Multivariate diagnosis of the variability of NIR spectrometers under industrial applications

P. Barreiro*[1], L. Henche[1], M. Ruiz-Altisent[1], N. Hernández[1] and A. Moya-González[2]

[1] Physical Properties Lab. Rural Engineering Dept. ETSIA-UPM. Avda. Complutense, s/n. 28040 Madrid. Spain
[2] Agrotécnica Extremeña, S. L. Ctra. Badajoz-Montijo, km 24,9. 06184 Pueblo Nuevo de Guadiana (Badajoz). Spain

Abstract

The transfer of NIR spectroscopy to industry relies on the possibility of real time identification of abnormal spectra as well as uncontrolled sources of variation. This study proposes an unsupervised procedure for the identification under an industrial application of daily events (general changes) and abnormal observations. It consists in defining a spectral database at the beginning of a season, performing a principal component (PC) analysis, and calculating the PC scores over time. Process control statistics (Hotelling T^2, Q) are used for multivariate supervision of the industrial application. Within this procedure 10,400 average spectra of onion bulbs were evaluated identifying events in 12 out of 66 work dates, as well as spectral trends throughout the season of 2002.

Key words: internal quality, PCA, unsupervised analysis, process control.

Resumen

Diagnóstico multivariante de la variabilidad de espectrofotómetros NIR en una aplicación industrial

La viabilidad de la transferencia de la espectroscopía infrarroja en el ámbito industrial depende de la posibilidad de identificar en tiempo real espectros anómalos, así como la existencia de fuentes de variación no controladas. Este estudio propone un procedimiento no supervisado para la identificación en una aplicación industrial de eventos diarios (sesgos generales en los espectros), así como de observaciones anómalas. El procedimiento consiste en la generación de una base de datos espectral al comienzo de la campaña, la aplicación de un análisis de componentes principales, y el cálculo posterior de las proyecciones sobre los componentes principales de los nuevos espectros a lo largo del tiempo. Se emplearon estadísticos habituales de control de procesos (Hotelling T^2 y Q) para hacer un seguimiento multivariante de la aplicación industrial. Con este procedimiento, a lo largo de la campaña 2002, se evaluaron 10.400 espectros medios de bulbos de cebolla, identificándose eventos en 12 de los 66 días de trabajo, así como cambios en la respuesta espectral.

Palabras clave: calidad interna, ACP, análisis no supervisado, control de procesos.

Introduction

The viability of Near Infra Red (NIR) Spectrometry for internal quality assessment in fruit and vegetables is accepted world wide even for real-time applications. However, the transfer of technology to the agro-industry is still a challenge due to a high number of uncontrolled sources of variation which modify the spectral information, and reduce the accuracy of estimations. Some of these sources of variation are: the internal temperature of the product and the spectrometer (Hernández-Sánchez *et al.*, 2003), the skin thickness (Krivoshiev *et al.*, 2000), and the presence of boundary layers and voids inside the product (Fraser *et al.*, 2003).

A main issue when developing a new NIR application is selection of the interaction mode between the light and the sample: reflectance, interactance or transmittance. The interactance mode, though it is the most difficult situation for online implementation, has shown encouraging results for obtaining good predictive models (Schaare and Fraser, 2000).

In the calibration of models, a difference should be stated between under and over fitted situations. The first ones do not take into account relevant information and thus lead to biased estimations, while the latter use non relevant (noisy) information leading to unrobust models when used under external validation (with new observations not used to calibrate the model) (Wortel *et al.,* 2001). In many applications there is a limitation in the transfer of predictive models through a season and also between years due to a poor validation performance, though these models may be extremely useful for classifying a reduced number of categories (Guthrie *et al.,* 1998).

Previous studies concerning the marketing process of high technology devices indicate that between 40 and 80% of new High-Tech products fail in the agro food industry. This rate is higher than that reported for other industrial sectors, and the poor product performance of these equipments is a main cause of failure (Karakaya and Kobu, 1994).

The 56th Technical Committee from the International Electrochemical Commission gives some crucial definitions for instrumented systems such as fault and failure (Mellor, 2001). Fault refers to something that is wrong with a system which normally leads to an event, that is, an occurrence or happening usually significant to the performance of a function. Failure is an event in time at which an item ceases to perform a required function. Arising from the Safety Instrument Systems, a new concept has been defined entitled Common Cause Fault (CCF). This term refers to a fault which causes the failure of multiple devices or processes. One of the main CCF is the loss of calibration of sensors, which in a wide sense may be due to drifts of the signal over long periods of time, mechanical and/or human faults (Summers and Raney, 1999). When the consequences of faults are delayed in time, these may be extremely difficult to identify before a failure occurs.

In this study we analyse the events occurred within the industrial use of a NIR spectrometer for the classification of onions according to their soluble solid content. This classification is performed within a genetic selection program. Selected onions have a high soluble solid content in order to optimise the dehydration procedure.

The objective of this study was to perform an unsupervised analysis of data to detect abnormalities in the NIR spectra due to: faults of the NIR spectrometer, specific spectral responses of specimens or changing environmental conditions, in order to avoid incorrect classification of samples.

Material and Methods

A NIR spectrometer (Hamamatsu PMA-11) that incorporates an indium, gallium and arsenic (InGaAs) detector array (256 elements) has been used, allowing spectral analysis in the 900 to 1,600 nm range, with a resolution of 2.7 nm. The minimum integration time for a single spectrum is 5 ms. In our case, the integration time was established in 70 ms for the applications with entire product when interactance (internal reflection) mode and spectra are acquired through an optical fiber in direct contact with the sample. This integration time guarantees a maximum signal around 90% of the dynamic range of the A/D card for the white reference, 80% for the samples (onion bulbs).

The spectrometer uses a mechanical shutter which closes the light path to the detector array when the integration time is attained. This device has been shown to be sensitive to a dusty environment such as most agro industrial factories, as a consequence of dust accumulation it can be stacked, leading to the failure of the equipment.

The plant material used for this study corresponds to 10,400 onion bulbs tested in the 2002 season (August to December) in the AGROTECNICA EXTREMEÑA S.L. facilities (Badajoz). This material was classified according to the average spectral data (three replicates per bulb in a selected area) into three categories of internal quality (low medium and high soluble solid content), by means of an estimation model developed in the previous season (2001). The estimation model ($R^2 = 0.75$) was adjusted with four sets of data (1,316 bulbs in total), following an iterative process which combines: stepwise multilinear regression for the different sets of data, comparison of selected variables, identification of frequently selected wavelengths and final readjustment of the model for the selected wavelengths. The model uses five wavelengths with scattering correction (subtraction of reflectance at 900 nm). This data pretreatment is important since it eliminates a main source of variation which is not relevant for our quality parameter. Therefore six wavelengths are used in total in the model allowing classification into the three mentioned quality categories with misclassification errors below 15% of individuals of the calibration set (Barreiro *et al.,* 2002).

A reduced set of 400 bulbs of the 10,400 bulbs was used in order to assess the predictive capacity of the model. For this validation set, measurements on internal quality were taken with a reference procedure: soluble

solids evaluation (measured in Brix degrees) by means of a refractometer (Barreiro *et al.,* 1999). Also the surface (non destructive) and internal (5 mm depth, destructive) temperatures of the bulbs were recorded to trace the effect of this source of variation. Measurements were performed with a PT100 probe.

One operator was trained in the use of the NIR spectrometer by members of the Physical Properties Laboratory (LPF). The LPF was in charge of supervising the functioning and maintenance of the equipment and detected the following events (fault and failure):

a) Malfunction of the mechanical shutter due to the aggression of the industrial environment This failure occurred the 4th of October 2002. The equipment was then transported to the LPF, repaired and installed back in the company on 12th October 2002.

b) Misalignment of the optical fibers. This fault occurs when the light slit of the optical fiber is not perfectly aligned with the monochromator; this situation was detected the day after reinstalling the equipment.

c) Excessive temperature fluctuations of the bulbs. In order to reduce this, the bulbs were stored near the NIR equipment 24 h before they were tested with the NIR spectrometer to avoid temperature differences between bulbs. Facilities lacked temperature control so the effect of temperature evolution during the season was not avoided.

d) Specific characteristics in size and structure of the bulbs. Despite all the bulbs belonging to the same cultivar which is under a genetic selection procedure, bulbs with abnormal shapes, ribbed bulbs or loose layers can produce relevant differences in the scattering of light inside the bulb tissue.

The analytical procedure followed by the LPF for the data recorded with this industrial application included the following steps (Fig. 1):

1. Validation of the estimation model with a set of 400 bulbs. This step requires the measurements of soluble solids (ºBrix) with a refractometer as well as other reference measurements such as temperature of the bulbs (ºC), or the maximum signal (intensity counts) of the NIR reference (barium sulphate disk).

2. Definition of a spectral database (342 average spectra corresponding to bulbs tested at the beginning of the season, August) and performance of a Principal Components Analysis (PCA) with spectral variables centred and scaled to unit variance to avoid the effect of magnitude. Definition of the spectral database consisted of choosing several test dates where the LPF
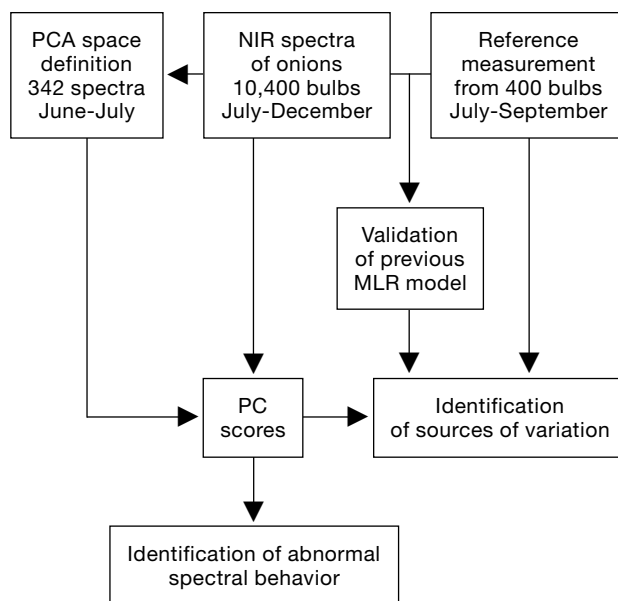

Figure 1. Scheme of the data analysis procedure followed in this study. PCA stands for Principal Component Analysis and MLR for multiple linear regression.

expert was present and no fault or failure had been recorded. From hereon this dataset is referred to as reference spectral database or reference database.

3. Projection of the validation set (400 average spectra) onto the PC space obtained from the reference spectral database, and study of correlations between the principal components (PC) scores and measured sources of variation (soluble solids, bulb temperatures, maximum signal from the spectral reference). Null correlation between PC scores is expected unless shifts in the spectra occur simultaneously affecting their spectral variables.

4. Projection of the remaining 9,658 average spectra onto the PC space obtained from the reference database and identification of abnormal spectra and changes in behaviour of the NIR spectrometer. Abnormal individuals are addressed by means of normal probability plots.

5. Calculation of process control statistics (Hotelling T^2 and Q) and plotting of process control charts with regard to the Upper Control Limits (UCLs).

6. Averaging the PC scores per day and calculation of the cumulative sum in order to summarize the change in equipment performance.

The Hotelling T^2 is a multivariate statistic which may be used as an event indicator. It is computed daily as stated in Eq. [1], where [X] is the vector containing the average PC scores per day, [m] is the vector containing the average expected values, and [S] is the covariance

matrix of the PC scores in a day which is expected to be the identity matrix.

$$T^2 = [X-m]^t [S] [X-m] \qquad [1]$$

Eq. [2]. shows the computation procedure for the *UCL* of the Hotelling T^2 statistic, where n is the number of observations tested per day, p the number of PC scores considered and F the critical value for a Fisher distribution with α confidence and p, n-p degrees of freedom.

$$UCL_\alpha(T^2) = p\,\frac{n-1}{n-p}\,F_\alpha(p, n-p) \qquad [2]$$

The Q statistic is developed to address atypical observations and is an indicator of how well each spectrum fits the PC model. It is computed as the normalized squared error between the original and the predicted spectra when using the PC scores. Under normal conditions the Q statistic has a multi-normal distribution and associated UCL may be estimated using a weighted chi-squared distribution (Simoglou *et al.*, 2000).

Table 1 refers to the number of specimens and spectra obtained in the 2002 season. In order to perform all mentioned steps, devoted Matlab programs (Mathworks Inc.) were developed.

Results

A PCA on the reference database (342 average raw spectra) was performed using the spectral variables (in our case 256 wavelengths) centred and scaled to unit variance. A variance of 99.82% from the original 256

Table 1. Summary of samples used for NIR and reference analysis: soluble solids (°Brix), internal and surface bulb temperature (°C)

Month	Number of test dates	Number of bulbs with NIR analysis	Number of bulbs with reference analysis
July	4	200	200
August	14	1,400	100
September	17	2,200	100
October	12	2,305	
November	16	2,940	
December	3	1,355	
Total	66	10,400	400

Failure of the shutter occurred on 4th October 2002. Equipment was reinstalled on 14th October 2002.

spectral variables was explained by means of five Principal Components (PC): 91.7% with PC1, 7.28% with PC2, 0.76% with PC3, 0.13% with PC4 and 0.08% with PC5. The first PC included all wavelengths since all 256 showed correlation coefficients above 0.8 with this factor. Correlation of the different wavelengths with the remaining PCs was always below 0.3.

Since the PCs are linearly correlated with the spectra variables, a normal behaviour of the latter leads to normal distributions of the PCs. Figure 2 shows the normal probability plots for the 10,400 average spectra considering PC1 and PC2 scores. Abnormal individuals are those which clearly separate from the expected normal behaviour. This segregation is clear for PC scores above and below 2 units; note that the mean value for both PC scores is 0.
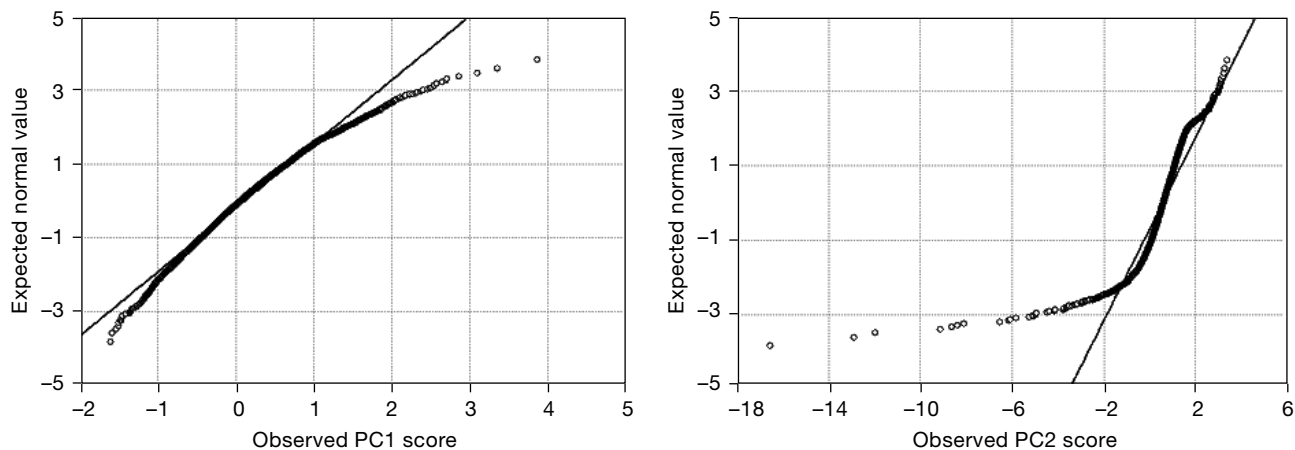

Figure 2. Normal probability plots for PC1 and PC2 scores (10,400 data). Abnormal values are those which clearly segregate from expected normal ones (line).

Figure 3 shows as crosses (+) and circles (o) the PC scores of the reference spectral database (342 bulbs) and specific test data (396) respectively. The top-left plot represents the PC1-PC2 plane while the top-right plot presents the PC1-PC3 plane. Most of the PC scores stay within the [–2,+2] interval. Isolated individuals can be found out of this range specially using the PC2 score. The test date corresponded to the 14th October 2002. This date showed a significant spectral shift due to incorrect reinstallation of the equipment after repair of the mechanical shutter, which was identified and corrected.

Since the representation of the PC scores produced interesting patterns, it was decided to perform on the validation set (400 bulbs) a correlation analysis using the PC scores together with reference variables: white level of the NIR reference (intensity counts), surface and internal temperature of the bulbs (ºC), observed and estimated soluble solids content (ºBrix). A very high correlation was found between the surface and internal temperature of the bulbs and PC5 (0.82 and 0.8 respectively), which was lower for PC1 (0.7). The white level (intensity counts) corresponding to the barium sulphate disk used for daily calibration of the spectrometer was not related to any of the sources of variation, revealing that this source of variation was manually controlled by the operators of the NIR equipment. One important feature emerging from this table was a high correlation between some of the PC scores (i.e. –0.82 between PC2 and PC3) for the validation set. This suggests the occurrence of new sources of variation that simultaneously affect the PC scores, and
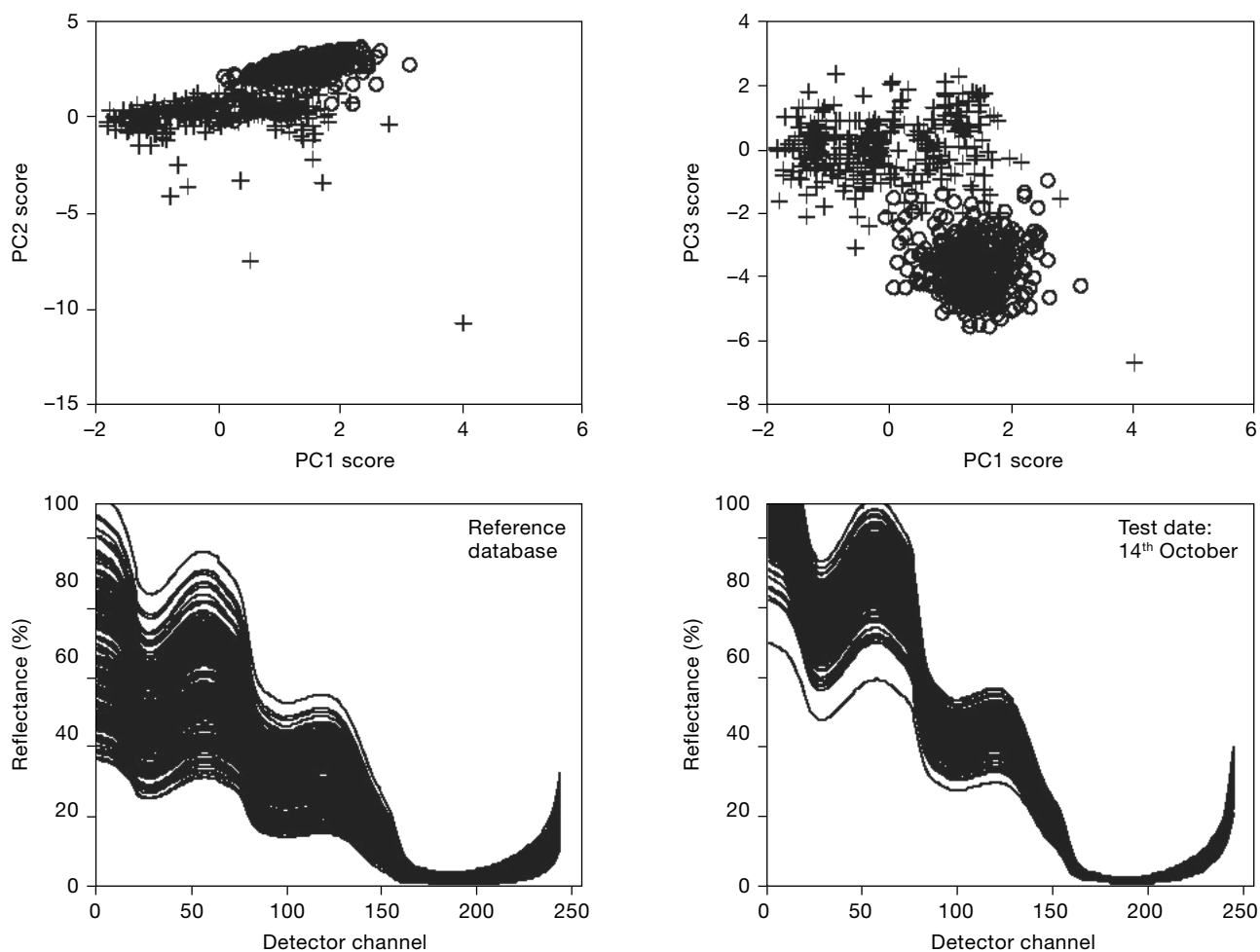

Figure 3. Example of result obtained under industrial use: crosses (+) represent PC score for the reference spectral database (342 bulbs) while circles (○) indicate the PC scores obtained for individuals tested on 14th October 2001 (396 bulbs). Bottom-left and bottom-right plot refer to the raw average spectra obtained for individuals from the original database and from 14th October 2001 respectively.

Table 2. Correlation matrix for the validation set (n = 400)

	White level	Ext. temp.	Int. temp.	Predicted SS	Observed SS	PC1	PC2	PC3	PC4	PC5
White level	1									
External Temp.	−0.56	1.00								
Internal Temp.	−0.52	0.93	1.00							
Predicted SS	−0.18	0.48	0.48	1.00						
Observed SS	0.06	0.05	0.10	**0.63**	1.00					
PC1	−0.34	**0.68**	**0.67**	0.44	0.12	1.00				
PC2	−0.14	0.51	0.43	0.19	−0.11	0.31	1.00			
PC3	0.36	**−0.67**	**−0.60**	−0.27	0.05	−0.21	**−0.82**	1.00		
PC4	−0.51	0.43	0.44	−0.19	−0.17	0.18	0.07	−0.37	1.00	
PC5	−0.51	**0.82**	**0.80**	**0.61**	0.26	**0.67**	0.28	−0.50	0.46	1.00

Variables included are: PC scores, external and internal temperature of the bulbs, the daily white level signal of the spectrometer, and the observed and estimated values of soluble solids (ºBrix). Independence of variables is found for correlation values below 0.13. Bold numbers enhance correlations above 0.6.

their spectral variables, and which were not acting for the reference database.

Also, in Table 2 the correlation between the internal quality parameter (soluble solids content) is explored in relation to the PCs and to the estimated soluble solid content using the model developed for the 2001 season. There was no statistically significant correlation between the observed soluble solid content and the PCs, while there was a significant correlation with the estimated soluble solid content according to the prediction model of 2001. This value (R = 0.63) is low when compared to that of the calibration set of the model (R = 0.86). Still only 2% of the individuals corresponding to low quality were erroneously classified as high class, and none of the high class bulb were rejected as belonging to a low class (data not shown). Although none of the PC scores showed a relevant correlation with the observed soluble solid content, the procedure used for model adjustment enables the relevant information concerning estimation of the quality parameter, soluble solid content, to be extracted.

Figure 4 shows the Hotelling T^2 distance computed for the combination of PC scores (1 to 5) along the 66 work dates in the industry. The upper control limit (UCL) for two different confidence intervals (90 and 95%) were also plotted considering an average of 150 individuals per date and events were identified for 12 out of 66 test dates. The highest abnormality was found for the 14th October 2002. Data for this date are also shown in Figure 3 as circles (○).

Figure 5 represents the Q statistic for all 10,400 average spectra of bulbs. The UCLs for 90% and 80%

confidence intervals are plotted to identify abnormality of isolated individuals. This procedure can be easily implemented for real time applications.

The evolution of the daily averages of isolated PC scores over time has also been studied. Figure 6 shows the cumulative sum of the daily averages for PC5 scores throughout the season. A clear trend for PC5 towards lower daily values was observed from July to December. This PC factor had been previously related to the surface and internal temperature of the bulbs, as
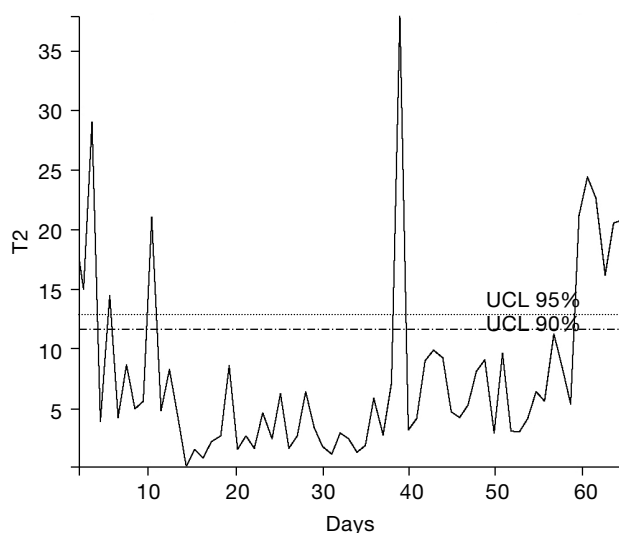

Figure 4. Hotelling T2 computed for the combination of the five PC considered (1 to 5). Events are found for 12 out of 66 test dates in the industry. The highest deviation is found for the 14th October 2002. This date also corresponds to the data shown in Figure 3 as circles (○). UCL stands for Upper Control Limit which has been computed for 95% and 90% confidence levels, and an average of 150 individuals tested per date.
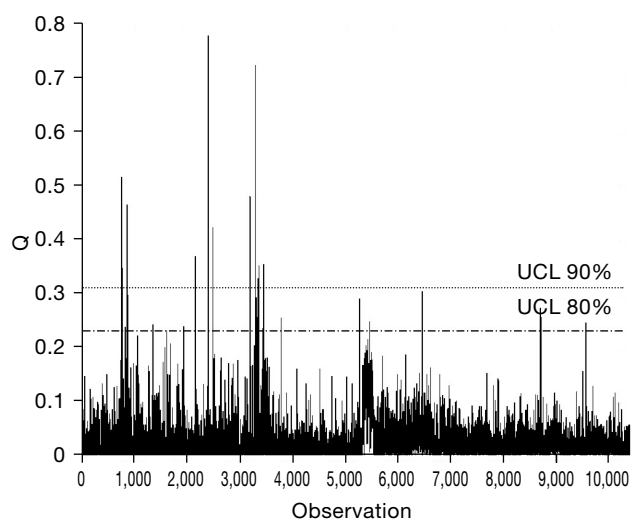
Figure 5. Q statistic computed for the combination of the five PC considered (1 to 5) for all 10,400 bulbs tested. UCL stands for Upper Control Limit which has been computed for 90% and 80% confidence levels. Values above UCL refer to significant mismatch between the original and predicted spectra within the PC model.

stated in Table 2. No clear time dependent trends were found for the remaining PC scores.

Discussion

The strategy for the use of estimation models presented in this work exploits the concept expressed by Wortel *et al.* (2001) as a first step in the transfer of
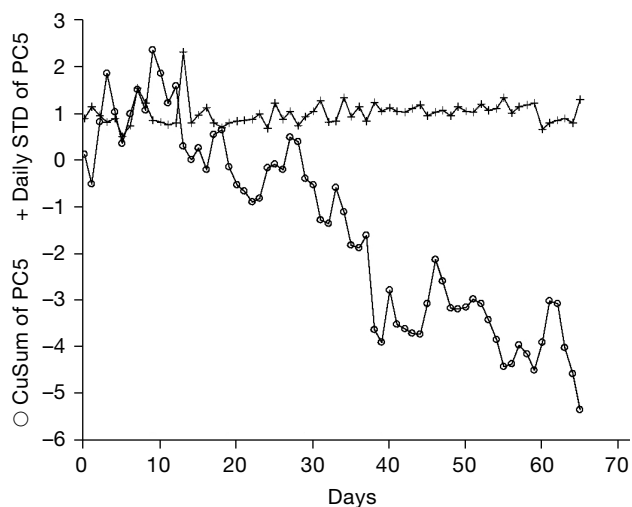

Figure 6. Time evolution of cumulating daily average (CUSUM) and standard deviation (STD) for PC5 score along 66 work dates in the industry. A temporal evolution is found which may relate to the temperature evolution over the season.

technology of the NIR technique to industry. We developed a not overfitted model in 2001 which had been tested with several calibration sets, and used it in the industry to classify into a reduced number of quality categories. To achieve accurate predictive models, wide representative calibration samples can be taken in industry. The amount of product processed in the industry per day allows a straight forward generation of wide spectral databases and preclassification of individuals. The results of even a coarse classification enables the selection of representative data sets to validate the model. Wortel *et al.* (2001) proposed the contamination of calibration data sets with noise in order to test the model robustness but with the current method new validation datasets can be obtained by daily application in the industry.

Schaare and Fraser (2002) evaluated, in the estimation of the internal quality of kiwifruits, several procedures to pre-process the spectra and even to select the most suitable mathematical parameters to obtain a predictive model. They concluded that variables related with the second derivative of the spectra are the best of all the possible transformations of spectral variables. From our point of view, this decision is risky in the estimation of some quality parameters, like soluble solids, since the derivation procedure leads to a high increase in the noise to signal ratio of data and may complicate further model transfer to the industry. Still the use of derivation may not be totally inappropriate when searching for main sources of variation in spectra like water (Ortiz *et al.*, 2001).

The lack of robustness in NIR models is well known in spectrometry. Guthrie *et al.* (1998), working with pineapples and melons, already stated the unfeasibility of using NIR models for quantitative estimations even within the same season though they proposed their use for classification purposes. This is also one of the conclusions of our study.

Fraser *et al.* (2003) emphasize the fact that internal boundary layers in the products are critical for transmittance measurements but not so much for other types of sample presentations such as reflectance or interactance. This point is also very relevant for onion bulbs since the internal layers lose tightness during the post harvest period, and will have to be taken into account when trying to shift from interactance to transmittance in order to facilitate online measurements.

A main result of this study is that the periodical generation of spectral databases under industrial applications, together with simple multivariate data analysis, allows a continuous unsupervised extraction

of NIR features. The application of world wide accepted multivariate techniques like PCA combined with process control statistics (Hotelling T^2) enables events to be identified in the daily process using PC scores. The presence of abnormal spectral responses of isolated bulbs can be highlighted as outlier dots in a normal probability plot though the main drawback concerning normal probability plots is that they may only be computed off-line when a whole population of data has been gathered. Instead, the use of the Q statistic commonly used for process control may easily be implemented in real time analysis for addressing spectra which are not well reproduced within the PC space.

Another important result or outcome of this work is the interest of studying the evolution in time of the PC scores of individuals. In this work a very close relationship is found between PC5 (0.08% of the spectral variance) and advance of the season (from July to December, 66 work dates). This PC factor (PC5) was also found to be highly correlated with the temperature of the onion bulbs in the validation set (r = 0.8). The effect of the temperature of the product on the NIR spectra is known. Hernández-Sánchez *et al.* (2003) estimated the bias in soluble solid prediction in apple due to a 30ºC temperature variation around 2ºBrix. This temperature variation is excessive when considering harvesting conditions, and even extreme when comparing fruits from cold storage and ambient conditions. In the case of onion bulbs over the postharvest season we expect a 10ºC variation range when no temperature control is used. Bulb temperature may then be integrated as an independent variable in the prediction models. Another interesting possibility to be explored is the use of specific PC information like PC5 for the internal correction of temperature in the spectra. A general conclusion of this work is that it is possible to address abnormalities in the NIR spectra emerging from non supervised analysis of industrial databases, which can then be related to new sources of variation.

Acknowledgements

References

BARREIRO P., MOLERO J., RUIZ-ALTISENT M., 1999. Informe 1ª fase. Estudios previos, medidas de referencia y caracterización óptica en bulbos de cebolla. Agrotecnica Extremeña S.L. Internal enterprise documentation.

BARREIRO P., HENCHE E.L., MUÑOZ M.A., 2002. Predicción de calidad interna en cebolla con espectroscopía NIR y desarrollo de un software de análisis. Agrotecnica Extremeña S.L. Internal enterprise documentation.

FRASER D.G., JORDAN R.B., KÜNNEMEYER R., MCGLONE V.A., 2003. Light distribution inside mandarin fruit during internal quality assessment by NIR spectroscopy. Postharvest Biol Tec 27, 185-196.

GUTHRIE J., WEDDING B., WALSH K., 1998. Robustness of NIR calibrations for soluble solids in intact melon and pineapple. J Near Infrared Spec 6, 259-265.

HERNÁNDEZ-SÁNCHEZ N., LURON S., ROGER J.M., BELLON-MAUREL V., 2003. Robustness of models based on NIR spectra for sugar content prediction in apples. J Near Infrared Spec 11, 97-107.

KARAKAYA F., KOBU B., 1994. New product development process: An investigation of success and failure in high technology firms. J Bus Vent 9(1), 49-66.

KRIVOSHIEV G.P., CHALUCOVA R.P., MOUKAREV M.I., 2000. A possibility for elimination of the interference from peel in nondestructive determination of the internal quality of fruit and vegetables by VIS/NIR spectroscopy. Lebensm Wiss u-Technol 33, 344-353.

MELLOR P., 2001. Definition of «latent fault». IEC/TC56/WG1. Available in http://www.cs.york.ac.uk/ hise/safety-critical-archive/2001/0238.html [28 Oct 2004].

ORTIZ C., BARREIRO P., CORREA E., RIQUELME F., RUIZ-ALTISENT M., 2001. Non-destructive identification of woolly peaches using impact response and NIR spectroscopy. J Agric Eng Res 78(3), 281-289.

SCHAARE P.N., FRASER D.G., 2000. Comparison of reflectance, interactance and transmission modes of visible-near infrared spectroscopy for measuring internal properties of kiwifruit. Postharvest Biol Tec 20, 175-184.

SIMOGLOU A., MARTIN E.B., MORRIS A.J., 2000. Multivariate statistical process control of an industrial fluidised-bed reactor. Control Eng Pract 8, 893-909.

SUMMERS A.E., RANEY G., 1999. Common cause and common sense. Designing failure out of your safety instrumented systems (SIS). ISA T 38, 291-299.

WORTEL V.A.L., HAMSEN W.G., WIEDEMANN S.C.C., 2001. Optimising multivariate calibration by robustness criteria. J Near Infrared Spec 9, 141-145.