Application of Near-Infrared Spectroscopy to statistical control in freeze-drying processes

European Journal of Pharmaceutics and Biopharmaceutics [ISSN: 0939-6411], 168, 26-37

DOI: 10.1016/j.ejpb.2021.08.009

Application of Near-Infrared Spectroscopy to statistical control in freeze-drying processes

Serena Bobba[1,2], Nunzio Zinfollino[2], Davide Fissore[1]

1. Dipartimento di Scienza Applicata e Tecnologia, Politecnico di Torino, corso Duca degli Abruzzi 25, 10129 Torino

2. Biotech Pharmaceutical Development Department, Merck Serono SpA, via Luigi Einaudi 11, 00012 Guidonia Montecelio (Roma)

Abstract

Batch freeze-drying of pharmaceutical products in vials may result in a high degree of intra-batch variability due to several reasons, e.g. non uniform heating rate in the drying chamber. Therefore, product quality in the final product has to be checked in a statistically significant number of samples, in particular in the stage of process development. Here, Fourier-Transform Near-Infrared Spectroscopy is proposed as a fast, non-destructive technique for an off-line Statistical Quality Control application. At first, results obtained in a batch where product features are satisfactory are used to identify a target quality threshold. Then, a statistical controller is developed in such a way that in a production run it is possible to quickly check if product quality exceeds the desired threshold or not. Two approaches based on multivariate analysis are presented: one employs the Hotelling T^2 and Mahalanobis statistics to calculate control charts, the other is an application of Partial Least Squares for discriminant analysis (PLS-DA). Control charts and PLS-DA were trained with samples obtained in a run where sucrose solution was processed and validated in other runs where the final product was known to have the desired qualitative characteristics or not. Overall, out-of-specification samples can be predicted by control charts and PLS-DA with 99% and 98% accuracy respectively. PLS-DA was shown to be able to better identify samples correctly processed, while the control charts where more accurate to identify vials where something went wrong. Focusing on residual moisture of the final product, all samples where it was higher than the target value were always correctly identified.

Keywords

Near-Infrared Spectroscopy, Freeze-drying, Statistical control, Multivariate analysis, Discriminant analysis, Control chart.

List of abbreviations

CL	Control Limit
CQA	Critical Quality Attributes
KF	Karl Fischer
LV	Latent Variable
NIR	Near-Infrared
NIRS	Near-Infrared Spectroscopy
PAT	Process Analytical Technology
PC	Principal Component
PCA	Principal Component Analysis
PLS	Partial Least Squares
PLS-DA	Partial Least Squares Discriminant Analysis
SNV	Standard Normal Variate
SPE	Squared Prediction Error
SQC	Statistical Quality Control
T2	Hotelling T2

List of symbols

A	number of latent variables
b	mean value of the SPE distribution
$\mathbf{B}$	regression matrix of PLS
$\mathbf{E}$	matrix of residuals (M x J)
e_j	residual of the j-th observation
$F_{\alpha(A,\ M\text{-}A)}$	F-distribution
J	number of wavelength
M	number of samples

P	loading matrix (J x A)
Q2	model predictivity parameter
r	correlation coefficient
R2	fraction of the total sum of squares explained
T	score matrix (M x A)
t_a	a-th score
W*	fictious loading matrix (J x A)
X	spectral data matrix (M x J)
Y	quality attributes matrix (M x 1)

Greeks

α	confidence interval
λ_a	variance of the a-th latent variable
υ	variance of the SPE distribution
χ^2_α	χ^2-distribution

1. Introduction

Freeze-drying is a low-temperature and low-pressure drying process, aimed at drying a product with minimal thermal stress [1]. It should be the preferable drying technique in the manufacturing of pharmaceutical products, especially for formulations that are not stable in liquid phase and contain active compounds sensitive to heat, e.g. biotech drug products [2,3]. Water removal allows long-term stability, and freeze drying additionally guarantees sterile conditions, that are essential in a pharmaceutical factory.

A freeze-drying cycle is essentially made up of three steps. Firstly, the (bio)pharmaceutical solution is frozen, to convert the liquid solvent into ice. Following, the sublimation of ice crystals occurs at very low temperature and pressure conditions (primary drying) and the product turns into a solid cake. Finally, during the secondary drying, the temperature is increased, while low pressure is maintained, to allow the desorption of bound water [3].

Because of the potential effect on the patients, the pharmaceutical field is strictly regulated by authorities, and final products have to meet stringent specifications [4]. The focus is on the critical quality attributes (CQAs) of the product, including, for example, the residual moisture (RM), the recovery of the biological activity of the active ingredient, the potency and purity of the product [5]. CQAs are usually evaluated off-line, in the final product, by means of destructive and time-consuming laboratory analysis, e.g. Karl Fischer (KF) titration and chromatographic assays. As a consequence, very few samples are analyzed, giving a measure that could be poorly representative of the entire batch. In fact, the intra-batch variability of the samples from a batch is an intrinsic feature of the process [6]. Pressure gradients between shelves [7], non-uniformity of the shelf surface temperature [8], and different radiative heat to the edge vials [9] are some of the phenomena responsible of the intra-batch heterogeneity. Besides, the non-uniform nucleation temperature affects the size of the ice crystals and, at the end, the structure of the dried product. Their effect is obviously more remarkable in large scale equipment of manufacturing lines [7].

In this framework, Fourier-Transform Near-infrared Spectroscopy (NIR Spectroscopy or NIRS) can be regarded as a powerful analytical technology as it is non-destructive and non-invasive and requires a short analysis time [10]. NIR Spectroscopy is a method that uses the near-infrared region of the electromagnetic spectrum (in the range 14300-4000 cm^{-1}), by measuring the absorbance of the system related to the vibrational transitions in molecules [4]. Details about NIR theory can be found in many literature reviews [11-14].

NIR radiation can be adsorbed only by vibrations that result in changes in dipole moment, hence

by molecules having atomic groups with permanent or induced dipoles. Molecules with O-H, N-H, C-H, S-H groups are thus strong NIR absorbers [4] and this peculiarity makes NIRS a proper technique for detecting molecules like water, sugars and proteins, i.e. the molecules of main interest in a freeze-dried drug product.

Actually, if NIRS is compared to other spectroscopic techniques that have been applied to freeze-dried products, e.g. Raman spectroscopy, it comes out that NIRS is particularly suitable for detecting the residual moisture [4,15]. This is one of the most crucial CQAs, since water strongly affects the stability of the product over its shelf-life [16]. Moreover, NIRS was proved to be able to investigate proteins features in freeze-dried products [17-20], thus making NIRS a suitable tool for investigating the end-of-batch quality of biotech drug products.

Freeze-dried samples can be analyzed in the sealed vials, as no sample pretreatment is required, and the cake can be scanned through the glass vial [21-23]. Also, samples can be evaluated in the immediate vicinity of the manufacturing line [4], making any shipments to laboratories for analysis unnecessary. Consequently, more samples can be tested, giving more representative information about the batch.

Several applications of NIRS were proposed in literature for the freeze-drying of pharmaceutical products, both off-line and in-line. In-line applications were mainly aimed at implementing NIRS as a Process Analytical Technology (PAT) tool, monitoring the product modifications while being freeze-dried [4,15,24]. The off-line applications could be of two types:

- quantitative, i.e. NIRS was used for measuring a specific physical or chemical property of the sample, e.g. the residual moisture [21-27];
- qualitative, i.e. NIRS was employed to describe a product by comparing some of its characteristics with those of a reference sample.

The quantitative methods require an extensive experimental campaign for calibrating a model relating the spectrum measured to the property of interest.

The qualitative applications are based on the capacity of NIRS to gather in a spectrum several characteristics of a product. This can be exploited to compare samples to a reference one and assess the overall equivalence between them.

In the pharmaceutical field NIRS was implemented, for example, for the quality control of incoming raw materials, of blend uniformity, in the evaluation of hydrate forms formation, and of interactions between excipients [20,28-32].

Conversely to the NIRS-based quantitative models, the qualitative NIRS applications for freeze-dried product have been poorly investigated, mainly to prove the ability of NIRS-based

methods to discriminate between different products and different conformational structures of some components. For example, Bai et al. demonstrated that a native protein structure could be distinguished from the structure denatured by the freeze-drying process [17]; Zhou et al. were able to distinguish the surface water from bound water in the crystal lattice of a freeze-dried drug product [33]; Hansen et al. proved the applicability of NIRS to classify different freeze-dried viral vaccines [34]; in a later work of Hansen et al. NIRS was used for evaluating the interactions of the excipient with the virus in some vaccines during a stability study [35]; Li et al. [23] and Azheruddin et al. [36] used NIR spectra to evaluate cake structure and collapse. At this point it has to be underlined that a qualitative approach has remarkable advantages over the quantitative methods:

- the calibration is much faster and less demanding, since it is just required the availability of some samples in compliance with the reference;
- lab analysis is performed only on very few samples, thus jeopardizing the statistical relevance of the obtained results.

Usually, a Quality-by-Design approach is adopted through the development of the design space [37], to assure a good quality of freeze-dried products at the end of the cycle. The design space encompasses the combinations of process variables and parameters proved to give a product in compliance with the quality targets [38], and the final product is considered in compliance with the specifications if the cycle has been run within the design space limits. However, it may happen that a cycle runs outside the design space for a certain time because of unexpected faults, and this raise uncertainty about the product final quality. In this case statistical control can be helpful to be performed off-line, using Statistical Quality Control (SQC) [39]. SQC is a quality control procedure where samples are compared with a reference to check if they are compliant or not with the target.

In this paper an at-line application of NIRS is presented, intended to SQC. Two multivariate methods were developed, based on the statistical comparison with a reference dataset, namely a multivariate control charts approach, built from a Principal Component Analysis (PCA), and a Partial Least Squares Discriminant Analysis (PLS-DA). According to these qualitative analysis at the end of the cycle, a freeze-dried product may be accepted as in compliance with the reference. Conversely, it may be rejected if out-of-specification. It has to be remarked that this approach is fairly different from the more traditional one using NIRS to evaluate the residual moisture in the final product and, thus, taking a decision about acceptance or rejection of the product by comparing the measured value with the target one. In fact, the final output of the system proposed and investigated in this paper will be an acceptance or rejection statement

for each vial, and not the residual moisture of the sample. The advantage of this approach is that it is not required to carry out any preliminary calibration (e.g. using the time-consuming and destructive KF titration) of the model that allows getting the residual moisture from the NIR spectrum. Moreover, the reference model may be obtained from the spectra collected in one of the validation runs carried out in the production freeze-dryer, without running any additional tests. Finally, the actual NIR spectrum of the processed sample is globally compared to the reference ones and, thus, all the reasons that may result in a different spectrum (e.g. cake collapse, erroneous formulation, …) may be accounted for.

2. Materials and Methods

2.1 Experimental procedures

2.1.1 Case study

To develop and test the algorithms proposed in this study, three types of samples were needed:

#1: reference samples, for building the control charts and training the PLS-DA;

#2: samples of the same product of group #1, correctly processed as the reference samples;

#3: samples where the product has not been correctly processed or contains some excipients not included in the formulation of the reference samples. Thus, these samples should be rejected.

Groups #2 and #3 of samples are needed for validation purposes. Part of the samples of group #3 is also needed for PLS-DA development, as it will be shown in the following.

The product tested is a sucrose 6%$_w$ (Merck Life Science, Darmstadt, Germany) aqueous solution, processed into 2R glass vial (Nuova Ompi, Piombino Dese, Italy), with a filling volume of 1 mL. For samples preparation ultra-pure water by a Millipore water system (IQ 7000, Merck Millipore, Burlington, USA) was used.

The product was freeze-dried in a pilot-scale freeze-drier (Lyofast Mini 1.1, IMA Life, Bologna, Italy) at Guidonia Montecelio (Italy) site of Merck Serono SpA. A single freeze-drying cycle was run, and the operating conditions of the cycle were the following:

- freezing at -45°C for 6 h, with an annealing step at -15°C for 2 h (cooling/heating rate set at 2°C/min);
- primary drying at -25°C and 5 Pa for about 34 h (heating rate set at 2°C/min);

- secondary dying at 35°C and 5 Pa for 15 h (heating rate set at 1°C/min).

Care was paid to check that ice sublimation was completed before rising shelf temperature from the value used in the primary drying to those of the secondary drying stage. Pressure difference between Pirani and Baratron gauges was used to this purpose, and only when the difference in the pressure values was lower than the accuracy of the tools, ice sublimation was considered to be completed,

About 900 vials were processed, arranged in metal frames with a honeycomb layout, in direct contact with the freeze-dryer shelves. At the end of the process, samples were automatically crimped at the end of the unloading line, to prevent any contaminations from the environment. Two groups of 270 vials each (#1 and #2) were set up by selecting samples randomly among those produced. The first set (#1) was adopted as the reference, the second one (#2) was used for the validation step. As both datasets were obtained from the same batch, it is expected that all samples of dataset #2 are correctly evaluated as meeting the quality specification of the product. In order to accept #1 as reference, the CQA target was a residual moisture lower than 1%. Therefore, KF titrations were performed on 25 vials of group #1. Since this analysis is usually performed on a limited number of samples, e.g. few units, the number of vials analyzed in this study was thought to be higher and more representative. Even though #2 was produced in the same batch of #1 and the same range of residual moisture values was expected, also 25 samples of #2 were analyzed by KF titrations, as a further evidence of #2 being in compliance with #1. The coulometric titrator employed was a CX30 Mettler Toledo (Columbus, USA), the reactants purchased by Sigma-Aldrich (Saint Louis, USA), and the Standard Operating Procedure of the company was followed to perform the analysis. As expected, the samples of #1 and #2 had similar residual moisture values, as shown in Table A1 in the Supplementary Material, ranging between 0.51% and 0.98%, lower than the maximum acceptable. As a further check, mainly related to the limited number of samples whose residual moisture was measured through KF titration, assuming some sort of normally distributed residual moisture, the obtained values were fitted with a normal distribution. The mean value (± the standard deviation) calculated are respectively 0.72% (± 0.09%) for group #1, and 0.73% (± 0.10%) for group #2. It appears that, for the batch size considered, no samples have a residual moisture higher than 1%, as expected, both in group #1 and #2, thus proving the suitability of batch #1 as reference batch, and of dataset #2 as a batch that the control system has to accept as the product meets the target quality.

A second freeze-drying cycle was performed in a lab scale freeze-dryer (Lyostar3, SP Scientific, Warminster, USA), in order to produce samples of the group #3, in all cases out-of-

specification. Three groups of samples were produced, as detailed in the following:

(a) 100 vials of sucrose 6%$_w$, which is the same formulation as in case #1. However, at the end of the freeze-drying cycle, these samples were humidified manually, as explained below, so that the only difference with the #1 was the higher residual moisture;

(b) 50 vials of trehalose 6%$_w$, having the same solid fraction as formulation #1, but a different excipient, though amorphous as sucrose;

(c) 50 vials of a mixture of sucrose 6%$_w$ and arginine 2%$_w$, hence with the same amount of sucrose as formulation #1, but with the addition of an amino-acid.

The above sample sets (a), (b), and (c) were used to set up, respectively, the #3a, #3b, #3c sample sets.

To prepare these solutions, sucrose and L-Arginine Monohydrochloride were supplied by Merck Life Science (Darmstadt, Germany), trehalose dihydrate by Sigma-Aldrich (Saint Louis, USA), and ultra-pure water by a Millipore water system (IQ 7000, Merck Millipore, Burlington, USA). Similarly to the #1 and #2 samples, also samples of these groups were freeze-dried in 2R vials, filled with 1 mL of solution. All samples were freeze-dried according to the same protocol as for the #1 and #2 samples: the goal of this test was to check if the system was able to properly identify out-of-specification vials and, thus, it was not necessary to carry out any process transfer operation from the first to the second freeze-dryer (i.e. to modify the shelf temperature/chamber pressure aiming to get the same product dynamics) as, in all cases, the product was not meeting target quality for several reasons.

For the humidification of the samples of group (a), given the weight of the dried cake and assuming an initial residual moisture equal to 0.5%, the amount of water necessary to get a final residual in the range 1.5% - 5.0% was calculated. Then, vials were unsealed, the calculated amount of water dropped on the internal face of the stoppers, laying on a flat surface, and the vials closed upside-down on the stoppers, i.e. with the bottom facing upside. This allowed the diffusion of water, avoiding the drops to get in direct contact with the cake, which would cause a partial collapse of the cake structure. Vials were left in the upside-down position for more than 24 h, to allow a complete diffusion of water. The time after humidification was proved in a previous study [55] with similar samples of freeze-dried sucrose to be long enough to allow an homogeneous distribution of water in the cake. Then vials were crimped and stored at room temperature with the vials not-humidified. The samples of group (a), after humidification, were selected as sample set #3a, made of 95 samples. To assess the effect of the humidification, 10 samples of #3a were analyzed by KF titrations, resulting in a range of residual moisture between

1.55% - 5.46%, as shown in Table A2 in the Supplementary Material. This means that the gap between the sets of samples considered acceptable and the set of samples non-compliant was about 0.57% in terms of residual moisture.

The #3b and #3c sample sets were made by selecting 35 and 25 samples respectively from the products of group (b) and (c). Additionally, also 5 samples of #3b and 5 of #3c underwent to KF titration, in order to verify that their residual mositure contents were comparable with the residual moisture of #1: results are shown in Table A2.

Since the KF titrations proved that #1, #3b, and #3c had similar residual moisture, the discrimination performed by the models presented for groups #3b and #3c was not affected by the residual moisture content. They were used to evaluate the ability of the algorithms to distinguish different amorphous excipients and to detect the presence of an amino-acid. On the other hand, #3a was meant to be different from #1 just because of the higher residual moisture, being essential proving that the models can distinguish samples with higher residual moisture.

2.1.2 Spectra acquisition

All the samples of the sets presented above (#1, #2, #3a, #3b, #3c) were analyzed by NIRS, in order to obtain the spectral datasets. A Fourier Transform NIR spectrometer (Antaris MX FT-NIR, Thermo Fischer Scientific, Waltham, USA) was used to this purpose: the instrument was equipped with a halogen NIR source and an InGaAs detector. Spectra were recorded in diffuse reflectance mode, with 32 scans, a resolution of 8 cm^{-1}, and a gain of 1, in the wavelength range 10000 - 4000 cm^{-1}.

For each sample, the spectrum was obtained as the average between two spectra, to reduce the noise and get a better-defined spectral path. The probe of the spectrometer was installed on the visual inspection station, just beside the unloading line (Figure A1 in the Supplementary Material), and samples were scanned through the side wall of the vials, i.e. the NIR beam spot pointed on the side of the freeze-dried cakes. A previous study [55] proved that comparable results could be obtained by processing spectra collected from the side wall or through the bottom of the vial. Thanks to spectral pretreatment techniques, which are described in the following, the effects due to the curvature of the glass are strongly minimized, thus legitimizing the use of spectra collected both through the side or the bottom of the vial. The setup adopted was suitable for:

(i) setting an at-line procedure that would speed up the analysis of the batch;

(ii) gathering in a single station two quality control analysis;

(iii) managing the acquisition time of the spectra, in order not to slow down the unloading

of the freeze-drying chamber, that would occur if vials were scanned directly on the unloading line.

2.2 Multivariate analysis

Each NIR spectrum is made of hundreds of values (one for each wavelength) and in case of off-line applications, as the one here presented, and several samples are analyzed. Also, one or more spectra are collected for each sample, aiming at repeating the analysis to get signals of better quality [40], or to acquire information from different spots of the sample [30,41]. Therefore, the amount of data obtained for a NIRS application is huge and chemometric analysis is needed to extract relevant information [4]. Among the chemometric methods multivariate analysis is widely applied. This is a statistical procedure for analyzing data when many (correlated)-variables are involved at a time, and several measurements are available for each sample [20]. Among the multivariate analysis techniques Principal Component Analysis (PCA) and Partial Least Squares (PLS) are widely used.

PCA and PLS are based, respectively, on principal components (PCs) or latent variables (LVs), which are not directly observed, but inferred from the variables that can be observed and measured. PCA is a linear algebra tool, which reduces a complex dataset, in this case the spectral dataset (**X**), to a simple structure in the space described by the PCs. In this space extracting relevant information is easier [42]. PLS is a linear regression method that models the relationship between the matrix of spectral data (**X**) and the matrix of attributes (**Y**). The main difference between two methods is that PCA describes the variability among the dataset, i.e. the variance, while PLS is focused on the covariance, that is the joint variability of the dataset and its attributes [43]. An extensive description of these techniques can be found in dedicated works [43-46].

2.2.1 Principal Component Analysis and multivariate control charts

PCA is based on the linear decomposition of the *M* x *J* data matrix **X** in the reference system made of *A* latent variables, in the form:

$$X = T P' + E \quad (1)$$

In this case, the dimension of the data matrix **X** corresponds to the number of samples (*M*) and to the number of signals (*J*) collected for each spectrum, i.e. to the *J* wavelengths scanned for each spectrum. **P** is the *A* x *J* loading matrix, and the loadings are the vectors constituting the

new reference system, along the direction where the variance is maximum [47]. The size of the loading matrix, i.e. the number of PCs, is a crucial point: the higher the number of PCs used, the more variance among **X** is explained, but overfitting must be avoided. **T** is the *M* x *A* matrix of the scores, that are the projections of the original observations in the new basis, i.e. they represent the spectra in the new reference system. The distances between the observations and their projection in the new reference system are the residuals, collected in the residuals matrix (**E**). The residuals are minimized by the decomposition to get a better representation of the observations in the new reference system [48].

PCA can be used for a discriminant analysis by means of two multivariate control charts. These are built accordingly to two statistics, the Hotelling T^2 (T2) and the Squared Prediction Error (SPE), obtained by calculating for each observation *i*:

$$T2_i = \sum_{a=1}^{A} \frac{t_a^2}{\lambda_a} \tag{2}$$

$$SPE_j = \sum_{j=1}^{J} e_j^2 \tag{3}$$

where t_a is the *a*-th score, λ_a the variance corresponding to the *a*-th PC and e_j is the residual of the *j*-th observation [39]. The T2 is related to the Mahalanobis distance of the *i*-observation from the origin of the new reference system, weighted by the corresponding variance λ_a. Thus, an observation with an anomalous value of T2 does follow the variance structure of the model, but it lies outside the distribution where most of the observations are, i.e. it has an extreme value of variance. The SPE is the perpendicular distance of the *i*-observation from the (hyper)-plane described by the loadings vectors, giving a measure of what the model does not fit. An observation with a high SPE has thus some variables that are not taken into account by the model, i.e. it does not follow the variance that is suitable to characterize the major part of the observations [49].

The control charts make use of control limits (CLs) that are some reference values calculated from the T2 and SPE statistics. If an observation has a T2 or SPE value higher than the respective CL, it may be considered as an outlier. CLs are calculated, for the first *A* PCs, as follows:

$$CL_{T2} = \frac{(M^2 - 1)A}{M(M - A)} F_{\alpha(A,I-A)} \tag{4}$$

$$\mathrm{CL_{SPE}} = \frac{\upsilon}{2b}\chi^2_\alpha\left(\frac{2b^2}{\upsilon}\right) \quad (5)$$

where α is the confidence interval of the χ^2_α distribution and of the $F_{\alpha(A,\ M\text{-}A)}$ distribution with (A, M-A) degrees of freedom, and b and υ are respectively the mean and the variance of the SPE distribution [48]. Some clarifications about the meaning of the confidence interval may be supportive. As an example, if α is equal to 99%, all the observations within the 99% of the distribution are considered, while if α is equal to 95% a smaller number of samples (within the 95% of the distribution) is taken. Consequently, only the observations that better fit the distribution are found within the CLs when a lower value of α is adopted, thus representing a more stringent condition.

2.2.2 Partial Least Squares Discriminant Analysis

PLS makes a linear decomposition, of both **X** and **Y** in this case, looking for some LVs able to model **X** and **Y** at the same time. Then, the matrix of regression (**B**) is calculated with the following linear combination [43]:

$$\mathrm{B} = \mathrm{W}^* \mathrm{Q}' \quad (6)$$

Q is the M x A loading matrix of **Y** describing the covariance in **Y**. **W*** is a J x A fictious loading matrix, made of the coefficients (or weights) that describe the covariance between **X** and **Y**. The fictious loading matrix $\mathbf{W}^*$ is needed to project the observations in the space of the PLS decomposition, i.e. to calculate the score matrix **T** [48]:

$$\mathrm{T} = \mathrm{X}\,\mathrm{W}^* \quad (7)$$

Finally, given a new dataset ($\mathbf{X}_{pred}$), its quality attributes ($\mathbf{Y}_{pred}$) can be predicted with the regression matrix [48]:

$$\mathrm{Y}_{pred} = \mathrm{X}_{pred} \cdot \mathrm{B} \quad (8)$$

A type of PLS is PLS for Discriminant Analysis (PLS-DA) [50]. It is focused on the separation of the observations in classes and, after training, it is able to associate a new observation to a specific group [34]. For this purpose, the **Y** matrix is filled with binary values, representing the membership of an observation to a class rather than to the other, i.e. the attribute associated to the observation is qualitatively rather than a precise number coming from a certain physical property. Actually, a strategy to turn the **Y** matrix into a categorical matrix, is introducing two dummy values (or flag values), i.e. binary values of Boolean type (true or false). Thus, **Y** turns out as an (M x 1) categorical matrix, i.e. an array of binary values describing a categorical

variable of **X** [51]. A regression matrix is obtained, and the outcome results in a predictive clustering of the observations, that can either be displayed in a score plot where the clusters may be distinguished, or adapted to a binary form in the $\mathbf{Y}_{pred}$ matrix.

A PLS-DA can be validated through a permutation test (or randomization test), aimed at testing the significance of the statistics obtained with the model [52,53], i.e. the model is not spurious and it does not fit just the calibration set. The categorical matrix of attributes **Y** undergoes a number of random permutations ($\mathbf{Y}_{rand}$), while the data matrix **X** is kept not permuted, and the test is based on the comparison between the original model and the permuted ones, by means of two parameters, Q2 and R2. The latter is the fraction of the total sum of squares explained by the model, i.e. it gives a measurement of the goodness of the fit in the training set, and Q2 is indicative of the model predictivity [34,53]. The Q2 and R2 values are plotted in a permutation plot, with respect to the Pearson correlation coefficient (*r*) , calculated for each permutation with the original **Y** and the permuted ($\mathbf{Y}_{rand}$). For a high validity of the model, the original Q2 and R2 values are expected to be higher and outside the distribution of the permuted ones [53]. Negative values of the indicators may also occur, as a further indication of the model validity [50]. For each permutation R2, Q2, and *r* are calculated with the following equations [54]:

$$\mathrm{R2\ (Q2)} = 1 - \frac{\sum_{i=1}^{M}\left(\mathrm{y}_i - \mathrm{y}_{pred,i}\right)^2}{\sum_{i=1}^{\mathrm{M}}(\mathrm{y}_i - \mathrm{Y_m})^2} \tag{9}$$

$$r = \frac{\mathrm{M}\sum_{i=1}^{M}\left(\mathrm{y}_i \cdot \mathrm{y}_{rand,i}\right) - \left(\sum_{i=1}^{M}\mathrm{y_i}\right)\cdot\left(\sum_{i=1}^{M}\mathrm{y}_{rand,i}\right)}{\left(\left[M\sum_{i=1}^{M}\mathrm{y}_i^2 - (\sum_{i=1}^{M}\mathrm{y}_i)^2\right]\left[M\sum_{i=1}^{M}\mathrm{y}_{rand,i}^2 - \left(\sum_{i=1}^{M}\mathrm{y}_{rand,i}\right)^2\right]\right)^{0.5}} \tag{10}$$

where $\mathbf{y}_i$, $\mathbf{y}_{pred,i}$ and $\mathbf{y}_{rand,i}$ correspond to the *i*-th observation of the **Y**, $\mathbf{Y}_{pred}$ and $\mathbf{Y}_{rand}$ matrices, and $\mathbf{Y_m}$ is the mean value of the **Y** matrix, paying attention to take the matrices of the calibration or validation set if R2 or Q2 is calculated.

3. Data processing

Processing of NIR spectra and data modeling were performed by means of algorithms written ad-hoc in MATLAB (R2019b), based on the algorithms described by McGregor and Kourti (1995) [46] and Wold (1992) [56].

3.1 Pretreatment of spectra

Prior to modeling, pretreatments of spectra are desirable in order to reduce some phenomena, e.g. light scattering, baseline shift, random noise [57]. Pre-processing of spectra was performed by means of standard normal variate (SNV) to study the spectrum in term of its variability from the mean value, to reduce the variability between samples due to light scattering, and to adjust the base-line shift [58]. SNV consists in subtracting the mean value of the spectrum from all its points and dividing for its standard deviation.

Attention was paid to reduce fictious modifications from the original data, and preserve the original information contained in the spectra. Therefore, other common pretreatment techniques, e.g. scatter corrections, derivates, and smoothing [42], were not implemented, since, for this application, the only SNV was found to return data suitable enough for high accuracy models.

After this, spectra have to be carefully evaluated as in some cases the NIR probe was not in perfect contact with the monitored vial and, thus, the spectrum appeared very different from the expected one. This issue is mainly related to our experimental setup, designed for proof-of-concept purposes, and is expected to be solved in an industrial application. These spectra, obtained after vial unloading from the freeze-dryer, must be removed from the analysis. This could be done manually. Aiming to speed up this step, we used at first PCA for removal of these spectra (considered as outlier) and then, we quickly checked that removed spectra were, effectively, the anomalous spectra, with a trend that could not be due to a different water content in the sample. No outliers of this type were evidenced in datasets #1 and #2, while, respectively, 6, 3 and 3 outliers were identified in datasets #3a, #3b and #3c. Figure A2, in the Supplementary Material, shows the spectra of all the sets involved, after pretreatments and removal of outliers over a wavelength range reduced to 9500 - 4100 cm^{-1}, since the signal collected at higher frequencies (wavelength lower than 4100 cm^{-1}) was too noisy. Here, the differences between the reference spectra and the other sets can be appreciated. #1 and #2 clearly differ form #3a for the intensity of the peak around 5150 cm^{-1}, a typical signal of water [22]. Only minimal differences can be seen with respect to #3b, while a very different shape characterize the spectra of #3c.

In Table 1 the subdivision of the spectral datasets between the calibration and validation sets is summarized, for both the methods proposed.

3.2 Multivariate control charts

3.2.1 Calculation of control limits

The first step for building the multivariate control charts was calculating the CLs, and this step can be considered as the calibration phase of the method. The spectra of the reference set #1 were used to this purpose.

They were, firstly, processed by PCA, which was functional for calculating the Hotelling T^2 and SPE statistics. Hence, for each observation, i.e. for each spectrum of the calibration set, the T2 and SPE values were calculated by means of equations (2) and (3). Following, the CLs were computed accordingly to equations (4) and (5), with a confidence interval α of 95% and 99%. In this way, two sets of CLs were obtained: the T2 and SPE CLs at α = 95%, and at α = 99%.

3.2.2. Validation

The validation of the method consisted of comparing each observation of the validation sets with the CLs calculated with the reference. For the sake of clearness, two validation sets were used (see Table 1) instead of a single one: the first was #2, the second was made of the datasets #3a, #3b and #3c. Therefore, from the validation with the #2 datasets, all samples were supposed to be found in compliance with the reference. Conversely, all the samples of the other datasets were expected to be classified as outliers.

For each observation of the validation sets, the T2 and SPE values were calculated by means of equations (2) and (3), and these values compared with the respective CLs. An observation was considered an outlier if its T2 or SPE value resulted higher than the corresponding CL calculated at 99% of the confidence interval. For this classification, the CLs at α = 99% were preferred to those at α = 95%. Indeed, a sample exceeding a CL at α = 95% might be an extreme observation, i.e. it lies in the outer region of the distributions, but still within the distributions, while a sample exceeding a CL at 99% is much more likely to be non-compliant. Lately, this is the reason why two confidence intervals were adopted, i.e. to distinguish between a potential extreme observation and a definitely outlier.

3.3 PLS-DA

3.3.1 Calibration

A PLS-DA has to be trained by a categorical binary matrix **Y**, which means that samples of both classes (compliant or not with the target value) have to be involved in the calibration step. It has to be underlined that this is a major difference with respect to the calibration step performed for building the control charts. In fact, for calculating the CLs of the control charts,

only the reference samples were required. Conversely, for calibrating the PLS-DA at least two groups are needed: one of samples in compliance with the CQAs and one or more referring to samples that should be rejected.

For this reason, the calibration set of the PLS-DA model was made not only by #1, but also by half of the spectra of the dataset #3, which were selected randomly (see Table 1). In this framework it has to be remarked that the two datasets considered in this part of the study are not composed by a "similar" number of samples, as group #1 is larger than group #3. Anyway, as this "unbalanced" set of data does not appeared to affect the results, no further analysis was considered to be necessary.

Following, the **Y** matrix was filled with a flag value (+1, meaning "true") if the corresponding spectrum belonged to #1, and with another flag value (-1, meaning "false") if the observation was from another set. The calibration step consisted in performing a PLS regression with the calibration set and the **Y** categorical matrix, and calculating the regression matrix **B** by means of equation (6). In order to perform the PLS with the number of LVs that would yield the best model, the misclassifications of the calibration and prediction sets were analyzed as a function of the number of LVs [59]. For this analysis the numbers of misclassifications were quantified by calculating the RMSEC, RMSEP [60], and RMSECV [61] (respectively the Root Mean Square Error of Calibration, of Prediction, and of Cross Validation). The minimum number of LVs giving minimal errors was then selected for further calculations.

3.3.2 Validation

For the validation step of the PLS-DA, the data matrix **X** was made of the spectra of #2, plus the remaining half of the dataset #3 (see Table 1). The regression matrix was applied accordingly to equation (8), to predict the matrix of attributes $\mathbf{Y}_{pred}$ associated to the validation set. The outcome of this prediction was managed in order to obtain a binary $\mathbf{Y}_{pred}$, so that each spectrum of the validation set could be identified by a flag value, i.e. +1 or -1 if the sample was classified as in compliance or as an outlier, respectively.

Another validation of the PLS-DA was carried out through a permutation set. The dataset involved in the calibration were used for calculating R2 and those involved in the validation for calculating Q2. Following, the **Y** matrix underwent to 100 random permutations: for each permutation a new PLS-DA model was calibrated and validated, and the R2, Q2, and *r* values calculated through equations (9) and (10). Finally, the R2 and Q2 distributions of the permuted models were compared with the original R2 and Q2 values through a permutation plot.

4. Results and discussion

4.1 Control charts performances

Building the control charts was the first step, by making use of the sample set #1. After SNV, #1 dataset was analyzed by a PCA over the range 9500 - 4100 cm^{-1}. The loading plot showed that the two PCs had lower loading values in correspondence of the outer frequencies, i.e. the highest and lowest frequencies were found to be less significative (as shown in Figure A3, in the Supplementary Material). Therefore, the wavelength range was reduced to 7700 - 4590 cm^{-1}, where higher loadings could be seen, in order to focus on the meaningful regions of the spectra and avoid further potentially misleading signals. At this point, the first 2 PCs of a new PCA were found able to explain more than 99% of the variance among the spectra, thus they were considered enough for modeling the spectra of #1. For the sake of completeness, the loading plot over the reduced wavelength range are also displayed in Figure A3.

To underline the capacity of a PCA to discriminate observations in clusters, the score plot of all the datasets was traced and reported in Figure A4 in the Supplementary Material. This score plot is useful to test preliminary the ability of the PCA to differentiate observations of different groups, since the control charts are basically built on the discriminant capacity of a PCA. The score plot was obtained by the first two PCs of the PCA and it underlines the division in two clusters. The main cluster represents the reference spectra (#1), where also #2 scores could be found. The second cluster, clearly spaced from the previous one, contains all the other observations and a partial sub-division can be seen between the samples of #3a, #3b and #3c. It is not surprising that all the challenge data (#3) lie so far from the reference data. In fact, moisture level in dataset #3a is highly different from that of the reference batch (see Table A2), as expected as this test was carried out for this goal. With respect to datasets #3b and #3c, the difference in moisture level is not as high, but, due to the different excipient, the shape of the spectra is very different, as it appears in Figure A2.

The CLs were calculated at a confidence interval α of 0.95 and 0.99, from the Hotelling T^2 and SPE statistics obtained after the decomposition of #1 in principal components. As expected, the CLs at $\alpha = 0.95$ were lower than the CLs at $\alpha = 0.99$, since adopting a smaller confidence interval means taking as acceptable a smaller part of the distributions.

The control charts of calibration are showed in Figure 1: the same observations used for the

calculation of the CLs are compared with the CLs. Provided that an observation was rejected if its T2 or SPE value exceeded the respective CL, 13.3% of the observations would have been rejected at $\alpha = 0.95$, while only 1.5% appeared as rejects with $\alpha = 0.99$. These percentages are considered as calibration errors, since all the samples used for calibrating were expected to be identified as compliant with the target. The difference between the calibration errors at $\alpha = 0.95$ and 0.99 is a direct consequence of accepting a smaller / larger part of the statistic distributions. As expected, the CLs at $\alpha = 0.99$, being less restrictive, seemed more suitable for discriminating the outliers, and, lately, adopted as the reference values.

Dataset #2, made of spectra in compliance with the reference, were processed to calculate the T2 and SPE values and compared with the CLs. The control charts reporting the validation with #2 dataset are given in Figure 2. It can be seen that almost all the observations had the T2 and SPE values below the CLs at $\alpha = 0.99$, and the percentage of misclassified samples was barely 1.2% (3 samples out of 245, i.e. almost 99% accuracy). Thus, #2 samples were classified, as expected, in compliance with the reference.

A second validation was performed with the other datasets, being different for the product contained (#3b and #3c) or for the residual moisture (#3a). The validation with the #3a, #3b and #3c dataset is presented in Figure 3. Here the totality of the observations (113 samples) was correctly identified as not in compliance with #1. It is worth to notice that also the #3a observations were correctly identified as non-compliant: the method proposed was perfectly able to discriminate samples with higher RM, which was a fundamental ability to be proven.

The role played by the SPE statistic is highlighted by the validation with these datasets, for which very high SPE values were calculated (up to more than 128). Actually, the datasets were made by spectra of samples with some variability not taken into account in the calibration of the model, i.e. the samples used for calculating the CLs did not contain any trehalose, nor arginine, and had a lower RM. This turned out in high SPE values, essentially meaning that there was some more variance lying in the structure of these datasets.

4.2 PLS-DA performances

The spectra of the training set (see Table 1) were pretreated by SNV and the categorical **Y** matrix was filled with proper flag values. After the removal of outliers by means of PCA, a total of 300 spectra were available for training the PLS-DA. The calibration dataset was analyzed through a PLS-DA over the range 9500 - 4100 cm^{-1}, and the first 4^{th} LVs turned out to be preferable for modeling. In fact, as displayed in Figure A5 in the Supplementary Material, the performance parameters, i.e. RMSEC, RMSEP, and RMSECV, dropped markedly till the

4th LV. Instead, the decrease in correspondence of the 5th LV was minimal. Therefore, the last LV was not used for performing the PLS-DA, as it did not appear to be very beneficial, and avoiding data overfitting was crucial.

The loading plot of the first four LVs (Figure A6 in the Supplementary Material) was investigated to select the more significative wavelength range. The trends of the loadings showed that they had higher values in a narrower wavelength range, comparable with the range of the PCA performed for building the control charts. Thus, the same wavelength rage was selected to perform the PLS-DA. In Figure A6, also the loadings of the LVs over the smaller wavelength range are displayed, for sake of completeness. Finally, the model was trained by a PLS-DA with four LVs over the range 7700 - 4590 cm^{-1}, and the regression matrix **B** calculated. The regression matrix was firstly applied to the training set itself, to assess the performances of the calibration. The outcome was provided as a stem plot (Figure 4), with upward and downward categorical values, accordingly to the observation being classified as in compliance or not. All the samples of #2 were correctly identified as in compliance with #1, and just few samples were misclassified, i.e. considered in compliance while they were not. The calibration error was just 1.7% (5 samples out of 300) and all the samples misclassified belonged to #3b.

The validation of the PLS-DA was performed on the validation dataset, made of #2 and the remaining half of spectra of dataset #3 (see Table 1). The spectra of the validation set were pre-processed similarly to the calibration set: pretreated by SNV, the outliers removed by a PCA with two PCs, and the wavelength range reduced to 7700 - 4590 cm^{-1}. Hence, 303 spectra were available for validating the model. The binary matrix $\mathbf{Y}_{pred}$ was calculated through the regression matrix **B** calculated in the calibration phase. The classification performed by the PLS-DA is presented in Figure A7 in the Supplementary Material, by mean of another categorical stem plot. Since the calibration and validation sets were made by almost the same number of samples, and samples of different products ordered similarly in the calibration / validation sets, Figure A7 looks very like to Figure 4, as also the calibration and validation performances were totally comparable. All the observations of #2 were correctly recognized as in compliance with the reference, as well as the big majority of the defective samples. Just few samples of #3c were wrongly identified as in compliance. However, the error of validation was still low (6 samples over 303, corresponding to about 98% accuracy). Nevertheless, it is relevant to notice that also the PLS-DA was able to distinguish all the #3a samples, proving to be reliable for discriminating samples with a residual moisture higher than the target set as a CQA.

It has to be noticed that all the samples misclassified by the PLS-DA were samples of group #3b, pointing out a sort of systematic error. Summing the number of errors both in the

calibration and in the validation step, 11 #3b samples out of 13 were misclassified, suggesting the PLS-DA was not very effective in distinguish the sucrose from the trehalose at that specific solid fraction and value of residual moisture. This may be due to the lower number of samples belonging to this dataset considered in the model development stage: it has to be taken into account that the subset #3b is a particularly challenging case study, as a totally different product is present in the vial. Actually, this case is not expected to occur in a real case study, in the sense that the model will be calibrated using data acquired in a validation run and, then, it will be used in production, trying to identify cases where the residual moisture is not meeting the target (like in set #3a), or when some impurity could be present (set in #3c). The case #3b was included in this study to challenge the model also in a very different situation that, in any case, is not expected to occur. The method based on multivariate control charts did not had to cope with this issue, since samples of #3b were all correctly classified. Though the control charts method and the PLS-DA performed similarly in terms of calibration and validation errors, the errors of the former looked like being related to the statistic and to the distributions, i.e. the misclassifications in the control charts had a random nature and might be resolved by adding a larger number of samples to the statistics. On the contrary, the PLS-DA seemed to have some intrinsic limitations affecting its performances in relation to the trehalose samples. A possible explanation could be the similarity between #1 and #3b spectra. In Figure A2, #3b already appeared much similar to #1 and #2. A further investigation was carried out by comparing #1 and some ad-hoc spectral sets, made by joining #1 and one of the non-compliant set of group #3 at a time, i.e. #1 and #3a, #1 and #3b, #1 and #3c. These sets were processed by a PLS with four LVs, as done for the development of the PLS-DA model. In Figure A8 in the Supplementary Material, the loadings of the 2nd LV, which was found to be particularly explanatory, are compared. The loading of #1 is clearly different from the loading describing the sets made by #1 and #3a and by #1 and #3c, meaning that the contribution introduced by #3a or #3c is significant enough to make the loadings change. On the contrary, the loading of the set made by #1 and #3b is much similar to the loading of #1. Hence, it seems that #3b, when joined to #1, is described similarly to the reference set, which, finally, may lead to misclassification issues.

An additional validation of the PLS-DA was carried out through a permutation test. The goodness of the fit in calibration and the ability of the prediction of the unpermuted model were estimated through the parameters R2 and Q2, which actually reached high values (0.934 and 0.888 respectively). On the other hand, the values of R2 and Q2 calculated for each permuted model were much lower (close to zero or even lower). Therefore, the PLS-DA model could be

considered not spurious, and the outcome of the permutation test can be seen in the permutation plot in Figure 5.

The classification carried out by the PLS-DA was also investigated in a score plot, which graphically resumed the information given by the categorical stem plots. All the four PCs used for the PLS-DA were able to perform a clear clustering of the samples, and two main clusters could always be found: one for the reference and the samples in compliance, the other for the rejected samples. In Figure 6 all the combinations with the 1st component are displayed as an example. Not only the first PCs gave a good clustering (which is something expected, since the 1st and 2nd PCs explained the majority of the variance), but also the 3rd and 4th PCs were proved to be relevant, with particularly clear clustering (Figure 7, upper right and lower left corners). However, in the plane described by the 2nd and 3rd PC, some overlapping turned up, as showed in Figure 6 in the lower right corner. In particular, the #3b scores were not grouped in a cluster, but appeared spread and partially mixed with the #1 and #2 scores. The overlapping of the #3b scores is representative of the misclassification of the #3b samples highlighted in the stem plots. Finally, it is important to notice that the clusters made by the validation samples were coincident with the clusters made by the decomposition of the calibration set. This was indicative of the goodness of the PLS-DA, being able to perform a decomposition of the reference samples in such a space where the projections of the new samples could be ordered with consistency.

5. Conclusions

In this paper an at-line application of NIRS was described, aimed at performing a qualitative SQC of freeze-dried samples. Two statistical methods were developed and tested. The first was based on multivariate control charts, obtained from the Hotelling T2 and the SPE statistic, calculated by means of a previous PCA; the second was a PLS-DA, which is a categorical binary version of the PLS. The two models were calibrated with a reference set, and the goal was classifying some new sample as in compliance or not with the reference.

Overall, both approaches turned out to have excellent performances, with very little errors, thus being able to discriminate samples with accuracy. In particular, according to the application pursued, i.e. to the characteristic of the samples that is looked at to perform the classification, a method is preferable rather than the other. In order to identify samples out-of-specification because of some issues in their composition (#3b and #3c) the method based on control charts is more reliable. The PLS-DA is better to confirm samples as in compliance (#2), since it

performed a perfect discrimination of all the samples in compliance with the reference. Both method, instead, were perfectly suitable to detect samples non in compliance because of the higher residual moisture content (#3a), which is a crucial attribute for freeze-dried products. Moreover, since these are statistical approaches, the performances of the algorithms presented can be further improved and consolidate by training the models with new sample datasets.
Additionally, NIRS could be implemented also in-line, as a PAT tool for the freeze-drying process. Thus, further developments in the direction of a NIRS application for in-line statistical process control could be done in following researches.

Acknowledgment

The contribution and financial support of Merck Serono SpA is gratefully acknowledged. The author also would like to thank Daniele Mari, Senior Laboratory Technician, Biotech Pharmaceutical Development Department, Merck Serono SpA, Guidonia Montecelio, Roma, for the precious help in the experimental activities.

References

1. R. Pisano, D. Fissore, Heat transfer in freeze-drying apparatus, in: M.A. Dos Santos Bernardes (Ed.), Developments in Heat Transfer, Rijeka, 2011, pp. 91–114.

2. D. Fissore, T. McCoy, Editorial: freeze-drying and process analytical technology for pharmaceuticals, Front. Chem., 6 (2018), pp. 1-2. https://doi.org/10.3389/fchem.2018.00622

3. D. Fissore, Freeze-drying of pharmaceuticals, in: J. Swarbrick (Ed.) Encyclopedia of Pharmaceutical Science and Technology. 4th ed, London, 2013, pp. 1723-1737

4. T. De Beer, A. Burggraeve, M. Fonteyne, S. Saerens, J.P. Remon, C. Vervaet, Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes, Int. J. Pharm., 417 (2011), pp. 32-47. . https://doi.org/10.1016/j.ijpharm.2010.12.012

5. V.R. Koganti, E.Y. Shalaev, M.R. Berry, T. Osterberg, M. Youssef, D.N. Hiebert, F. A. Kanka, M. Nolan, R. Barrett, G. Scalzo, G. Fitzpatrick, N. Fitzgibbon, S. Luthra, L. Zhang, Investigation of design space for freeze-drying: use of modeling for primary drying segment of a freeze-drying cycle, AAPS PharmSciTech, 12 (2011), pp. 854-861. https://doi.org/10.1208/s12249-011-9645-7

6. D. Fissore, A.A. Barresi, Scale-up and process transfer of freeze-drying recipes, Dry Technol., 29 (2011), pp. 1673–1684. https://doi.org/10.1080/07373937.2011.597059

7. A. A. Barresi, P. Pisano, V. Rastto, D. Fissore, D. L. Marchisio, Model-based monitoring and control of industrial freeze-drying processes: effect of batch nonuniformity, Dry Technol., 28 (2010), pp. 577-590. https://doi.org/ 10.1080/07373931003787934

8. S. Rambhatla, S. Tchessalov, M. J. Pikal, Heat and mass transfer scale-up issues during freeze-drying, III: control and characterization of dryer differences via operational qualification tests, AAPS PharmSciTech, 7 (2006), E1-E10. https://doi.org/ 10.1208/pt070239

9. S. Rambhatla, M. J. Pikal, Heat and mass transfer scale-up issues during freeze-drying, I: atypical radiation and the edge vial effect, AAPS PharmSciTech, 4 (2003), pp. 1-10. https://doi.org/ 10.1208/pt040214

10. J. Luypaert, D. L. Massart, Y. Vander Heyden, Near-infrared spectroscopy applications in pharmaceutical analysis, Talanta, 72 (2007), pp. 865-883. https://doi.org/10.1016/j.talanta.2006.12.023

11. E. W. Ciurczak, J. K. Drennen, Pharmaceutical and Medical Applications of Near-infrared Spectroscopy, New York, 2002.

12. M. Blanco, J. Coello, H. Iturriga, S. Maspoch, C. de la Pezuela, Near-infrared spectroscopy in the pharmaceutical industry, Analyst, 123 (1998), pp. 135R-150R. https://doi.org/10.1039/A802531B

13. B. G. Osborne, Near-infrared Spectroscopy in Food Analysis, in: R. A. Meyer (Ed.), Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation, 2006, pp. 1-14.

14. C. Pasquini, Near infrared spectroscopy: a mature analytical technique with new perspectives – a review, Anal. Chim. Acta, 1026 (2018), pp. 8-36. https://doi.org/ 10.1016/j.aca.2018.04.004

15. T. De Beer, P. Vercruysse, A. Burggraeve, T. Quinten, J. Ouyang, X. Zhang, C. Vervaet, J. P. Remon, W. R. G. Baeyens, In-line and real-time process monitoring of a freeze drying process using

Raman and NIR spectroscopy as complementary Process Analytical Technology (PAT) tools, J Pharm Sci, 98 (2009), pp. 3430-3446. https://doi.org/10.1002/jps.21633

16. J. C. May, Regulatory Control of Freeze-Dried Products: Importance and Evaluation of Residual Moisture, in: L. Rey, J. C. May (Eds.), Freeze-Drying/lyophilization of Pharmaceutical and Biological Products Second Edition, Revised and Expanded, New York, 2004.

17. S. Bai, R. Nayar, J. F. Carpenter, M. C. Manning, Noninvasive determination of protein conformation in the solid state using Near Infrared (NIR) spectroscopy, J. Pharm. Sci., 94 (2005), pp. 2030-2038. https://doi.org/ 10.1002/jps.20416

18. D. S. Katayama, J. F. Carpenter, K. P. Menard, T. W. Randolph, Mixing properties of lyophilized protein systems: a spectroscopic and calorimetric study, J. Pharm. Sci., 98 (2009), pp. 2954-2969. https://doi.org/ 10.1002/jps.21467

19. S. Pieters, T. De Beer, J. C. Kasper, D. Boulpaep, O. Waszkiewicz, M. Goodarzi, C. Tistaert, W. Friess, J. Remon, C. Vervaet, Y. Vander Heyden, Near-Infrared spectroscopy for in-line monitoring of protein unfolding and its interactions with lyoprotectants during freeze-drying, Anal. Chem., 84 (2012), pp. 947-955. https://doi.org/10.1021/ac2022184

20. G. Reich, Near-infrared spectroscopy and imaging: basic principles and pharmaceutical applications, Adv. Drug. Deliver Rev., 57 (2005), pp. 1109-1143. https://doi.org/10.1016/j.addr.2005.01.020

21. J. A. Jones, I. R. Last, B. F. Macdonald, K. A. Prebblet, Development and transferability of near-infrared methods for determination of moisture in a freeze-dried injection product, J Pharmaceut Biomed, 11 (1993), pp. 1227-1231. https://doi.org/10.1016/0731-7085(93)80108-D

22. Y. Zheng, X. Lai, S. W. Bruun, H. Ipsen, J. N. Larsen, H. Løwenstein, I. Søndergaard, S. Jacobsen, Determination of moisture content of lyophilized allergen vaccines by NIR spectroscopy, J. Pharmaceut. Biomed.,46 (2008), pp. 592-596. https://doi.org/

23. Y. Li, Q. Fan, S. Liu, L. Wang, Simultaneous analysis of moisture, active component and cake structure of lyophilized powder for injection with diffuse reflectance FT-NIR chemometrics, J Pharmaceut Biomed, 55 (2011), pp. 216-219. https://doi.org/10.1016/j.jpba.2010.12.028

24. T. De Beer, M. Wiggenhorn, R. Veillon, C. Debacq, Y. Mayeresse, B. Moreau, A.Burggraeve, T. Quinten, W. Friess, G. Winter, C. Vervaet, J. P. Remon, W. R. G. Baeyens, Importance of using complementary process analyzers for the process monitoring, analysis, and understanding of freeze drying, Anal. Chem., 81 (2009), pp. 7639-7649. https://doi.org/10.1021/ac9010414

25. I. R. Last, K. A. Prebble, Suitability of near-infrared methods for the determination of moisture in a freeze-dried injection product containing different amounts of the active ingredient, J. Pharmaceut. Biomed., 11 (1993), pp. 1071-1076. https://doi.org/10.1016/0731-7085(93)80084-E

26. M. W. J. Derksen, P. J. M. van der Oetelaarb, F. A. Maris, The use of near-infrared spectroscopy in the efficient prediction of a specification for the residual moisture content of a freeze-dried product, J. Pharmaceut. Biomed., 17 (1998), pp. 473-480. https://doi.org/10.1016/S0731-7085(97)00216-1

27. M. Clavaud, Y. Roggo, K. Degardin, P.-Y. Sacre, P. Hubert, E. Ziemons, Moisture content determination in an antibody-drug conjugate freeze-dried medicine by near-infrared spectroscopy: a

case study for release testing, J Pharm Biomed Anal, 131 (2016), pp. 380-390. https://doi.org/10.1016/j.jpba.2016.09.014

28. J. Aaltonen, K. C. Gordon, C. J. Strachan, T. Rades, Perspectives in the use of spectroscopy to characterise pharmaceutical solids, Int. J. Pharmaceut., 364 (2008), pp. 159-169. https://doi.org/10.1016/j.ijpharm.2008.04.043

29. J. F. Carpenter, J. H. Crowe, An infrared spectroscopic study of the interactions of carbohydrates with dried proteins, Biochemistry, 28 (1989), pp. 3916-3922. https://doi.org/10.1021/bi00435a044

30. H. Grohganz, M. Fonteyne, E. Skibsted, T. Falck, B. Palmqvist, J. Rantanen, Role of excipients in the quantification of water in lyophilized mixtures using NIR spectroscopy, J. Pharmaceut. Biomed., 49 (2009), pp. 901–907. https://doi.org/10.1016/j.jpba.2009.01.021

31. M. A. Mensink, P. Van Bockstal, S. Pieters, L. De Myer, H. V. Frijlink, K. van der Voort Maarschalk, W. L. J. Hinrichs, T. De Beer, In-line near infrared spectroscopy during freeze-drying as a tool to measure efficiency of hydrogen bond formation between protein and sugar, predictive of protein storage stability, Int. J. Pharmaceut., 496 (2015), pp. 792-800. https://doi.org/10.1016/j.ijpharm.2015.11.030

32. Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jent, A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies, J. Pharm. Biomed. Anal., 44 (2007), pp. 683-700. https://doi.org /10.1016/j.jpba.2007.03.023

33. G. X. Zhou, G. E. Zhihong, J. Dorwart, B. Izzo, J. Kukura, G. Bicker, J. Wyvratt, Determination and differentiation of surface and bound water in drug substances by Near Infrared spectroscopy, J. Pharm. Sci., 92 (2003), pp. 1058-1065. https://doi.org/10.1002/jps.10375

34. L. Hansen, T. De Beer, S. Pieters, Y. Vander Heyden, C. Vervaet, J. P. Remon, J. P. Montenez, R. Daoussi, Near-infrared spectroscopic evaluation of lyophilized viral vaccine formulations, Biotechnol. Prog., 29 (2013), pp. 1573-1586. https://doi.org/10.1002/btpr.1807

35. L. Hansen, J. Van Renterghem, R. Daoussi, C. Vervaet, J. P. Remon, T. De Beer, Spectroscopic evaluation of a freeze-dried vaccine during an accelerated stability study, Eur J Pharm Biopharm, 104 (2016), pp. 89-100. https://doi.org/10.1016/j.ejpb.2016.04.010.

36. M. Azheruddin, R. Gosselin, A. Cournoyer, Identifying collapse in freeze-dried products via NIR spectroscopy, Am. Pharm. Rev. (2021), available on-line at https://www.americanpharmaceuticalreview.com/Featured-Articles/575216-Identifying-Collapse-in-Freeze-Dried-Products-Via-NIR-Spectroscopy/ (accessed 16 July 2021).

37. Guideline Pharmaceutical Development Q8(R2). ICH Harmonised Tripartite 2009. https://database.ich.org/sites/default/files/Q8_R2_Guideline.pdf (accessed 4 January 2021).

38. D. Fissore, R. Pisano, Computer-aided framework for the design of freeze-drying cycles: optimization of the operating conditions of the primary drying stage, Processes, 3 (2015), pp. 406-421. https://doi.org/10.3390/pr3020406

39. M. Stuart, E. Mullins, E. Drew, Statistical quality control and improvement, Eur. J. Oper. Res., 88 (1996), pp. 203-214. https://doi.org/10.1016/0377-2217(95)00069-0

40. M. Clavaud, Y. Roggo, K. Degardin, P. Y. Sacre, P. Hubert, E. Ziemons, Global regression model for moisture content determination using near-infrared spectroscopy, Eur. J. Pharm. Biopharm., 119 (2017), pp. 343-352. https://doi.org/10.1016/j.ejpb.2017.07.007

41. C. R. Muzzio, N. G. Dini, L. D. Simionato, Determination of moisture content in lyophilized mannitol through intact glass vials using NIR micro-spectrometers, Braz. J. Pharm. Sci., 47 (2011), pp. 289-297. https://doi.org/10.1590/S1984-82502011000200010

42. Shlens J. A Tutorial on Principal Component Analysis, 2014. https://arxiv.org/pdf/1404.1100.pdf (accessed 12 November 2020).

43. S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemometr. Intell. Lab., 58 (2001), pp. 109-130. https://doi.org/10.1016/S0169-7439(01)00155-1

44. H. Hotelling, Analysis of a complex of statistical variables into principal components, J. Educ. Psychol., 24 (1993), pp. 417-441. https://doi.org/10.1037/h0071325

45. P. Nomikos, J. M. MacGregor, Monitoring batch processes using multiway principal component analysis, Technometrics, 40 (1994), pp. 1361-1375. https://doi.org/10.1002/aic.690400809

46. J. F. McGregor, T. Kourti, Statistical process control of multivariate processes, Control Eng. Pract., 3 (1995), pp. 403-414. https://doi.org/10.1016/0967-0661(95)00014-L

47. D. Colucci, J. M. Prats-Montalban, D. Fissore, A. Ferrer, Application of multivariate image analysis for on-line monitoring of a freeze-drying process for pharmaceutical products in vials, Chemometr Intell Lab, 187 (2019), pp. 19-27. https://doi.org/10.1016/j.chemolab.2019.02.004

48. D. Colucci, Infrared Imaging: a New Process Analytical Technology for Real Time Monitoring and Control of a Freeze-Drying Process, Doctoral Dissertation (2019)

49. S. A. Vander Wiel, W. T. Tucker, F. W. Faltin, N. Doganaksoy, Algorithmic statistical process control: concepts and an application, Technometrics, 34 (1992), pp. 286-297. https://doi.org/10.1007/978-1-4612-2856-1_35

50. L. Eriksson, E. Johansson, J. Trygg, C. Vikstrom, Multi- and Megavariate Data Analysis. Basic Principles and Applications. Third revised edition, Malmo, 2013.

51. P. Nomikos, J. F. MacGregor, Multivariate SPC charts for monitoring batch processes, Technometrics, 37 (1995) . https://doi.org/10.2307/1269152

52. M. Perez-Enciso, M. Tenenhaus, Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach, Hum. Genet., 112 (2003), pp. 581-592. https://doi.org/10.1007/s00439-003-0921-9

53. L. Eriksson, J. Jaworska, A. P. Worth, M. T. D. Cronin, R. M. McDowell, P. Gramatica, Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs, Environ. Health. Perspect., 111 (2003), pp. 1361–1375. https://doi.org/10.1289/ehp.5758

54. R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C. P. Varghese, R. K. Agrawal, Validation of QSAR models - strategies and importance, Int. J. Drug. Des. Discovery, 2 (2011), pp. 511-519

55. S, Bobba, N. Zinfollino, D. Fissore, On the development of a robust model for the determination of the residual moisture in freeze-dried products using Near-Infrared Spectroscopy [Manuscript submitted for publication].

56. S. Wold, Nonlinear partial least square modeling II. Spline inner relation, Chemometr. Intell. Lab. 14(1992), pp. 71-84, https://doi.org/10.1016/0169-7439(92)80093-J.

57. G. Reich, Near-infrared spectroscopy and imaging: basic principles and pharmaceutical applications, Adv Drug Deliver Rev, 57 (2005), pp. 1109-1143. https://doi.org/10.1016/j.addr.2005.01.020

58. A. Rinnan, F. Van der Berg, S. B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, Trends Anal. Chem., 28 (2009), pp. 1201-1222. https://doi.org/10.1016/j.trac.2009.07.007.

59. C. Tan, J. Wang, T. Wu, X. Qin, M. Li, Determination of nicotine in tobacco samples by near-infrared spectroscopy and boosting partial least squares, Vib. Spectrosc., 54 (2010), pp. 35-41, https://doi.org/10.1016/j.vibspec.2010.06.002.

60. C.R. Muzzio, N.G. Dini, L.D. Simionato, Determination of moisture content in lyophilized mannitol through intact glass vials using NIR micro-spectrometers, Brazilian J. Pharm. Sci. 47 (2011), pp. 289-297, https://doi.org/10.1590/S1984-82502011000200010.

61. R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C.P. Varghese, R.K. Agrawal, Validation of QSAR models - strategies and importance, Int. J. Drug Discov. 2 (2011), pp. 511-519, https://www.researchgate.net/publication/284566093_Validation_of_QSAR_Models_-_Strategies_and_Importance (accessed 27 January 2021).

List of Tables

Table 1. Number of samples of each set (calibration and validation) used in this study.

List of Figures

Figure 1. Calibration step for building the control charts, calculated with the 245 spectra (on the horizontal axes) of the #1 sample set. T2 and SPE values are reported on the vertical axes. The CLs (solid horizontal lines) were calculated from the T2 distribution (upper charts) and from the SPE distribution (lower charts). The CLs calculated at 0.95 of confidence interval α are shown on the left, the CLs at α 0.99 are on the right.

Figure 2. Validation of the PCDA with the #2 dataset (245 samples axis). The CLs at 0.95 of confidence interval (dashed line) and at 0.99 (solid line) are shown, as calculated in the calibration step. The Hotelling T^2 statistic is presented above, the SPE statistic below. Black bars correspond to outlier observations.

Figure 3. Validation of the control charts method with dataset #3. Samples 1 – 69 (light grey bars) refer to #3a dataset, 70 - 96 (dark grey bars) to dataset #3b, and 97 - 113 (white bars) to dataset #3c. The values of T2 (above) and of SPE (below) of each sample are compared with the respective CLs calculated at α 0.95 (dashed line) and 0.99 (solid line) of confidence interval.

Figure 4. Categorical stem plot obtained through the application of the PLS-DA to the calibration set. Samples 1 - 245 refer to #1 dataset, samples 246 - 279 to #3a dataset, samples 280 – 292 to #3b dataset, and samples 293 - 300 to #3c dataset. The samples corresponding to a stem pointing upward are classified as in compliance, those represented by a stem pointing downward are identified as outliers (indicated on the vertical axis as COMPL and OUT, respectively). The observations from #1 are marked by a circle, those from the dataset #3 by a diamond.

Figure 5. Permutation plot showing the values of R2 (triangles) and Q2 (circles) calculated for each permutation as a function of the correlation coefficient (r) between the categorical attributes matrix (**Y**), and the **Y** permuted. Grey symbols refer to the values of R2 and Q2 of the original model.

Figure 6. Score plots obtained by the PLS-DA over the 7700 - 4590 cm^{-1} range. The calibration set is represented by gray diamonds (samples from #1) and grey squares (dataset #3); the validation set is represented by white marks: diamonds represent the #2 samples, triangles the #3a samples, circles stand for #3b samples, and squares for the #3c ones.

Table 1. Number of samples of each set (calibration and validation) used in this study.

Sample set	Control charts		PLSDA	
	Calibration set	**Validation set**	**Calibration set**	**Validation set**
#1	245	-	245	-
#2	-	245	-	245
#3a	-	79	34	35
#3b	-	27	13	14
#3c	-	17	8	9
Total samples	245	368	300	303

Figure 1. Calibration step for building the control charts, calculated with the 245 spectra (on the horizontal axes) of the #1 sample set. T2 and SPE values are reported on the vertical axes. The CLs (solid horizontal lines) were calculated from the T2 distribution (upper charts) and from the SPE distribution (lower charts). The CLs calculated at 0.95 of confidence interval α are shown on the left, the CLs at α 0.99 are on the right.

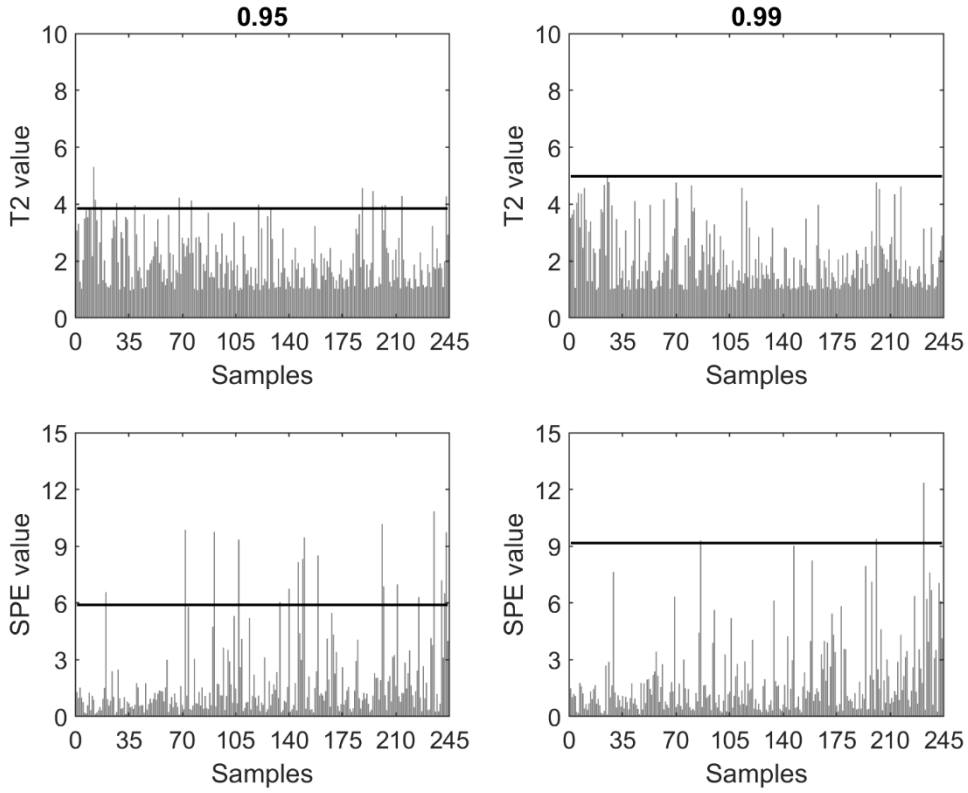

Figure 2. Validation of the PCDA with the #2 dataset (245 samples axis). The CLs at 0.95 of confidence interval (dashed line) and at 0.99 (solid line) are shown, as calculated in the calibration step. The Hotelling T^2 statistic is presented above, the SPE statistic below. Black bars correspond to outlier observations.

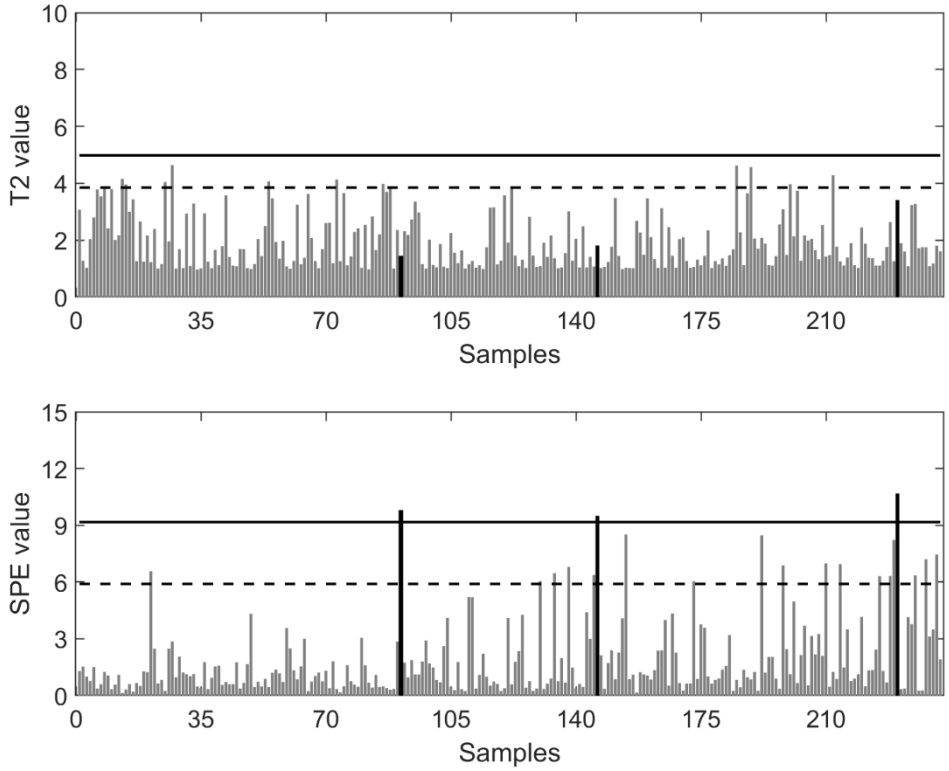

Figure 3. Validation of the control charts method with dataset #3. Samples 1 – 69 (light grey bars) refer to #3a dataset, 70 - 96 (dark grey bars) to dataset #3b, and 97 - 113 (white bars) to dataset #3c. The values of T2 (above) and of SPE (below) of each sample are compared with the respective CLs calculated at α 0.95 (dashed line) and 0.99 (solid line) of confidence interval.

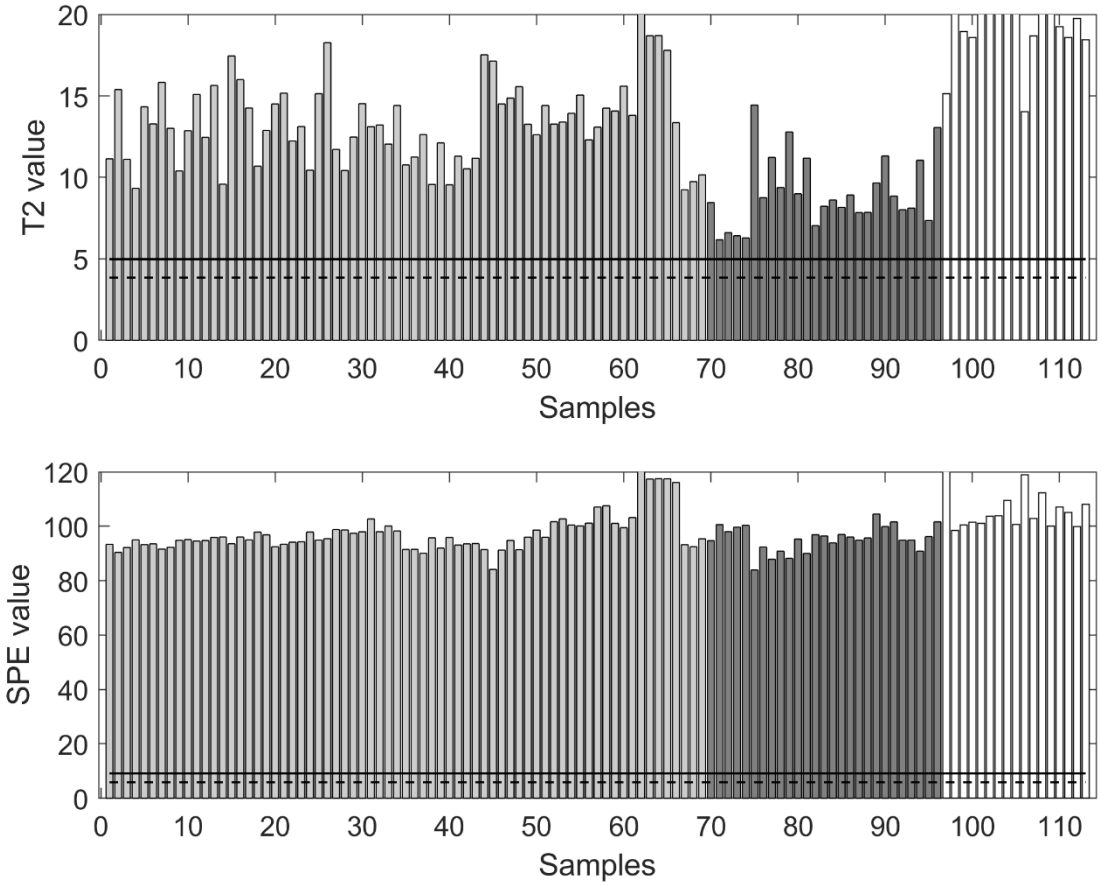

Figure 4. Categorical stem plot obtained through the application of the PLS-DA to the calibration set. Samples 1 - 245 refer to #1 dataset, samples 246 - 279 to #3a dataset, samples 280 – 292 to #3b dataset, and samples 293 - 300 to #3c dataset. The samples corresponding to a stem pointing upward are classified as in compliance, those represented by a stem pointing downward are identified as outliers (indicated on the vertical axis as COMPL and OUT, respectively). The observations from #1 are marked by a circle, those from the dataset #3 by a diamond.

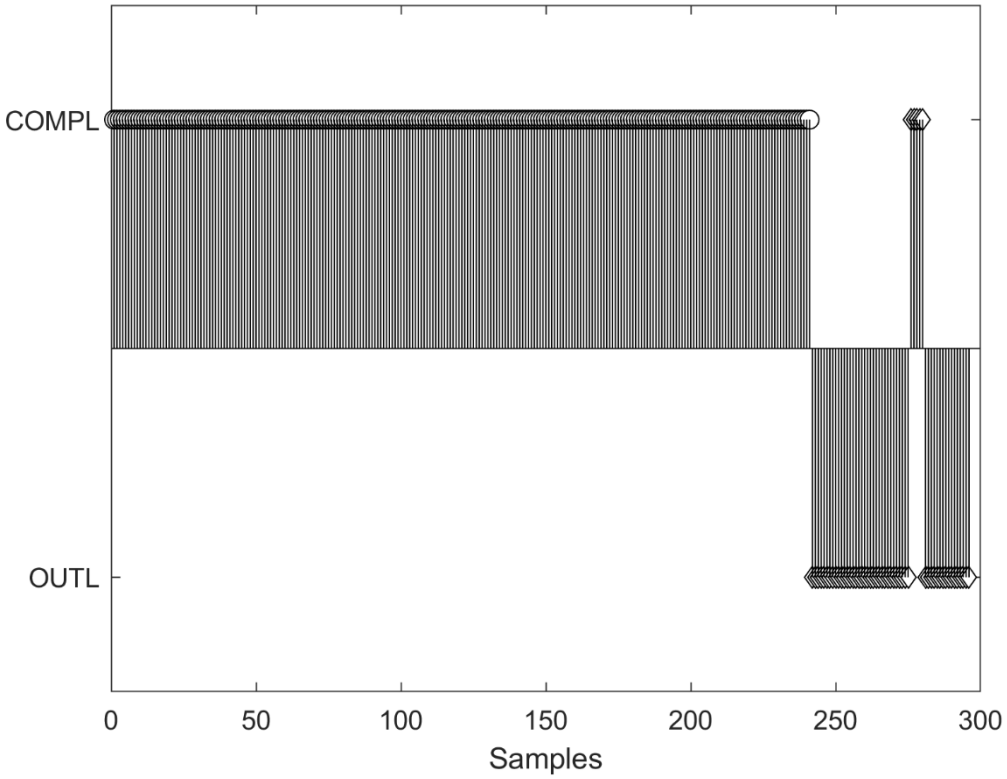

Figure 5. Permutation plot showing the values of R2 (triangles) and Q2 (circles) calculated for each permutation as a function of the correlation coefficient (*r*) between the categorical attributes matrix (**Y**), and the **Y** permuted. Grey symbols refer to the values of R2 and Q2 of the original model.

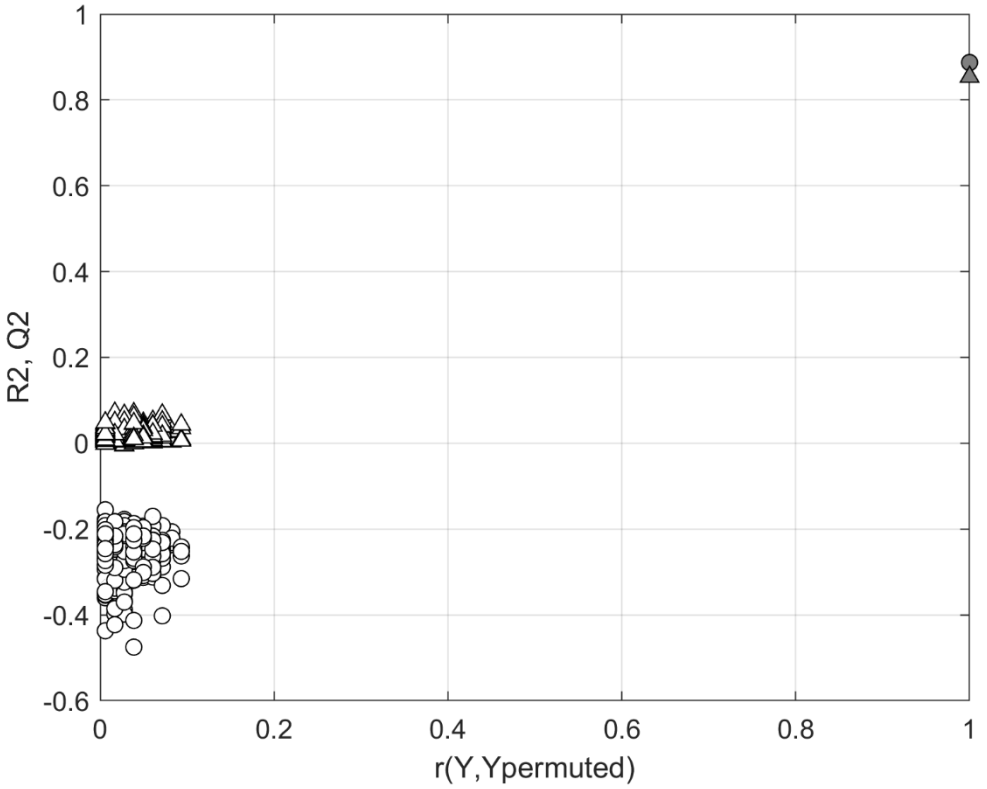

Figure 6. Score plots obtained by the PLS-DA over the 7700 - 4590 cm^{-1} range. The calibration set is represented by gray diamonds (samples from #1) and grey squares (dataset #3); the validation set is represented by white marks: diamonds represent the #2 samples, triangles the #3a samples, circles stand for #3b samples, and squares for the #3c ones.

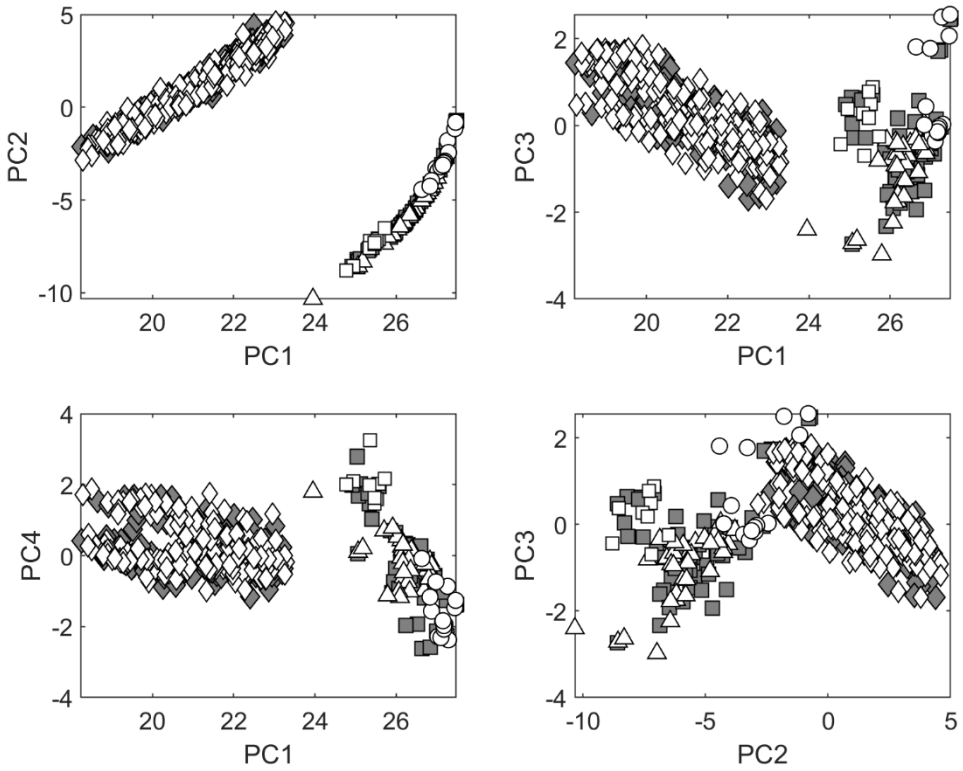

Figure A1. Experimental setup for spectra acquisition: NIR spectrometer probe is placed on the visual inspection station, while a sealed vial is being scanned.

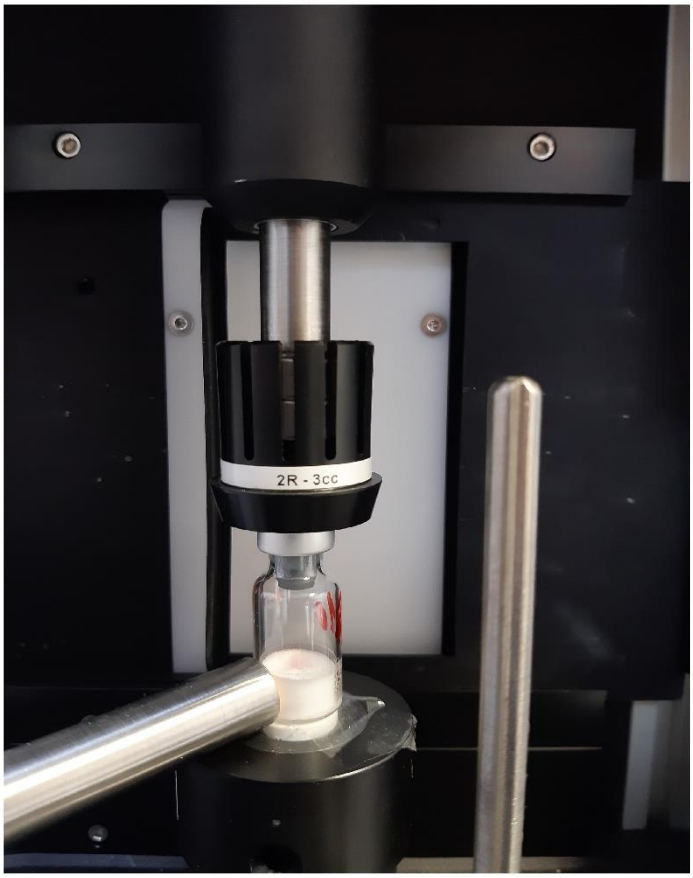

Figure A2. Spectral datasets (given as absorbance vs. frequency), after pre-processing, over the wavelength range 9500 - 4100 cm^{-1}: #1 and #2 (490 spectra), #3a (79 spectra), #3b (27 spectra), #3c (17 spectra).

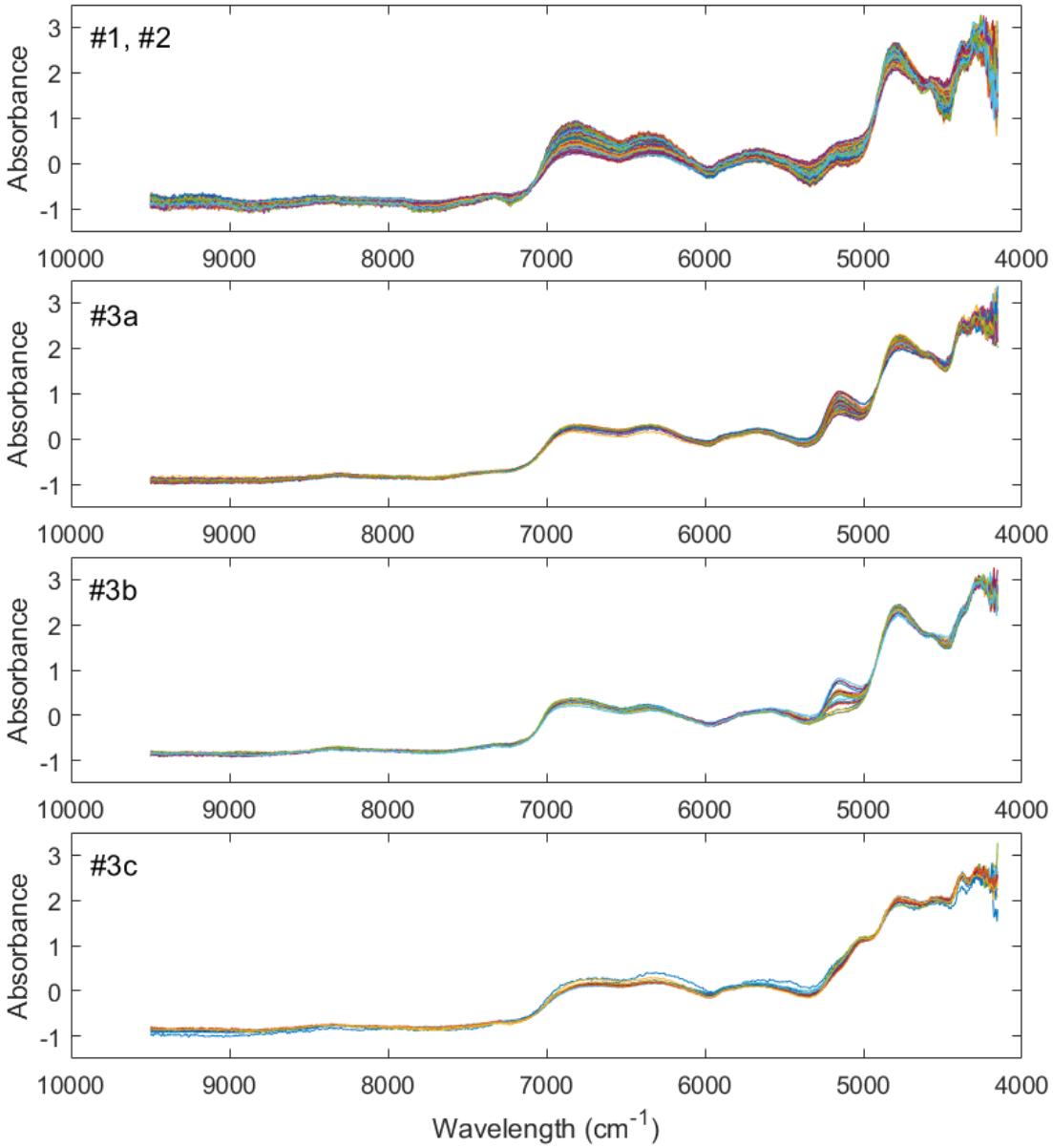

Figure A3. Loading plot of the first 2 PCs of the PCA performed on #1, over the wavelength ranges 9500 - 4100 cm^{-1} (on the left) and 7700 - 4590 cm^{-1} (on the right).

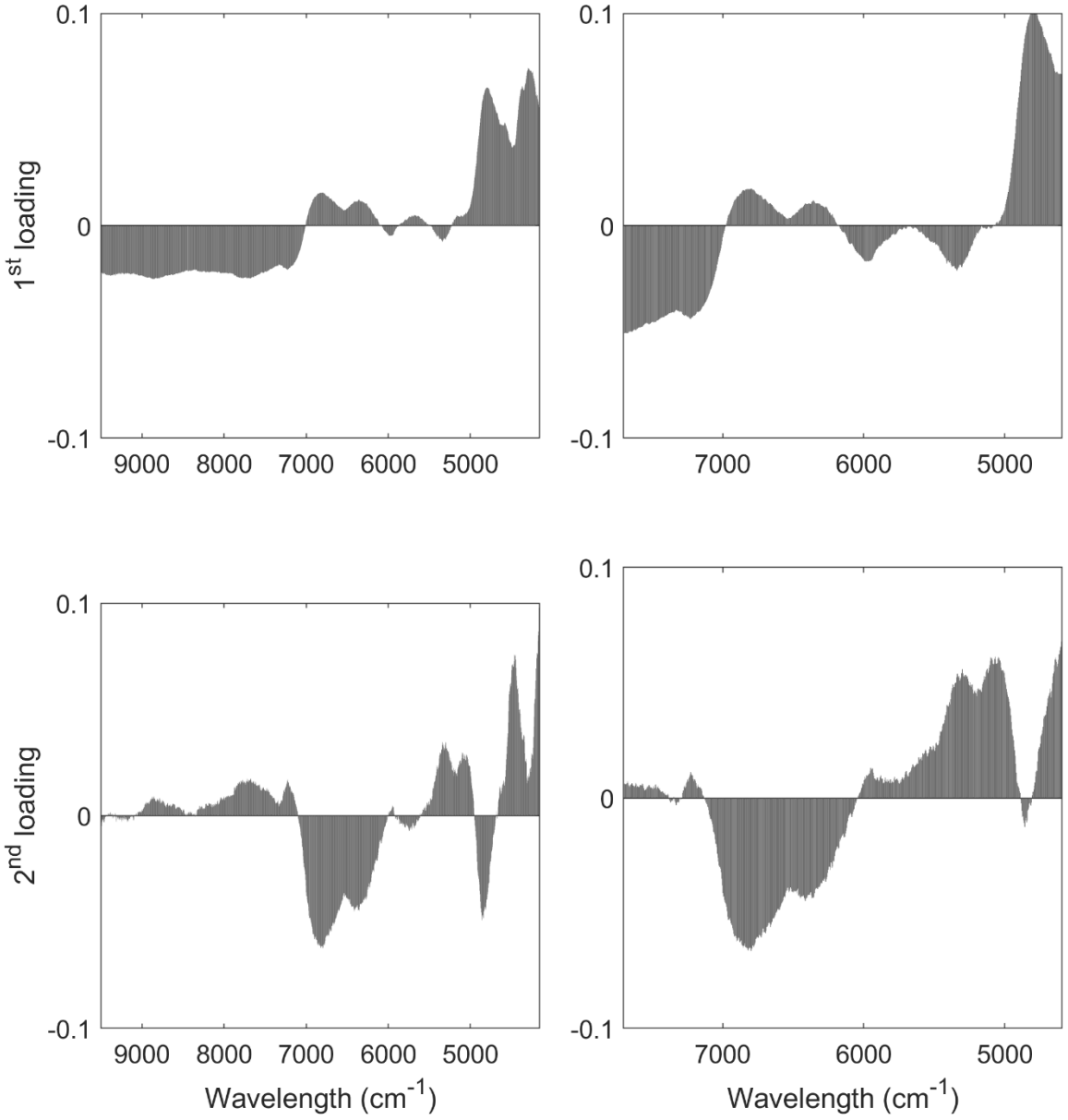

Figure A4. Score plot of the PCA analysis of all the spectral datasets considered in this study. Black triangles refer to #1 observations, while green circles to the #2 spectra. Blue, orange and black circles refer to the observations of datasets #3a, #3b and #3c respectively.

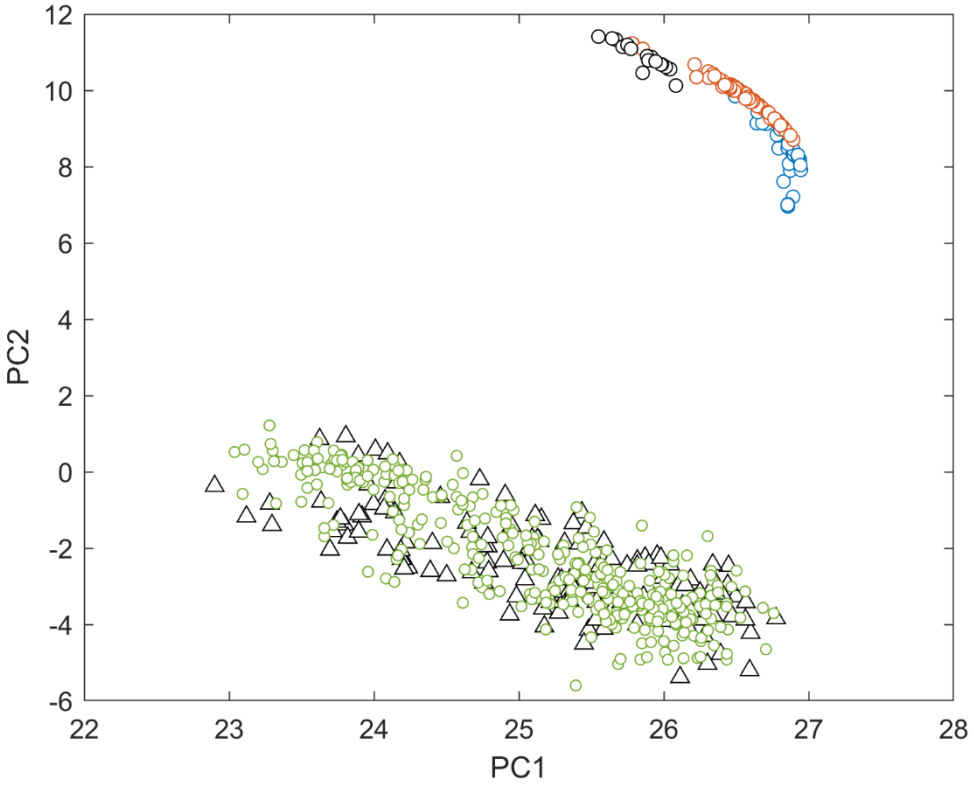

Figure A5. Influence of the number of LVs on the performance of the PLS-DA.

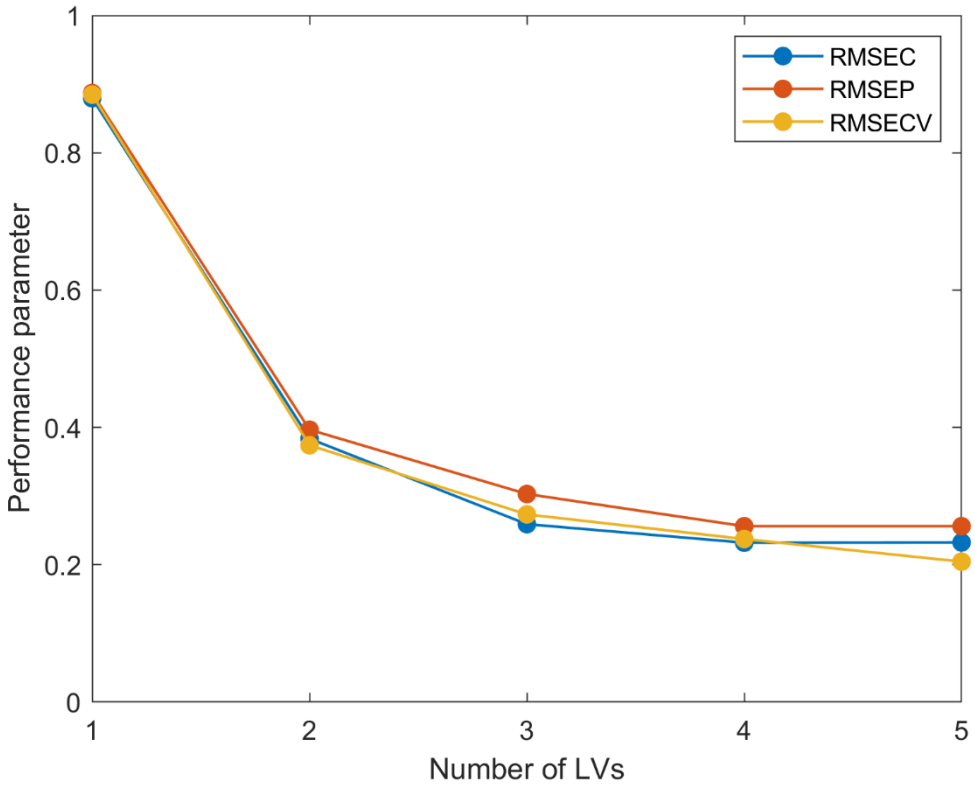

Figure A6. Loading plot of the first four PCs of the PLSDA, over the wavelength ranges 9500 - 4100 cm^{-1} (on the left) and 7700 - 4590 cm^{-1} (on the right).

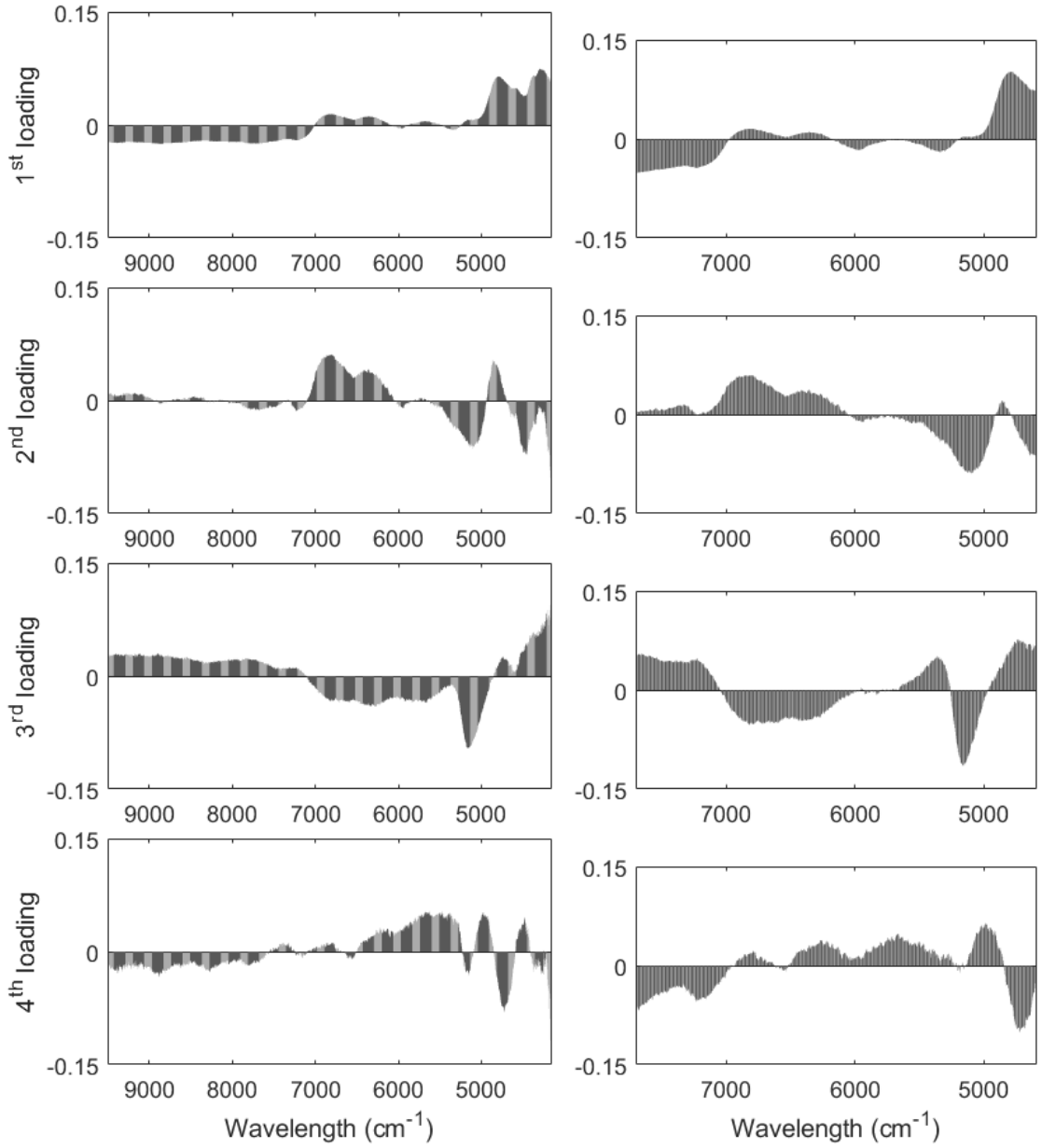

Figure A7. Application of the PLSDA to the validation set: samples are recognized as in compliance with the reference or as outliers (COMPL or OUT on the vertical axis, indicated by the stems). Samples 1-245 refer to #2 dataset, samples 246-280 to #3a dataset, samples 281-294 to #3b dataset, samples 295-303 to #3c dataset. The #2 samples are marked by circles, those from the dataset #3 by diamond.

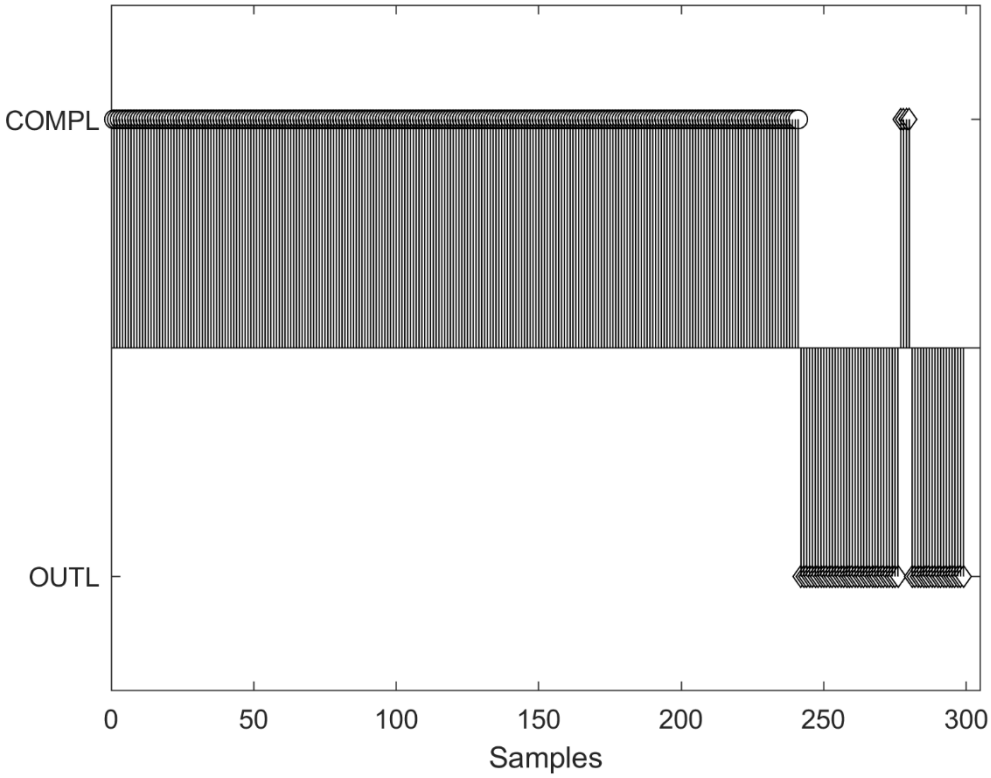

Figure A8. Loadings of the 2nd PC over the wavelength 7700 - 4590 cm^{-1} describing (in blue) the reference set made by #1 and (in orange) the ad-hoc sets made by: (a) #1 and #3a, (b) #1 and #3b, (c) #1 and #3c.

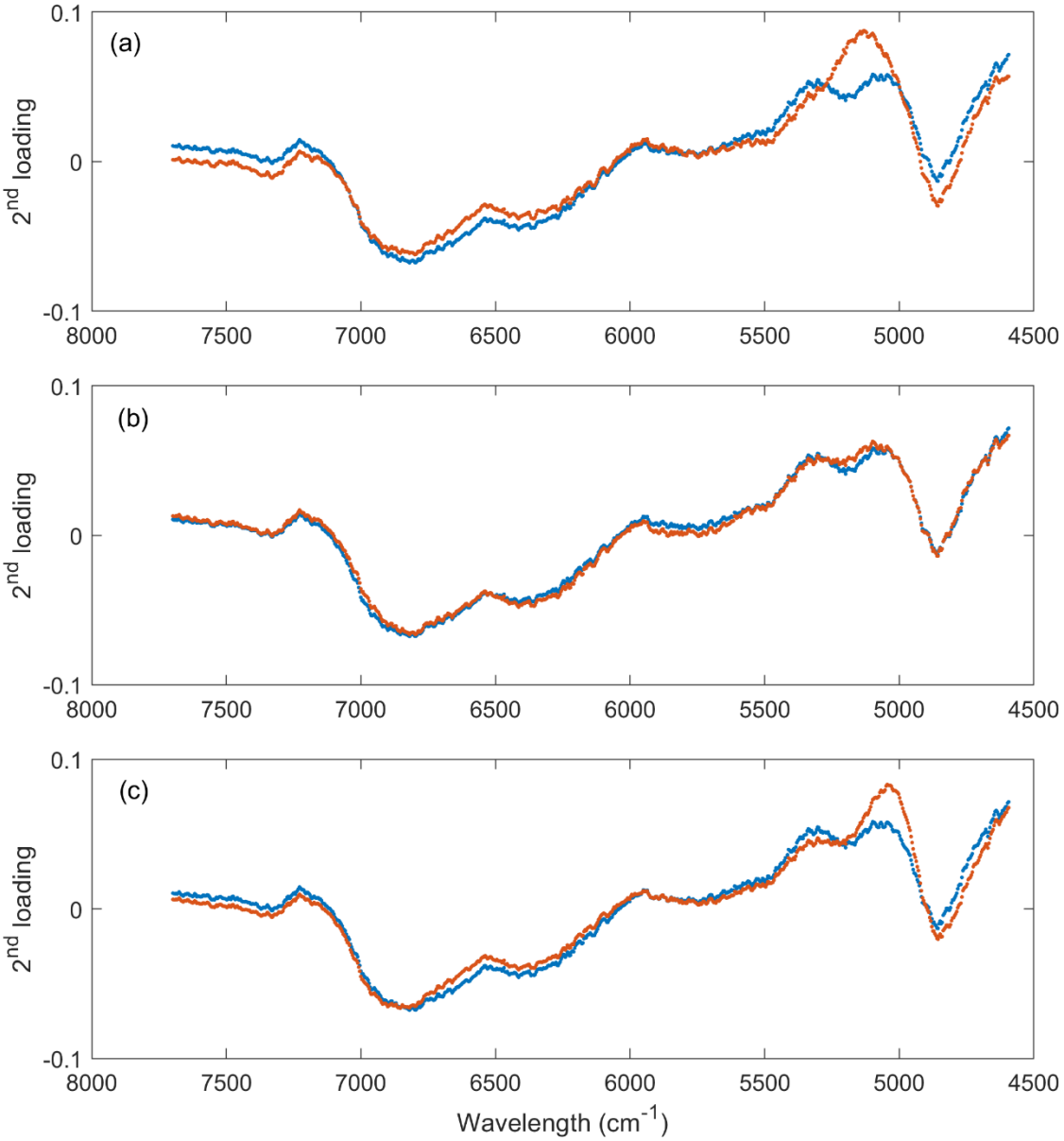

Table A1. Values of RM, measured by KF, of samples of datasets #1 and #2.

dataset #1		dataset #2	
Sample	**RM, %**	**Sample**	**RM, %**
1	0.64%	1	0.64%
2	0.82%	2	0.54%
3	0.52%	3	0.78%
4	0.59%	4	0.81%
5	0.68%	5	0.59%
6	0.77%	6	0.69%
7	0.75%	7	0.72%
8	0.70%	8	0.76%
9	0.51%	9	0.75%
10	0.74%	10	0.78%
11	0.77%	11	0.71%
12	0.63%	12	0.64%
13	0.77%	13	0.59%
14	0.74%	14	0.63%
15	0.88%	15	0.95%
16	0.77%	16	0.96%
17	0.62%	17	0.79%
18	0.67%	18	0.73%
19	0.79%	19	0.66%
20	0.67%	20	0.68%
21	0.82%	21	0.80%
22	0.74%	22	0.69%
23	0.77%	23	0.82%
24	0.82%	24	0.77%
25	0.74%	25	0.78%

Table A2. Values of RM, measured by KF, of samples of datasets #3a, #3b and #3c.

dataset #3a		dataset #3b		dataset #3c	
Sample	**RM, %**	**Sample**	**RM, %**	**Sample**	**RM, %**
1	3.02%	1	0.54%	1	1.03%
2	2.75%	2	1.25%	2	0.41%
3	2.40%	3	0.54%	3	1.12%
4	3.02%	4	0.40%	4	0.53%
5	1.55%	5	0.75%	5	0.34%
6	4.60%				
7	2.06%				
8	5.46%				
9	2.00%				
10	2.93%				