ANEXO IV

Strategies for constructing the calibration set for a near infrared spectroscopic quantitation method

M. Blanco y M. A. Romero

Enviado para su publicación

STRATEGIES FOR CONSTRUCTING THE CALIBRATION SET FOR A NEAR INFRARED SPECTROSCOPIC QUANTITATION METHOD

M. Blanco* and M.A. Romero

Departament de Química, Unitat de Química Analítica, Facultat de Ciències,

Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain

ABSTRACT

Two different approaches to constructing the calibration set were tested in the development and validation of a near infrared spectroscopic quantitation method. The two approaches use as many types of samples, namely: laboratory samples obtained by mixing the ingredients of the pharmaceutical preparation concerned, and doped samples obtained by under- and overdosing production samples. The ensuing models were validated with a view to determining their fitness for purpose. However, spectral differences between the laboratory samples and doped samples resulted in spurious predictions in quantifying samples of one type using the model developed from samples of the other. Such differences were studied in depth and various procedures tested with a view to their minimization or suppression.

Keywords: near infrared spectroscopy, NIR, calibration, Pharmaceutical analysis.

1. Introduction

The last two decades of the XX century have no doubt seen the greatest expansion of near infrared spectroscopy (NIRS). The simplicity, precision and expeditiousness of this technique, in addition to an improved knowledge of the chemometric tools required to apply NIRS-based methodologies, have extended its use to virtually all industrial areas.

The pharmaceutical industry has shown special interest in the NIRS technique on account not only of its expeditiousness and non-destructive character, but also, especially, of its flexibility for both qualitative analysis (*e.g.* in the identification of raw materials and finished products,[1, 2] reaction monitoring in process control operations,[3,4] monitoring of blending processes,[5] control of film coating procedures [6]) and quantitative analysis (*e.g.* in the determination of active principles in commercially available preparations,[7,8] moisture[9] or even polymorphs[10]).

Proper development of an analytical methodology entails using samples representative of that to be analyzed and spanning an adequate concentration range. This is especially important in the NIRS technique, where spectra depend not only on the chemical properties of the sample but also on physical properties of its matrix including particle size, shape and distribution, or degree

of compaction, all of which significantly affect the spectroscopic signal. Consequently, the calibration samples used should be representative of chemical variability (in the concentrations of the active principle and excipients) and physical variability (associated to the manufacturing process and arising from particle size, the degree of compaction, *etc.*). Meeting both requirements in constructing a calibration model is usually difficult as the active principle and excipient concentrations are very close to the nominal value in virtually all samples of the pharmaceutical preparation. A number of procedures have been developed with a view to overcoming this problem, however.

One approach involves preparing laboratory samples by mixing accurately weighed amounts of the active principle and excipients in appropriate proportions in order to expand the concentration range spanned to the desired bounds. This is probably the most simple and convenient choice; also, it allows one to design a sample preparation approach that minimizes correlation, facilitates the development of robust models and provides highly reliable reference values (weighings) for the analyte. However, laboratory samples are not obtained using the same procedure as production samples, so none of the physical variability in the manufacturing process is included in the calibration process. The solution usually involves expanding the calibration set with production samples in order to incorporate matrix variability.[11]

One other approach involves preparing synthetic samples at a pilot plant reproducing the operations of the production plant. This method is much more labour-intensive and expensive; also, its feasibility depends on the particular type of sample and on the concentration of the active principle in the preparation, and the ensuing model may incorporate some source of variability not present in the production samples. This method is usually employed in the analysis of tablets using transmission measurements.

A third choice involves under- and overdosing production samples with small amounts of the excipients and active principle, respectively, in order to extend the original concentration range. This method is somewhat more laborious than preparing laboratory samples but undoubtedly more expeditious than the pilot plant method. The procedure has the advantage that differences between doped (under- or overdosed) samples and production samples are smaller than with laboratory samples as the addition of small amounts of the excipients or active principle does not alter matrix effects, so the ensuing calibration models are usually more simple. Correct doping requires that samples be in powdered or granular form, which may entail applying some sample pretreatment.

This paper compares the performance of two calibration procedures based on

laboratory and doped samples in the development and validation of a method for the determination of an active principle in a pharmaceutical preparation.

2.Experimental

2.1 Hardware and software

NIRRS spectra were recorded on a NIRSystems 6500 near infrared spectrophotometer from Foss NIRSystems (Raamsdonksveer, The Netherlands) equipped with a reflectance detector and a model AP6641ANO4P fibre-optic probe. The instrument was governed via a PC computer running the software Vision 2.22, also from Foss NIRSystems, for data acquisition. Laboratory samples were homogenized in a Turbula Type T2C Mixer from WAB (Basel, Switzerland).

Spectral pretreatments and multivariate calibration were both done using Unscrambler v.7.5 from CAMO (Trondheim, Norway).

2.2. Samples

The pharmaceutical preparation studied was a granulate with anti-inflammatory action containing nimesulide as active principle and sucrose as major excipient.

Laboratory samples were prepared by weighing the different components of the preparation in pure form and mixing them in variable proportions to span a concentration range ± 50% around the nominal content in the active principle. The mixtures were blended to homogeneity and their NIR spectra recorded.

Doped samples were obtained by supplying production samples of known concentration with also known amounts of the active principle (overdosed samples) or excipient (underdosed samples). Following doping, the samples were homogenized and their NIR spectra recorded. As with the synthetic samples, an active principle concentration range ± 50% around the nominal content was thus encompassed.

All samples (production specimens, the active principle and excipients) were supplied by Laboratorios Menarini (Badalona, Spain).

An overall 29 production samples (from as many different batches), 21 laboratory samples and 32 doped samples were used.

2.3. UV reference method

The active principle (nimesulide) content in the production samples was determined by ultrasonicating *ca.* 0.25 g of sample with 35 ml of MeOH for 10 min. The solution was then diluted to 50 ml with MeOH and a 5 ml aliquot was supplied with 5 ml of water and 2.5 ml of 1 M HCl, and made to 50 ml with 1:1 MeOH/H_2O. The nimesulide content in this solution was determined by applying multiple linear regression (MLR) to

the 220–450 nm region of its UV spectrum, using the spectrum for pure nimesulide as standard.

3. Results and Discussion

3.1. Construction of the calibration models

Both models were constructed using the PLS1 algorithm. Their predictive ability was assessed via the relative mean square error of prediction (RMSEP). Samples were split between the calibration and prediction sets to construct the two models, based on laboratory and doped samples.

In order to improve the predictive ability of the models, each set was expanded with a given number of production samples that were the same for both calibration sets, as well as with those added to both prediction sets. Table 1 shows the characteristics of the best models for each type of sample, as well as the results they provided.

Both models were very similar. In fact, both were constructed from first-derivative spectra, had the same number of PLS components and even similar predictive abilities; however, the model based on doped samples performed slightly better. In principle, both models provided accurate results and were suitable for determining the active principle. However, their actual suitability was ascertained by validation.

3.2.Validation of the models

Validating an analytical method entails determining whether it fulfills its intended purpose (*i.e.* its "fitness for purpose"). The two NIR methods corresponding to the previous calibration models were validated in order to determine whether they would allow the accurate quantitation of the active principle in the pharmaceutical preparation. To this end, their selectivity, accuracy, repeatability, intermediate precision, linearity and robustness were determining, following the ICH guidelines. [12]

Table 1. Characteristics of the models based on laboratory samples and production samples.

	Laboratory model		Doped Model	
	Calibration	Prediction	Calibration	Prediction
Samples	10 lab. 6 prod.	4 lab. 8 prod.	12 dop. 6 prod.	6 dop. 8 prod.
Wavelenth range	1100-2200 nm		1100-2200 nm	
Spectra pretreatment	1st derivative		1st derivative	
Number of factors	4		4	
RMSEP	1.4	1.7	0.9	1.0

3.2.1. Selectivity

The selectivity of a NIR method cannot be assessed as in other analytical methods. The proposed procedure involves identifying the pharmaceutical preparation in a library including various classes

corresponding to the preparation and its pure components. Because this identification step does not require the use of the quantitation models, we used the same validation procedure with both.

Five different samples from as many batches per product belonging to each class defining the spectral library constructed for this purpose were identified. The correlation ranges were established from the extreme values obtained in the identifications. The production samples were always identified as the pharmaceutical preparation, with a threshold of 0.98. Although the correlation coefficient of the production samples with sucrose was always high, none was confused with the excipient, so all were accurately identified (Table 2).

3.2.2. Linearity

The linearity of a multivariate calibration model is evaluated by plotting the results for a series of samples spanning a given concentration range against their reference values. The linearity of the two calibration models was assessed by using samples of the same type in each calibration (*i.e.* doped samples with the doped model and laboratory samples with the laboratory model). The linearity results are shown in Table 3. As can be seen, both models were linear throughout the concentration range studied.

3.2.3. Accuracy

Ten production samples were used to compare the active principle concentrations provided by both models with the reference values. A paired *t*-test of differences was conducted to this end that revealed the NIR values not to be significantly different from the reference values. As can be seen from Table 3, both models provided accurate values.

Table 2. Correlation coefficients obtained in the identification of unknown samples using the spectral library. Correlation ranges are the extreme values of five samples from each class identified in the library.

Library classes	Samples analyzed		
	Pharm. Preparation	Nimesulide	Sucrose
Pharm. Preparation	0.999 to 1.000	-0.142 to -0.120	0.957 to 0.979
Nimesulide		0.997 to 0.998	-0.229 to (-0.152)
Sucrose			0.996 to 0.999

Table 3. Results obtained in the validation of both calibration models for the determination of the active principle nimesulide.

VALIDATION RESULTS

ASPECTS	PROCEDURE	LABORATORY MODEL	DOPPED MODEL
Linearity	Nir Value= a + b·Ref.Value (a = 0 ; b = 1)	7 laboratory samples conc.range: 35-65 mg/g b = 1.02 ± 0.06 a= -2.18 ± 2.83 corr. = 0.999	10 laboratory samples conc.range: 35-65 mg/g b = 1.01 ± 0.09 a= -0.75 ± 4.81 corr. = 0.994
Accuracy	Paired t-test of NIR values and reference values of production batches	15 samples avg.diff.= -0.77 std. dev.= 3.54 t exp = 1.6 t critical = 2.14	15 samples avg.diff.= -0.37 std. dev.= 1.63 t exp = 1.13 t critical = 2.14
Repeatability	Production sample analysed 6 times by the same operator Calculation of % C. V.	$\bar{x}$ = 50.57 std. dev.= 1.41 % C. V. = 2.82	$\bar{x}$ = 50.54 std. dev. = 0.37 % C. V. = 0.72
Intermediate Precision	Production sample analysed 3 days by 2 different operators Calculation of % C. V. and ANOVA	$\bar{x}$ = 52.65 std. dev.= 2.16 % C. V. = 4.10 No signifficant effect of day and operator	$\bar{x}$ = 51.50 std. dev.= 0.68 % C. V. = 1.33 No signifficant effect of day and operator
Robustness	Comparison between NIR values and reference values (paired t-test) of production samples over a period of one and a half years	30 samples avg.diff.= 0.2 std. dev.= 2.8 t exp = 0.64 t critical = 2.03	30 samples avg.diff.= 0.15 std. dev.= 1.39 t exp = 0.91 t critical = 2.03

3.2.4. Repeatability

Repeatability was evaluated by having the same operator replicate the determination of the active principle in the same sample at least six times on the same day. Table 3 shows the results, alongside their standard deviations and percent coefficients of variation (% CV).

3.2.5. Intermediate precision

Intermediate precision was determined to establish between-day and between-operator variability. To this end, two operators replicated the determination of the nimesulide content in the same sample on three different days. The data thus obtained were subjected to a statistical study to determine their standard deviation and % CV, as well as to a variance analysis intended to establish whether either effect was significant. Table 3 shows the results obtained for both models.

3.2.6. Robustness

Robustness was assessed by checking the results obtained in the determination of nimesulide with both models using samples collected over a period of one year and a half. Table 3 shows the results of a test of differences between the results and the reference values. As can be seen, both models provided results consistent with the reference values, so both can be assumed to be robust.

From Table 3 it follows that both models are suitable for quantifying the active principle as the two provide acceptable results for each validation parameter. The model using doped samples provides slightly better results as regards repeatability and intermediate precision; both, however, are comparable in terms or accuracy and linearity.

3.3. Doped or laboratory samples?

Because both models passed the validation tests, both were in principle suitable for quantifying the active principle in the pharmaceutical preparation.

An attempt at quantifying samples of both types with the two PLS models revealed that samples of the same type predicted with both models yielded different curves. By way of example, Fig. 1 illustrates the result of quantifying doped samples using both models (note the two types of curve obtained). While the concentrations provided by the doped model were consistent with the reference values, those obtained with the laboratory model were not. Table 4 gives the figures of merit of the curves. The outcome was similar with laboratory samples: two clearly different curves.

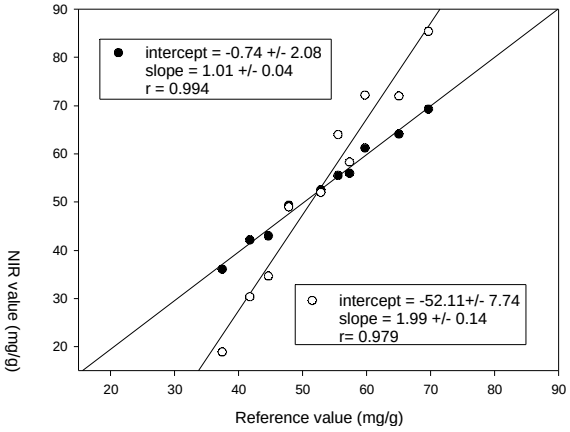

Figure 1. Quantitation of doped samples using the two PLS models: Doped model (·) and Laboratory model(_).

The differential response of the two types of samples was addressed by using various spectral pretreatments (*viz.* derivatives, SNV and MSC); all, however, led to results identical with those of Fig. 1, so the initial differences persisted and required deeper analysis with a view to establishing to what extent the doped samples departed from the laboratory samples.

Table 4. Figures of merit of the prediction curves for a set of doped samples using doped and laboratory models (constructed using PLS and MLR with different combinations of scores from the PCA).

Model	slope	intercept	r
PLS 4 factors			
Laboratory model	1.99 ± 0.14	-52.11 ± 7.74	0.979
Doped model	1.01 ± 0.04	-0.74 ± 2.08	0.994
MLR with PCs 1-2-3-4			
Laboratory model	2.02 ± 0.10	-54.66 ± 5.49	0.990
Doped model	1.01 ± 0.03	-1.13 ± 1.84	0.997
MLR with PCs 1-2-3			
Laboratory model	2.01 ± 0.10	-53.82 ± 5.47	0.985
Doped model	1.02 ± 0.04	-1.36 ± 2.07	0.996
MLR with PCs 2-3-4			
Laboratory model	1.84 ± 0.09	-52.09 ± 4.77	0.992
Doped model	1.02 ± 0.03	-0.91 ± 1.75	0.996
MLR with PCs 2-3			
Laboratory model	1.73 ± 0.05	-44.38 ± 2.86	0.997
Doped model	1.02 ± 0.03	-1.45 ± 1.76	0.997
MLR with PC 2			
Laboratory model	1.466 ± 0.54	-28.60 ± 29.38	0.69
Doped model	0.476 ± 0.176	27.84 ± 9.54	0.69
MLR with PC 3			
Laboratory model	2.38 ± 2.19	-95.89 ± 118.27	0.359
Doped model	0.13 ± 0.12	46.28 ± 6.40	0.359

Based on the nature of the variable reduction procedure used by PLS, whose decomposition involves the simultaneous use of variables and concentrations matrices, we chose principal component analysis (PCA), as the most suitable method for assessing the differences between doped and laboratory samples. Figure 2 shows the distribution in the first two principal components (PCs) in a PCA conducted on the three types of sample used in this work (*viz.* laboratory, doped and production), using first derivative spectra in the 1100-2200 nm range. Thus, production and doped samples were very similar, whereas, the first PC differentiates laboratory samples from both.

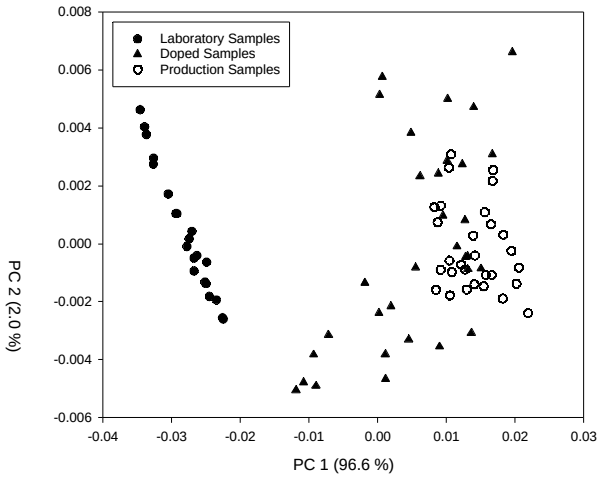

Figure 2. First and second principal components of the PCA conducted on Laboratory samples (·), Doped samples (_) and Production samples (_). First-derivative spectra in the 1100–2200 nm range has been used.

All samples were aligned along the second PC in terms of concentration, so this PC must contain important information regarding the nimesulide content in the three types of sample. No separation between samples according to origin was established by higher-order PCs.

In order to determine which PCs provided the best description for the active principle concentration, a score versus concentration plot was constructed for each PC (Fig. 3). The first PC clearly distinguished between laboratory and doped samples but exhibited no significant correlation with the concentration. On the other hand, the second and third PC exhibited good correlation, with no differences between samples. Finally, the fourth PC seemingly contained no information about the active principle content.

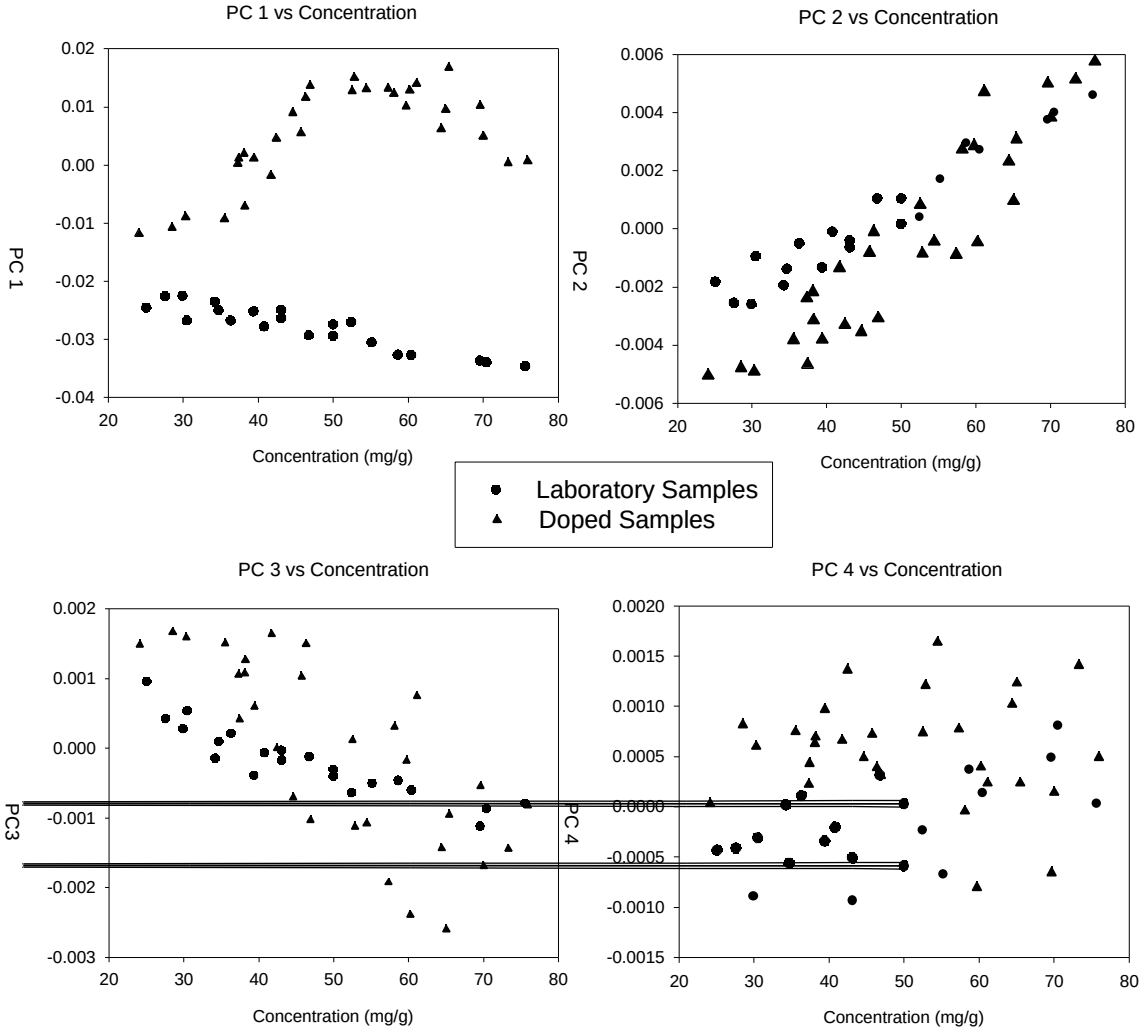

Figure 3. Distribution of PCA scores of all samples versus its nimesulide concentrations.

In order to ascertain the origin of the differences observed in applying the PLS models, multiple linear regression (MLR) models, using different combinations of scores as variable data, were constructed. The different MLR calibration models were applied to the same set of doped samples, comparing the results of prediction with laboratory and doped models constructed with the same scores combination). As can be seen from Fig. 4, two different curves were obtained in all cases; however, only the second and third PC were correlated with the concentration)the first and fourth contained no quantitative information. Table 4 shows the figures of merit of the models, which confirm the previous assertion and also that the models excluding the first and fourth PC are not statistically different from those including them. Also, the models based only on the second or the third PC cannot accurately predict the active principle content in both types of sample. A study of loadings also failed to provide a relationship allowing the above-described differences to be ascribed to specific physical or chemical properties.

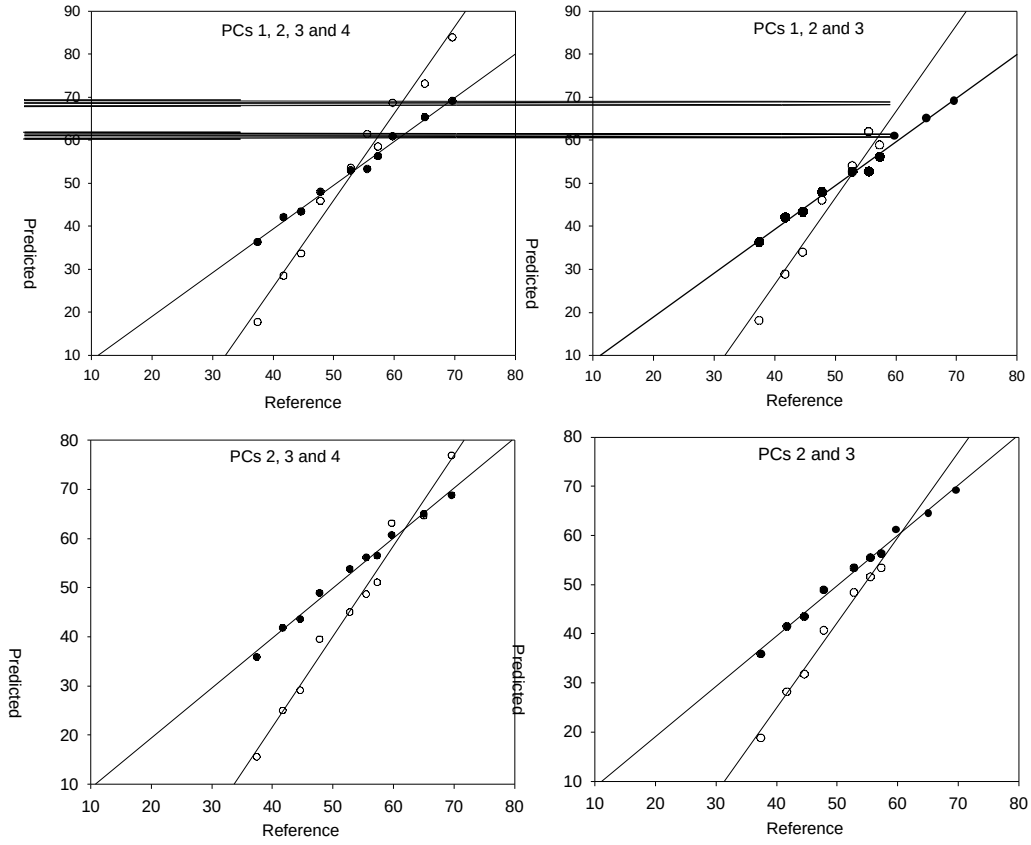

Figure 4. Prediction of doped samples by using MLR models constructed from doped samples (·) and laboratory samples (_).

Based on the foregoing, some remaining physical information about the samples is still contained in the second and third PCs and cannot be eliminated by the usual chemometric treatments. Since any calibration models that are to provide accurate predictions must contain these two PCs, the differences between laboratory and doped samples cannot be removed by this procedure.

Finally, models containing both types of samples (laboratory and doped) have been constructed. However these models required using as many as 6 PLS factors and provided poor prediction ability and dependent of the number of samples of each type included in the calibration set.

4. Conclusions

The use of laboratory samples or doped samples is equally effective with a view to constructing calibration models for the accurate determination of an active principle in a pharmaceutical preparation. Both types of model provide accurate calibration and prediction results. The results obtained in the validation of both calibration models confirm their fitness for purpose.

The differences between doped and laboratory samples cannot be ascribed to specific physico–chemical properties. Attempts at suppressing or minimizing such differences by using various mathematical treatments, MLR calibration to remove the information corresponding to the differences and even calibrations involving both types of samples, all failed. As a result, the validation procedure should always use samples of the same type as those employed to construct the model in order to avoid the deviations observed in this work (particularly in assessing linearity).

Acknowledgements

The authors are grateful to Spain's DGICyT for funding this research within the framework of Project BQU2000-0234. M.A. Romero acknowledges additional funding from Spain's Ministry of Education and Culture in the form of a researcher training grant. Finally, the authors wish to thank Laboratorios Menarini (Badalona, Spain) for kindly supplying samples.

References

[1] C. I. Geräusser and K. A. Kovar, Appl. Spectrosc. 51 (10) (1997)1504

[2] W. Plugge and C. Van der Vlies, J. Pharm. Biomed. Anal. 11 (1993)435.

[3] C. Coffey, B. E. Cooley Jr. and D. S. Walker, Anal. Chim. Acta 395 (1999)335.

[4] Z. Ge, B. Buchanan, J. Timmermans, D. DeTora, D. Ellison and J. Wyvratt, Process Control Qual. 11, (4) (1999)277.

[5] S. Sekulic, J. Wakeman, P. Doherty and P. A. Hayley, J. Pharm. Biomed. Anal.17 (1998) 1285.

[6] J. D. Kirsch and J. K. Drennen, J. Pharm. Biomed. Anal.13 (1995) 1273.

[7] D. Trafford, R. D. Jee, A. C. Moffat and P. Graham, Analyst 124 (1999)163.

[8] P. Corti, G. Ceramelli, E. Dreassi and S. Matii, Analyst 124 (1999) 755.

[9] I. R. Last and K. A. Prebble, J. Pharm. Biomed. Anal. 14 (1993)1071.

[10] M. Blanco and A. Villar, Analyst 125 (2000) 2311.

[11] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch and C. de la Pezuela, Anal. Chim. Acta 333 (1996) 147.

[12] *ICH Q2B: Validation of Analytical procedures: Methodology*, Consensus Guideline, International Conference on Harmonisation (ICH), 1998.