Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Pharmaceutical and Biomedical Analysis 36 (2004) 777–786
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Near infrared spectroscopy for qualitative comparison of pharmaceutical batches

Y. Roggo*, C. Roeseler, M. Ulmschneider

Bau 65, Raum 516, Hoffmann-La Roche, Grenzacherstrasse, 4070 Basel, Switzerland

Received 11 May 2004; received in revised form 16 August 2004; accepted 16 August 2004
Available online 21 September 2004

Abstract

Pharmaceuticals are produced according to current pharmacopoeias, which require quality parameters. Tablets of identical formulation, produced by different factories should have the same properties before and after storage. In this article, we analyzed samples having two different origins before and after storage (30 °C, 75% relative moisture). The aim of the study is to propose two approaches to understand the differences between origins and the storage effect by near infrared spectroscopy. In the first part, the main wavelengths are identified in transmittance and reflectance near infrared spectra in order to identify the major differences between the samples. In this paper, this approach is called fingerprinting. In the second part, principal component analysis (PCA) is computed to confirm the fingerprinting interpretation. The two interpretations show the differences between batches: physical aspect and moisture content. The manufacturing process is responsible for the physical differences between batches. During the storage, changes are due to the increase of moisture content and the decrease of the active content.

Keywords: NIRS; Batch control; Solid form; PCA; Fingerprinting

1. Introduction

Productions of pharmaceuticals need to respect quality parameters to ensure both quality and safety. Usually these parameters are set to certain ranges depending on the manufacturing conditions. Hence, the industry needs to sort products that are out of specification.

Near infrared spectroscopy (NIRS) is an advantageous method to evaluate quality [1]. NIRS is a rapid and non-destructive technique requiring no sample preparation. As a consequence, the number of pharmaceutical NIRS applications is high: for example it is used to identify tablets in bulk and non-invasively inside individual blister pack cells, for the determination of moisture in lyophilized products through the bases of vials or for the validation of blending processes [2,3]. Moreover, Yoon et al. [4] demonstrated the possibility to determine the site of production of pharmaceutical product by NIRS.

In this article, we show how NIRS can be used to understand the differences between samples produced in two different production plants and control the stability during the storage. Generally, NIRS and chemometrics are used to classify [5] or quantify [6] samples by the help of a reference measurement. In our study, as no chemical analysis was performed on the samples, except the NIR measurement, solutions to interpret spectra need to be found. In the first part, even if the exploitation of raw NIR spectra is usually the application field of multivariate data analysis, the spectra are interpreted with the fingerprinting concept, an additional method complementary to multivariate analysis. This approach is similar to the infrared spectra interpretation. We analyzed spectra to find out specific frequencies for the detection of the active ingredient or the excipients. In the second part, principal component analysis (PCA) is applied on our

* Corresponding author. Tel.: +41 61 68 81 336; fax: +41 61 68 87 408.
E-mail addresses: yves.roggo@roche.com, y_roggo@hotmail.com (Y. Roggo).

0731-7085/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2004.08.009

Table 1
Sample description

Origin	A	B
Before storage	12 samples	23 samples
After storage	9 samples	9 samples

data sets: scores interpretation shows the difference between samples and loadings analysis explains the differences by the identification of the main wavelengths.

2. Material and methods

2.1. Samples

A set of 53 samples from two different production origins was analyzed. More details concerning the samples are given in Table 1. The general composition is strictly identical: the tablets contain lactose monohydrate, sodium starch glycolate, microcrystalline cellulose, magnesium stearate and the active compound. Fig. 1 gives the chemical formula of the active ingredient. The storage conditions are as follows: 30 °C, 75% relative moisture, during 12 months. In this article, the batches will be called: A (12 samples), A_{stored} (9 samples), B (23 samples) and B_{stored} (9 samples).

2.2. Near infrared analyses

The samples were analyzed with different near infrared methods in order to determine the influence of measurement.

Each of the two techniques was optimized separately:

- The transmittance analyses were carried out with a Bruker spectrometer (MPA-NIR-FT type) with the range 12,000–5500 cm^{-1}, i.e. 833–1818 nm. A spectra results of the mean of 32 scans with a resolution of 8 cm^{-1}.
- Concerning the reflectance measurement, the same spectrometer was used. The spectra were acquired with the range 12,000–4000 cm^{-1}, i.e. 833–2500 nm, the number of scans was 32 and the resolution was 8 cm^{-1}.

The samples were analyzed three times. Tablets were placed directly on the window in the NIR measuring device and centered over the light beam.

2.3. Chemometric methods

2.3.1. Principal component analysis

Principal component analysis (PCA) forms the basis for multivariate data analysis. The most important PCA application is to reduce the number of variables and represent a multivariate data table in a low-dimensional space [7,8]. Thus, the new variables (loadings) are linear combinations of the original absorbances.

In this study, the data was mean centered as follows: $x_{\text{centered}\,i,j} = x_{i,j} - x_{\cdot,j}$ with $x_{\text{centered}\,i,j}$ corrected absorbance (sample i and wavelength j), $x_{i,j}$ the raw data and $x_{\cdot,j}$ the mean of the absorbances at the wavelength j.

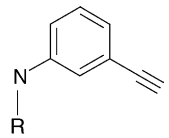

Fig. 1. Chemical structure of the active ingredient.

The new co-ordinates were computed as follows: **T** = **Xc**·**P** with **T**: score matrix, **P** the loading matrix and **Xc** the mean centered spectral matrix. We used NIPALS algorithm [9] (Non-linear Iterative Partial least Square) for the determination of loadings and scores.

2.3.2. Multiplicative scatter correction (MSC)

MSC improves the linearity of the relation between the constituents and the spectral values [10]. As a result, this method is interesting in order to eliminate the scatter effects. To compute MSC, a regression model is computed by least square method: $x_i = a + b\bar{x}i + e_i$ (with a and b being the model coefficients and e_i the model error at the wavelength i). The corrected values are calculated as follows: $x_{i,\text{corrected}} = (x_i - a)/b$.

2.4. Software

The software package for the data acquisition was Opus (Bruker). All data were exported as JCAMP files and computed with the Unscrambler software (v 7.8, Camo).

2.5. NIR tables

Frequency–structure correlation charts exist for NIRS. A lot of data were collected over years on pure or specially prepared compounds. Several NIRS wavelengths attribution tables were used [11–14]. Concerning our samples, the main wavelengths and their interpretations are summarized in Table 2.

Table 2
Interpretation of the main wavelengths

Component	Wavelengths (nm)/wavenumbers (cm^{-1})
Water	1900–1950 nm, 1410 nm, 1154 nm (large bands); 5263–5128 cm^{-1}, 7092 cm^{-1}, 8640 cm^{-1}
Starch or cellulose	925 nm, 1227 nm, 1370 nm, 2308 nm (large bands); 10,810 cm^{-1}, 8149 cm^{-1}, 7299 cm^{-1}, 4333 cm^{-1}
CH aromatic	865 nm, 1310 nm, 1637 nm, 2153 nm (large bands); 11,556 cm^{-1}, 7629 cm^{-1}, 6108 cm^{-1}, 4644 cm^{-1}
≡CH	1054 nm, 1554 nm; 9488 cm^{-1}, 6426 cm^{-1}
CH_2, CH_3	1133 nm (large bands); 7498 cm^{-1}

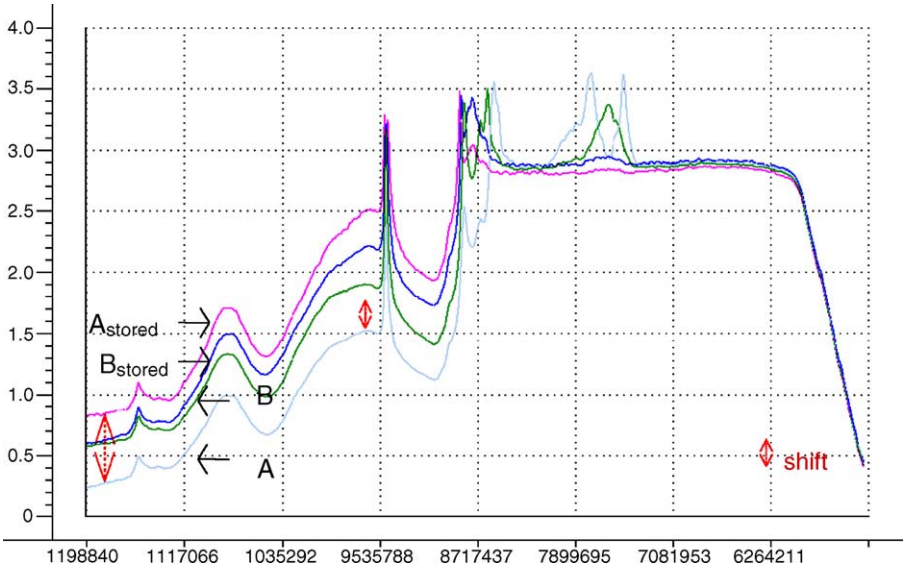

Fig. 2. Mean spectra of the different batches analyzed in transmittance.

3. Results and discussions

In the first part, our intention is to introduce the concept of fingerprinting by NIR, which correspond to the interpretation of major wavelengths. As a consequence, we examined spectra to find out specific frequencies. In the second part, PCA was applied to interpret data. Finally we compared the two different approaches, summed up and explained the main differences between the sample origins and the main effect of storage.

3.1. Comparison of the two origins by fingerprinting

3.1.1. Raw spectra interpretation

Figs. 2 and 3 show the mean of raw spectra analyzed in transmittance and reflectance, respectively. In Fig. 2, an absorbance shift can be seen on the range of 800–1600 nm (i.e. 12,500–6250 cm^{-1}) and in Fig. 3, the absorbance shift concerns the whole spectra range. Samples produced by factory A have higher absorbances. Moreover, a clear difference can be detected at 1310 nm (7633 cm^{-1}). This wavelength can be

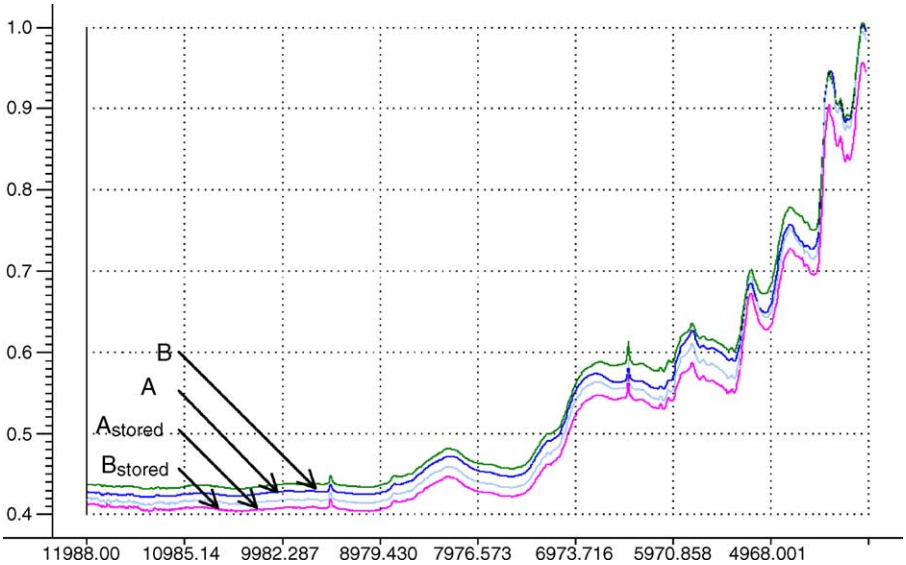

Fig. 3. Mean spectra of the different batches analyzed in reflectance.

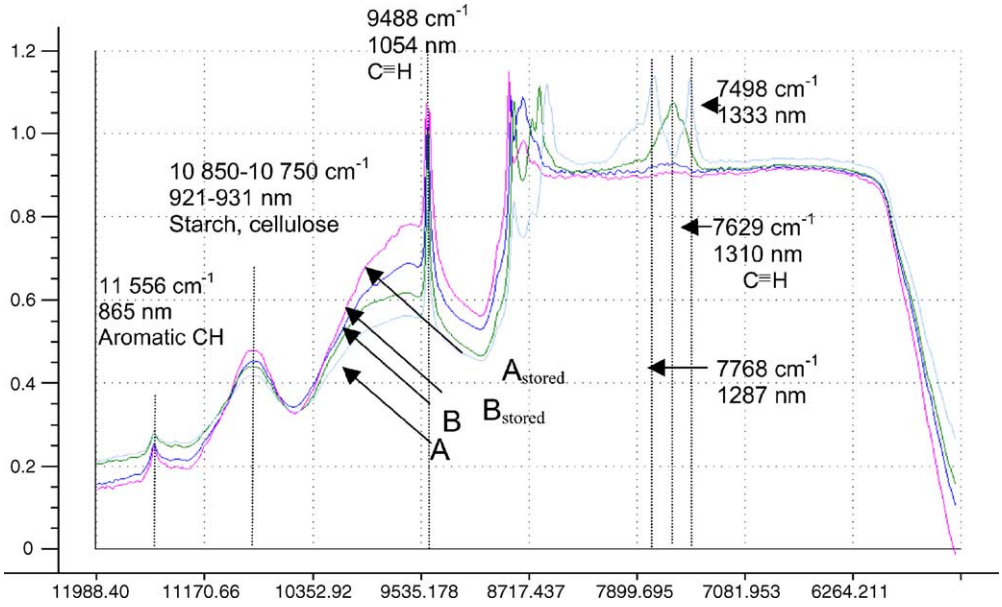

Fig. 4. MSC corrected spectra of the different batches analyzed in transmittance.

attributed to aromatic CH. The spectral difference between 1100 and 1160 nm (respectively, 9090 and 8620 cm^{-1}) is also important. Nevertheless, the interpretation of this range is not easy with NIR tables.

The raw spectra analysis underlines an absorbance shift between the samples A and B. Differences between raw spectra are due to scattering effects and physical properties of the samples. These differences can be due to particle sizes, aspects of surface or density of tablets, which modified the optical pathlength.

Despite an exorbitant amount of efforts toward the evaluation of chemometrics methods, which allow the extraction of significant information, there is a significant number of applications in which inspection of the raw spectra may be very informative [15]. As we saw, raw spectra contain information about both the physical and chemical properties of the samples.

In earlier applications, the effects of physical parameters on the NIR spectrum were often considered as a disadvantage. However, it can be useful to determine the differences between samples, which have the same composition but a different aspect.

3.1.2. Corrected spectra interpretation

In order to minimize the physical differences, spectra were normalized with MSC correction [16]. Hence, this transformation removes the absorbance shift and allows us to compare absorbances. Even if spectra are pre-treated, some differences still remain and we consider that these differences are due to sample chemistry. Figs. 4 and 5 show the means of the MSC corrected spectra of the four batches, in transmittance and reflectance mode, respectively.

First of all, the main wavelengths were attributed with the help of the NIR table (Table 2). On one hand, some wavelengths can be attributed to the active ingredient: 865 and 1310 nm (i.e. 11,556 and 7629 cm^{-1}) corresponding to the group ≡CH, 1054 and 1556 nm (i.e. 9488 and 6426 cm^{-1}) corresponding to –CH aromatic. On the other hand, some wavelengths can be assigned to the excipient and residual water: 1950 and 1410 nm (5128 and 7092 cm^{-1}) are wavelengths specific of water, 925 and 2308 nm (10,810 and 4333 cm^{-1}) are assigned to starch and cellulose.

After the wavelength identification, we tried to determine the main spectral differences. In transmittance, the main differences are the absorbances at 1310 and 925 nm, which are characteristic of CH aromatic and starch or cellulose. In reflectance, an absorbance difference at 1950 nm shows a difference of water content between the two production origins.

For registration purpose NIRS was not considered as a primary method. NIRS still suffers from the comparison with the traditional mid-infrared spectroscopy. In any case, the theoretical background is the same: vibrational spectroscopy. One recurrent difficulty for near infrared spectroscopists is to convince of the measurement specificity in absence of particular band assignments to molecular vibrations like in the mid-infrared. In the NIR frequency region, there is a large degree of inter-correlation combinations and overtones, which are correlated to fundamental vibrations originating in the mid-infrared region. However, it seems that the pharmaceutical product spectra can be interpreted with the help of the fingerprinting.

3.2. Chemometrics to underline the difference between batches from the two origins

First, we demonstrated the possibility to use fingerprinting: several wavelengths can characterize the ac-

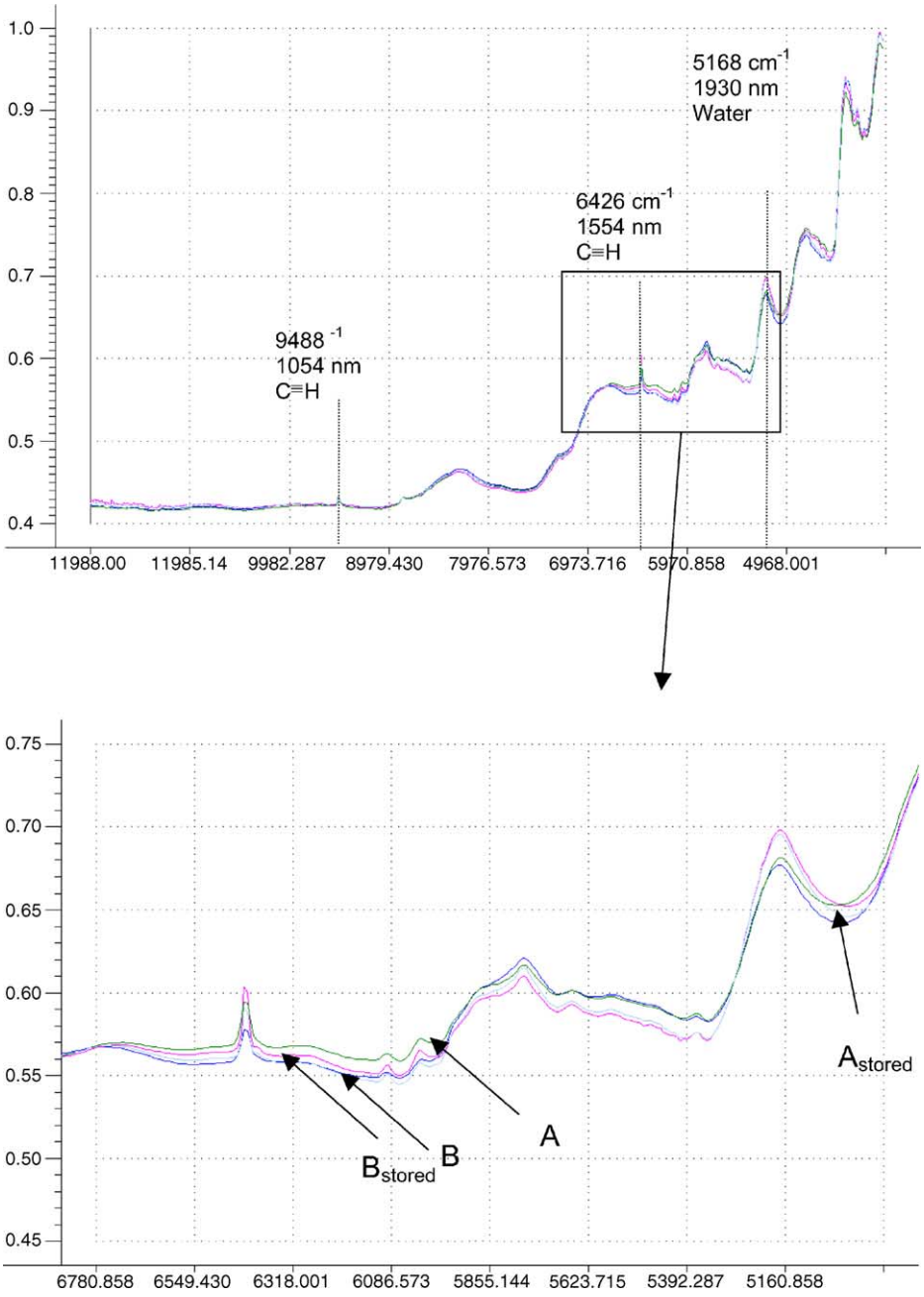

Fig. 5. MSC corrected spectra of the different batches analyzed in reflectance.

tive ingredient. We decided to use principal component analysis to confirm our spectral interpretations, extract more information and explain the origin of the chemical differences between batches. In this part, only the results concerning the samples without storage will be presented.

3.2.1. Principal component (PC) analysis of the transmittance data

Two PCA analyses were performed: one on the raw spectra and one on the MSC corrected spectra. Fig. 6 shows the PCA results with raw data. The plane defined by the first two principal components (PC) contains 88% of the variance of the data set. Fig. 6A shows the difference between tablets produced by the two different ways. The first principal component separates the samples by origin.

Loadings and spectra have the same dimension and loadings can be explained like the spectra. Fig. 6A and B show the first two loadings. Concerning PC 1, the loadings are high between 800 and 1000 nm. It is due to scattering effects and physical differences between samples of the two origins. The wavelengths, which have the highest influence on the first PC are: 1054, 1154 and 1310 nm, which correspond

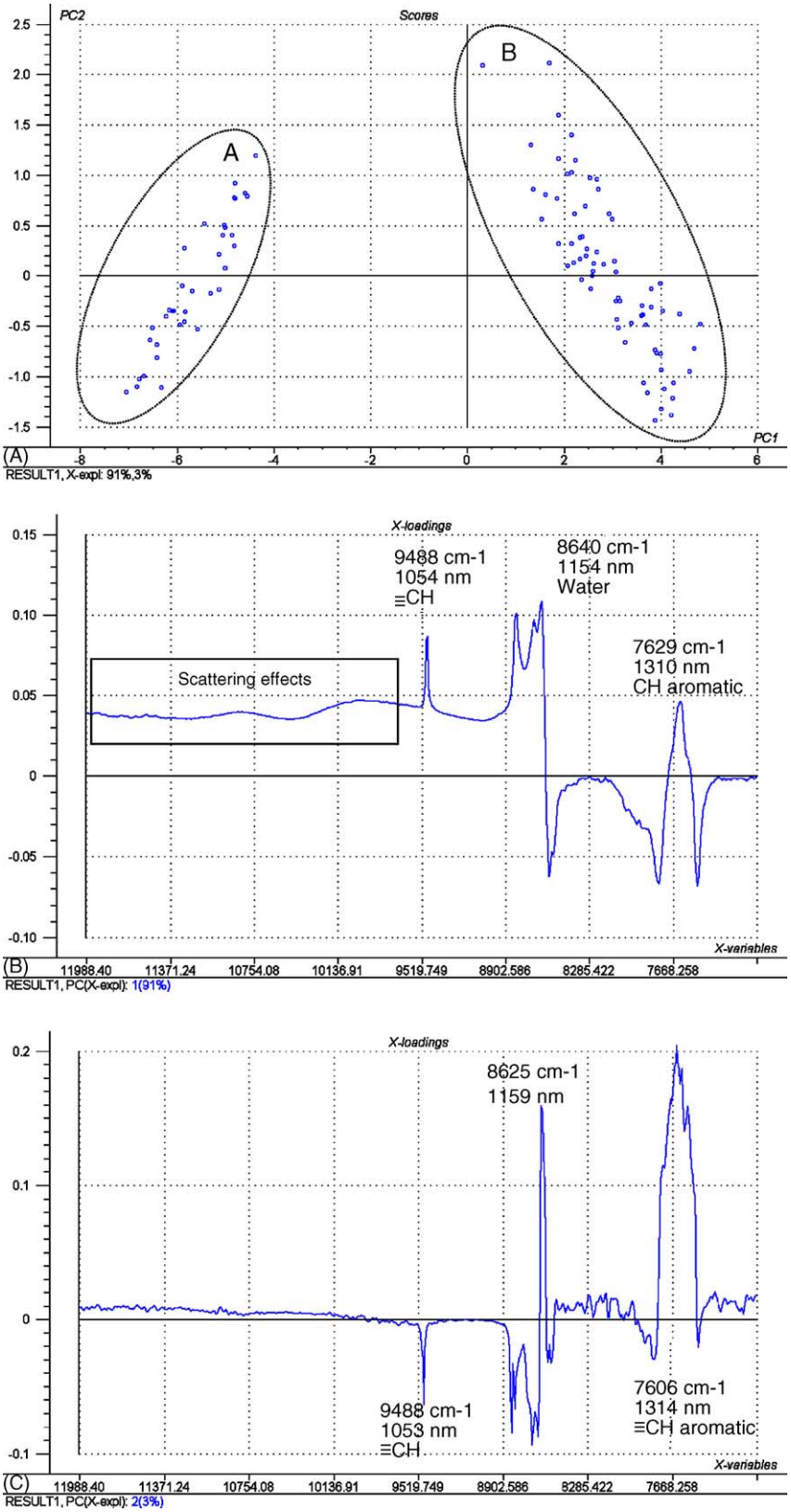

Fig. 6. PCA on raw transmittance spectra—samples without storage.

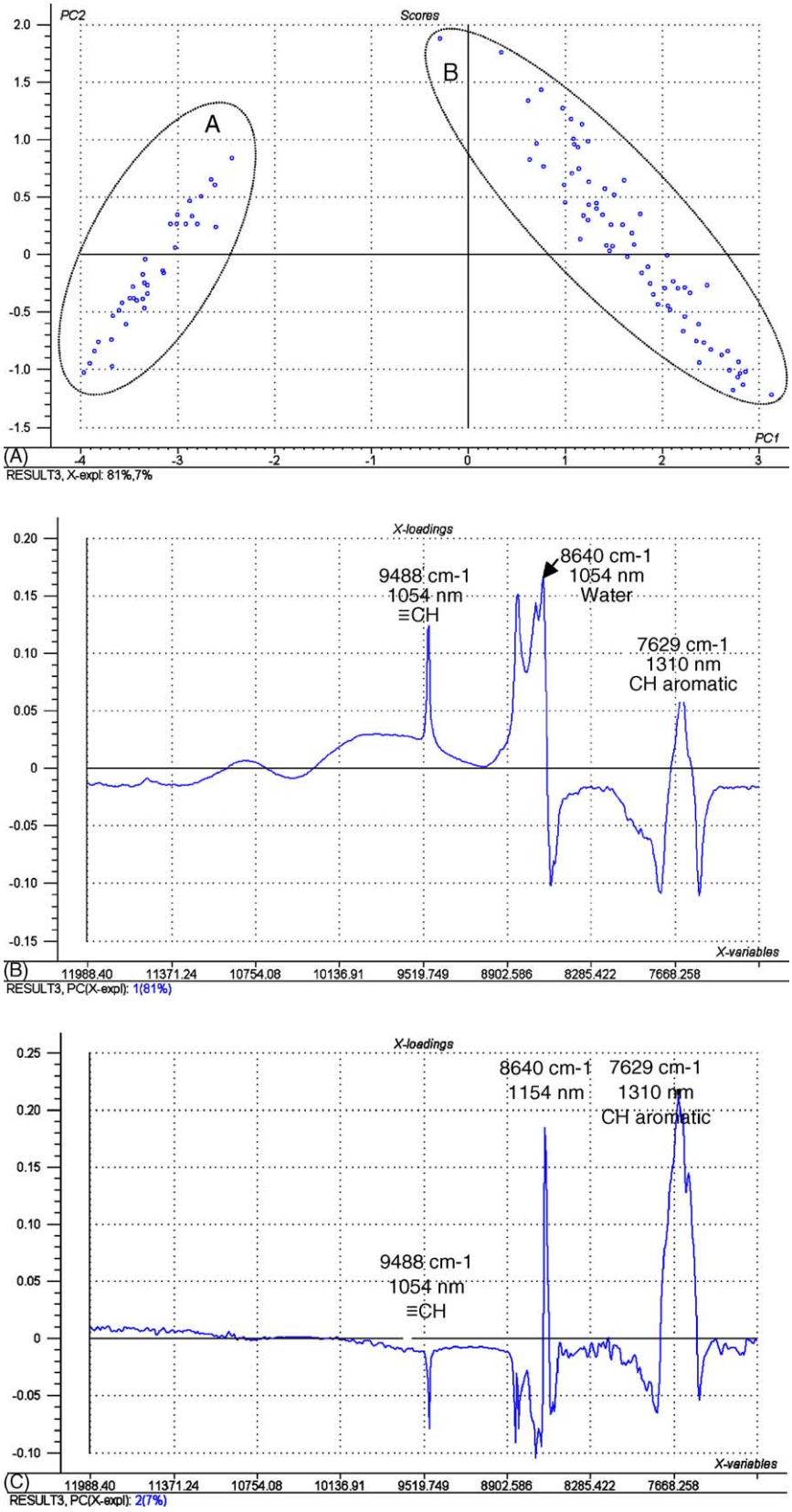

Fig. 7. PCA on MSC corrected transmittance spectra—samples without storage.

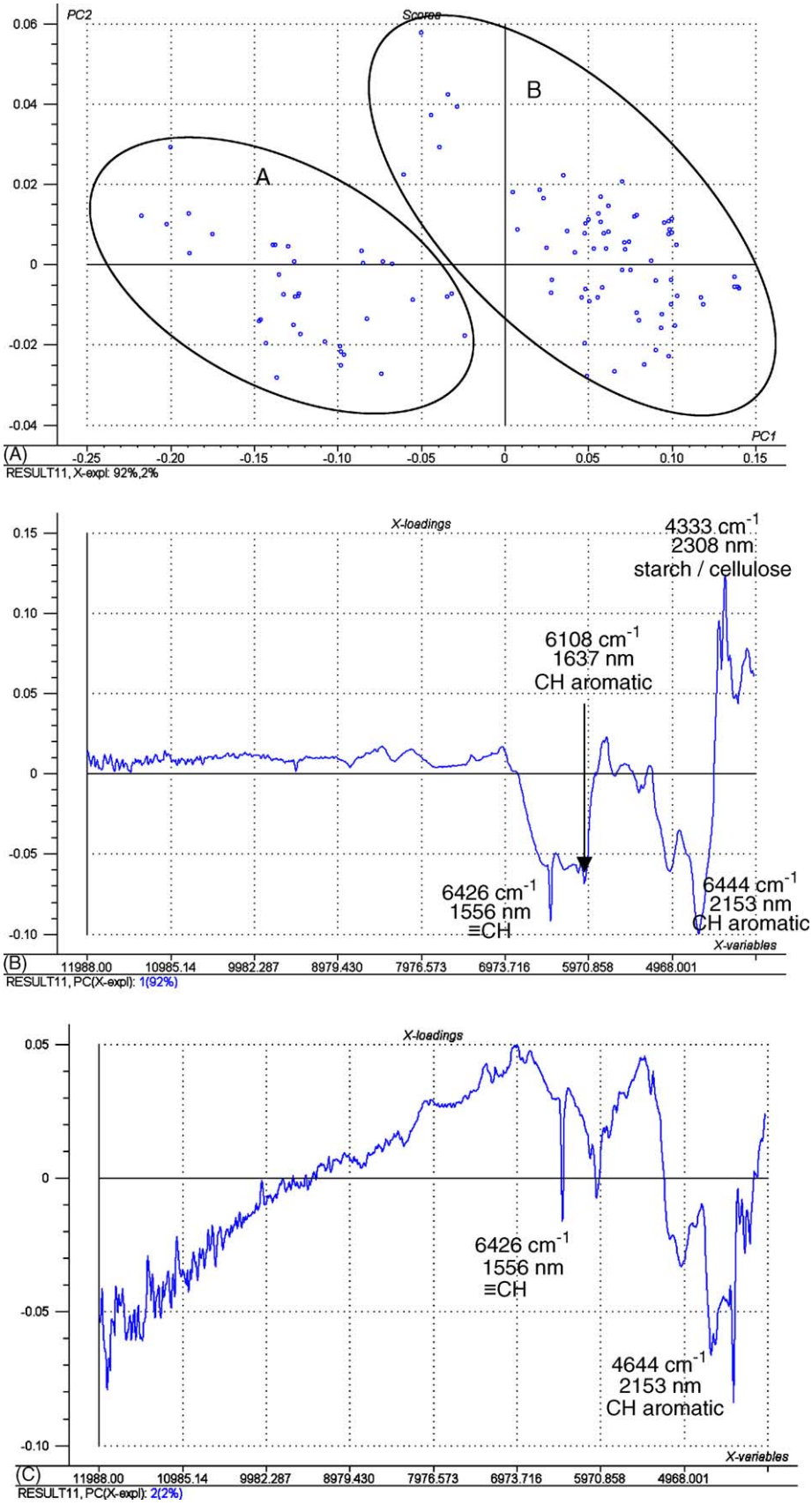

Fig. 8. PCA on MSC corrected reflectance spectra—samples without storage.

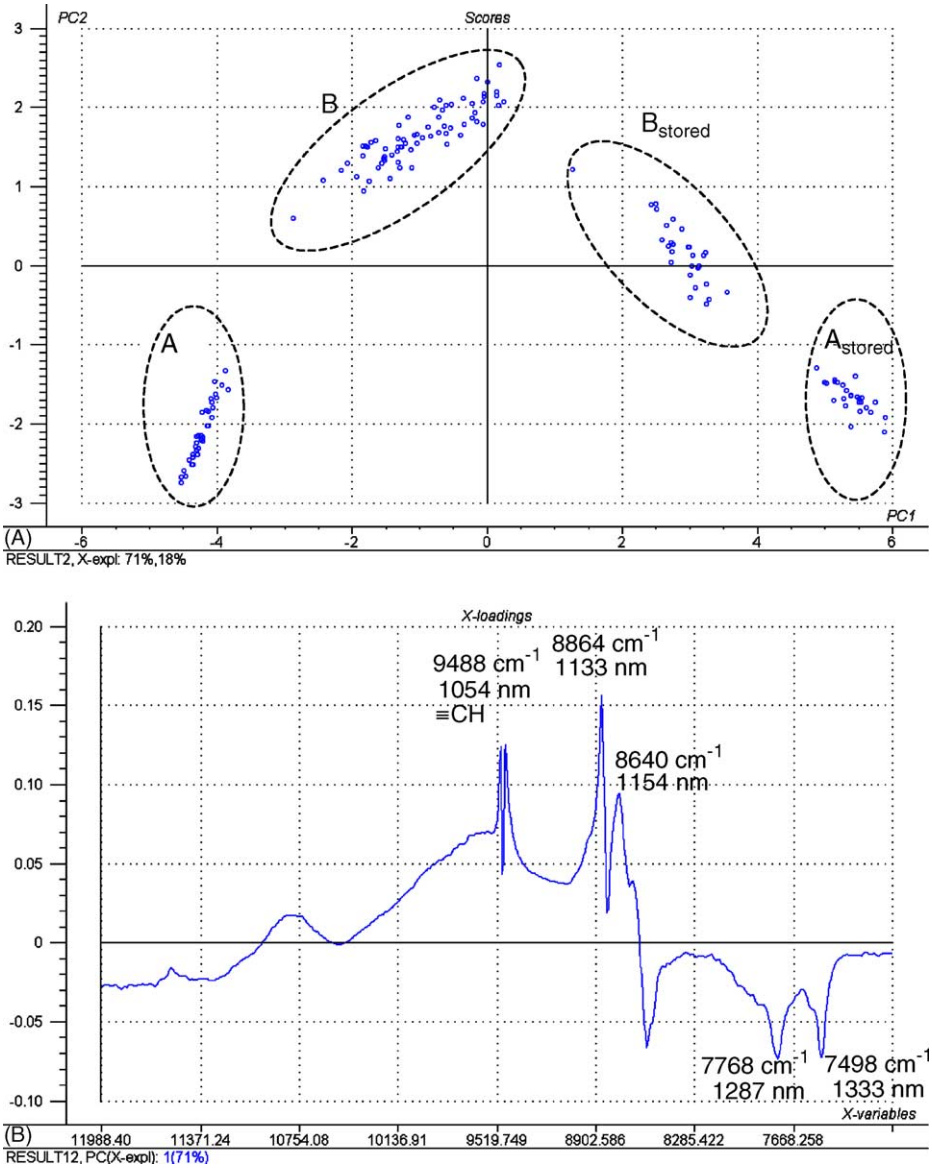

Fig. 9. Comparison of samples with or without storage by PCA on MSC corrected transmittance spectra.

to ≡CH, water and CH aromatic, respectively. The second PC has high loadings at 1153 (water), 1147 (CH_2,CH_3) and 1054 nm (≡CH).

The PCA confirms that the differences between samples A and B are due to changes in the active ingredient (≡CH and CH aromatic absorbances are modified). These modifications may be due to the increase of the water content in A tablets. We assume that water is introduced by the excipients: starch and cellulose.

Fig. 7 shows the PCA performed on the MSC corrected data. MSC was applied to corrected spectra and removed the influence of scattering effects. The scores (Fig. 7A) are similar to those of the raw spectra (Fig. 6A). The differences between the two origins are underlined. The loadings of PC 2 are identical to those of Fig. 7C. Nevertheless, on the first PC, we identify the changes and the effects of MSC (Fig. 7B). A wavelength (1069 nm assign to starch or cellulose) was hidden by scattering effects and is extracted by MSC treatment.

3.2.2. Principal component (PC) analyses of the reflectance MSC corrected data

Fig. 8 shows the results of PCA on reflectance MSC corrected data. As performed before, the scores and the loadings were analyzed. The first two principal components separate the four batches clearly. PC1 discriminates the A samples from the B ones. This analysis confirms the differences between the samples.

The first PC (Fig. 8B) can be interpreted by the following wavelengths: 2158 and 1637 (CH aromatic), 1554 (≡CH), 1910 and 1410 nm (water). This component underlines a difference in the chemical composition of the two origins.

Table 3
Comparison PCA vs. fingerprinting

Methods	Drawbacks	Advantages
Fingerprinting	Analysis of a low number of spectra Difficulties to interpret small spectral variation	Speed Easy to understand and explain Comparable to IR analysis
PCA	Statistical method are considered sometime as a black box by infrared spectroscopist Method required a chemometric knowledge	Interpretation of large database Graphical representation the samples Loadings can be analyzed like spectra

3.3. Comparison of the two spectra measurement

Reflectance measurements penetrate only 1–4 mm of the front surface of solid samples. This low penetration of energy into a sample features greater variation when measuring non-homogenous samples.

In transmittance, the entire pathlength of the sample is integrated into the spectral measurement and the errors due to sample homogeneity are reduced. For fine particles, the front surface scatter brings a loss of energy transmitted and a decrease in the signal to noise ratio. However, the main drawback is only absorbances at the lower wavelength 800–1400 nm, i.e. the upper energy level, can be interpreted. If the particle size is sufficiently small, the instrument will not transmit enough energy to the detectors. In fact, the two ways of measurement are complementary in our research. However, some studies [17,18] prove that transmission measurements give more accurate quantitative models.

3.4. Effect of the storage

Fig. 9 shows the principal component analysis of MSC corrected spectra analyzed in transmittance. The effect of storage can be underlined. The main differences are due to the water content of the sample linked to the extreme storage conditions and active compound, whose concentration decreased during the storage.

4. Conclusion

Statistical methods and spectra interpretation (fingerprinting concept) show the differences between tablets produced by two different factories A and B. The differences between B and A are summarized as follows. On one hand, the physical differences and the distinction in water content are due to the excipients, especially the starch and cellulose. On the other hand, the active ingredient (CH aromatic and ≡CH peaks) is not exactly the same depending on the production site.

We prove that spectroscopic techniques could be helpful to discriminate between apparently identical materials from different origins. Near infrared spectroscopy helps us to identify and understand a manufacturing problem and follow the evolution of samples during storage.

In this article, two methods for spectra exploitation were compared. PCA is a powerful method. However, fingerprinting can be introduced for the interpretation of transmittance data. The two interpretation methods lead to the same conclusions. The advantages and drawbacks of the two methods are summarized in Table 3. Even if we prefer the PCA approach, the fingerprinting fits to interpret spectra and can be more easily explained to authorities.

References

[1] M. Blanco, I. Villaroya, TrAC Trends Anal. Chem. 21 (2002) 240–250.
[2] B. McDonald, K. Prebble, J. Pharm. Biomed. Anal. 11 (1993) 1077–1085.
[3] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, C. De la Pezuela, Analyst (Camb., UK) 123 (1998) 135R–150R.
[4] W. Yoon, R. Jee, A. Charville, G. Lee, A. Mofat, J. Pharm. Biomed. Anal. 34 (2004) 933–944.
[5] Y. Roggo, L. Duponchel, J.-P. Huvenne, Anal. Chim. Acta 477 (2003) 187–200.
[6] Y. Roggo, L. Duponchel, B. Noé, J.-P. Huvenne, J. Near Infrared Spectrosc. 10 (2002) 137–150.
[7] L. Lebart, A. Morineau, M. Piro, Statistiques exploratoires multidimensionnelles, Dunod, Bordas, Paris, 1997, pp. 32–34.
[8] M. Danzart, Statistique descriptive, in: SSHA&ISHA, Analyse sensorielle. Manuel méthodologique, Tec&Doc, Paris, 1990, pp. 209–235.
[9] P. Geladi, B. Kowalski, Anal. Chim. Acta 185 (1986) 1–17.
[10] T. Naes, T. Isaksoon, Anal. Chem. 62 (1990) 664–669.
[11] N. Colthup, L. Daly, S. Wiberley, Introduction to Infrared and Raman Spectroscopy, 3rd ed., Academic press, London, 1990.
[12] A.G. Büchi, ANATEC, Table with characteristic absorbance in NIR, 1997.
[13] D. Burns, E. Ciurczak, Handbook of Near Infrared Spectroscopy, Dekker, New York, 1992.
[14] Instrumentation Research Laboratory, Identification of wavelengths of near infrared absorbers, BARC, USDA, Beltville, MD, USA.
[15] S. Sekulic, J. Wakeman, P. Doherty, P. Hailey, J. Pharm. Biomed. Anal. 17 (1998) 1285–1309.
[16] C. Pizarro, I. Esteban-Diez, A.-J. Nistal, J.-M. Gonzalez-Saiz, Anal. Chim. Acta 509 (2004) 217–227.
[17] R. Schneider, K.-A. Kovar, Forensic Sci. Int. 134 (2003) 187–295.
[18] S. Thosar, R. Forbess, N. Ebube, Y. Chen, R. Reibinovitz, M. Kemper, G. Reier, T. Wheatley, A. Shukla, Pharm. Dev. Technol. 6 (2001) 19–29.