Review Article

Eric Deconinck*, Celine Duchateau, Margot Balcaen, Lies Gremeaux, and Patricia Courselle

Chemometrics and infrared spectroscopy – A winning team for the analysis of illicit drug products

Abstract: Spectroscopic techniques such as infrared spectroscopy and Raman spectroscopy are used for a long time in the context of the analysis of illicit drugs, and their use is increasing due to the development of more performant portable devices and easy application in the context of harm reduction through drug checking or onsite forensic analysis. Although these instruments are routinely used with a spectral library, the importance of chemometric techniques to extract relevant information and give a full characterisation of samples, especially in the context of adulteration, is increasing. This review gives an overview of the applications described in the context of the analysis of illicit drug products exploiting the advantages of the combination of spectroscopy with chemometrics. Next to an overview of the literature, the review also tries to emphasize the shortcomings of the presented research papers and to give an incentive to what is needed to include chemometrics as a part of the

daily routine of drug checking services and mobile forensic applications.

1 Introduction

The new EU drugs strategy and action plan (2021–2025) confirms the need for balanced and evidence-based responses to the drug phenomenon to address the continued and escalating challenges to both public health and security arising from the availability and use of a wide range of substances [1,2]. Different initiatives were already taken under different forms, e.g., legislative initiatives [3,4] and initiatives focusing on harm reduction, a principle that is today deeply implemented in the EU Drug Strategy [2]. These initiatives are aimed at ensuring a more pragmatic approach to dealing with drug use, through different interventions with a purpose to reduce the health-related risks [5,6]. Another initiative in this context is drug checking. Drug checking is defined as a harm reduction-oriented service that allows users of psychoactive substances to submit their samples for chemical analysis and receive timely feedback regarding their composition and the associated risks of using them [7–11]. The major advantages here are that, at the level of the user, short- and long-term adverse effects might be avoided, while at the level of society, more insight into the drug market is provided, e.g., the circulation of lethal doses or new substances can be rapidly detected. This is then followed by communication to care providers and relevant authorities, such as policymaking, leading to more informed decision-making and generating a better picture of the drug market [12,13]. The first drug checking service was established in the Netherlands in 1992, but already in 2017, a global review identified 31 drug checking services spread over 20 different countries [14]. All these services come with their own impact and limitations [15]. One of these limitations is determined by the analytical

---

* Corresponding author: Eric Deconinck, Scientific Direction Chemical and Physical Health Risks, Service of Medicines and Health Products, Sciensano, J. Wytsmanstraat 14, B-1050, Brussels, Belgium; RD3 Unit of Pharmacognosy, Bioanalysis and Drug Discovery, Faculty of Pharmacy, Campus Plaine, Université Libre de Bruxelles, 1050 Brussels, Belgium, e-mail: Eric.Deconinck@sciensano.be, tel: +322-642-51-70, fax: +322-642-53-27
Celine Duchateau: Scientific Direction Chemical and Physical Health Risks, Service of Medicines and Health Products, Sciensano, J. Wytsmanstraat 14, B-1050, Brussels, Belgium; RD3 Unit of Pharmacognosy, Bioanalysis and Drug Discovery, Faculty of Pharmacy, Campus Plaine, Université Libre de Bruxelles, 1050 Brussels, Belgium
Margot Balcaen, Lies Gremeaux: Scientific Direction Epidemiology and Public Health, Section Lifestyle and Chronic Diseases, Sciensano, J. Wytsmanstraat 14, B-1050, Brussels, Belgium
Patricia Courselle: Scientific Direction Chemical and Physical Health Risks, Service of Medicines and Health Products, Sciensano, J. Wytsmanstraat 14, B-1050, Brussels, Belgium

capacities and instruments, the drug checking services dispose of, to analyse the samples brought to them by users [16].

In general, it can be said that effective testing relies on low-cost and rapid screening methods. For this, most drug checking services make use of colour testing and spectroscopic methods like ultraviolet-visible spectroscopy, Fourier transformed-infrared (FT-IR), or Raman spectroscopy [5,7,17,18]. In the context of drug checking, colour reagents test kits are most often used as an initial step. These tests are purely presumptive in nature, but still are fairly accurate in identifying a compound or mixture of compounds, especially when a standard operating procedure comprising a series of tests is used [5]. Although all the techniques mentioned earlier have their limitations, most of them are not able to identify multiple drugs in the sample, determine purity of the products, and identify new substances that are not included in the colour reagent kits or the spectral libraries/databases of the spectroscopic instruments. Some of the spectroscopic instruments come with software allowing deconvolution of the spectra to identify multiple compounds in one sample; however, this is limited to the main components. Low-level substances will not be detected due to the interference of the spectrum of the more abundant compounds in the samples. The more it has to be emphasised that not all samples are suited to be analysed through these spectroscopic techniques, e.g. infrared inactive substances or salts or coloured samples causing fluorescence in Raman spectroscopy [16,17].

Next to the context of drug checking, analysis of illicit drugs is of course also performed by law enforcement. Here, the samples are often analysed in specialised forensic laboratories where chromatographic methods hyphenated with mass spectrometry (MS) are the golden standard. Needless to say this traditional lab testing comes with high costs due to expensive equipment and the necessity of highly trained personnel. This way of work is also more time consuming and therefore does not correspond to the idea of onsite drug testing, nor in the context of drug checking or in the context of law enforcement [5,17]. Indeed, even if law enforcement has the analytical capacities of forensic laboratories at their disposal, they are as well in need of more rapid and onsite analysis to identify seizures or to stop products at the borders. Many spectroscopic techniques are non-destructive preserving evidence for future reference. Therefore, law enforcement is turning toward the aforementioned techniques, especially FT-IR and Raman, most often combined with extensive databases or spectral libraries. An evolution that is also pushed forward due to the development of portable devices, with increasing spectral resolution.

However in the context of both drug checking, as forensic analysis, spectroscopic techniques like FT-IR and Raman could be better exploited by combining them with chemometrics. Chemometrics is defined as "a chemical discipline that uses mathematics, statistics and formal logic (a) to design or select optimal experimental procedures; (b) to provide maximum relevant chemical information by analysing chemical data; and (c) to obtain knowledge about a chemical system" [19]. Although the combination of spectroscopy and chemometrics was already applied in a lot of different domains [20], as in the analysis of illicit drugs, the applications stay often on the level of scientific publications and do not seem to find their way to the routine testing in both drug checking and in forensic analysis of illicit drug samples. It is the conviction of the authors that a more extensive implementation of chemometrics in the software of the spectroscopic instruments, combined with guidance on how to create chemometric models based on reference standards, reference samples, and real-life samples, could benefit on-site testing in both harm reduction settings and in the context of law enforcement.

This review aims to give an overview of the possibilities of combining spectroscopy with chemometrics, more precisely modelling, in the context of the (on-site) analysis of illicit drug samples, as described in the literature. First, a short theoretical overview is given of the techniques applied in the articles related to this review, followed by a critical overview of the different applications and possibilities. Ultimately this review hopes to contribute to a more practical application of chemometrics in the routine on-site analysis of samples in the context of drug checking as law enforcement.

2 Theory

2.1 (Near) Infrared spectroscopy

Infrared spectroscopy (IR) is based on the direction of infrared light on a molecule, which absorbs this energy.

When using Mid-IR light (wavenumber range 4,000–400 cm^{-1}), the absorbance of this energy causes molecular bonds to stretch (stretching vibration) and bend and brings the molecule into a so-called excited state. Afterwards, the molecule returns to the ground state by emitting a photon, which is detected by the infrared spectrophotometers and transformed into an electric signal. The strength and the atoms of the bond determine how much energy is needed to reach the excited state and allows the identification of

functional groups within the 4,000–1,800 cm^{-1} region. The fingerprint region (1,800–400 cm^{-1}) [21] is most valuable to identify compounds or mixtures of compounds and can be used for identification by comparing the spectrum with mid-IR spectral libraries.

Where before mid-IR spectroscopy was employed by transmittance using potassium bromide tablets, modern instruments are equipped with an attenuated total reflectance (ATR) sampler. In this case, the spectrum is measured through reflectance, needing only a small amount of powdered sample, without the necessity of further sample pre-treatment [22].

The near-infrared region is situated in the wavenumber range from 12,500 to 4,000 cm^{-1}. Spectroscopy in this region relies on the presence of polarised chemical bonds and the modification of the dipole moment by the absorption of a photon from the light source, causing molecular bond vibrations [23].

NIR spectra have fingerprint properties and are very useful for qualitative analyses. An advantage is the property of NIR to penetrate further into a sample than Mid-IR spectroscopy. Conversely, the spectra are rather complex, and therefore, multivariate calibration techniques are often used to extract information.

In the case of both types of spectroscopies, Fourier transformation is used as it allows faster acquisition, simultaneous measurement of all wavelengths with a more accurate wavelength scale, better spectral resolution, and better sensitivity compared to dispersive systems.

2.2 Raman spectroscopy

Raman spectroscopy is based on the inelastic scattering of light, resulting in an energy change between a photon from a monochromatic light source (often a laser) and the sample under investigation. When such a beam of monochromatic light is directed at a sample molecule, a part of the light will be absorbed and transmitted by the molecule, while another part will be scattered [23–25].

Raman scattering is a weak phenomenon and has some limitations, like its poor sensitivity. A problem that can be partly solved by the use of surface-enhanced Raman spectroscopy (SERS). Indeed, SERS can increase the limit of detection by dramatically enhancing the Raman signal. This is achieved by absorbing the analytes of interest on a metallic surface such as metal electrodes or colloidal metal nanoparticles. Often, silver or gold particles are used for this purpose. This process can cause an electromagnetic enhancement factor of up to 10^{11} and so a

Raman signal enhancement of several orders of magnitude [26].

2.3 Chemometric methods applied for illicit drug products analysis

This section presents a short theoretical overview of the chemometric procedures and algorithms that were encountered during the literature review described beneath. It is not the intention to be exhaustive, but to limit to the techniques most often applied. When a technique was only encountered once, a reference with the theoretical principles was added in the application section.

2.3.1 Data pre-processing

In the context of spectroscopic techniques such as infrared and Raman spectroscopy, a whole range of data pre-processing techniques are available. In general, it can be said that all these techniques have the same purpose, i.e., reducing variability due to secondary effects that cannot be related to the sample itself, diminish noise in the data, and enhance the most important signals in the spectrum. In general, pre-processing of data will enhance the quality of chemometric models and increase their meaning and robustness.

A first very important step in the process of dealing with spectral data is baseline correction, a step that is often automatically performed by the software of the vendors of the instruments. Other popular pre-treatment techniques, often applied for spectral data, are standard normal variate (SNV), derivatives, normalisation, scaling, and smoothing. SNV is a multivariate pre-processing technique allowing to correct for baseline shifts in the spectra. These baseline shifts are often a consequence of interference, e.g. fluorescence in Raman spectroscopy [19]. Derivative pre-processing is among the most applied techniques in spectroscopy. The first derivative of a spectrum removes constant off sets from the spectrum, while the second derivative eliminates these off sets and linear drifts. Although often applied, the disadvantage is that the original structure/appearance of the data is changed, rendering interpretation of the spectra more complicated [19,27]. Derivative pre-processing is often combined with the use of smoothing algorithms. One algorithm, particularly popular in infrared and Raman spectroscopy, is Savitsky–Gollay filtering. The idea here is to reduce or even eliminate noise in the data by fitting the data points using a low-degree polynomial using

linear least squares [19,20]. Other types of pre-processing are scaling techniques, e.g., autoscaling. Here, each column in a data matrix is centred followed by a division by the column standard deviation. This procedure is particularly used when different variables have different ranges for their data. A related scaling technique is normalisation. Here, the range of intensities at each wavelength is brought back on a scale from 0 to 1. Another popular scaling technique is multiplicative scatter correction (MSC). In MSC, the scale and the offset of the spectra are corrected to fit as closely as possible a reference spectrum [20]. The disadvantage of scaling techniques is that the models build with scaled data have to be recalculated each time when a new sample/spectrum is added to the data matrix [19].

2.3.2 Hierarchical clustering (HCA)

HCA is an unsupervised clustering technique based on the (dis)similarities in the recorded data, e.g., spectra. Several similarity measures can be used. The most classical approaches are based on spatial distances like the Euclidean or the Mahalonobis distance or based on correlation coefficients. Also, some more advanced measures exist like the popular Wards algorithm. Clustering can be performed in an agglomerative or a divisive way, and the results are often presented as a dendrogram in which the length of the branches represents the (dis)similarity between the objects for the recorded data (Figure 1) [19,28].
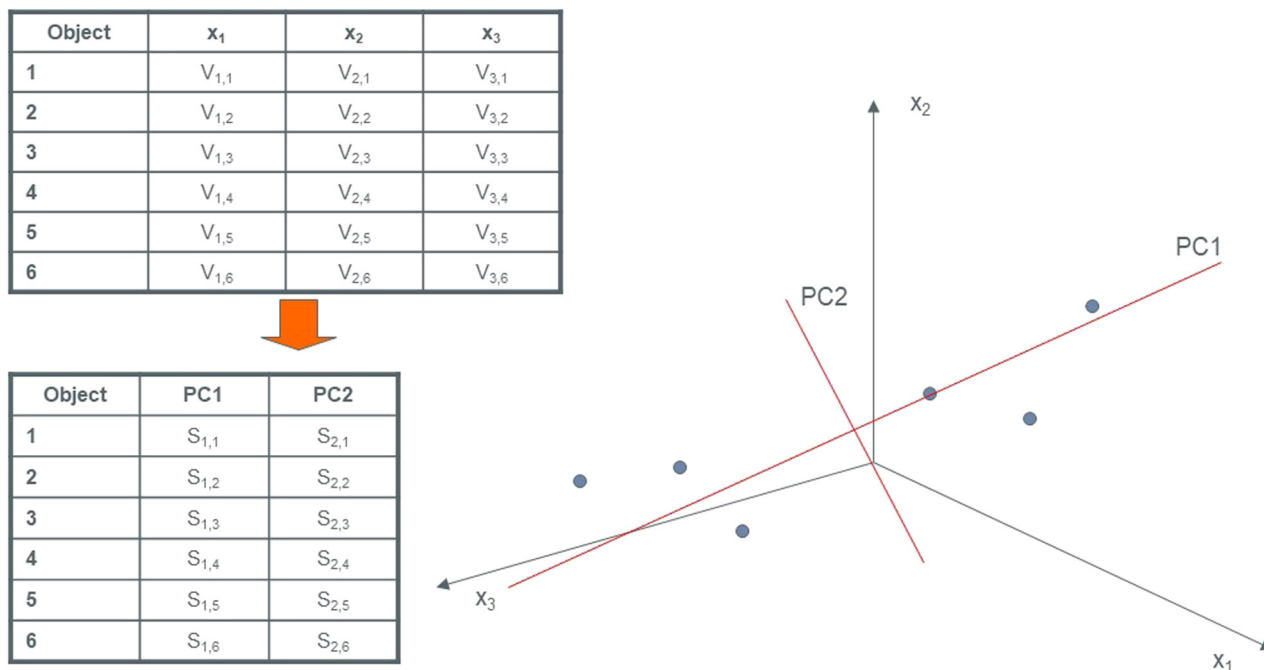
2.3.3 Principal component analysis (PCA)

PCA is an unsupervised projection technique and allows feature reduction of multi-dimensional data sets to obtain a two- or three-dimensional representation of the data space. To do this, PCA calculates latent variables, which are defined as linear combinations of the manifest variables, optimising the variance in the data. Following this definition, the first principal component (PC1) will be defined in the direction of the highest variation in the data, PC2 in the direction of the highest variation around PC1, PC3 in the direction of the highest variation around the plane PC1–PC2, and so on. Figure 2 gives a simplified representation of the procedure for the reduction from three to two dimensions.

The regression coefficients on the different PCs are called the loadings and are a measure for the importance of a manifest variable on a certain PC. The projection of an object (sample) on a PC is called the score of the object. In this way, PCA allows a visual interpretable image of the data and allows to distinguish clusters of samples, as well as defining those variables most important for the distinction [19]. Contrary to what is presented often in the literature, PCA is not a classification technique, but a data exploration technique. However, PCA can be used as a part of supervised techniques like soft independent modelling of class analogy (SIMCA) (see further) for classification or principal component regression (PCR) for continuous response variables. The latter is
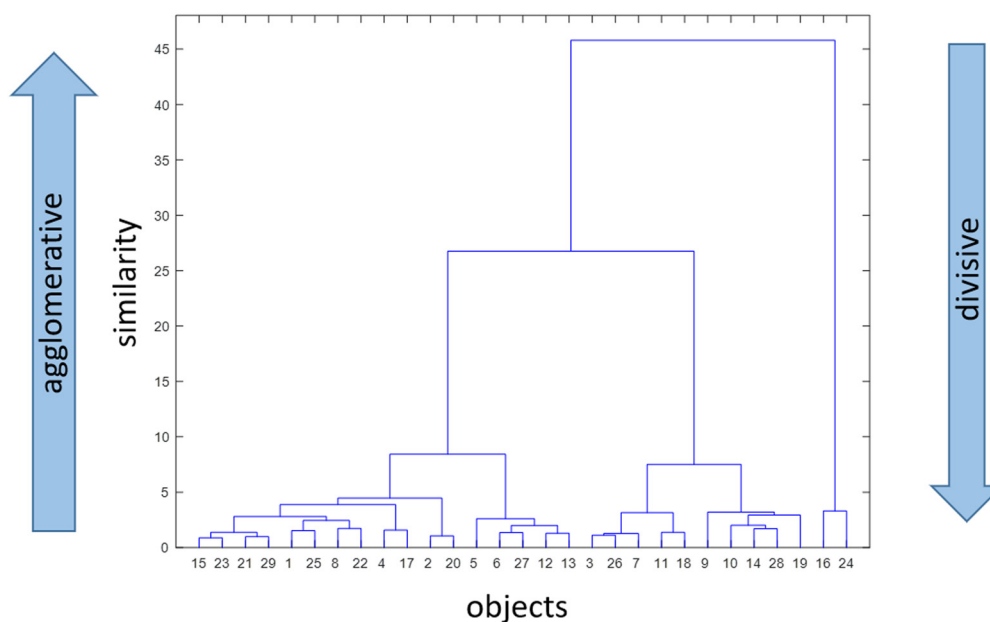
Figure 1: Example of a dendrogram obtained through HCA. The length of the branches indicates the similarity between the object (how smaller the branches the more similar the sample).

Object	x_1	x_2	x_3
1	$V_{1,1}$	$V_{2,1}$	$V_{3,1}$
2	$V_{1,2}$	$V_{2,2}$	$V_{3,2}$
3	$V_{1,3}$	$V_{2,3}$	$V_{3,3}$
4	$V_{1,4}$	$V_{2,4}$	$V_{3,4}$
5	$V_{1,5}$	$V_{2,5}$	$V_{3,5}$
6	$V_{1,6}$	$V_{2,6}$	$V_{3,6}$

Object	PC1	PC2
1	$S_{1,1}$	$S_{2,1}$
2	$S_{1,2}$	$S_{2,2}$
3	$S_{1,3}$	$S_{2,3}$
4	$S_{1,4}$	$S_{2,4}$
5	$S_{1,5}$	$S_{2,5}$
6	$S_{1,6}$	$S_{2,6}$

Figure 2: The principle of a PCA analysis.

in principle a simple linear regression, using the scores on a selection of PCs (latent variables) as input for the model instead of the manifest variables [19].

2.3.4 K-nearest neighbours (*k*-NN)

One of the simplest chemometric modelling techniques is k-NN. It is a classification technique where classification rules are defined by neighbouring training set objects (spectra of samples). Neighbourhoods can be calculated based on Euclidean distance or correlation between an unknown object and each of the training set objects. The object is then attributed to the class to which the majority of the *k* neighbouring objects belong. k-NN is particularly powerful in binary classification problems, and *k* is the only parameter to be optimised [19].

2.3.5 Linear discriminant analysis (LDA)

LDA is also a feature reduction and projection technique, defining latent variables as linear combinations of the manifest variables. Unlike PCA, LDA is a supervised technique and defines the latent variables in the direction of the maximum discrimination between the known classes. New unknown samples can then be classified according to the region they are projected in, in the data space defined by the latent variables [19].

Bayes discriminant analysis is an adapted version of LDA using the Bayesian approach for the definition of the discrimination functions and so the classification rules [29].

2.3.6 SIMCA

SIMCA is a supervised classification technique emphasising the similarity within the classes rather than focussing on discrimination between the classes. This type of classification is called disjoint class modelling. SIMCA models each class separately using PCA, defining a space around the training samples of the class by the Euclidian distance towards the SIMCA model and the Mahalanobis distance determined in the space of scores following the hotelling T2-distribution. A new sample is attributed to the considered class when its projection is located within the space defined around the training samples of this class [19].

2.3.7 Partial least squares (discriminant analysis)

Partial least squares (PLS) is a supervised projection technique based on the same principles as PCA. In PLS, latent variables are calculated as linear combinations of manifest variables, but here they are defined in such a way that they maximise the co-variance with a response variable.

PLS is one of the most utilised regression techniques and therefore applied to continuous response variables such as dosage or concentration [19]. PLS-discriminant analysis (PLS-DA) is a variation of PLS that allows to work with categorical response variables and therefore using PLS as a classification technique [19].

Orthogonal-PLS or also called orthogonal projection to latent structures is a variation of the PLS algorithm in which the co-variance with the response variable is optimised for the first latent variable, while the following latent variables capture the remaining variance around the first and so are by definition orthogonal (cfr. PCA). The fact that O-PLS models separately the variations in the manifest variables, correlated and uncorrelated with the response, reduces the model complexity and allows a better interpretation of the loadings and so the importance of the manifest variables in describing the response [30,31].

2.3.8 Support vector machines (SVM)

SVM was originally designed for binary classification problems. Therefore, SVM calculates hyperplanes to define decision boundaries to separate data points belonging to different classes. This method is able to treat simple linear as well as more complex non-linear problems by mapping the original data points to a high-dimensional feature space such that the classification problem becomes easier to solve. The mapping is done using a Kernel function [32]. Variations on the algorithm exist allowing to solve multi-class problems as well as regression problems. More details about SVM and Kernel functions can be found in the literature [33,34].

2.3.9 Artificial neural networks (ANN)

ANNs are non-linear modelling techniques that mimic in fact the functioning of the brain, using a series of transfer functions or neurons, where the output of the first function (layer) becomes the input for the second and so on. Different types of ANN exist, and they can be used to solve both classifications as regression problems. For more details, we refer to the literature [19,35,36].

2.3.10 Multivariate curve resolution-alternating least squares (MCR-ALS)

MCR-ALS is a popular resolution method that starts from the assumption that a spectrum of a mixture results from

the weighted sum of the spectra of the pure compounds. The method is especially interesting when there is no a priori knowledge about the chemical system under investigation and is most often used for the chemometric treatment of the data from spectral imaging such as Raman and NIR imaging.

MCR-ALS works in an iterative way and allows to screen samples for their composition since no *a priori* knowledge of the system is necessary. Of course, the algorithm will need a data set of pure spectra to compare with and to screen for. More details about MCR-ALS can be found in the literature [37–39].

2.3.11 Genetic algorithms

One important aspect of chemometric modelling is feature selection. Indeed, in a data set, in our case of spectral data, always a part of the variables is related to noise or experimental error or is just redundant in solving the classification or regression problem at hand.

Genetic algorithms are such a feature selection method that is heuristically based on the evolutionary ideas of natural selection. For implementing a genetic algorithm, a series of strings of variables or candidate solutions are necessary. For each candidate, a model is created and evaluated, after which reproduction is performed using three consecutive steps, i.e., selection, recombination, and mutation. This continues in an iterative way until a pre-defined number of iterations is performed or a feasible set of variables is selected [40,41].

3 Applications for the analysis of illicit drug products

This section is divided into several subsections according to the drugs analysed in the different applications described in the literature and the subject of this review.

3.1 Cocaine

Cocaine is the second most commonly used and seized drug in Europe and is in fact an alkaloid stimulant, extracted from native species of South-America, *Erythroxylon coca* and *Erythroxylon novogranatense* [42]. The popularity of cocaine as a party drug and so the high number of seizures are probably responsible for the fact that most studies in

literature, combining chemometrics and spectroscopy are focussed on cocain samples. The analysis of cocaine samples also represents some interesting challenges like the differentiation between samples containing the cocaine base and the ones containing the salt (often the hydrochloride salt), the range of the purity of the samples, generally higher than 30% w/w but going up to more than 90% w/w, and the different adulterants used to either increase the volume and the profit (e.g., sugar as mannitol or inositol) or to complement the pharmacological effects of cocaine (e.g., lidocaine, phenacetin, caffeine, benzocaine, levamisole, boric acid, and hydroxyzine) [43]. Some of these adulterants, like levamisole, are not innocent and could lead by their self to medical harms [44]. The necessity of fast screening, onsite testing, and the need for non-destructive analysis led to different applications for spectroscopic techniques described in the literature, which are often combined with chemometric data analysis to overcome the challenges described earlier. Table 1 presents the most important features of the applications found in the literature. The readers should be aware that the figures of merit are not always comparable since they were obtained in different ways for different goals, and therefore, they are directed to the respective references for all the details. Pérez-Alfonso et al. [45] applied PLS to NIR spectra of seized cocaine samples for purity estimation and obtained models with good accuracy, compared to the results obtained through gas chromatography (GC). To obtain this, the authors created three models, a first including all samples and two others based on only bulk samples (higher purity) and individual dose samples (lower purity). To create these models the authors applied the first derivative and vector normalisation as data pre-treatment before PLS modelling. Eliaerts et al. worked a lot on the use of SVM in combination with different spectroscopic techniques (NIR, MIR, and Raman) for the analysis of cocaine samples. They proposed a sampling and fast analysis tool for large cocaine seizures based on FT-IR using an ATR interface for spectra recording and SVM [44]. After recording, the spectra were pre-processed using SNV followed by the creation of two models using SVM-DA, one for the identification of cocaine and one for levamisole. Once cocaine was identified, the positive samples were used for SVM modelling in regression mode. The results were compared with the results obtained through traditional analysis using GC-MS and flame ionisation detection (FID), and it could be concluded that the proposed approach ATR-FT-IR-SVM allowed immediate and reliable information about homogeneity of the units, drug presence, and purity, reducing the number of samples needing confirmatory analysis and keeping sufficient high confidence to use the obtained results in the court. A similar study of the same group [46] compared the ATR-IR-SVM approach with the use of more traditional modelling with PLS(-DA). The results showed that the models using SVM-DA had significantly less false positives and negatives than the PLS-DA models, though the regression models, to estimate the purity of the cocaine samples were similar for both modelling approaches. They also repeated the experiments using the SVM approach of the previously mentioned studies, but this time comparing NIR, FT-IR, and Raman spectroscopy for the same purpose. They had to conclude that the best classification models were obtained using FT-IR, while the better quantification/regression was obtained using NIR spectra [47]. It was also tried to apply the ATR-FT-IR approach on smuggling cases, where it is the intention of smugglers to hide cocaine from identification, e.g., by adding certain substances. However, here the approach could be used as a kind of triage, but not as confirmatory analysis. Chromatographic methods remain necessary in these cases [48]. Kranenburg et al. tried to develop an approach based on a handheld near-infrared device, using a multistage model based on *k*-NN and ANN. Multistage means that when the results of the first model *k*-NN are inconclusive, the sample will be predicted based on the other submodels, in this study an ANN model. The authors used a combination of SNV and the first derivative as preprocessing, followed by outlier detection based on PCA. From there on, modelling started with first *k*-NN and further ANN [49]. The disadvantage of this study is that it will only identify cocaine samples when they have a purity of more than 20% w/w, and the models do not give an estimation of the purity. It is true that most of the time the purity of cocaine samples is above 40%, though samples with less than 20% w/w cannot be excluded and were already encountered, e.g., at music festivals. The same group also evaluated the use of a portable Raman spectrometer for this purpose and compared the performance of the built-in software with the outputs of an additional PLS and PLS-DA modelling. They had to conclude that the performance of the instrument was adequate and confirmed by the PLS models. The additional modelling could improve the performance, especially for samples returning an inconclusive conclusion by using software [50]. Again, also here no quantitative information about the samples was obtained, except of the comparison to a threshold. da Silva et al. remarked correctly that when using spectroscopy with modelling techniques as PLS and SVM, a high amount of samples are necessary to train the models and obtain high accuracy [51]. The authors proposed the use of MCR-ALS after the selection of the region 600–3,726 cm^{-1}

Table 1: Summary table of spectroscopy–chemometric applications for cocaine samples

Spectroscopic technique	Data pre-treatment	Chemometric technique	Target molecules	No of samples	Purity range cocaine	Qualitative analysis	Correct classification rate	Quantitative analysis	Prediction errors	Reference
ATR-FT-IR	SNV	SVM-DA/SVM	Cocaine/levamisole*	481/349*	56–96% w/w	+	100%	+	5% w/w/6% w/w*	[44]
ATR-FT-IR	SNV	SVM-DA/PLS-DA/SVM/PLS	Cocaine	482	4–99% w/w	+	99%	+	5.4–7.1% w/w	[46]
ATR-FT-IR	SNV	SVM-DA	Cocaine	482	0.8–35% w/w	+	99%	–	–	[48]
ATR-FT-IR	Normalisation	MCR-ALS	Cocaine, caffeine, levamisole, boric acid, aminopyrine, benzocaine, sodium bicarbonate, phenacetine, lidocaine, starch	728	19.9–99% w/w	+	66.7–100%	+	10.2% w/w (highest error, boric acid)	[51]
ATR-FT-IR	Genetic algorithm	kNN	Cocaine salt and base	513	NA	+	100%	–	–	[40]
ATR-FT-IR	Wavenumber importance index	kNN, LDA, ANN	Cocaine salt and base	513	NA	+	100%	–	–	[52]
ATR-FT-IR	Normalsiation, 1st derivative, mean centring	PLS-DA, SVM-DA	Cocaine salt and base	513	NA	+	100%	–	–	[53]
ATR-FT-IR	SNV, mean centring	PCA, PLS-DA	Cocaine salt and base	91	1–73.6% w/w	+	97–83%	–	–	[54]
ATR-FT-IR	SNV, orthogonal signal correction, 1st derivative, mean centring	PLS	Cocaine	275	NA	–	–	+	2.5% w/w	[55]
ATR-FT-IR	SNV, orthogonal signal correction, 1st derivative, mean centring	PLS/PLS-DA	Cocaine salt and base, phenacetine, benzocaine, aminopyrine, lidocaine, caffeine, levamisole	1,085	24.2–99.9% w/w	+	99.7%	+	0.23–1.66% w/w	[56]
ATR-FT-IR	SNV	PLS	Cocaine, caffeine, lidocaine, procaine, phenacetin, mannitol	Simulated samples	68.4–75.9% w/w	–	–	+	0.09–1.74% w/w	[57]
ATR-FT-IR/NIR/ Raman	SNV	SVM-DA/SVM	Cocaine	364	4–99% w/w	+	100% (FT-IR)	+	3.79% (RMSECV)	[47]
NIR	First derivative and vector nomalisation	PLS	Cocaine	78	11.4–86.4% w/w	–	–	+	3–4% w/w	[45]

(continued)

Table 1: *(continued)*

Spectroscopic technique	Data pre-treatment	Chemometric technique	Target molecules	No of samples	Purity range cocaine	Qualitative analysis	Correct classification rate	Quantitative analysis	Prediction errors	Reference
NIR	SNV + first derivative	kNN/ANN	Cocaine	166	19.1–99.4% w/w	+	100–90.6%	–	–	[49]
NIR	Second derivative	SIMCA/PLS	Cocaine, phenacetin, levamisole	240	0.5–99% w/w	+	100%	+	4.7–8.2% w/w	[58]
NIR	NA	Randomised trees	Cocaine	2,047	0–100% w/w	+	99.4%	+	15% w/w	[59]
Raman	SNV + Savitsky–Golay smoothing + baseline correction	PLS/PLS-DA	Cocaine	676	19–99% w/w	+	90.5%	–	–	[50]

of the FT-IR spectrum and normalisation of the data to solve this problem. Indeed, MCR-ALS is a technique allowing qualitative and quantitative estimation based on less calibration data and can also provide an estimate of the pure spectral profile of unknown substances. The study showed the potential of this approach since it allowed to obtain rapid information on the composition of the samples and dosage estimations. The advantages over classical modelling such as PLS were the need for less calibration data and the possibility to identify unknown components, by estimating pure profiles and comparing them with commercial libraries [51]. Anzanello et al. [40,52] worked on the optimisation of models based on FT-IR spectra to classify cocaine samples as base or salt forms. Therefore, they used not only classical chemometric modelling techniques such as *k*-NN and LDA but also ANN. The optimisation consisted of the selection of the wavenumbers of the spectra and hence reducing the noise present in the spectral data. Two methods were proposed, one based on the calculation of a wavenumber importance index using the Bhattacharyya distance [52] and another based on a genetic algorithm. Both approaches allowed indeed to obtain a slight improvement of the classification rates of the models differentiating between salt and base forms, though it can be asked if the added complexity to the application is in balance with the gained accuracy of the models. Marcelo et al., for example, used the complete fingerprint region of the FT-IR spectra and after pre-processing that comprised normalisation, first derivative and mean centring were able to obtain models with 100% correct classification rates using PLS-DA and SVM-DA, respectively, to differentiate between salt and base containing samples. They also used PCA and HCA more for a further profiling of the samples to determine purity and adulterants, with the ultimate goal to link seizures to the same origin or trafficking network [53]. Similar applications using ATR-FT-IR with PCA and PLS to determine chemical form and dilution [54] and quantify the cocaine content in seized samples [55] were also published. The group of Grobério et al. also extended their approach based on FT-IR and PLS to obtain a complete characterisation of samples, creating distinct models for chemical form (PLS-DA), cocaine content (PLS), the degree of oxidation of the samples (PLS), and the quantification of the adulterants phenacetine, benzocaine, aminopyrine, lidocaine, caffeine, and levamisole (PLS). All models resulted in acceptable figures of merit, except for the model for levamisole [56], and so they proved to be able to almost completely characterise cocaine samples based on the measurement of the FT-IR spectrum, thanks to chemometrics. A similar study was performed by Materazzi et al. [57]. They used PCA, PLS, and PCR to obtain quantitative results about cocaine content and adulterant concentrations

(caffeine, lidocaine, procaine, phenacetin, and mannitol). The best models were obtained using PLS with SNV as a spectral pre-processing method and allowed accurate predictions of the different concentrations, e.g., for cocaine, an overall R^2 value of 0.9999 was obtained between the predicted results with the model and the results obtained by GC-MS.

Hespanhol et al. [58] tried to solve the different questions using a low-cost portable NIR spectrometer that can be easily linked to a computer through an USB interface. They managed to classify samples according to the presence of cocaine and its chemical form using SIMCA classification models and to determine cocaine content and the content of two adulterants (phenacetin and levamisole) using PLS. For all models, the second derivative of the measured spectra was used as input for the models. Also Coppey et al. [59] evaluated the use of a low-cost NIR device, linked to an application on a smartphone. On the basis of a selection of the region in the spectrum of interest, they created classification models (cocaine vs non-cocaine containing samples). Therefore, they tried a series of chemometric classifiers and found an optimal model using a set of randomised tree methods. However, no further explanation about the chemometrics used was given in the article. The final model showed a sensitivity of 0.994 and a selectivity of 1. They also created quantitative models with a correlation of 0.964 (dispersion of relative error <15%) with the results obtained by GC-MS, but again it is not clear which modelling algorithm was employed.

3.2 Amphetamines and 3,4-methylenedioxymethamphetamine (MDMA/ecstasy)

The second group of the drug of abuse products for which applications of spectroscopy hyphenated with chemometrics were described are amphetamines, and especially MDMA, which is the second most used party drug in night-life settings [60]. Amphetamines are in fact a group of related compounds with β-phenylisopropylamine as the basic structure. Some of these molecules were marketed as medicines, e.g., for nasal congestion or weight control, though they are more and more abandoned and even prohibited in many countries. Most of the amphetamine analogues have central nervous stimulating and anorexinogenic properties and can be classified as psychoanaleptica. The analogues containing a 3,4-methylene-dioxy substituted phenyl ring, like MDMA,

also have hallucinogenic and mood-modifying properties and are considered psychodysleptica [61]. Concerning the analytical challenges, the major issue is not only the high diversity of molecules and analogues, part of this group, but, as for cocaine, also the diversity of adulterants and diluents present in the seized products [62]. Table 2 summarizes the most important feature of the applications described in the literature, which are discussed hereunder.

The study by Praisler used PCA and SIMCA to create screening techniques for amphetamine-based drugs of abuse. The models are based on the analysis of the samples with GC, hyphenated with FT-IR. In the first article [61], they used this strategy to create an automated system to distinguish between amphetamines and other drugs of abuse, but also among amphetamines according to the substitution patterns. This also allows to classify new molecules according to their pharmacological activities, i.e., hallucinogenic and stimulants. For the exploratory analysis (PCA), they used the mean-centred FT-IR spectra, while the SIMCA models were created based on the scaled spectra using the standard deviation. In two follow-up articles [63,64], the screening strategy was extended to distinguish between the main stimulant and hallucinogenic amphetamines and so to distinguish between, at that time, amphetamines used as medicines and as drugs of abuse as well as to give a first classification of "negative" samples. These samples contained amphetamine-like molecules, not present in a database, and so classified as negative by the SIMCA model. Through the introduction of a second closest class, a first idea can be given about its use as a stimulant or as a hallucinogenic. Another follow-up study [65] optimised the screening models for amphetamines by also taking into account GC-FT-IR signals due to degradation products of some thermally unstable amphetamine analogues, improving the specificity and sensitivity of the models. The same group used also ANN for the same purpose. In this case, the spectra obtained through GC-FT-IR were normalised and used as input to train ANN models. The best ANN model was able to distinguish between stimulant and hallucinogenic amphetamines and non-amphetamines with a false negative rate of 0% [66]. Working on the same project, Gosav et al. also presented an extension of this article, combining IR and MS data to distinguish between the same groups. Therefore, they trained ANNs using the spectral data after autoscaling and mean centring, and they limited the number of input variables using PCA and so eliminate the less relevant data. The final model was able to classify unknown samples with high true positive and low false negative rates [67]. Goh et al. [68] used PCR for the quantification of methamphetamine in mixtures with

Table 2: Summary table of spectroscopy–chemometric applications for amphetamine and MDMA samples

Spectroscopic technique	Data pre-treatment	Chemometric technique	Target molecules	No of samples	Purity range	Qualitative analysis	Correct classification rate	Quantitative analysis	Prediction errors	Reference
(GC-)FT-IR	Mean centring, scaling based on variable standard deviation	PCA, SIMCA	Amphetamines	NA	NA	+	93.9%	–	–	[61,63–65]
(GC-)FT-IR	Normalisation	ANN	Amphetamines	NA	NA	+	85.71%	–	–	[66]
(GC-)FT-IR	Autoscaling, mean centring	ANN	Amphetamines	NA	NA	+	94.17%	–	–	[67]
ATR-FT-IR	Normalisation	PLS	MDMA	96	0.01–78.6% w/w	–	–	+	3.8% w/w	[71]
ATR-FT-IR, NIR	SNV	PLS-DA, PLS	MDMA	267	18.1–72.7% w/w	+	96%	+	5% w/w	[72]
FT-IR, ATR-FT-IR, Raman	Baseline zeroing, normalisation	PCR	Methamphetamine	NA	NA	–	–	+	4% w/w (ATR-FT-IR)	[68]
NIR	Second derivative	PLS	Amphetamine, MDMA, MDE	120	27–42% w/w	–	–	+	2.1% w/w (MDMA)	[69]
NIR	1/x Transformation, second derivative, Savitsky–Golay smoothing			120	54.5–94.18% w/w (MDMA)// 54.7–85.9% w/w	–	–	+	0.9% w/w (MDMA)// 0.6% w/w (MDE)	[70]
Raman	Second derivative	PLS	Amphetamines	10	8–69% w/w	–	–	+	<1% w/w	[62]

caffeine, glucose, and paracetamol and compared the use-fulness of FT-IR in transmittance mode, ATR-FT-IR, and Raman spectroscopy. They had to conclude that the FT-IR approaches clearly outperformed Raman. For FT-IR base-line correction, zeroing and normalisation were necessary before modelling, where for ATR-FT-IR, only normalisation was applied, although it should be mentioned that ATR-FT-IR is more influenced by heterogenic samples and by differences in the particle size. Issues that should be taken into account when creating the calibration models and during the analysis of real samples. Although the authors did not use real samples, but prepared and there-fore simplified samples, they showed that the content of methamphetamine in the mixture could be predicted using PCR with errors around 3% or 4%. Further, only few appli-cations of Raman spectroscopy for the analysis of amphe-tamines can be found in the literature. One application described a quantification of amphetamine samples using PLS, based on Raman spectra, obtained after dissolving the samples in an acidic solvent, containing an internal stan-dard [62]. As input for the model, the second derivative of the spectra was used and resulted in quantitative results in good agreement with the results obtained by liquid chro-matography, which served as a reference method.

Next to the applications for amphetamines, also spe-cific literature is available on the use of spectroscopy and chemometrics for the characterisation of ecstasy tablets, i.e., MDMA tablets. Sonderman and Kovar [69] presented a PLS approach with cut-off values to differentiate among placebo, amphetamine, and ecstasy samples using the second derivative as spectral pre-treatment for the recorded NIR spectra. A second model was created for the "ecstasy group" using PLS on the raw spectra for the distinction between MDMA and the structurally related compound N-ethyl-3,4-methylenedioxyamphetamine (MDE). Although the approach can be applied on the tablets as such, they clearly showed that better results were obtained with crushed tablets due to homogeneity and production pro-blems. A study of the same group [70] extended this PLS approach by applying PLS regression models to quantify the active compounds in ecstasy tablets, i.e., MDMA and MDE. Models were respectively calculated with spectra measured in transmission mode and in reflectance mode. For the first a $1/x$ transformation was applied, why for the latter an additional second derivative of the spectral data was necessary. Both models gave quantitative predictions in line with the results of a reference liquid chromato-graphy method. A similar study was performed by Hughes et al. [71], but this time using ATR-FT-IR spectra with nor-malisation as pre-treatment. Using PLS with the major characteristic peaks of the spectrum, a model was achieved

able to quantify MDMA with a prediction error of 3.8 and so obtaining a cost-saving alternative for standard chromato-graphy methods. Deconinck et al. [72] made use of PLS-DA models to distinguish between ecstasy tablets and other tablets, followed by a PLS regression model for the quanti-fication of MDMA in the tablets. For that purpose, three different spectral data sets were compared, i.e., the FT-IR and NIR spectra on the crushed tablets and the NIR spectra on the intact tablets. All spectral data was pre-treated using SNV. The results showed that the best approach, for onsite testing, was obtained using the NIR spectra on the intact tablets for discrimination of the ecstasy tablets from the others, followed by a quantitative PLS model on the NIR spectra of the crushed tablets. The qualitative model showed a correct classification rate of 96%, while the quantitative model was able to predict the MDMA content within the error limits of 5%.

3.3 Heroin

Heroin is an opiate drug, synthesised from morphine through acetylation, also called 3,6-diacetyl morphine [73]. Heroin represents about 8% of the world's drug sei-zures and presents itself as a white powder with bitter taste in its pure form [73,74]. Though pure heroin is rarely found as a street drug and most often colours vary between white and dark brown, depending on the impu-rities left from the production process and the additives present. The wide variety of the purity of heroin samples represents the highest risk due to the difficulty of dosing and so the risk of overdose. A study in Serbia showed that heroin purity can vary between 1.7% and 58.8% w/w [74]. Diluents that are often found are inert powders like chalk, flour, talcum powder, and D-glucose, and also the phar-macological active products paracetamol and caffeine [73,74]. Also here the presence of multiple additives and the wide range in purity represents the major chal-lenge for onsite analysis using spectroscopic approaches.

Only a limited number of approaches combining spectroscopy and chemometrics for heroin characterisa-tion could be found and are summarised in Table 3. The first application used FT-IR in a transmittance mode with KBr tablets and applied HCA to cluster samples according to their origin [75]. Moros et al. [76] compared the perfor-mance of three PLS models and created using the full NIR spectra obtained for heroin samples and using two dif-ferent kinds of variable selection techniques and so only using part of the spectra. The sample set covered a range of purity from 6% to 34% w/w, and they were able to

Table 3: Summary table of spectroscopy–chemometric applications for heroin samples

Spectroscopic technique	Data pre-treatment	Chemometric technique	Target molecules	No of samples	Purity range	Qualitative analysis	Correct classification rate	Quantitative analysis	Prediction errors	Reference
ATR-FT-IR	Mean centring	PLS	Heroin	NA	NA	–	–	+	4.3% w/w	[74]
ATR-FT-IR	Baseline correction, multivariate scatter correction, standard normal variables, Savitzky–Golay smoothing	ANN	Heroin	NA	NA	+	86.7%	–	–	[78]
NIR	Variable selection	PLS	Heroin	31	6–34% w/w	–	–	+	1.5–5% w/w	[76]
NIR	Window selection, SNV, detrending	PLS	Heroin	31	6–34% w/w	–	–	+	1.3% w/w	[73]
NIR	Mean centring, SNV	PLS	Heroin	NA	NA	–	–	+	0.09–0.12% w/w	[77]
NIR	NA	Gradient boosting classifiers and regressors	Heroin	416	3–67% w/w	+	99.8%	+	<15% w/w	[59]

obtain quantitative predictions with errors within the range of 1.5% and 5% w/w. The same group also described a similar study based on a selected spectral window of the recorded NIR spectra of the same sample set. Based on this window, they were able to create a PLS model with a prediction error of 1.3% w/w, a quality coefficient of 10%, and a residual predictive deviation of 5.4 [73]. Melucci et al. [77] compared a portable and benchtop NIR instrument for the quantification of heroin, caffeine, and paracetamol, in self-prepared ternary mixtures. They used the SNV pre-treated spectra after centring to create PLS models for each of the three components and obtained good predictive models for both instruments. Unfortunately, this approach was not tested on real samples, which are often more complex than the ternary mixtures, the approach was applied. Coppey et al. [59] evaluated the already previously mentioned low-cost NIR spectrometer for the identification and quantification of heroin samples. They therefore selected the region of interest for heroin from the spectra and applied a group of gradient-boosting classifiers to obtain a classification model with a sensitivity of 0.998 and a selectivity of 1. Further, they created a quantitative model, showing a correlation of 0.964 (dispersion of relative error <15%) with the results obtained with GC-MS. Unfortunately, no details on the chemometric algorithms used are given. In a more recent study, Stevanović et al. [74] did a similar thing using standard mixtures of heroin, caffeine, paracetamol, sucrose, and ᴅ-glucose and applying ATR-FT-IR. The most characteristic vibrational bands were chosen for each of the investigated components, and these were used in PCA, as exploratory analysis to reveal hidden correlations between the samples, and in HCA to cluster the mixtures with similar composition and finally in PLS for the estimation of the heroin content. A PLS model with a prediction error of 4.3 was obtained in the range of heroin dosages from 15% to 50% w/w. Also He et al. [78] did a study on standard mixtures of heroin with its most often encountered adulterants. They only used binary mixtures, which is a simplification of the real situation, and used ANN modelling and linear fitting analysis to obtain classifiers for the heroin mixtures according to the adulterant or diluent present.

3.4 Cannabis

Cannabis sativa L. is one of the oldest plants used in traditional medicine in central and Northeast Asia [79]. Also known as marijuana, this plant is known as the number one drug of abuse worldwide. The psychotropic

properties are mainly due to the presence of the cannabinoid delta-9-tetrahydrocannabinol (Δ9-THC) [80]. Recently, the plant gained also interest for its medicinal use, resulting in the first products registered as medicine and derived products based on medicinal claims of other cannabinoids, present in the plant, came into the market. The most sold and popular products are the ones based on cannabidiol (CBD). The latter products are therefore limited in their Δ9-THC content. Next to this, *Cannabis* varieties are also cultivated for industrial fibres and animal food and should legally have a Δ9-THC below 0.2% m/m in most European countries [81]. All these different applications of the plant make legislation complicated and therefore, especially for the legal limits of the Δ9-THC content, surveillance is necessary. The biggest challenge here for the application of spectroscopic methods is different as for the other drugs, since it is a herbal matrix, with high complexity, containing more than 500 different compounds as flavonoids, monoterpenes, sesquiterpenes, steroids, and cannabinoids [82]. Table 4 presents an overview of the applications found in the literature that are discussed in this review.

A first application describes the application of supervised (PLS-DA and SVM-DA) and unsupervised techniques (PCA and HCA) on NIR spectral data to differentiate and classify cannabis plants from a greenhouse according to their growing stage [83]. The authors obtained the best differentiation with PCA after normalisation of the NIR spectra followed by the first derivative. SVM-DA gave the best classification model according to the growth stage, using the same spectral pre-treatment as for PCA. The intention was to apply these models to rapidly gain information about indoor cultivation time and establish a connection among cultivation sites, trafficked seeds, and trafficking routes of cannabis [83]. Another application describes the use of PLS regression to quantify the major cannabinoids in a raw plant material [84]. The authors linked the spectra obtained with dispersive and FT-NIR with the quantitative data obtained with GC-FID. They used PLS on the spectral data after applying SNV and the first derivative. With both types of NIR spectroscopy, they obtained distinct regression models for each of the considered cannabinoids allowing them to accurately predict the respective content in the samples. No significant differences could be observed for the models based on the respective methods. Duchateau et al. [85] compared a benchtop NIR and a handheld NIR device to classify cannabis samples according to European and Swiss law, i.e., they tried to classify into three classes: below 0.2% m/m Δ9-THC, above 0.2% m/m Δ9-THC (European directive), and above 1.0% m/m Δ9-THC (Swiss law). They measured 189 samples with both instruments and applied different pre-treatment techniques and modelling techniques (k-NN,

SIMCA, and PLS-DA) for this purpose. For the benchtop data, the second derivative followed by SNV was chosen as pre-treatment of the data, while for the handheld data, SNV and the first derivative were chosen. The authors first created a binary model according to the European regulation and selected a SIMCA model as optimal with an accuracy of 91% and 93% for an external set for benchtop and handheld data, respectively. For the tertiary models, PLS-DA gave the best results with 91% and 95% of accuracy. These models and especially the ones with the handheld device can be used by inspectors, e.g., in CBD shops or even in agriculture [85]. A totally different application was presented by Pereira et al. [86]. They applied near-infrared hyperspectral imaging to a representative setting of Brazilian flora to detect illegal cannabis plantations. In this feasibility study, they recorded hyperspectral images of several settings, and after the application of smoothing algorithms (Savitzky–Golay) and SNV, they applied a PCA approach for the selection of the most discriminative bandwidths. Further, they developed a 1-class SIMCA model for the classification of cannabis plants and others. They obtained a model with a sensitivity of 89.45% and a specificity of 97.60% for an external test set. The idea is to implement this approach using a drone, allowing screening at cultivation sites [86]. Very recently, Deidda et al. [87] compared two handheld NIR devices for the quantification of Δ9-THC in cannabis inflorescences and resins and compared it with the results obtained by liquid chromatography. They recorded NIR spectra with both devices and build genetic algorithm-PLS and machine learning models [88] after pre-treatment of the spectra with the second derivative and SNV. They could observe that one instrument outperformed the other, and they point at the selection of the device. Technical features should be taken into account as well as the sample analysis window, especially in this case where samples can be very heterogeneous. With the best device they were able to obtain good correlation between the predicted values with the NIR models and chromatographic data and this for an external test set. They proved a good agreement between the results using the bland-Altman statistic [89].

Next to NIR applications, Geskovski et al. [90] recently explored the possibilities of using ATR-FT-IR as process analytical technology in the production of cannabis products for medicinal use. They created a PLS model linking selected bands of the IR spectrum to the quantitative results for, respectively, Δ9-THC and CBD, obtained by HPLC. They selected the most important bands based on the literature and pre-processed the spectral data using Savitsky–Golay smoothing and the second derivative. At

Table 4: Summary table of spectroscopy–chemometric applications for cannabis samples

Spectroscopic technique	Data pre-treatment	Chemometric technique	Target molecules	No of samples	Purity range	Qualitative analysis	Correct classification rate	Quantitative analysis	Prediction errors	Reference
ATR-FT-IR	SNV, second derivative	PLS	Δ9-THC, CBD	2	0.32–20.4% w/w/ 0.18–22.85% w/w	–	–	+	1.3–3.8% w/w	[90]
NIR	Normalisation, first derivative	PLS-DA, SVM-DA	Cannabis plants	29	NA	+	100%	–	–	[83]
NIR	SNV, first derivative	PLS	Cannabinoids	189	0–22% w/w (Δ9-THC)	–	–	+	1.72% w/w (Δ9-THC)	[84]
NIR	SNV, second derivative, Savitzsky–Golay smoothing	kNN, SIMCA, PLS-DA	Cannabis plants	189	NA	+	95%	–	–	[85]
NIR	SNV, second derivative	GA-PLS, machine learning models	Δ9-THC	14	6.4–20.2% w/w	–	–	+	1.46% w/w	[87]
NIR (imaging)	Savitzsky–Golay smoothing, SNV	SIMCA	Cannabis plants	NA	NA	+	89.4%	–	–	[86]
Raman	SNV, normalisation, first derivative	O-PLS-DA	Cannabis plants	86	0.07–10.3% w/w	+	95–100%	–	–	[91]
Raman	SNV, normalisation	O-PLS-DA	Cannabis plants	212	0.09–4.09% w/w (Δ9-THC)	+	100%	–	–	[92]

the end, they obtained PLS calibration models that were able to monitor the Δ9-THC and CBD concentration in dried flowers and extracts with prediction errors between 1.33% and 3.79%. Sanchez et al. applied Raman spectroscopy to differentiate between hemp and different varieties of cannabis [91], as well as to differentiate among hemp, cannabis, and CBD-rich hemp [92]. In both applications, the Raman spectra were recorded for a set of samples and then pre-treated with SNV and normalised. For the former application, an additional first derivative was calculated. Further, they applied O-PLS-DA models to classify the different samples and obtained accuracies of 100% for the differentiations among hemp, cannabis, and CBD-rich hemp, as well as accuracies between 95% and 100% to differentiate among three cannabis varieties.

3.5 New psychotropic substances (NPS)

Next to the "classical" drugs described earlier, new substances and designer molecules, created to mimic the effects of known and regulated drugs, are detected. They are called NPS or even legal highs. This phenomenon, that is present since over a decade and mainly a reaction of circumventing legislations, has a huge impact on the users, since it are often analogues or derivatives of molecules, which are not studied and so neither the toxic dose nor the effects are predictable, resulting in higher risks and mortalities [93,94]. One example is the fentanyl crisis in North America [95,96]. Fentanyl and fentanyl analogues are considered as one of the major concerns in this context due to their prevalence, diversity, and potency, though there are also non-fentanyl synthetic opioids emerging, which adapt and diversify according to the evolution of the drug market and the national and international legislations [94,97]. Apart from the emerging synthetic opioids, reported NPS belong also to the synthetic cannabinoids, the synthetic cathinones, tryptamines, piperazine derivatives, piperidines, pyrrolidines, and the phenethylamines [93,98].

The applications found in the literature combining spectroscopy and chemometrics are summarised in Table 5. Due to the diversity of the applications and the way data were analysed, the quality features mentioned in the table are not always immediately comparable, and the readers are directed to the respective references for the details. A first application was described using ATR-FT-IR spectra to discriminate NBOMe's, a generic denomination for phenethylamines presenting a 2-methoxybenzyl group replacing a hydrogen on the amine, in blotters [99]. The authors used a simple discriminant analysis to distinguish among blotters containing NBOMEs (as group), lysergic acid diethyamide (LSD), mescaline, and blanks. Although the disadvantage of the model was that they could not differentiate between the three kinds of NBOMEs found in their sample set, the model allowed the detection of NBOMEs and LSD in lower concentrations than the library comparisons. A similar study was performed Custódio et al. [100]. In this study, two PLS-DA models were created that could be used in a two-step approach. The first model distinguishes between drug free and positive blotter, while the second model classifies the blotter as containing NBOMe's or NBOH molecules. To do so, the authors applied first a selection of spectral regions of the FT-IR spectrum for both models, followed by the application of the second derivative and mean centring before PLS-DA modelling. For both models, correct classifications of more than 96% were obtained compared to the results obtained with GC-MS [100]. A remark here is that the authors performed outlier testing and removing, which is a correct approach in chemometric modelling, though when doing so an explanation should be given, since if this outlying behaviour is due to the nature of the sample, future similar samples will not be identified correctly by the model, since this variation was not included in the training of the model. A very similar study using a handheld NIR device was performed by de Oliveira Magalhães et al. [101]. They collected the spectral data, and after SNV and mean centring, PLS-DA and SIMCA were applied for the same purpose. The PLS-DA models clearly outperformed the SIMCA models with correct classification for the model distinguishing between NPS containing blotters and other of 100% and for the model making difference between NBOMe' and NBOH of 97.1% [101]. In the context of the fentanyl pandemic, Xu et al. [102] presented an approach based on infrared spectral data to distinguish between analogues of fentanyl according to their functional groups. They performed first an exploratory analysis with PCA, followed by creating distinct linear models for four functional groups (amide, aniline, benzene, and piperidine). On the basis of these models, the authors were able to obtain an overall classification accuracy of 92.5%. Combining the binary models, to detect the fentanyl analogues containing two or more of the functional groups, diminished the accuracy rate to 79.4%. Another application using mid-IR spectroscopy and chemometrics for the analysis of fentanyl was recently described by Tobias et al. [103]. They created a quantitative PLS model using quantitative data on fentanyl content of heroin samples, obtained by nuclear magnetic resonance (qNMR), and used it in routine analysis at their point-of-care drug checking service to follow the trend in the fentanyl content and dosage in heroin samples in function of time. Although the authors used the vendor's software for

Table 5: Summary table of spectroscopy–chemometric applications for NPS samples

Spectroscopic technique	Data pre-treatment	Chemometric technique	Target molecules	No of samples	Purity range	Qualitative analysis	Correct classification rate	Quantitative analysis	Prediction errors	Reference
ATR-FT-IR	NA	Discriminant analysis	NBOMe's	39	NA	+	91%	–	–	[99]
ATR-FT-IR	Second derivative, mean centring	PLS-DA	NBOMe's, NBOH	158	NA	+	96%	–	–	[100]
FT-IR	Rescaling to unit standard deviation	Generalised linear models	Fentanyl analogues	NA	NA	+	92.5%	–	–	[102]
FT-IR	NA	PLS	Fentanyl analogues	3,621	4.5–10.4% w/w	–	–	+	NA	[103]
IR	NA	SIMCA	Synthetic cannabinoids	NA	NA	–	100%	–	–	[104]
NIR	SNV, mean centring	PLS-DA, SIMCA	NBOMe's, NBOH	101	NA	+	97.1%	–	–	[101]
NIR	SNV	PCA	Synthetic cannabinoids, NBOMe's	13	NA	+	NA (clustering)	–	–	[105]
Raman (SERS)	AsLS, autoscaling	PCA, PLS	Synthetic cannabinoids, diphenidines, aminoindanes, cathinones	NA	NA	+	NA (clustering)	–	–	[106]
Raman (SERS)	NA	PCA, PLS-DA	Fentanyl and morphine analogues	NA	NA	+	NA (clustering)	–	–	[107]
Raman (SERS)	MSC, the second derivative with Savitzky–Golay smoothing, mean centring	PLS-DA	Fentanyl analogues and precursors	NA	NA	+	96–96.7%	–	–	[108]
Raman (SERS)	NA	PLS	Fentanyl analogues	NA	NA	–	–	+	NA	[26]
Raman (SERS)	Second derivative, Savitzky–Golay smoothing, normalisation	PLS-DA, PLS	Fentanyl analogues	NA	NA	+	NA	+	NA	[109]

modelling and not much details on the model are described, it is good and one of the few routine implementations of chemometrics in the context of harm reduction and drug checking. de Castro et al. [104] used theoretical chemistry to predict the IR spectrum for 41 synthetic cannabinoids. Based on these theoretical spectra, they used PCA to define six different clusters based on the differences in substitution. Further classification models were build using SIMCA for these six clusters/classes, which showed a 100% accuracy. In this case, chemometrics were used to prove that theoretical chemistry can help in predicting the IR spectrum of molecules for which less or no analytical data are available. In this context, the approach, with the checks with PCA and SIMCA for validity, could be generalised for all NPS or suspected molecules.

An application of NIR was described by Risoluti et al. [105], where they measured NIR spectra for simulated and confiscated samples containing synthetic cannabinoids and NPS of the phenethylamine group. They applied PCA to differentiate between the two classes of NPS despite the influence of the complex matrices.

Raman spectroscopy was also applied for the detection and characterisation of NPS. Muhamadali et al. [106] used both Raman spectroscopy as SERS for the identification and discrimination between synthetic cannabinoids, diphenidines, aminoindanes, and cathinones. The spectral data for the solid samples were first base line corrected using asymmetric least squares (AsLS) algorithms and autoscaling. Further, PCA was used to cluster the different classes. The authors were able to obtain a good clustering for all classes with the spectra measured directly on the powdered samples as well as on sample solutions. For the latter solutions, they performed also the analysis with SERS allowing the clustering as well as a good quantitative predictability using PLS. This latter part of the study was a preliminary study to apply SERS for the detection of NPS in biological fluids, e.g., in urine, which is out of the scope of this review. A very similar approach was used by Wang et al. [107] for the differentiation of fentanyl and morphine analogues as well as the detection of fentanyls in heroin samples. Also, here, the samples were brought in solution, and PCA plots were used for clustering the different types of molecules. The authors used also PLS-DA, but only used the score plot, which gave similar results as PCA. They did not create real classification models. Also, SERS was used to differentiate between fentanyl and two of its precursors. Therefore, PLS-DA was used in a hierarchical way (i.e., three different models, where the output of the first goes to the second, etc.), and Raman spectra were recorded through the SERS technique using standard solutions of the analytes. The spectra were pre-processed using MSC, the second derivative with Savitzky–Golay smoothing, and mean centring. The models were able to classify the different analytes although no application on real samples was described [108]. In the context of the fentanyl crisis, Smith et al. [26] evaluated the use of SERS for fentanyl detection and quantification. For the latter, they applied a PLS regression model. Unfortunately, a lot of details are not obtained about the model. It was only shown that a determination coefficient of 0.96 could be obtained between real and predicted concentrations. Next to this, the authors also evaluated the impact of the presence of heroin and glucose. It was possible to detect fentanyl, but not to create a (semi-)quantitative model for fentanyl in the presence of these two molecules. On the other hand, good calibration models could be obtained for four analogues of fentanyl. Although for the same purpose, detection and quantification of fentanyl in heroin and cocaine samples, Wang et al. [109] followed a slightly different approach. They used SERS but with a portable Raman spectrometer. The proposed spectra were pre-treated using the second derivative with Savitzky–Golay smoothing, followed by normalisation. For chemometric techniques, they used PCA and PLS-DA for the detection of fentanyl in binary mixtures fentanyl/heroin and fentanyl/cocaine and PLS for quantitative analysis. In contrary to what was done in the other studies, they took the whole spectrum into account for modelling, improving significantly the detection limits, comparing to the use of single intensities (marker band approach), and improving the quantitative analysis compared to univariate analysis based on single intensities at a characteristic wavelength of fentanyl. It must be mentioned that this was only done on standard solutions and that no real samples were taken into account or used to validate the approach. In most of the papers using SERS, modelling was used, but either for clustering or to lower the detection limits, that is why these papers do not give prediction error or other validation parameters for their models (Table 5).

3.6 Multi-analyte analysis

All previously mentioned applications always concern a certain drug product or a set of related compounds. This is not the real situation occurring at drug checking services or in on-site analysis in a forensic context. Indeed, often the analysts receive a product within the best case, which is an indication of what it might contain. This is the reason why more attention should go to multi-analyte and preferably untargeted approaches, allowing the fast

Table 6: Summary table of spectroscopy–chemometric applications for multiple analytes

Spectroscopic technique	Data pre-treatment	Chemometric technique	Target molecules	No of samples	Qualitative analysis	Correct classification rate	Quantitative analysis	Prediction errors	Reference
ATR-FT-IR	SNV, centring	PLS-DA	5-MeO-MIPT, MDMA, and MDA, methamphetamine, cathinones	92	+	100%	−	−	[112]
ATR-FT-IR	MSC, SNV, Savitzky–Golay smoothing	Bayes discriminant analysis	Heroin, methamphetamine, ketamine	NA	+	95%	−	−	[113]
ATR-FT-IR, NIR	SNV, 1st derivative, log10 transformation	PLS-DA	Cocaine, amphetamine, ketamine, other	287	+	93.1%	+	2.81–5.78% w/w	[114]
NIR	First Derivative, SNV	SIMCA, SVM-DA	Methamphetamine, ketamine, cocaine, heroin, other	836	+	87.7–90%	+	<3.6% w/w	[111]
Raman	First derivative, automated polynomial method	PCA	Cocaine, heroin, MDMA	NA	+	NA (clustering)	+	3.2–8.0% w/w	[27,115]

identification of diverse illicit compounds at the same time [110]. Several such approaches were already described in the literature and more and more, an evolution towards this untargeted screening occurs. Table 6 summarizes the most important features of the found applications.

A first tryout for an untargeted screening was found in the work of Praisler et al. [63,64]. In both studies, the authors created SIMCA models to distinguish between amphetamines analogues and between illicit used amphetamines and the ones, at that time, used for medicinal purposes. As mentioned earlier, these models made use of GC-FT-IR data. In both papers, they also tried to extent the obtained SIMCA models to classify the negative samples as other hallucinogens or stimulants, sympathomimetic agents, and narcotics. However, the authors acknowledge the lesser performance of the models for the latter purpose, and it was already a first step in the untargeted screening of illicit drug products [64]. After this initial attempt, the next study allowing the analysis of more than one drug compound in one analysis was presented by Liu et al. [111]. These authors used NIR spectra of four drug types (methamphetamine, ketamine, cocaine, and heroin) to build qualitative and quantitative models. Therefore, the spectra were pre-treated using a first-order derivative and SNV. An exploratory analysis with PCA showed a clear distinction between the four groups of samples, allowing to proceed to the calculation of classification models. SIMCA and SVM-DA were used as classification algorithm, where SIMCA proved to be the best option in this case. The false positive rate was zero, while the false negative rate was 10.5%, due to the purity of some samples. The authors included also a set of NPS samples, which were all unclassified in the SIMCA model and therefore selected for further analysis. Once classified, PLS quantitative models were created for the four targeted groups with prediction errors for the validation set below 3.6%. Pereira et al. [112] made use of ATR-FT-IR to screen ecstasy tablets. Ecstasy tablets should in fact contain MDMA, though often other compounds as certain NPS are present. The spectra were pre-treated using SNV and centring, and further, PLS-DA models were built in a hierarchical way. Here, the presented approach is more multi-analyte since no class of "other" or "for further analysis" was considered. First, a PLS-DA was created classifying tablets containing 5-MeO-MIPT, methyl amphetamine derivatives (MDMA and MDA), methamphetamine, and cathinones (methylone, ethylone, and PV-8). In a next step, two separate submodels were calculated for the discrimination of MDMA and MDA and within the class of the cathinones for methylone, ethylone, and PV-8. The main model resulted in 100% correct classification rates for the

four classes. Also, the two submodels gave 100% correct classification. It must be said that the cathinone model was only validated through cross-validation due to a lack of samples for the three targeted components.

Also He et al. [113] presented a multi-analyte approach using ATR-FT-IR for the discrimination between heroin, methamphetamine, and ketamine samples. For this purpose, they prepared mixtures of different concentrations of the active compounds and their most often encountered cutting agents, and so no real samples were analysed in this study. After the preparation of the samples, ATR-FT-IR spectra were measured and pre-treated using baseline correction, MSC, SNV, and Savitzky–Golay smoothing. After this, PCA and factor analysis were performed to reduce the dimensionality of the data, i.e., the latent variables obtained with both methods were used for modelling. Classification models were calculated using tree-based modelling approaches, Bayes discriminant analysis, and SVM. In the end, the best model was obtained using Bayes discriminant analysis and PCA for dimensionality reduction with correct classification rates for the three classes above 95%. Deconinck et al. [114] proposed a strategy based on a combination of mid-IR and NIR to discriminate among cocaine, amphetamine, ketamine, and samples containing other substances. Therefore, they created first PLS-DA models with the MID-IR spectra pre-treated with SNV and with the NIR spectra pre-treated with SNV, followed by the first derivative. The best model was obtained with the mid-IR spectra with a correct classification rate of 93.1% for an external test set. Further, distinct PLS models were built for each of the three targeted components. Here, the best results were obtained with the NIR spectra. For amphetamine, the best model used a log10 transformation as pre-treatment for the spectra and showed a correlation of 0.96 with the results obtained with the reference method and a prediction error of 2.81. For cocaine, SNV and the first derivative were used on the NIR spectra and resulted in a model with a correlation of 0.89 and a prediction error of 5.78. For ketamine, NIR with SNV correction resulted in the optimal PLS model with a correlation of 0.99 and a prediction error of 4.62. The proposed strategy can be summarised as follows: measure the mid-IR spectrum and classify according to the PLS-DA model. Samples classified as "other" must be sent to the laboratory for further analysis. For the other samples, the NIR spectrum is measured, and the dosage purity is predicted using the respective PLS models obtained. This approach can also be combined with a similar approach developed for ecstasy tablets [72], allowing the characterisation of MDMA tablets, cocaine, amphetamine, and ketamine samples on site using both mid-IR and NIR.

Also, Raman spectroscopy was used in the context of multiple drug detection. Ryder [115] made use of PCA for the discrimination among self-made solid samples of cocaine, heroin, and MDMA, taking into account their most often encountered cutting agents and diluents. The authors used the first derivate as pre-processing and showed that the selection of 2–3% of the spectral data corresponding to the most intensive peaks related to the studied drugs allowed a clear discrimination of the drugs and this by limiting substantially the calculation times. In a follow-up article [27], the same data set was used, but this time, the first derivative was compared to an automated polynomial method [116] as pre-processing. The authors showed that the polynomial method had performances equal to the first derivative both in discrimination between the three drugs using PCA as in their quantification with PCR. The authors could obtain correlations between real and predicted values higher than 0.9418 for all three drugs. However, results were only validated using cross-validation, which is a weakness of this study. The modelling results between both pre-processing strategies were equal, although the authors emphasised that the polynomial approach allowed to retain the shape of the original spectrum, rendering interpretability of the models easier.

4 Discussion

From the literature mentioned earlier, it can be seen clearly that the use of chemometrics in combination with spectroscopic techniques can resolve an important part of the limitations of spectroscopic instruments, especially in data interpretation. Indeed, where a simple comparison to libraries is impacted by the presence of interfering compounds, chemometrics may allow to extract information from the sample spectrum to identify the compound of interest. The more, chemometrics allows to objectify the decisions on identification and to extract (semi)-quantitative information from the spectra.

Although an extensive series of applications proves the utility of chemometrics and the fact that vendor's software often has some kind of chemometric tools included, the integration in routine analysis of illicit substances is often very limited. This is in contrast with pharmaceutical analysis, where chemometrics is a necessary step not only in process analytical technology but also in authenticity checking of suspected products using spectroscopy [38,117,118]. The reason for this limited use could be the fact that chemometrics is perceived as being very

complex and personnel, especially of drug checking services, is not trained to deal with chemometrics. Therefore, in general, spectroscopy is only used for qualitative analysis, using spectral libraries. Due to the enormous gain that can be achieved in the context of drug checking and forensic analysis, there is a need for accessible trainings, focussed on the implementation of chemometrics, rather than on theory. Users of spectroscopy should start to think of chemometrics as another analytical tool, which can be learned and used, without knowing all the details of the algorithms behind it. Also, the trainings by vendors at the acquisition of a new instrument should include a training on the possibilities of the chemometric tools offered by their software and give an indication on the possibilities of exporting data to different commercial chemometric software.

A disadvantage of chemometric modelling is the necessity of having a sample set, representing the real-life situation and so the variability that may occur. This is a shortcoming in the majority of the papers cited earlier and also the problem in classic library search, since here only acceptable matches can be found if the library contains representative spectra for the unknown. The problem with chemometric models is similar, though these models allow the extraction of data from the spectrum, linked to the active ingredient, reducing the effect of mixtures in which illicit drugs are often present [43]. Generally, models, described in the literature, are calculated using only limited sample sets, and if large sample sets are used, they are often quite homogeneous with a narrow range of excipients and adulterants taken into account [119]. Some articles tried to solve these problems by using self-made preparations containing the drug(s) of interest in different concentrations in mixtures with the most often encountered cutting agents and adulterants [68,77,113,115]. For sure this can solve the problem of having too few samples to create significant models, though it simplifies the problem and may result in good-performing models, failing in real-life situations. Therefore, models should always be build using real-life samples, analysed through a classical analytical pathway (e.g., chromatography). If this is not possible at least, the models should be validated using real-life samples for both qualitative and quantitative models. Validation of chemometrics is always an important aspect. Ideally always an external test set of samples should be used. This test set can consist of an independent sample set or can be selected from the original data/sample set using dedicated algorithms, where it is always important that the test set selection is performed in a randomised and objective way. Manual selection could lead to a bias in the selection. When sample sets

are small, cross-validation could give an indication of the performance of the model, but in that case, the model cannot be considered validated for application. In fact, as said earlier, for routine implementation of chemometric models, the models should be based on a large sample set, and so allowing the selection of an external test set. Also, due to the context of illicit drugs, products evolve. This means new adulterants and new excipients may occur, representing new variations and interferences, not covered by the models. Therefore, it is necessary that the models are regularly and continuously updated with new samples to guarantee their performance and robustness and to make sure they cover the variability in the samples on the market.

Another aspect that is important in the creation of routine applications for the analysis of illicit drug samples through spectroscopy is the selection of the instruments. First, a choice has to be made among mid-IR, NIR, and Raman spectroscopy. Often this choice will be based on the available technique, reports in the literature, and the existence/availability of a library, as well as on historical reasons, e.g., the experience of a lab with a certain technique in other fields. Something that is less explored in the literature is the complementarity of different spectroscopic techniques. Often when different techniques are compared to solve a same problem, they are compared based on the separate results obtained by the chemometric models based on the respective spectra. Mid-IR, NIR, and Raman could give complementary results and so the combined data could give better characterisation of the samples [120]. This can be done by using mid-IR for qualitative purposes and NIR for quantitative as was done by Deconinck et al. [72] or even by combining the spectral data of two or three techniques into one data set [121]. The use of more than one technique will render the analysis more complex as well as the onsite implementation of the developed approaches. Also, here, a compromise will have to be found between complexity, needs, and quality of chemometric models/analytical data.

Once the technique chosen, the choice of the instrument is also important. Different instruments can differ in some technical aspects making them more or less suited for the analysis of certain samples. These differences will have more impact with the increasing complexity of the samples. This was shown by Deidda et al. [87] for the analysis of Δ9-THC in cannabis samples using handheld devices and by Liu et al. for a series of other common drugs [122]. They tested two devices differing in signal-to-noise ratio and light source, and they clearly showed better results with one of the instruments. The differences in the instruments are clearly shown in the analysis of the complex matrices of cannabis samples Nowadays, a

whole range of handheld instruments exist, in different price categories, sizes, spectral ranges, and so on, which can complicate the choice of the most-suited instrument.

Another aspect in this context is that chemometric models are often related to an instrument, meaning that applying a model built with spectral data of one instrument is not immediately transferable to another. This can be problematic when laboratories use different types of handheld devices. In that case, a model has to be created for each instrument separately. Recent research showed that also for the transferability of calibration models chemometrics can offer solutions [123–125].

From the review of the literature, it is clear that a wide range of chemometric approaches can be applied for the same problem. Almost all papers described in this review differ in data pre-treatment, in modelling technique or the combination of both of them. Now it must be clear that presenting one combination of pre-treatment and modelling technique to solve all the classification and regression problems in this context is utopic. The performance of a chemometric approach is largely dependent on the spectral data and the complexity of the samples analysed. Sure, the choice of the approach is partly based on the preferences of the researcher, but often it is based on testing different combinations of pre-processing and modelling techniques to select the best approach. It must also be said that based on the reviewed applications, classical chemometric tools such as PCA, SIMCA, and PLS usually show a very good performance, and more complex chemometrics are therefore often not necessary. Although it can be acknowledged that some applications are used to prove the performance/utility of a certain algorithm, first choice should go to classical, more conceivable tools, to increase routine applicability and use by analysts with only a limited background in chemometrics. More complex or alternative tools could then be applied in case the 'standard' techniques are failing. Another aspect in the choice or definition of a chemometric approach is whether or not to use of variable selection. Only few applications, reviewed here, applied variable selection [40,51,52,59,73,76,88,112, 115]. Variable selection or window selection (selection of a limited part or limited parts of the spectrum) can be valuable in certain contexts, though for the routine application in the analysis of illicit samples, this could have some disadvantages. A wide variety of samples is offered at drug checking services and forensic laboratories. These samples can differ both in active compounds as in excipients, cutting agents, and adulterants. Selection of only a limited number of wavelengths or ranges of wavelengths could result in bad prediction when excipients, cutting agents, or adulterants occur, not taken into account during the selection of the wavelengths. In this case, taking the whole spectrum, except the uncharacteristic regions, e.g., related to water, should be the better option for the qualitative analysis. Once the sample is identified, variable selection could have a positive impact on the regression model (quantitative), and here, it should be kept in mind that interference, not taken into account, during model calculations is possible.

Variable or window selection also complicates untargeted analysis, since often those variables/windows will be selected for targeted compounds. In routine applications, it is important to have models able to screen for multiple components or even complete untargeted screening, even if the latter will be very difficult using spectroscopy and chemometrics. The option of using latent variables calculated with PCA seems a better alternative to reduce noise and data input in the models, since here the importance of some regions in the spectra is diminished, but not eliminated. So, in short, from the perspective of the application in routine drug checking or forensic analysis, the use of variable and window selection would not be recommended. However, once samples are identified and quantitative models are created based on samples of the same 'type' variable and window selection could be beneficial, being always careful about new, emerging, and unexpected interfering compounds or matrices.

Concerning Raman spectroscopy, several papers pointed at the limited sensitivity and the high limits of detection as a major disadvantage of this technique. This means Raman spectroscopy can be very useful in screening for illicit drugs, but only if the molecule is present in significant amounts. The use of SERS is one possibility to circumvent this disadvantage. As said earlier, SERS allows to increase sensitivity by enhancing the Raman signal by several orders of magnitude. However, the number of applications for drug products [26,106,108,126,127], described in the literature, is very limited at this time, and the applications focus more on the analysis of illicit drugs in biological fluids, tissues, etc. [106,128–132]. This seems logical since here the question of sensitivity is even more important. In drug products, one may ask if sensitivity is a major issue, since low concentration of known illicit drugs will often not have detrimental effects from a public health of view, except for NPS and of course fentanyl analogues. Applications combining SERS with chemometrics are very scares and here there is an application field to be studied. It has to be investigated if chemometrics can help in solving some of the disadvantages of SERS and perhaps rendering it more interesting in the context of drug checking and on-site forensic analysis. It has to be clear that the

implementation of SERS as a mobile analytical tool is not that straightforward as classical Raman spectroscopy, although portable approaches are emerging [133,134].

To conclude this discussion of the literature, it should be noted that the combination of spectroscopy and chemometrics can (and was sporadically) also be used in other applications concerning illicit drugs. Examples are the detection of illicit drugs in beverages [120,135–137], in biological fluids [120], on fingerprints and fingernails [120], on banknotes [120,138] and on textile [120,138,139]. Since these applications will not be performed in the context of drug checking or mobile forensic analysis, they were considered out of scope for this review, but they are certainly of equal importance and required for supporting the monitoring of rapidly changing trends in drug use and drug market. It demonstrates once again the power of the combination and makes this still an interesting field in analytical chemistry.

5 Conclusion

With this review, it was the intention to demonstrate the possibilities of chemometrics in spectroscopy for the analysis of illicit drug products, especially in the context of drug checking and mobile forensic analysis. Although this combination was explored regularly in the literature and clearly showed advantages, chemometrics does not often find its way to routine applications for several reasons. One of the major issues is that the applications mentioned earlier are developed in research laboratories focussing on one type of illicit drug samples or one family of illicit drugs. In real life, drug checking services and mobile forensic analysts encounter a whole series of samples in different forms: powders, tablets, and liquids often without indication of the identity of the product. The continuous emergence of NPS and masking compounds even complicates the situation. As a consequence, to be applicable in routine, there is a need of chemometric models, based on spectral data and representative samples of the retail drug market, able to identify multiple drug compounds. It is utopic to believe that a complete untargeted screening could be possible using only spectroscopy. If spectroscopic techniques with the help of chemometrics could be able to identify and characterise the most encountered drugs and on the other hand identify those samples that do not fit the models, money and time could be saved, since only the samples classified as "outlier to the model" should be send to the laboratory for more thorough screening and analysis, using chromatography and MS.

References

[1]　Council of Europe. EU Drugs Strategy 2021–2025 – Introduction – Aim, foundations and approach. Off J Eur Union; 2021.

[2]　Council of Europe. EU Drugs Action Plan 2021–2025. Off J Eur Union; 2021.

[3]　Belgisch staatsblad, Belgium, Brussels. Koninklijk besluit houdende regeling van verdovende middelen en psychotrope stoffen; 2017.

[4]　Bao Y, Meng S, Shi J, Lu L. Control of fentanyl-related substances in China. Lancet Psychiatry. 2019;6:e15. doi: 10.1016/S2215-0366(19)30218-4.

[5]　Fregonese M, Albino A, Covino C, Gili A, Bacci M, Nicoletti A, et al. Drug checking as strategy for harm reduction in recreational contests: Evaluation of two different drug analysis methodologies. Front Psychiatry. 2021;12:596895. doi: 10.3389/fpsyt.2021.596895.

[6]　Kammersgaard T. Harm reduction policing: From drug law enforcement to protection. Contemp Drug Probl. 2019;46:345–62. doi: 10.1177/0091450919871313.

[7]　Gerace E, Seganti F, Luciano C, Lombardo T, Di Corcia D, Teifel H, et al. On-site identification of psychoactive drugs by portable Raman spectroscopy during drug-checking service in electronic music events: Drug checking by portable Raman instrument. Drug Alcohol Rev. 2019;38:50–6. doi: 10.1111/dar.12887.

[8]　Giné CV, Vilamala MV, Measham F, Brunt TM, Bücheli A, Paulos C, et al. The utility of drug checking services as monitoring tools and more: A response to Pirona et al. Int J Drug Policy. 2017;45:46–7. doi: 10.1016/j.drugpo.2017.05.018.

[9]　Goncalves R, Titier K, Latour V, Peyré A, Castaing N, Daveluy A, et al. Suitability of infrared spectroscopy for drug checking in harm reduction centres. Int J Drug Policy. 2021;88:103037. doi: 10.1016/j.drugpo.2020.103037.

[10]　McCrae K, Tobias S, Tupper K, Arredondo J, Henry B, Mema S, et al. Drug checking services at music festivals and events in a Canadian setting. Drug Alcohol Depend. 2019;205:107589. doi: 10.1016/j.drugalcdep.2019.107589.

[11]　Tupper KW, McCrae K, Garber I, Lysyshyn M, Wood E. Initial results of a drug checking pilot program to detect fentanyl

adulteration in a Canadian setting. Drug Alcohol Depend. 2018;190:242–5. doi: 10.1016/j.drugalcdep.2018.06.020.

[12] Harper L, Powell J, Pijl EM. An overview of forensic drug testing methods and their suitability for harm reduction point-of-care services. Harm Reduct J. 2017;14:52. doi: 10.1186/s12954-017-0179-5.

[13] Maghsoudi N, McDonald K, Stefan C, Beriault DR, Mason K, Barnaby L, et al. Evaluating networked drug checking services in Toronto, Ontario: Study protocol and rationale. Harm Reduct J. 2020;17:9. doi: 10.1186/s12954-019-0336-0.

[14] Barratt MJ, Kowalski M, Maier LJ, Ritter A. Global review of drug checking services operating in 2017. Drug Policy Model Program Bull. 2018;24:1–23.

[15] Guirguis A, Moosa I, Gittins R, Schifano F. What about drug checking? systematic review and netnographic analysis of social media. Curr Neuropharmacol. 2020;18:906–17. doi: 10.2174/1570159X18666200413142632.

[16] Wallace B, Hills R, Rothwell J, Kumar D, Garber I, Roode T, et al. Implementing an integrated multi-technology platform for drug checking: Social, scientific, and technological considerations. Drug Test Anal. 2021;13:734–46. doi: 10.1002/dta.3022.

[17] Johnson CS, Stansfield CR, Hassan VR. Festival testing: A survey of suspected drugs seized from New Zealand music festivals, December 2018 to March 2019. Forensic Sci Int. 2020;313:110367. doi: 10.1016/j.forsciint.2020.110367.

[18] McCrae K, Tobias S, Grant C, Lysyshyn M, Laing R, Wood E, et al. Assessing the limit of detection of Fourier-transform infrared spectroscopy and immunoassay strips for fentanyl in a real-world setting. Drug Alcohol Rev. 2020;39:98–102. doi: 10.1111/dar.13004.

[19] Massart DL, editor. Handbook of chemometrics and qualimetrics. Amsterdam; New York: Elsevier; 1997.

[20] Mark H, Workman J. Chemometrics in spectroscopy. Amsterdam: Elsevier/Academic Press; 2007.

[21] Workman J, Springsteen AW. Applied spectroscopy: A compact reference for practitioners. Massachusetts: Academic Press; 1998.

[22] Chem.libretext.org. ATR-FTIR; 2020.

[23] Dumont E, De Bleye C, Sacré P-Y, Netchacovitch L, Hubert P, Ziemons E. From near-infrared and Raman to surface-enhanced Raman spectroscopy: Progress, limitations and perspectives in bioanalysis. Bioanalysis. 2016;8:1077–103. doi: 10.4155/bio-2015-0030.

[24] Bumbrah GS, Sharma RM. Raman spectroscopy – Basic principle, instrumentation and selected applications for the characterization of drugs of abuse. Egypt J Forensic Sci. 2016;6:209–15. doi: 10.1016/j.ejfs.2015.06.001.

[25] Wang W, Zhang H, Yuan Y, Guo Y, He S. Research progress of Raman spectroscopy in drug analysis. AAPS PharmSciTech. 2018;19:2921–8. doi: 10.1208/s12249-018-1135-8.

[26] Smith M, Logan M, Bazley M, Blanchfield J, Stokes R, Blanco A, et al. A Semi-quantitative method for the detection of fentanyl using surface-enhanced Raman scattering (SERS) with a handheld Raman instrument. J Forensic Sci. 2021;66:505–19. doi: 10.1111/1556-4029.14610.

[27] Leger MN, Ryder AG. Comparison of derivative preprocessing and automated polynomial baseline correction method for classification and quantification of narcotics in solid mixtures. Appl Spectrosc. 2006;60:182–93. doi: 10.1366/000370206776023304.

[28] Custers D, Courselle P, Apers S, Deconinck E. Chemometrical analysis of fingerprints for the detection of counterfeit and falsified medicines. Rev Anal Chem. 2016;35:145–68. doi: 10.1515/revac-2016-0013.

[29] Hamsici OC, Martinez AM. Bayes optimality in linear discriminant analysis. IEEE Trans Pattern Anal Mach Intell. 2008;30:647–57. doi: 10.1109/TPAMI.2007.70717.

[30] Brown SD, Tauler i Ferré R, Walczak B. Comprehensive chemometrics: Chemical and biochemical data analysis. Amsterdam: Elsevier (Firm); 2020.

[31] Marini F. Orthogonal PLS (O-PLS) and related algorithms. J Chemom. 2020;34:e3214. doi: 10.1002/cem.3214.

[32] Deconinck E, Sacré PY, Courselle P, De Beer JO. Chemometrics and chromatographic fingerprints to discriminate and classify counterfeit medicines containing PDE-5 inhibitors. Talanta. 2012;100:123–33. doi: 10.1016/j.talanta.2012.08.029.

[33] Luts J, Ojeda F, Van de Plas R, De Moor B, Van Huffel S, Suykens JAK. A tutorial on support vector machine-based methods for classification problems in chemometrics. Anal Chim Acta. 2010;665:129–45. doi: 10.1016/j.aca.2010.03.030.

[34] Suykens J. Kernel methods. Comprehensive chemometrics. Amsterdam: Elsevier; 2009. p. 555–66. doi: 10.1016/B978-0-444-64165-6.02021-8.

[35] Lavine BK, Blank TR. Feed-forward neural networks. Comprehensive chemometrics. Amsterdam: Elsevier; 2009. p. 571–86. doi: 10.1016/B978-044452701-1.00026-0.

[36] Marini F. Neural networks. Comprehensive chemometrics. Amsterdam: Elsevier; 2009. p. 477–505. doi: 10.1016/B978-044452701-1.00128-9.

[37] Rebiere H, Martin M, Ghyselinck C, Bonnet P-A, Brenier C. Raman chemical imaging for spectroscopic screening and direct quantification of falsified drugs. J Pharm Biomed Anal. 2018;148:316–23. doi: 10.1016/j.jpba.2017.10.005.

[38] Rebiere H, Ghyselinck C, Lempereur L, Brenier C. Investigation of the composition of anabolic tablets using near infrared spectroscopy and Raman chemical imaging. Drug Test Anal. 2016;8:370–7. doi: 10.1002/dta.1843.

[39] Rutan SC, de Juan A, Tauler R. Introduction to multivariate curve resolution. Comprehensive chemometrics. Amsterdam: Elsevier; 2009. p. 249–59. doi: 10.1016/B978-044452701-1.00046-6.

[40] Anzanello MJ, Yamashita G, Marcelo M, Fogliatto FS, Ortiz RS, Mariotti K, et al. A genetic algorithm-based framework for wavelength selection on sample categorization: Wavelength selection on forensic sample categorization. Drug Test Anal. 2017;9:1172–81. doi: 10.1002/dta.2138.

[41] Lavine BK, Davidson CE. Multivariate approaches to classification using genetic algorithms. Comprehensive chemometrics. Amsterdam: Elsevier; 2009. p. 619–46. doi: 10.1016/B978-044452701-1.00032-6.

[42] Penido CAFO, Pacheco MTT, Zângaro RA, Silveira L. Identification of different forms of cocaine and substances used in adulteration using near-infrared Raman spectroscopy and infrared absorption spectroscopy. J Forensic Sci. 2015;60:171–8. doi: 10.1111/1556-4029.12666.

[43] Mainali D, Seelenbinder J. Automated fast screening method for cocaine identification in seized drug samples using a portable fourier transform infrared (FT-IR) instrument. Appl

Spectrosc. 2016;70:916–22. doi: 10.1177/0003702816638305.

[44] Eliaerts J, Meert N, Van Durme F, Samyn N, De Wael K, Dardenne P. Practical tool for sampling and fast analysis of large cocaine seizures. Drug Test Anal. 2018;10:1039–42. doi: 10.1002/dta.2364.

[45] Pérez-Alfonso C, Galipienso N, Garrigues S, de la Guardia M. A green method for the determination of cocaine in illicit samples. Forensic Sci Int. 2014;237:70–7. doi: 10.1016/j.forsciint.2014.01.015.

[46] Eliaerts J, Dardenne P, Meert N, Van Durme F, Samyn N, Janssens K, et al. Rapid classification and quantification of cocaine in seized powders with ATR-FTIR and chemometrics: ATR-FTIR with chemometrics for rapid classification and quantification of cocaine in seized powders. Drug Test Anal. 2017;9:1480–9. doi: 10.1002/dta.2149.

[47] Eliaerts J, Meert N, Dardenne P, Baeten V, Pierna J-AF, Van Durme F, et al. Comparison of spectroscopic techniques combined with chemometrics for cocaine powder analysis. J Anal Toxicol. 2020;44:851–60. doi: 10.1093/jat/bkaa101.

[48] Eliaerts J, Meert N, Van Durme F, Dardenne P, Charles S, De Wael K, et al. Challenges for cocaine detection in smuggling samples. Forensic Sci Int. 2021;319:110534. doi: 10.1016/j.forsciint.2020.110534.

[49] Kranenburg RF, Verduin J, Weesepoel Y, Alewijn M, Heerschop M, Koomen G, et al. Rapid and robust on-scene detection of cocaine in street samples using a handheld near-infrared spectrometer and machine learning algorithms. Drug Test Anal. 2020;12:1404–18. doi: 10.1002/dta.2895.

[50] Kranenburg RF, Verduin J, Ridder R, Weesepoel Y, Alewijn M, Heerschop M, et al. Performance evaluation of handheld Raman spectroscopy for cocaine detection in forensic case samples. Drug Test Anal. 2021;13:1054–67. doi: 10.1002/dta.2993.

[51] da Silva AF, Grobério TS, Zacca JJ, Maldaner AO, Braga JWB. Cocaine and adulterants analysis in seized drug samples by infrared spectroscopy and MCR-ALS. Forensic Sci Int. 2018;290:169–77. doi: 10.1016/j.forsciint.2018.07.006.

[52] Anzanello MJ, Kahmann A, Marcelo MCA, Mariotti KC, Ferrão MF, Ortiz RS. Multicriteria wavenumber selection in cocaine classification. J Pharm Biomed Anal. 2015;115:562–9. doi: 10.1016/j.jpba.2015.08.008.

[53] Marcelo MCA, Mariotti KC, Ferrão MF, Ortiz RS. Profiling cocaine by ATR–FTIR. Forensic Sci Int. 2015;246:65–71. doi: 10.1016/j.forsciint.2014.11.011.

[54] Rodrigues NVS, Cardoso EM, Andrade MVO, Donnici CL, Sena MM. Analysis of seized cocaine samples by using chemometric methods and FTIR spectroscopy. J Braz Chem Soc. 2013;24:507–17. doi: 10.5935/0103-5053.20130066.

[55] Grobério TS, Zacca JJ, Talhavini M, Braga JWB. Quantification of cocaine hydrochloride in seized drug samples by infrared spectroscopy and PLSR. J Braz Chem Soc. 2014. doi: 10.5935/0103-5053.20140164.

[56] Grobério TS, Zacca JJ, Botelho ÉD, Talhavini M, Braga JWB. Discrimination and quantification of cocaine and adulterants in seized drug samples by infrared spectroscopy and PLSR. Forensic Sci Int. 2015;257:297–306. doi: 10.1016/j.forsciint.2015.09.012.

[57] Materazzi S, Gregori A, Ripani L, Apriceno A, Risoluti R. Cocaine profiling: Implementation of a predictive model by ATR-FTIR coupled with chemometrics in forensic chemistry. Talanta. 2017;166:328–35. doi: 10.1016/j.talanta.2017.01.045.

[58] Hespanhol MC, Pasquini C, Maldaner AO. Evaluation of a low-cost portable near-infrared spectrophotometer for *in situ* cocaine profiling. Talanta. 2019;200:553–61. doi: 10.1016/j.talanta.2019.03.091.

[59] Coppey F, Bécue A, Sacré P-Y, Ziemons EM, Hubert P, Esseiva P. Providing illicit drugs results in five seconds using ultra-portable NIR technology: An opportunity for forensic laboratories to cope with the trend toward the decentralization of forensic capabilities. Forensic Sci Int. 2020;317:110498. doi: 10.1016/j.forsciint.2020.110498.

[60] EMCCDA. European drug report 2021: Trends and developments; 2021.

[61] Praisler M. Identification of novel illicit amphetamines from vapor-phase FTIR spectra – a chemometrical solution. Talanta. 2000;53:155–70. doi: 10.1016/S0039-9140(00)00461-6.

[62] Katainen E, Elomaa M, Laakkonen U-M, Sippola E, Niemelä P, Suhonen J, et al. Quantification of the amphetamine content in seized street samples by Raman spectroscopy. J Forensic Sci. 2007;52:88–92. doi: 10.1111/j.1556-4029.2006.00306.x.

[63] Praisler M, Dirinck I, Van Bocxlaer JF, De Leenheer AP, Massart DL. Computer-aided screening for hallucinogenic and stimulant amphetamines with gas chromatography-fourier transform infrared spectroscopy (GC-FTIR). J Anal Toxicol. 2001;25:45–56. doi: 10.1093/jat/25.1.45.

[64] Praisler M. Pattern recognition techniques screening for drugs of abuse with gas chromatography–Fourier transform infrared spectroscopy. Talanta. 2000;53:177–93. doi: 10.1016/S0039-9140(00)00460-4.

[65] Praisler M, Van Bocxlaer J, De Leenheer A, Massart DL. Chemometric detection of thermally degraded samples in the analysis of drugs of abuse with gas chromatography–Fourier-transform infrared spectroscopy. J Chromatogr A. 2002;962:161–73. doi: 10.1016/S0021-9673(02)00536-8.

[66] Gosav S, Praisler M, Van Bocxlaer J, De Leenheer AP, Massart DL. Class identity assignment for amphetamines using neural networks and GC–FTIR data. Spectrochim Acta A Mol Biomol Spectrosc. 2006;64:1110–7. doi: 10.1016/j.saa.2005.11.033.

[67] Gosav S, Praisler M, Birsa ML. Principal component analysis coupled with artificial neural networks—a combined technique classifying small molecular structures using a concatenated spectral database. Int J Mol Sci. 2011;12:6668–84. doi: 10.3390/ijms12106668.

[68] Goh CY, Van Bronswijk W, Priddis C. Rapid nondestructive on-site screening of methylamphetamine seizures by attenuated total reflection fourier transform infrared spectroscopy. Appl Spectrosc. 2008;62:640–8. doi: 10.1366/000370208784658002.

[69] Sondermann N, Kovar K-A. Screening experiments of ecstasy street samples using near infrared spectroscopy. Forensic Sci Int. 1999;106:147–56. doi: 10.1016/S0379-0738(99)00195-4.

[70] Schneider RC, Kovar K-A. Analysis of ecstasy tablets: comparison of reflectance and transmittance near infrared spectroscopy. Forensic Sci Int. 2003;134:187–95. doi: 10.1016/S0379-0738(03)00125-7.

[71] Hughes J, Ayoko G, Collett S, Golding G. Rapid quantification of methamphetamine: Using attenuated total reflectance fourier transform infrared spectroscopy (ATR-FTIR) and chemometrics. PLoS One. 2013;8:e69609. doi: 10.1371/journal.pone.0069609.

[72] Deconinck E, Van Campenhout R, Aouadi C, Canfyn M, Bothy JL, Gremeaux L, et al. Combining attenuated total reflectance-infrared spectroscopy and chemometrics for the identification and the dosage estimation of MDMA tablets. Talanta. 2019;195:142–51. doi: 10.1016/j.talanta.2018.11.027.

[73] Moros J, Galipienso N, Vilches R, Garrigues S, Guardia MD. Nondestructive direct determination of heroin in seized illicit street drugs by diffuse reflectance near-infrared spectroscopy. Anal Chem. 2008;80:7257–65. doi: 10.1021/ac800781c.

[74] Stevanović NR, Jovanović M, Marini F, Ražić S. Chemometric approach to a rapid attenuated total reflection fourier transform infrared analysis of complex heroin-based mixtures. Appl Spectrosc. 2021;75:545–55. doi: 10.1177/0003702820969715.

[75] Cai X, Wu G. Preliminary study on identification of heroin from different route with clustering analysis by fourier transform infrared spectroscopy. Guang Pu Xue Yu Guang Pu Fen Xi Guang Pu. 2007;27:2441–4.

[76] Moros J, Kuligowski J, Quintás G, Garrigues S, de la Guardia M. New cut-off criterion for uninformative variable elimination in multivariate calibration of near-infrared spectra for the determination of heroin in illicit street drugs. Anal Chim Acta. 2008;630:150–60. doi: 10.1016/j.aca.2008.10.024.

[77] Melucci D, Monti D, D'Elia M, Luciano G. Rapid *in situ* repeatable analysis of drugs in powder form using reflectance near-infrared spectroscopy and multivariate calibration: Rapid *in situ* repeatable analysis of drugs. J Forensic Sci. 2012;57:86–92. doi: 10.1111/j.1556-4029.2011.01945.x.

[78] He X, Wang J, Niu F, Fan L, Teng X, Zhang C, et al. Characterization of heroin and its additives by attenuated total reflection (ATR) – fourier transform infrared spectroscopy (FTIR) and multivariate analysis. Anal Lett. 2020;53:2656–70. doi: 10.1080/00032719.2020.1751181.

[79] Andre CM, Hausman J-F, Guerriero G. Cannabis sativa: The plant of the thousand and one molecules. Front Plant Sci. 2016;7:19. doi: 10.3389/fpls.2016.00019.

[80] Mead A. The legal status of cannabis (marijuana) and cannabidiol (CBD) under U.S. law. Epilepsy Behav. 2017;70:288–91. doi: 10.1016/j.yebeh.2016.11.021.

[81] European Monitoring Centre for Drugs and Drug Addiction. Cannabis legislation in Europe; 2017.

[82] Farag S, Kayser O. The cannabis plant: Botanical aspects. Handbook of cannabis and related pathologies. Amsterdam: Elsevier; 2017. p. 3–12. doi: 10.1016/B978-0-12-800756-3.00001-6.

[83] Borille BT, Marcelo MCA, Ortiz RS, de Cássia Mariotti K, Ferrão MF, Limberger RP. Near infrared spectroscopy combined with chemometrics for growth stage classification of cannabis cultivated in a greenhouse from seized seeds. Spectrochim Acta A Mol Biomol Spectrosc. 2017;173:318–23. doi: 10.1016/j.saa.2016.09.040.

[84] Sánchez-Carnerero Callado C, Núñez-Sánchez N, Casano S, Ferreiro-Vera C. The potential of near infrared spectroscopy to estimate the content of cannabinoids in Cannabis sativa L.: A comparative study. Talanta. 2018;190:147–57. doi: 10.1016/j.talanta.2018.07.085.

[85] Duchateau C, Kauffmann J, Canfyn M, Stévigny C, De Braekeleer K, Deconinck E. Discrimination of legal and illegal *Cannabis spp.* according to European legislation using near infrared spectroscopy and chemometrics. Drug Test Anal. 2020;12:1309–19. doi: 10.1002/dta.2865.

[86] Pereira JFQ, Pimentel MF, Amigo JM, Honorato RS. Detection and identification of Cannabis sativa L. using near infrared hyperspectral imaging and machine learning methods. A feasibility study. Spectrochim Acta A Mol Biomol Spectrosc. 2020;237:118385. doi: 10.1016/j.saa.2020.118385.

[87] Deidda R, Coppey F, Damergi D, Schelling C, Coïc L, Veuthey J-L, et al. New perspective for the in-field analysis of cannabis samples using handheld near-infrared spectroscopy: A case study focusing on the determination of Δ9-tetrahydrocannabinol. J Pharm Biomed Anal. 2021;202:114150. doi: 10.1016/j.jpba.2021.114150.

[88] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. ArXiv12010490 Cs; 2018.

[89] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet Lond Engl. 1986;1:307–10.

[90] Geskovski N, Stefkov G, Gigopulu O, Stefov S, Huck CW, Makreski P. Mid-infrared spectroscopy as process analytical technology tool for estimation of THC and CBD content in Cannabis flowers and extracts. Spectrochim Acta A Mol Biomol Spectrosc. 2021;251:119422. doi: 10.1016/j.saa.2020.119422.

[91] Sanchez L, Filter C, Baltensperger D, Kurouski D. Confirmatory non-invasive and non-destructive differentiation between hemp and cannabis using a hand-held Raman spectrometer. RSC Adv. 2020;10:3212–6. doi: 10.1039/C9RA08225E.

[92] Sanchez L, Baltensperger D, Kurouski D. Raman-based differentiation of hemp, cannabidiol-rich hemp, and Cannabis. Anal Chem. 2020;92:7733–7. doi: 10.1021/acs.analchem.0c00828.

[93] Pirona A, Bo A, Hedrich D, Ferri M, van Gelder N, Giraudon I, et al. New psychoactive substances: Current health-related practices and challenges in responding to use and harms in Europe. Int J Drug Policy. 2017;40:84–92. doi: 10.1016/j.drugpo.2016.10.004.

[94] Sharma KK, Hales TG, Rao VJ, NicDaeid N, McKenzie C. The search for the "next" euphoric non-fentanil novel synthetic opioids on the illicit drugs market: current status and horizon scanning. Forensic Toxicol. 2019;37:1–16. doi: 10.1007/s11419-018-0454-5.

[95] Karch L, Tobias S, Schmidt C, Doe-Simkins M, Carter N, Salisbury-Afshar E, et al. Results from a mobile drug checking pilot program using three technologies in Chicago, IL, USA. Drug Alcohol Depend. 2021;228:108976. doi: 10.1016/j.drugalcdep.2021.108976.

[96] Ti L, Tobias S, Lysyshyn M, Laing R, Nosova E, Choi J, et al. Detecting fentanyl using point-of-care drug checking technologies: A validation study. Drug Alcohol Depend. 2020;212:108006. doi: 10.1016/j.drugalcdep.2020.108006.

[97] Blanckaert P, Balcaen M, Vanhee C, Risseeuw M, Canfyn M, Desmedt B, et al. Analytical characterization of "etonitazepyne," a new pyrrolidinyl-containing 2-benzylbenzimidazole opioid sold online. Drug Test Anal. 2021;13:1627–34. doi: 10.1002/dta.3113.

[98] Piorunska-Sedlak K, Stypulkowska K. Strategy for identification of new psychoactive substances in illicit samples using attenuated total reflectance infrared spectroscopy. Forensic Sci Int. 2020;312:110262. doi: 10.1016/j.forsciint.2020.110262.

[99] Coelho Neto J. Rapid detection of NBOME's and other NPS on blotter papers by direct ATR-FTIR spectrometry. Forensic Sci Int. 2015;252:87–92. doi: 10.1016/j.forsciint.2015.04.025.

[100] Custódio M, Magalhães L, Arantes L, Braga J. Identification of synthetic drugs on seized blotter papers using ATR-FTIR and PLS-DA: Routine application in a forensic laboratory. J Braz Chem Soc. 2021;32:513–22. doi: 10.21577/0103-5053.20200205.

[101] de Oliveira Magalhães L, Arantes LC, Braga JWB. Identification of NBOMe and NBOH in blotter papers using a handheld NIR spectrometer and chemometric methods. Microchem J. 2019;144:151–8. doi: 10.1016/j.microc.2018.08.051.

[102] Xu M, Wang C-H, Terracciano AC, Masunov AE, Vasu SS. High accuracy machine learning identification of fentanyl-relevant molecular compound classification via constituent functional group analysis. Sci Rep. 2020;10:13569. doi: 10.1038/s41598-020-70471-7.

[103] Tobias S, Grant CJ, Laing R, Arredondo J, Lysyshyn M, Buxton J, et al. Time-series analysis of fentanyl concentration in the unregulated opioid drug supply in a canadian setting. Am J Epidemiol. 2022;191:241–7. doi: 10.1093/aje/kwab129.

[104] de Castro JS, Rodrigues CHP, Bruni AT. *In silico* infrared characterization of synthetic cannabinoids by quantum chemistry and chemometrics. J Chem Inf Model. 2020;60:2100–14. doi: 10.1021/acs.jcim.9b00871.

[105] Risoluti R, Materazzi S, Gregori A, Ripani L. Early detection of emerging street drugs by near infrared spectroscopy and chemometrics. Talanta. 2016;153:407–13. doi: 10.1016/j.talanta.2016.02.044.

[106] Muhamadali H, Watt A, Xu Y, Chisanga M, Subaihi A, Jones C, et al. Rapid detection and quantification of novel psychoactive substances (NPS) using Raman spectroscopy and surface-enhanced Raman scattering. Front Chem. 2019;7:412. doi: 10.3389/fchem.2019.00412.

[107] Wang K, Xu B, Wu J, Zhu Y, Guo L, Xie J. Elucidating fentanyls differentiation from morphines in chemical and biological samples with surface-enhanced Raman spectroscopy. Electrophoresis. 2019;40:2193–203. doi: 10.1002/elps.201900004.

[108] Mirsafavi R, Moskovits M, Meinhart C. Detection and classification of fentanyl and its precursors by surface-enhanced Raman spectroscopy. Analyst. 2020;145:3440–6. doi: 10.1039/C9AN02568E.

[109] Wang L, Vendrell-Dones MO, Deriu C, Doğruer S, de B, Harrington P, et al. Multivariate analysis aided surface-enhanced raman spectroscopy (MVA-SERS) multiplex quantitative detection of trace fentanyl in illicit drug mixtures using a handheld Raman spectrometer. Appl Spectrosc. 2021;75:1225–36. doi: 10.1177/00037028211032930.

[110] Cawley A. Non-targeted screening of drugs. Drug Test Anal. 2021;13:892–3. doi: 10.1002/dta.3053.

[111] Liu C, Han Y, Min S, Jia W, Meng X, Liu P. Rapid qualitative and quantitative analysis of methamphetamine, ketamine, heroin, and cocaine by near-infrared spectroscopy. Forensic Sci Int. 2018;290:162–8. doi: 10.1016/j.forsciint.2018.07.008.

[112] Pereira LSA, Lisboa FLC, Coelho Neto J, Valladão FN, Sena MM. Screening method for rapid classification of psychoactive substances in illicit tablets using mid infrared spectroscopy and PLS-DA. Forensic Sci Int. 2018;288:227–35. doi: 10.1016/j.forsciint.2018.05.001.

[113] He X, Wang J, You X, Niu F, Fan L, Lv Y. Classification of heroin, methamphetamine, ketamine and their additives by attenuated total reflection-Fourier transform infrared spectroscopy and chemometrics. Spectrochim Acta A Mol Biomol Spectrosc. 2020;241:118665. doi: 10.1016/j.saa.2020.118665.

[114] Deconinck E, Aït-Kaci C, Raes A, Canfyn M, Bothy J, Duchateau C, et al. An infrared spectroscopic approach to characterise white powders, easily applicable in the context of drug checking, drug prevention and on-site analysis. Drug Test Anal. 2021;13:679–93. doi: 10.1002/dta.2973.

[115] Ryder AG. Classification of narcotics in solid mixtures using principal component analysis and Raman spectroscopy. J Forensic Sci. 2002;47:275–84. doi: 10.1520/JFS15244J.

[116] Lieber CA, Mahadevan-Jansen A. Automated method for subtraction of fluorescence from biological Raman spectra. Appl Spectrosc. 2003;57:1363–7. doi: 10.1366/000370203322554518.

[117] Gaião Calixto M, Alves Ramos H, Veríssimo LS, Dantas Alves V, Medeiros D, Alencar Fernandes FH, et al. Trends and application of chemometric pattern recognition techniques in medicinal plants analysis. Crit Rev Anal Chem. 2021;1–13. E-pub ahead of print, doi: 10.1080/10408347.2021.1953370.

[118] Storme-Paris I, Rebiere H, Matoga M, Civade C, Bonnet P-A, Tissier MH, et al. Challenging near infrared spectroscopy discriminating ability for counterfeit pharmaceuticals detection. Anal Chim Acta. 2010;658:163–74. doi: 10.1016/j.aca.2009.11.005.

[119] O'Connell M-L, Ryder AG, Leger MN, Howley T. Qualitative analysis using Raman spectroscopy and chemometrics: A comprehensive model system for narcotics analysis. Appl Spectrosc. 2010;64:1109–21. doi: 10.1366/000370210792973541.

[120] West MJ, Went MJ. Detection of drugs of abuse by Raman spectroscopy. Drug Test Anal. 2011;3:532–8. doi: 10.1002/dta.217.

[121] Rebiere H, Grange Y, Deconinck E, Courselle P, Acevska J, Brezovska K, et al. European fingerprint study on omeprazole drug substances using a multi analytical approach and chemometrics as a tool for the discrimination of manufacturing sources. J Pharm Biomed Anal. 2022;208:114444. doi: 10.1016/j.jpba.2021.114444.

[122] Liu C, He H, Xu L, Hua Z. New qualitative analysis strategy for illicit drugs using Raman spectroscopy and characteristic

peaks method. Drug Test Anal. 2021;13:720–8. doi: 10.1002/dta.2963.

[123] Eliaerts J, Meert N, Dardenne P, Van Durme F, Baeten V, Samyn N, et al. Evaluation of a calibration transfer between a bench top and portable Mid-InfraRed spectrometer for cocaine classification and quantification. Talanta. 2020;209:120481. doi: 10.1016/j.talanta.2019.120481.

[124] Ni L, Chen H, Hong S, Zhang L, Luan S. Near infrared spectral calibration model transfer without standards by screening spectral points with scale invariant feature transform from master samples spectra. Spectrochim Acta A Mol Biomol Spectrosc. 2021;260:119802. doi: 10.1016/j.saa.2021.119802.

[125] Yang J, Lou X, Yang H, Yang H, Liu C, Wu J, et al. Improved calibration transfer between near-Infrared (NIR) spectro-meters using canonical correlation analysis. Anal Lett. 2019;52:2188–202. doi: 10.1080/00032719.2019.1604725.

[126] Mao J, Kang Y, Yu D, Zhou J. Surface-enhanced Raman spectroscopy integrated with aligner mediated cleavage strategy for ultrasensitive and selective detection of methamphetamine. Anal Chim Acta. 2021;1146:124–30. doi: 10.1016/j.aca.2020.12.028.

[127] Tay L-L, Poirier S, Ghaemi A, Hulse J, Wang S. Iodide func-tionalized paper-based SERS sensors for improved detection of narcotics. Front Chem. 2021;9:680556. doi: 10.3389/fchem.2021.680556.

[128] Dong R, Weng S, Yang L, Liu J. Detection and direct readout of drugs in human urine using dynamic surface-enhanced Raman spectroscopy and support vector machines. Anal Chem. 2015;87:2937–44. doi: 10.1021/acs.analchem.5b00137.

[129] Fang W, Zhang B, Han F-Y, Qin Z-N, Feng Y-Q, Hu J-M, et al. On-site and quantitative detection of trace methampheta-mine in urine/serum samples with a surface-enhanced Raman scattering-active microcavity and rapid pretreatment device. Anal Chem. 2020;92:13539–49. doi: 10.1021/acs.analchem.0c03041.

[130] Sivashanmugan K, Squire K, Tan A, Zhao Y, Kraai JA, Rorrer GL, et al. Trace detection of tetrahydrocannabinol in body fluid via surface-enhanced Raman scattering and prin-cipal component analysis. ACS Sens. 2019;4:1109–17. doi: 10.1021/acssensors.9b00476.

[131] Weng S, Dong R, Zhu Z, Zhang D, Zhao J, Huang L, et al. Dynamic surface-enhanced Raman spectroscopy and Chemometric methods for fast detection and intelligent identification of methamphetamine and 3, 4-Methylenedioxy methamphetamine in human urine. Spectrochim Acta A Mol Biomol Spectrosc. 2018;189:1–7. doi: 10.1016/j.saa.2017.08.004.

[132] Yu B, Ge M, Li P, Xie Q, Yang L. Development of surface-enhanced Raman spectroscopy application for determination of illicit drugs: Towards a practical sensor. Talanta. 2019;191:1–10. doi: 10.1016/j.talanta.2018.08.032.

[133] Raveendran J, Docoslis A. Portable surface-enhanced Raman scattering analysis performed with microelectrode-tem-plated silver nanodendrites. Analyst. 2020;145:4467–76. doi: 10.1039/D0AN00484G.

[134] Wang H, Xue Z, Wu Y, Gilmore J, Wang L, Fabris L. Rapid SERS quantification of trace fentanyl laced in recreational drugs with a portable Raman module. Anal Chem. 2021;93:9373–82. doi: 10.1021/acs.analchem.1c00792.

[135] Brewster VL, Edwards HGM, Hargreaves MD, Munshi T. Identification of the date-rape drug GHB and its precursor GBL by Raman spectroscopy. Drug Test Anal. 2009;1:25–31. doi: 10.1002/dta.11.

[136] Eliasson C, Macleod NA, Matousek P. Non-invasive detection of cocaine dissolved in beverages using displaced Raman spectroscopy. Anal Chim Acta. 2008;607:50–3. doi: 10.1016/j.aca.2007.11.023.

[137] Ye J, Wang S, Zhang Y, Li B, Lu M, Qi X, et al. Surface-enhanced shifted excitation Raman difference spectroscopy for trace detection of fentanyl in beverages. Appl Opt. 2021;60:2354. doi: 10.1364/AO.418579.

[138] Ali EMA, Edwards HGM, Scowen IJ. Rapid *in situ* detection of street samples of drugs of abuse on textile substrates using microRaman spectroscopy. Spectrochim Acta A Mol Biomol Spectrosc. 2011;80:2–7. doi: 10.1016/j.saa.2010.11.001.

[139] Pérez-Alfonso C, Galipienso N, Garrigues S, de la Guardia M. Preliminary results on direct quantitative determination of cocaine in impregnated materials by infrared spectroscopy. Microchem J. 2018;143:110–7. doi: 10.1016/j.microc.2018.07.026.