POLITECNICO DI TORINO
Repository ISTITUZIONALE

Evaluation of the robustness of a novel NIR-based technique to measure the residual moisture in freeze-dried products

Original
Evaluation of the robustness of a novel NIR-based technique to measure the residual moisture in freeze-dried products / Bobba, Serena; Zinfollino, Nunzio; Fissore, Davide. - In: JOURNAL OF PHARMACEUTICAL SCIENCES. - ISSN 0022-3549. - STAMPA. - 111:5(2022), pp. 1437-1450. [10.1016/j.xphs.2021.10.015]

Availability:
This version is available at: 11583/2933178 since: 2022-04-19T15:07:32Z

Publisher:
Elsevier

Published
DOI:10.1016/j.xphs.2021.10.015

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
Elsevier postprint/Author's Accepted Manuscript

© 2022. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/.The final authenticated version is available online at:
http://dx.doi.org/10.1016/j.xphs.2021.10.015

(Article begins on next page)

08 May 2026

Journal of Pharmaceutical Sciences
https://www.jpharmsci.org/article/S0022-3549(21)00556-6/fulltext
https://doi.org/10.1016/j.xphs.2021.10.015

Evaluation of the robustness of a novel NIR-based technique to measure the residual moisture in freeze-dried products

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

(Bio)pharmaceutical products freeze-dried in vials must meet stringent quality specifications: among these, the residual moisture (RM) is crucial. The most common techniques adopted for measuring the RM are destructive, e.g. Karl Fisher titration, thus few samples from each batch are tested. Being a high intra-batch variability an intrinsic feature of batch freeze-drying, a high number of samples needs to be tested to get a representative measurement. Near-Infrared (NIR) spectroscopy was extensively applied in the past as a non-invasive method to quantify the RM. In this paper, an accurate Partial Least Square (PLS) model was developed and calibrated with a single product, focusing on a small but significative wavelength range of NIR spectra (model SR), characteristic of the water and not of the product. The salient feature of this approach is that the model SR appears to provide fairly accurate estimates with the same product but at a higher concentration, with other excipients and in presence of an amino acid at high concentration, without requiring any additional calibration with KF analysis, as in previous techniques; the irrelevance of the vial shape was also shown. This approach was compared to a simpler one, based on a single-variable linear regression, and to more complex one, using a wider wavelength range or calibrating the PLS model with several products. Model SR definitely ended up as the most accurate, and it appeared to have a great potential as a robust model, suitable also for products that were not involved in the calibration step.

Keywords

Freeze-drying, Near-Infrared Spectroscopy, Multivariate analysis, Partial Least Square, Mathematical modeling, Robust model, Residual moisture, Karl Fischer titration.

List of abbreviations

CQA	Critical Quality Attribute
CV	Cross Validation
DP	Drug Product
KF	Karl Fischer
MVA	Multivariate Analysis
NIR	Near-Infrared
NIRS	Near-Infrared Spectroscopy
PC	Principal Component
PCA	Principal Component Analysis
PLS	Partial Least Square
RM	Residual Moisture
RMSEC	Root Mean Square Error of Calibration
RMSECV	Root Mean Square Error of Cross Validation
RMSEP	Root Mean Square Error of Prediction
SNV	Standard Normal Variate
SOP	Standard Operating Procedure

List of symbols

A	number of latent variables
$\mathbf{B}$	regression matrix of PLS
I	number of quality attributes
J	number of wavelengths
M	number of samples
Q2	model predictivity parameter
R2	fraction of the total sum of squares explained
$\mathbf{W}^*$	matrix of transformed weights (J x A)
$\mathbf{X}$	spectral data matrix (M x J)
$\mathbf{Y}$	matrix of quality attributes (M x I)

1. Introduction

In the last decade the biotech market strongly grew up and, despite the challenges the biotech products pose, like high development and quality control costs, they are expected to get 50% of shares in pharma companies by 2024[1]. Frequently biopharmaceutical drugs are prepared in liquid form and they need to be stabilized, which is done by converting them into solid products. This is done by means of low-temperature drying processes, that fit the needs of these products, typically very sensitive to heat stress[2,3]. Freeze-drying turns out as a fundamental downstream process in the manufacturing of (bio)pharmaceuticals: in fact, in 2016 about 50% of the biopharmaceuticals in the Food and Drug Administration and European Medicines Agency lists were processed by freeze-drying[4].

Since the pharmaceutical industry is highly controlled by regulatory agencies, final products must meet rigorous specifications[5,6], in particular with respect to the Critical Quality Attributes (CQAs). Above all, the residual moisture (RM) is a major CQA for freeze-dried products[7], since it strongly affects the stability of the active component and the collapse of the dried product during storage[8-11].

At the end of the process, a typical freeze-dried product has a low RM, even lower than 1%[10,12], but a wide distribution of RM values can be found in the vials of the same batch[13]. In fact, pressure gradients between shelves[14], non-uniformity of temperature of shelf surface[15], and different heat fluxes to the vials according to their position in the dryer[16,17] are responsible of non-uniform pressure and temperature conditions in the drying chamber, particularly in large scale equipment[14]. As a consequence, taking also into account non-uniform nucleation temperature[18], sublimation rate and product temperature are different from sample to sample[19] and, thus, the RM. Karl Fischer (KF) titration is the most common method employed in the pharmaceutical industry to measure the RM[20-22], despite the downsides it presents: it is a destructive and time-consuming technique, since handling of the samples is required for sample pre-treatment. Thus, only few samples from each batch are tested.

In this framework, Near-Infrared Spectroscopy (NIR spectroscopy or NIRS) was used as a method to measure the RM in freeze-dried samples in place of KF titrations: samples handling and pre-treatment are not required, and the product can be analyzed through the sealed vial. Therefore, NIRS showed up as a safe, fast, and non-destructive method[23-25].

NIRS is a spectroscopic technique developed in the range 14300-4000 cm^{-1} (700-2500 nm) of the electromagnetic spectrum, i.e. in the near-infrared region: samples are irradiated by a NIR beam and a fraction of NIR energy is absorbed by the sample[26-29]. Atomic groups like O-H, N-H, C-H, S-H have a dipole moment[30], thus molecule like water, sugar and proteins are strong NIR absorber. These

are the most common components in a freeze-dried drug product (DP), and for this reason NIRS is regarded as a suitable technique to be applied in the pharmaceutical field. In particular, water gives high signals around 6900 cm^{-1} and 5150 cm^{-1} [31], which justify the choice of NIRS as the technique alternative to KF for the determination of the RM[12,24,32-39].

When NIR is used for RM measurement a calibration step is required by processing the NIRS spectra with the values obtained using a reference analytical technique (in this case KF titrations), and identifying a regression model. Following, the model is validated with samples of the same product not employed for the calibration. An attempt to minimize the experimental activity required for model development was based on the use of samples with different products in the calibration step: Grohganz et al. developed a model in presence of different mixtures of mannitol - sucrose and applied it to samples containing proteins and other excipients[32], and Mainali et al. developed a model with three DPs and validated it with another DP[40]. By following this approach, in case the model has been properly formulated, it can be used for a wider set of products. Another step in the direction of the minimization of the experimental activities for adjusting a NIRS model to a new product was done by Clavaud et al.[41] and by Grohganz et al. (2010)[42]. The former included in the calibration a large range of product formulations and process parameters, according to a statistical design of experiments. The latter validated a model varying the composition, excipients, and solid fraction. However, these approaches still required a consistent experimental activity.

Within this framework, here an algorithm was developed to minimize the experimental effort required to calibrate the NIR-based tool for RM measurement, aiming to get an algorithm that could be applied also when the product processed is different from the one used in the calibration step (thus shortening the calibration step). In particular, the cases of a different concentration of the excipient and of the presence of another component in the mixture (an amino-acid, in this case), were considered. The algorithm was built by means of a Partial Least Square (PLS) regression, which is by far the most common multivariate method employed when NIRS data have to be handled[11,12,23,27,43]. Great attention was paid to the selection of the parameters used for training the algorithm, rather than to the number of the products used in the calibration set. The idea behind is to focus the regression on a region of the NIR spectra that could be specific for the water signals in several products, thus not taking into account the parts of the spectra more related to the product being analyzed. Using a small region of the spectra allows to develop a model with a low number of variables, which is of outmost importance to adapt the model to different products.

The first aim of this study was showing the better performance of a PLS model developed on a small wavelength range. The second was challenging its robustness, i.e. its suitability and adaptability to a product not involved for developing purposes. In fact, a model that could be used in case other

products and systems are considered without further calibration would be a significant advancement, shortening drastically the calibration step. The benefits of a PLS model obtained from a small part of the spectra are stressed by a comparison against a simpler method, based on a single-variable linear regression, and a more complex one, using a wider wavelength range or calibrating the PLS model with several products.

2. Materials and Methods

2.1 Multivariate analysis

NIR spectra can gather lot of information about the product, both chemical and physical. Each spectrum is made by hundreds of signals, i.e. one value of absorbance for each wavelength scanned. For developing a NIRS-based application typically hundreds of spectra are collected. As a consequence, a huge amount of data has to be handled, and not all of these data are equally significative for the application considered[30]. This is the reason why Multivariate Analysis (MVA) is required to extract relevant information[43].

Principal Component Analysis (PCA) and Partial Least Square (PLS) are much common among the MVA techniques and are both based on latent variables (or principal component, PCs). PCA is a linear algebra tool, able to reduce a complex dataset into a simpler structure[44]. PLS is a linear regression method, able to model the relationship between the (spectral) dataset **X** and one or more of its attributes (**Y**)[45].

At this point, the substantial difference between a classical linear regression and a PLS regression has to be briefly pointed out. The classical linear regression may appear easier to compute, as it makes use of a single variable, i.e. of the value of absorbance at a specific wavelength. However, choosing properly a single wavelength from an entire spectrum to describe a property is not trivial, as relevant information can be gathered in several signals, and a slightly shift of the peaks may occurs, e.g. water signals may be significative all around 5150 cm^{-1}. Moreover, at low water concentration, signals of any interferents may overlap the signal of the water bands. On the contrary, the regression carried out by PLS is still linear, but it allows to account for a higher numbers of wavelengths, thus taking into account any interferences with water signals.

In this study, PCA was used to perform the removal of outliers from spectral datasets (as explained in section 2.3.1), while all the modeling was carried out using PLS. In the followings, a short introduction to the theoretical principles of PLS is presented. Dedicated works can provide a detailed and inclusive description of the techniques mentioned above[45-48].

2.1.1 Partial Least Squared regression

PLS carries out a decomposition of the data matrix **X** and of the matrix of attributes **Y** in a new reference system, i.e. the space described by *A* latent variables (where *A* is the number of PCs adopted by the decomposition). The dimensions of **X** (*M* x *J*) correspond to the number of spectra (*M*) and to the number of wavelength (*J*) scanned for each spectrum. **Y** contains the values of a number equal to *I* of attributes, in this case the values of RM, for each one of the *M* observations corresponding to the data matrix **X**.

Both **X** and **Y** are decomposed in the respective loading and score matrices. The number of loadings[49], i.e. of the PCs, is a crucial choice: a higher number of PCs can explain more variance in **X**, but if too many PCs are employed, the data matrix **X** may be overfitted. The scores represent the spectra in the new reference system, and can be predicted as a linear combination of the data with some coefficients (or weights), collected in the *J* x *A* matrix of weights $\mathbf{W}^*$[45]. The calibration step of a PLS model is finalized when the regression matrix **B** (*M* x *I*) is obtained, by means of the linear combination between $\mathbf{W}^*$ and the loading matrix of **Y**. Given a new dataset ($\mathbf{X}_{pred}$), its quality attributes ($\mathbf{Y}_{pred}$) can be predicted through the **B** matrix. If the quality attributes corresponding to the observations of $\mathbf{X}_{pred}$ are known, e.g. because they were measured by analytical methods, they can be compared with the predicted ones, thus evaluating the performances of the regression.

2.1.2 Evaluation of the performances

In order to assess the calibration and prediction performances of the PLS model, the root mean square error of calibration and of prediction (respectively, RMSEC and RMSEP) were adopted. These parameters describe the degree of agreement between the measured value of a quality attribute ($\mathbf{y}_i$) and the corresponding value calculated by the model ($\mathbf{y}_{pred,i}$)[50]. In this case, $\mathbf{y}_i$ is the value of RM of the *i*-th sample measured by KF titration, and $\mathbf{y}_{pred,i}$ is the RM value calculated by the model for the same *i*-th sample. RMSEC and RMSEP were calculated with the following equations:

$$\mathrm{RMSEC} = \left[\frac{\sum_{i=1}^{M^{\mathrm{c}}}\left(\mathrm{y}_i^{\mathrm{c}} - \mathrm{y}_{pred,i}^{\mathrm{c}}\right)^2}{M^{\mathrm{c}} - A - 1}\right]^{0.5} \quad (1)$$

$$\mathrm{RMSEP} = \left[\frac{\sum_{i=1}^{M^{\mathrm{v}}}\left(\mathrm{y}_i^{\mathrm{v}} - \mathrm{y}_{pred,i}^{\mathrm{v}}\right)^2}{M^{\mathrm{v}}}\right]^{0.5} \quad (2)$$

where $\mathbf{y}_i^{\mathbf{c}}$ and $\mathbf{y}_{pred,i}^{\mathbf{c}}$ are referred to the *i*-th sample of the calibration set, $\mathbf{y}_i^{\mathbf{v}}$ and $\mathbf{y}_{pred,i}^{\mathbf{v}}$ to the *i*-th sample of the validation set, M^{c} and M^{v} to the number of samples making up respectively the calibration and the validation sets[51].

Another method implemented to validate a PLS regression was cross-validation (CV), aimed at performing an internal validation of the model being developed. CV was carried out with a ν-fold approach. This means that the dataset was split in ν groups of approximately the same size, by selecting the observations randomly from the main dataset. Following, the regression is trained on all the groups, except one, that is left out and used for the validation of the regression just calculated. This procedure is repeated until all the subgroups have been left out once. The entire cycle of CV can be performed several times in order to use a different random split of the dataset in the ν groups[50,52]. The CV cycle comes out in two parameters: R2, which gives a measurement of the goodness of the fit of the calibration set, and the root mean square error of cross validation (RMSECV), used to estimate the predictive ability of the model. RMSECV can be estimated through equation (1), being $\mathbf{y}_{\boldsymbol{i}}^{\mathbf{v}}$ and $\mathbf{y}_{\boldsymbol{pred,i}}^{\mathbf{v}}$ referred to the *i*-th sample of the group left out[50], while R2 is calculated with the following equation[51]:

$$R2 = 1 - \frac{\sum_{i=1}^{M^c}\left(y_i^c - y_{pred,i}^c\right)^2}{\sum_{i=1}^{M^c}(y_i^c - y_m^c)^2} \quad (3)$$

where $\mathbf{y}_{\boldsymbol{i}}^{\mathbf{c}}$ and $\mathbf{y}_{\boldsymbol{pred,i}}^{\mathbf{c}}$ are referred to the *i*-th sample of the calibration groups and $\mathbf{y}_{\boldsymbol{m}}^{\mathbf{c}}$ is their mean value.

To evaluate the performances in the external validations, the Q2 parameter was calculated similarly to R2, by implementing equation (3) with the values of RM measured and calculated for all the samples of the dataset involved in the external validation. Thus, Q2 represents the goodness of the fitting in prediction.

The parameters presented above were employed to compare the models, as indicative of the goodness of the fit in calibration and of the predictive ability. An accurate model is expected to turn out values of R2 and Q2 close to 1 and low RMSEC, RMSEP, and RMSECV values: in literature values in the order of 0.10 - 0.20 were presented as representative of good performances[13,32,34].

Finally, the comparison between the measured and the calculated quality attributes can be pictured in a correlation plot, where the $\mathbf{Y}_{\boldsymbol{pred}}$ is drawn over the measured values **Y**. The bisect line of the correlation plot represent the ideal identity between the predicted and the measured values of the quality attribute. An accurate PLS regression is expected to turn out in a correlation plot where the distribution of the observations is well aligned along the bisect line, with the observations laying near it.

2.2 Experimental procedures

2.2.1 Case Study

Some freeze-drying cycles were carried out to get the samples used for model development. In the first part of the study, a sucrose 6%$_w$ aqueous solution, freeze-dried into 2R glass vial (Nuova Ompi, Piombino Dese, Italy) with a filling volume of 1 mL, was considered for samples preparation: sucrose was supplied by Merck Life Science (Darmstadt, Germany) and ultra-pure water was obtained by a Millipore water system (IQ 7000, Merck Millipore, Burlington, USA). The choice of this product was aimed at mimicking the placebo solution of a commercial freeze-dried product by Merck Serono SpA. Vials were placed according to a honeycomb layout surrounded by metal frames, in direct contact with the freeze-drier shelves, and processed in a lab-scale freeze-drier (Lyostar3, SP Scientific, Warminster, USA) in the laboratories of the Guidonia Montecelio (Italy) site of Merck Serono SpA. Two freeze-drying cycle were carried out with the same process conditions:

- freezing at -45°C for 6 h, with an annealing step at -15°C for 2 h;
- primary drying at -25°C and 5 Pa for 30 h;
- secondary dying at 35°C and 5 Pa for 10 h;
- all the cooling / heating rate were set at +/- 2°C/min, except for the heating rate in the transition from the primary to the secondary drying, set at +1°C/min.

About 200 vials were processed in each test, and all samples were crimped at the end of the cycle to prevent any humidification from the environment. From each batch 65 samples were taken randomly, hence the sample set used as case study (labelled as S6 in the followings) was made of 130 vials.

For calibrating the model, samples in a wider range of moisture were needed. Therefore, 115 vials were humidified on purpose after the freeze-drying step, in order to get a RM in the range 1% - 5% (15 vials were left unaltered, thus their water content was representative of the RM obtained at the end of the freeze-drying process). For the humidification, a procedure already implemented for the manual humidification of freeze-dried samples[53] was followed:

- the amount of water required to get a target value of RM was calculated, given the weight of the cake and assuming an initial RM of 0.5%;
- each vial was unsealed, the stopper placed on a flat surface, and the amount of water calculated dropped on the internal face of the stopper;
- vials were closed upside-down on the stopper, i.e. with the bottom pointing upside, thus, avoiding the direct contact of the drop with the cake, which would cause a partial collapse of the cake;
- vials were left in the upside position for more than 24 h, to allow the diffusion of water.

After the humidification, vials were sealed and stored at room temperature.

In order to test the robustness of the model further sample sets, having specific different features, were necessary. The validations carried out with these samples are called “external validation”,

meaning that they were performed using datasets from products different from those used for the calibration. Analogously, the respective datasets are called “external datasets”. The followings samples were thus produced:

(a) 10 samples of sucrose 6%$_w$ in 6R vial, the same product of S6, but in a different vial format (sample set VF);

(b) 25 samples of sucrose 9%$_w$ (sample set S9), thus having a solid fraction higher than S6;

(c) 35 samples of trehalose 6%$_w$ (sample set T6), with the same solid fraction as S6, but a different amorphous excipient;

(d) four different mixtures sucrose - arginine, containing different concentrations of arginine:

- 10 samples of sucrose 3%$_w$ - arginine 3%$_w$, (50% of arginine, sample set SA50);
- 10 samples of sucrose 6%$_w$ - arginine 2% $_w$, (25% of arginine, sample set SA25);
- 20 samples of sucrose 6%$_w$ - arginine 1%w, (14.3% of arginine, sample set SA14);
- 10 samples of sucrose 6%$_w$ - arginine 0.5%$_w$, (7.7% of arginine, sample set SA07).

Samples of group (a) were intended to evaluate if the vial format, i.e. the shape of the container, could affect the characteristics of the spectra collected. Since the NIR beam is affected by phenomena of diffraction and light scattering, the aim was investigating if the spectra (of the same product) collected through different vial formats were significantly different. Samples of groups (b) and (c) were used to evaluate the applicability of the PLS regression to a different product, i.e. if the model could be applied for samples with a different solid fraction or made of a different (but still amorphous) excipient. Finally, the goal of samples of group (d), generically referred all together as sample set SA, was testing the effect of an additional amino acid and evaluating up to what extent the amino acid could interfere according to its concentration.

All the above listed sample sets were produced in 2R vial with 1 mL filling volume, except for the sample set VF. Since samples of VF were produced in 6R vials (Nuova Ompi, Piombino Dese, Italy), the filling volume was 2 mL, in order to get the same height of liquid in the vial and keep the same freeze-drying process conditions (in terms of temperature, pressure, and duration) as applied to the freeze-drying process of S6. For the preparation of these solutions, ultra-pure water by a Millipore water system (IQ 7000, Merck Millipore, Burlington, USA) was used, sucrose and L-Arginine Monohydrochloride were supplied by Merck Life Science (Darmstadt, Germany), and trehalose dihydrate by Sigma-Aldrich (Saint Louis, USA).

A freeze-drying cycle was run to produce the sample sets VF, S9, TS, and SA with the same process conditions listed above.

All the samples were crimped at the end of the cycle and stored at room temperature. In order to validate the model in all the range of RM used for the calibration, also the samples of these sets were

humidified, applying the procedure described above. Then, five samples from each one of the sample sets VF, S9 and T6 and two from each sample sets SA were left unaltered, while the others were humidified in the same range of RM (1% - 5%) adopted for the humidification of S6.

2.2.2 Spectra acquisition

The samples listed above were all analyzed through a Fourier Transform NIR spectrometer (Antaris MX FT-NIR, Thermo Fischer Scientific, Waltham, USA), equipped with an InGaAs detector and a halogen NIR source. Spectra were recorded in the full wavelength range 10000 - 4000 cm^{-1}, in diffuse reflectance mode, with 32 scans for spectrum, and a gain of 1. In order to reduce the noise and get a well-defined path, for each sample the spectrum was obtained as the average between 96 scans, and the whole acquisition took about 1 min for each sample.

All the samples were numbered and scanned through the side wall of the vial, with the NIR beam pointing on the side of the freeze-dried cake (vertical layout), except for the samples of S6 that were scanned not only in vertical layout, but also in horizontal layout, i.e. through the bottom of the vial, with the NIR beam pointing on the bottom of the cake (Fig. A1 in Appendix A). Therefore, from the sample set S6 two spectral datasets were obtained: dataset S6 made of the spectra collected in vertical layout (used as the case study), and the dataset S6-H, made of the spectra collected in horizontal layout (used as an external dataset). The aim of dataset S6-H was investigating in what extent the information collected by the NIR beam through the side of the vial differed from the information collected through the bottom.

2.2.3 Karl Fischer titration

After the collection of the spectra, KF titrations were performed for all the samples. A coulometric titrator was employed (CX30 Mettler Toledo, Columbus, USA), and, to perform the analysis, the instructions of the Standard Operating Procedure (SOP) of the company were followed. According to the SOP, the equipment was verified before every campaign of titrations by the titration of a standard solution (Honeywell HYDRANAL Water Standard 1.0, Fischer Scientific, Milano, Italy). For each sample, two titrations were performed, and the value of RM calculated as their average, in order to reduce the uncertainty related to the analytical method.

The samples not humidified had a RM in the range 0.21%-0.80%, while the humidification procedure turned out samples with a water content up to 5.46%.

2.3 Data processing

In this study, two algorithms were used, based on PLS analysis, one for developing the model, the

second to assess the robustness of the model previously developed, by means of an external validation.

(A) For model development, the dataset is divided into the calibration and validation sets, a regression is performed using the calibration set, to get the regression matrix **B**, and the RMSEC is calculated. Following, the **B** matrix is applied to the validation set, performing an "internal validation" of the model, and the RMSEP is estimated. As a further internal validation of the model, CV is carried out and the RMSECV and R2 are calculated.

(B) The external validations are done on the entire external dataset, i.e. on a dataset made of samples of a product different from the samples used for calibrating the model. The regression matrix calculated previously is used, and the performance is assessed by calculating the RMSEP and the Q2.

As a graphical outcome, correlation plots of the RM measured vs the RM calculated were drawn.

As a general specification, all the PCA and PLS decompositions presented in the followings were carried out with two PCs, since at least more than 98% of the (co)variance among the data of each dataset was always described by this way. In particular, concerning the PLS regressions, usually the 1st PC was found able to describe the majority of the covariance (slightly less than 80%), and approximately a further 20% was up to the 2nd PC. As a confirmation of this, decompositions performed with a higher number of PCs did not result in lower values of RMSEC and RMSECV and, thus, did not lead to significantly improved performances. Another reason explaining why PLS decompositions were preferentially performed with two PCs was avoiding the overfitting of the calibration set, as a higher number of PCs would restrict the model validity to that specific product, thus reducing the prediction ability of the model when applied to samples of different products.

The procedure for selecting the range of wavelengths used for model development consisted in performing a preliminary PLS decomposition (with two PCs) on the entire wavelengths range. Following, the loadings of this decomposition were analyzed to identify the most significant variables, characterized by higher peaks in the trend of the loadings. Finally, a PLS regression (again with two PCs) was performed on the selected wavelength range, in order to get the regression matrix. For processing NIR spectra and modeling, scripts were written using MATLAB (R2019b), based on the algorithms of Shlens (2020)[44] and Wold (1992)[54].

2.3.1 Pretreatment of spectra

Pretreatments are mathematical corrections of the spectra required to reduce some interfering phenomena affecting the NIR beam, e.g. baseline shift, light scattering, and random noise[43].

Pre-processing of the spectra was carried out by using a standard normal variate (SNV) approach,

consisting in subtracting the mean value of the spectrum from all its points, and dividing for its standard deviation. SNV is useful to reduce the variability between samples due to light scattering, to adjust the base-line shift, and to focus the study of the spectrum on points with higher variability from the mean value[55]. Since for this application the SNV method was found to turn out data suitable enough for high accuracy models, other common pretreatment techniques, e.g. scatter corrections, derivates and smoothing[55], were not applied.

Before modeling, the wavelength range was reduced to 10000 - 4200 cm^{-1}, since the signals collected at higher frequencies appeared too noisy and potentially misleading. Beside this, the spectra obtained in some cases appeared very different from the expected ones, mainly due to issues in the signal acquisition or to a non-perfect contact between the NIR probe and the vial wall. All these spectra may be considered as "outliers", and have to be removed from the following analysis. This may be done manually, looking at all the spectra obtained and deleting the "anomalous" ones, but, with the goal to speed up this step, PCA was used to remove them (considered as outliers). After this, the removed spectra were checked to be effectively those where the trend could not be related to an unexpected water content. As an example, Fig. A2 in Appendix A shows the spectra of S6 after pretreatments and removal of outliers. The number of spectra making up each dataset after the removal of outliers is summarized in Table 1. All the models developed in the following were based on spectral data pre-processed in the same way.

2.3.2 Model developing

Dataset S6 was used for developing three models:

- Model SR, based on a PLS regression on a small wavelength range;
- Model LR, based on a linear regression considering a single wavelength;
- Model WR, based on a PLS regression on a large wavelength range.

Model SR is the goal of this study and models LR and WR were considered for comparison purposes, to point out the major robustness of model SR.

The performance parameters (RMSEC, RMSEP, RMSECV, R2) were calculated for each model, and compared to assess the effect of the different number of variables, i.e. of wavelengths, considered.

Dataset S6 was split in two sets, in order to perform the calibration with a set and the internal validation with the other. The calibration set was made by 60 spectra (about 50% of S6), randomly selected from S6, and the validation set by the remaining 61 spectra. Since the samples of S6 had a RM content in the range 0.53% - 4.92%, this was the range of applicability of the developed models. In order to develop the models, the preliminary PLS was carried out in the range 10000 - 4250 cm^{-1}, and the significance of the wavelengths evaluated from the loading plot.

For building model LR the most significant signal had to be selected. A peak in the loading of the 2nd PC was found 5150 cm^{-1} (see Fig. A3(a) in Appendix A), corresponding to the signal of water[31,56], thus it was taken as the variable for calculating the linear regression. The values of absorbance at 5150 cm^{-1} were thus collected for each spectrum of the calibration set, and a linear regression was calculated by fitting these values of absorbance over the corresponding value of RM measured by KF titrations (see Fig. A4, upper graph, in Appendix A). The linear regression was performed by the MATLAB function *polyfit*, which applies the least squares method to calculate the coefficients of a polynomial that is the best fit for the data given[57]. In this case, a polynomial of first degree, i.e. a straight line fit, was computed as regression line. Once obtained the coefficients of the linear regression, they were applied to the values of absorbance at 5150 cm^{-1} of the validation spectra, in order to predict the RM and perform the internal validation of model LR. In order to compare the performance of the linear regression with the PLS regressions, also model LR underwent to CV to calculate RMSECV and R2, and equations (1) and (2) were applied to calculate the RMSEC and RMSEP.

To build model WR, the wavelength range 7100 - 4250 cm^{-1} was selected, since here the loadings of the preliminary PLS decomposition showed high signals (see Fig. A3(a) in Appendix A). A PLS regression was performed on the selected range with the calibration set. The internal validation of model WR was then performed, obtaining the predicted values of RM. The RMSEP and RMSEC were calculated, as well as the RMSECV and R2 after CV.

However, the wavelength range adopted for model, though made of significant signals, still contained a lot of information in addition to the signal specific for water (at 5150 cm^{-1}). Hence, attention was paid in selecting a smaller range of signals, wide enough to collect what was relevant for the water quantification but, at the same time, to leave out the signals not so specific for water, which may mix up the calibration. Therefore, to develop model SR the wavelength range was further reduced to 5290 - 4785 cm^{-1}, where just the highest loadings peaks were found. This range was supposed to be most significant for representing the water content. It has to be underlined that this range is just around the main signal of water at 5150 cm^{-1}. Similarly to the procedure described for the model WR, model SR was built performing another PLS regression on the reduced wavelength range with the calibration set, obtaining the **B** regression matrix, and the internal validation performed using the validation set through the application of **B**. RMSEC and RMSEP were calculated, as well as RMSECV and R2.

2.3.3 External validations

Challenging the robustness of model SR was the second goal of this paper. The performances of the external validations were assessed by means of RMSEP and Q2, since calibrations and CV were not

needed (thus RMSEC, RMSECV, and R2 were not calculated).

For some datasets, as it will be shown in the followings, an "ad-hoc" PLS regression, i.e. a PLS regression calibrated specifically for a product and validated with samples containing the same product, was also developed: this was supposed to give the best performances, accordingly to the goodness of the experimental data and their aims were:

- comparing the performances of the model SR with the best achievable performance;
- discriminating if any poor performances were due to the little applicability of model SR (low performances of the external validation, but high for the ad-hoc model) or to the poor accuracy of the experimental data (low performances of both the external validation and the ad-hoc model).

The ad-hoc models were developed similarly to what presented for model SR and WR. The processing applied to each external dataset is listed in the followings.

Dataset S6-H:

- external validation of model SR: 111 spectra with a RM content in the range 0.53% - 4.92% (the lower number of spectra with respect to dataset S6 is due to a higher number of outliers removed from the S6-H dataset);
- ad-hoc PLS regression (model H): 55 spectra for calibration, 56 spectra for the internal validation. Since the product was the same as dataset S6, the same wavelength range used for model SR (5290 - 4785 cm^{-1}) was adopted for developing model H.

Dataset VF:

- external validation of model SR: 10 spectra with a RM content in the range 0.21% - 5.46%;
- because of the lower number of samples, an ad-hoc PLS regression was not performed.

Dataset S9:

- external validation of model WR and of model SR: 24 spectra with a RM content in the range 0.40% - 3.87%;
- ad-hoc PLS regression (model SF, which stands for solid fraction): 12 spectra for calibrating, 12 spectra for the internal validation. Being the samples made of sucrose as dataset S6 (with just a different solid fraction), the same range used for model SR (5290 - 4785 cm^{-1}) was adopted.

Dataset T6:

- external validation of model WR and of model SR: 30 spectra with a RM content in the range 0.51% - 5.40%;
- ad-hoc PLS regression (model T): 15 spectra for calibrating, 15 spectra for the internal validation. For the selection of the wavelength range, a preliminary PLS decomposition was

performed in the range 10000 - 4250 cm^{-1}. The loadings (Fig. A3(b) in Appendix A) were found similar to the loadings of S6, and a similar range was selected (5300 - 4740 cm^{-1}). For model development, another PLS regression was performed in this range with two PCs.

Dataset SA:

- external validation of model SR: 50 samples, with RM in the range 0.53% - 4.31%;
- since the external validation of model SR turned out a considerably high error associated to the samples with higher arginine concentration, the external validation of model WR was not carried out. Consequently, only the datasets SA07 and SA14 (20 spectra with RM in the range 0.53% - 3.84%) were taken for an external validation of model SR and of model WR, to calculate some performances comparable with the other validations;
- an ad-hoc PLS regression (model A), with just the datasets SA07 and SA14: 10 spectra for calibrating, 10 spectra for the internal validation. A preliminary PLS regression was performed on the range 10000 - 4250 cm^{-1} to get the loading plot (Fig. A3(c) in Appendix A). In this case, quite significative differences were found from the loadings of S6, though the range, with the highest values of loadings where in the same wavelength range observed for dataset S6. Thus, a similar range was selected (5295 - 4745 cm^{-1}) to perform the PLS regression with two PCs for model development.

2.3.4 Global model

An ad-hoc PLS regression (model G) was performed considering all the products previously listed joined in a single dataset (GL). In case a certain product was represented by a considerably higher number of samples than the other products, it might affect the PLS regression, i.e. forcing the regression to give more weight to that specific product. Therefore, only a fraction (about 30%) of samples of S6 (36 spectra) was considered, selected randomly, since dataset S6 was much larger than the others involved in model G. In this way all the products were represented by a comparable number of spectra. Dataset GL, thus, resulted in 110 spectra, with a RM in the range 0.40% - 5.40%, and it was split between the calibration and validation sets (respectively 55 and 55 spectra). The ad-hoc regression was performed in the wavelength range 5300 - 4740 cm^{-1}, made by the union of the ranges adopted for the single ad-hoc PLS regressions (with the exception of the range used for model WR), since this range was likely to include the wavelengths already found significative for the description of the single products. The model G underwent to internal validation and to CV.

Since model G was calibrated with different products, it was assumed to provide the best performances in prediction, at least in the case different products were represented by the same model. Therefore, the external validations of model SR were compared with the performances of model G,

for evaluating its robustness.

Additionally, a linear regression (model G-LR) was developed as described for model LR, at 5150 cm^{-1}, and underwent to internal validation and CV (the curve of calibration of model G-LR is shown in Figure A4 in Appendix A). The aim was investigating to what extent a univariate linear regression could be suitable for the simultaneous modeling of different products, with respect to the multivariate PLS method. For the same purpose, an external validation of model LR was carried out with dataset GL, thus testing the robustness of a linear regression method.

All the performance parameters estimated for the calculations described above are listed in Table 2 and discussed in the following section.

3. Results and discussion

3.1 Model development

Models LR, SR, and WR are compared in the followings to establish which is the most suitable method for predicting the RM content.

Model WR turned out to have the poorest performances (RMSEC and RMSEP equal to 0.314 and 0.300), especially in comparison with model SR. In fact, model SR was extremely accurate, both in calibration and in the prediction (RMSEC and RMSECV equal to 0.118 and 0.122 respectively, R2 equal to 0.988). Taking in mind that the experimental error expected from a KF analysis may be up to 0.3%[32,58,59], model SR could be considered very accurate. Surprisingly, model LR had better performance than model WR, with RMSEC and RMSEP equal to 0.179 and 0.161 respectively, though not as accurate as the SR.

These findings are confirmed by the correlation plots in Fig. 1. The correlation plot of model WR (Fig. 1(c) and (d)) turned out a much wider distribution of the observations around the bisect line, representative of poor accuracy of the model. A similar, though less evident, situation can be seen in Fig. 1(a) and (b), regarding model LR. Conversely, the same observations in the correlation plot of model SR (Fig. 1(e) and (f)) appeared much better aligned along the bisect line, as expected for a model of high accuracy.

Another evidence of the benefits coming from the selection of a smaller wavelength range is displayed in the score plots of Fig. 2, where the scores are grouped by their RM content, as measured by KF titrations. The scores obtained from the PLS decomposition over the smaller range appear well aligned and ordered bottom-up according to the increasing value of the RM. The same observations processed by model WR resulted in a less precise distribution, with scores overlapping.

Therefore, the comparison of these three models pointed out that a PLS regression on a small, but

significative, wavelength range was a well suitable technique to predict the RM content from NIR spectra.

3.2 Effect of the vial

Model SR was applied to dataset S6-H. The external validation of model SR with S6-H spectra yielded in very good performance parameters (RMSEP and Q2 equal to 0.144 and 0.986). The performance parameters of the external validation of SR are comparable to those obtained from the ad-hoc model H, that turned out to be equally accurate. Therefore, model SR could be equally applied to spectra collected with vertical and horizontal layout, and a specific model for modeling spectra collected in horizontal layout was not needed. Fig. 3(b) shows the correlation plot obtained from the application of model SR to the S6-H dataset, and Fig. 3(c) and (d) the outcome of the ad-hoc model H. It can be seen that the observations are distributed similarly along the bisect line, in all the correlation plots, which is indicative of similar performances. Additionally, in Fig. 3(a) a score plot of the observations of S6 and S6-H is displayed. It points out that the observations of S6-H were comparable to those of S6, since no clusters could be found. This is another evidence that the two datasets could be processed properly by the same model.

The fact that the spectra collected in vertical layout and in horizontal layout were similar was an important outcome, meaning that the spectra collected from two different spots of the cake are basically equivalent. Therefore, the fraction of cake irradiated by the NIR beam could be assumed representative of the larger cake volume, at least in the case study here investigated.

Model SR was also applied to dataset VF, in order to further investigate to what extent the vial shape, i.e. the vial format, could affect the applicability of the model. The validation of model SR with the samples of dataset VF is showed in Fig. A5. Though this validation yielded in slightly less accurate performances (RMSEP equal to 0.139), the fitting of the prediction was still very accurate (Q2 equal to 0.985). Therefore, the suitability of model SR for samples produced in a different vial format could be assessed. This outcome confirmed what expected from literature, since the aim of the pretreatment technique, among the others, is to solve the diffraction and light scattering effects, which can also be due to the different vial shape.

3.3 Robustness of the model

3.3.1 Different solid fraction

Here the effect of a different solid fraction was evaluated by means of dataset S9.

The external validation of model WR with dataset S9 was less accurate, especially at lower RM,

where the distribution of the observations in the correlation plots was found significantly spread (Fig. 4(a)). On the other hand, the external validation of SR (Fig. 4(b)) and the ad-hoc model SF (Fig. 4(c) and (d)) appeared comparable, though the former slightly less accurate because of a bit cluttered displacement of the observations at lower RM content. The performance parameters in Table 2 confirm these findings: the RMSEP of the external validation performed with model WR (equal to 0.338) was considerably higher than that obtained with model SR (equal to 0.138), and also Q2 was barely equal to t 0.853. However, the ad-hoc model SF turned out to be extremely accurate in calibration (RMSEC and R2 equal to 0.093 and 0.993) and in prediction (RMSEP equal to 0.078). The comparison between the performance of the external validation of model SR and of model SF was useful to assess the goodness of the experimental data: since the ad-hoc model SF was very accurate, the lower performances of model SR were up to the algorithms itself, i.e. not to any faults of the experimental data. Nevertheless, it has to be said that the performance of model SR applied to S9 was not poor, but just less accurate when compared to the much accurate performances of model SF. Therefore, this external validation suggested that model SR could be applied to samples with higher solid fraction, obtaining a prediction of the RM content sufficiently accurate, although higher accuracy may be obtained from an ad-hoc model.

3.3.2 Different excipient

Firstly, the failure of the external validation of model WR applied to dataset T6 appeared clear. In Fig. 5(a) the correlation plots show a muddled distribution, especially at lower RM. The poor performance parameters confirm these findings (RMSEP equal to 0.410). Due to the limited number of samples available, a more extended calibration might be advisable. Looking at the loading plot in Fig. A3(b), differences could be found in the peak of the 1st PC around 6350 cm^{-1}, and in the regions 7000-5500 cm^{-1} and 4850-4250 cm^{-1} of the 2nd PC. This is fundamental to be noticed, since it could be useful to explain the poorer performances coming out from the external validation of model WR and, lately, to validate the thesis that, for predicting the RM content, using a smaller range may lead to a more robust PLS regression. In fact, the range used for calibrating model WR comprehended the wavelength ranges where differences in the loading trends of S6 and T6 were found, thus its lower robustness appeared reasonable.

The performances of the external validation of model SR and of the ad-hoc model T were basically comparable (RMSEP equal to 0.147 and 0.148 respectively). This means that whatever the goodness of the experimental data was, model SR could be applied to samples of trehalose giving performances equivalent to those of an ad-hoc model. To confirm this, in the correlation plots of model SR (Fig. 5(b)) and of the ad-hoc model T (Fig. 5(c) and (d)), the observations appeared well aligned along the

bisect line.

3.3.3 Addition of an amino acid

The score plot in Fig. 6(a) was obtained by the PLS regression carried out for model SR with dataset SA. Here the division in clusters of the samples is highlighted, according to the different concentration of the amino acid. In the score plot also the scores of dataset S6 are displayed, and the higher the arginine concentration, the farther the scores of SA lay from the clusters of S6. Only the samples of SA07 and SA14 are partially overlapped to the S6 cluster. Therefore, it could be deduced that only samples with lower amino acid concentration could be described (at least partially) by model SR. The outcome of the external validation of model SR to dataset SA is showed in Fig. 6(b). The samples of SA07 and SA14 are nearer to the bisect line, though a constant gap could be seen at any RM content, while the observations of SA25 and SA50 are progressively farther and more spread. The constant gap characterizing SA07 and SA14 could be considered a systematic error, as an evidence of some lacks in the structure of model SR, that actually was not calibrated with any proteinaceous component. Therefore, it appeared that the addition of an amino acid in any of the concentrations here tested represented the introduction of a major variability.

Consequently, to evaluate the robustness of model SR, only SA07 and SA14 were taken for the external validations. The application of model WR to SA07 and A14 ended up certainly in a lower quality prediction, with RMSEPM and Q2 equal to 1.441 and 0.825. It is worth to underline that, from the loading plot of Fig. A3(c), differences from the loadings of S6 could be found: both the loadings of the 1st and 2nd PCs had different trends in the range 7000 - 5300 cm^{-1}, and a significant peak could be found in the trend of the loading of the 2nd PC at about 4490 cm^{-1}. These major differences are in a range involved in model WR, thus this is likely the reason why a model built on a large range has little robustness, at least in the case of RM prediction.

The application of model SR to datasets SA07 and SA14 yielded in better performances, though not very accurate (RMSEP equal to 0.321). Obviously, better performances were obtained from the ad-hoc model A, with RMSEP and RMSECV equal to 0.125 and 0.132. Also, the gap between the distribution of the observations and the bisect line in the correlation plot did no longer occur (Fig. 6(c) and (d)). This confirmed that the lower accuracy of the SR validation was not related to the goodness of the experimental data, but it was up to the intrinsic robustness of model SR. However, few samples were available to calibrate model A, thus an even more accurate modeling could be carried out involving more samples.

Some considerations about the range selected for model A are worth to be pointed out. Though the

range selected for model A (5295 - 4745 cm^{-1}) was not so different from the range used by model SR (5290 - 4785 cm^{-1}), the loading plot showed some differences in the loadings that turned out to be quite significative: the trends of the loadings of the 1st PC and the shapes of the peak around 5150 cm^{-1} described by the 2nd PC were a little different. Therefore, the lack in the intrinsic robustness of model SR should not be related to the selection of the wavelength range. Instead, the robustness of model SR was affected by the too large variability introduced by arginine, being the lowest concentration tested in this study still too strong impacting.

3.4 Comparison with a global model

Model G was calibrated with several products and its performance was used as a comparison with the outcome of the external validations carried out with model SR.

Firstly, some considerations on model G are needed to evaluate its performance. The model was found very accurate in calibration, with RMSEC and R2 approaching those estimated for the calibration of model SR (equal to 0.121 and 0.981). The performance in prediction (RMSEP and RMSECV equal to 0.139 and 0.168) were slightly lower than what found for model SR. However, this was expected, since model G had to describe various characteristics of different products at the same time. For this reason, model G could still be considered accurate. In fact, a rather well aligned distribution of observations in the correlation plots of Fig. 7 (a) and (b) could also be seen.

By comparing the RMSEP values of model G and of the external validations of model SR, it turned out that the performance of model SR applied to dataset S9 was basically the same as model G. A similar conclusion could be drawn regarding the external validation of model SR with dataset T6, since the RMSEP of the external validation was just 5.8% higher than the RMSEP of model G. Higher differences were noticed relative to the external validation of model SR with datasets SA07 and SA14, being its RMSEP more than twice the RMSEP of model G. Consequently, performing a calibration with multiple products seemed beneficial just in case a major variability among products has to be accounted, e.g. the introduction of a proteinaceous component in a considerably high concentration. For minor differences, as the case of datasets S9 and T6, a multi-products model was not really necessary, and a robust model like model SR could provide much similar performances.

The model SR was applied to the global dataset and the prediction yielded in a RMSEP and Q2 equal to 0.193 and 0.970. This was not the best performance assessed for model SR, though it could be considered accurate enough considering that a large variety of product not involved in the SR calibration was accounted for. Also, as expected, the accuracy of this prediction was in-between of the (better) performances found for S6, S9, and T6 and the (wore) performances regarding datasets SA07 and SA14. Conversely, the external validation of model WR turned out in a much higher

RMSEP, equal to0.641. Lately, this is the evidence of the little robustness and, thus, applicability of a PLS regression developed on a larger wavelength range.

Since model LR developed for dataset S6 resulted in surprisingly good accuracy, further investigations on the robustness of a linear regression method were advisable. For this reason, model LR was applied to dataset GL as an external validation with multiple products, and it yielded in a high RMSEP (equal to 0.599) and low Q2. This outcome pointed out the very little robustness of a linear regression method, not even comparable to the robustness proved fora PLS method like model SR..

Additionally, model G-LR was compared with the ad-hoc model G. Beginning with the curve of calibration in Fig. A4, a poorer accuracy could be seen: the variability of the values of RM measures, though placed over the values of absorbance accordingly to a linear trend, was too high for a good fitting. Hence, also in this case, the performances of the linear regression method resulted much less accurate, with higher error both in calibration and in validation. In fact, model G-LR had an RMSEC more than twice the RMSEC of model G, and much higher RMSEP and RMSECV (equal to 0.349 and 0.312). The minor accuracy of model G-LR could also be guessed by the correlation plots of Fig. 7 (c) and (d), where more spread distributions of observations are displayed, especially at lower RM values.

4. Conclusions

In this paper NIRS was applied to develop a model suitable for the prediction of the RM content of freeze-dried products, focusing particularly on its robustness and suitability for different products not involved in the calibration procedure. A key point was proving the major robustness of a PLS model developed considering a small part of the spectra: when a larger wavelength range is used for calibrating a model, more characteristics specific of the product used for the calibration are taken into account. Hence, when it comes to apply the model to different products, too many not essential signals might be misleading for the prediction.

Developing a robust PLS model by selecting variables accurately, rather than making use of huge datasets for its the calibration, was a relevant advancement. Indeed, a single product (sucrose $6\%_w$) was employed for the calibration of this PLS regression (model SR) and it was developed with just two PCs, over a small wavelength range. A similar PLS regression was developed on a wider range (model WR), but model SR turned out to be extremely accurate for sucrose $6\%_w$, much more than model WR.

Model SR was adopted with datasets of spectra collected from vials placed in different layout, and to

samples produced in different vial format: the accuracy of the outcome suggested that the vial shape did not affect significatively the collection of the spectra.

Then, model SR was applied to samples containing a higher solid fraction of sucrose and samples made of a different excipient, still yielding pretty much accurate performance, much more accurate than the performance of model WR when applied to the same datasets. Hence, model SR had the great advantage to be applicable to different products without further calculation and relevant loss of accuracy. Lower accuracy resulted from the application of model SR to samples made by a mixture sucrose - arginine, since the concentrations of arginine used in this study represented a source of major variability. A global model, calibrated with multiple products, gave basically the same results as model SR, except for the samples containing arginine, which appeared better described. These findings proved the suitability of model SR for samples different from those used for its calibration, unless too significant variations are introduced: in this case an ad-hoc multi-products calibration could be beneficial.

Besides, the PLS regression was compared to a linear regression method developed considering a single wavelength. If the linear regression appeared quite accurate when applied to the same single product used for its calibration, a failure of its performance was evident when its robustness was tested with different products. Similarly, a linear regression method was showed to be unable to model multiple different products at the same time. Therefore, the PLS method appeared clearly more suitable.

Finally, it has to be remarked that the accuracy of this method is strongly dependent on the accuracy of the KF titration used for model calibration: the proposed approach aims avoiding the repetition of the calibration step when the vial is changed, when the solid fraction is changed, and when the active pharmaceutical ingredient, an amino acid in the considered case study, is added, thus allowing carrying out the calibration step with just the excipient.

Acknowledgment

The authors thankfully acknowledge Merck Serono SpA for the contribution and financial support. The assistance in the experimental activities of Daniele Mari, Senior Laboratory Technician, Biotech Pharmaceutical Development Department, Merck Serono SpA, Guidonia Montecelio, Roma, is also gratefully appreciated.

References

1. Reh G. 2020 Global life sciences outlook. Deloitte Insights 2020. Available at: https://www2.deloitte.com/content/dam/Deloitte/pl/Documents/Reports/pl_2020_global_life_sciences_outlook.pdf. Accessed January 26, 2021.
2. Langford A, Bhatnagar B, Walters R, Tchessalov S, Drying technologies for biopharmaceutical applications: recent developments and future direction. Dry Technol 2018;36(6):677–684. https://doi.org/10.1080/07373937.2017.1355318.
3. Harguindeguy M, Fissore D, Temperature/end point monitoring and modelling of a batch freeze-drying process using an infrared camera. Eur J Pharm Biopharm 2021;158:113-122. https://doi.org/10.1016/j.ejpb.2020.10.023.
4. Fissore D, McCoy T. Editorial: freeze-drying and process analytical technology for pharmaceuticals. Front Chem 2018;6(622):1-2. https://doi.org/10.3389/fchem.2018.00622.
5. Oddone I, Pisano R, Bullich R, Stewart P. Vacuum-induced nucleation as a method for freeze-drying cycle optimization. Ind Eng Chem Res 2014;53(47):18236-18244. https://doi.org/10.1021/ie502420f.
6. De Beer T, Burggraeve A, Fonteyne M, Saerens S, Remon JP, Vervaet C. Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes. Int J Pharm 2011;417(1-2):32–47. https://doi.org/10.1016/j.ijpharm.2010.12.012.
7. Koganti VR, Shalaev EY, Berry MR, Osterberg T, Youssef M, Hiebert DN, Kanka FA, Nolan M, Barrett R, Scalzo G, Fitzpatrick G, Fitzgibbon N, Luthra S,Zhang L. Investigation of design space for freeze-drying: use of modeling for primary drying segment of a freeze-drying cycle. AAPS PharmSciTech 2011;12(3):854-861. https://doi.org/10.1208/s12249-011-9645-7.
8. Pikal MJ, Dellerman KM, Roy ML, Riggin RM. The effects of formulation variables on the stability of freeze-dried human growth hormone. Pharm Res 1991;8(4):427-436. https://doi.org/10.1023/a:1015834724528.
9. Hu W-W, Wang Z, Hollister SJ, Krebsbach PH. Localized viral vector delivery to enhance in situ regenerative gene therapy. Gene Ther 2007;14:891-901. https://doi.org/10.1038/sj.gt.3302940.
10. Tang X, Pikal MJ. Design of freeze-drying processes for pharmaceuticals: practical advice. Pharm Res 2004;21(2):191-200. https://doi.org/10.1023/B:PHAM.0000016234.73023.75.
11. Roggo Y, Chalus P, Maurer L, Lema-Martinez C, Edmond A, Jent N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. J Pharm Biomed Anal 2007;44(3):683-700. https://doi.org/10.1016/j.jpba.2007.03.023.
12. Wiggenhorn M, Winter G. The current state of PAT in freeze drying. Eur Pharm Rev 2005;

8(1):38-44. Available at: https://www.europeanpharmaceuticalreview.com/article/2950/the-current-state-of-pat-in-freeze-drying/ . Accessed September 7, 2021
13. Kauppinen A, Toiviainen M, Korhonen O, Aaltonen J, Järvinen K, Paaso J, Juuti M, Ketolainen J. In-line multipoint near-infrared spectroscopy for moisture content quantification during freeze-drying. Anal Chem 2005;85(4):2377-2384. https://doi.org/10.1021/ac303403p.
14. Barresi AA, Pisano R, Rasetto V, Fissore D, Marchisio DL. Model-based monitoring and control of industrial freeze-drying processes: effect of batch nonuniformity. Dry Technol 2010;28(5):577-590. https://doi.org/10.1080/07373931003787934.
15. Rambhatla S, Tchessalov S, Pikal MJ. Heat and mass transfer scale-up issues during freeze-drying, III: control and characterization of dryer differences via operational qualification tests. AAPS PharmSciTech 2006;7(2):E1-E10. https://doi.org/10.1208/pt070239.
16. Rambhatla S, Pikal MJ. Heat and mass transfer scale-up issues during freeze-drying, i: atypical radiation and the edge vial effect. AAPS PharmSciTech 2003;4(2):1-10. https://doi.org/10.1208/pt040214.
17. Fissore D, Harguindeguy M, Ramirez DV, Thompson TN. Development of freeze-drying cycles for pharmaceutical products using a micro freeze-dryer. J Pharm Sci 2019;109(1):797-806. https://doi.org/10.1016/j.xphs.2019.10.053.
18. Arsiccio A, Barresi AA, De Beer T, Oddone I, Van Bockstal P-J, Pisano R. Vacuum Induced Surface Freezing as an effective method for improved inter- and intra-vial product homogeneity. Eur J Pharm Biopharm 2018;128:210-219. https://doi.org/10.1016/j.ejpb.2018.04.002.
19. Fissore D, Barresi AA. Scale-up and process transfer of freeze-drying recipes. Dry Technol 2011;29(14):1673–1684. https://doi.org/10.1080/07373937.2011.597059
20. Connors KA. The Karl Fischer titration of water. Drug Dev Ind Pharm 1988;14(14):1891-1903. https://doi.org/10.3109/03639048809151996.
21. Cachet T, Hoogmartens J. The determination of water in erythromycin by Karl Fischer titration. J Pharm Biomed Anal 1998;6(5):461-72. https://doi.org/10.1016/0731-7085(88)80013-x.
22. Water: semi-micro determination, European Pharmacopoeia 8th edition Volume 1, chapter 2.5.12, 2013.
23. Luypaert J, Massart DL, Vander Heyden Y. Near-infrared spectroscopy applications in pharmaceutical analysis. Talanta 2007;72(3):865-883. https://doi.org/10.1016/j.talanta.2006.12.023.
24. Clavaud M, Roggo Y, Degardin K, Sacre PY, Hubert P, Ziemons E. Global regression model for moisture content determination using near-infrared spectroscopy. Eur J Pharm Biopharm 2017;119:343-352. https://doi.org/10.1016/j.ejpb.2017.07.007
25. Mazivila SJ, Olivieri AC. Chemometrics coupled to vibrational spectroscopy and spectroscopic

imaging for the analysis of solid-phase pharmaceutical products: A brief review on non-destructive analytical methods. Trends Analytical Chem 2018;108:74-87. https://doi.org/10.1016/j.trac.2018.08.013
26. Ciurczak EW, Drennen JK. Pharmaceutical and Medical Applications of Near-infrared Spectroscopy, Marcel Dekker, New York: Marcel Dekker, Inc; 2002.
27. Blanco M, Coello J, Iturriga H, Maspoch S, De La Pezuela C. Near-infrared spectroscopy in the pharmaceutical industry. Analyst 1998;123(8):135R-150R. https://doi.org/10.1039/A802531B.
28. Osborne BG. Near-infrared Spectroscopy in Food Analysis. In: Meyer RA, ed. *Encyclopedia of Analytical Chemistry : Applications, Theory and Instrumentation*, John Wiley & Sons, Ltd; 2006:1-14.
29. Pasquini C. Near infrared spectroscopy: a mature analytical technique with new perspectives - A review. Anal Chim Acta 2018;1026:8-36. https://doi.org/10.1016/j.aca.2018.04.004.
30. De Beer T, Burggraeve A, Fonteyne M, Saerens S, Remon JP, Vervaet C. Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes. Int J Pharm 2011;417(1-2):32–47. https://doi.org/10.1016/j.ijpharm.2010.12.012.
31. Zheng Y, Lai X, Bruun SW, Ipsen H, Larsen JN, Løwenstein H, Søndergaard I, Jacobsen S. Determination of moisture content of lyophilized allergen vaccines by NIR spectroscopy. J Pharmaceut Biomed 2008;46(3):592-596. https://doi.org/10.1016/j.jpba.2007.11.011.
32. Grohganz H, Fonteyne M, Skibsted E, Falck T, Palmqvist B, Rantanen J. Role of excipients in the quantification of water in lyophilized mixtures using NIR spectroscopy. J Pharm Biomed Analysis 2009;49(4):901–907. https://doi.org/10.1016/j.jpba.2009.01.021.
33. Zheng Y, Lai X, Bruun SW, Ipsen H, Larsen JN, Løwenstein H, Søndergaard I, Jacobsen S. Determination of moisture content of lyophilized allergen vaccines by NIR spectroscopy. J Pharm Biomed Analysis 2008;46(3):592-596. https://doi.org/10.1016/j.jpba.2007.11.011.
34. Li Y, Fan Q, Liu S, Wang L. Simultaneous analysis of moisture, active component and cake structure of lyophilized powder for injection with diffuse reflectance FT-NIR chemometrics. J Pharm Biomed Analysis 2011;55(1):216-219. https://doi.org/10.1016/j.jpba.2010.12.028.
35. Jones JA, Last IR, Macdonald BF, Prebblet KA. Development and transferability of near-infrared methods for determination of moisture in a freeze-dried injection product. J Pharmaceut Biomed 1993;11(11-12):1227-1231. https://doi.org/10.1016/0731-7085(93)80108-D.
36. Last IR, Prebble KA. Suitability of near-infrared methods for the determination of moisture in a freeze-dried injection product containing different amounts of the active ingredient. J Pharm Biomed Analysis 1993;11(11/12):1071-1076. https://doi.org/10.1016/0731-7085(93)80084-E.
37. Derksen MWJ, Van der Oetelaarb PJM, Maris FA. The use of near-infrared spectroscopy in

the efficient prediction of a specification for the residual moisture content of a freeze-dried product. J Pharm Biomed Analysis 1998;17(3):473-480. https://doi.org/10.1016/S0731-7085(97)00216-1.
38. McCoy T, Affleck R, Khamar D. Considering Residual Moisture Measurement in Lyophilized Drug Product. Am Pharm Rev 2019;22(7). Available at: https://www.americanpharmaceuticalreview.com/Featured-Articles/559534-Considering-Residual-Moisture-Measurement-in-Lyophilized-Drug-Product/. Accessed August 26, 2021.
39. Clavaud M, Roggo Y, Degardin K, Sacré P-Y, Hubert P, Ziemons E. 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 2016;30(131):380-390. https://doi.org/10.1016/j.jpba.2016.09.014.
40. Mainali D, Li J, Yehl P, Chetwyn N. Development of a comprehensive near infrared spectroscopy calibration model for rapid measurements of moisture content in multiple pharmaceutical products. J Pharm Biomed Analysis 2014;95:169-175. http://dx.doi.org/10.1016/j.jpba.2014.03.001.
41. Clavaud M, Lema-Martinez C, Roggo Y, Bigalke M, Guillemain A, Hubert P, Ziemons E, Allmendinger A. Near-Infrared Spectroscopy to Determine Residual Moisture in Freeze-Dried Products: Model Generation by Statistical Design of Experiments. J Pharm Sci 2020;109(1):719-729. https://doi.org/10.1016/j.xphs.2019.08.028.
42. Grohganz H, Gildemyn D, Skibsted E, Flink JM, Rantanen J. Towards a robust water content determination of freeze-dried samples by near-infrared spectroscopy. Anal Chim Acta 2010;676(1-2):34-40. https://doi.org/10.1016/j.aca.2010.07.035.
43. Reich G. Near-infrared spectroscopy and imaging: basic principles and pharmaceutical applications. Adv Drug Deliver Rev 2005;57(8):1109-1143. https://doi.org/10.1016/j.addr.2005.01.020.
44. Shlens J. A Tutorial on Principal Component Analysis, 2014. Available at: https://arxiv.org/pdf/1404.1100.pdf. Accessed November 12, 2020.
45. Wold S, Sjostroma M, Eriksson L. PLS-regression: a basic tool of chemometrics. Chemometr Intell Lab 2001;58(2):109-130. https://doi.org/10.1016/S0169-7439(01)00155-1.
46. Hotteling H. Analysis of a complex of statistical variables into principal components. J Educ Psychol 1993;24(6):417-441. https://doi.org/10.1037/h0071325.
47. Nomikos P, MacGregor JM. Monitoring batch processes using multiway principal component analysis. AlChEm J 1994;40(8):1361-1375. https://doi.org/10.1002/aic.690400809.
48. McGregor JF, Kourti T. Statistical process control of multivariate processes. Control Eng Pract

1995;3(3):403-414. https://doi.org/10.1016/0967-0661(95)00014-L.
49. Colucci D, Prats-Montalban JM, Fissore D, Ferrer A. Application of multivariate image analysis for on-line monitoring of a freeze-drying process for pharmaceutical products in vials. Chemometr Intell Lab 2019;187:19-27. https://doi.org/10.1016/j.chemolab.2019.02.004.
50. Beebe KR, Pell RJ, Seasholtz MB. Chemometrics: a practical guide. Gemperline P, ed. New York, John Wiley & Sons; 1998.
51. Muzzio CR, Dini NG, Simionato LD. Determination of moisture content in lyophilized mannitol through intact glass vials using NIR micro-spectrometers. Braz J Pharm Sci 2011;47(2):289-297. https://doi.org/10.1590/S1984-82502011000200010.
52. Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK. Validation of QSAR models - strategies and importance. Int J Drug Des Discovery 2011;2(3):511-519. Available at: https://www.researchgate.net/publication/284566093_Validation_of_QSAR_Models_-_Strategies_and_Importance. Accessed January 27, 2021.
53. S. Bobba, N. Zinfollino, D. Fissore, Application of Near-Infrared Spectroscopy to statistical control in freeze-drying processes. Eur J Pharm Biopharm 2021;168:26-37. https://doi.org/10.1016/j.ejpb.2021.08.009.
54. Wold s. Nonlinear partial least square modeling II. Spline inner relation. Chemometr Intell Lab 1992;14(1-3):71-84. https://doi.org/10.1016/0169-7439(92)80093-J.
55. Rinnan A, Van der Berg F, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. Trends Anal Chem 2009;28(10):1201-1222. https://doi.org/10.1016/j.trac.2009.07.007.
56. De Beer T, Wiggenhorn M, Veillon R, Debacq C, Mayeresse Y, Moreau B, Burggraeve A, Quinten T, Friess W, Winter G, Vervaet C, Remon JP, Baeyens WRG. Importance of using complementary process analyzers for the process monitoring, analysis, and understanding of freeze drying. Anal Chem 2009;81(18):7639-7649. https://doi.org/10.1021/ac9010414.
57. MathWorks, Help Center Documentation. Available at: https://it.mathworks.com/help/matlab/ref/polyfit.html. Accessed February 5, 2021.
58. Ward HW, Sistare FE. On-line determination and control of the water content in a continuous conversion reactor using NIR spectroscopy. Anal Chim Acta 2007;595(1-2):319-322. https://doi.org/10.1016/j.aca.2007.03.020.
59. Clua-Palau G, Jo E, Nikolic S, Coello J, Maspoch S. Finding a reliable limit of detection in the NIR determination of residual moisture in a freeze-dried drug product. J Pharm Biomed Anal 2020;183:13-163. https://doi.org/10.1016/j.jpba.2020.113163.

List of Tables

List of Figures

Table 1. Number of samples for each product, after removal of outliers, in the datasets used in this study.

Product / dataset	S6	S6-H	VF	S9	T6	SA	GL
sucrose 6%$_w$	121	-	-	-	-	-	36
sucrose 6% $_w$ (horizontal layout)	-	111	-	-	-	-	-
sucrose 6% $_w$ (vial 6r)	-	-	10	-	-	-	-
sucrose 9% $_w$	-	-	-	24	-	-	24
trehalose 6% $_w$	-	-	-	-	30	-	30
sucrose – arginine (50% $_w$)	-	-	-	-	-	10	-
sucrose – arginine (25% $_w$)	-	-	-	-	-	20	-
sucrose – arginine (14% $_w$)	-	-	-	-	-	10	10
sucrose – arginine (7% $_w$)	-	-	-	-	-	10	10
Total samples	121	111	10	24	30	50	110

Table 2. Performance parameters of the models developed and of the external validations carried out in this study. The type of algorithm used, the wavelength range (cm^{-1}), and the dataset involved are specified.

Dataset	Type	Model	Range	RMSEC	RMSEP	RMSECV	R2	Q2
S6	(A)	LR	5150	0.179	0.143	0.161	0.979	-
	(A)	WR	7100-4250	0.314	0.300	0.286	0.934	-
	(A)	SR	5290-4785	0.118	0.123	0.122	0.988	-
S6-H	(B)	SR	5290-4785	-	0.144	-	-	0.986
	(A)	H	5290-4785	0.116	0.143	0.131	0.987	-
VF	(B)	SR	5290-4785	-	0.139	0.155	-	0.985
S9	(B)	WR	7100-4250	-	0.338	-	-	0.853
	(A)	SR	5290-4785	-	0.138	-	-	0.986
	(B)	SF	5290-4785	0.093	0.078	0.092	0.993	-
T6	(B)	WR	7100-4250	-	0.410	-	-	0.963
	(A)	SR	5290-4785	-	0.147	-	-	0.984
	(B)	T	5300-4740	0.158	0.148	0.15	0.989	-
SA07 SA14	(B)	WR	7100-4250	-	1.441	-	-	0.825
	(A)	SR	5290-4785	-	0.321	-	-	0.974
	(B)	A	5295-4745	0.054	0.125	0.132	0.989	-
GL	(A)	G	5300-4740	0.121	0.139	0.168	0.981	-
	(B)	SR	5290-4785	-	0.193	-	-	0.970
	(B)	WR	7100-4250	-	0.641	-	-	0.855
	(B)	LR	5150	-	0.599	0.314	-	0.927
	(A)	G-LR	5150	0.261	0.349	0.312	0.935	-

Figure 1. Correlation plots between the RM (%) measured vs RM (%) calculated, obtained from the calibration (left column) and validation (right column) sets of S6 processed by: (a) and (b) model LR, (c) and (d) model WR, (e) and (f) model SR.

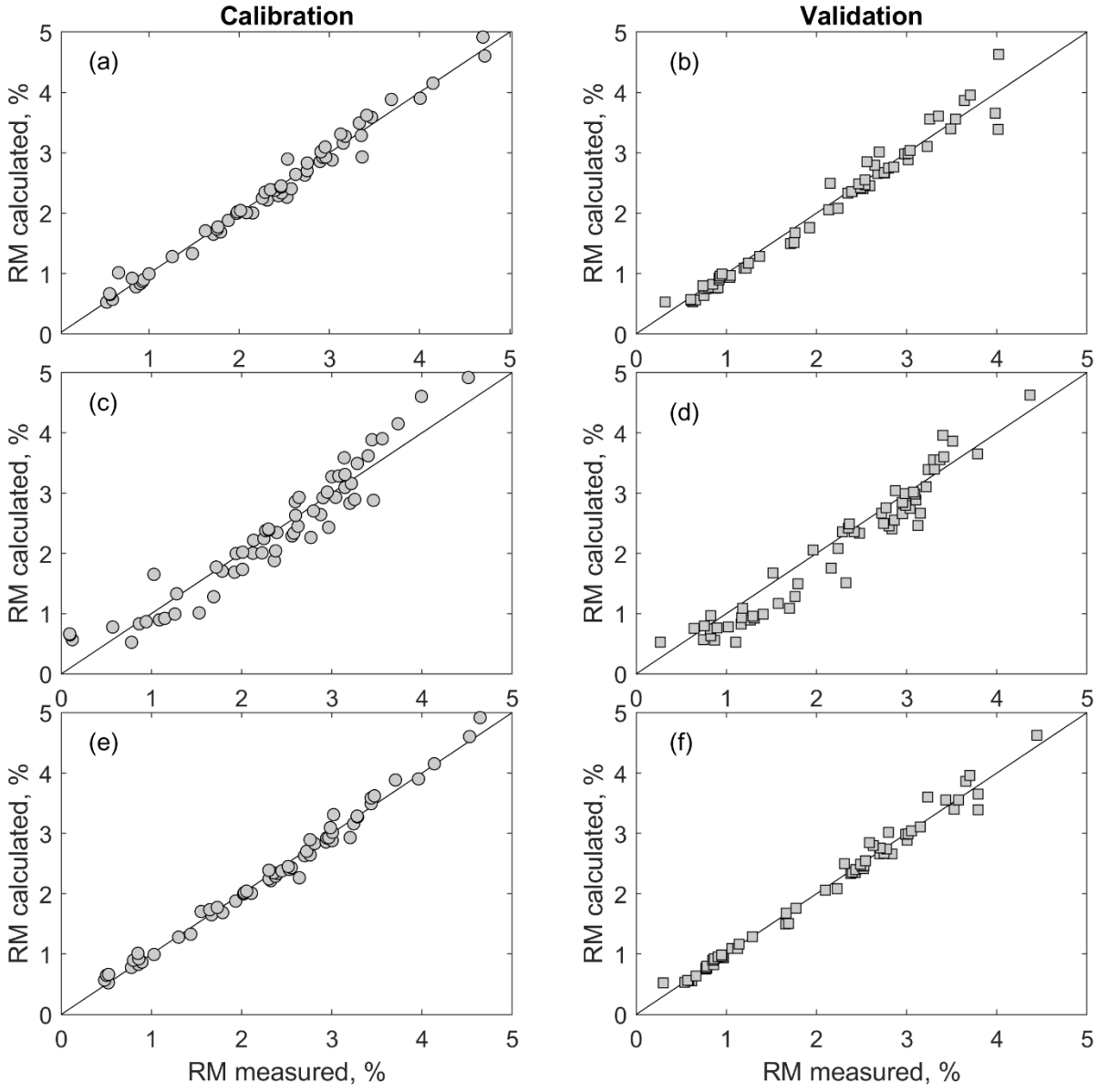


Figure 2. Score plot of dataset S6 obtained by PLS decomposition in the range 7100 - 4250 cm^{-1} of model WR (left) and in the range 5290 - 4785 cm^{-1} of model SR (right). The scores are displayed in the plane described by the 1st and 2nd PC and are ordered according to the RM content measured by KF analysis: RM lower than 2.0% (white circles), RM between 2.0% and 3.0% (gray triangles), and RM higher than 3.0% (black squares).

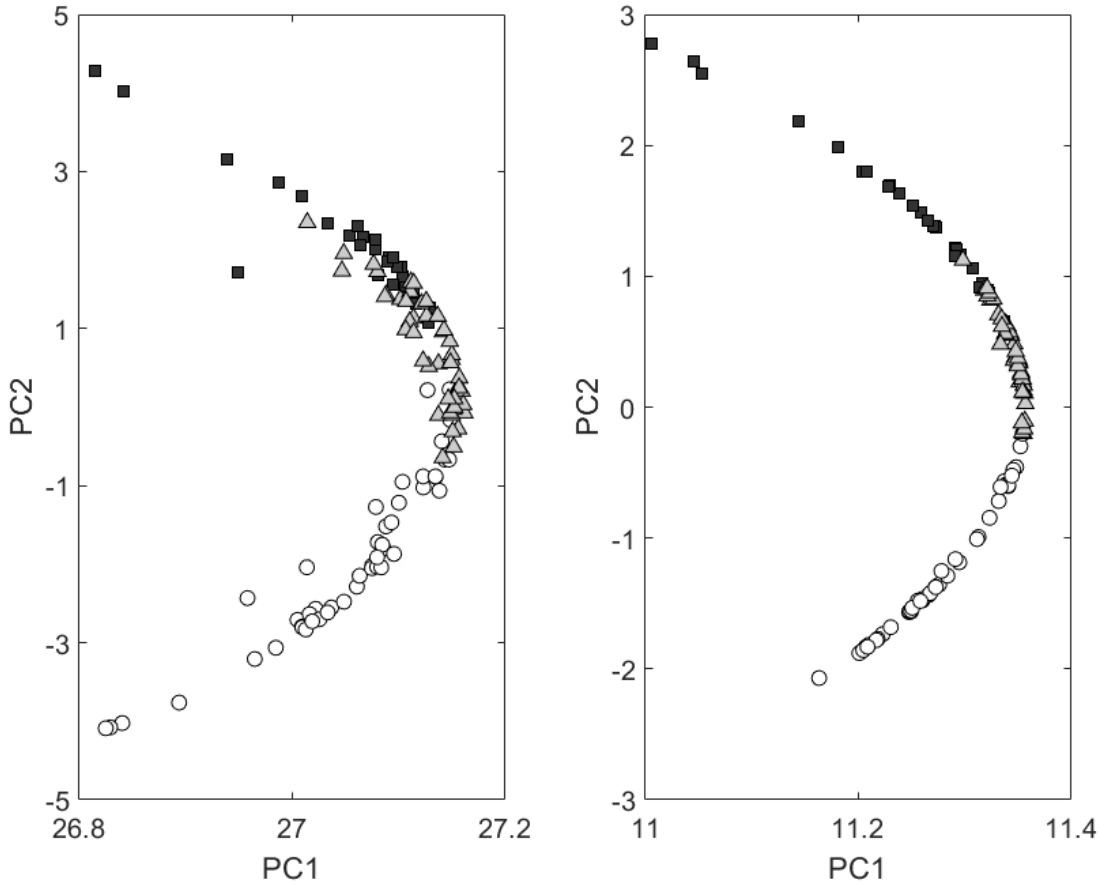


Figure 3. Analysis of dataset S6-H: (a) score plot with the scores of datasets S6 (white circles) and S6-H (gray circles) obtained by PLS decomposition in the range of model SR; (b) external validation of model SR; (c) and (d) calibration and validation correlation plots of the ad-hoc model H. Scores are displayed in the plane described by the 1st and 2nd PCs.

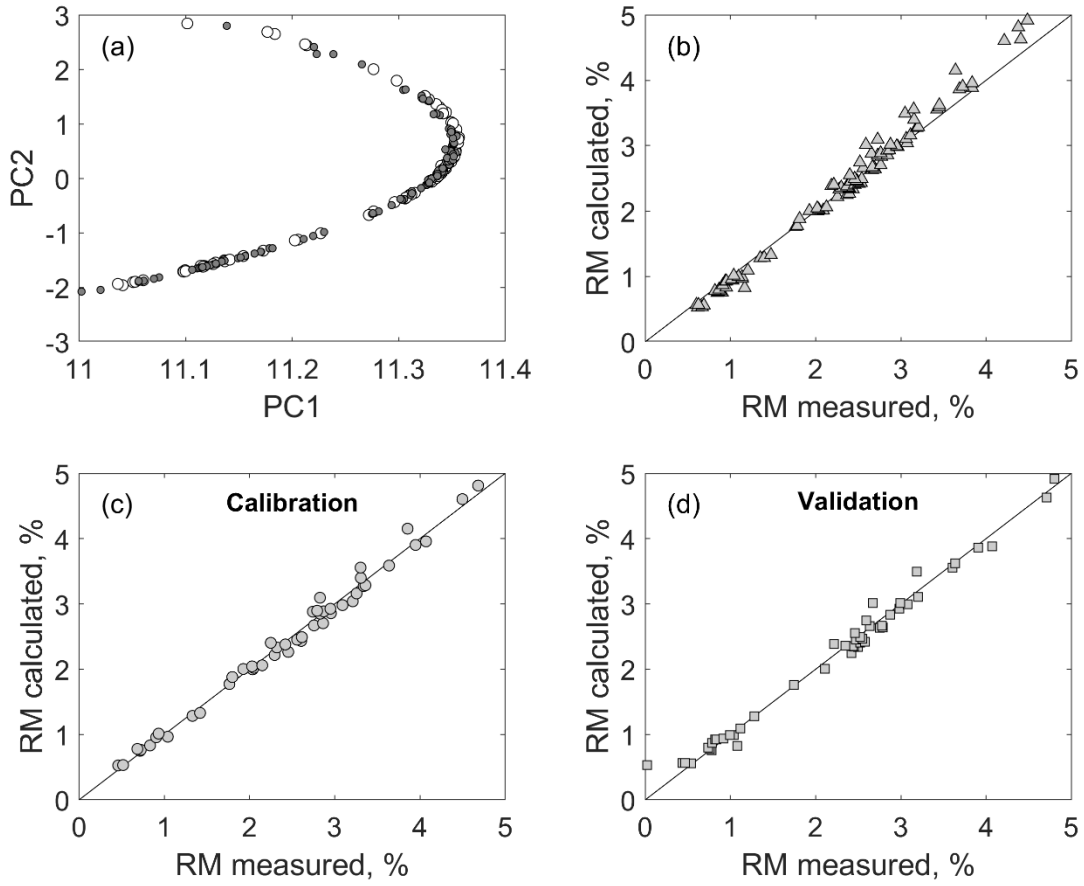


Figure 4. Analysis of dataset S9: (a) external validation of model WR; (b) external validation of model SR; (c) and (d) calibration and validation correlation plots of the ad-hoc model SF.

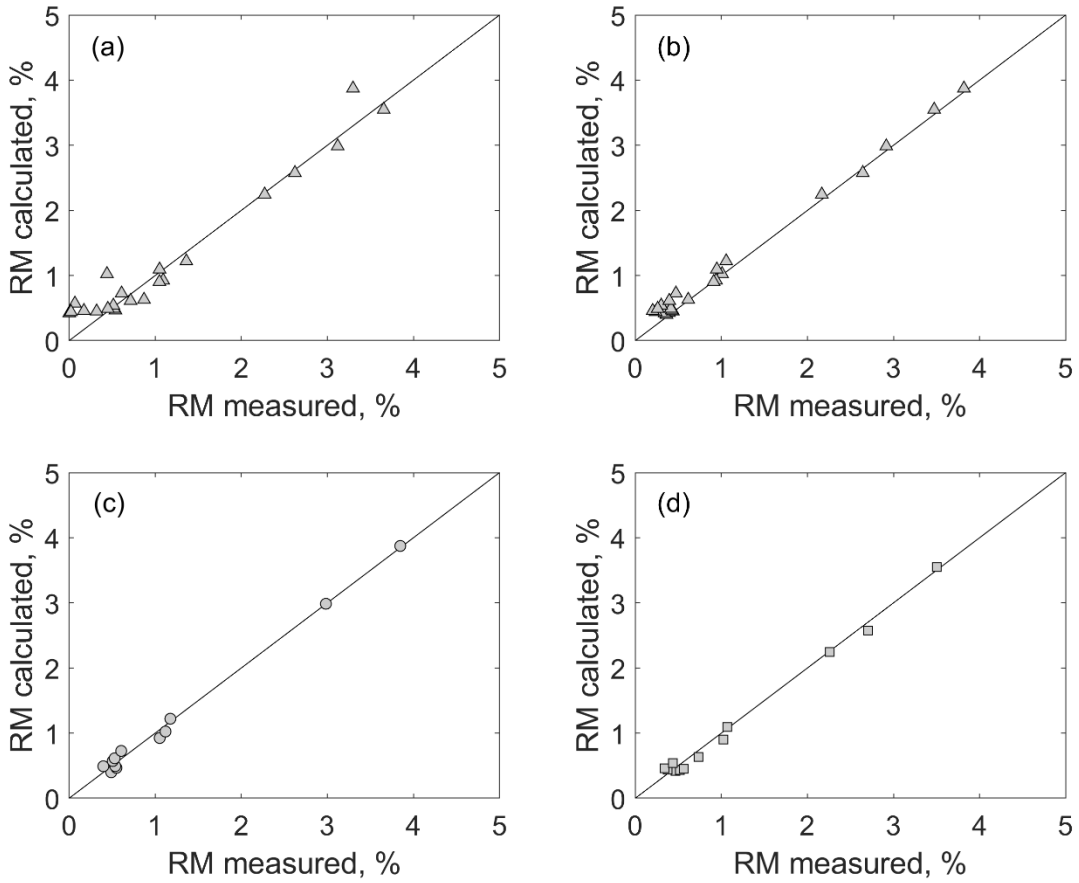


Figure 5. Analysis of dataset T6: (a) external validation of model WR; (b) external validation of model SR; (c) and (d) calibration and validation correlation plots of the ad-hoc model T.

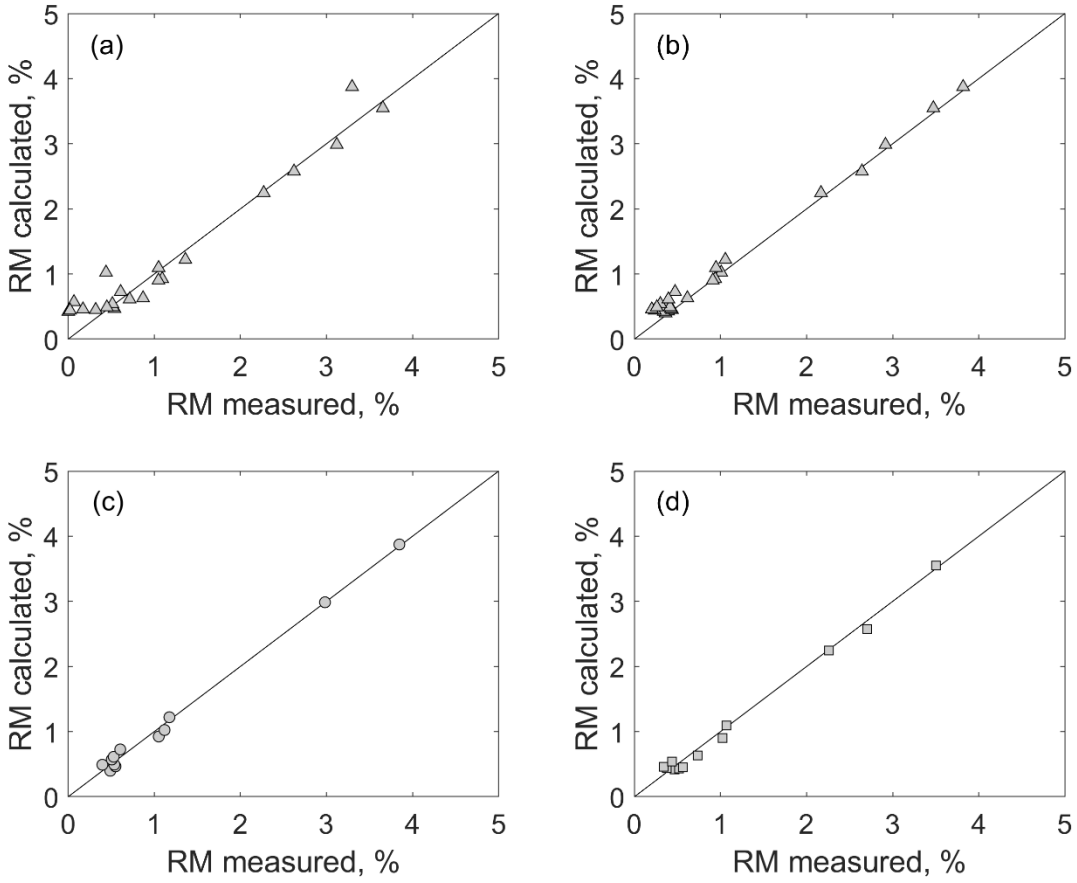


Figure 6. Analysis of sample set SA: (a) score plot obtained by the PLS decomposition in the range of model SR of the datasets S6 (white circles), SA07 (gray circles), SA14 (gray triangles), SA25 (gray squares), and SA50 (gray diamonds); (b) external validation of model SR with the datasets SA07, SA14 (white circles and triangles), SA25, and SA50 (gray squares and diamonds); (c) and (d) calibration and validation correlation plots of the ad-hoc model A. The scores are presented in the plane described by the 1st and 2nd PC.

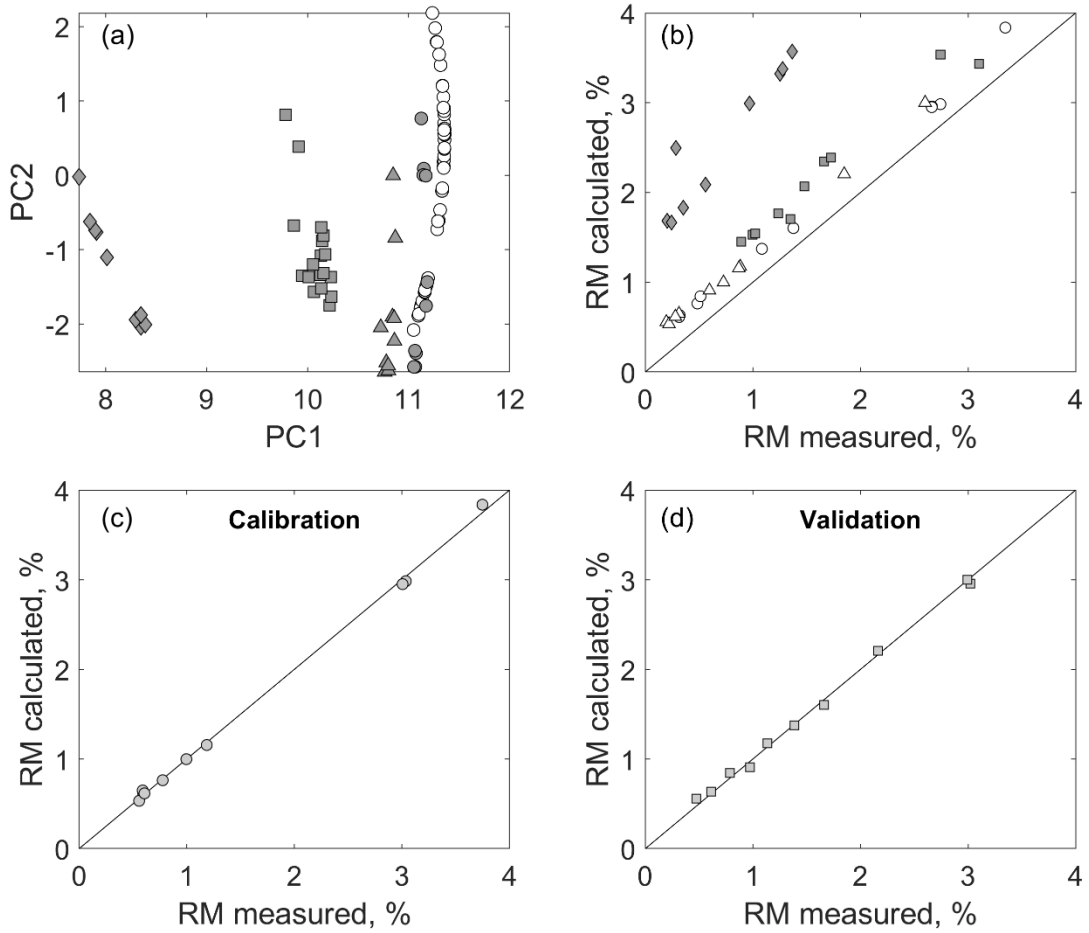


Figure 7. Analysis of dataset GL: (a) and (b) calibration and validation of the ad-hoc model G; (c) and (d) calibration and validation of the ad-hoc model G-LR. All the correlation plots are given as RM (%) calculated vs RM (%) measured.

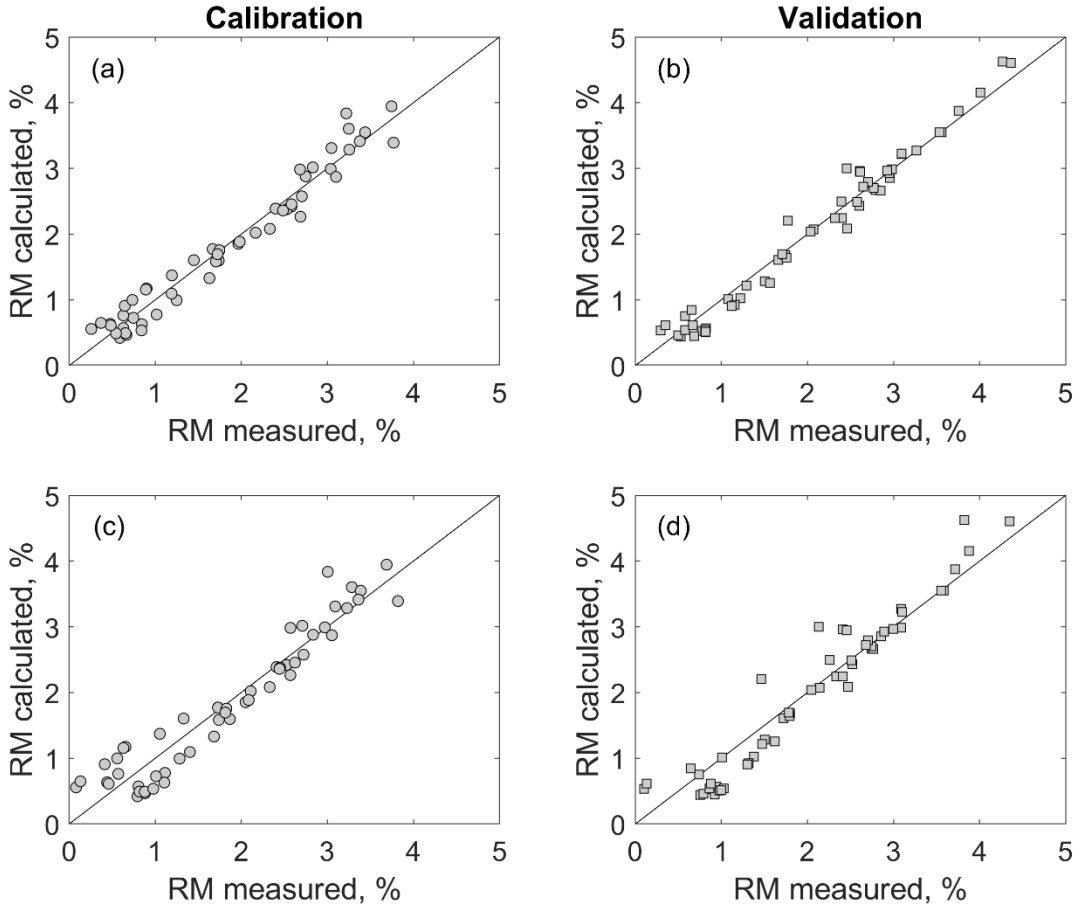


Figure A1. Acquisition of spectra: the NIR probe points the side wall or the bottom of a sealed vial, denoted, respectively, as vertical layout (above) and horizontal layout (below).

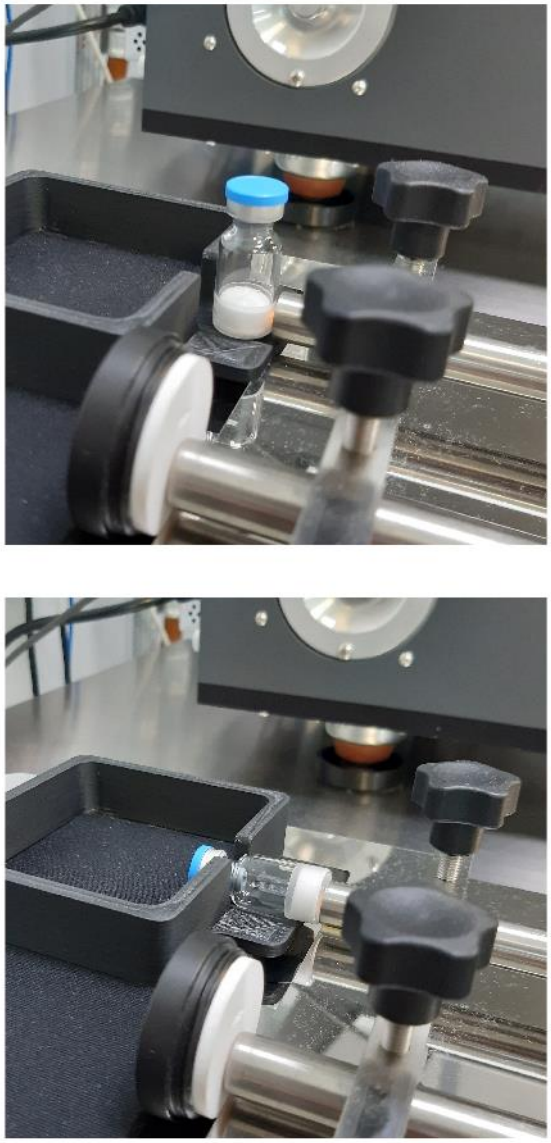


Figure A2. Spectral dataset S6 after pre-processing over the wavelength range 10000 - 4150 cm^{-1}.

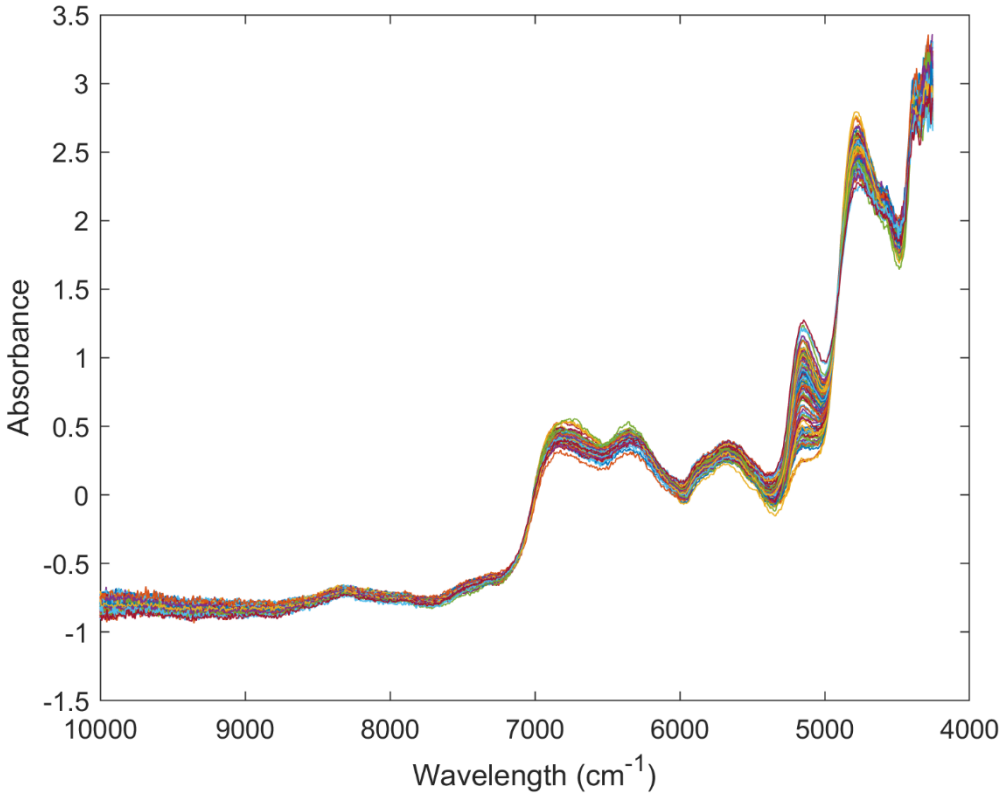


Figure A3. Loading plots of the PLS decomposition in the wavelength range 10000 - 4250 cm^{-1}, of the datasets (a) S6, (b) T6, (c) SA7 and SA14. In blue the loading of the 1st PC, in orange the loading of the 2nd PC.

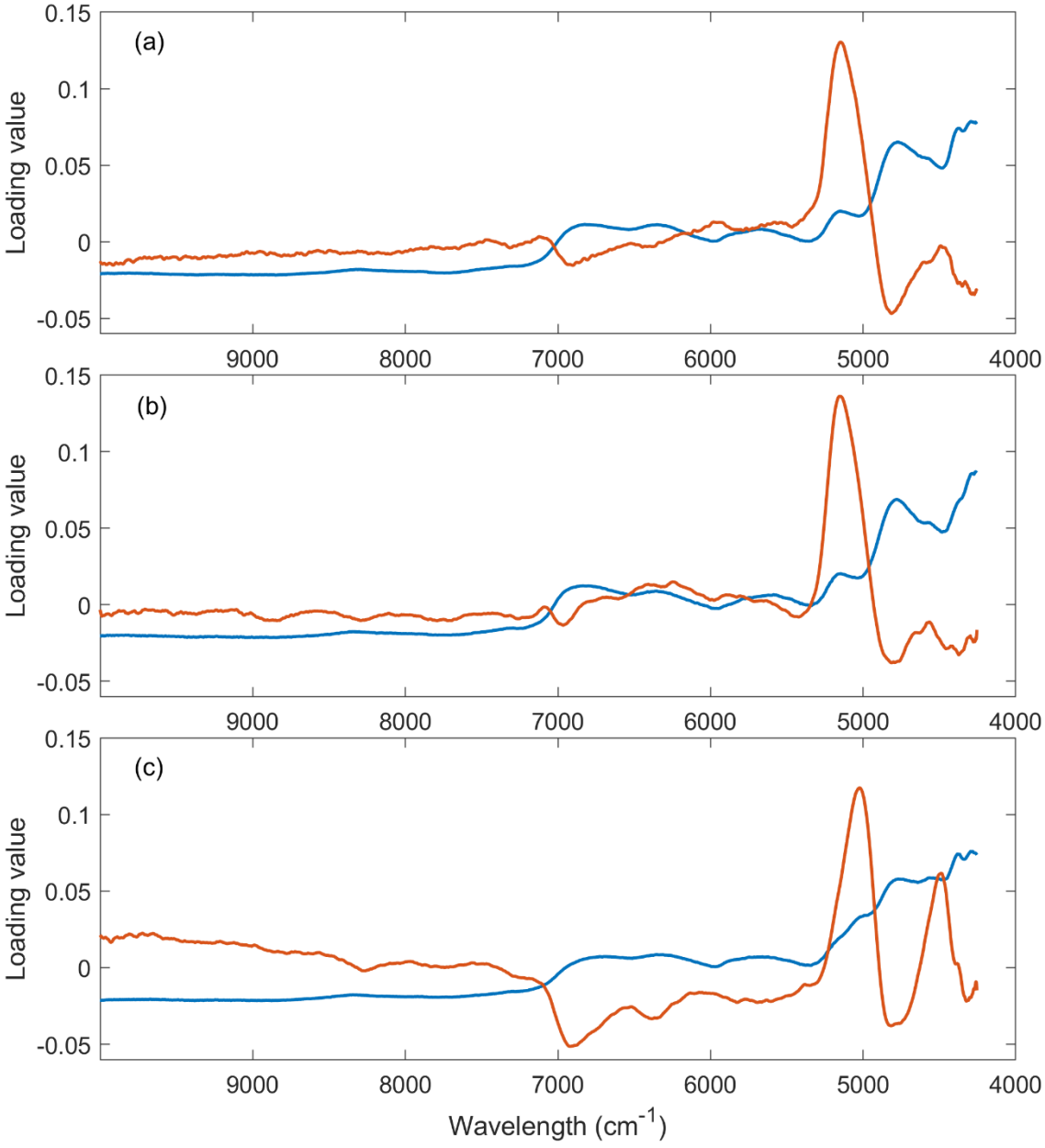
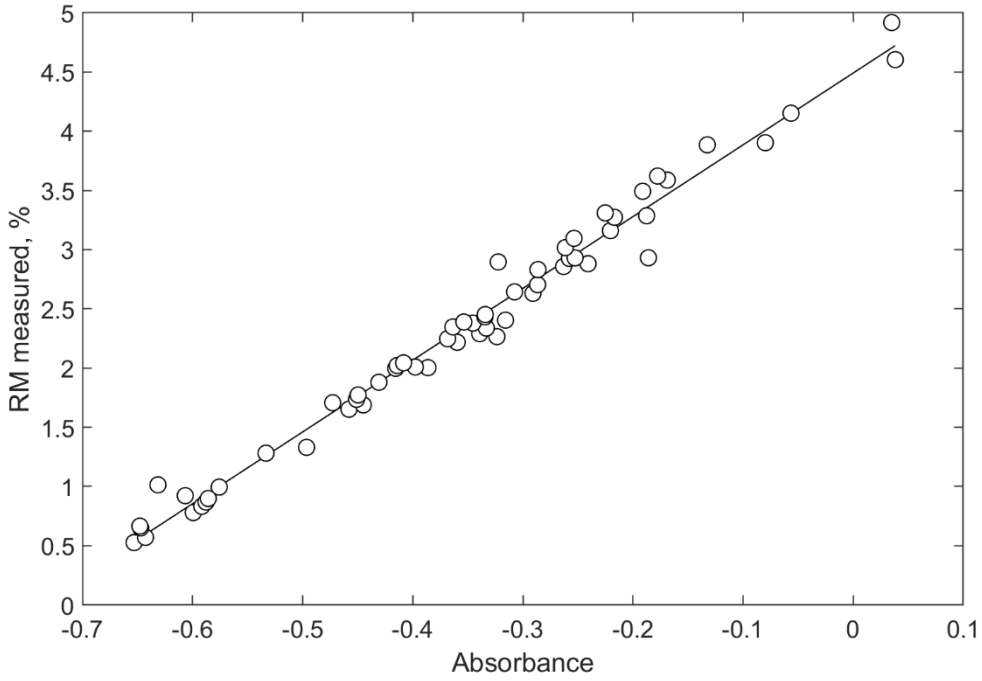


Figure A4. Curve of calibration of model LR (above) and of model G-LR (below), given as RM measured (%) vs absorbance. Negative values of absorbance are due to the SNV applied to spectra as pre-treatment. The solid black line represents the fitting curve.



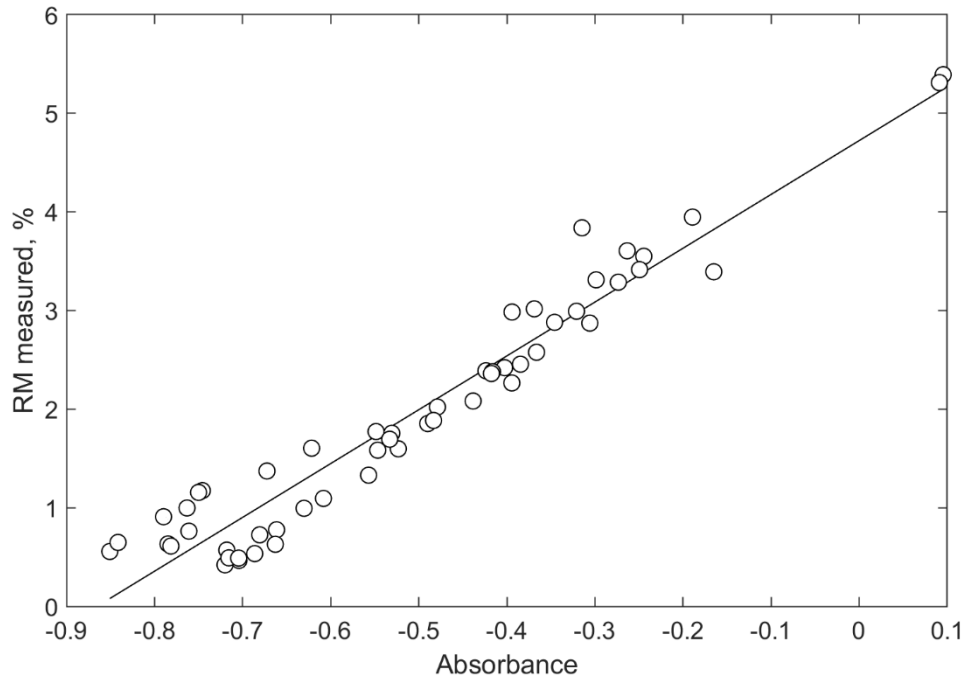


Figure A5. External validation of model SR with the samples of dataset VF.