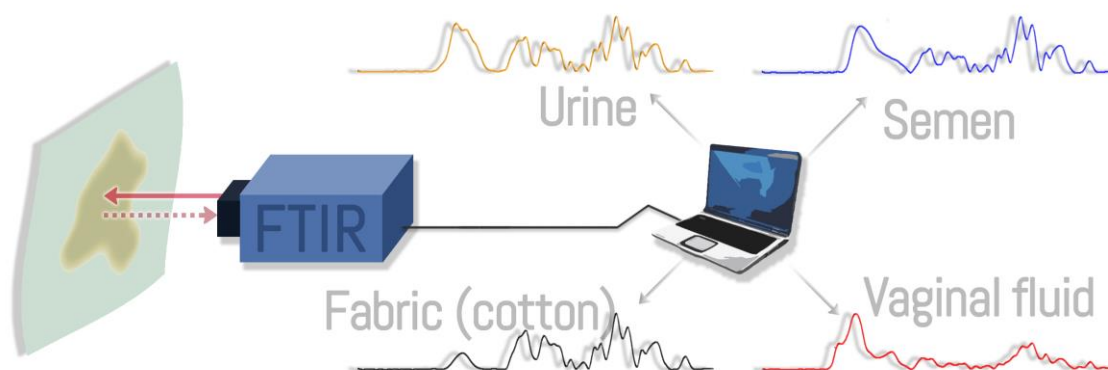


Differentiation of body fluid stains on fabrics using external reflection Fourier transform infrared spectroscopy (FT-IR) and chemometrics

Félix Zapata, M^a Ángeles Fernández de la Ossa, Carmen García-Ruiz*

Department of Analytical Chemistry, Physical Chemistry and Chemical Engineering and University Institute of Research in Police Sciences (IUICP), University of Alcalá, Ctra. Madrid-Barcelona km 33.600, 28871 Alcalá de Henares (Madrid), Spain.

E-mail: carmen.gruiz@uah.es; felix.zapata@uah.es



Cite: F. Zapata, M. A. Fernández de la Ossa, C. García-Ruiz, Differentiation of body fluids stains on fabrics by external reflection Fourier transform infrared (FTIR) spectroscopy and chemometrics, *Appl. Spectrosc.* 70(4) (2016) 654-665. DOI: [10.1177/0003702816631303](https://doi.org/10.1177/0003702816631303)



Abstract

Body fluids are evidence of great forensic interest due to the DNA extracted from them, which allows genetic identification of people. This study focuses on the discrimination among semen, vaginal fluid and urine stains (main fluids in sexual crimes) placed on different coloured cotton fabrics by external reflection FTIR spectroscopy combined with chemometrics. Semen-vaginal fluid mixtures and potential false positive substances commonly found in daily life such as soaps, milk, juices and lotions were also studied. Results demonstrated that the IR spectral signature obtained for each body fluid allowed its identification and the correct classification of unknown stains by means of Principal Components Analysis (PCA) and Soft Independent Modelling of Class Analogy (SIMCA). Interestingly, results proved that these IR spectra did not show any bands due to the colour of the fabric and no substance of those present in daily life which were analysed, provided a false positive.

Keywords: External reflection FTIR spectroscopy; PCA; Semen; SIMCA; Urine; Vaginal fluid.

1. Introduction

Because DNA can be used for individual identification, body fluids have become extremely useful evidence in the forensic field. Stains of body fluids are subjected to different presumptive and confirmatory tests¹ prior to DNA analysis. The study of semen, vaginal fluid and urine stains is particularly crucial because forensic investigators have to face them in sexual abuse cases. In these cases, fabric evidence stained by semen, vaginal fluid or urine are commonly found. Semen has been studied for many years and the available confirmatory tests allow forensic practitioners to identify it in a stain.^{1,2} However, vaginal fluid and urine have been scarcely studied and nowadays they do not have available confirmatory tests. Both fluids usually appear in the victim's underwear mixed with semen, making the identification of DNA from semen traces left by the perpetrator difficult. Consequently, in these cases the genetic profile of the victim may be the only result obtained after performing the DNA extraction and DNA typing, even though the presence of semen in the stain has been verified by a confirmatory test. Therefore, it is a current forensic need to develop a methodology allowing to differentiate among semen, vaginal fluid and urine samples in order to optimize the extraction of the DNA from the perpetrator.

The presumptive and confirmatory tests currently used by forensic crime investigators to identify these body fluid stains have several limitations and disadvantages.^{1,2} Therefore other techniques, mainly spectroscopic techniques, have been evaluated for the identification of body fluids. These techniques present several attractive features as their non-destructive, non-invasive, fast and portable nature and their applicability to analyse multiple body fluids.² Ultraviolet-Visible (UV-Vis) spectroscopy was the first spectroscopic technique applied to the localization of body fluids stains^{3,4} and evolved to the alternate light sources (ALS), routinely used by forensic community. However, ALS lack selectivity, produce false positives⁵ and have limitations on identifying stains on coloured backgrounds.^{6,7}

For these reasons, infrared (IR) wavelengths are being investigated by using IR and Raman spectroscopy in order to evaluate their application to the identification of body fluids. A large part of the studies is focused on analysing body fluids with clinical and medical purposes⁸⁻¹⁵, instead of forensic aims. Although IR and Raman spectroscopy

are accepted techniques at forensic laboratories to analyse evidence such as drugs¹⁶⁻¹⁸ and explosives^{16,19,20}, their application to the analysis of body fluids still remains challenging. Therefore, further investigation related to the identification of body fluids using IR and Raman spectroscopy must be accomplished in order to put it on practice for solving real forensic cases. Up to now, Raman spectroscopy has been mainly researched by Lednev's group²¹. As a summary of all their work, the representative and characteristic Raman spectra from blood, semen, saliva, vaginal fluid and sweat have been established²² and multivariate classification models which correctly classify unknown fluids, even in mixtures, have been developed.^{23,24} Stains of body fluids made on different fabrics such as glass slides, cotton, denim or tile have also been analysed.²⁵ On the contrary, IR spectroscopy applied to the forensic identification of body fluids has been less studied than Raman spectroscopy. Studies reported in literature were mainly focused on analysing blood stains², apart from two of them, which were focused on several body fluids including blood, semen, saliva, vaginal fluid and urine.^{26, 27} In fact, the first one²⁶ was the first study where those five body fluids were analysed by attenuated total reflection (ATR) Fourier Transform infrared spectroscopy (FTIR) and it represented a preliminary research of body fluids by IR spectroscopy. Since only one stain and donor per body fluid were analysed in the reported work, further research using IR spectroscopy is necessary. Moreover, white 100% cotton was the unique fabric where stains were deposited and no chemometric methods were applied in order to extract information from the data.²⁶ Recently, a second study focused on the discrimination of blood, saliva, semen and vaginal fluid by ATR-FTIR has been published.²⁷ Results demonstrate the capability of ATR-FTIR to distinguish these body fluids according to their characteristic FTIR spectrum. However, the study was accomplished by depositing the fresh body fluids on to the ATR stage and analysing them during different time intervals until they were completely dry.²⁷ Therefore, the challenge of identifying stains of these body fluids on real casework materials such as clothes remains unsolved.

Besides ATR-FTIR spectroscopy, external reflection IR spectroscopy might be a useful novel methodology to differentiate stains of body fluids, especially when they are located on the surface of different materials including clothes. External reflection FTIR spectroscopy is also a non-destructive, non-invasive, fast, portable and easy-to-use technique where no sample preparation is required. In addition, external reflection

spectroscopy benefits from the fact that no contact between sample and instrument is required to work, contrary to ATR where the contact sample-ATR crystal is essential. As consequence, external reflection can be potentially used for stand-off detection (at distance), specially the detection of samples which are difficult to be pressed against the ATR crystal. Particularly, stains on clothes and other materials usually offer difficulties because of the formation of folds and creases which hinder a proper contact. Furthermore, the capability of detecting stains at some distance can potentially contribute to evidence recovery. External reflection spectroscopy is due to a combination of diffuse and specular reflection processes since, theoretically, the amount of reflection type occurring on the sample depends on the roughness of the sample surface. In addition, its combination with chemometric analyses such as principal component analysis (PCA) and soft independent modelling of class analogy (SIMCA), is a powerful strategy for spectra differentiation and classification. As consequence, our study focuses on analysing semen, vaginal fluid and urine stains collected from several volunteers to establish the external reflection FTIR spectral signature of stains of these body fluids on white cotton fabrics prior to their detection on coloured cotton pieces from T-shirts and jeans and their differentiation from daily products, which can potentially provide false positives by visual inspection. PCA and SIMCA models are applied to classify the experimental data and show if the spectral interpretation made visually can be corroborated.

2. Experimental

2.1. Samples

The body fluids studied were human semen, vaginal fluid and urine obtained from 10 anonymous donors (5 men, 5 women). Some cosmetic substances and foodstuffs including sunflower oil, cold cream, sunscreen, soap, milk, mayonnaise, lotions, toothpaste, soup, vaseline, yogurt, orange juice and pineapple juice were analysed in order to check potential false positives. Semen-vaginal fluid mixture stains made by adding fresh semen to both fresh and dry vaginal fluid stains, which were then completely dried overnight, were also analysed.

Samples were prepared by placing 0.6 mL of each body fluid (semen and urine), cosmetic substance or foodstuff on pieces of 11x8 cm² cut from T-shirts and jeans made of 100% cotton. Vaginal fluid samples were directly prepared by the donor placing the body fluid on the cotton pieces. In order to know the mass of body fluid which was deposited on the fabric and subsequently analysed in the dry stain, one stain of each body fluid was weighted just after the addition of body fluid and after drying, 24 hours later. In summary, 76 stains (62 from body fluids and 14 from cosmetic substances and foodstuffs) with a diameter of approximately 2-3 cm were created in the middle of the mentioned cotton pieces. Specifically, stains of each body fluid came from five different donors as summarized in Table I.

Table I. Donors and stains distribution on fabrics.

Donor	Stains distribution on fabrics
1	3 stains of semen on white cotton
2	3 stains of semen on white cotton
3	3 stains of semen on white cotton
4	3 stains of semen on white cotton
5	3 stains of semen on white cotton
6	3 stains of vaginal fluid on white cotton
7	3 stains of vaginal fluid on white cotton
8	3 stains of vaginal fluid on white cotton
9	3 stains of vaginal fluid on white cotton
10	3 stains of vaginal fluid on white cotton
1	3 stains of urine on white cotton
2	3 stains of urine on white cotton
6	3 stains of urine on white cotton
7	3 stains of urine on white cotton
8	3 stains of urine on white cotton
1	5 stains of semen on coloured fabrics
6	5 stains of vaginal fluid on coloured fabrics
6	5 stains of urine on coloured fabrics

All stains were made on white cotton fabric except for those which were made on coloured cotton fabrics including red, blue, green, black and blue jeans. These stains were studied in order to test whether coloured fabrics are problematic substrates when identifying stains of body fluids by external reflection IR spectroscopy.

The statistical distribution of all the stains were subsequently visualized by PCA in order to study the differentiation of body fluids even in coloured fabrics. Finally, the

stains were divided in two groups. One of them was used to create the SIMCA classification model which was subsequently applied to the second group used as a test set in order to classify these stains. The first group included stains from 4 donors whereas the second group included stains from the fifth donor. Furthermore, the test set also included two independent semen-vaginal fluid mixtures and 14 stains of cosmetic substances and foodstuffs in order to check their classification by the SIMCA classification model.

All stains were dried completely overnight in a biological safety cabinet before being analysed.

The unstained cotton fabrics (white and coloured) were also analysed to study the cotton response. Likewise IR spectral signature of cotton was achieved and cotton was used in the PCA and SIMCA analyses.

2.2. Instrumentation for MIR spectra collection

IR external reflection measurements were performed on a Bruker ALPHA FT-IR spectrometer (Ettlingen, Germany) using an ALPHA-R external reflection accessory equipped with the OPUS software version 7.2 (Bruker, Billerica, MA). Log (1/R) spectra were recorded from 4000 to 400 cm^{-1} with a resolution of 4 cm^{-1} (data spacing 2.04 cm^{-1}) and signal averaged over 24 accumulations. The 3-term Blackmann-Harris function was applied as an apodization function. The sampling spot, i.e. the diameter of the beam at the sample, was 5 mm. All stains were analysed in several places in order to ensure that the IR spectrum of body fluid was collected. For the stains on white fabrics 10 random points along the stain were measured in order to cover different places along the visibly stained area. On the other hand, after noticing that all points provided almost the same spectra, only 3 random points were analysed for the stains on coloured fabrics which were subsequently studied. All points were measured in triplicate. Data were saved as an OPUS file for further analysis. All spectra were ratioed against the single-beam spectrum measured from a gold-coated mirror. It was cleaned with alcohol wipes between samples. In addition, the transmission spectra of these body fluids were measured by replacing the external reflection accessory with the transmission one in order to know exclusively the spectrum of the body fluid (without cotton) and compare both types of spectra. Stains of these body fluids measured by transmission were

prepared by following the same procedure previously explained, but depositing them on KBr plates instead of cotton. Unfortunately, vaginal fluid was not analysed by transmission because it could not be conveniently deposited on the KBr plate. Therefore, only semen and urine were analysed by transmission FTIR.

2.3. Data treatment

IR spectral measurements of dry stains were processed performing various treatments. In OPUS 7.2 (Bruker, Billerica, MA) a range of 800-2000 cm^{-1} was selected before performing a rubber-band baseline correction²⁸ (n=64 points and 15 iterations) and a range normalization of 0-2 (intensity value). The rubber-band baseline correction consists in finding a convex envelope of the spectrum and subtracting the convex part of the envelope lying below the spectrum from the spectrum in the portion to be corrected²⁸. These treated spectra were imported as a matrix form to The Unscrambler X 10.1 (Oslo, Norway) software. Then, a smoothing task was performed to reduce the random noise, prior to the spectral differentiation of body fluids, using a Savitzky-Golay algorithm (polynomial order 2 and 15 smoothing points in a symmetric kernel). These values were selected because they provided the best results in the chemometric analyses after checking that no relevant spectral information was lost by applying the smoothing.

After the spectral treatment, the IR spectra of cotton, semen, vaginal fluid and urine were studied. Since the main bands of the three body fluids appeared between 1500 and 1800 cm^{-1} , the subsequent multivariate analyses PCA and SIMCA were performed considering this range from the spectra. First, PCA, an unsupervised technique of visualization, was performed using non-linear iterative partial least squares (NIPALS) algorithm, leverage correction and the standardization 1/SDev. Interestingly, the typical standard deviation for the strongest bands in the spectra was 0.08 for cotton fabric, 0.25 for vaginal fluid, 0.07 for semen and 0.13 for urine. In fact, PCA was performed to visualize how similar or different the spectra were and to obtain a first overview of the main sources of variability. PCA allowed the reduction of the matrix dimensions consisting of hundreds of wavenumber values from the spectra into three principal components (PC) without losing relevant information and showing them in graphical plots (scores). The distribution of samples in relation to PCs was based on loadings,

which related wavenumber values (original variables) and new PCs. Both scores and loadings are necessary to explain the correlation between samples and PCs information.

Finally, a SIMCA classification procedure was performed establishing the four classes previously mentioned (cotton, semen, vaginal fluid and urine). This procedure was achieved by calculating separately the PCA of each class in order to use each PCA to define the four classes in the SIMCA model. This process was mandatory since SIMCA is a modelling procedure of discrete classes where a rule which allows discriminating between the inclusion or non-inclusion of a sample in a given class is needed, instead of a rule which discriminates among classes. Therefore, an unknown sample can be classified in one class, in no class or, even, in several classes, because its inclusion in each class is checked separately. In order to perform the SIMCA model it was necessary to use some samples to create the model whereas the remaining samples were considered as test samples. The samples used to create the model (calibrating set) consisted of 6 samples of cotton (one sample from each coloured cotton including the white one), 17 stains of semen, 17 stains of vaginal fluid and 17 stains of urine (12 on white cotton from 4 different donors and 5 stains on each coloured cotton). Note that every stain was analysed in several places obtaining several spectra for each sample. These samples were divided in cotton, semen, vaginal fluid and urine classes and a PCA was calculated for each class by using the same parameters of the global PCA. The number of PCs considered for each class to create SIMCA were five for cotton, three for vaginal fluid, three for urine and four for semen. This choice was selected because they were the number of PCs needed to explain in each class at least the 95% of the information contained in the data. Regarding the test set, those samples consisted of 6 samples of cotton (one from each colour), 3 stains of semen, 3 stains of vaginal fluid, 3 stains of urine, 14 stains of cosmetics and foodstuff and 2 stains of semen-vaginal fluid mixture. These 3 stains of each body fluid, which belonged to a fifth donor, were tested in order to validate the model for identifying stains of semen, vaginal fluid and urine from other people besides the 4 donors considered to create the model. In fact, SIMCA was performed in order to check if the identification of unknown stains (even mixtures) was possible by means of an easy chemometric classification tool where discrete classes were defined by separate PCAs based on the IR spectral signature of each body fluid.

3. Results and discussion

3.1. Determination of the mass of body fluids analysed in dry stains on fabrics.

Two stains per body fluid were weighted before and after drying in order to determine the mass of body fluid analysed by external reflection FTIR spectroscopy prior to perform their spectral characterization. The water content for the three body fluids was also calculated. These results are summarized in Table II. The approximate volume of 0.5 mL of semen and urine deposited to create the stains, weighed around 500 mg. On the contrary, the amount of vaginal fluid deposited on the stains was less controlled because it was impossible to deposit a specific volume. However, by weighting these stains we could know the mass of vaginal fluid deposited in stains of that size, a mass which was in the range of 140-180 mg. After drying the amount of sample weighted in each stain was around 20 mg for urine, 20-35 mg for vaginal fluid and around 41 mg for semen. That depended on the water content of body fluids, which was found to be around 91% for semen, 96% for urine and a wider range from 80% to 86% for vaginal fluid, which notably changed depending on the day of the menstrual cycle.

Table II. Summarized results about the mass of body fluids in stains on fabrics.

Stains	Mass before drying (mg)	Mass after drying (mg)	Water content	Content of non-volatile components
Semen 1	516.2 ± 0.2	42.5 ± 0.2	91.8%	8.2%
Semen 2	459.9 ± 0.2	40.8 ± 0.2	91.1%	8.9%
Urine 1	598.2 ± 0.2	22.3 ± 0.2	96.3%	3.7%
Urine 2	521.8 ± 0.1	16.4 ± 0.1	96.9%	3.1%
Vaginal fluid 1	141.3 ± 0.2	19.8 ± 0.2	86.0%	14.0%
Vaginal fluid 2	178.2 ± 0.2	35.1 ± 0.2	80.3%	19.7%

3.2. IR spectral characterization of semen, vaginal fluid and urine stains on fabrics

In order to determine the IR spectral signature of semen, vaginal fluid and urine stains on white cotton fabrics, a spectral average of the 450 spectra measured for each fluid (which came from considering the 30 spectra from the ten points in each stain along the 15 stains from the 5 donors) was calculated. These spectral signatures are shown in Figure 1. Since the stains of these fluids were made on cotton, the bands of cotton appeared in all spectra. Almost all these bands were located in the interval 800-1500 cm^{-1} . This interval was named "fabric range". As can be seen in Figure 1, specific bands

for semen, vaginal fluid and urine were also observed in the spectra, specifically in the interval 1500-1800 cm^{-1} , which was designated as "ID range". The main bands of cotton as well as the specific bands of urine, semen and vaginal fluid are indicated in Figure 1.

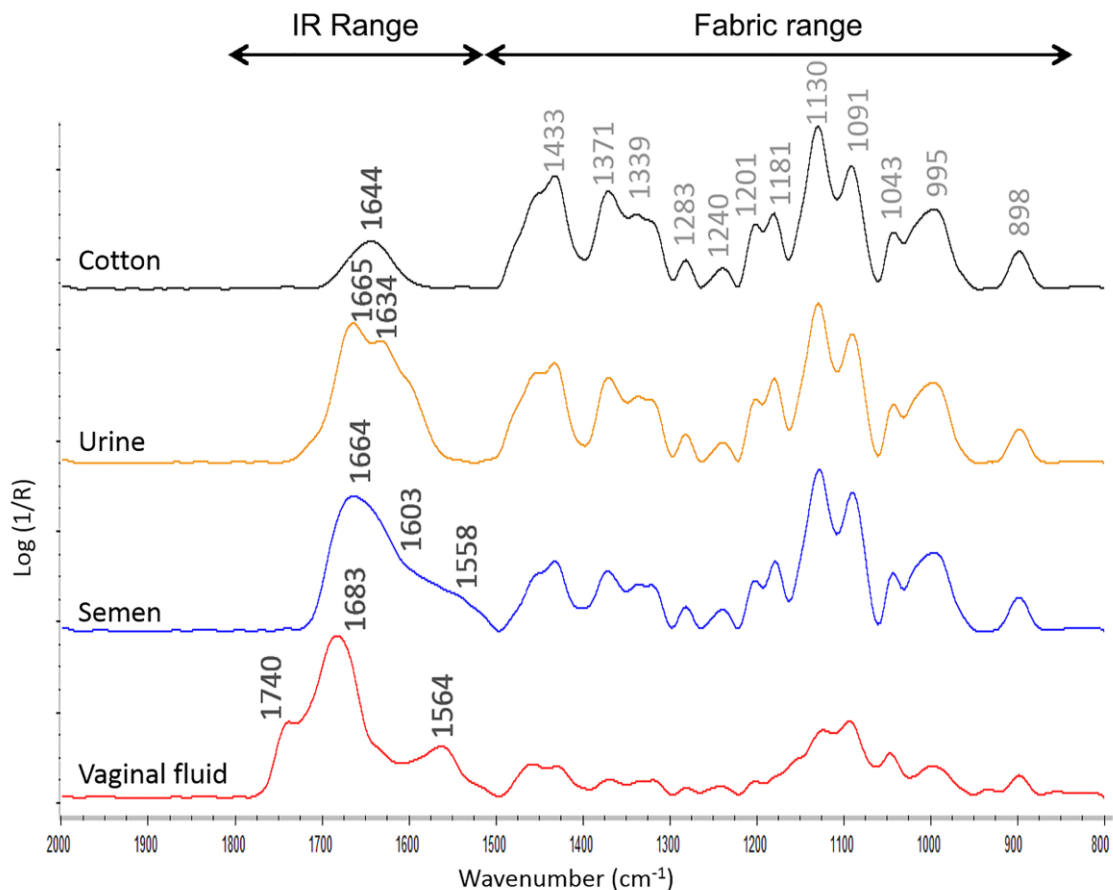


Figure 1. IR spectral signature by using external reflection of cotton, and stains of urine, semen and vaginal fluid on cotton.

The spectrum of cotton was not subtracted from the spectra of semen, vaginal fluid and urine stains because cotton did not supply exactly the same contribution to all stains. Even in the same stain, the contribution of cotton bands to the spectra differed along the stain which it was attributed to the amount of body fluid placed in each part of the stain. For that reason, the indiscriminate subtraction of the cotton contribution was unfeasible.

As can be seen in Figure 1, the main IR bands of cotton appeared at wavenumbers below 1500 cm^{-1} except for one band located at 1644 cm^{-1} which was due to water absorbed by the cotton fabric. Since the spectral region of interest in this work is 1500-1800 cm^{-1} , only that band from cotton above 1500 cm^{-1} was further studied. Although this band from water may interfere at first appearance with the identification of bands

from body fluids, the spectral signatures from body fluids are quite different, more intense and complex. In fact, the spectrum of vaginal fluid stains was significantly different to the cotton spectrum. It presented specific bands located at 1740, 1683 and 1564 cm^{-1} . On the contrary, the cotton contribution to semen and urine spectra was higher than its contribution to vaginal fluid spectrum. In spite of that, characteristic bands were observed for semen and urine. The spectrum of semen showed a wide band (1664, 1603, 1558 cm^{-1}) along the “ID range” whereas characteristic bands of urine were located at 1665 and 1634 cm^{-1} . We think that the reason why bands of cotton are not so evident in vaginal fluid stains, contrary to semen and urine stains, lies in the lower water content of vaginal fluid whose stains hide the cotton fabric to a greater extent than urine and semen stains. Therefore, the signal from cotton in vaginal fluid stains is lower. By considering the water content in each body fluid, we checked that vaginal fluid has the lowest water content, so a higher amount of residues remain on the surface of the dry stain. Moreover, those fluids with higher water content such as semen and urine were soaked up to a greater extent the fabric leaving a lower amount of body fluid on the surface.

Semen and urine were also measured by transmission FTIR spectroscopy after being deposited on a KBr plate and dried. The transmission spectra from semen and urine are shown in Figure S1 (See supporting material).

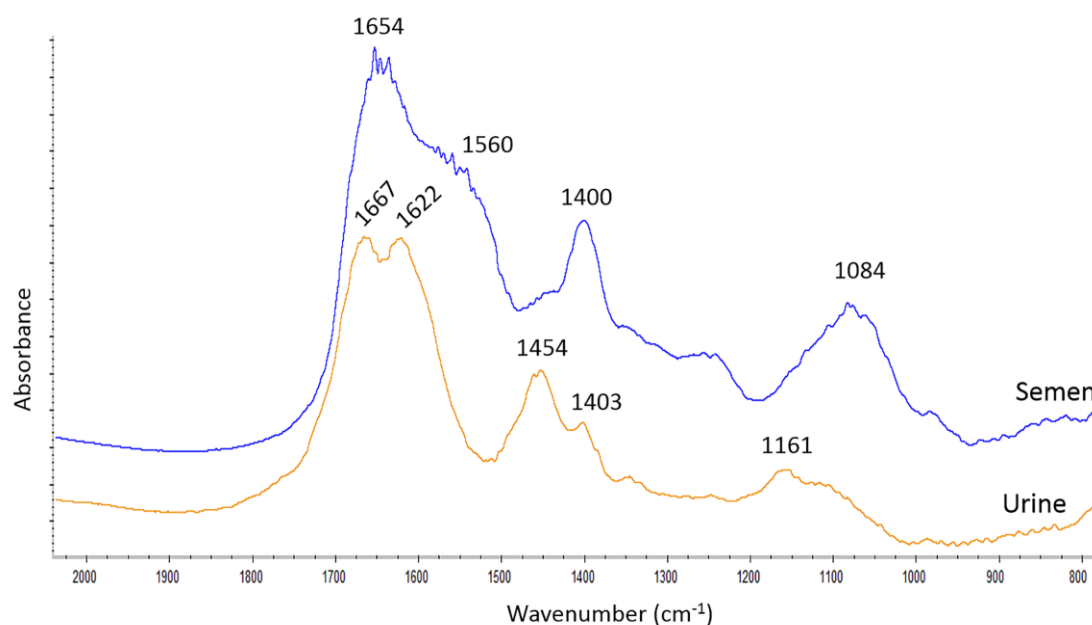


Figure S1 (Supplemental material). Absorbance IR spectra by using transmission mode from dry stains of semen and urine on KBr plates.

It is important to mention that by comparing both types of spectra (external reflection and transmission) we realized that they matched each other along the range 1500-1800 cm^{-1} though small shifts ($\leq 10 \text{ cm}^{-1}$) were found for some bands in semen and urine. On the other hand, the relative intensities of some bands were quite different between both types of spectra as can be seen in urine spectra (Figure 1 vs Figure S1). Interestingly, bands from semen and urine below 1500 cm^{-1} became visible since cotton fabric was not present.

Because of these small shifts between reflection and transmission modes, it was difficult to accomplish an accurate assignment of the bands based on the known spectral ranges found in literature for some functional groups. Therefore, only a general bands interpretation was made. Thus, vaginal fluid bands were probably the result of COO^- (of lactate) vibration and C=O stretching and N-H bending (of the amino acids that make up the proteins such as phosphatase, peptidase and immunoglobulins present in vaginal fluid). Bands from semen probably belonged to C=O stretching and N-H bending vibrations from the amino acids contained in proteins as phosphatase, prostate antigen, semenogelin and immunoglobulins which are in semen. Finally, the main bands in urine probably belonged to urea, the main component in the dried stains of urine, due to C=O stretching and N-H bending vibrations (from NH_2 groups)²⁹. A more detailed study accomplished by Orphanou²⁷ about the composition of body fluids by FTIR spectroscopy demonstrates the large contribution of proteins such as acid phosphatase and albumin to semen spectrum or acid phosphatase and lysozyme to vaginal fluid spectrum. Nevertheless, ATR FTIR spectroscopy was used in that study instead of external reflection. The different interaction between radiation and the sample when using different reflection accessories is probably the reason why we have found some significant differences from our external reflection spectra to those ATR spectra reported by Elkins²⁶ and Orphanou²⁷.

Taking the obtained results into account, a visual inspection of the external reflection spectral signature of the three body fluids allowed their characterization by a few defined spectral bands because the spectral signatures were different among each other even to the naked eye. Consequently, it was possible to differentiate each body fluid stain on fabric by its IR spectrum.

After establishing the IR spectra of the studied body fluids, the stains of these three body fluids on coloured cotton fabrics were subsequently tested. First the coloured cotton fabrics (without stain) were analysed. Fortunately, the spectra from the coloured cotton fabrics were almost identical (Figure 2a). A new band was observed for blue jeans, which seemed to demonstrate that there are slight differences between denim fabric and cotton fabric, in spite of being both 100% cotton. Regarding stains on coloured cotton fabrics, they provided the characteristic signature of the respective body fluid. As an example, the stains of semen deposited on the different coloured fabrics are shown in Figure 2b.

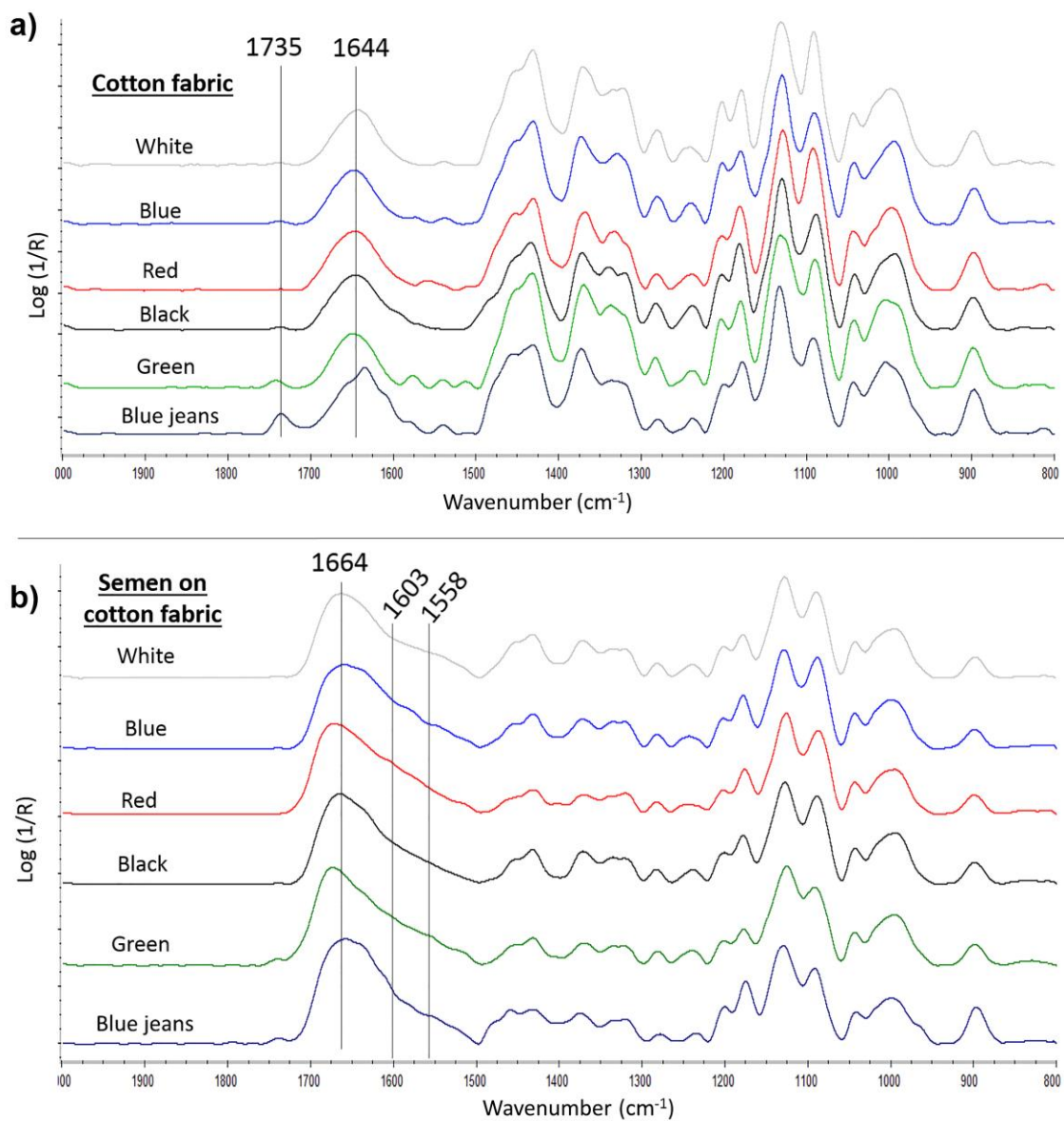


Figure 2. IR spectra of coloured cotton fabrics: white, blue, red, black, green and blue jeans (a), and stains of semen on these coloured cotton fabrics (b).

It is known that most dyes are IR active, so the lack of interference due to colour is probably due to the fact that dyes are present at such low concentration in fabrics that they are undetectable by IR spectroscopy. Therefore IR spectroscopy is suitable to detect a body fluid stain on the fabric but not the dyes from the fabric. This study is especially interesting since it has already been demonstrated that the colour of the fabric presents several disadvantages by UV-Vis spectroscopy^{6,7} but it can be overcome by external reflection FTIR spectroscopy.

3.3. Differentiation of semen, vaginal fluid and urine stains on fabrics

In order to facilitate the discrimination among stains of body fluids by external reflection FTIR spectroscopy even on coloured fabrics, and to check its discrimination capability in mixtures of several body fluids, different multivariate chemometric approaches were performed. These statistic approaches were undergone by considering the “ID range” from spectra, between 1500 and 1800 cm^{-1} .

A global PCA was calculated including all the spectra of stains of body fluids on white and coloured fabrics and the (white and coloured) cotton pieces. Thereby, it was possible to represent all the spectra in a three dimensional plot, where PCs represented the combination of the most discriminative wavenumber values. Figure 3a shows the PCA scores plot from cotton and stains of body fluids on cotton fabrics. PC1, PC2 and PC3 summarized the 95% of the PCA model. On the other hand, Figure 3b shows the scores plot also including the lotions and foodstuffs. Figure 4 shows the loadings plots from PC1, PC2 and PC3. The relationship between PCs and body fluids can be tentatively evaluated by observing Figure 3a and Figure 4 together, and considering the spectral signatures of body fluids previously showed in Figure 1. In particular, PC1 that summarized the 77% of the initial variation seemed to mainly differentiate vaginal fluid (loading bands below zero) from urine (band over zero) and semen (bands over and below zero) as shown in Figure 4. That is why the scores plot, displayed in Figure 3a, showed the vaginal fluid samples along negative values of PC1, urine samples on positive values of PC1 and semen samples close to the zero value like cotton samples. Subsequently, as can be seen in Figure 3a and Figure 4, PC2 that summarized the next 11% of the variation mainly contributed to differentiate semen (band over zero) from urine and cotton, both close to zero value of PC2. Vaginal fluid was also affected by PC2 because its bands seem to be along both negative and positive values.

Consequently, as shown in Figure 3a, vaginal fluid samples were widely dispersed along the PC2 axis. Finally, PC3, which summarized the next 7% of the variation, helped to separate the three body fluids a bit more, though its effect was much smaller than the previous ones of PC1 and PC2.

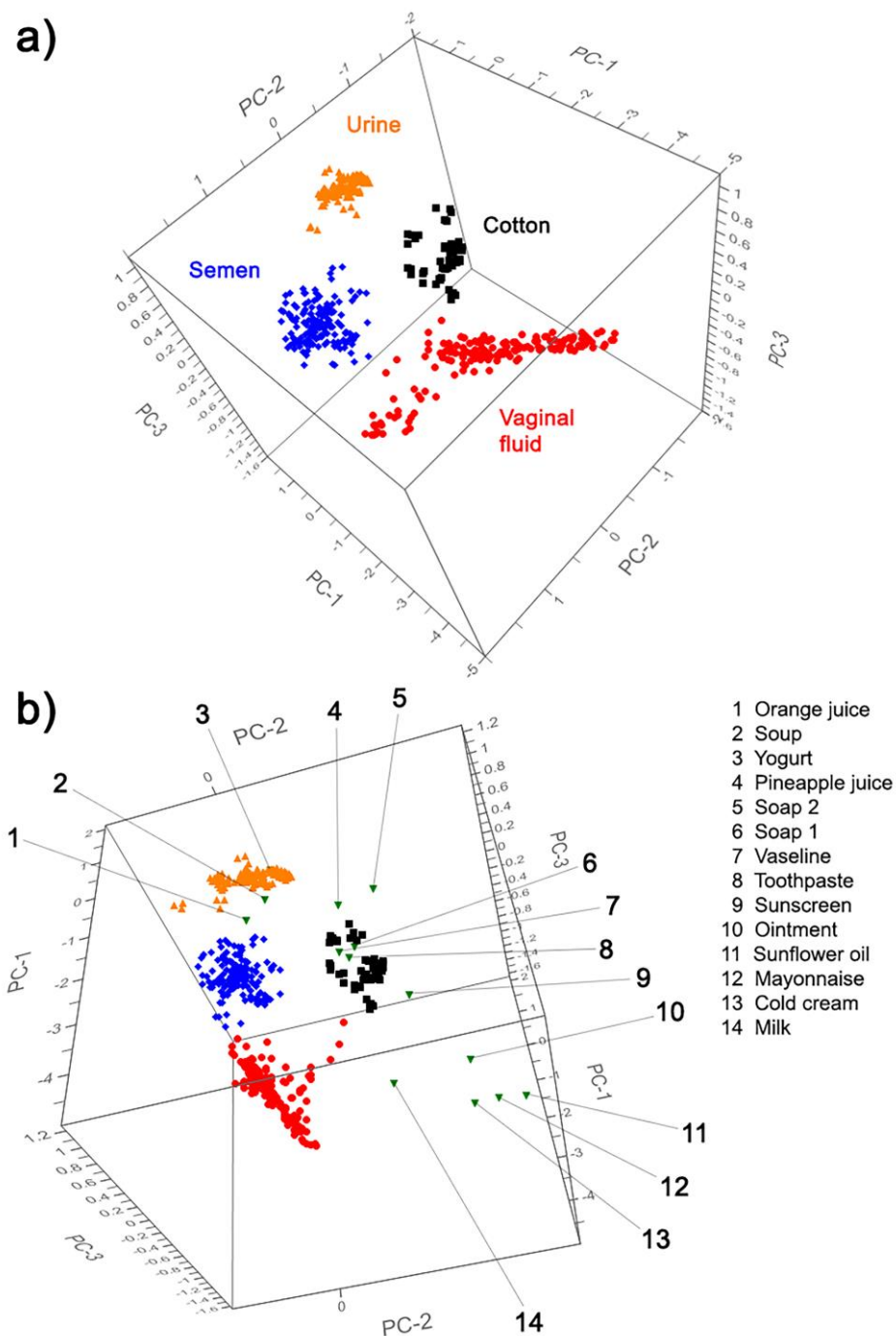


Figure 3. a) Three-dimensional scores PCA plot of the IR spectra from the unstained white and coloured cotton pieces (black ■), vaginal fluid stains (red ●), urine stains (orange ▲) and semen stains (blue ◆) placed on white and coloured cotton. The PCA model is explained by PC1 (77%), PC2 (11%) and PC3 (7%); b) Also including scores from lotions and foodstuffs samples (green ▼) within the PCA plot.

In summary, according to Figure 3a, the pieces of cotton, semen, urine and vaginal fluid stains were clustered in separate groups. This result confirmed that the differentiation of these three body fluids based on their IR spectra was possible by using external reflection FTIR spectroscopy and chemometrics, even though the stains of the body fluids were on coloured cotton fabrics.

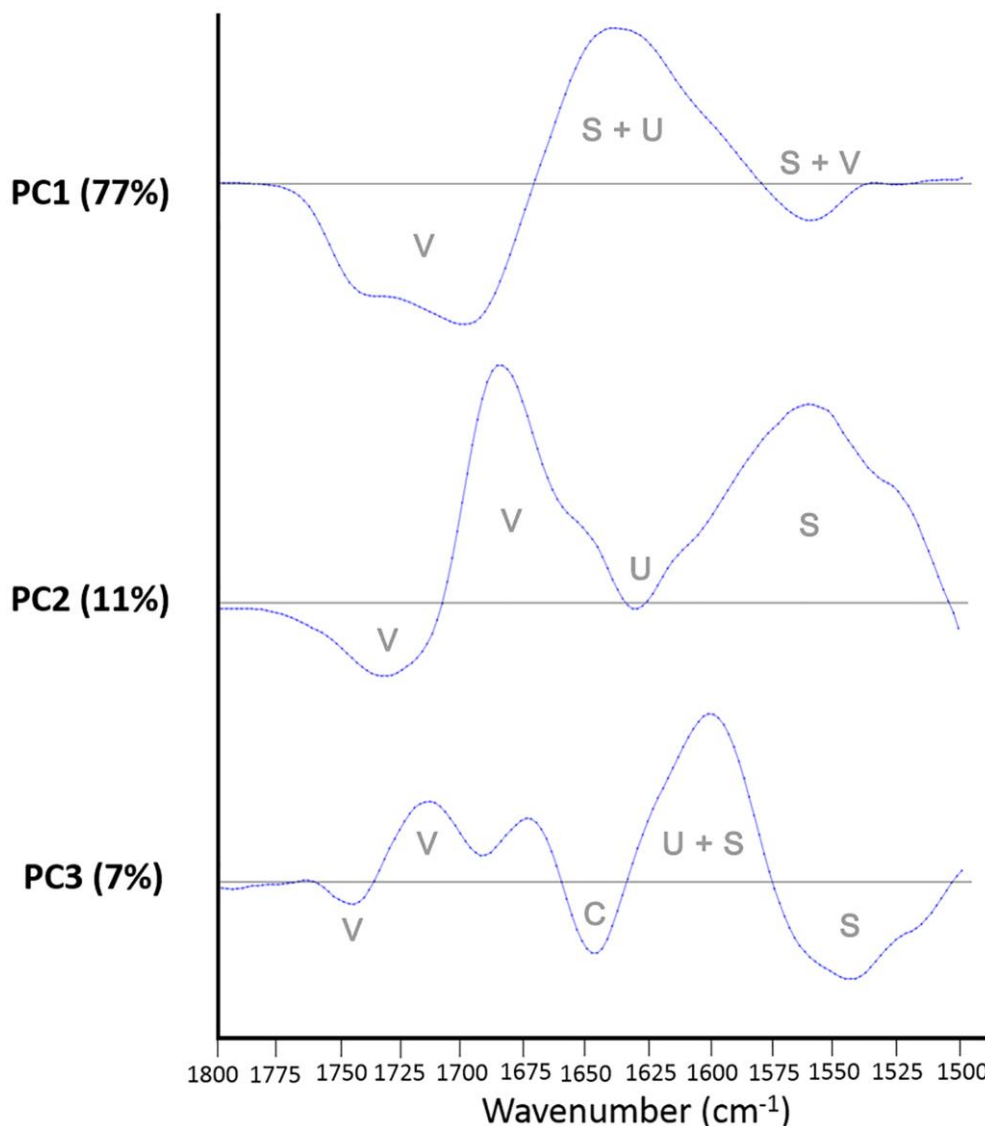


Figure 4. Line loadings plot from PC1 (above), PC2 (centre) and PC3 (below). The tentatively relationship between loadings features and body fluids are indicated as “V” for vaginal fluid, “S” for semen and “U” for urine. Cotton is indicated as “C”.

As was previously explained, stains on different coloured fabrics (white, black, blue, red, green and blue jeans) were included in the PCA. As previously stated, PCA also confirmed that the colour of the fabric did not affect the differentiation of body fluids by using external reflection FTIR spectroscopy and PCA analysis. In addition, most stains of lotions and foodstuffs were displayed far away from body fluids. Because of the

perspective some of them (orange juice (1), soup (2) and yogurt (3)) seem to be quite close to urine group. However, 1 and 2 are above the urine group and 3 is below the urine group (the reason why is almost completely covered). In fact, there is enough distance among their scores to avoid their confusion. Regarding soap 1 (6), vaseline (7) and toothpaste (8) seem to be quite close to cotton. Again, they are below cotton group although it is difficult to notice due to the displayed perspective in Figure 3b.

Considering Figure 3 again, it was also observed that vaginal fluid stains were the body fluid samples with the highest variability according to their wide dispersion. This dispersion seems to be related to two factors. First, it was known that the stains of vaginal fluid had a different quantity of fluid since its deposition was less controlled than those ones for semen and urine. Therefore, some of the dispersion is explained by the differences in the amount of vaginal fluid. We observed that the effect of decreasing the quantity of the deposited fluid produced a shift towards a higher intensity of cotton fabric. The lower the quantity of fluid, the higher the contribution of cotton. Consequently, those spectra are closer to cotton cluster. This result implies that the worst that could occur is a stain without enough quantity of fluid which would be placed close to the cotton group (false negative) leading to necessary subsequent analyses. On the other hand, there is also a dispersion which is not explained by the amount of fluid because it does not tend towards cotton cluster. That dispersion seems to be explained by little differences among the IR spectra from vaginal fluid stains which may be due to a different chemical composition based on the vaginal fluid physiologic process, where vaginal fluid composition varies with the phases of the menstrual cycle. However, the dispersion that appeared within each cluster, considering also vaginal fluid, did not affect the differentiation observed among classes, which were separate enough to be clearly distinguished. Figure 3 shows that there is no overlapping among clusters. Most stains of the same body fluid were closely distributed and every stain was closer to its analogous stains than to the other stains. This fact seems to indicate that the within-cluster variability is lower than inter-cluster variability, which is obviously needed to discriminate among different samples.

3.4. Classification of unknown stains on fabrics

The most ambitious goal of this work was focused on developing a precise and accurate SIMCA classification model to differentiate and classify unknown stains of the three

body fluids studied. The SIMCA model was calculated using separate PCAs of each class (cotton, vaginal fluid, urine and semen), which were made considering white cotton and the stains from four of the donors. Subsequently, SIMCA was applied to the remaining stains used as test set in order to specifically classify three stains of each body fluid from a fifth donor, coloured cotton samples, stains of the substances that may provide false positives and stains of semen-vaginal fluid mixtures.

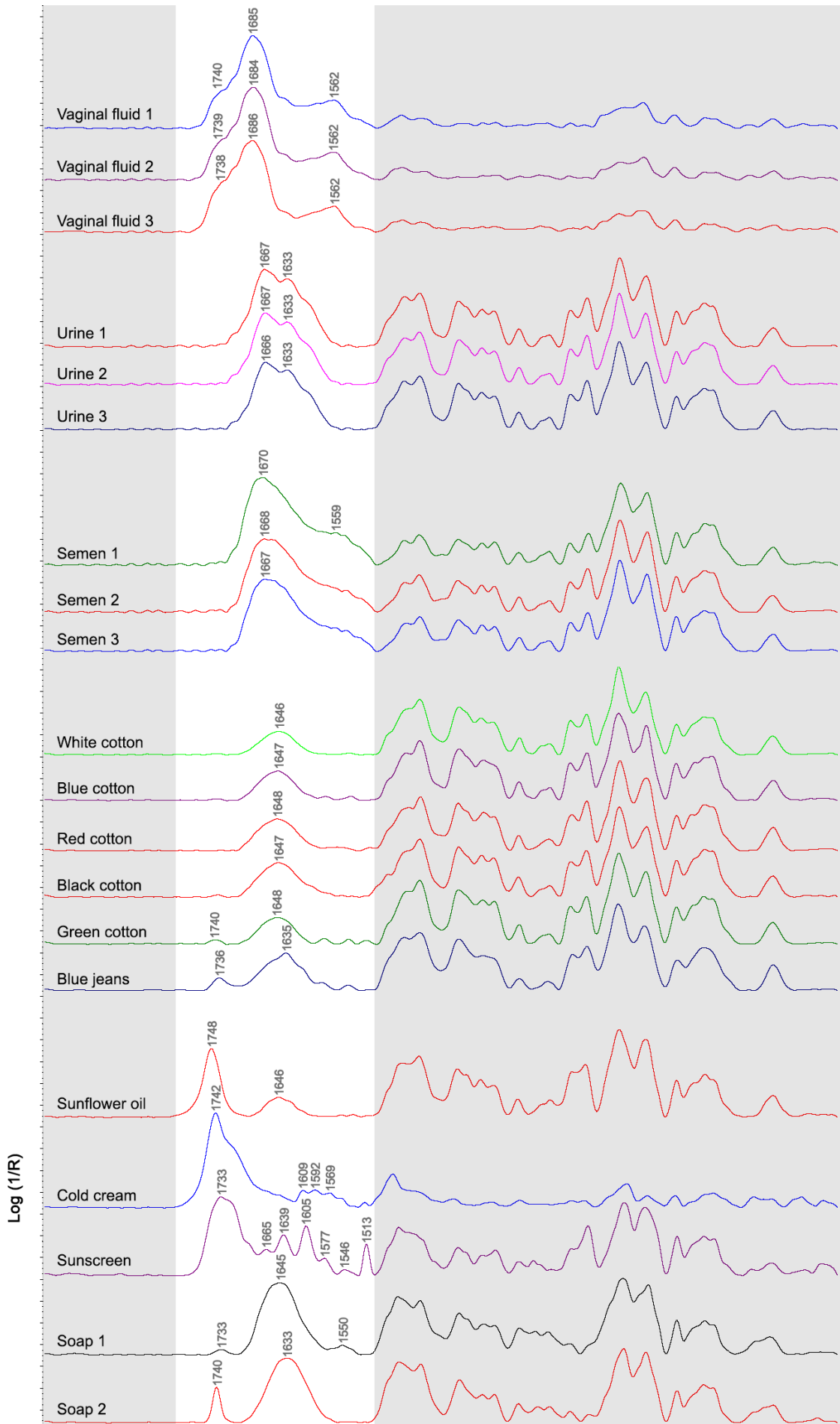
Table III shows the results of the SIMCA classification for these test samples. The classification of the samples in each class was established according to two measures: the sample-to-model distance (S_i) and the sample leverage (H_i). “ S_i ” is a measure of how far the sample lies from the modelled class and it is computed as the square root of the sample residual variance. A small distance means that the sample is well described by the class model. The S_i is the basis of the statistical criterion used to decide whether a sample can be classified as a member of the class or not³⁰. It is an intrinsic value of each sample in the SIMCA classification. “ H_i ” is a measure of how far the projection of a sample onto the model is from the class centre. In fact, it expresses how different the sample is from the other members, regardless of how well it can be described by the class model³⁰. The “ H_i ” has a linear relationship to the Hotelling T^2 statistic, which is related to the Mahalanobis distance. Its critical limit is based on an F-test and that is why H_i accepts different significance levels. In a classification methodology, the significance level is the probability of rejecting a sample from a class when it belongs to that class. However it is more important in this cases to reduce the opposite error: including a sample in a class when it does not belong to that class; because that would mean a false positive. In methods of identification, false negatives are preferred to false positives. In this case, a significance limit of 5% was established.

As Table III shows, the SIMCA model accurately classified the stains of vaginal fluid, urine, semen and the unstained cotton samples. In addition, the 14 potential false positive substances were not classified in any class although some of them were false positives by UV-Vis spectroscopy.⁵ Finally, the SIMCA model partly classified semen-vaginal fluid mixture stains as vaginal fluid. Mixture stains were only classified as vaginal fluid instead of both semen and vaginal fluid. Erroneously, semen was not detected in the mixture. The spectra from all of these test samples are shown in Figure S2, (see supporting material).

Table III. SIMCA classification matrix result of the prediction of test samples. The asterisks indicate the classes where samples were classified after establishing a significance limit of 5%. The spectra from these samples are shown in Figure S2 in supporting material.

Test set	Cotton	Vaginal fluid	Urine	Semen
Vaginal fluid 1		*		
Vaginal fluid 2		*		
Vaginal fluid 3		*		
Urine 1			*	
Urine 2			*	
Urine 3			*	
Semen 1				*
Semen 2				*
Semen 3				*
White Cotton	*			
Blue Cotton	*			
Red Cotton	*			
Black Cotton	*			
Green Cotton	*			
Blue Jeans	*			
Sunflower oil				
Cold cream				
Sunscreen				
Soap 1				
Soap 2				
Milk				
Mayonnaise				
Toothpaste				
Soup				
Ointment				
Vaseline				
Yogurt				
Orange juice				
Pineapple juice				
Mixture 1		*		
Mixture 2		*		

Finally, it is important to highlight that no sample was classified into an erroneous class. According to these preliminary results, if the SIMCA model classified a stain in a particular fluid class, it would mean that that fluid is present in the stain. However, the non-identification of a fluid in a stain could also take place, as happened with mixture stains, where semen was not identified.



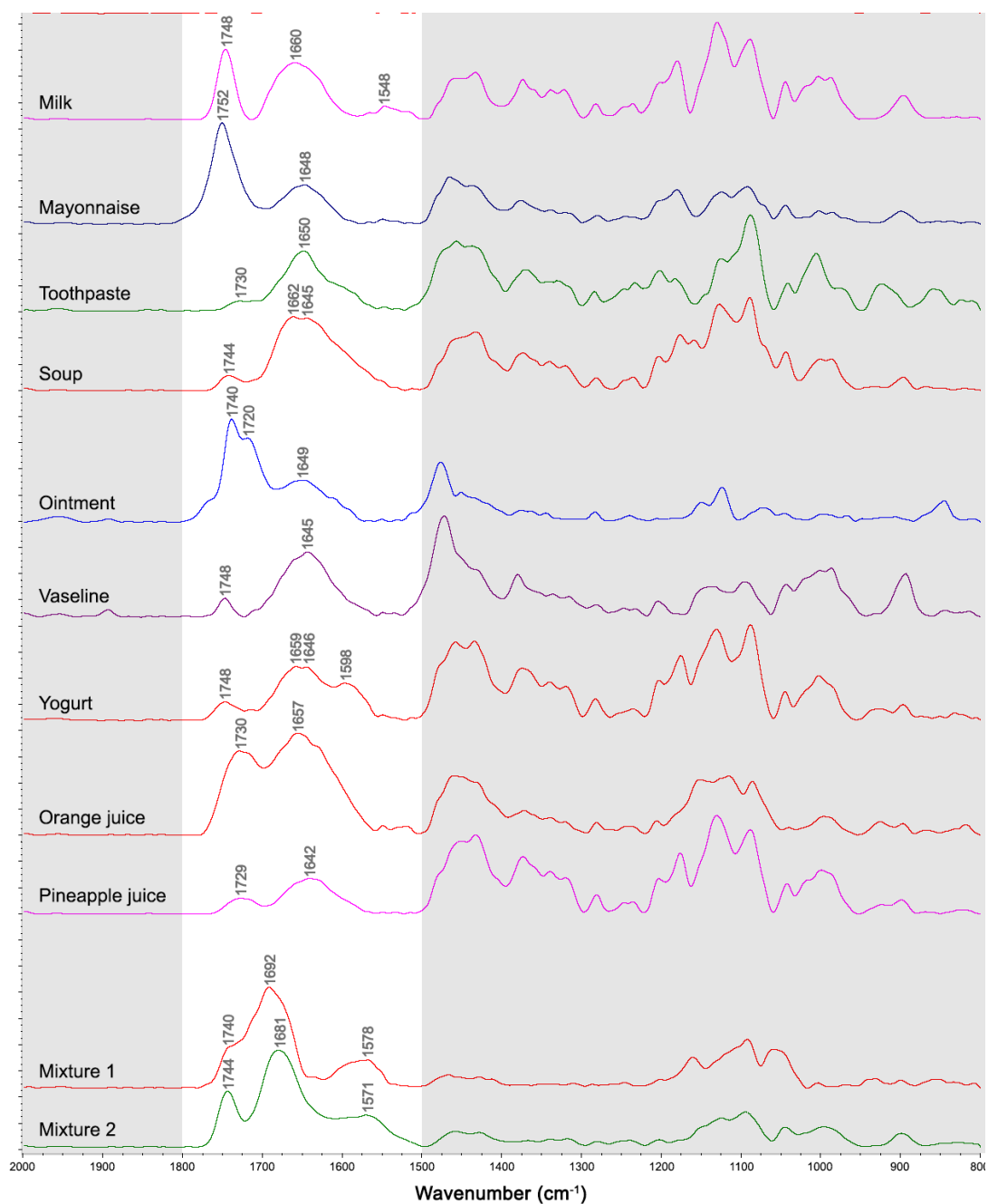


Figure S2 (Supplemental material). External reflection FTIR spectra from test samples used for the SIMCA classification in table III.

4. Conclusions and future trends

In this study, dry stains containing about 20-40 mg of vaginal fluid, semen and urine on cotton fabrics were detected and differentiated by external reflection FTIR spectroscopy. By visually comparison of their spectra it was observed that they were different so this technique has interesting potential to identify stains of body fluids.

Although the bands assignment provided information about the major compounds in the fluid, the characterization of the body fluid was achieved by its IR spectral signature using the 1500-1800 cm^{-1} range.

The differentiation of stains on fabrics of the three body fluids studied (semen, vaginal fluid and urine) was confirmed both visually and by chemometrics. In fact, spectra from unknown samples could be added to the PCA model to statistically see its proximity to each body fluid cluster. In addition, it was also demonstrated that the colour of the fabric on which the stain was does not influence the body fluid identification by external reflection FTIR spectroscopy. The weak external reflection power, which does not allow the differentiation of coloured fabrics, may also be a limitation when detecting body fluids when its concentration is low.

Finally, a SIMCA classification model was applied to the identification and classification of unknown stains of body fluids on cotton fabrics by external reflection FTIR spectroscopy. Semen, urine and vaginal fluid stains were not misclassified and no cosmetic substance or foodstuff (of the 14 analysed) constituted a false positive. Unfortunately, mixture stains were only classified as one of its components (vaginal fluid). Therefore, further research is needed to improve and refine the classification model, particularly with semen-vaginal fluid mixtures where only the female's fluid was identified. This should be achieved by analysing stains with different amount of fluid and stains from more donors.

External reflection measurements can be used for standoff detection. Interestingly, in this study, stains were analysed at a distance of only 1.5 cm from the spectrometer because of the configuration of the instrument. However, external reflection can be designed to work at longer distances because it does not require a physical contact between sample and instrument. This great advantage makes external reflection particularly suitable for evidence recovery in which finding and locating the evidence together with a preliminary identification is very desirable. Consequently, as a future trend, it seems very promising to combine this IR spectroscopy mode with imaging and photography by means of hyperspectral imaging (HSI), which allows the measuring of the IR spectrum of every pixel in the image. Using this approach, the differences in chemical composition could be visualized in the image.

Acknowledgements. Authors thank Bruker Company for Alpha-FTIR spectrometer and donors for their voluntary collaboration. The authors thank Paula Guíu for her assistance in the experimental measurements. F. Zapata thank the University of Alcalá for his pre-doctoral grant.

References

- 1 K. Virkler, I.K. Lednev. "Analysis of Body Fluids for Forensic Purposes: From Laboratory Testing to Non-Destructive Rapid Confirmatory Identification at a Crime Scene". *Forensic Sci. Int.* 2009. 188(1-3): 1-17.
- 2 F. Zapata, M.A. Fernández de la Ossa, C. García-Ruiz. "Emerging Spectrometric Techniques for the Forensic Analysis of Body Fluids". *TrAC, Trends Anal. Chem.* 2015. 64: 53-63.
- 3 M.J. Auvdel. "Comparison of Laser and Ultraviolet Techniques Used in the Detection of Body Secretions". *J. Forensic Sci.* 1987. 32(2): 326-345.
- 4 M. Stoilovic. "Detection of Semen and Blood Stains Using Polilight as a Light Source". *Forensic Sci. Int.* 1991. 51(2): 289-296.
- 5 D.G. Nelson, K.A. Santucci. "An Alternate Light Source to Detect Semen". *Acad. Emerg. Med.* 2002. 9(10): 1045-1048.
- 6 N. Vandenberg, R.A.H. van Oorschot. "The Use of Polilight in the Detection of Seminal Fluid, Saliva, and Bloodstains and Comparison with Conventional Chemical-Based Screening Tests". *J. Forensic Sci.* 2006. 51(2): 361-370.
- 7 A. Fiedler, J. Rehdorf, F. Hilbers, L. Johrdan, C. Stribl, M. Benecke. "Detection of Semen (Human and Boar) and Saliva on Fabrics by a Very High Powered UV-/VIS-Light Source". *Open Forensic Sci. J.* 2008. 1: 12-15.
- 8 C.E. Thorn, S.J. Matcher, I.V. Meglinski, A.C. Shore. "Is Mean Blood Saturation a Useful Marker of Tissue Oxygenation?" *Am. J. Physiol.: Heart Circ. Physiol.* 2009. 296(5): 1289-1295.
- 9 S. Kloß, B. Kampe, S. Sachse, P. Rösch, E. Straube, W. Pfister, M. Kiehntopf, J. Popp. "Culture Independent Raman Spectroscopic Identification of Urinary Tract Infection Pathogens: A Proof of Principle Study". *Anal. Chem.* 2013. 85(20): 9610-9616.

- 10 K. Gilany, R.S.M. Pouracil, M.R. Sadeghi. "Fourier Transform Infrared Spectroscopy: a Potential Technique for Noninvasive Detection of Spermatogenesis". *Avicenna J. Med. Biotechnol.* 2014. 6(1): 47-52.
- 11 Z. Huang, X. Chen, Y. Chen, J. Chen, M. Dou, S. Feng, H. Zeng, R. Chen. "Raman Spectroscopic Characterization and Differentiation of Seminal Plasma". *J. Biomed. Opt.* 2011. 16(11): 110501/1-110501/3.
- 12 B.R. Wood, L. Chiriboga, H. Yee, M.A. Quinn, D. McNaughton, M. Diem. "Fourier Transform Infrared (FTIR) Spectral Mapping of the Cervical Transformation Zone, and Dysplastic Squamous Epithelium". *Gynecol. Oncol.* 2004. 93(1): 59–68.
- 13 E.M. Kanter, E. Vargis, S. Majumder, M.D. Keller, E. Woeste, G.G. Rao, A. Mahadevan-Jansen. "Application of Raman Spectroscopy for Cervical Dysplasia Diagnosis". *J. Biophotonics* 2009. 2(1-2): 81-90.
- 14 J.L. Pezzaniti, T-W. Jeng, L. McDowell, G.M. Oosta. "Preliminary Investigation of Near-Infrared Spectroscopic Measurements of Urea, Creatinine, Glucose, Protein, and Ketone in Urine". *Clin. Biochem.* 2001. 34(3): 239-246.
- 15 W.R. Premasiri, J.C. Lee, L.D. Ziegler. "Surface-Enhanced Raman Scattering of Whole Human Blood, Blood Plasma and Red Blood Cells: Cellular Processes and Bioanalytical Sensing". *J. Phys. Chem. B.* 2012. 116(31): 9376-9386.
- 16 N.N. Daéid. Review Papers, 17th Interpol International Forensic Science Managers Symposium. Lyon, France. 2013. Pp. 280-610.
- 17 F.P. Smith. *Handbook of Forensic Drug Analysis*, Elsevier, 2005.
- 18 D.E. Bugay. "Characterization of the Solid-State: Spectroscopic Techniques". *Adv. Drug Delivery Rev.* 2001. 48(1): 43-65.
- 19 J. Yinon. *Counterterrorist Detection Techniques of Explosives*, Amsterdam, The Netherlands: Elsevier, 2007.
- 20 J.I. Steinfeld, J. Wormhoudt. "Explosives Detection: A Challenge for Physical Chemistry". *Annu. Rev. Phys. Chem.* 1998. 49: 203-232.
- 21 K. Virkler, I.K. Lednev. "Raman Spectroscopy Offers Great Potential for the Nondestructive Confirmatory Identification of Body Fluids". *Forensic Sci. Int.* 2008. 181(1-3): e1-e5.
- 22 V. Sikirzhytski, A. Sikirzhytskaya, I.K. Lednev. "Multidimensional Raman Spectroscopic Signatures as a Tool for Forensic Identification of Body Fluid Traces: A Review". *Appl. Spectrosc.* 2011. 65(11): 1223-1232.
- 23 V. Sikirzhytski, K. Virkler, I.K. Lednev. "Discriminant Analysis of Raman Spectra for Body Fluid Identification for Forensic Purposes". *Sensors* 2010. 10(4): 2869-2884.

- 24 V. Sikirzhytski, A. Sikirzhytskaya, I.K. Lednev. "Advanced Statistical Analysis of Raman Spectroscopic Data for the Identification of Body Fluid Traces: Semen and Blood Mixtures". *Forensic Sci. Int.* 2012. 222(1-3): 259-265.
- 25 I.K. Lednev. Application of Raman Spectroscopy for an Easy-to-Use, on-Field, Rapid, Nondestructive, Confirmatory Identification of Body Fluids. Report to the U.S. Department of Justice. 2012.
- 26 K.M. Elkins. "Rapid Presumptive "Fingerprinting" of Body Fluids and Materials by ATR FT-IR Spectroscopy". *J. Forensic Sci.* 2011. 56(6): 1580-1587.
- 27 C. Orphanou. "The Detection and Discrimination of Human Body Fluids Using ATR FT-IR Spectroscopy". *Forensic Sci. Int.* 2015. 252: e10-e16.
- 28 M. Pirzer, J. Sawatzki. Method and Device for Correcting a Spectrum. US Patent: US 7,359,815 B2. Apr. 15, 2008.
- 29 E. Pretsch, P. Bühlmann, C. Affolter, A. Herrera, R. Martinez. Structure Determination of Organic Compounds. Berlin, Germany: Springer-Verlag, 2000.
- 30 The Unscrambler X 10.1 Software (Oslo, Norway). Help contents. 2009.