Deborah Custers, Patricia Courselle, Sandra Apers and Eric Deconinck*

Chemometrical analysis of fingerprints for the detection of counterfeit and falsified medicines

DOI 10.1515/revac-2016-0013
Received March 31, 2016; accepted July 1, 2016; previously published online September 22, 2016

Abstract: Counterfeit medicines pose a major threat to public health worldwide. These pharmaceuticals are mostly manufactured without respecting Good Manufacturing Practices. Moreover, they are not subjected to any form of quality control, and therefore their safety, efficacy, and quality cannot be guaranteed. Extensive research on counterfeit medicines has already been performed and published in literature. This review aims at providing an updated overview of the use of fingerprints and subsequent multivariate (chemometrical) data analysis in the field of counterfeit medicine detection. Fingerprinting could be a useful tool in the analysis of counterfeit medicines because it generates a holistic view of a sample, rather than focusing on specific and predefined characteristics, such as identification and quantification of present active pharmaceutical ingredients. This review first provides an introduction into the counterfeiting problem. Next, the concept of fingerprinting and the basic principles of chemometrics are explained, followed by a description of the successful application of fingerprints in the field of Pharmacognosy. The last part of this review provides an overview describing the use of fingerprints in counterfeit medicine research.

Keywords: chemometrics; counterfeit medicines; fingerprints.

***Corresponding author: Eric Deconinck,** Section Medicinal Products, Division of Food, Medicines and Consumer Safety, Scientific Institute of Public Health (WIV-ISP), J. Wytsmanstraat 14, Brussels B-1050, Belgium, e-mail: Eric.Deconinck@wiv-isp.be
Deborah Custers: Section Medicinal Products, Division of Food, Medicines and Consumer Safety, Scientific Institute of Public Health (WIV-ISP), J. Wytsmanstraat 14, Brussels B-1050, Belgium; and Research Group NatuRA (Natural Products and Food – Research and Analysis), Department of Pharmaceutical Sciences, University of Antwerp, Universiteitsplein 1, Wilrijk B-2610, Belgium
Patricia Courselle: Section Medicinal Products, Division of Food, Medicines and Consumer Safety, Scientific Institute of Public Health (WIV-ISP), J. Wytsmanstraat 14, Brussels B-1050, Belgium
Sandra Apers: Research Group NatuRA (Natural Products and Food – Research and Analysis), Department of Pharmaceutical Sciences, University of Antwerp, Universiteitsplein 1, Wilrijk B-2610, Belgium

Introduction

Counterfeit medicines pose a major threat to public health worldwide (World Health Organization [WHO] 2005, International Medical Products Anti-Counterfeiting Taskforce [IMPACT] 2008). Not only developing countries are subjected to the distribution of counterfeit medicines, but industrialized countries such as European countries, the United States, and Japan are exposed to pharmaceutical forgery as well (European Alliance for Access to Safe Medicines [EAASM] 2008). These counterfeited medicines are mostly manufactured by uncontrolled or street laboratories without respecting Good Manufacturing Practices (Sacre et al. 2011a). They are not subjected to any form of quality control (Jackson et al. 2012, Deconinck et al. 2014a), and therefore their safety, efficacy, and quality cannot be guaranteed (EAASM 2008, Höllein et al. 2016). The consequences of the use of counterfeit medicines may vary from therapeutic failure to the occurrence of serious adverse events and even death (Gautam et al. 2009).

The WHO defines a counterfeit medicine as "one which is deliberately and fraudulently mislabeled with respect to identity and/or source. Counterfeiting can apply to both branded and generic products, and counterfeit products may include products with the correct ingredients or with the wrong ingredients, without active ingredients, with insufficient active ingredients, or with fake packaging" (WHO 2005). Counterfeit products copy the brand name, and their appearance also resembles that of a genuine product (Vredenbregt et al. 2006, Deconinck et al. 2013c). Furthermore, contents of counterfeit medicines are highly unreliable because their source is unknown; they can range from inactive and useless formulations to harmful and toxic products (IMPACT 2008). Health risks might be due to the presence of incorrect active pharmaceutical ingredients (APIs), the absence of API, an incorrect dosage, the presence of high concentrations of potential toxic secondary components, and fake packaging or documentation (EAASM 2008). However, the definition by the WHO does not apply to the majority of illegal drugs encountered in industrialized countries because most of them do not copy the brand name and packaging of genuine medicines (Sacre et al. 2010). Moreover, despite

the reached consensus on this definition, several countries apply different definitions in their respective laws, thereby impeding both international cooperation and implementation of procedures to tackle the trade in counterfeit medicines (WHO 2005). Therefore, the WHO proposed some modifications, resulting in the use of the term "substandard/spurious/falsely labeled/falsified/counterfeit" medicines (WHO 2016b). Spurious, falsely labeled, falsified, and counterfeit medicines are recognized to be substandard, but not all substandard medicines are necessarily spurious, falsely labeled, falsified, or counterfeit. Substandard medicines (or out-of-specification products) constitute a particular group of illegal medicines. They are defined by the WHO as "a genuine medicine produced by manufacturers authorized by the national medical regulatory authority which does not meet the quality specifications set for them by national standards". This means that these kinds of medicines are produced in a legal way by the marketing authorization holder, but they should be destroyed because of failure to meet the set quality requirements. However, these batches sometimes do become available on the market because of unscrupulous people getting hold of these rejected batches (Wertheimer and Norris 2009, Deconinck et al. 2013c, WHO 2016a). Another variant of substandard medicines is drug diversion, i.e. expired genuine medicines that are repacked and sold (Wertheimer and Norris 2009, Degardin et al. 2014, Höllein et al. 2016).

Besides counterfeit and substandard medicines, other groups of frequently encountered illegal pharmaceuticals are imitation products and adulterations. Imitations do not copy the brand name or the packaging, but they do claim the presence of a certain API. Most of them are manufactured in Asian countries because these countries do not recognize the European and American patent laws. As a consequence, these products are produced legally in these countries but imported illegally in Europe and the United States (Vredenbregt et al. 2006, Sacre et al. 2010, Deconinck et al. 2013c). Adulterations are herbal products or dietary supplements that contain undeclared synthetic APIs; these APIs are added to increase the efficacy of these products (Rebiere et al. 2012, Justa Neves and Caldas 2015). These adulterated products represent a major hazard to public health because consumers are not aware of the presence of synthetic APIs because of fraudulent labeling with regard to the ingredients. This could lead to unexpected side effects or interactions with other medication a patient could be taking (Deconinck et al. 2013c).

As already mentioned, medicine counterfeiting is a global threat to public health (Degardin et al. 2014). Unfortunately, assessing the true extent of pharmaceutical counterfeiting is particularly difficult because of its illicit and clandestine nature (EAASM 2008, Höllein et al. 2016). Moreover, the size of the problem differs from region to region. The WHO estimates that approximately 1% of the total medicine market of industrialized countries is covered by counterfeit medicines. The extension of the Internet certainly contributes to the increasing threat posed by these pharmaceuticals in these developed areas. Research has shown that approximately 50% of all medicines, purchased online from websites that cover up their true identity, are forged (IMPACT 2008, Deconinck et al. 2012c, Delepierre et al. 2012). Online buying of medicines is sometimes preferred because it is quicker, purchases can be made anonymously, and prices are often thought to be lower compared with official pharmacies (Degardin et al. 2014). In countries of the former Soviet Union, approximately 20% of the medicine market would be covered by forged pharmaceuticals. On average, 30% of the medicines sold in African countries and parts of Asia and Latin America is suspected to be counterfeit, with percentages differing from 15% to 60% depending on the region (IMPACT 2008, Delepierre et al. 2012). Moreover, the types of medicines that are most commonly counterfeited also differ from region to region (IMPACT 2008). Counterfeiters mainly target expensive and high-consumption medicines (Deisingh 2005). In industrialized countries, the primary targets for counterfeiting are commonly called "life style drugs" and comprise phosphodiesterase type 5 inhibitors for the treatment of erectile dysfunction, slimming products, anabolic hormones, products for the treatment of hair loss, and even narcotics (Deconinck et al. 2013a,c, Degardin et al. 2014). Furthermore, counterfeiters are also targeting other types of medicines such as expensive cancer treatments and antiviral pharmaceuticals (Degardin et al. 2014). In developing countries, life-saving medicines, such as antibiotics and medicines for the treatment of tuberculosis, malaria, and HIV/aids are mostly counterfeited (Deconinck et al. 2013c, Degardin et al. 2014, Höllein et al. 2016). This represents a major threat to public health; treating diseases associated with high untreated mortality rates, e.g. malaria, tuberculosis, aids, and meningitis, with counterfeit or substandard medicines increases morbidity and mortality substantially. Moreover, the use of substandard/counterfeit antibiotics (which often contain subtherapeutic dosages of API or which do not release the API correctly) increases the risk of developing microbial resistance, which could even undermine the efficacy of genuine antibiotics (Caudron et al. 2008, Gautam et al. 2009, Wertheimer and Norris 2009, Deconinck et al. 2013c, 2014a, Hajjou et al. 2015). In general, the increasing prevalence of counterfeit

medicines jeopardizes patients' safety, and patients might start to severely distrust health care systems, health care professionals, the pharmaceutical industry, and drug regulatory authorities (Deconinck et al. 2013c, Hajjou et al. 2015).

The global character of this particular problem suggests that the production of counterfeit medicines has become a structured criminal industry consisting of manufacturers, wholesalers, distributors, and local sellers (Degardin et al. 2014). This globally organized industry is the result of several factors that contribute to the problem of medicine counterfeiting. Since this problem is affecting countries worldwide, a global harmonized legal framework is necessary which provides a universally accepted definition of counterfeit medicines and pharmaceutical crime. Unfortunately, as already stated previously, most member states of the WHO have their own understanding of a counterfeit medicine, making international cooperation very difficult (WHO 2005). A second contributing factor, which mostly affects developing countries, is the lack of effective regulatory systems and market control. Developing countries lack the means, e.g. restricted laboratory capacities, weak analytical infrastructure, insufficiently trained personnel, etc. to assess the quality, efficacy, and safety of medicines, thereby paving the way for counterfeit medicines (Wertheimer and Norris 2009, Deconinck et al. 2013c, Höllein et al. 2016). In addition, corruption is abundant in these areas, which facilitates the spread of fake pharmaceuticals even more (Wertheimer and Norris 2009, Degardin et al. 2014). Developing countries are also confronted with a demand for medicines which exceeds the supply. In these circumstances, unscrupulous people try to make a profit out of these situations by distributing counterfeit medicines (EAASM 2008, Wertheimer and Norris 2009). In addition, genuine medicines are expensive and often unaffordable for the population, which forces patients to seek cheaper alternatives on the street (EAASM 2008, Wertheimer and Norris 2009, Degardin et al. 2014). A very important factor that encourages trade in counterfeit medicines are the huge profits that are made; counterfeit medicines are sold at relatively high prices, while production costs are kept to a minimum. Counterfeiters are also less likely to be caught, unlike drug traffickers, because of weak legislation and enforcement. Moreover, penalties for pharmaceutical counterfeiting are less severe, making the counterfeiting of medicines financially rewarding and largely risk free (Delepierre et al. 2012, Deconinck et al. 2013c, Degardin et al. 2014). The WHO has estimated in 2010 that approximately 75 billion USD was spent worldwide on counterfeit medicines (WHO 2010); a survey conducted by Pfizer in the same year estimated that European citizens purchased medicines from illegal sources for a total amount of 10.5 billion EUR (World Courier 2014).

Nevertheless, despite all challenges that have to be faced to tackle the forgery of medicines, several international initiatives were launched. Fourteen major companies founded the Pharmaceutical Security Institute in 2002, which aims to collect data to identify the true extent of pharmaceutical counterfeiting and to provide assistance in the coordination of international investigations (Pharmaceutical Security Institute 2015). The WHO also launched an initiative, i.e. the IMPACT. Its task is to build coordinated networks between countries with the aim of putting an end to the production, trading, and selling of counterfeit medicines worldwide (WHO 2006). The EAASM, founded in 2007, is a pan-European initiative committed to the patient's safety by ensuring access to safe and legitimate medicines and promoting the elimination of counterfeit and substandard medicines (EAASM 2015). In 2008, Interpol coordinated the first Pangea operation, which is an annual international initiative intending to cease the online sale of counterfeit pharmaceuticals and to raise awareness of the dangers of buying medicines online (Interpol 2015). The European Parliament and the Council of Europe published the Falsified Medicines Directive (Directive 2011/62/EU) on July 1, 2011. This directive aims at improving the protection of public health using harmonized measures to ensure the safety of medicines and the strict control of medicine trade. These newly defined measures include obligatory safety features on the packaging of medicines, a common EU logo to identify legal online pharmacies, more stringent inspections of API producers, and rigorous requirements of record keeping by wholesale distributors (Directive 2011/62/EU of the European Parliament and of the Council 2011). On October 28, 2011, Europe casted its Medicrime convention. This convention is unique because it is the first binding international agreement on the criminalization of production and distribution of counterfeit medicines and similar crimes that threat public health. It has three major aims: (1) to impose penalties for offences described in the convention, (2) to protect victims, and (3) to promote national and international cooperation. This convention is not limited to European countries only but can be signed and ratified by any state worldwide. Up till now, the convention has been signed by 26 countries; in six of these countries, the convention already entered into force (Council of Europe 2014). The European Directorate for Quality of Medicines launched the "Track and Trace" project, which aims to implement a system of drug traceability. This project tries to involve all manufacturers of

raw materials active in one or more of the member states (European Directorate for the Quality of Medicines and HealthCare [EDQM] 2011).

Extensive research on counterfeit medicines has already been performed, resulting in the publication of several reviews. Some of these reviews focus on a specific technique such as nuclear magnetic resonance (NMR) (Holzgrabe and Malet-Martino 2011) or liquid chromatography (LC) (Deconinck et al. 2013c), others on specific groups of counterfeited medicines (Venhuis and de Kaste 2012), or they provide a more general overview of the counterfeiting problem along with the different analytical techniques that could be applied (Deisingh 2005, Martino et al. 2010, Höllein et al. 2016). This review aims at providing an updated overview of the use of fingerprints and subsequent data analysis by means of chemometrics in the field of counterfeit medicine research. Fingerprints, and chemometric analysis, could provide a useful tool in the analysis of counterfeit medicines because it generates a holistic view of a sample rather than focusing on specific and predefined characteristics, such as identification and quantification of present APIs. In addition to the present APIs, the fingerprint approach takes secondary substances, such as excipients and impurities, into consideration as well, which are considered to be equally important compared with the APIs. Consequently, the fingerprinting method can generate additional information that could aid competent regulatory authorities in their decision making and in protecting public health.

Fingerprints and their analysis using chemometrics

Fingerprints

In recent years, samples are increasingly being characterized by their chemical fingerprints. This approach is commonly referred to as non-targeted analysis (Daszykowski and Walczak 2011). A fingerprint can be defined as a characteristic profile that visualizes the chemical composition of the considered sample. Its aim is to construct a specific pattern of recognition, which will be entirely evaluated during data analysis (Alaerts et al. 2010a, Deconinck et al. 2013b,c, Goodarzi et al. 2013, Donno et al. 2015). Indeed, a non-targeted strategy assumes that unique information about the chemical composition of samples is captured and subsequently used for comparative analysis. The non-targeted analysis of samples benefits from several advantages, as opposed to targeted analysis. It generates a comprehensive view of the composition of a sample in contrast to targeted analysis, which describes a sample by a limited number of carefully selected and quantified compounds. The latter approach therefore assumes preliminary knowledge about the samples being studied, which is in reality often not the case when analyzing complex samples. Furthermore, it also requires the use of chemical standards. When using the non-targeted strategy, no preliminary knowledge about the samples is required, and depending on the aim of the data analysis, no chemical standards are needed (Daszykowski and Walczak 2011).

Fingerprints can be acquired by spectroscopic, chromatographic, or electrophoretic techniques (Alaerts et al. 2010a, Deconinck et al. 2013b,c). Chromatographic fingerprints are the most informative type of fingerprints. Since information concerning the composition of a sample is spread over time, information on present individual compounds can be extracted from the fingerprint. However, mobile phase changes, aging of analytical columns, and instrumental instability complicate the chemometrical analysis of the acquired fingerprints (Alaerts et al. 2010a, Deconinck et al. 2013b, Yang et al. 2013). Spectroscopic fingerprints are widely used for the identification of bulk materials (Deconinck et al. 2013c). Both the European and the United States Pharmacopoeia make use of infrared (IR) spectra to compare fingerprint regions of the spectra obtained from samples with reference spectra (EDQM 2015, United States Pharmacopoeial Convention 2015). These kinds of fingerprints are influenced by all compounds present in the samples because samples are analyzed in a whole without the spread of information over time (Deconinck et al. 2013c). Depending on the aim of the study, this can either be an advantage or a disadvantage.

No matter which analytical technique is used to acquire fingerprints, vast amounts of data are generated from which it is difficult to extract meaningful information. However, chemometrics or multivariate data handling provides the necessary tools to manage and analyze the data (Deconinck et al. 2013c).

Chemometrics

Chemometrics can be used for exploratory, classification (i.e. pattern recognition), and multivariate regression purposes (Alaerts et al. 2010a, Tistaert et al. 2011b). Chemometric analysis of fingerprints usually consists of three important parts: (1) data pretreatment, (2) unsupervised analysis, and (3) supervised analysis (Alaerts et al. 2010a, Tistaert et al. 2011b). To enable multivariate analysis of the

acquired fingerprints, all fingerprints are represented in a $m \times n$ data matrix $\boldsymbol{X}$ (Figure 1). Each fingerprint constitutes a single row in data matrix $\boldsymbol{X}$; consequently, m equals the number of acquired fingerprints. The number of columns is designated by n. Each column contains the signal intensity for a given measuring point (most often called variable), e.g. when acquiring chromatographic fingerprints, the considered variables are the time points, and each column represents the registered detector signal for one particular time point. For spectroscopic fingerprints, the variables are the included wavelengths, and each column represents the detector signal at that particular wavelength (Alaerts et al. 2010a, Tistaert et al. 2011b, Goodarzi et al. 2013).

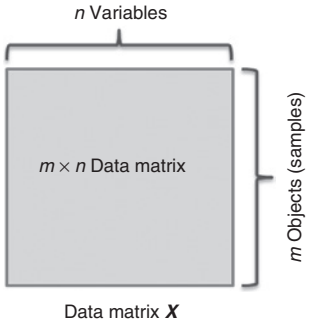

Figure 1: Schematic representation of how fingerprints are organized in a data matrix suitable for multivariate data analysis.

Data preprocessing

Prior to unsupervised and supervised data analysis, an appropriate pretreatment of the data needs to be performed to remove experimental variation and to improve the quality of the data set (Tistaert et al. 2011b). When recording chromatographic fingerprints, shifts in retention time can occur because of column aging, mobile phase changes, and instability of the chromatographic system. As a result, the recorded fingerprints often need to be aligned. Several techniques are developed for this purpose, such as correlation optimized warping (COW) (Vest Nielsen et al. 1998, Pravdova et al. 2002, Tomasi et al. 2004, van Nederkassel et al. 2006a), parametric time warping (Eilers 2004, van Nederkassel et al. 2006a), dynamic time warping (Pravdova et al. 2002, Tomasi et al. 2004), semi-parametric time warping (van Nederkassel et al. 2006a,b), target peak alignment (van Nederkassel et al. 2006b), and fuzzy warping (Walczak and Wu 2005). Among these alignment techniques, COW is most often used and proved to be most effective (Daszykowski et al. 2010, Tistaert et al. 2011b). However, it should also be noted that COW is a time-consuming technique and lacks ease of use since two input parameters need to be optimized by a trial-and-error approach (Tistaert et al. 2011b).

In addition to peak alignment, other data pretreatment techniques might be required, such as normalization, column centering, and paretoscaling. These latter preprocessing methods aim to minimize the unwanted variability and to maximize the observed information, i.e. the differences between the samples (Alaerts et al. 2010a, Tistaert et al. 2011b). Normalization is a preprocessing technique that scales each column to a constant total by dividing each particular column by a value that represents the general intensity of the column. When applying column centering, the column mean is subtracted from each value in a particular column. This basic preprocessing method is an essential part of several multivariate analysis techniques such as principal component analysis (PCA). Paretoscaling constitutes a subtraction of the column mean from each variable followed by a division by the square root of the column standard deviation (Alaerts et al. 2010a). Data preprocessing needs to be performed with the necessary caution because it will largely influence the results that will be acquired during the following steps of the chemometric data analysis (Tistaert et al. 2011b).

Unsupervised data analysis

The second step in the chemometric data analysis consists of an exploratory analysis, using unsupervised techniques. Unsupervised methods only use the information present in data matrix $\boldsymbol{X}$ or, put differently, they make use of the acquired fingerprints as their sole source of information (Massart et al. 1997, Alaerts et al. 2010a, Tistaert et al. 2011b, Goodarzi et al. 2013). This concept

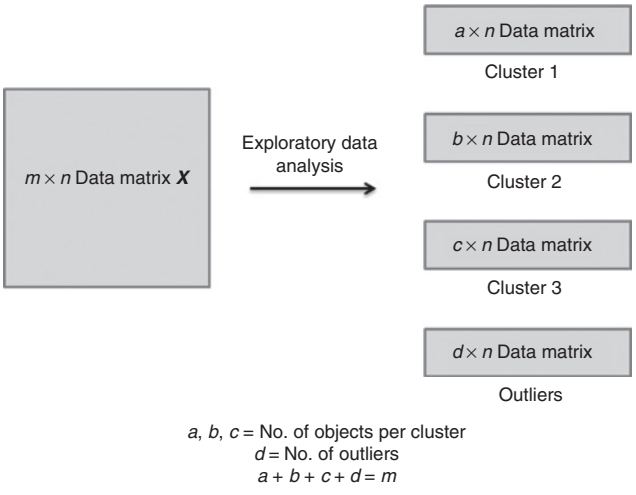

Figure 2: Visualization of the exploratory data analysis principle.

is visualized in Figure 2. Exploratory data analysis aims at revealing differences among the investigated samples and to identify the signals that contribute most to the observed differences (Daszykowski and Walczak 2011). Such an approach could be used to explore whether an underlying clustering of samples or outliers is present (Alaerts et al. 2010a, Tistaert et al. 2011b, Goodarzi et al. 2013). Commonly applied techniques are PCA, projection pursuit (PP) (which are both projection techniques), and cluster analysis (which is a clustering technique). Clustering and projection techniques have both proven their usefulness for the exploratory analysis of fingerprints.

Projection methods

Projection techniques have the valuable feature of visualizing high-dimensional data by projecting them into a low-dimensional space using only a few latent variables. Thereby, they help to summarize the data structure (Daszykowski and Walczak 2011).

Principal component analysis

PCA is a variable reduction technique that allows to visualize the information contained in the fingerprints (i.e. $m \times n$ data matrix $\boldsymbol{X}$) by projecting the original high-dimensional variables into a low-dimensional space. This low-dimensional space is defined by new orthogonal latent variables, which are commonly referred to as principal components (PCs). These PCs are linear combinations of the original variables; they are defined in a way to describe the variance present in data matrix $\boldsymbol{X}$. The first constructed PC explains the highest variance in the data; the second one explains the highest residual variance remaining after the construction of the first PC. The same principle is repeated for the third PC around the plane, defined by the first and second PCs. PCA results in two matrices: a score matrix and a loading matrix. All m objects (i.e. samples) are projected on these newly defined PCs; these projections are called scores. They can be visualized using a score plot, thereby providing information about the (dis)similarities among the objects. The loadings represent a measure for the weight of each original variable, which indicates that variables with higher loadings contribute more to the construction of a given PC. A loading plot can have its usefulness in exploring which individual variables are responsible for the present inter-sample variances. This information can often be linked to certain signals or chemical components in the fingerprints (Massart et al. 1997, Daszykowski et al. 2003, Daszykowski 2007, Daszykowski and Walczak 2011, Goodarzi et al. 2013).

Projection pursuit

PP will project high-dimensional data into a low-dimensional space as well. This low-dimensional space is defined by a few latent variables, which are called PP features. The concept of PP is to find several directions in the data space that will lead to "interesting" low-dimensional projections revealing useful information about the data structure (e.g. to reveal the presence of sample grouping or outliers). This is achieved by maximizing a projection index (PI), which describes the inhomogeneity of the data. One of the most popular PIs is entropy. Other PIs used in literature are yenyukov's index and kurtosis. By testing different PIs, different aspects of the data structure can be explored (Daszykowski et al. 2003, 2005, Stanimirova et al. 2005, Daszykowski 2007). Since the definition of the latent variables differs for PCA and PP, PP could be complementary to PCA in revealing the data structure. Consequently, PP could visualize groups of samples that could not be observed by PCA and vice versa (Deconinck et al. 2012c).

Clustering techniques

Clustering techniques represent a second important tool in the exploratory analysis of data. The different clustering methods are generally divided in two subgroups: hierarchical and non-hierarchical clustering (Alaerts et al. 2010a, Tistaert et al. 2011b).

Non-hierarchical clustering

Non-hierarchical methods, also called partitioning methods, divide the objects (i.e. samples) into several mutually exclusive groups in a way that objects assigned to one group are more similar to one another than to objects assigned to other groups. This acquisition of mutually exclusive groups is the result of maximizing a certain similarity measure over the considered groups. The most popular similarity measure used is the variance criterion, which is formulated as follows (Drab and Daszykowski 2014):

$$E = \sum_{k=1}^{k} \sum_{i \in c_k} \| x_i - c_k \|^2$$

where x_i represents the ith object, c_k is the coordinate of the kth cluster center, and k is the number of cluster centers.

The partitioning of objects is obtained by defining several cluster centers k (e.g. randomly). For each object, the nearest cluster center is identified to which it is subsequently assigned. Next, for each cluster separately, the new center coordinates are estimated, i.e. the mean of the coordinates of the objects belonging to the considered cluster is calculated. Afterwards, the membership of all objects is verified,

thereby taking into account the actual cluster centers that were calculated in the previous step. If necessary, the objects have to be reassigned to another cluster. These latter two steps are repeated until all objects are finally assigned to the correct cluster (Drab and Daszykowski 2014).

Hierarchical clustering

Hierarchical methods are the most often applied clustering techniques (Alaerts et al. 2010a). Hierarchical clustering analysis (HCA) can best be described as a technique that identifies homogeneous subgroups in a way that objects belonging to the same group/cluster are more similar to one another than to objects from other clusters (Almeida et al. 2007). It attempts to reveal the data structure, which is displayed in the form of a tree (i.e. a so-called dendrogram). The construction of this dendrogram is an iterative and multi-step procedure using either an agglomerative or a divisive approach (Figure 3). The agglomerative strategy consists of a sequential merging of objects; the divisive method sequentially splits clusters until they all contain a single object. However, the former is most applied in literature. The agglomerative approach starts with each object in a separate cluster. During each consecutive step, the two most similar samples and/or clusters are combined to form larger clusters, thereby taking a certain similarity measure into account. This sequential merging continues until all objects are part of one large cluster. This entails that if m objects are present in the data set, then m–1 merging steps have to be performed. The resulting dendrogram shows the established hierarchy of sample similarities. A series of clustered samples is listed along the horizontal axis; the vertical axis provides information about the similarity among the considered samples. Samples that are grouped in the lowest branches of the dendrogram are the most similar (Vandeginste et al. 1998, Van Gyseghem et al. 2006, Almeida et al. 2007, Daszykowski and Walczak 2011, Tistaert et al. 2011b, Drab and Daszykowski 2014).

An agglomerative HCA is characterized by two parameters: (1) the similarity measure and (2) the linkage technique. To quantify the similarity, several distance measures can be selected, e.g. Euclidean distance, Mahalanobis distance, and Standardized Euclidean distance. The linkage techniques are used to decide which objects or clusters should be joined. The most popular linkage techniques are single linkage, complete linkage, average linkage, Ward's algorithm, and centroid linkage.

Single linkage

The single linkage is often called the nearest neighbour; two clusters are merged when the calculated distance between an object from one cluster and an object from the other cluster is the smallest one, or

$$\text{Single linkage: } \min_i \| a_i - b_i \|,$$

where a_i and b_i are all objects in clusters a and b.

Complete linkage

Complete linkage is often called the furthest neighbour. This linkage technique takes into consideration the largest

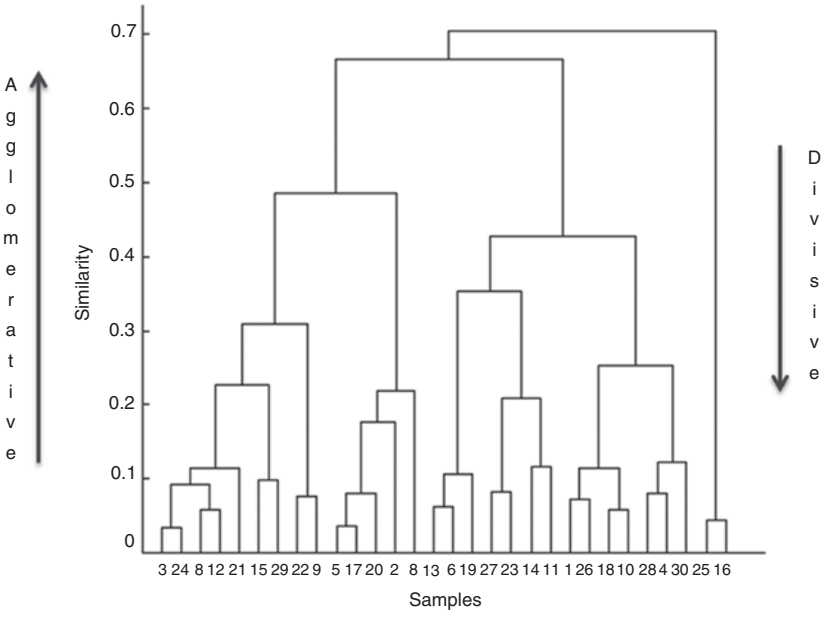

Figure 3: Representation of a dendrogram acquired by HCA.

distances calculated between objects from one cluster and objects from the other; the two clusters to be joined are characterized by the "smallest" largest distance, or

$$\text{Complete linkage: } \max_i \| a_i - b_i \|,$$

where a_i and b_i are all objects in clusters a and b.

Average linkage

A third often used linkage technique is the average linkage, which defines the distance between two clusters as an average of the single linkage and the complete linkage distances. The two clusters exhibiting the "smallest" average distance are subsequently merged:

$$\text{Average linkage: } \text{mean}_i \| a_i - b_i \|,$$

where a_i and b_i are all objects in clusters a and b.

Ward's algorithm

Ward's algorithm is based on the inner squared distance of clusters, meaning that two clusters are merged for which the lowest increase in total within-group error sums of squares is observed:

$$\text{Ward's algorithm: } \frac{\sqrt{2m_a m_b}}{\sqrt{m_a + m_b}} \| c_a - c_b \|,$$

where c_a and c_b are the centers of clusters a and b, and m_a and m_b are the number of objects in both clusters.

Centroid linkage

The fifth linkage technique, i.e. centroid linkage, is based on the distance between the cluster centers:

$$\text{Centroid linkage: } \| c_a - c_b \|,$$

where c_a and c_b are the centers of clusters a and b.

The outcome of an HCA can be very different depending on the data distribution and the selected similarity measure and linkage technique. For exploratory purposes, a data set can be studied by using different combinations of similarity measures and linkage techniques and by comparing the resulting classifications. It should, however, be kept in mind that the interpretation of the resulting dendrograms is rather intuitive (Vandeginste et al. 1998, Smolinski et al. 2002, Daszykowski and Walczak 2011, Goodarzi et al. 2013, Drab and Daszykowski 2014).

Supervised data analysis

Supervised techniques are featured by the use of additional information on the samples. Besides the information present in the measured fingerprints (i.e. data matrix $\boldsymbol{X}$), a response vector $\boldsymbol{y}$ is used containing the supplementary information (Massart et al. 1997, Alaerts et al. 2010a, Tistaert et al. 2011b). Supervised data analysis is used for multivariate regression and classification/discrimination purposes; the concept of the latter is shown in Figure 4. The main difference between both approaches is the definition of the $m \times 1$ response vector $\boldsymbol{y}$ (m equals the number of acquired fingerprints). When solving a classification problem, the response vector $\boldsymbol{y}$ is categorical, meaning that $\boldsymbol{y}$ represents the classes of the samples (e.g. genuine versus counterfeit or different origins of plants). In case of multivariate regression, the information contained in $\boldsymbol{y}$ is continuous, i.e. an activity of the samples expressed as a value and determined using a reference method (e.g. antioxidant capacity of samples). Supervised techniques make use of a training or calibration set, consisting of samples for which the classes or response values are known a priori, to construct the classification or calibration model (Alaerts et al. 2010a).

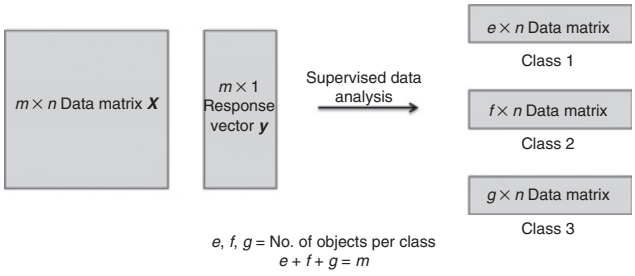

Figure 4: Visualization of the supervised data analysis principle for classification purposes.

Multivariate regression

Multivariate regression models are used to model a property of interest, e.g. antioxidant or cytotoxic activity of samples, in function of the measured fingerprints. After the selection of a calibration set, a regression model is constructed, which is subsequently used to predict the property of interest for new objects (Alaerts et al. 2010a, Tistaert et al. 2011b). Techniques applied for this purpose are, among others, multiple linear regression (MLR) and partial least squares (PLS).

MLR produces a linear model that describes a quantitative property (indicated by the response vector $\boldsymbol{y}$) by means of independent variables (data matrix $\boldsymbol{X}$). However, the application of MLR requires that the number of objects m (i.e. fingerprints) is larger than the number of variables n. This condition is often not met when analyzing fingerprints. Therefore, a variant of MLR was developed, i.e. stepwise MLR, which includes a stepwise selection procedure

to decide on which variables should be included in the regression model (Massart et al. 1997, Caetano et al. 2005).

PLS is a regression technique that models the response vector ***y*** by means of latent variables (so-called PLS factors), which are defined in a way to maximize the covariance between data matrix ***X*** and ***y***. On the basis of these latent variables, a vector of PLS regression coefficients can be computed relating the fingerprints of the objects to the dependent ***y*** values. This vector of regression coefficients can subsequently be used to predict ***y*** for unknown objects (Vandeginste et al. 1998, Chemometrics in Food Chemistry 2013). PLS was originally developed for regression purposes. However, depending on the formulation of response vector ***y***, it can also be used for discrimination purposes; this technique (explained more in detail further on) is then commonly referred to as partial least squares-discriminant analysis (PLS-DA). Several variants of this regression method have been developed and used in literature such as orthogonal projection to latent structures (O-PLS). O-PLS not only defines PLS factors representing the maximum covariance between ***X*** and ***y*** but also removes data from the ***X*** matrix which is not correlated to ***y*** (Vandeginste et al. 1998, Trygg and Wold 2002). Uninformative variable elimination-partial least squares (UVE-PLS) is a variant of the classic PLS algorithm as well. It aims to remove variables that are not more informative for modeling than noise, i.e. the so-called uninformative variables, in addition to the construction of PLS factors (Centner et al. 1996). Partial robust M-regression (PRM) is a robust variant of PLS, which diminishes possible negative influences of the objects on the regression coefficients (Serneels et al. 2005, Daszykowski et al. 2007).

Such techniques are often used for modeling in the field of Pharmacognosy, as shown in later sections. However, in the field of fingerprint analysis of counterfeit medicines, only classification and discrimination techniques are applied. Therefore, an in-depth description of regression techniques is beyond the scope of this review. Multivariate classification and discrimination techniques, on the other hand, will be described more in detail in the next sections.

Classification and discrimination

The aim of classification and discrimination (often referred to as pattern recognition) is to classify samples in classes, predefined in the response vector ***y***, based on the acquired fingerprints. On the basis of a training set with a priori known classification, a classification model is constructed which is subsequently validated by an independent test set. This test set consists of additional objects with known classification and is used to validate the predictive ability of the constructed model (Tistaert et al. 2011b). The main difference between classification and discrimination techniques lies in the means to classify samples. Classification methods result in "soft" classification rules allowing samples to be assigned to one class, to multiple classes, or to any class; discrimination techniques construct "hard" classification rules, indicating that a sample will always be allocated to only one class (Vandeginste et al. 1998, Chemometrics in Food Chemistry 2013).

Supervised data analysis in the field of counterfeit medicine fingerprinting mostly consists of the use of pattern recognition techniques; modeling techniques are applied to predict the nature of the analyzed samples based on their fingerprints (e.g. genuine versus counterfeit) (Deconinck et al. 2013c). Therefore, the most commonly applied techniques will be briefly described. More detailed information on the considered techniques can be found in the corresponding references.

Soft independent modeling of class analogy

Soft independent modeling of class analogy (SIMCA) is a classification technique that models each class of objects separately by defining the optimal number of PCs (which are derived from PCA) required to describe each class individually by usage of a cross-validation procedure. Next, classification rules are constructed by taking two critical values into account: (1) the Euclidean distance towards the SIMCA model and (2) the Mahalanobis distance calculated in the space of scores. These two critical values determine a restricted space around the included objects. The position of a new object is calculated using the scores and loadings of the created PCA models. If the object is situated within the restricted space, defined by the Euclidean and Mahalanobis distances, then the object is assigned to that particular class. SIMCA is a soft classification method, meaning that an object can be assigned to one or more existing classes or to any (Vandeginste et al. 1998, Practical Guide to Chemometrics 2006, Deconinck et al. 2012c, Chemometrics in Food Chemistry 2013).

Classification and regression tree analysis

The classification and regression tree (CART) analysis is a supervised non-parametric discrimination technique that can be used to solve both classification and regression problems, depending on the definition of the response vector ***y***. A classification tree is produced, which is used to solve classification problems, if ***y*** is categorical. When ***y*** is continuous, a regression tree is created, which can be

used to solve regression problems (see the Section Supervised data analysis) (Deconinck et al. 2005, 2006).

A CART analysis consists of three steps. The first step is the creation of the maximum tree, starting at the tree root containing all objects, using a binary split procedure. In each step of this splitting procedure, a mother group is split in two daughter groups, based on one variable and its split value. The most appropriate variable and split value, i.e. the variable and the split value resulting in the highest decrease in impurity between the mother and the daughter groups, are selected using an algorithm that considers all possible variables and split values. This impurity can be defined by different split criteria such as the Gini, Twoing, and Deviance indexes. The splitting procedure is repeated until the maximum tree is created, i.e. the tree in which each end node (leaf) contains one object or a predefined number of objects or homogeneous groups (Deconinck et al. 2005, 2006, 2012b).

This maximal tree shows overfitting because of overgrowing. Therefore, it is pruned in the second step by cutting terminal branches. As a result, several smaller and less complex trees are derived from the maximal tree (Deconinck et al. 2005, 2006, 2012b).

The final step consists of selecting the optimal tree, based on the evaluation of the predictive error by means of a cross-validation procedure (Deconinck et al. 2005, 2006, 2012b). CART has the advantage of a clear visualization and a straightforward interpretation of the acquired model because it results in a manifest set of decision rules (Ronowicz et al. 2013).

Two variants of the classic CART methodology were developed. One variant is actually an unsupervised method because no response vector $\boldsymbol{y}$ is included during the data analysis. This method is called auto-associative multivariate regression trees (AAMRT) and uses the explanatory variables as response variables as well. When the response vector $\boldsymbol{y}$ consists of more variables (i.e. more than one response variable per sample), the second CART variant can be used; this is an extended version of CART and is called multivariate regression trees. More detailed information on both variants is provided by Breiman et al. (1984), Questier et al. (2005), and Smyth et al. (2006).

k-Nearest neighbours

k-Nearest neighbours (k-NN), a discrimination technique, constructs a classification model by calculating the Euclidean distance between an unknown object and each of the objects of the training set. This signifies that if the training set includes m samples, then m distances are calculated. Subsequently, the k nearest objects to the unknown object are selected and a majority rule is applied; the unknown object is assigned to the class to which the majority of the k neighbouring objects of the training set belong. The number of nearest neighbours (k) to be included in the construction of a classification model has to be determined by optimization. Several k-NN models are built using different values for k. The best model is selected based on a cross-validation procedure (Vandeginste et al. 1998, Wehrens 2011, Chemometrics in Food Chemistry 2013).

Partial least squares-discriminant analysis

PLS-DA aims to differentiate between groups of objects of which the group membership is indicated by a response vector $\boldsymbol{y}$ containing categorical variables. A PLS-DA model is acquired by constructing so-called PLS factors, which are linear combinations of the original variables. These variables are constructed in a way that they represent maximum covariance between the original variables and the response variable $\boldsymbol{y}$. To obtain the best performing PLS-DA model, its complexity, i.e. the number of PLS factors, is optimized using a cross-validation procedure (Wold et al. 2001, Barker and Rayens 2003, Daszykowski et al. 2008, Chemometrics in Food Chemistry 2013, Deconinck et al. 2014b).

Other techniques

Besides the described classification and discrimination techniques, other supervised methods aiming at classifying samples are also occasionally encountered in literature. Linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) both define several discriminant functions to reduce the number of variables. These discriminant functions are combinations of the original variables; they are defined in a way to separate the classes of the objects as optimally as possible. In case of LDA, the discriminant functions are linear; when applying QDA, the functions are quadratic (Wu et al. 1996, Vandeginste et al. 1998). Canonical discriminant analysis (CDA) creates a new vector space that maximizes the distances between the present classes, thereby generating a better separation of the classes compared with the original data space (Bertrand et al. 1990, Petrakis et al. 2008).

The support vector machine (SVM) method is a machine learning technique that attempts to discriminate between classes of samples by constructing a separation hyperplane. The most optimal separation hyperplane is generated by maximizing the margin between the hyperplane and the respective sample classes. This method has the advantage of solving both linear and nonlinear discrimination problems depending on the used type of kernel function (Drucker et al. 1999, Belousov et al. 2002).

Usage of fingerprints in Pharmacognosy

The fingerprint strategy has already proven its usefulness in the field of Pharmacognosy for several purposes: (1) identification of plants, (2) classification of plants and differentiation of related species, (3) stability testing, (4) quality control, and (5) prediction of pharmacological activities or identification of potential active compounds (Deconinck et al. 2013b,c, Goodarzi et al. 2013). Pharmacognosy is referred to as the research area that investigates medicinal substances derived from plant, animal, and mineral materials in their natural, crude, or unprepared state and explores naturally occurring structure-activity relationships (Kinghorn 2001). The ability of the fingerprinting method to generate a comprehensive view of a sample's composition provides a useful means to analyze plants because their complex composition is not only unknown but also variable because of several factors such as cultivating and harvesting conditions and geographic area of cultivation (Goodarzi et al. 2013). Chromatographic fingerprinting has been accepted by the WHO as an identification and qualification technique for herbal products (Tistaert et al. 2011b, Goodarzi et al. 2013). Since a comprehensive review of the use of fingerprints in the field of Pharmacognosy is beyond the scope of this publication, only a limited number of exemplary studies demonstrating the usefulness of multivariate analysis of fingerprints in this particular research field are described. The use of the fingerprinting strategy for the quality control of plants has already been extensively reviewed (Liang et al. 2009, 2010, Alaerts et al. 2010a, Jiang et al. 2010, Song et al. 2013, Yang et al. 2013).

Alaerts et al. (2014) made use of high-performance liquid chromatography coupled to a photodiode array detector (HPLC-PDA) to acquire fingerprints that were subsequently used for the distinction, identification, and quality control of four different *Artemisia* species (*Artemisia vulgaris*, *Artemisia absinthium*, *Artemisia annua*, and *Artemisia capillaris*). These four species could be distinguished by PCA, and a reliable prediction model could be constructed when applying SIMCA. In addition, samples of different quality could be distinguished (Alaerts et al. 2014).

Besides identification, herbal fingerprinting has also shown its utility in the classification and differentiation of related species. PCA, PP, HCA, and similarity analysis by means of correlation coefficients were used to analyze fingerprints, acquired by HPLC-PDA, which resulted in a reliable discrimination between rhizoma *Chuanxiong* and rhizoma *Ligustici* (Alaerts et al. 2010b). Viaene et al. (2015) compared different discriminant and classification techniques to distinguish between two genera (i.e. *Mallotus* and *Phyllanthus*) on the one hand and six species (*Mallotus apelta*, *Mallotus paniculatus*, *Phyllanthus emblica*, *Phyllanthus reticulatus*, *Phyllanthus urinaria*, and *Phyllanthus amarus*) on the other hand, based on HPLC-UV fingerprints. For the aimed binary classification, LDA, QDA, and CART generated good results; for the six-class classification problem, LDA, QDA, and SIMCA resulted in a perfect discrimination (Viaene et al. 2015). The potential use of near-infrared (NIR) spectroscopic fingerprints was evaluated to determine the origin of *Cryptomeria japonica* varieties growing in southern Brazil. The use of NIR spectroscopy in combination with PCA proved sensitive enough to detect minor differences in the chemical composition of trees from the same species growing in the same location, but with different origins (Nisgoski et al. 2016). Zhou et al. (2015) used HPLC-PDA fingerprinting for the purpose of authenticity testing. Both raw and roasted seeds of *Descurainia sophia* were analyzed; PCA and PLS-DA proved to be suitable to distinguish between both types of seeds. Furthermore, fingerprints were measured from the seeds of *Plantago depressa* because these are often used as an adulterant for *D. sophia*; HCA proved to be capable of distinguishing seeds from both species (Zhou et al. 2015).

As already mentioned, the quality control of plants and herbal products is a very important issue, which is often addressed by means of fingerprinting. For instance, the quality of green tea samples was investigated based on a similarity analysis of chromatographic fingerprint profiles by Alaerts et al. (2012). Different correlation and distance measures were tested to evaluate their suitability. It was shown that similarity analysis based on correlation and congruence coefficients could serve as useful tools to identify outlying green tea samples, which possibly do not meet applicable quality standards (Alaerts et al. 2012). Li et al. (2013) demonstrated that NIR spectroscopic and chromatographic fingerprinting methods are effective and helpful for the quality control of *Lonicerae japonicae flos*. The quality control of *Ganoderma lucidum* was performed by means of HPLC-PDA fingerprinting in combination with PCA, HCA, SIMCA, and PLS-DA by Chen et al. (2008). Gas chromatographic (GC) fingerprints were explored for quality control purposes of *Scutellaria barbata* and its discrimination from two common adulterants, i.e. *Oldenlandia diffusa* and *Lobelia chinensis* (Pan et al. 2011). The combination of chromatographic fingerprinting and multivariate analysis by PCA, HCA, and MLR were shown to be useful for the quality control of *Salvia miltiorrhiza* (Zhang et al. 2015). A strategy for the quality control of *Pericarpium citri reticulatae* and *Pericarpium citri reticulatae viride*,

based on HPLC-PDA fingerprinting and chemometric analysis by PCA and PLS-DA, was proposed by Yi et al. (2007).

The use of herbal fingerprinting can also aim at predicting pharmacological activities or identifying potential active compounds. An example of the latter was provided by Masoum et al. (2015). Gas chromatography-mass spectrometry (GC-MS) fingerprints were analyzed by O-PLS, which led to the identification of compounds potentially responsible for the antioxidative activity of the essential oil of Thyme (Masoum et al. 2015). A vast amount of work has been performed on predicting the antioxidant capacity of green tea extracts based on chromatographic fingerprints. Several multivariate regression techniques have been used for this particular purpose, such as PLS and UVE-PLS (van Nederkassel et al. 2005), PRM (Daszykowski et al. 2007), O-PLS, and stepwise MLR (Dumarey et al. 2008, 2010). Another species that is extensively studied in literature within this context is *Mallotus*. O-PLS was performed on a set of chromatographic fingerprints acquired for 17 different *Mallotus* species, and examination of the regression coefficients of the acquired calibration model resulted in the detection of several peaks, which are potentially responsible for the antioxidant capacity (Tistaert et al. 2009). Four of these peaks of interest could subsequently be identified using HPLC-MS and turned out to be quercitrin, kaempferol-3-0-L-rhamnosyl, and mallonanoside A and B (Tistaert et al. 2012b). Tistaert et al. (2011a) used two dissimilar chromatographic systems to develop fingerprints of a set of *Mallotus* samples. To indicate the potentially antioxidant peaks, a multivariate calibration model was constructed by applying O-PLS. It was shown that the dissimilar chromatographic systems provided complementary information. Several compounds coeluting on one system were separated on the dissimilar system, and the corresponding calibration models revealed additional information on the contribution of these compounds to the antioxidant activity of the samples (Tistaert et al. 2011a). These *Mallotus* extracts were also evaluated for their cytotoxic activity. The antiproliferative activity was linked to chromatographic fingerprints and analyzed by means of O-PLS. The evaluation of the acquired regression coefficients led to the recognition of seven potential cytotoxic compounds, two of which were identified as malloapelta A and malloapelta B (Tistaert et al. 2012a). The antioxidant capacity of herbal products containing *Ginkgo biloba* extract was investigated by means of chromatographic fingerprints based on which a calibration model was created by usage of the regression trees method. This model has been shown to be capable of predicting the antioxidant activity of *Ginkgo biloba* products (Ronowicz et al. 2013). Also, *Turnera diffusa* was explored for its antioxidant properties by means of multivariate analysis of the acquired HPLC-PDA fingerprints. The antioxidant capacity could be predicted based on PLS regression. Moreover, peaks, resulting from compounds suspected to be responsible for the antioxidant property, could be designated on the fingerprints by exploring the loading plots and variable importance in projection (Lucio-Gutierrez et al. 2012). Islam et al. (2012) analyzed chromatographic fingerprints of *Epimedium koreanum* extract by means of CDA; the acquired model allows to predict the estrogenic activity of the analyzed samples.

Fingerprints in the analysis of counterfeit medicines

Many published studies concerning the analysis of counterfeit medicines are based on the identification and quantification of the APIs. This approach has the major disadvantage that products can be considered relatively safe, for they might contain the correct API in the correct dosage, while in actual fact potentially toxic secondary components can be present, such as impurities and residual solvents (Deconinck et al. 2013c). As shown in the previous sections, fingerprints have proven their usefulness for the classification of plants. This application could provide an interesting strategy for the detection and characterization of counterfeit medicines as well because it can meet the limitations faced when analyzing only the present APIs. Several analytical techniques have been tested and, consequently, gained their place in the array of techniques available for the detection of counterfeit medicines. Both chromatographic and spectroscopic techniques, aiming at acquiring characteristic fingerprints, have been developed and described in literature. The next sections provide a brief overview of studies using fingerprints and chemometric analysis in the field of counterfeit medicine detection.

Spectroscopic techniques

Spectroscopic techniques are featured by several interesting advantages; they are fast, a minimum or no sample preparation is required, some of them are nondestructive, and they provide fingerprints of the entire sample matrix. The analytical setup of spectroscopic fingerprints is typically to acquire a characteristic fingerprint from both genuine and counterfeit medicines and compare them using chemometric approaches (Scafi and Pasquini 2001, Deisingh 2005, Deconinck et al. 2013c).

Near-infrared spectroscopy

Applications in the NIR region are gaining popularity, as demonstrated by several reviews (Reich 2005, Luypaert et al. 2007, Roggo et al. 2007, de Bleye et al. 2012). NIR spectroscopy allows easy analysis of excipients, which constitute the bulk of a tablet. It is possible that counterfeit medicines contain the correct API in the correct dosage, but the excipient composition will seldom be the same as the genuine sample. This feature makes NIR very valuable for the analysis of counterfeit medicines. Furthermore, NIR also allows to explore physical properties of tablets/capsules (Deisingh 2005). The usefulness of NIR spectroscopy is demonstrated by several published studies.

Vredenbregt et al. (2006) made use of NIR spectroscopy to analyze genuine, counterfeit and imitation samples of Viagra® (Pfizer, New York City, USA). Based on PCA analysis and wavelength correlation of the acquired spectra, it was possible to check the homogeneity of a batch to distinguish counterfeits and imitations from genuine Viagra®, to screen for the presence of sildenafil citrate, and to detect whether similar samples have already been analyzed in the past (Vredenbregt et al. 2006). A study testing NIR spectroscopy for its ability to discriminate between genuine and counterfeit artesunate antimalarial tablets was published by Dowell et al. (2008). PLS has been demonstrated to generate a perfect distinction between genuine and counterfeit samples. Based on a PLS regression model, the wavelengths which mostly influence the acquired discrimination were identified (Dowell et al. 2008). NIR spectroscopy was also tested for its discriminating abilities by Storme-Paris et al. (2010). During this study, hard capsules containing fluoxetine and tablets containing ciprofloxacin were used as standard reference. First, several foreign generic products were analyzed. The acquired results showed that the foreign generics could be distinguished from the standard references by SIMCA. Afterwards, several counterfeit and imitation samples were included in the study. PCA turned out to be capable of differentiating the genuine samples from the counterfeit/imitation ones (Storme-Paris et al. 2010). Genuine and counterfeit Viagra® (Pfizer, New York City, USA) and Cialis® (Eli Lilly, Indianapolis, USA) tablets were analyzed using NIR spectroscopy by the authors' research group (Custers et al. 2016c). In first instance, tablets were removed from the blister and measured as a whole. Both SIMCA and PLS-DA were capable of making a perfect distinction between genuine and counterfeit samples for both pharmaceuticals based on the acquired NIR spectra. Afterwards, all tablets were measured a second time, leaving the samples in the blister, thereby keeping the blister intact. Even in this analytical setup, perfect models differentiating between genuine and counterfeit tablets were acquired (Custers et al. 2016c).

Sabin et al. (2013) used NIR chemical imaging (NIR-CI) for the characterization of tablets containing sildenafil citrate derived from various sources. The authors made use of multivariate curve resolution-alternating least squares (MCR-ALS), which decomposes the hyperspectral data, i.e. data matrix $\boldsymbol{X}$, into the product of a matrix containing the concentration profiles and a matrix consisting of optimized spectra (Tauler 1995, Jaumot et al. 2005, Sabin et al. 2013). They showed that MCR-ALS could be successfully applied to distinguish genuine from counterfeit tablets (Sabin et al. 2013).

A four-stage concept to differentiate between genuine and counterfeit antihypertensive products based on NIR-CI was developed by Puchert et al. (2010). The four stages comprised (1) visual inspection, (2) PCA analysis of the NIR-CI data, (3) classification of samples based on PLS, and (4) multivariate linear image signature data analysis, which consisted of a summation and unfolding of the multidimensional data. This study showed a higher variability in the spatial distribution of the API and some excipients, thereby allowing a clear distinction between genuine and counterfeit samples (Puchert et al. 2010).

Attenuated total reflectance-Fourier transform IR and mid-IR spectroscopy

Mid-IR spectroscopy was explored for its discriminating abilities by our research group (Custers et al. 2016c). The coating of all samples, i.e. genuine and counterfeit Viagra® (Pfizer, New York City, USA) and Cialis® (Eli Lilly, Indianapolis, USA) samples, was accurately scratched off the tablets and analyzed using IR spectroscopy in the mid-IR region (4000–400 cm^{-1}). Genuine samples could be differentiated perfectly from counterfeit samples using PLS-DA and SIMCA (Custers et al. 2016c).

To discriminate between authentic and counterfeit Viagra® (Pfizer, New York City, USA) and Cialis® (Eli Lilly, Indianapolis, USA) samples, Ortiz et al. (2013a) measured the respective fingerprints by means of attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy. For both data sets, a successful distinction between genuine and counterfeit medicines was obtained by using PCA. In addition, the similarity match method was applied to demonstrate that a mixture of powders from a common source might have been used to manufacture counterfeit tablets from distinct seizures (Ortiz et al. 2013a). ATR-FTIR was also evaluated for its discriminating abilities

by Deconinck et al. (2014c). One of the most encountered adulterants when analyzing herbal or dietary supplements intended for slimming purposes is sibutramine. Based on the ATR-FTIR analysis of these kinds of slimming aids and the subsequent analysis of the acquired fingerprints by means of k-NN, all adulterated dietary supplements could be detected (Deconinck et al. 2014c). Anzanello et al. (2013) proposed a method to reduce the number of ATR-FTIR wavelengths (variables) in order to ameliorate the classification of genuine and counterfeit Viagra® (Pfizer, New York City, USA) and Cialis® (Eli Lilly, Indianapolis, USA) samples since the high amount of wavelengths might reduce the performance of chemometric techniques. First, PCA was applied and a variable importance index was included. Next, a backwards iterative procedure of variable elimination was started, which was guided by the importance index. After each elimination step, a k-NN model was constructed by classifying samples as genuine or counterfeit, and the respective classification accuracy was monitored. Their results have shown that for both the Viagra® (Pfizer, New York City, USA) and the Cialis® (Eli Lilly, Indianapolis, USA) data sets, an increase of classification accuracy could be obtained when eliminating low-informative wavelengths (Anzanello et al. 2013). Our research group also explored the usefulness of ATR-FTIR in the field of counterfeit medicines detection (Custers et al. 2015). ATR spectra were measured for genuine and counterfeit samples of Viagra® (Pfizer, New York City, USA) and Cialis® (Eli Lilly, Indianapolis, USA), generic Viagra® samples, and several placebo samples, which were previously shown to contain no sildenafil or tadalafil. SIMCA enabled the construction of a model, which discriminated between samples containing only (1) sildenafil or (2) tadalafil, (3) both sildenafil and tadalafil, and (4) none of both APIs. Furthermore, the SIMCA model could be expanded to a seven-class model, additionally discriminating between genuine Viagra® and Cialis® and generic products of Viagra® as well (Custers et al. 2015).

Raman spectroscopy

Raman spectroscopy has gained its place in the array of techniques used for pharmaceutical analysis (Vankeirsbilck et al. 2002). Recently, a systematic study exploring the benefits and limitations of Raman spectroscopy combined with chemometric techniques for analyzing pharmaceutical samples was performed by Neuberger and Neusüß (2015). This study was conducted based on a large set of model tablets differing in API, excipients, colour dye, or coating. It was shown that Raman spectroscopy, combined with chemometric analysis, is a powerful and fast tool for characterizing samples suspected to be counterfeit. The observed suitability of Raman spectroscopy is due to the fact that the acquired Raman spectra are molecular fingerprints that are strongly associated with the chemical composition of the samples. Therefore, samples with differing coatings or excipients and varying amounts of API can easily be distinguished (Neuberger and Neusüß 2015).

de Veij et al. (2008) proposed Raman spectroscopy as a fast and reliable method for the detection of counterfeit Viagra® (Pfizer, New York City, USA) tablets. Raman spectra were measured from genuine and counterfeit samples and subsequently analyzed using PCA and HCA. It could be concluded that a combined approach of PCA and HCA with Raman spectroscopy allows an automated approach to discriminate between genuine and counterfeit Viagra® tablets (de Veij et al. 2008). Raman microscopy imaging was applied on a sample set consisting of genuine, counterfeit, and imitation products of Viagra® (Pfizer, New York City, USA). The full spectra were used as fingerprints in a chemometric analysis consisting of PCA, k-NN, and SIMCA. Both k-NN and SIMCA were able to generate models that perfectly described and predicted the genuine/illegal nature of all samples (Sacre et al. 2011b). Based on Raman spectroscopy, an analytical strategy to detect and classify counterfeit medicines was developed by Degardin et al. (2011). The first step of the analytical strategy enables the identification of pharmaceutical preparations and the detection of their counterfeits based on SVMs and correlation testing. In case of detecting a counterfeit, the second step consists in its chemical profiling. This second step made use of PCA and correlation distance measures to classify samples in one of the counterfeit classes, which are defined based on the chemical composition of the respective constituent counterfeits (Degardin et al. 2011). Raman microscopy was successfully applied by Kwok and Taylor (2012a) to discriminate genuine from counterfeit Cialis® (Eli Lilly, Indianapolis, USA) tablets. Multivariate curve resolution was used to analyze the acquired spectra and to identify the present APIs and excipients. Comparison of the spectra acquired for genuine and counterfeit samples revealed both similarities and dissimilarities regarding the identity of the excipients, the quantity of API, and the distribution of the components (Kwok and Taylor 2012a). Furthermore, the authors used Raman spectroscopy and two-dimensional correlation spectroscopy to analyze the packages of both genuine and counterfeit Cialis® (Eli Lilly, Indianapolis, USA) samples (Kwok and Taylor 2012b). The study focused on the white and yellow colour regions from which different components could be distinguished. These differences in components enable a clear distinction between genuine and counterfeit samples (Kwok and Taylor 2012b). Lu et al.

(2013) tested a portable Raman spectroscopy system to distinguish between genuine and counterfeit hypoglycemic tablets. The combination of Raman spectroscopy with data analysis by means of local straight-line screening (LSLS) proved to be a successful high-throughput screening approach (Lu et al. 2013). More information about the used LSLS methodology can be found in the studies of Lu et al. (2007) and Zhu et al. (2009).

Nuclear magnetic resonance spectroscopy

NMR spectroscopy is often used in the analysis of counterfeit medicines because of its capability to elucidate the structure of unknown molecules such as impurities and structural analogues of the APIs (Holzgrabe and Malet-Martino 2011). Analogues are the result of minor structural modifications of the parent molecule (i.e. an approved API). Counterfeiters make use of such analogues in an attempt to circumvent legal prosecutions (Poon et al. 2007). Furthermore, NMR allows to perform trace analysis and semiquantitative analyses. When analyzing counterfeit medicines, mostly ^{1}H and ^{13}C NMR spectroscopy are used (Holzgrabe and Malet-Martino 2011). Its use in the field of pharmaceutical analysis has been reviewed by Holzgrabe and Malet-Martino (2011) and Holzgrabe et al. (2005). However, NMR is characterized by some notable disadvantages; it requires not only very expensive equipment but also well-trained scientists to operate the equipment and to interpret the data (Deconinck et al. 2013c).

Silvestre et al. (2009) used quantitative isotopic ^{13}C NMR to define site-specific isotopic profiles of aspirin and paracetamol. Their study has shown that isotopic profiling can have its merit in detecting counterfeiting and patent infringement in the pharmaceutical industry. ^{13}C NMR was also used by Bussy et al. (2011) to investigate the isotopic fingerprints of different ibuprofen samples. This strategy may be used to characterize batch-to-batch production, specific manufacturing processes, and the origin of raw materials used in the process.

Combination of spectroscopic techniques

An approach that is increasingly becoming popular in the analysis of counterfeit medicines is to combine different spectroscopic techniques. The combination of different spectra can improve the classification and predictive properties of the respective models compared with those based on only one type of spectrum. The combination of multiple spectroscopic techniques could be a valuable approach in the fight against counterfeit medicines because the quality of counterfeit products is rising (Deisingh 2005, Deconinck et al. 2013c, Neuberger and Neusüß 2015). An example of such a study is provided by Sacre et al. (2010). A set of genuine, counterfeit, and imitation samples of both Viagra® (Pfizer, New York City, USA) and Cialis® (Eli Lilly, Indianapolis, USA) were analyzed using Raman spectroscopy, NIR spectroscopy, and FTIR spectroscopy in the mid-IR region. The acquired fingerprints were analyzed using PLS-DA. For the Viagra® set, the best discrimination between genuine and counterfeit samples was acquired when combining the fingerprints from FTIR and NIR spectroscopy; the combination of NIR and Raman spectroscopy provided the best results for the Cialis® data set. It could be concluded that combining spectra from different techniques improved the classification and predictive properties of the models compared with the models calculated based on only one type of spectrum (Sacre et al. 2010). Deconinck et al. (2012b) made use of these data sets as well and analyzed them by means of CART. For both the Viagra® (Pfizer, New York City, USA) and the Cialis® (Eli Lilly, Indianapolis, USA) sets, each data set, acquired by the three spectroscopic techniques, is analyzed multiple times: (1) all three data sets individually, (2) all possible pairs of the data sets, and (3) all three data sets combined. It was not only attempted to differentiate between genuine and counterfeit medicines; counterfeit samples were also classified according to their classification defined by the Dutch National Institute for Public Health and the Environment (RIVM). The CART models obtained based on FTIR and NIR spectroscopic analysis of the Viagra® samples resulted in a good classification of the counterfeit samples. Furthermore, for both models, a perfect distinction between genuine and counterfeit samples was obtained. The combination of all three spectroscopic data sets did not result in an amelioration of the acquired classification models. For the Cialis® data set, the best CART model was acquired by combining the NIR and the Raman spectra (Deconinck et al. 2012b). Li et al. (2014) published an interesting study showing the merit of combining multiple spectroscopic techniques as well. They demonstrated the usefulness of Raman and NIR spectroscopy for the identification of anisodamine tablets. Raman spectroscopy proved useful for identifying counterfeit tablets using the similarity coefficient method; NIR spectroscopy proved capable of differentiating tablets originating from different manufacturing plants based on a PLS-DA model. Another published study compared NIR and Raman spectroscopy as rapid screening methods to distinguish genuine from counterfeit Lipitor® (Pfizer, New York City, USA). PLS-DA analysis of the generated profiles has shown that both spectroscopic techniques are suitable

for the aimed differentiation. Moreover, PLS-DA models based on NIR and Raman fingerprints also proved capable of distinguishing between counterfeits containing either atorvastatin or lovastatin (de Peinder et al. 2008). da Silva Fernandes et al. (2012) have shown that combining chemometric analysis of spectra acquired by NIR or fluorescence spectroscopy is a powerful means to detect tablets adulterated with glibenclamide in a nondestructive way. SIMCA and PLS-DA generated a perfect classification based on the NIR data. Unfolded PLS-DA showed a 100% correct discrimination using the fluorescence spectra.

Other spectroscopic techniques are also found in literature; for instance, the measurement of characteristic fingerprints of counterfeit medicines by X-ray fluorescence spectrometry and subsequent analysis by chemometric tools is described by Ortiz et al. (2012). Isotopic ratios, acquired by means of isotope-ratio mass spectrometry, and subsequent analysis by chemometrics were evaluated as a tool for discriminating APIs originating from different geographical areas or synthetic processes (Deconinck et al. 2008). Rodomonte et al. (2010) made use of colorimetry to analyze and compare the packaging of genuine and counterfeit samples of Viagra® (Pfizer, New York City, USA), Cialis® (Eli Lilly, Indianapolis, USA), and Levitra® (Bayer, Leverkusen, Germany). A combination of image processing and statistical analysis was used by Jung et al. (2012) to distinguish between genuine and counterfeit samples of Viagra® (Pfizer, New York City, USA) and Cialis® (Eli Lilly, Indianapolis, USA).

Chromatographic techniques

Chromatographic techniques are often used for counterfeit fingerprinting as well, owing to its widespread use and standard availability in nearly all laboratories for medicinal control. These techniques generate highly informative fingerprints since they spread the information about the chemical composition of a sample over time, thereby allowing compounds to be individually detected and revealing additional information (such as quantitative or structural information) depending on the coupled detector (Deconinck et al. 2013c).

Liquid chromatography

LC has certain advantages that make it highly suitable for fingerprint analysis. Overall, it is easy to operate, and it allows different analytical setups (e.g. reverse-phase chromatography, hydrophilic interaction chromatography, and ion-exchange chromatography), which enables highly different analytes to be analyzed. Furthermore, this technique is characterized by high resolution, selectivity, and sensitivity (Deconinck et al. 2013c).

Sacre et al. (2011a) performed a study during which impurity fingerprints of genuine, counterfeit, and imitation products of Viagra® (Pfizer, New York City, USA) and Cialis® (Eli Lilly, Indianapolis, USA) were measured using HPLC-PDA and analyzed by means of chemometrics (i.e. k-NN and SIMCA). It was shown that k-NN results in suitable models capable of predicting the genuine or counterfeit nature of samples. Chromatographic fingerprints were also recorded for a set of genuine and counterfeit Viagra® (Pfizer, New York City, USA) and Cialis® (Eli Lilly, Indianapolis, USA) samples by Deconinck et al. (2012c). These fingerprints were subsequently evaluated for their use in the detection and classification of genuine and counterfeit medicines based on exploratory and modeling techniques. Besides differentiation between genuine and counterfeit medicines, this study also focused on differences between classes of counterfeit samples as defined by RIVM. PP revealed differences between genuine and counterfeit medicines; differences among the distinct groups of counterfeit samples were disclosed by HCA. SVM and SIMCA resulted for both sample sets in the best models that were able to generate a perfect discrimination between genuine and counterfeit medicines and high correct classification rates for the classification in the different counterfeit classes (Deconinck et al. 2012c). The chemometric treatment of HPLC fingerprints has also shown its usefulness to link paracetamol-containing drugs to their synthesis pathway. Paracetamol was extracted from the pharmaceutical formulation and subsequently analyzed using trace-enrichment HPLC. The analysis of the acquired fingerprints by PP, HCA, and AAMRT led to the successful distinction between four different synthesis pathways. This study shows that the chemometric analysis of impurity fingerprints can have its merit in the detection of patent infringement (Dumarey et al. 2009). Schneider and Wessjohann (2010) made use of HPLC-UV and HPLC-MS/MS to obtain impurity profiles of Xenical® (Roche, Basel, Switzerland) and two of its generic products [Cobese® (Ranbaxy Laboratories Limited, New Delhi, India) and Orsoten® (KRKA, Novo Mesto, Slovenia)]. Their study has shown that the generic formulations contain higher levels of impurities than the original product. Impurity profiles from morphine samples, derived from different origins, were measured by Acevska et al. (2015) using the HPLC-PDA method described in the *European Pharmacopoeia* for the evaluation of "related substances" and an own developed HPLC-PDA-MS method. However, the latter proved to be more suitable because

it allows unambiguous identification of the impurities. PCA and HCA resulted in distinct clusters of the morphine samples according to their origin. Using O-PLS, a discrimination between the different classes was obtained, thereby allowing the identification of the origin of the morphine samples and, accordingly, revealing falsifications of these products (Acevska et al. 2015). Our research group measured impurity profiles of genuine and counterfeit samples of Viagra® (Pfizer, New York City, USA) and Cialis® (Eli Lilly, Indianapolis, USA) and generic products of Viagra® by means of HPLC-PDA and HPLC-MS (Custers et al. 2016a,b). Both types of fingerprints, i.e. PDA and MS fingerprints, were tested individually and in combination for their discriminating properties. For both the Viagra® and the Cialis® data set, good diagnostic models were already acquired based on the two types of fingerprints separately using either k-NN, SIMCA or PLS-DA. However, taking all three modeling techniques into account, the combination of PDA and MS fingerprints resulted in less classification errors between genuine/generic and counterfeit medicines compared with the PDA and MS data separately (Custers et al. 2016a,b).

Also, ultraperformance liquid chromatography (UHPLC) was tested in the field of counterfeit medicine fingerprinting. Ortiz et al. (2013b) analyzed both genuine and counterfeit samples of Viagra® (Pfizer, New York City, USA) and Cialis® (Eli Lilly, Indianapolis, USA) by means of UHPLC coupled to MS. Based on PCA and HCA, they demonstrated UHPLC-MS to be a valuable tool to differentiate between genuine and counterfeit samples. Moreover, samples could also be grouped in five classes depending on the API content.

Gas chromatography

Although GC is widely available, its use as a fingerprinting method is rather limited. However, GC has proven its usefulness for detecting volatile compounds, such as residual solvents, in counterfeit medicines (Deconinck et al. 2013c).

The authors' research group developed and validated a headspace GC-MS method to detect and quantify residual solvents in counterfeit medicines. This method was shown to be fast and suitable for routine analysis, and only limited sample preparation was required (Deconinck et al. 2012a). It was subsequently applied to a set of genuine and counterfeit Viagra® (Pfizer, New York City, USA) and Cialis® (Eli Lilly, Indianapolis, USA) samples, which has shown that all counterfeit samples have higher residual solvent content compared with the genuine samples. In addition, the presence of toxic residual solvents, such as dichloromethane and carbon tetrachloride, was detected in several counterfeit samples (Deconinck et al. 2013a). The chromatograms, acquired for this sample set, were subsequently used as fingerprints in a chemometrical data analysis aiming at classifying counterfeit samples based on their residual solvent content. This study demonstrated that suitable and reliable diagnostic models can be obtained based on SIMCA; counterfeit samples containing carbon tetrachloride can be distinguished from samples containing only ethanol or 2-propanol. Therefore, it was shown that GC-MS fingerprints combined with chemometric analysis can result in a suitable prediction model, which gives a prime notion of the public health risks counterfeit products might pose (Custers et al. 2014).

Combination of spectroscopic and separation techniques

Anzanello et al. (2014) tested several analytical methods and chemometric techniques (i.e. PCA, SVM, and k-NN) to identify which analytical techniques provide the most relevant data to distinguish genuine from counterfeit medicines. These techniques were tested on a sample set consisting of genuine and counterfeit Viagra® (Pfizer, New York City, USA) and Cialis® (Eli Lilly, Indianapolis, USA) samples. The included analytical methods were physical profiling, X-ray fluorescence, direct infusion electrospray ionization mass spectrometry, UHPLC coupled to MS, and ATR-FTIR. Based on this analytical setup, the authors found UHPLC-MS, physical profiling and ATR-FTIR to yield the highest classification accuracy (Anzanello et al. 2014). A strategy for the classification of counterfeit medicines was proposed by Been et al. (2011). NIR and Raman spectroscopy were used to establish a database containing the spectra of both counterfeit and genuine samples. All samples were analyzed by GC-MS and FTIR as well to determine the respective chemical profiles. Based on the acquired spectra and the chemical profiles, unsupervised chemometric techniques were used to identify classes of counterfeits which were subsequently used in the supervised data analysis. An approach based on distance measures and receiver operating characteristics curves proved to be most successful (Been et al. 2011). Rodionova et al. (2010) tested NIR spectroscopy, GC-MS, HPLC-PDA-MS and capillary electrophoresis coupled to UV aiming at differentiating between genuine and counterfeit ampoules of dexamethasone. The acquired NIR spectra were analyzed using PCA and SIMCA, and this combination was shown to be valuable to distinguish genuine ampoules from counterfeit

ones. Of the tested separation techniques, all methods, but GC-MS, revealed differences in impurities, thereby allowing to differentiate between both types of ampoules (Rodionova et al. 2010).

Practical application of counterfeit fingerprinting: investigation of common sources

Several studies included in the overview of counterfeit medicine fingerprinting show a very useful application of fingerprinting, i.e. the investigation of potential common sources of counterfeit and illegal pharmaceuticals. As already mentioned in the Introduction section, medicine counterfeiting is a global problem which is most likely fuelled by a structured criminal industry consisting of manufacturers, wholesalers, distributors and local sellers (Degardin et al. 2014). As a consequence, there is a substantial possibility that among the large amounts of seized counterfeit and illegal medicines, several products are derived from common sources. Several analytical techniques have already been used to investigate the sources of counterfeit medicines. Ortiz et al. (2013a) showed by means of ATR-FTIR that powders from a common source might have been used in the manufacturing of counterfeit tablets from distinct seizures. ^{13}C NMR was also used to explore the origin of raw materials used in the manufacturing process of different ibuprofen samples (Bussy et al. 2011). Another technique, used in this research context, was isotope-ratio mass spectrometry, which enabled to discriminate APIs originating from different geographical areas (Deconinck et al. 2008). Dumarey et al. (2009) analyzed several paracetamol-containing samples by means of trace-enrichment HPLC, which showed that the acquired fingerprints reveal information about the underlying synthesis pathways. Information of synthesis pathways could reveal cases of patent infringement and/or illegal production of APIs. The same separation technique, coupled to MS, was used by Acevska et al. (2015) to acquire impurity fingerprints of morphine samples, which allow the identification of the origin of the considered samples and, accordingly, revealing falsifications.

This mode of working with fingerprints provides exceptionally interesting information about counterfeit and illegal medicines because these kinds of data are particularly useful to support competent authorities and police/customs services charged with closing down the distribution of counterfeit medicines.

Discussion and conclusion

The aim of this review was to provide a general overview of the role of fingerprints and their multivariate analysis in the detection of counterfeit medicines. It has been shown that the fingerprinting approach is a valuable tool for the analysis of counterfeit and illegal pharmaceuticals. Both chromatographic and spectroscopic techniques have been proven useful for this purpose. Fingerprints allow to discriminate genuine medicines from counterfeit ones; depending on the used analytical technique, fingerprints identify and quantify APIs and other predominant signals. Furthermore, they provide a complete image of the product.

Each of these techniques is featured by their respective advantages and disadvantages. Spectroscopic techniques, for instance, are a "whole sample approach" which can be both an advantage and a disadvantage. The advantage is that by means of multivariate data analysis, a spectrum of a counterfeit medicine can easily be compared with that of a genuine, thereby allowing a fast discrimination between genuine and counterfeit pharmaceuticals. The disadvantage of this approach is that when aiming at detecting certain compounds, the respective compounds need to be present in a considerable amount. Furthermore, the matrix of a sample can mask the compound of interest or possibly interferes with its detection. However, it should be mentioned that when measuring fingerprints, the mentioned disadvantage is not very significant because one is generally interested in the entire profile and not in the presence of specific compounds. When analyzing adulterated samples, the mentioned disadvantage does reduce the applicability of spectroscopic techniques because, in this particular case, one is interested in the presence and quantification of certain compounds. Notwithstanding, spectroscopic techniques are fast, some of them are non-destructive, and no or little sample preparation is required. Therefore, spectroscopic techniques remain very useful tools in the fingerprint analysis of counterfeit medicines.

The comparison of the different spectroscopic techniques shows that Raman spectroscopy, alone or in combination with other spectroscopic techniques, is a very widely applied technique. This can be attributed to the fact that Raman is more specific for API and chemical content compared with other spectroscopic techniques. The information obtained by Raman is due to fundamental vibrational modes, especially for aromatic compounds. Furthermore, Raman is also characterized by high peak resolution (Degardin et al. 2011).

Chromatographic techniques have proven useful for the fingerprint analysis of counterfeit pharmaceuticals

as well. By separating the compounds of a sample, each compound can be detected individually, which generates highly informative fingerprints. The disadvantages of such techniques are that they are destructive and more time consuming compared to spectroscopic techniques. However, separation techniques are more suitable when one wants to identify and quantify certain signals present in the fingerprint. Therefore, it is necessary to previously reflect on what information one wants to acquire and select an appropriate technique in function of this information needed since chromatographic and spectroscopic techniques both have their respective advantages and disadvantages.

When analyzing fingerprints, proper data analysis is indispensable to extract useful information from the acquired data. Chemometrics (multivariate analysis) provides the means to perform suitable analysis of the large amounts of data, which fingerprints very often constitute. However, chemometrics include a large variety of techniques for data preprocessing, exploratory pattern recognition and supervised pattern recognition. It is important to select appropriate techniques; therefore, a good knowledge of current techniques is necessary.

To conclude, both spectroscopic and chromatographic techniques have their merit in the fingerprint analysis of counterfeit medicines. However, which technique is best used depends on the purpose of the study. If one attempts to make a simple distinction between genuine and counterfeit medicines, IR spectroscopic techniques might suffice. If, however, one wants to obtain additional information, such as identification of APIs or structural information, one probably opts for the use of separation techniques, often coupled to MS or NMR.

Unfortunately, despite all efforts to tackle the distribution of counterfeit medicines and methods described in literature to distinguish genuine from counterfeit pharmaceuticals, the presence of the latter cannot be entirely excluded. Therefore, authorities, health practitioners and patients should remain vigilant, and patients should be discouraged to purchase their medicines from dubious sources. As a consequence, the development of efficient analytical techniques for the detection and risk evaluation of counterfeit medicines should continue. These techniques play a major role in supporting health authorities and in fighting pharmaceutical crime.

References

Acevska, J.; Stefkov, G.; Cvetkovikj, I.; Petkovska, R.; Kulevanova, S.; Cho, J.; Dimitrovska, A. Fingerprinting of morphine using chromatographic purity profiling and multivariate data analysis. *J. Pharm. Biomed. Anal.* **2015,** *109*, 18–27.

Alaerts, G.; Dejaegher, B.; Smeyers-Verbeke, J.; Vander Heyden, Y. Recent developments in chromatographic fingerprints from herbal products: set-up and data analysis. *Comb. Chem. High Throughput. Screen.* **2010a,** *13*, 900–922.

Alaerts, G.; Merino-Arevalo, M.; Dumarey, M.; Dejaegher, B.; Noppe, N.; Matthijs, N.; Smeyers-Verbeke, J.; Vander Heyden, Y. Exploratory analysis of chromatographic fingerprints to distinguish rhizoma Chuanxiong and rhizoma Ligustici. *J. Chromatogr. A* **2010b,** *1217*, 7706–7716.

Alaerts, G.; Van, E. J.; Pieters, S.; Dumarey, M.; van Nederkassel, A. M.; Goodarzi, M.; Smeyers-Verbeke, J.; Vander Heyden, Y. Similarity analyses of chromatographic fingerprints as tools for identification and quality control of green tea. *J. Chromatogr. B. Analyt. Technol Biomed. Life Sci* **2012,** *910*, 61–70.

Alaerts, G.; Pieters, S.; Logie, H.; Van, E. J.; Merino-Arevalo, M.; Dejaegher, B.; Smeyers-Verbeke, J.; Vander Heyden, Y. Exploration and classification of chromatographic fingerprints as additional tool for identification and quality control of several Artemisia species. *J. Pharm. Biomed. Anal.* **2014,** *95*, 34–46.

Almeida, J. A. S.; Barbosa, L. M. S.; Pais, A. A. C. C.; Formosinho, S. J. Improving hierarchical cluster analysis: a new method with outlier detection and automated clustering. *Chemometr. Intell. Lab. Syst.* **2007,** *87*, 208–217.

Anzanello, M. J.; Ortiz, R. S.; Limbergerb, R. P.; Mayorga, P. A multivariate-based wavenumber selection method for classifying medicines into authentic or counterfeit classes. *J. Pharm. Biomed. Anal.* **2013,** *83*, 209–214.

Anzanello, M. J.; Ortiz, R. S.; Limberger, R.; Mariotti, K. A framework for selecting analytical techniques in profiling authentic and counterfeit Viagra and Cialis. *Forensic Sci. Int.* **2014,** *235*, 1–7.

Barker, M.; Rayens, W. Partial least squares for discrimination. *J. Chemom.* **2003,** *17*, 166–173.

Been, F.; Roggo, Y.; Degardin, K.; Esseiva, P.; Margot, P. Profiling of counterfeit medicines by vibrational spectroscopy. *Forensic Sci. Int.* **2011,** *211*, 83–100.

Belousov, A. I.; Verzakov, S. A.; von Frese, J. Applicational aspects of support vector machines. *J. Chemometrics* **2002,** *16*, 482–489.

Bertrand, D.; Courcoux, P.; Autran, J. C.; Meritan, R.; Robert, P. Stepwise canonical discriminant analysis of continuous digitalized signals: application to chromatograms of wheat proteins. *J. Chemometrics* **1990,** *4*, 413–427.

Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. Classification and Regression Trees; Wadsworth & Brooks: Monterey, 1984.

Bussy, U.; Thibaudeau, C.; Thomas, F.; Desmurs, J. R.; Jamin, E.; Remaud, G. S.; Silvestre, V.; Akoka, S. Isotopic finger-printing of active pharmaceutical ingredients by 13C NMR and polarization transfer techniques as a tool to fight against counterfeiting. *Talanta* **2011,** *85*, 1909–1914.

Caetano, S.; Decaestecker, T.; Put, R.; Daszykowski, M.; Van Bocxlaer, J.; Vander Heyden, Y. Exploring and modelling the responses of electrospray and atmospheric pressure chemical ionization techniques based on molecular descriptors. *Anal. Chim. Acta* **2005,** *550*, 92–106.

Caudron, J. M.; Ford, N.; Henkens, M.; Mace, C.; Kiddle-Monroe, R.; Pinel, J. Substandard medicines in resource-poor settings: a problem that can no longer be ignored. *Trop. Med. Int. Health* **2008,** *13*, 1062–1072.

Centner, V.; Massart, D. L.; de Noord, O. E.; de, J. S.; Vandeginste, B. M.; Sterna, C. Elimination of uninformative variables for multivariate calibration. *Anal. Chem.* **1996,** *68*, 3851–3858.

Chemometrics in Food Chemistry; 1st Edition. Federico Marini., Ed. Elsevier: Amsterdam, 2013.

Chen, Y.; Zhu, S. B.; Xie, M. Y.; Nie, S. P.; Liu, W.; Li, C.; Gong, X. F.; Wang, Y. X. Quality control and original discrimination of *Ganoderma lucidum* based on high-performance liquid chromatographic fingerprints and combined chemometrics methods. *Anal. Chim. Acta* **2008,** *623*, 146–156.

Council of Europe. Counterfeiting of medical products (MEDICRIME). http://www.coe.int/medicrime (2014) (last accessed 30.03.2016).

Custers, D.; Canfyn, M.; Courselle, P.; De Beer, J. O.; Apers, S.; Deconinck, E. Headspace-gas chromatographic fingerprints to discriminate and classify counterfeit medicines. *Talanta* **2014,** *123*, 78–88.

Custers, D.; Cauwenbergh, T.; Bothy, J. L.; Courselle, P.; De Beer, J. O.; Apers, S.; Deconinck, E. ATR-FTIR spectroscopy and chemometrics: an interesting tool to discriminate and characterize counterfeit medicines. *J. Pharm. Biomed. Anal.* **2015,** *112*, 181–189.

Custers, D.; Krakowska, B.; De Beer, J. O.; Courselle, P.; Daszykowski, M.; Apers, S.; Deconinck, E. Testing of complementarity of PDA and MS detectors using chromatographic fingerprinting of genuine and counterfeit samples containing sildenafil citrate. *Anal. Bioanal. Chem.* **2016a**, *408*, 1643–1656.

Custers, D.; Krakowska, B.; De Beer, J. O.; Courselle, P.; Daszykowski, M.; Apers, S.; Deconinck, E. Chromatographic impurity fingerprinting of genuine and counterfeit Cialis® as a means to compare the discriminating ability of PDA and MS detection. *Talanta* **2016b,** *146*, 540–548.

Custers, D.; Vandemoortele, S.; Bothy, J. L.; De Beer, J. O.; Courselle, P.; Apers, S.; Deconinck, E. Physical profiling and IR spectroscopy: simple and effective methods to discriminate between genuine and counterfeit samples of Viagra® and Cialis®. *Drug Test. Anal.* **2016c,** *8*, 378–387.

da Silva Fernandes, R.; da Costa, F. S.; Valderrama, P.; Marco, P. H.; de Lima, K. M. Nondestructive detection of adulterated tablets of glibenclamide using NIR and solid-phase fluorescence spectroscopy and chemometric methods. *J. Pharm. Biomed. Anal.* **2012,** *66*, 85–90.

Daszykowski, M. From projection pursuit to other unsupervised chemometric techniques. *J. Chemom.* **2007,** *21*, 270–279.

Daszykowski, M.; Walczak, B. Methods for the exploratory analysis of two-dimensional chromatographic signals. *Talanta* **2011,** *83*, 1088–1097.

Daszykowski, M.; Walczak, B.; Massart, D. L. Projection methods in chemistry. *Chemometr. Intell. Lab. Syst.* **2003,** *65*, 97–112.

Daszykowski, M.; Stanimirova, I.; Walczak, B.; Coomans, D. Explaining a presence of groups in analytical data in terms of original variables. *Chemometr. Intell. Lab. Syst.* **2005,** *78*, 19–29.

Daszykowski, M.; Vander Heyden, Y.; Walczak, B. Robust partial least squares model for prediction of green tea antioxidant capacity from chromatograms. *J. Chromatogr. A* **2007,** *1176*, 12–18.

Daszykowski, M.; Danielsson, R.; Walczak, B. No-alignment-strategies for exploring a set of two-way data tables obtained from capillary electrophoresis-mass spectrometry. *J. Chromatogr. A* **2008,** *1192*, 157–165.

Daszykowski, M.; Vander Heyden, Y.; Boucon, C.; Walczak, B. Automated alignment of one-dimensional chromatographic fingerprints. *J. Chromatogr. A* **2010,** *1217*, 6127–6133.

de Bleye, C.; Chavez, P. F.; Mantanus, J.; Marini, R.; Hubert, P.; Rozet, E.; Ziemons, E. Critical review of near-infrared spectroscopic methods validations in pharmaceutical applications. *J. Pharm. Biomed. Anal.* **2012,** *69*, 125–132.

de Peinder, P.; Vredenbregt, M. J.; Visser, T.; de Kaste, D. Detection of Lipitor counterfeits: a comparison of NIR and Raman spectroscopy in combination with chemometrics. *J. Pharm. Biomed. Anal.* **2008,** *47*, 688–694.

de Veij, M.; Deneckere, A.; Vandenabeele, P.; de Kaste, D.; Moens, L. Detection of counterfeit Viagra with Raman spectroscopy. *J. Pharm. Biomed. Anal.* **2008,** *46*, 303–309.

Deconinck, E.; Hancock, T.; Coomans, D.; Massart, D. L.; Vander Heyden, Y. Classification of drugs in absorption classes using the classification and regression trees (CART) methodology. *J. Pharm. Biomed. Anal.* **2005,** *39*, 91–103.

Deconinck, E.; Zhang, M. H.; Coomans, D.; Vander Heyden, Y. Classification tree models for the prediction of blood-brain barrier passage of drugs. *J. Chem. Inf. Model.* **2006,** *46*, 1410–1419.

Deconinck, E.; van Nederkassel, A. M.; Stanimirova, I.; Daszykowski, M.; Bensaid, F.; Lees, M.; Martin, G. J.; Desmurs, J. R.; Smeyers-Verbeke, J.; Vander Heyden, Y. Isotopic ratios to detect infringements of patents or proprietary processes of pharmaceuticals: two case studies. *J. Pharm. Biomed. Anal.* **2008,** *48*, 27–41.

Deconinck, E.; Canfyn, M.; Sacre, P. Y.; Baudewyns, S.; Courselle, P.; De Beer, J. O. A validated GC-MS method for the determination and quantification of residual solvents in counterfeit tablets and capsules. *J. Pharm. Biomed. Anal.* **2012a,** *70*, 64–70.

Deconinck, E.; Sacre, P. Y.; Coomans, D.; De Beer, J. O. Classification trees based on infrared spectroscopic data to discriminate between genuine and counterfeit medicines. *J. Pharm. Biomed. Anal.* **2012b,** *57*, 68–75.

Deconinck, E.; Sacre, P. Y.; Courselle, P.; De Beer, J. O. Chemometrics and chromatographic fingerprints to discriminate and classify counterfeit medicines containing PDE-5 inhibitors. *Talanta* **2012c,** *100*, 123–133.

Deconinck, E.; Canfyn, M.; Sacre, P. Y.; Courselle, P.; De Beer, J. O. Evaluation of the residual solvent content of counterfeit tablets and capsules. *J. Pharm. Biomed. Anal.* **2013a,** *81–82*, 80–88.

Deconinck, E.; De Leersnijder, C.; Custers, D.; Courselle, P.; De Beer, J. O. A strategy for the identification of plants in illegal pharmaceutical preparations and food supplements using chromatographic fingerprints. *Anal. Bioanal. Chem.* **2013b,** *405*, 2341–2352.

Deconinck, E.; Sacre, P. Y.; Courselle, P.; De Beer, J. O. Chromatography in the detection and characterization of illegal pharmaceutical preparations. *J. Chromatogr. Sci.* **2013c,** *51*, 791–806.

Deconinck, E.; Andriessens, S.; Bothy, J. L.; Courselle, P.; De Beer, J. O. Comparative dissolution study on counterfeit medicines of PDE-5 inhibitors. *J. Pharm. Biomed. Anal.* **2014a,** *4*, 250–257.

Deconinck, E.; Bothy, J. L.; Desmedt, B.; Courselle, P.; De Beer, J. O. Detection of whitening agents in illegal cosmetics using attenuated total reflectance-infrared spectroscopy. *J. Pharm. Biomed. Anal.* **2014b,** *98*, 178–185.

Deconinck, E.; Cauwenbergh, T.; Bothy, J. L.; Custers, D.; Courselle, P.; De Beer, J. O. Detection of sibutramine in adulter-

ated dietary supplements using attenuated total reflectance-infrared spectroscopy. *J. Pharm. Biomed. Anal.* **2014c,** *100*, 279–283.

Degardin, K.; Roggo, Y.; Been, F.; Margot, P. Detection and chemical profiling of medicine counterfeits by Raman spectroscopy and chemometrics. *Anal. Chim. Acta* **2011,** *705*, 334–341.

Degardin, K.; Roggo, Y.; Margot, P. Understanding and fighting the medicine counterfeit market. *J. Pharm. Biomed. Anal.* **2014,** *87*, 167–175.

Deisingh, A. K. Pharmaceutical counterfeiting. *Analyst* **2005,** *130*, 271–279.

Delepierre, A.; Gayot, A.; Carpentier, A. Update on counterfeit antibiotics worldwide, public health risks. *Med. Mal Infect.* **2012,** *42*, 247–255.

Directive 2011/62/EU of the European Parliament and of the Council of 8 June 2011 amending Directive 2001/83/EC on the community code relating to medicinal products for human use, as regards the prevention of the entry into the legal supply chain of falsified medicinal products. *Official Journal of the European Union* **2011**, *54*, 74–87.

Donno, D.; Boggia, R.; Cerutti, A. K.; Guido, M.; Mellano, M. G.; Prgomet, Z.; Beccaro, G. L. Phytochemical fingerprint and chemometrics for natural food preparation pattern recognition: an innovative technique in food supplement quality control. *J. Food Sci. Technol.* **2015,** *53*, 1071–1083.

Dowell, F. E.; Maghirang, E. B.; Fernandez, F. M.; Newton, P. N.; Green, M. D. Detecting counterfeit antimalarial tablets by near-infrared spectroscopy. *J. Pharm. Biomed. Anal.* **2008,** *48*, 1011–1014.

Drab, K.; Daszykowski, M. Clustering in analytical chemistry. *J. AOAC Int.* **2014,** *97*, 29–38.

Drucker, H.; Wu, D.; Vapnik, V. N. Support vector machines for spam categorization. *IEEE Trans. Neural Netw.* **1999,** *10*, 1048–1054.

Dumarey, M.; van Nederkassel, A. M.; Deconinck, E.; Vander Heyden, Y. Exploration of linear multivariate calibration techniques to predict the total antioxidant capacity of green tea from chromatographic fingerprints. *J. Chromatogr. A* **2008,** *1192*, 81–88.

Dumarey, M.; van Nederkassel, A. M.; Stanimirova, I.; Daszykowski, M.; Bensaid, F.; Lees, M.; Martin, G. J.; Desmurs, J. R.; Smeyers-Verbeke, J.; Vander, H. Y. Recognizing paracetamol formulations with the same synthesis pathway based on their trace-enriched chromatographic impurity profiles. *Anal. Chim. Acta* **2009,** *655*, 43–51.

Dumarey, M.; Smets, I.; Vander, H. Y. Prediction and interpretation of the antioxidant capacity of green tea from dissimilar chromatographic fingerprints. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2010,** *878*, 2733–2740.

Eilers, P. H. Parametric time warping. *Anal. Chem.* **2004,** *76*, 404–411.

European Alliance for Access to Safe Medicines (EAASM). The Counterfeiting Superhighway. http://www.eaasm.eu/eaasm-reports (2008) (last accessed 30.03.2016).

European Alliance for Access to Safe Medicines (EAASM). About EAASM. http://www.eaasm.eu/about-the-eaasm (2015) (last accessed 30.03.2016).

European Directorate for the Quality of Medicines and HealthCare (EDQM). EDQM begins development of a demonstrated of its' future 'Track and Trace' system. http://www.edqm.eu/medias/fichiers/The_EDQM_begins_development_of_a_demonstrator_of_i.pdf (2011) (last accessed 30.03.2016).

European Directorate for the Quality of Medicines and HealthCare (EDQM), 2.2.40. Near-infrared spectroscopy, *European Pharmacopoeia*, 8.3 ed., Strasbourg, 2015.

Gautam, C. S.; Utreja, A.; Singal, G. L. Spurious and counterfeit drugs: a growing industry in the developing world. *Postgrad. Med. J.* **2009,** *85*, 251–256.

Goodarzi, M.; Russell, P. J.; Vander Heyden, Y. Similarity analyses of chromatographic herbal fingerprints: a review. *Anal. Chim. Acta* **2013,** *804*, 16–28.

Hajjou, M.; Krech, L.; Lane-Barlow, C.; Roth, L.; Pribluda, V. S.; Phanouvong, S.; El-Hadri, L.; Evans, L.; III, Raymond, C.; Yuan, E.; Siv, L.; Vuong, T. A.; Boateng, K. P.; Okafor, R.; Chibwe, K. M.; Lukulay, P. H. Monitoring the quality of medicines: results from Africa, Asia, and South America. *Am. J. Trop. Med. Hyg.* **2015,** *92*, 68–74.

Höllein, L.; Kaale, E.; Mwalwisi, Y. M.; Schulze, M. H.; Holzgrabe, U. Routine quality control of medicines in developing countries: analytical challenges, regulatory infrastructures and the prevalence of counterfeit medicines in Tanzania. *Trends Analyt. Chem.* **2016,** *76*, 60–70.

Holzgrabe, U.; Deubner, R.; Schollmayer, C.; Waibel, B. Quantitative NMR spectroscopy – applications in drug analysis. *J. Pharm. Biomed. Anal.* **2005,** *38*, 806–812.

Holzgrabe, U.; Malet-Martino, M. Analytical challenges in drug counterfeiting and falsification – The NMR approach. *J. Pharm. Biomed. Anal.* **2011,** *55*, 679–687.

International Medical Products Anti-Counterfeiting Taskforce (IMPACT). Counterfeits drugs kill. http://www.gphf.org/images/downloads/impactbrochure.pdf (2008) (last accessed 30.03.2016).

Interpol. Pangea operation. http://www.interpol.int/Crime-areas/Pharmaceutical-crime/Operations/Operation-Pangea (2015) (last accessed 30.03.2016).

Islam, M. N.; Kim, U.; Kim, D. H.; Dong, M. S.; Yoo, H. H. High-performance liquid chromatography-based multivariate analysis to predict the estrogenic activity of an Epimedium koreanum extract. *Biosci. Biotechnol. Biochem.* **2012,** *76*, 923–927.

Jackson, G.; Patel, S.; Khan, S. Assessing the problem of counterfeit medications in the United Kingdom. *Int. J. Clin. Pract.* **2012,** *66*, 241–250.

Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemometr. Intell. Lab. Syst.* **2005,** *76*, 101–110.

Jiang, Y.; David, B.; Tu, P.; Barbin, Y. Recent analytical approaches in quality control of traditional Chinese medicines – a review. *Anal. Chim. Acta* **2010,** *657*, 9–18.

Jung, C. R.; Ortiz, R. S.; Limberger, R.; Mayorga, P. A new methodology for detection of counterfeit Viagra® and Cialis® tablets by image processing and statistical analysis. *Forensic Sci. Int.* **2012,** *216*, 92–96.

Justa Neves, D. B.; Caldas, E. D. Dietary supplements: international legal framework and adulteration profiles, and characteristics of products on the Brazilian clandestine market. *Regul. Toxicol. Pharmacol.* **2015,** *73*, 93–104.

Kinghorn, A. D. Pharmacognosy in the 21st century. *J. Pharm. Pharmacol.* **2001,** *2*, 135–148.

Kwok, K.; Taylor, L. S. Analysis of counterfeit Cialis® tablets using Raman microscopy and multivariate curve resolution. *J. Pharm. Biomed. Anal.* **2012a,** *66*, 126–135.

Kwok, K.; Taylor, L. S. Analysis of the packaging enclosing a counterfeit pharmaceutical tablet using Raman microscopy and two-dimensional correlation spectroscopy. *Vib. Spectrosc.* **2012b,** *61*, 176–182.

Li, W.; Cheng, Z.; Wang, Y.; Qu, H. Quality control of *Lonicerae japonicae flos* using near infrared spectroscopy and chemometrics. *J. Pharm. Biomed. Anal.* **2013,** *72*, 33–39.

Li, L.; Zang, H.; Li, J.; Chen, D.; Li, T.; Wang, F. Identification of anisodamine tablets by Raman and near-infrared spectroscopy with chemometrics. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2014,** *127*, 91–97.

Liang, X. M.; Jin, Y.; Wang, Y. P.; Jin, G. W.; Fu, Q.; Xiao, Y. S. Qualitative and quantitative analysis in quality control of traditional Chinese medicines. *J. Chromatogr. A* **2009,** *1216*, 2033–2044.

Liang, Y.; Xie, P.; Chau, F. Chromatographic fingerprinting and related chemometric techniques for quality control of traditional Chinese medicines. *J. Sep. Sci.* **2010,** *33*, 410–421.

Lu, F.; Li, S.; Le, J.; Chen, G.; Cao, Y.; Qi, Y.; Chai, Y.; Wu, Y. A new method for testing synthetic drugs adulterated in herbal medicines based on infrared spectroscopy. *Anal. Chim. Acta* **2007,** *589*, 200–207.

Lu, F.; Weng, X.; Chai, Y.; Yang, Y.; Yu, Y.; Duan, G. A novel identification system for counterfeit drugs based on portable Raman spectroscopy. *Chemometr. Intell. Lab. Syst.* **2013,** *127*, 63–69.

Lucio-Gutierrez, J. R.; Garza-Juarez, A.; Coello, J.; Maspoch, S.; Salazar-Cavazos, M. L.; Salazar-Aranda, R.; Waksman de, T. N. Multi-wavelength high-performance liquid chromatographic fingerprints and chemometrics to predict the antioxidant activity of *Turnera diffusa* as part of its quality control. *J. Chromatogr. A* **2012,** *1235*, 68–76.

Luypaert, J.; Massart, D. L.; Vander Heyden, Y. Near-infrared spectroscopy applications in pharmaceutical analysis. *Talanta* **2007,** *72*, 865–883.

Martino, R.; Malet-Martino, M.; Gilard, V.; Balayssac, S. Counterfeit drugs: analytical techniques for their identification. *Anal. Bioanal. Chem.* **2010,** *398*, 77–92.

Masoum, S.; Mehran, M.; Ghaheri, S. Identification of potential antioxidant compounds in the essential oil of thyme by gas chromatography with mass spectrometry and multivariate calibration techniques. *J. Sep. Sci.* **2015,** *38*, 410–417.

Massart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J. Handbook of Chemometrics and Qualimetrics – Part A. Elsevier: Amsterdam, 1997.

Neuberger, S.; Neusüß, C. Determination of counterfeit medicines by Raman spectroscopy: systematic study based on a large set of model tablets. *J. Pharm. Biomed. Anal.* **2015,** *112*, 70–78.

Nisgoski, S.; Schardosin, F. Z.; Batista, F. R. R.; Bolzon de Muniz, G. I.; Carneiro, M. E. Potential use of NIR spectroscopy to identify Cryptomeria japonica varieties from southern Brazil. *Wood Sci. Technol.* **2016,** *50*, 71–80.

Ortiz, R. S.; Mariotti, K. C.; Schwab, N. V.; Sabin, G. P.; Rocha, W. F.; de Castro, E. V.; Limberger, R. P.; Mayorga, P.; Bueno, M. I.; Romao, W. Fingerprinting of sildenafil citrate and tadalafil tablets in pharmaceutical formulations via X-ray fluorescence (XRF) spectrometry. *J. Pharm. Biomed. Anal.* **2012,** *58*, 7–11.

Ortiz, R. S.; Mariotti, K. C.; Fank, B.; Limberger, R. P.; Anzanello, M. J.; Mayorga, P. Counterfeit Cialis and Viagra fingerprinting by ATR-FTIR spectroscopy with chemometry: can the same pharmaceutical powder mixture be used to falsify two medicines? *Forensic Sci. Int.* **2013a,** *226*, 282–289.

Ortiz, R. S.; Mariotti, K. C.; Holzschuh, M. H.; Romao, W.; Limberger, R. P.; Mayorga, P. Profiling counterfeit Cialis, Viagra and analogs by UPLC-MS. *Forensic Sci. Int.* **2013b,** *229*, 13–20.

Pan, R.; Guo, F.; Lu, H.; Feng, W. W.; Liang, Y. Z. Development of the chromatographic fingerprint of Scutellaria barbata D. Don by GC-MS combined with chemometrics methods. *J. Pharm. Biomed. Anal.* **2011,** *55*, 391–396.

Petrakis, P. V.; Agiomyrgianaki, A.; Christophoridou, S.; Spyros, A.; Dais, P. Geographical characterization of greek virgin olive oils (cv. Koroneiki) using 1H and 31P NMR fingerprinting with canonical discriminant analysis and classification binary trees. *J. Agric. Food Chem.* **2008,** *56*, 3200–3207.

Pharmaceutical Security Institute. http://www.psi-inc.org (2015) (last accessed 30.03.2016).

Poon, W. T.; Lam, Y. H.; Lai, C. K.; Chan, A. Y.; Mak, T. W. Analogues of erectile dysfunction drugs: an under-recognised threat. *Hong Kong Med. J.* **2007,** *13*, 359–363.

Practical Guide to Chemometrics; 2nd Edition. Paul Gemperline., Ed. Taylor & Francis Group: Boca Raton, 2006.

Pravdova, V.; Walczak, B.; Massart, D. L. A comparison of two algorithms for warping of analytical signals. *Anal. Chim. Acta* **2002,** *456*, 77–92.

Puchert, T.; Lochmann, D.; Menezes, J. C.; Reich, G. Near-infrared chemical imaging (NIR-CI) for counterfeit drug identification – a four-stage concept with a novel approach of data processing (Linear Image Signature). *J. Pharm. Biomed. Anal.* **2010,** *51*, 138–145.

Questier, F.; Put, R.; Coomans, D.; Walczak, B.; Vander Heyden, Y. The use of CART and multivariate regression trees for supervised and unsupervised feature selection. *Chemometr. Intell. Lab. Syst.* **2005,** *76*, 45–54.

Rebiere, H.; Guinot, P.; Civade, C.; Bonnet, P. A.; Nicolas, A. Detection of hazardous weight-loss substances in adulterated slimming formulations using ultra-high-pressure liquid chromatography with diode-array detection. *Food Addit. Contam. Part A. Chem. Anal. Control Expo. Risk Assess.* **2012,** *29*, 161–171.

Reich, G. Near-infrared spectroscopy and imaging: basic principles and pharmaceutical applications. *Adv. Drug Deliv. Rev.* **2005,** *57*, 1109–1143.

Rodionova, O.; Pomerantsev, A.; Houmoller, L.; Shpak, A.; Shpigun, O. Noninvasive detection of counterfeited ampoules of dexamethasone using NIR with confirmation by HPLC-DAD-MS and CE-UV methods. *Anal. Bioanal. Chem.* **2010,** *397*, 1927–1935.

Rodomonte, A. L.; Gaudiano, M. C.; Antoniella, E.; Lucente, D.; Crusco, V.; Bartolomei, M.; Bertocchi, P.; Manna, L.; Valvo, L.; Alhaique, F.; Muleri, N. Counterfeit drugs detection by measurement of tablets and secondary packaging colour. *J. Pharm. Biomed. Anal.* **2010,** *53*, 215–220.

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*, 683–700.

Ronowicz, J.; Kupcewicz, B.; Budzisz, E. Chemometric analysis of antioxidant properties of herbal products containing Ginkgo biloba extract. *Cent. Eur. J. Biol.* **2013,** *8*, 374–385.

Sabin, G. P.; Lozano, V. A.; Rocha, W. F.; Romao, W.; Ortiz, R. S.; Poppi, R. J. Characterization of sildenafil citrate tablets of dif-

ferent sources by near infrared chemical imaging and chemometric tools. *J. Pharm. Biomed. Anal.* **2013,** *85*, 207–212.

Sacre, P. Y.; Deconinck, E.; De Beer T.; Courselle, P.; Vancauwenberghe, R.; Chiap, P.; Crommen, J.; De Beer, J. O. Comparison and combination of spectroscopic techniques for the detection of counterfeit medicines. *J. Pharm. Biomed. Anal.* **2010,** *53*, 445–453.

Sacre, P. Y.; Deconinck, E.; Daszykowski, M.; Courselle, P.; Vancauwenberghe, R.; Chiap, P.; Crommen, J.; De Beer, J. O. Impurity fingerprints for the identification of counterfeit medicines – a feasibility study. *Anal. Chim. Acta* **2011a,** *701*, 224–231.

Sacre, P. Y.; Deconinck, E.; Saerens, L.; De, B. T.; Courselle, P.; Vancauwenberghe, R.; Chiap, P.; Crommen, J.; De Beer, J. O. Detection of counterfeit Viagra® by Raman microspectroscopy imaging and multivariate analysis. *J. Pharm. Biomed. Anal.* **2011b,** *56*, 454–461.

Scafi, S. H.; Pasquini, C. Identification of counterfeit drugs using near-infrared spectroscopy. *Analyst* **2001,** *126*, 2218–2224.

Schneider, A.; Wessjohann, L. A. Comparison of impurity profiles of Orlistat pharmaceutical products using HPLC tandem mass spectrometry. *J. Pharm. Biomed. Anal.* **2010,** *53*, 767–772.

Serneels, S.; Croux, C.; Filzmoser, P.; Van Espen, P. J. Partial robust M-regression. *Chemometr. Intell. Lab. Syst.* **2005,** *79*, 55–64.

Silvestre, V.; Mboula, V. M.; Jouitteau, C.; Akoka, S.; Robins, R. J.; Remaud, G. S. Isotopic 13C NMR spectrometry to assess counterfeiting of active pharmaceutical ingredients: site-specific 13C content of aspirin and paracetamol. *J. Pharm. Biomed. Anal.* **2009,** *50*, 336–341.

Smolinski, A.; Walczak, B.; Einax, J. W. Hierarchical clustering extended with visual complements of environmental data set. *Chemometr. Intell. Lab. Syst.* **2002,** *64*, 45–54.

Smyth, C.; Coomans, D.; Everingham, Y.; Hancock, T. Auto-associative multivariate regression trees for cluster analysis. *Chemometr. Intell. Lab. Syst.* **2006,** *80*, 120–129.

Song, X. Y.; Li, Y. D.; Shi, Y. P.; Jin, L.; Chen, J. Quality control of traditional Chinese medicines: a review. *Chin J. Nat. Med.* **2013,** *11*, 596–607.

Stanimirova, I.; Daszykowski, M.; Van Gyseghem, E.; Bensaid, F. F.; Lees, M.; Smeyers-Verbeke, J.; Massart, D. L.; Vander Heyden, Y. Chemometrical exploration of an isotopic ratio data set of acetylsalicylic acid. *Anal. Chim. Acta* **2005,** *552*, 1–12.

Storme-Paris, I.; Rebiere, H.; Matoga, M.; Civade, C.; Bonnet, P. A.; Tissier, M. H.; Chaminade, P. Challenging near infrared spectroscopy discriminating ability for counterfeit pharmaceuticals detection. *Anal. Chim. Acta* **2010,** *658*, 163–174.

Tauler, R. Multivariate curve resolution applied to second order data. *Chemometr. Intell. Lab. Syst.* **1995,** *30*, 133–146.

Tistaert, C.; Dejaegher, B.; Nguyen, H. N.; Chataigne, G.; Riviere, C.; Nguyen, T. H.; Chau, V. M.; Quetin-Leclercq, J.; Vander Heyden, Y. Potential antioxidant compounds in Mallotus species fingerprints. Part I: indication, using linear multivariate calibration techniques. *Anal. Chim. Acta* **2009,** *652*, 189–197.

Tistaert, C.; Dejaegher, B.; Chataigne, G.; Van, M. C.; Quetin-Leclercq, J.; Vander Heyden, Y. Dissimilar chromatographic systems to indicate and identify antioxidants from Mallotus species. *Talanta* **2011a,** *83*, 1198–1208.

Tistaert, C.; Dejaegher, B.; Vander Heyden, Y. Chromatographic separation techniques and data handling methods for herbal fingerprints: a review. *Anal. Chim. Acta* **2011b,** *690*, 148–161.

Tistaert, C.; Chataigne, G.; Dejaegher, B.; Riviere, C.; Nguyen, H. N.; Chau, V. M.; Quetin-Leclercq, J.; Vander Heyden, Y. Multivariate data analysis to evaluate the fingerprint peaks responsible for the cytotoxic activity of Mallotus species. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2012a,** *910*, 103–113.

Tistaert, C.; Dejaegher, B.; Chataigne, G.; Riviere, C.; Hoai, N. N.; Van, M. C.; Quetin-Leclercq, J.; Vander Heyden, Y. Potential antioxidant compounds in Mallotus species fingerprints. Part II: fingerprint alignment, data analysis and peak identification. *Anal. Chim. Acta* **2012b,** *721*, 35–43.

Tomasi, G.; van den Berg, F.; Andersson, C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *J. Chemometrics* **2004,** *18*, 231–241.

Trygg, J.; Wold, S. Orthogonal projections to latent structures (O-PLS). *J. Chemom.* **2002,** *16*, 119–128.

United States Pharmacopoeial Convention. Spectrophotometry and Light-Scattering; 38th Edition. United States Pharmacopoeia: Rockville, 2015.

Van Gyseghem, E.; Dejaegher, B.; Put, R.; Forlay-Frick, P.; Elkihel, A.; Daszykowski, M.; Heberger, K.; Massart, D. L.; Vander Heyden, Y. Evaluation of chemometric techniques to select orthogonal chromatographic systems. *J. Pharm. Biomed. Anal.* **2006,** *41*, 141–151.

van Nederkassel, A. M.; Daszykowski, M.; Massart, D. L.; Vander Heyden Y. Prediction of total green tea antioxidant capacity from chromatograms by multivariate modeling. *J. Chromatogr. A* **2005,** *1096*, 177–186.

van Nederkassel, A. M.; Daszykowski, M.; Eilers, P. H.; Vander Heyden, Y. A comparison of three algorithms for chromatograms alignment. *J. Chromatogr. A* **2006a,** *1118*, 199–210.

van Nederkassel, A. M.; Xu, C. J.; Lancelin, P.; Sarraf, M.; Mackenzie, D. A.; Walton, N. J.; Bensaid, F.; Lees, M.; Martin, G. J.; Desmurs, J. R.; Massart, D. L.; Smeyers-Verbeke, J.; Vander Heyden, Y. Chemometric treatment of vanillin fingerprint chromatograms. Effect of different signal alignments on principal component analysis plots. *J. Chromatogr. A* **2006b,** *1120*, 291–298.

Vandeginste, B. G. M.; Massart, D. L.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J. Handbook of Chemometrics and Qualimetrics – Part B; Elsevier Science: Amsterdam, 1998.

Vankeirsbilck, T.; Vercauteren, A.; Baeyens, W.; Van der Weken, G. Applications of Raman spectroscopy in pharmaceutical analysis. *Trends Anal. Chem.* **2002,** *21*, 869–877.

Venhuis, B. J.; de Kaste, D. Towards a decade of detecting new analogues of sildenafil, tadalafil and vardenafil in food supplements: a history, analytical aspects and health risks. *J. Pharm. Biomed. Anal.* **2012,** *69*, 196–208.

Vest Nielsen, N. P.; Carstensen, J. M.; Smedsgaard, J. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *J. Chromatogr. A* **1998,** *805*, 17–35.

Viaene, J.; Goodarzi, M.; Dejaegher, B.; Tistaert, C.; Hoang Le, T. A.; Nguyen, H. N.; Chau, V. M.; Quetin-Leclercq, J.; Vander, H. Y. Discrimination and classification techniques applied on Mallotus and Phyllanthus high performance liquid chromatography fingerprints. *Anal. Chim. Acta* **2015,** *877*, 41–50.

Vredenbregt, M. J.; Blok-Tip, L.; Hoogerbrugge, R.; Barends, D. M.; de Kaste, D. Screening suspected counterfeit Viagra and

imitations of Viagra with near-infrared spectroscopy. *J. Pharm. Biomed. Anal.* **2006,** *40*, 840–849.

Walczak, B.; Wu, W. Fuzzy warping of chromatograms. *Chemometr. Intell. Lab. Syst.* **2005,** *77*, 173–180.

Wehrens, R. Chemometrics with R; Springer: London, 2011.

Wertheimer, A. I.; Norris, J. Safeguarding against substandard/counterfeit drugs: mitigating a macroeconomic pandemic. *Res. Social. Adm. Pharm.* **2009,** *5*, 4–16.

Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometr. Intell. Lab. Syst.* **2001,** *58*, 109–130.

World Courier. Securing the Global Pharmaceutical Supply Chain against the Threat of Counterfeit Drugs. http://www.worldcourier.com/insights-events/market-insights/detail/securing-global-pharmaceutical-supply-chain (2014) (last accessed 30.03.2016).

World Health Organization (WHO). Combating Counterfeit Drugs: A Concept Paper for Effective International Collaboration. http://www.who.int/medicines/services/counterfeit/CombatingCounterfeitDrugs_Conceptpaper.pdf (2005) (last accessed 30.03.2016).

World Health Organization (WHO). IMPACT – Who We Are. http://www.who.int/mediacentre/news/releases/2006/pr69/en/ (2006) (last accessed 30.03.2016).

World Health Organization (WHO). Growing Threat from Counterfeit Medicines. http://www.who.int/bulletin/volumes/88/4/10-020410/en/ (2010) (last accessed 30.03.2016).

World Health Organization (WHO). Definitions of SSFFC Medical Products. http://www.who.int/medicines/regulation/ssffc/definitions/en/ (2016a) (last accessed 30.03.2016).

World Health Organization (WHO). Substandard, Spurious, Falsely Labelled, Falsified, Counterfeit (SSFFC) Medical Products. http://www.who.int/medicines/regulation/ssffc/en/ (2016b) (last accessed 30.03.2016).

Wu, W.; Mallet, Y.; Walczak, B.; Penninckx, W.; Massart, D. L.; Heuerding, S.; Erni, F. Comparison of regularized discriminant analysis, linear discriminant analysis and quadratic discriminant analysis applied to NIR data. *Anal. Chim. Acta* **1996,** *329*, 257–265.

Yang, D. Z.; Yin, X. X.; Ong, C. N.; Tang, D. Q. Multidimensional information-based HPLC technologies to evaluate traditional Chinese medicine. *J. Chromatogr. Sci.* **2013,** *51*, 716–725.

Yi, L. Z.; Yuan, D. L.; Liang, Y. Z.; Xie, P. S.; Zhao, Y. Quality control and discrimination of pericarpium citri reticulatae and pericarpium citri reticulatae viride based on high-performance liquid chromatographic fingerprints and multivariate statistical analysis. *Anal. Chim. Acta* **2007,** *588*, 207–215.

Zhang, D.; Duan, X.; Deng, S.; Nie, L.; Zang, H. Fingerprint analysis, multi-component quantitation, and antioxidant activity for the quality evaluation of Salvia miltiorrhiza var. alba by high-performance liquid chromatography and chemometrics. *J. Sep Sci* **2015,** *38*, 3337–3344.

Zhou, X.; Tang, L.; Wu, H.; Zhou, G.; Wang, T.; Kou, Z.; Li, S.; Wang, Z. Chemometric analyses for the characterization of raw and processed seeds of *Descurainia sophia* (L.) based on HPLC fingerprints. *J. Pharm. Biomed. Anal.* **2015,** *111*, 1–6.

Zhu, X.; Zhang, Z.; Lu, F.; Wu, Y.; Qi, Y. Modified local straight-line screening to detect synthetic drugs in adulterated herbal medicines. *Appl. Spectrosc.* **2009,** *63*, 471–476.