\mathbf{S}^T\mathbf{I} = \mathbf{V}\Lambda^{-2}\mathbf{V}^T\mathbf{S}^T\mathbf{I} = \mathbf{V}\Lambda^{-2}\mathbf{V}^T\mathbf{S}^T \quad (8)$$

Eq. (8) is a PCR model, but because the weights identified for an I-component PLS model span the same space as $\mathbf{V}$, $\hat{\mathbf{B}}_{PLS\text{-}Class}$ identified by PCR and PLS will be identical for this application.

Substituting Eq. (5) into the last expression of Eq. (8) gives

$$\hat{\mathbf{B}}_{PLS\text{-}Class} = \mathbf{V}\Lambda^{-2}\mathbf{V}^T\mathbf{V}\Lambda\mathbf{U}^T = \mathbf{V}\Lambda^{-1}\mathbf{U}^T = \mathbf{V}\Lambda^{-1}\mathbf{U}^T\mathbf{U}\Lambda^{-1}\mathbf{V}^T\mathbf{V}\Lambda^{-1}\mathbf{U}^T$$
$$= \mathbf{V}\Lambda^{-1}\mathbf{U}^T(\mathbf{U}\Lambda\mathbf{V}^T\mathbf{V}\Lambda\mathbf{U}^T)^{-1} = \mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1} \quad (9)$$

Classical least squares regression is a well-known method for calibration. It is very useful in hyperspectral analysis because of the simple chemical interpretation it allows (Gallagher, 2007; Ravn et al., 2008). The only requirement of CLS is that the pure spectra of all the analytes must be available and that any mixture spectrum can be described as a linear combination of these spectra. Once the pure spectra are known ($\mathbf{S}$) the concentrations in a hyperspectral unfolded sample, $\mathbf{x_{new}}$, can be easily calculated by direct regression as is indicated in Eq. (10):

$$\hat{\mathbf{y}}_{\boldsymbol{new}} = \mathbf{x}_{\boldsymbol{new}}\mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1} \quad (10)$$

Comparing the last two equations (Eqs. (9) and (10)) it can be noticed that the last expression of Eq. (9) [$\hat{\mathbf{B}}_{PLS\text{-}Class} = \mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1}$] is the CLS estimate of $\hat{\mathbf{B}}_{PLS\text{-}Class}$. Therefore, the CLS and PLS-Class models are identical and will provide *exactly* the same results.

3. Example

3.1. Experimental

The data set used in this demonstration is described in detail in Amigo and Ravn (2009) and Ravn et al. (2008). A five-compound conventional pharmaceutical tablet formulation was used to produce the data set analysed (active pharmaceutical ingredient (API),

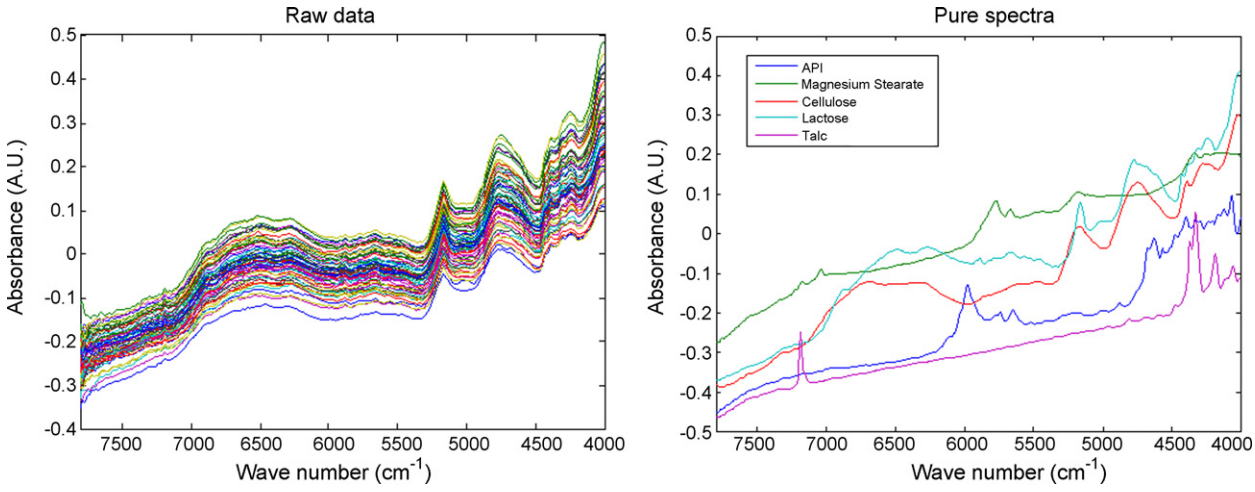

Fig. 1. Unfolded sample and pure spectra.

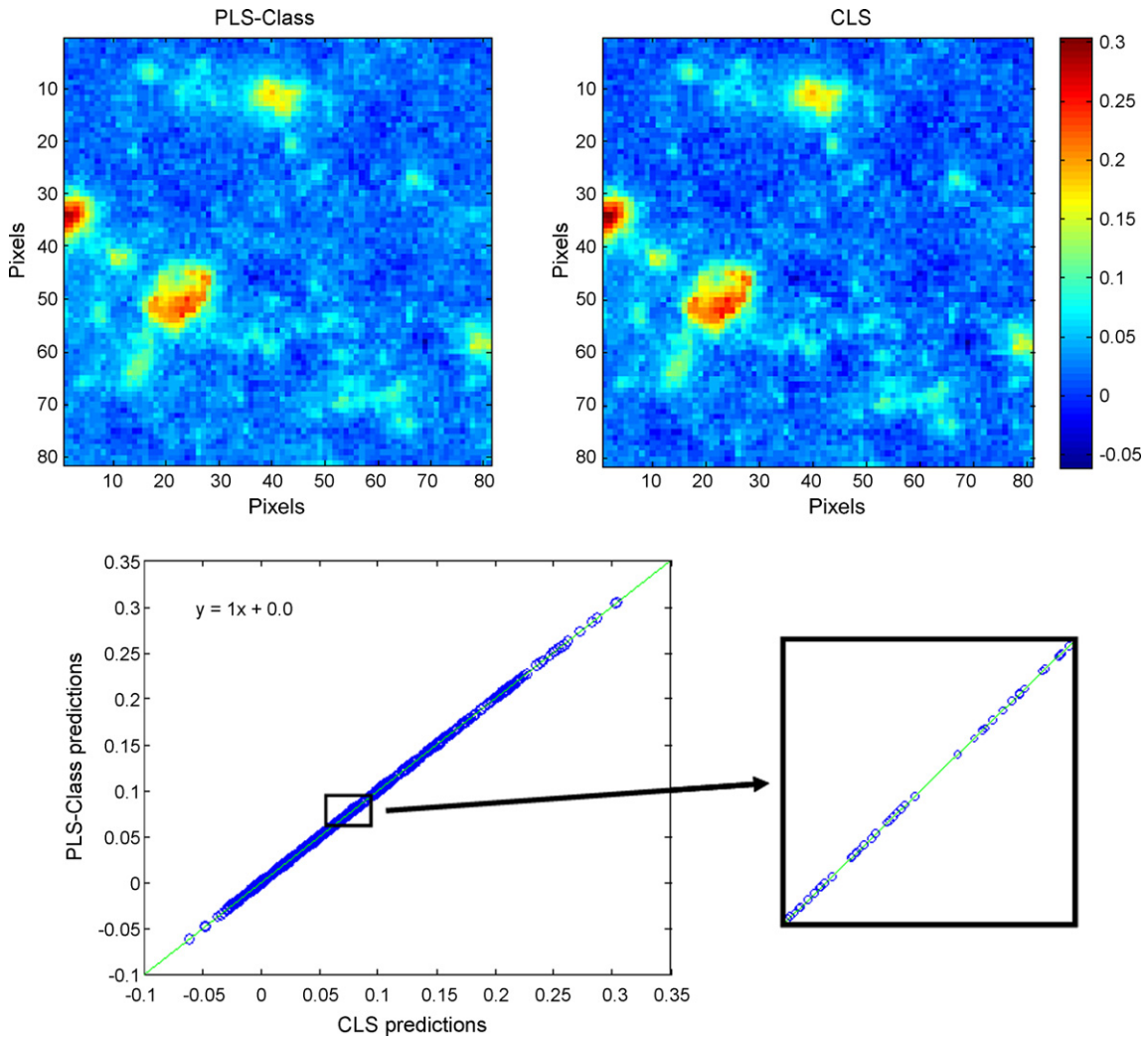

Fig. 2. Upper: concentration surface for API predicted by PLS-Class and CLS methods. Bottom: CLS predictions against PLS-Class predictions for API.

6.3%; microcrystalline cellulose, 20.0% (w/w); lactose, 71.5%; magnesium stearate, 0.75%; and talc, 1.5% (w/w)). Pure compound reference tablets of the five components were also produced. The sample selected was analysed on a NIR line mapping system (Spectrum Spotlight 350 FT-NIR Microscope, PerkinElmer, UK). An area of 2 mm × 2 mm was analysed using pixel size 25 μm × 25 μm, covering 6561 pixels in total. The spectrum of each pixel was collected from wavelength region 7800–4000 cm^{-1} using a 16 cm^{-1} spectral resolution, providing a total of 476 wavelength channels (Fig. 1).

The PLS-Class calibration model was performed by using the pure five spectra as **X**, and constructing a square diagonal dummy matrix **I** with five classes, denoting as 1 the belonging of each spectrum to each class (as in Eq. (4)). The prediction of the five components of the tablet was performed with this PLS-Class model. On the other hand, the CLS prediction was developed by direct regression of $\mathbf{X}_{\mathbf{new}}$ by using the pure spectra matrix **S**.

PLS-Class algorithm was performed by using the PLS2 algorithm implemented in PLS-Toolbox (Eigenvector Research Inc.). CLS for hyperspectral analysis algorithm from references (Amigo et al., 2008) was used. This algorithm is freely available at http://www.models.kvl.dk/users/jose_manuel_amigo/index.htm (February 2009). Both algorithms work under MatLab v. 7.5 (MatLab(R)).

3.2. Results and discussion

Only results for the prediction of the active pharmaceutical ingredient are depicted in Fig. 2. The color denotes the intensity for the concentration of API obtained for each pixel (moving from deep blue, low concentration to light red, highest concentration). As can be observed, the predictions are exactly the same for the two methods and the same relationship was obtained for the other four components. This result verifies that both methods offer the same response.

In a practical application, the PLS-Class model involves two different steps (calculation of the calibration model, $\hat{\mathbf{B}}_{PLS\text{-}Class}$, and prediction of the concentrations of the new sample). CLS, on the other hand, is typically performed by just a direct regression of the new sample onto the pure spectra.

4. Conclusion

It has been demonstrated, mathematically and with an example, that the PLS classification methodology (PLS-Class) provides *exactly* the same results as CLS. There are many advantages in avoiding the use of the PLS engine for building models as above. First of all, the PLS-Class model is not a low-rank model and therefore does not provide the usual benefits of PLS. More importantly, CLS is

a very well-defined method and it offers theoretical and practical tools, e.g., for calculating well-defined uncertainties of estimated parameters (see Appendix A and Draper and Smith, 1981; Gallagher, 2007).

While the comparison in this paper has been using hyperspectral samples, the results are generic and valid for any type of data. The use of the term PLS-Class model, though, is almost exclusively seen in hyperspectral imaging.

Acknowledgement

José Manuel Amigo wants to thank the Danish Research Council for Technology and Production Sciences for his post-doctoral fellowship.

Appendix A. Theoretical background of the calculation of uncertainty boundaries in CLS model

Having an unfolded hyperspectral sample, $\mathbf{X}(M \times J)$, and the pure spectra matrix, $\mathbf{S}$ $(J \times I)$, the CLS model to calculate the individual concentration of each component in each pixel can be expressed as follows:

$$\hat{\mathbf{y}}^T = \mathbf{x}^T\mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1} \quad \text{(A.1)}$$

The errors in the individual measurements of $\mathbf{x}$ are assumed to dominate the estimation error. The differential of Eq. (A.1) is

$$(d\hat{\mathbf{y}}^T) = (d\mathbf{x}^T)\mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1} \quad \text{(A.2)}$$

Therefore, the error covariance for the CLS model estimate of the concentrations is given by (Draper and Smith, 1981):

$$E(d\hat{\mathbf{y}}\ d\hat{\mathbf{y}}^T) = (\mathbf{S}\mathbf{S}^T)^{-1}\mathbf{S}^T\,E(d\mathbf{x}\ d\mathbf{x}^T)\mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1} \quad \text{(A.3)}$$

where $E()$ is the expectation operator. For CLS, it is assumed that the noise on each channel is of similar magnitude and uncorrelated, therefore:

$$E(d\mathbf{x}\ d\mathbf{x}^T) = \sigma^2\mathbf{I} \quad \text{(A.4)}$$

Substituting Eq. (A.4) into the variance of the estimation gives

$$E(d\hat{\mathbf{y}}\,d\hat{\mathbf{y}}^T) = \sigma^2(\mathbf{S}\mathbf{S}^T)^{-1}\mathbf{S}\mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1} \quad \text{(A.5)}$$

$$E(d\hat{\mathbf{y}}\ d\hat{\mathbf{y}}^T) = \sigma^2(\mathbf{S}\mathbf{S}^T)^{-1} \quad \text{(A.6)}$$

Eq. (A.6) accounts for the final estimation of the uncertainty in CLS models. Further information can be found in the supplied references (Draper and Smith, 1981; Gallagher, 2007).

References

Amigo, J.M., Bautista, M., Cruz, J., Coello, J., Maspoch, S., Blanco, M., 2008. Study of pharmaceutical samples by NIR chemical image and multivariate analysis. Trends Anal. Chem. 27, 696–713.
Amigo, J.M., Ravn, C., Direct quantification and distribution assessment of major and minor components in pharmaceutical tablets by NIR-chemical imaging. Eur. J. Pharm. Sci. (2009), doi:10.1016/j.ejps.2009.01.001.
Chevallier, S., Bertrand, D., Kohler, A., Courcoux, P., 2006. Application of PLS-DA in multivariate image analysis. J. Chemometr. 20, 221–229.
Clarke, F., 2004. Extracting process-related information from pharmaceutical dosage forms using near infrared microscopy. Vib. Spectrosc. 34, 25–35.
de Juan, A., Tauler, R., Dyson, R., Marcolli, C., Rault, M., Maeder, M., 2004. Spectroscopic imaging and chemometrics: a powerful combination for global and local sample analysis. Trac-Trends Anal. Chem. 23, 70–79.
Draper, N., Smith, H., 1981. Applied Regression Analysis, 2nd edition. John Wiley & Sons, Ltd., New York.
Furukawa, T., Sato, H., Shinzawa, H., Noda, I., Ochiai, S., 2007. Evaluation of homogeneity of binary blends of poly(3-hydroxybutyrate) and poly(L-lactic acid) studied by near infrared chemical imaging (NIRCI). Anal. Sci. 23, 871–876.
Gallagher, N.B., 2007. Detection, classification and quantification in hyperspectral images using classical least squares models. In: Grahn, H., Geladi, P. (Eds.), Techniques and Applications of Hyperspectral Image Analysis. John Wiley & Sons, West Sussex, England, pp. 181–201.
Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Downey, G., Frias, J.M., 2007. Hyperspectral imaging—an emerging process analytical tool for food quality and safety control. Trends Food Sci. Technol. 18, 590–598.
Henson, M.J., Zhang, L., 2006. Drug characterization in low dosage pharmaceutical tablets using Raman microscopic mapping. Appl. Spectrosc. 60, 1247–1255, http://www.models.kvl.dk/users/jose_manuel_amigo/index.htm.
Kohler, A., Bertrand, D., Martens, H., Hannesson, K., Kirschner, C., Ofstad, R., 2007. Multivariate image analysis of a set of FTIR microspectroscopy images of aged bovine muscle tissue combining image and design information. Anal. Bioanal. Chem. 389, 1143–1153.
Lyon, R.C., Lester, D.S., Lewis, E.N., Lee, E., Yu, L.X., Jefferson, E.H., Hussain, A.S., 2002. Near-infrared spectral imaging for quality assurance of pharmaceutical products: analysis of tablets to assess powder blend homogeneity. AAPS Pharm. Sci. Tech. 3, 1–17.
Ma, H., Anderson, C.A., 2008. Characterization of pharmaceutical powder blends by NIR chemical imaging. J. Pharm. Sci. 97, 369–373.
Martens, H., Naes, T., 1984. Multivariate calibration. 1. Concepts and distinctions. Trac-Trends Anal. Chem. 3, 204–210.
Martens, H., Naes, T., 1989. Multivariate Calibration. John Wiley & Sons, Chichester.
MatLab(R),2006. Version 7.0. The Matworks, Inc., Massachusetts.
Naes, T., Martens, H., 1984. Multivariate calibration. 2. Chemometric methods. Trac-Trends Anal. Chem. 3, 266–271.
PLS-Toolbox 2004. Version 3.5. Eigenverctor Research, WA, USA.
Ravn, C., Bro, R., Skibsted, E., 2008. Near-infrared chemical imaging (NIR-CI) on pharmaceutical solid dosage forms—comparing common calibration approaches. J. Pharm. Biomed. Anal. 48, 554–561.
Weiyong, L., Woldu, A., Kelly, R., 2008. Measurement of drug agglomerates in powder blending simulation samples by near infrared chemical imaging. Int. J. Pharm. 350, 369–373.
Westenberger, B.J., Ellison, C.D., Fussner, A.S., Jenney, S., Kolinski, T.G., Lipe, T.G., Lyon, R.C., Moore, T.W., Revelle, L.K., Smith, A.P., Spencer, J.A., Story, K.D., Toler, D.Y., Wokovich, A.M., Buhse, L.F., 2005. Quality assessment of internet pharmaceutical products using traditional and non-traditional analytical techniques. Int. J. Pharm. 306, 56–70.