Talanta Open 3 (2021) 100027

Contents lists available at ScienceDirect

Talanta Open

journal homepage: www.elsevier.com/locate/talo

Determination of the quality of metronidazole formulations by near-infrared spectrophotometric analysis

Abdoul Karim SAKIRA[a,*], Corenthin MEES[b], Kris De BRAEKELEER[b], Cédric DELPORTE[b], Josias YAMEOGO[c], Moussa YABRE[a], Touridomon Issa SOME[a], Pierre Van ANTWERPEN[b], Dominique MERTENS[b], Jean Michel KAUFFMANN[b]

[a] Laboratoire de Toxicologie, Environnement et Santé (LATES), Université Joseph KI-ZERBO, 03 BP 7021 03 Ouagadougou, Burkina Faso
[b] Unité de Pharmacognosie, Bioanalyse et Médicaments (RD3), Université Libre de Bruxelles, Campus de la Plaine, ULB CP205/05, boulevard du Triomphe, Ixelles, 1050 Bruxelles, Belgique
[c] Laboratoire de Développement du Médicament (LADME), Université Joseph KI-ZERBO, 03 BP 7021 03 Ouagadougou, Burkina Faso

ARTICLE INFO

Keyswords:
NIRS
Metronidazole
Quality control
PCA
SIMCA
PLS-R

ABSTRACT

In the quality control of medicines and the fight against the phenomenon of poor quality medicines, there is an urgent need for rapid and broad spectrum methods for screening these types of medicines. In the present work, we have used near infrared spectroscopy combined with multivariate data analysis to develop chemometric models for the classification and quantification of metronidazole in Burkina Faso pharmaceutical formulations. For this purpose, drug samples were collected in drugstores located in different Burkina Faso border zones. Four product classes were defined based on the national nomenclature: 3 classes for the generic drugs (C1, C3, and C4) and one class for the reference (C2) drugs. The exploratory analysis using PCA identified two clusters of drugs within class C1. Discrimination was confirmed by the developed and optimised DD-SIMCA model, with only one target class. The quality control of the samples from product class C1 was proven to be very satisfactory with specificities and sensitivities of 100%. The quantification models developed with the PLS-R method were successfully applied for the determination of the active ingredient content in the samples, with acceptable relative bias between 0.15 and 12.7 % with respect to the dose determined by the HPLC method. The RMSEC was estimated at 13.57 (R^2, 0.9937), the RMSECV at 18.07 (R^2, 0.9888) and the RMSEP at 13.69 (R^2, 0.9941).The models developed and the results obtained are promising for routine quality control of similar formulations of metronidazole.

1. Introduction

The quality of medicines remains a major determinant of the success of any health policy [1, 2]. Poor quality medicines can take several forms: total absence of active substances, incorrect dose of active substances, and presence of toxic derivatives in the drug composition and poor dissolution profile. Such formulations lead to therapeutic failures and expose consumers to various risks of intoxication [3]. In developing countries, the persistence of poor quality or falsified medicines is partly linked to the inadequacy of regulatory provisions and their application, but also to the absence and/or inadequacy of an effective and efficient system for monitoring the pharmaceutical market [4]. In Burkina Faso, very little referenced data are available on the quality of medicines, although inspection reports indicate the presence of illegal medicines in the licit circuit. The recent national surveys available, based only on organoleptic analyses, have revealed several quality defects with complaints recorded on the effectiveness of certain medicines (*Unpublished results*). The drug quality control system in Burkina Faso is organized around the National Public Health Laboratory, which carries out systematic import controls in the legal sales channel. It is based on a post-marketing surveillance plan targeting only antimalarials and the use of the basic Minilab screening technique and confirmatory tests using chromatographic methods. The other therapeutic classes are not covered, and therefore the physico-chemical quality profile of the latter is unknown. In the quality control of medicines and the fight against the phenomenon of poor quality medicines, there is an urgent need for rapid and broad spectrum methods for screening these types of medicines [1]. Instrumental analytical techniques for drug analysis include separation methods whose performance are established and included in the standards: High Performance Liquid Chromatography, Gas Chromatog-

* Corresponding author.
E-mail addresses: karim_sakira@yahoo.fr (A.K. SAKIRA), corenthin.mees@ulb.be (C. MEES), kris.de.braekeleer@ulb.be (K.D. BRAEKELEER), cedric.delporte@ulb.be (C. DELPORTE), josiasbyg@yahoo.fr (J. YAMEOGO), yabr.moussa@yahoo.fr (M. YABRE), pierre.van.antwerpen@ulb.be (P.V. ANTWERPEN), dominique.mertens@ulb.be (D. MERTENS), jean-michel.kauffmann@ulb.be (J.M. KAUFFMANN).

https://doi.org/10.1016/j.talo.2020.100027
Received 8 August 2020; Received in revised form 4 December 2020; Accepted 4 December 2020

raphy, Electrochemistry, and High Performance Thin Layer Chromatography [4, 5]. In recent years, however, there has been a growing interest in spectroscopic techniques with Near infrared spectroscopy (NIRS), Raman resonance spectroscopy, NMR and mass spectroscopy. Vibrational and near infrared spectroscopy in particular seem very well suited to our objectives because of the following advantages: fast analysis, almost no sample preparation, non-destructive, relatively modest cost (portable equipment) and adapted to the context of developing countries [5–10]. In addition to these benefits, there is no use of solvents or toxic reagents and no environmental impact. This study is focused on samples of metronidazole, an anti-parasitic antibiotic widely used for the management of diseases with sensitive anaerobic germs. In order to guarantee the distribution channel of the drug and verify the quality of the different approved classes, samples were collected in rural border areas that could be infiltrated by other parallel or illicit drugs. It is also to be recalled that generics are normally characterised by a pharmaceutical quality comparable to the corresponding genuine medicines, but they do not necessary contain the same excipients. From a spectral point of view, discrimination of genuines and generics is not always successful and certainly more difficult compared to discrimination between genuines and falsified [11–13]. The feasibility of a method to discriminate between different brands of metronidazole tablets collected in China, based on the use of Partial Least Squares Discriminant Analysis (PLS/DA) has already been proposed [14]. Over shorter wavelength ranges, *Zhao and al,* have also demonstrated that it is possible to monitor metronidazole powder samples using the artificial neural network (ANNs) [15]. Partial Least Squares Regression was also used on near infrared spectra recorded by transflectance measurements on six (06) dermatological formulations including one based on metronidazole for their quality control [16]. The near-infrared spectrum, difficult to interpret in general, is characterized by the overlapping of broad absorption bands resulting from different harmonic vibrations and their combinations [17, 18]. Extraction of relevant information requires coupling with multivariate analysis [1]. Conventional exploratory chemometric tools, classification tools, and quantification tools, such as Principal Component Analysis (PCA), Soft Independent Modelling of Class Analogy (SIMCA and DD-SIMCA) and PLS-R have been explored with the objective to develop chemometric models for the classification and quantification of metronidazole in Burkina Faso pharmaceutical formulations.

2. Materials and methods

2.1. Chemicals and drugs products

The reagents and standards used for the implementation of the HPLC reference method were: methanol (LiChrosolv^{R} Merck), Milli-Q water (18.2 MΩ.cm, 25°C) and the metronidazole standard (Purity ≥ 99% Sigma Aldrich).

Drug samples were collected in drugstores located in Burkina Faso different border zones. Taking into account the different product names registered by national regulations, four product classes have been established in order to verify their respective quality: Class 1 (generics C1), Class 2 (reference drug C2), Class 3 (generics C3) and Class 4 (generics C4).

Class 2 products are considered as a reference. Class 1, 3 and 4 are copies of class 2. The primary packaging of all these products consists of transparent PVC/Alu blisters. Class 1 tablets have all the same physical aspect and are sold in bulk without secondary packaging and without patient information leaflet. These products are subdivided into two subclasses 1-A and 1-B because they differ in their excipient composition (information obtained from national pharmaceutical regulations). Class 2, 3, and 4 drugs are more expensive and are sold in the pharmacy in their secondary carton-type packaging with a patient information leaflet. The physical aspect of Class 2, 3 and 4 tablets are comparable with those of Class 1: round, biconvex and white. The other characteristics of the products are presented in Table 1 below.

2.2. Sample preparation for NIRS

2.2.1. Samples for classification

Spectra were recorded directly on the tablets through the blisters using a benchtop Infrared Spectrophotometer with an optical fiber probe. Twenty spectra were collected on twenty different tablets for each sample. All the blisters used were transparent to NIR irradiation, and preliminary measurements showed that there was no noticeable difference between the spectra collected with and without blisters.

2.2.2. Samples for quantification

A sample matrix was prepared by mixing equal amounts of powder of all the classes of samples (C1, C2, C3, and C4). Another matrix made of representative amounts of all the excipients was used as diluent (matrix reconstituted without metronidazole). The calibration samples were prepared as follows: an amount of sample matrix equivalent to a dose of 100 mg, 200 mg, 300 mg and 400 mg metronidazole was diluted with the excipient matrix to a mass of approximately 790 mg. For the highest concentration level of 600 mg metronidazole a mass of approximately 790 mg sample matrix was used. Calibration samples were homogenized during 45 min using a tubular type homogenizer. The calibration samples were prepared in triplicate. The spectra were collected through transparent vials using the optical fiber probe module which allows measurements in diffuse reflectance mode. Three spectra were measured for each content level. For metronidazole 500 mg test samples, the spectra were directly collected through the blister without any preparation. For metronidazole 250 mg samples, however, a small preparation step was necessary to standardize the mass of the samples; it consists in diluting the sample powder in a matrix of excipients to obtain an average mass of approximately 700 mg. After this preparation, the spectra were collected on this prepared powder.

2.3. Near infrared spectroscopy

All samples were scanned with the Frontier MIR/NIR Spectrometer^{TM} (PerkinElmer^{TM}, USA). Spectra were collected according to the diffuse reflection mode. Data acquisition was performed with Spectrum software. The spectra have been digitized and made compatible with MATLAB software (**R**2018**b**, Mathworks). The instrumental parameters have been set as follows: measuring range 10 000 to 4000 cm^{-1}, spectral resolution (8 cm^{-1}), and number of scans (32 scans). After eliminating irrelevant information areas, the 9012 cm^{-1} to 4022 cm^{-1} domain was selected for qualitative analysis. For the development of the calibration model, we considered an absorption zone specific to metronidazole (from 6062 cm^{-1} to 5840 cm^{-1}).

2.4. HPLC reference method for quantification

The assay of the samples collected (500 mg and 250 mg metronidazole tablets) and the assay of the calibration samples was performed using the method described in the monograph “Metronidazole Tablets” of the United States Pharmacopeia (USP).The reversed-phase Liquid Chromatography (LC) method was carried out on an Agilent Technologies system (Infinity 1290) equipped with a PDA detector and a Zorbax C8 (150 × 4.2 mm 3.5μm) column, kept at 15°C. Chromatography of the samples was performed in an isocratic mode with binary mobile phase methanol/water (20:80 % v/v) at 1 mL per min. The UV detection was at 254 nm. The injection volume was 10 μL. All the sample solutions were prepared in mobile phase for a nominal concentration of 0.5 mg/mL metronidazole. For this purpose, 10 tablets were grounded and a suitable amount of the powder was stirred for 30 minutes with mobile phase and finally diluted to 100.0 mL with mobile phase. Samples were filtered through a 0.45 μm membrane filter. The filtered samples were transferred into glass vials for injection into the chromatograph. The standard solution was prepared in the mobile phase to contain 0.5 mg/mL of metronidazole. The results have been interpreted according to USP

Table 1
Characteristics of the samples under study

Class	Galenic form	Diameter (mm)	Thickness (mm)	Dose (mg)	Expiry date	Ratio metronidazole/tablet weight (% m/m) calculated	Batches
Class 1-A (generic)	tablet	13	3	500	04/2021	75.8	9
	tablet	7	1	250	04/2021	78.2	
Class 1-B (generic)	tablet	13	3	500	01/2021	90.5	
Class 2 (reference)	coated tablet	12.5	2.5	500	01/2023	73.2	2
Class 3 (generic)	tablet	13	3	500	11/2020	84.2	1
Class 4 (generic)	tablet	12	3	500	04/2022	68.4	1

standards, with content compliance between 90 and 110% of the label claim.

2.5. Chemometrics methods

2.5.1. Data pretreatment

Appropriate pretreatments of the near-infrared spectra were necessary to eliminate irrelevant information which is mainly due to differences in physical characteristics of the samples. These differences were responsible for uncontrolled baseline variations.

Two preprocessing methods were applied: standardization followed by derivation. The combination of pretreatments is commonly used in chemometrics [13, 19–22]. First the raw spectra were standardized using the Standard Normal Variate transformation (SNV). SNV was applied on each spectrum and consisted of the subtraction of the mean and division by the standard deviation of the spectral reading. Spectra were standardized to have a mean of zero and standard deviation of one. After SNV, the second derivative was applied using the convolution method of Savitzky and Golay [23] with a second order polynomial and a window size of 17. The second derivative enhanced the small spectral differences which were characteristic for each of the samples studied.

2.5.2. Exploratory data analysis by Principal Component Analysis (PCA)

Principal Component Analysis (PCA) reduces the dimensions of the original data space by using a smaller and more efficient abstract space of latent variables (PCs), where the data can be displayed while keeping the same information as the original space [22, 24]. PCA is commonly used to perform an unsupervised exploratory data analysis [25]. Similarities and differences between the spectra are observed by their clustering tendency. Each spectrum is visualized as a point in a two or three dimensional plot defined by the selected principal components (PCs) [13]. Usually the first three principal components are the most informative and explain the variance in the data. PCA was applied here to test whether the spectral data could differentiate between the types of samples defined based on the knowledge of the distribution channel and the results obtained with HPLC. The clusters formed in the PCA plot allow defining classes for the supervised classification [26].

2.5.3. Selection of a test set for external validation of the classification and calibration models

The duplex algorithm was used to select seventy-five per cent of the spectra to build the model and 25 per cent of the spectra to test and validate the model. The two spectra with highest Euclidean distance were selected and placed in a first set. The next two spectra furthest from each other were put in a second set (test set). This procedure continued until 25 percent of the spectra were selected in the test set. The first set and the remaining not selected spectra were used to generate the training set [27].

2.5.4. Soft Independent modelling of class analogy (SIMCA and DD-SIMCA)

The main goal of authentication is to determine whether a drug product is what it is declared to be and that its quality is conform to the regulatory requirements. Soft independent modelling of class analogy (SIMCA) is a class modelling method or one-class classifier commonly used in chemometrics for authentication purpose. The original version of SIMCA, proposed by S. Wold [28] has numerous modifications mostly related to the way of constructing the acceptance boundaries. A well-known modification is DD-SIMCA.

The original SIMCA algorithm is a supervised classification technique referred to a class-modeling approach. It uses samples with known origin (training samples) to derive a classification rule which allows classifying new samples (test samples) with unknown origin in one of the classes [26, 29]. SIMCA considers different classes which are modelled individually by a separate principal component (PC) model. The number of PCs was selected for each class using a leave-one-out cross-validation procedure. In order to determine the limits around samples belonging to a class, two critical values were taken into account: The Euclidean distances towards the SIMCA model assessed by the residual Q statistic and the Mahalanobis distances measured in the space of the scores and assessed by the Hotelling T^2 statistic [26, 30, 31]. The scores and the loadings from the PCA model were used to calculate the position of a new sample. Then, the new sample was classified to the most similar class of the model using a confidence limit set at 95% [31, 32]. The SIMCA model was built on all the drug products studied. The model aimed to classify new samples into one of the product classes.

The Data Driven SIMCA algorithm (DD-SIMCA) has also been used to create a classification model using only one target class [33, 34]. Similarly to the original SIMCA, first a Principal Component Analysis (PCA) was applied to the training data from the target class [29]. In a second step, scores distances (SD), hi, and orthogonal distances (OD), vi, were calculated. The SD represents the position of a sample within the score space, and the OD characterizes the distance of the sample to the score space [29, 35]. Each type of distance is modelled using a scaled chi-squared distribution instead of the Fisher or Hotelling T^2 statistic used in the original SIMCA models. The two scaled chi-squared distributions were used to define an acceptance area (critical distance) for a given type I error or level of confidence 1-α [29, 34]. The data-driven distribution parameters are the mean values v_0 and h_0, and the numbers of the degrees of freedom N_h and N_v for the SD $\{hi\}$ and OD $\{vi\}$ respectively [29]. The acceptance area was used to predict new samples and classify them in two categories: 'regular' samples which are attributed to the target class and the 'external' samples which were located out of the acceptance area and are considered as non-members of the target class [36].

The quality of the DD-SIMCA model depends on the selection of the number of PCs for the target class and the critical α-value. By default, the α-value is calculated so that all samples in the training set are inside the acceptance area. The classification performance of the DD-SIMCA model is expressed by the sensitivity and specificity, which were calculated with the Matlab GUI tool for data driven SIMCA [36].

2.5.5. Partial least squares regression

Partial Least Squares regression (PLS-R) is a projection technique in which latent variables are defined as linear combinations of manifest variables. These combinations reflect the highest (PLS factor 1) or highest remaining (PLS factor 2, 3,…) co-variance between the data (NIR spectra) and a continuous response such as the concentration. Quality measures of the PLS-R models are: the root-mean-square error of calibration (RMSEC), the root-mean-square error of cross-validation (RM-

SECV), the root-mean-square error of prediction (RMSEP) and the coefficient of determination (R^2). The RMSEC is a measure of how well the model fits the calibration data, when all calibration samples are included in the model [37]. The root-mean-square error of cross-validation (RMSECV) measures the ability of the model to predict samples of the calibration set which were not used to build the model. For leave-one-out-cross-validation, samples are successively deleted from the calibration set. Calibration performance is tested by predicting the deleted sample and comparing the predicted values with the reference values. The root-mean-square error of prediction (RMSEP) is an estimate of the model's prediction ability for an independent set of samples (test set) which were not included in the calibration model. The optimal number of PLS factors in the calibration model is selected by leave-one-out cross-validation and corresponds with the model with the smallest RMSECV value.

The calibration model is supposed to be reliable and stable when cross validated (RMSECV) and fitting results (RMSEC) are similar and comparable, which means that the prediction performance is not influenced by samples taken out of the calibration set during cross-validation. The true predictive performance of the calibration model (RMSEP) should confirm the performance achieved by internal validation (RMCECV) [38].The coefficient of determination, R^2, is calculated between the measured and predicted values. The values for R^2 should be as close as possible to 1.0 [37, 39]. After the establishment of the optimal PLS-R model, the method was validated by estimating the classical performance criteria of specificity, linearity, repeatability, and trueness.

2.6. Software

Data treatments were performed by using Matlab version R2018b (The Mathworks, Natick, USA). The algorithm of PCA, Duplex, SNV, Savitzky-Golay, and PLS were part of the ChemoAC toolbox (Freeware, ChemoAC Consortium, Brussel, Belgium, version 4.1). The toolbox for SIMCA was downloaded from the Matlab Central [40]. For DD-SIMCA - A Matlab GUI tool for data driven SIMCA was applied [36].

3. Results and discussion

3.1. Identification and quantification of metronidazole with HPLC

Using the HPLC method described in the monograph "Metronidazole Tablets" of the United States Pharmacopeia (USP), the active pharmaceutical ingredient (API) was identified in the 13 samples studied. All products complied with the 90 -110 % requirement for the metronidazole content with a recovery range between 95 % and 105 % with respect to the label claim.

3.2. Qualitative analysis with NIR

Preliminary studies have isolated a spectral band specific to metronidazole (Fig. 1). In this spectral band (6062 - 5840 cm^{-1}), no influence from excipients, blister package or sample vial was observed. The intensity of this spectral band increases with the concentration of metronidazole in the calibration samples (Fig. 2 and Fig. 3). Superimposable profiles of metronidazole spectra were observed in this zone, depending on whether the measurement was made directly through the blister pack, directly on the bare tablet or on a sample powder through a vial (Fig. 4).

3.2.1. Unsupervised analysis

The score plots obtained with different data pretreatment procedures were compared and the best score plot was obtained with SNV followed by the second derivative. The three first principal components (PC1 = 73.06 %, PC2 = 18.42 % and PC3 = 3.99 %) explained more than 95 % of the total variance. Fig. 5 shows the score plot for the spectra of the studied samples in the space spanned by the first (PC1) and third (PC3) principal component. Three clearly separated clusters were observed: one for the 2 batches of the reference product C2, one for the batch of product C3 and one cluster for the products C1 (9 batches) and C4 (1 batch). The same clusters were observed in the PC1-PC2 and PC2-PC3 score plots shown in the supplementary data. HPLC analysis confirmed that all products studied contained between 95–105% of the labelled amount of metronidazole. The ratio of metronidazole/tablet weight was between 68 % for C4 products and 90% for C1 products. The discrimination between the products C2, C3 and C4 was due to differences in excipients or can be due the blister presence in C2, C3 and C4 or the coating material present in of the C2 tablets and not in the other tablets: the C3 products contained talc which gave a typically sharp peak at 7185 cm^{-1} in the NIR spectra. Talc was not present in the other products studied.The score plot after PCA on the spectral data of C1 generic samples showed a homogeneous cloud of points in the space spanned by the first two principal components which explain more than 70% of total variance (PC1 = 65.3% and PC2 = 9.66%). Fig. 6 shows the score plot in the space spanned by PC1 and PC3: two clusters were observed which were clearly separated according to PC3. The discrimination between the samples can be attributed to differences in excipients or production area: one cluster corresponded to samples of China and the other to samples from India.

NIR spectroscopy combined with PCA showed only a faint clustering tendency between the four product classes studied. The discrimination between products was mainly due to qualitative and quantitative differences in excipients, which represented maximum 30% of the tablet weight. The clustering tendency of the NIR spectral fingerprints confirmed the HPLC results: there were no substandard or falsified products collected in the Burkina Faso rural border areas. The NIR spectral profiles of the studied products were very similar and mainly characterized by the presence of the active pharmaceutical ingredient metronidazole.

3.2.2. Supervised analysis

SIMCA classification models were built using the NIR spectra pretreated with SNV followed by the second derivative. Based on the clustering tendency observed in the PCA score plots, four classes were initially defined, one for the genuine (C2) and one for each of the generics C1, C3 and C4. From a regulatory point of view, there are two classes of medicines, namely reference or genuine medicines and generics. Generics are copies of originators with a mandatory requirement of identical qualitative and quantitative composition of the active substance(s). In addition, and compared to the reference, similarity of bioequivalence, and safety parameters should also be demonstrated for generics. SIMCA models were built for classification by drug product. The main goal of the SIMCA model was the authentication of the drug products to be sure that its quality conforms the regulatory requirements.

The Duplex algorithm was used to select 25% of the samples for the test set (external validation), the 75% remaining samples formed the training set to build the model. For each class 10 PCs were selected using leave-one -out cross validation. The smallest number of PCs was selected for each class. The correct classification rate or accuracy obtained for the training samples was 90.3%. The predictive performance of the model for the test set was almost identical with a correct classification rate of 89.0%. Detailed analysis of the results for the test set showed that all samples in class 1 (C1) were well identified. Reference samples in class 2 (C2) were nine times out of ten well ranked and one time out of ten ranked as generic C1. Class 3 (C3) samples were well identified and were not confused or assimilated to any other class. Samples in class 4 were 4 times out of 5 assimilated to C1 and 1 time out of 5 assimilated to C2.

The NIR spectral profiles of Class 4 (C4) samples were very similar to those of class 1 samples as shown in the PCA score plots. Therefore Class 4 and Class 1 samples were combined in one class. A SIMCA model with three classes was built: class 1 and class 4 drugs, class 2 reference drugs, and class 3. The model established with 4 PCs for each class was characterized by a better correct classification rate of 97.4 %. The success rate for prediction was estimated to 100%. The C4 samples are not

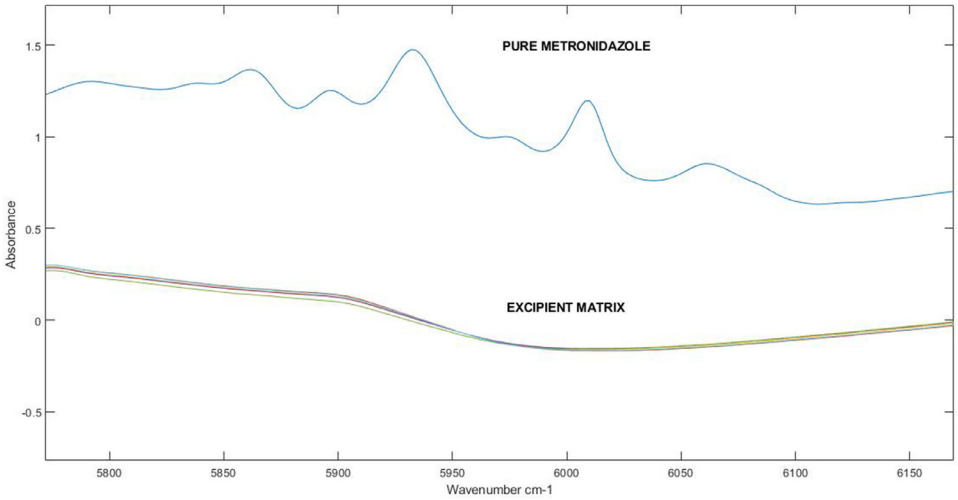

Fig. 1. NIR spectrum of pure metronidazole superimposed on the spectrum of the excipient matrix between 5850 and 6200 cm^{-1}.

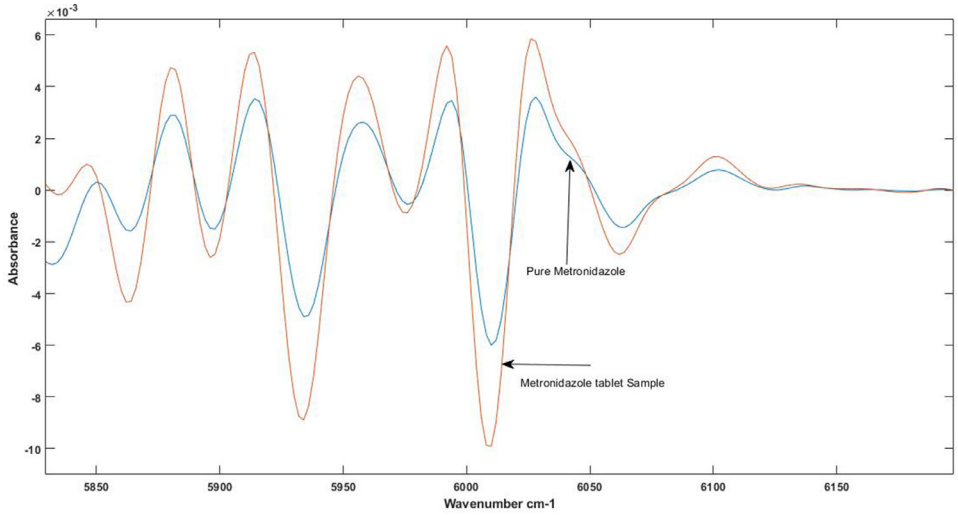

Fig. 2. Superimposed pre-processed NIR spectra (SNV+2nd order derivative) of pure metronidazole and metronidazole tablet sample.

Table 2
Original SIMCA model parameters.

Parameters	PC Number	CCR (%) * for training set	CCR (%)* for test set
Ratio of test set	2/3 – 2/3	4 or 5	100	100
	1/4 – 3/4	4	98.1	100
	1/4 – 3/4	7	100	100

*to training set

ccr = correct classification rate

superimposed on the reference product, but rather on the generic drugs C1.

PCA analysis on the spectral data of C1 products showed two separated clusters: one for products of India and one for products of China. The original SIMCA and its variant (DD-SIMCA) were applied to classify the C1 product according to the manufacturer's origin. Several SIMCA classification models have been established by varying the number of PCs for both classes, and the number of samples in test and training set (see Table 2). A correct classification rate of 100 % for test and training samples was obtained with the original SIMCA method.

The classification was confirmed by the DD-SIMCA algorithm. The target class of C1-B generics was modelled with 5 PCs and the boundary was set using $\alpha = 0.0001$. The DD-SIMCA classification performs with a sensitivity and specificity of 100% (see Fig. 7).

It appeared thus from these results, that SIMCA models can be used for the authentication of drug products with very similar NIR spectral profiles. First the original SIMCA approach can be applied to develop a model with all product classes: C1A, C1B, C2, C3 and C4. If the classification parameters showed that some classes coincide (C1-A, C1-B and C4) these classes could be combined to create a new SIMCA model with less classes. In a second stage, for products which were situated in the same class, a DD -SIMCA model could be applied using one of the products as target, such as C1-B products in this study. In this way one can discriminate between drug products with very similar spectral profiles: C1-B was clearly separated from C1-A samples.

The same strategy could also be applied to discriminate C1 products from C4 products using C4 products for instance as target class.

3.3. Quantitative analysis with NIR

The NIR spectra for samples were pretreated: SNV followed by second derivative and a characteristic spectral zone was selected, where the absorbance was very well correlated to the metronidazole content: 6062 cm^{-1} - 5840 cm^{-1} (see Fig. 3).

It was shown that the spectral band specific to metronidazole (6062 - 5840 cm^{-1}) was not influenced by the nature of the sample container (Fig. 4). This zone has therefore been exploited for the construction of a

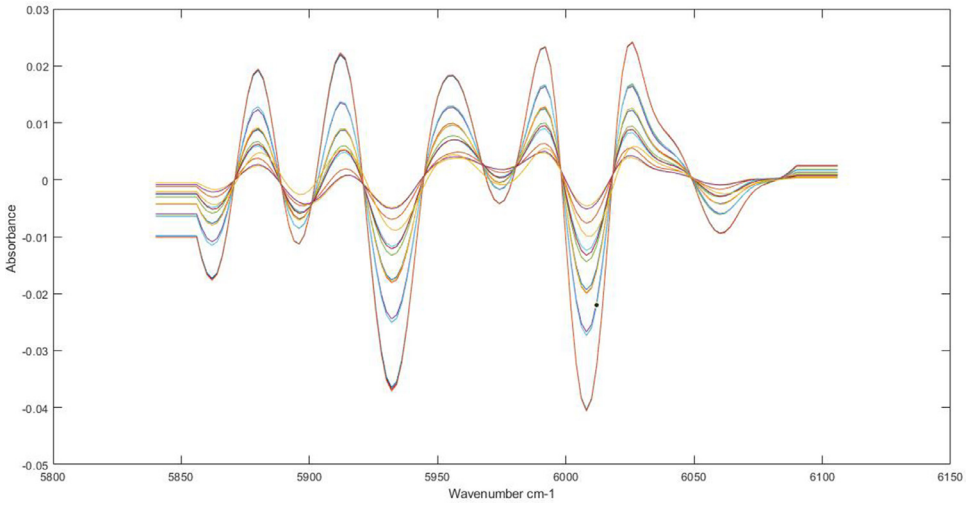

Fig. 3. Superimposed pre-processed NIR spectra (SNV+2nd order derivative) of metronidazole at 100, 200, 300, 400 and 600 mg levels.

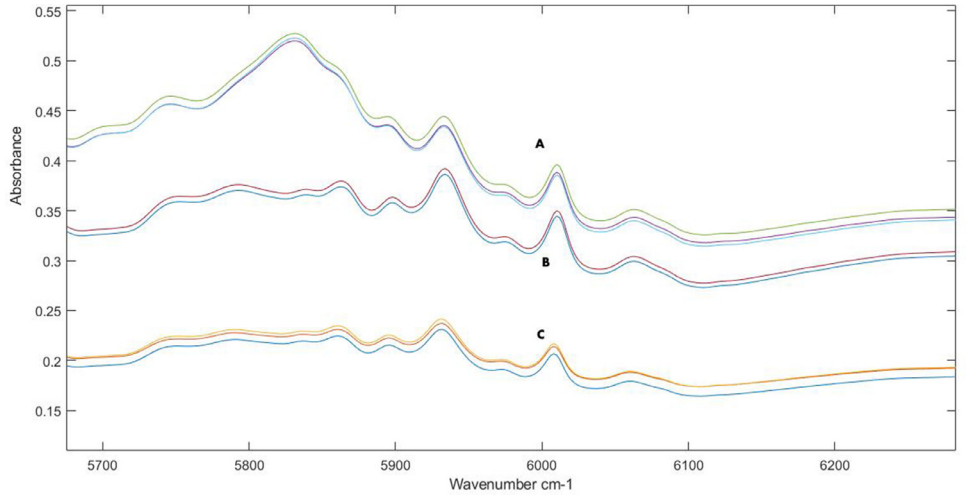

Fig. 4. Overlay spectra of metronidazole samples recorded under the following conditions: A, Direct measurement of samples in blister packs without preparation - B, Spectrum measurement directly on the bare tablet - C, Spectrum measurement through the vial (metronidazole powder sample).

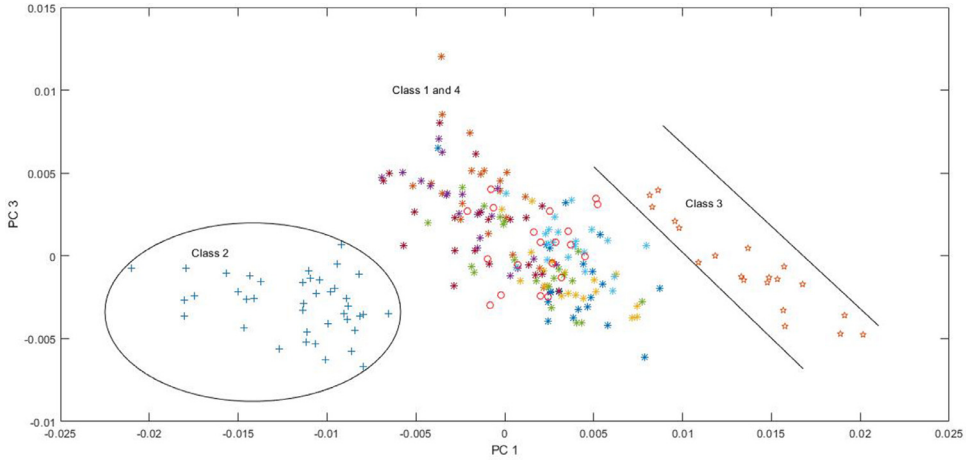

Fig. 5. PCA (PC1/PC3) score plot obtained with the NIR spectra for the metronidazole products studied after SNV followed by the second derivative.

quantitative model applied to the direct determination of metronidazole contents.

3.3.1. *Partial Least Squares regression*

A Partial Least Squares regression model (PLS-R) for the 500 mg and 250 mg metronidazole tablets was constructed from a set of 15 calibration samples: 5 content levels and 3 samples per content level. The nominal values for the content levels were: 100, 200, 300, 400 and 600 mg metronidazole. Each calibration sample was measured with NIR in triplicate meaning that in total 45 spectra were used to construct the calibration model. The concentration of the calibration samples was determined with the HPLC method described in the monograph “Metronidazole Tablets” of the United States Pharmacopeia (USP). Several PLS regression models were obtained and tested with different performance

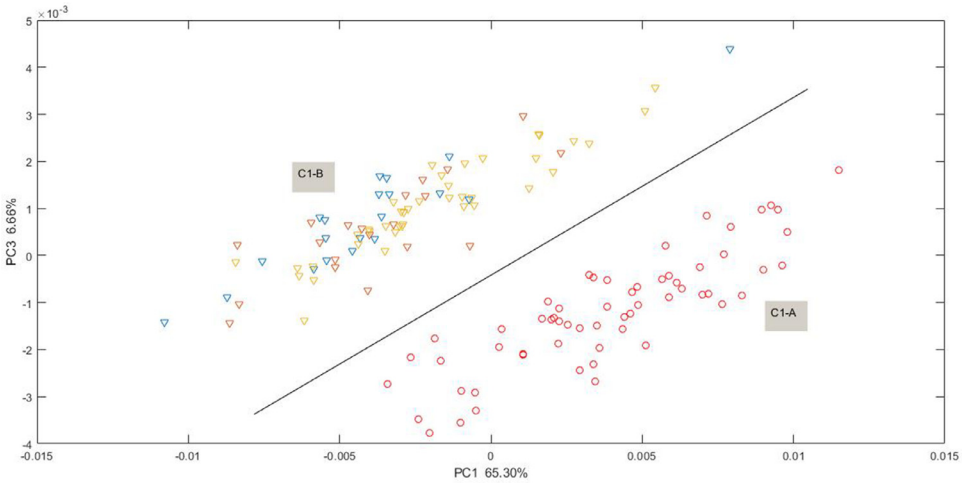

Fig. 6. PCA score plot obtained with the NIR spectra for the C1 products after SNV followed by the second derivative.

Table 3
PLS-R model parameters calculated according to the selected bands.

Range cm^{-1}	nLV	RMSEC	R^2	RMSECV	R^2	RMSEP	R^2
5980 - 5900	5	19.87	0.9797	26.15	0.9649	14.08	0.9971
6020 – 5850	4	15.94	0.9563	20.34	0.9777	21.32	0.9776
5988 - 5888	4	14.75	0.9924	17.55	0.9893	18.43	0.9934
	9	12.95	0.9942	17.38	0.9895	21.27	0.9899
6004 - 5872	4	13.57	0.9937	18.07	0.9888	13.69	0.9941

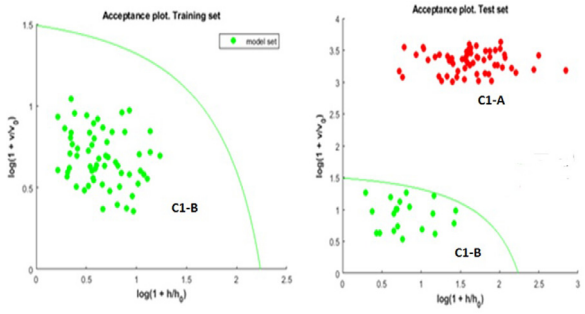

Fig. 7. DD-SIMCA to discriminate between C1 generics according to origin (o: C1-B generics; o: C1-A generics).

results (see Table 3). Satisfactory prediction results were obtained for all the spectral zones studied but the best result was obtained for zone 6004 – 5872 cm^{-1}. The calibration parameters RMSEC, RMSECV and RMSEP were comparable and the R^2 values were close to one, meaning that the PLS-R model is reliable and stable. The metronidazole content was determined with the HPLC reference method and compared to the result obtained with the PLS-R model based on the NIR measurement. The results are presented in Table 4 and Fig. 8 B for the 13 drug products studied: 9 batches of C1, 2 batches of C2, 1 batch of C3 and 1 batch of C4. Acceptable results were obtained for the PLS-R model. All the contents predicted by near infrared spectrophotometry are between 90 and 110% except for a single sample for which the predicted value is equal to 85.34%.

For 500 mg metronidazole tablets NIR spectra were measured through the blister and no sample preparation was required. However, the spectra obtained for the C1-A drug products (of Chinese origin) with 250 mg and 500 mg metronidazole were similar. The estimated ratio of API mass to tablet mass is 78.2% for the 250 mg samples (average mass, 319.2 mg) and 75.8% for the 500 mg samples (an average mass of 659.6 mg). The similarity of the NIR spectra obtained is due to the similarity of the ratio of API mass to tablet mass and the presence of the same excipients. The PLS-R model was constructed based on calibration samples with a content metronidazole expressed in 700 mg tablet weight (the average weight for 500 mg metronidazole tablets). The strategy of the preparation step for the 250 mg samples, consisting of standardization by weight, made it possible to use the PLS-R model and to quantify these samples in a satisfactory manner. The special feature of the present work is the fact that with metronidazole, a quantitative PLS-R calibration model could be developed on a powder and applied to tablets forms for content determination.

3.4. Validation of the PLS-R model

For the estimation of the performance criteria of the established model, external samples were used.

3.4.1. Specificity

We have already shown that the absorption band from 6062 to 5840 cm^{-1} is specific to metronidazole and could be used to identify the active ingredient metronidazole in the tablet formulations. The high ratio of metronidazole/tablet weight results in a spectral band due to metronidazole where almost no contribution of excipients is observed. The model is specific, detects and quantifies metronidazole exclusively.

3.4.2. Linearity

The method was found to be linear in the range of 100 mg - 600 mg metronidazole. The linear regression line (y = a + bx; y = dose predicted with PLS-R model and x = dose determined with HPLC reference method) had the following characteristics:

$$\text{Intercept } a = -18.1\left(IC_{99\%} = [-38.6; +2.45]\right), \text{slope } b = 1.03, \left(IC_{99\%} = [+0.97; +1.09]\right), r^2 = 0.994.$$

3.4.3. Precision

The repeatability of the model was evaluated by measuring 10 times the contents of the same sample of metronidazole tablet. The repeatability relative standard deviation is 2.2%.

Table 4
Relative bias range for NIRS metronidazole content prediction.

Metronidazole samples code	HPLC value (mg)	NIR Value (mg)	Relative bias (%)
E1	507.7	488.5	3.78
E2	478.2	433.6	9.33
E3	474.3	495.5	4.47
E4	484.1	468.1	3.31
E5	475.3	474.6	0.15
E6	477.9	487.2	1.95
E7	481.6	449.7	6.62
E8	501.6	458.9	8.51
E9	525.2	458.7	12.7
E10	515.7	492.2	4.56
E11	507.9	504.6	0.65
E12	251.0	252.7	0.68
E13	243.0	244.3	0.53

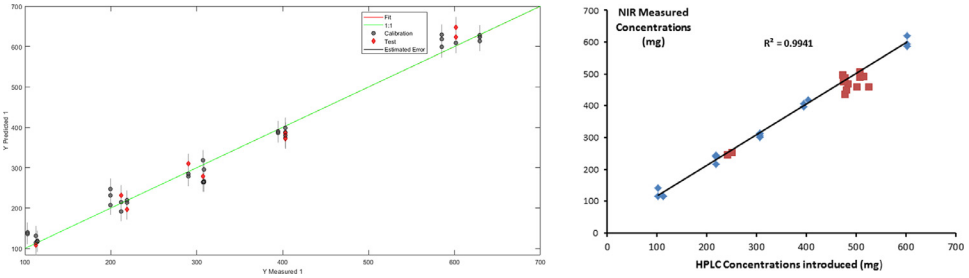

Fig. 8. (A): Established PLS-R calibration model, (A): Y Measured 1 = HPLC Measured Concentrations; Y Predicted 1 = NIR Concentrations predicted. Prediction of the calibration samples selected by the Duplex algorithm
(B): Quantification of the 13 metronidazole products studied (B).

3.4.4. Trueness

Trueness was assessed for the two content levels studied by calculating the relative bias between the models predicted content and the content established using the reference method. The relative bias was maximum = 13% for the 500 mg level and maximum 0.67% for the 250 mg level.

4. Conclusion

The classification performance of SIMCA models was evaluated for the authentication of the samples of four product classes C1, C2, C3 and C4. Initially 4 classes were defined in the SIMCA model, one for each of the product classes. The NIR spectra for samples of class C3 products were well separated from the other classes due to the presence of the excipient talc. The samples of C3 in the test set were all well classified by the SIMCA model. Nevertheless, the model performed less good for the classification of samples from class C4: 4 of the 5 test samples were classified as C1 samples. The PCA score plot also showed that C1 and C4 products had very similar NIR spectra. The SIMCA model in which C1 and C4 products were included in the same class, showed a classification rate for the test samples of 100%. The use of SIMCA on the NIR spectra for only C1 products (reduced data set) showed that the classification technique performed better for the classification of products with very similar spectra. The test set samples for products C1A (India) and C1B (China) were all well classified, the correct classification rate was 100%. The use of DD-SIMCA with one target class of C1B samples was also able to discriminate C1B samples from C1A. These models were able to identify the origin of the manufacturer with classification performance of 100% for the sensitivity and specificity.

The NIR fingerprints of the four product classes were very similar and mainly characterized by the API metronidazole. Only small spectral differences due to excipients were responsible for the discrimination of the products. Outlying products, which could be attributed to falsified or substandard products, were not observed. These results were confirmed by the HPLC reference method described in the USP monograph, “Metronidazol Tablets”, which identified metronidazole as the API and confirmed the dose: between 95 and 105 % of the label claim for all products studied.

The PLS-R model for the quantification of metronidazole was developed and applied to the 13 samples of products studied. The dose of metronidazole determined with the NIR spectral band specific for metronidazole using the PLS-R model gave acceptable results: relative bias between 0.15 and 12.7 % with respect to the dose determined by the HPLC method. This method demands no or almost no sample preparation, is very quickly and could be easily applied in the field by inspection authorities to detect substandard products. However, the application of the model on a larger number of samples and of more varied origin would make it possible to assess certain performance criteria such as the robustness of the method.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Abdoul Karim SAKIRA: Conceptualization, Formal analysis, Investigation, Writing - original draft. **Corenthin MEES:** Formal analysis, Investigation, Methodology, Software. **Kris De BRAEKELEER:** Formal analysis, Data curation, Methodology, Software, Validation, Visualization. **Cédric DELPORTE:** Resources, Project administration, Investigation. **Josias YAMEOGO:** Conceptualization. **Moussa YABRE:** Formal analysis. **Touridomon Issa SOME:** Conceptualization, Funding acquisition. **Pierre Van ANTWERPEN:** Resources, Funding acquisition. **Dominique MERTENS:** Project administration. **Jean Michel KAUFFMANN:** Conceptualization, Resources, Supervision, Project administration.

Acknowledgements

We sincerely thank the ARES-CCD (Académie de Recherche et d’Enseignement Supérieur de la Commission de Coopération au Développement), the financial support of this study, and the whole team

of the "Pharmacognosy, Bioanalysis and Drugs Discovery" unit (ULB, Brussel) for its collaboration and technical support.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.talo.2020.100027.

References

[1] H. Rebiere, P. Guinot, D. Chauvey, C. Brenier, Fighting falsified medicines: the analytical approach, J. Pharmaceut. Biomed. Anal. 142 (2017) 286–306, doi:10.1016/j.jpba.2017.05.010.

[2] K. Dégardin, Y. Roggo, P. Margot, Understanding and fighting the medicine counterfeit market, J. Pharmaceut. Biomed. Anal. 87 (2014) 167–175, doi:10.1016/j.jpba.2013.01.009.

[3] World Health Organization, Fact sheet No 275: Medicines: spurious/falsely labelled/ falsified/counterfeit (SFFC) medicines, 2013. Available from: http://www.who.int/mediacentre/factsheets/fs275/en/

[4] L. Höllein, E. Kaale, Y.H. Mwalwisi, M.H. Schulze, U. Holzgrabe, Routine quality control of medicines in developing countries: analytical challenges, regulatory infrastructures and the prevalence of counterfeit medicines in Tanzania, TrAC Trends Anal. Chem. 76 (2016) 60–70, doi:10.1016/j.trac.2015.11.009.

[5] B. Krakowska, D. Custers, E. Deconinck, M. Daszykowski, Chemometrics and the identification of counterfeit medicines—a review, J. Pharmaceut. Biomed. Anal. 127 (2016) 112–122, doi:10.1016/j.jpba.2016.04.016.

[6] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, N. Pou, Influence of the procedure used to prepare the calibration sample set on the performance of near infrared spectroscopy in quantitative pharmaceutical analyses, Analyst 126 (7) (2001) 1129–1134, doi:10.1039/B102090K.

[7] J. Luypaert, D.L. Massart, Y. Vander Heyden, Near-infrared spectroscopy applications in pharmaceutical analysis, Talanta 72 (3) (2007) 865–883, doi:10.1016/j.talanta.2006.12.023.

[8] M.H. Ferreira, J.W. Braga, M.M. Sena, Development and validation of a chemometric method for direct determination of hydrochlorothiazide in pharmaceutical samples by diffuse reflectance near infrared spectroscopy, Microchem. J. 109 (2013) 158–164, doi:10.1016/j.microc.2012.03.008.

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

[10] O.Y. Rodionova, A.L. Pomerantsev, NIR-based approach to counterfeit-drug detection, TrAC Trends Anal. Chem. 29 (8) (2010) 795–803, doi:10.1016/j.trac.2010.05.004.

[11] B.A. Olsen, M.W. Borer, F.M. Perry, R.A. Forbes, Screening for counterfeit drugs using near-infrared spectroscopy, Pharmaceut. Technol. 26 (6) (2002) 62–71.

[12] M.J. Vredenbregt, L. Blok-Tip, R. Hoogerbrugge, D.M. Barends, D. De Kaste, Screening suspected counterfeit Viagra® and imitations of Viagra® with near-infrared spectroscopy, J. Pharmaceut. Biomed. Anal. 40 (4) (2006) 840–849, doi:10.1016/j.jpba.2005.07.048.

[13] I. Storme-Paris, H. Rebiere, M. Matoga, C. Civade, P.A. Bonnet, M.H. Tissier, P. Chaminade, Challenging near infrared spectroscopy discriminating ability for counterfeit pharmaceuticals detection, Anal. Chim. Acta 658 (2) (2010) 163–174, doi:10.1016/j.aca.2009.11.005.

[14] H. Chen, Z. Lin, C. Tan, Non destructive discrimination of pharmaceutical preparations using near-infrared spectroscopy and partial least-squares discriminant analysis, Anal. Lett. 51 (4) (2018) 564–574 http://dx.doi.org/, doi:10.1080/00032719.2017.1339070.

[16] L.B. Schlegel, M. Schubert-Zsilavecz, M. Abdel-Tawab, Quantification of active ingredients in semi-solid pharmaceutical formulations by near infrared spectroscopy, J. Pharmaceut. Biomed. Anal. 142 (2017) 178–189 http://dx.doi.org/doi:, doi:10.1016/j.jpba.2017.04.048.

[15] L. Zhao, Y. Dou, H. Mi, M. Ren, Y. Ren, Non-destructive determination of metronidazole powder by using artificial neural networks on short-wavelength NIR spectroscopy, Spectrochimica Acta Part A: Mol. Biomol. Spectrosc. 66 (4-5) (2007) 1327–1332, doi:10.1016/j.saa.2006.06.022.

[17] M. Jamrógiewicz, Application of the near-infrared spectroscopy in the pharmaceutical technology, J. Pharmaceut. Biomed. Anal. 66 (2012) 1–10, doi:10.1016/j.jpba.2012.03.009.

[18] L. Alvarenga, D. Ferreira, D. Altekruse, J.C. Menezes, D. Lochmann, Tablet identification using near-infrared spectroscopy (NIRS) for pharmaceutical quality control, J. Pharmaceut. Biomed. Anal. 48 (1) (2008) 62–69, doi:10.1016/j.jpba.2008.05.007.

[19] F. Been, Y. Roggo, K. Degardin, P. Esseiva, P. Margot, Profiling of counterfeit medicines by vibrational spectroscopy, Forensic Sci. Int. 211 (1-3) (2011) 83–100, doi:10.1016/j.forsciint.2011.04.023.

[20] J.A.F. Pierna, O. Abbas, P. Dardenne, V Baeten, Discrimination of Corsican honey by FT-Raman spectroscopy and chemometrics, Base. [En ligne], numéro 1 15 (2011) (2011) 75–84 https://popups.uliege.be:443/1780-4507/index.php?id=6895.

[21] A. Guillemain, K. Dégardin, Y. Roggo, Performance of NIR handheld spectrometers for the detection of counterfeit tablets, Talanta 165 (2017) 632–640.

[22] C. Mees, F. Souard, C. Delporte, E. Deconinck, P. Stoffelen, C. Stévigny, JM Kauffmann, K. De Braekeleer, Identification of coffee leaves using FT-NIR spectroscopy and SIMCA, Talanta 177 (2018) 4–11.

[23] P.A. Gorry, General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method, Anal. Chem. 62 (6) (1990) 570–573, doi:10.1021/ac00205a007.

[24] E. Saccenti, J. Camacho, On the use of the observation-wise k-fold operation in PCA cross-validation, J. Chemomet. 29 (8) (2015) 467–478, doi:10.1002/cem.2726.

[25] D.L. Massart, B.G. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, C.K Mann, Handbook of chemometrics and qualimetrics: Part A, Appl. Spectrosc. 52 (1998) 302A.

[26] O.Y. Rodionova, L.P. Houmøller, A.L. Pomerantsev, P. Geladi, J. Burger, V.L. Dorofeyev, A.P. Arzamastsev, NIR spectrometry for counterfeit drug detection: a feasibility study, Anal. Chim. Acta 549 (1-2) (2005) 151–158, doi:10.1016/j.aca.2005.06.018.

[27] R.D. Snee, Validation of regression models: methods and examples, Technometrics 19 (4) (1977) 415–428, doi:10.1080/00401706.1977.10489581.

[28] Svante Wold, Michael Sjostrom, SIMCA: A method for analyzing chemical data in terms of similarity and analogy, in: B.R. Kowalski (Ed.), Chemometrics Theory and Application, American Chemical Society Symposium Series 52, Washington, D.C., American Chemical Society, 1977, pp. 243–282, doi:10.1021/bk-1977-0052.ch012.

[29] O.Y. Rodionova, A.V. Titova, A.L. Pomerantsev, Discriminant analysis is an inappropriate method of authentication, TrAC Trends Anal. Chem. 78 (2016) 17–22, doi:10.1016/j.trac.2016.01.010.

[30] De Maesschalck, R., Candolfi, A., Massart, D. L., & Heuerding, S. (1999). Decision criteria for soft independent modelling of class analogy applied to near infrared data. Chemometrics and Intelligent Laboratory Systems, 47(1), 65-77. 10.1016/S0169-7439(98)00159-2

[31] N. Kumar, A. Bansal, G.S. Sarma, R.K. Rawal, Chemometrics tools used in analytical chemistry: an overview, Talanta 123 (2014) 186–199, doi:10.1016/j.talanta.2014.02.003.

[32] I.C. Yang, C.Y. Tsai, K.W. Hsieh, C.W. Yang, F. Ouyang, Y.M. Lo, S. Chen, Integration of SIMCA and near-infrared spectroscopy for rapid and precise identification of herbal medicines, J. Food Drug Anal. 21 (3) (2013) 268–278, doi:10.1016/j.jfda.2013.07.008.

[33] O.Y. Rodionova, A.L. Pomerantsev, Application of the curve resolution method to the preprocessing spectral data in two-layer systems, J. Anal. Chem. 71 (1) (2016) 56–61, doi:10.1134/S1061934816010111.

[34] O.Y. Rodionova, A.V. Titova, N.A. Demkin, K.S. Balyklova, A.L. Pomerantsev, Qualitative and quantitative analysis of counterfeit fluconazole capsules: a non-invasive approach using NIR spectroscopy and chemometrics, Talanta 195 (2019) 662–667, doi:10.1016/j.talanta.2018.11.088.

[35] Rodionova, O. Y., Balyklova, K. S., Titova, A. V., & Pomerantsev, A. L. (2018). Application of NIR spectroscopy and chemometrics for revealing of the 'high quality fakes' among the medicines. Forensic Chem., 8, 82-89. 10.1016/j.forc.2018.02.004

[36] Y.V. Zontov, O.Y. Rodionova, A.L. Pomerantsev, S.V. Kucheryavskiy, DD-SIMCA–a MATLAB GUI tool for data driven SIMCA approach, Chemometr. Intell. Lab. Syst. 167 (2017) 23–28, doi:10.1016/j.chemolab.2017.05.010.

[37] J.U. Porep, D.R. Kammerer, R. Carle, On-line application of near infrared (NIR) spectroscopy in food production, Trends Food Sci. Technol. 46 (2) (2015) 211–230, doi:10.1016/j.tifs.2015.10.002.

[38] C. De Bleye, P.F. Chavez, J. Mantanus, R. Marini, P. Hubert, E. Rozet, E. Ziemons, Critical review of near-infrared spectroscopic methods validations in pharmaceutical applications, J. Pharmaceut. Biomed. Anal. 69 (2012) 125–132, doi:10.1016/j.jpba.2012.02.003.

[39] L.L. Wang, Y.W. Lin, X.F. Wang, N. Xiao, Y.D. Xu, H.D. Li, Q.S. Xu, A selective review and comparison for interval variable selection in spectroscopic modeling, Chemometr. Intell. Lab. Syst. 172 (2018) 229–240, doi:10.1016/j.chemolab.2017.11.008.

[40] C.A. Nunes, V.O. Alvarenga, A. de Souza Sant'Ana, J.S. Santos, D. Granato, The use of statistical software in food science and technology: Advantages, limitations and misuses, Food Res. Int. 75 (2015) 270–280, doi:10.1016/j.foodres.2015.06.011.