



Human odor and forensics: Towards Bayesian suspect identification using GC x GC–MS characterization of hand odor

Vincent Cuzuel, Roman Leconte, Guillaume Cognon, Didier Thiebaut, Jérôme Vial, Charles Sauleau, Isabelle Rivals

► To cite this version:

Vincent Cuzuel, Roman Leconte, Guillaume Cognon, Didier Thiebaut, Jérôme Vial, et al.. Human odor and forensics: Towards Bayesian suspect identification using GC x GC–MS characterization of hand odor. *Journal of Chromatography B Biomedical Sciences and Applications*, 2018, 1092, pp.379-385. 10.1016/j.jchromb.2018.06.018 . hal-02395700

HAL Id: hal-02395700

<https://hal.science/hal-02395700>

Submitted on 5 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Human odor and forensics: towards Bayesian suspect identification using GCxGC-MS characterization of hand odor

Authors: Vincent Cuzuel ^{a,c}, Roman Leconte ^b, Guillaume Cognon ^a, Didier Thiébaut ^c, Jérôme Vial ^c, Charles Sauleau ^a, Isabelle Rivals* ^b

Corresponding author: Isabelle Rivals, isabelle.rivals@espci.fr, +33140794545

^a Institut de Recherche Criminelle de la Gendarmerie Nationale, Caserne Lange, 5 boulevard de l'Hautil, BP 20312 Pontoise, 95037 Cergy Pontoise CEDEX, France

^b Equipe de Statistique Appliquée, ESPCI Paris, PSL Research University, INSERM, UMRS 1158 Neurophysiologie Respiratoire Expérimentale et Clinique, 10 rue Vauquelin, Paris, France

^c UMR CBI 8231, CNRS, Laboratoire Sciences Analytiques Bioanalytiques et Miniaturisation, ESPCI Paris, PSL Research University, 10 rue Vauquelin, Paris, France

Highlights

- Subject identification by hand odor using GCxGC-MS and Bayesian testing is proposed.
- Bayesian testing enables to decide if two odor samples stem from the same person.
- The AUC, sensitivity and sensitivity of the classifiers is evaluated on a panel of 139 subjects.

Abstract

A new method for identifying people by their odor is proposed. In this approach, subjects are characterized by a GCxGC-MS chromatogram of a sample of their hand odor. The method is based on the definition of a distance between odor chromatograms and the application of Bayesian hypothesis testing. Using a calibration panel of subjects for whom several odor chromatograms are available, the densities of the distance between chromatograms of the same person, and between chromatograms of different persons are estimated. Given the distance between a reference and a query chromatogram, the Bayesian framework provides an estimate of the probability that the corresponding two odor samples come from the same person. We tested the method on a panel that is fully independent from the calibration panel, with promising results for forensic applications.

Keywords

Bayesian inference, comprehensive two-dimensional gas chromatography, forensics, human hand odor, mass spectrometry.

1. Introduction

Forensic profiling of human hand odor using analytical devices and statistical tools is of particular interest to confirm the information provided by dogs in courts of law. Indeed, the identification dogs can perform remains challenged because of their inability to testify. For some years, several teams have tried to get a better understanding of the human odor and have aimed at characterizing it using analytical devices. The results and conclusions of these studies on odor sampling, sample analysis and data processing, three steps that are essential to a successful odor identification strategy, were recently reviewed¹.

The sampling procedure can either be performed directly, i.e. with contact, or indirectly, i.e. without contact: direct sampling necessitates putting an adsorbent phase on the skin of the subject in order to collect odor compounds², whereas indirect sampling is based on air suction around the object of interest³. In both cases, appropriate protocols can be implemented to reduce the contamination by exogenous substances. Most of the time, they consist in prewashing the sampling zone with tap water or perfume-free soap and subsequent “natural” drying. Subjects can be asked to use special soap and deodorant or no deodorant at all up to a week before the study^{4,5}. Some authors asked their subjects not to eat specific foods (spices, garlic) or to do some exercise before sampling in order to increase the production of sweat^{5,6}. To our knowledge, there is no validated, let alone standardized protocol yet.

The analytical separation is often performed by gas chromatography (GC) coupled with mass spectrometry (MS)^{7,8}. Still, the information collected might not be sufficient, especially when hundreds of compounds need to be monitored within short-time analysis. Thus, the use of two-dimensional gas chromatography (GCxGC) is a relevant alternative and was already shown to provide more information than classic GC⁹.

Finally, the statistical processing is essential to extract the information relevant for identification. To our knowledge, only few studies were carried out on a large panel, one of the

largest being that collected by Colon-Crespo et al. involving 105 subjects, 54 women and 51 men belonging to 3 different ethnic groups (Caucasian, Hispanic and east Asian), with ages in the range 18-77². In their study, Curran et al.⁷ worked with a panel of 60 people, 30 men and 30 women, but since only one sample (hands) per person was collected, the comparison of samples from the same individual was not possible. Brown et al.⁸ worked with a more reduced panel (20 people), but sampled each subject 3 times from different body locations. However, though samples were taken in triplicate, average profiles were used, so that the degree of similarity between samples from the same person was again not evaluated. Gallagher et al.⁵ sampled the backs and arms of 25 people twice to study the significance of the effects of three factors, sampling area (back/arm), sex (male/female) and age (young/old): significant differences were found for certain compounds across age and site of collection, but not between men and women. The study conducted by Penn et al.⁴ is more thorough, their panel including 197 individuals who were sampled five times each, once every fortnight, over a ten-week period. The results of this study suggest that using pattern recognition on the entire profile pattern, rather than on a small set of compounds, could be essential to perform proper identification.

The aim of the present paper is to propose an attempt, not at classifying people into categories by gender and/or ethnicity, not at attributing chromatograms to a finite set of persons for which numerous odor samples are available, but at identification. For this purpose, we gathered an important and diversified panel of subjects sampled several times, and adopted a Bayesian approach, which does not suffer from several drawbacks of the frequentist approach. The first main drawback is the asymmetry between the null and alternative hypotheses, since evidence can essentially be collected against the null hypothesis and not in favor of it, and the second one is that very small p-values can be obtained despite weak evidence against the null hypothesis^{10,11}. For these reasons, Barcaru and Vivò-Truyols used Bayesian hypothesis testing to spot differences between the GCxGC-MS chromatograms of pairs of diesel samples¹². Bayesian approaches are successfully used in many fields of forensics, especially to assess the probative value of DNA^{13,14}. Though there are large differences between DNA and odor profiles, the ultimate question is essentially the same: given a reference profile (the one collected on the crime scene), does the query profile (that of the suspect) correspond to the same person?

The proposed procedure can be outlined as follows. First, the available panel, in which subjects are described by several GCxGC-MS chromatograms of their hand odor reduced to the intensities of a large set of compounds, is to be split into a calibration set for training and validation, and an independent test set for performance estimation. Second, the calibration set is used to estimate the distributions of a distance between two chromatograms when they correspond to the same subject, i.e. under the null hypothesis, and when they belong to two different subjects, i.e. under the alternative hypothesis. The posterior probability of the null hypothesis given the distance between two chromatograms is computed using Bayes' formula, and the performance of the corresponding classifier is estimated using the independent test set, in terms of area under the receiver operating curve, sensitivity and specificity. The issues of the choice of a distance and of the selection of the most appropriate compounds to enhance the performance will be discussed.

2. Material and methods

2.1. Subjects

A panel of 119 subjects was set up. The subjects gave their consent for the analyses of the samples and filled an information sheet. The data were anonymized before the analysis.

This panel gathers 61 men and 58 women, with 39 subjects aged 10 to 23, 39 subjects aged 24 to 36, and 41 subjects aged 37 to 81. They also belong to 3 different phototype groups, the phototype measuring the susceptibility of the skin to tan depending on the amount of melanin in the skin, and is linked to the vitamin D biosynthesis by the skin: the first group (25 subjects) burns easily and is seldom or never tanned, the second group (79 subjects) can get tanned more or less quickly and also burned, whereas the last group (15 subjects) never burns when exposed to the sun, as categorized by Fitzpatrick¹⁵ (actually we grouped Fitzpatrick's six phototypes by pairs, the first group bringing phototypes I and II together, etc.).

These subjects were sampled four times to make sure that at least three repetitions per subject were available, should analytical problems occur.

2.2. Analytical separation and detection

2.2.1. Sorbent phase for compounds trapping

Sorb-Stars® were purchased from Action Europe (Sausheim, France) for direct sampling. This sorbent is a patented silicon-based polymeric phase and is subject to specific conditioning

processes to avoid contaminations as much as possible. The Sorb-Star® is a 2 cm long cylinder with a cylindrical section of 2 mm in diameter, has a density of 1.12 g/cm³ and is compliant with FDA 177.2600. It is physiologically safe, and even suitable for applications in the food industry.

2.2.2. Sample collection

Care was taken that the sampling conditions were the same for all subjects. To this end, identical sample collection kits of Sorb-Stars®, special soap (Topialyse, SVR laboratory), nitrile gloves and information sheets were dispatched to different samplers in France, who were given precise instructions for the sampling protocol. The subjects were asked to wash their hands for 30 seconds with the provided soap, to rinse them carefully with clear water for 1 minute, to dry them with a paper towel and to rub them for 2 minutes. 5 minutes later, 4 Sorb-Stars® were placed into the hands of the subjects, who rubbed them for 15 minutes. In the meantime, a blank was obtained by placing an open vial containing a sorbstar® in front of the sampled person.

2.2.3. Analytical devices

The Sorbstars® were thermodesorbed prior to GCxGC-MS analysis. The development of the analytical method and its optimization were the topic of two previous studies^{16,17}.

The purge and trap system Versatile Sample Preparator (VSP4000) was purchased from Innovative Messtechnik GmbH (Vohenstrauß, Germany). Volatile substances were purged from the Sorbstar® by the carrier gas of the GC. This concentration step is done by adsorption on a Tenax TA® in the system trap by freezing out at -30 °C. After incubation of the sample at 190 °C and completion of the purging process (20 mL/min during 20 minutes), the concentrated substances were transferred by fast thermal desorption from the trap onto a transfer line, heated at 280 °C, and then separated by GC.

The thermodesorption device was coupled with a GCxGC-MS Q2010Plus purchased from Shimadzu (Kyoto, Japan). A ZB-1MS column (30 m x 0.25 mm, 0.25 µm) (Phenomenex, Torrance, USA) coupled with a ZB-1701 column (1.5 m x 0.1 mm, 0.1 µm) (Phenomenex, Torrance, USA) were used to conduct the chromatographic separation. The modulation was performed with a N₂ cooled Zoex ZX1 thermal modulator (Zoex, Houston, USA), and the modulation time was set at 8 seconds. The initial temperature was set to 40 °C for 1 minute, then raised to 250 °C at 2.5 °C/min, and held for 1 minute at 250 °C. The mass spectrometer

was used with the electronic ionization source (70 eV) heated at 200 °C, the acquisition being performed in scan mode. The scan range was 29-250 m/z, and the sampling frequency 50Hz. Figure 1 displays a typical chromatogram collected on a 25-year-old man's hand.

2.3. Chromatogram processing

Data were acquired, converted to .mzXML file with GC Real Time Analysis 4.20 (Shimadzu software), and then processed with MatlabTM (Natick, MA, USA) version 9.3.0.713579 (R2017b), its Statistics and Machine Learning Toolbox version 11.2 and its Bioinformatics Toolbox version 4.9.

A preliminary manual processing of 25 chromatograms obtained on a subset of subjects of both genders sampled several times at different time instants enabled to draw up a list of several hundreds of peaks. A library was built to store their retention times, their linear retention index, their mass spectrum, and the name of the corresponding compound when it could be identified using the NIST library. Indeed, a compound does not need to be formally identified for the comparison of chromatograms, whereas the availability of its mass spectrum is compulsory. We also checked whether compounds described in the literature as constituents of the human hand odor¹ were present in this library, otherwise they were included. This preliminary work led us to a customized library of 600 compounds, identified or not, which were looked for in each chromatogram using a "home-made" Matlab script¹⁸.

As a result, each chromatogram was characterized by a list of 600 descriptors/compounds and their associated intensity (apex of the detected peak). In order to compensate for uncontrolled variations of the total intensity of the chromatograms, the sum of these intensities was normalized to unit value. Not knowing whether all 600 compounds are relevant for identification, such a normalization might be questionable. Thus, the possibility to avoid the problem by working on the binarized intensities, i.e. 1 if the compound is present or 0 if it is absent from the sample, was also investigated.

3. Statistical treatment

The aim is, given two odor chromatograms, to decide whether the corresponding odors belong to the same person or not. In the Bayesian framework, the two hypotheses are random events whose probability, given the observed data, can be estimated.

3.1. Bayesian hypothesis testing

In the following, H_0 refers to the null hypothesis that the two chromatograms are similar and correspond to the same person, and H_1 to the alternative hypothesis that they do not. Given the *a priori* probabilities $P(H_0)$ and $P(H_1)$, the Bayesian formula yields the posterior probability of H_0 given some observed data D :

$$P(H_0 | D) = \frac{f(D | H_0)P(H_0)}{f(D | H_0)P(H_0) + f(D | H_1)P(H_1)}$$

where the $f(D | H_i)$ are the distributions of the observed data under H_0 and H_1 , or likelihoods. Jeffreys (1939) developed an absolute scale to evaluate the degree of confidence in the null hypothesis outside a decisional framework based on the posterior probability of H_0 using the likelihood ratio ^{19, 20} defined as:

$$LR(D) = \frac{f(D | H_0)}{f(D | H_1)}$$

which is independent from the *a priori* probabilities. If $LR > 10$, the confidence is considered as strong, and if $LR > 100$, as decisive.

3.2. Choice of a distance between two chromatograms

Actually, the observed data D consists of the two 600-vectors of intensities representing the two chromatograms. Here, we propose to reduce the information contained in these two vectors to a scalar that is relevant for the discrimination between H_0 and H_1 , i.e. to a scalar distance d between the two vectors of normalized or binarized intensities. We will consider the Euclidean distance and the distances based on Pearson and Spearman's correlation coefficients (one minus the correlation coefficient), the latter differing in the case of normalized intensities only. Compared to the Euclidean distance, the correlation-based distances are robust with respect to shifts and linear transformations of the peak intensities, and since Spearman's correlation coefficient is able to capture a monotonic nonlinear association, as opposed to Pearson's linear correlation coefficient²¹, the Spearman correlation-based distance is also expected to be more robust with respect to nonlinear variations of the peak intensities.

3.3. Estimating the distributions of the distance under H_0 and H_1

In order to apply Bayes' formula, we need to have an insight into the distribution of the distance d between chromatograms under H_0 (same person) and under H_1 (different persons). For this purpose, a part of the panel was used as calibration set, while the rest of the panel served as test set, i.e. for independent performance estimation. The calibration set was used to build two types of couples of chromatograms: couples chromatograms obtained on the same person and couples of chromatograms obtained on two different persons. Using the former, an empirical estimate of $f(d|H_0)$ was obtained, while $f(d|H_1)$ was estimated using the latter. To be able to retrieve $(d|H_i)$ whatever the value of d , the empirical densities were fitted with a Gaussian mixture distribution using Matlab's function "fitgmdist".

3.4. Compound selection

We also studied the advantage gained by selecting the compounds that contribute most to having different densities under H_0 and H_1 . To this end, for each compound c , we computed the absolute values of the difference in intensity for couples of chromatograms corresponding to the same person (sample corresponding to H_0), and those for couples of chromatograms corresponding to different persons (sample corresponding to H_1). We tested for a difference between these two samples with Wilcoxon's non parametric test when using normalized intensities, and with Fisher's exact test when using binarized intensities, both tests being one-sided since smaller differences in intensity under H_0 are desired. These tests lead to a p -value p_c for each compound, and the compounds for which $-\log_{10}(p_c)$ was larger than a given threshold π were selected, the value of π being established by cross-validation on the calibration set.

3.5. Performance estimation

We assumed that the plausibility of the two hypotheses H_0 and H_1 was equal prior to taking the data into account, i.e. we took prior probabilities $P(H_0) = P(H_1) = 0.5$. Varying the decision threshold on $P(H_0|d)$, the sensitivity and the specificity were estimated on the calibration and test sets, and used to compute the corresponding areas under the receiver operating characteristic "ROC" curve (AUC)²². The likelihood ratio, which is independent from the *a priori* probabilities, will also be evaluated as a function of the distance d .

4. Results and discussion

The data was first split into calibration and test sets, the split being made so as to respect the gender and phototype proportions, with subject of all ages in the two sets, and chromatograms of the same subject being put in the same set. The calibration set includes 260/75 chromatograms/individuals, while there are 139/44 chromatograms/individuals in the test set (due to analytical problems, the four samples per subject were not systematically exploitable).

The homogeneity of the two sets can be grasped with a principal component analysis (PCA), see Figure 2. Figure 2 also shows a limited fidelity: the chromatograms of the same subject are not necessarily very similar, see for example subject 61 of the calibration set (bottom left, 3 chromatograms) and subject 20 of the test set (top right, 4 chromatograms). As a matter of fact, though the four samples of a given subject were performed at the same time, the GCxGC-MS analyses of these samples were randomized, so that the dispersion of the results reflects the analytical variability. The latter can be estimated using the 111 subjects for which at least 2 chromatograms are available to compute relative standard deviations of the compounds normalized intensities. The median of the mean relative standard deviations is indeed quite large with a median of 47.1% and interquartile range [5.2% ; 70.2%].

The procedure was first applied on the 600 compounds, i.e. without compound selection, either on the corresponding normalized intensities, or on their binarized version (0 if the compound is absent, 1 if it is present). The chosen distance was either the Euclidean distance, or the distances based on Pearson's and Spearman's correlation coefficients. The histograms under H_0 (341 couples) and H_1 (33 329 couples) were both fitted with a Gaussian mixture of two components. The corresponding results in terms of AUC on the calibration set and on the test set (in the latter, there are 173 couples of the same person, i.e. under H_0 , and 9 418 couples under H_1) are summarized in Table 1. These results show that binarization significantly improves the performance of the classifier using the Euclidian distance, but better results are obtained with the normalized intensities when using the distance based on the non-parametric Spearman correlation coefficient. This is very satisfactory, because it indicates that the quantitative information supplied by GCxCG-MS is likely to be properly exploited despite the analytical variability, provided an adequate distance is chosen. With a distance that does not compensate for shifts and nonlinearities (Euclidean), or only for shifts

(Pearson's), the possibly improperly normalized and/or nonlinearly transformed intensities cannot be handled, so that binarization improves the performance. With a more robust distance that compensates both for shifts and nonlinearities (Spearman's), it becomes possible to extract useful information from the normalized intensities despite their imperfections. The estimates of the likelihoods and of the posterior probability of H_0 and the likelihood ratio obtained with the Spearman correlation based distance between normalized intensities are shown in Figure 3.

Note that the distribution of the distance under H_0 is bimodal, and that its largest mode is close to the lowest mode of the distribution under H_1 . In order to weaken this undesired mode, we investigated the possibility of keeping only the compounds that were most discriminant between H_0 and H_1 , i.e. which contributed most to having different distributions of the distance d between chromatograms under H_0 and H_1 . They were chosen as the compounds having the most significantly different absolute differences between intensities under H_1 and H_0 in the sense of the appropriate statistical test (Wilcoxon's test for normalized intensities, Fisher's test for binarized ones), with $-\log_{10}(p\text{-value})$ larger than a given threshold π . The value of π was fixed using 3-fold cross-validation on the calibration set. The corresponding results in terms of AUC are summarized in Table 2. These results demonstrate the benefit of compound selection for all distance choices. The best classifier using normalized intensities still outperforms the one using binarized intensities, however, the latter proves more parsimonious: it uses 82 compounds, against 146 for the classifier using the normalized intensities and Spearman's based distance. Note that the two classifiers have 67 compounds in common, i.e. the great majority of those of the classifier on the binarized intensities. The chemical relevance of these lists of compounds will be the subject of a future paper.

The estimates of the likelihoods and of the posterior probability of H_0 as well as the likelihood ratio obtained with the correlation distance on the binarized intensities are shown in Figure 4, and those obtained with Spearman's correlation distance on the normalized intensities are shown in Figure 5. In both cases, as sought for with the compound selection, the distributions of the distance d between chromatograms under H_0 and H_1 are now very different (the second undesired mode of the distribution under H_0 having almost vanished and the plateau of the likelihood ratio having disappeared), hence the increase in performance.

These results must also be analyzed in terms of sensitivity and specificity. The classifier of Figure 5 achieves the best mean of sensitivity and specificity on the calibration set with a sensitivity of 89.1% and a specificity of 94.9%, and on the test set with a sensitivity of 85.9% and a specificity of 95.0%. Though both sensitivity and specificity should be raised for a forensic application, these results are already promising in view of the identification of a suspect by his olfactory fingerprint.

A first limitation of this work is that, for practical reasons, the subjects were sampled at a single time point, so that the variability of the data is essentially due to the analytic variability. A second one is that the chromatograms being compared are of the same nature, i.e. obtained on samples provided by directly sampling the subject, whereas in real life, only the query chromatogram will be obtained this way, the reference chromatogram of the odor on the crime scene being indirect. Finally, the odor collected on the crime scene might also be contaminated by other odors than that of a person. However, the excellent results we have obtained with data of limited reproducibility (as shown on the PCA of Figure 2), to the point that a performance close to the best one could be obtained using only the knowledge of presence/absence of the compounds, are encouraging. Moreover, a normalization on specific compounds (and not on all 600 compounds), or the use of ratios between the abundance of these compounds, might also contribute to improve the results on the non binarized, quantitative data.

A point that might offer an interesting perspective is the extension of the proposed method to a vector distance, using a naïve Bayes classifier. The components of this vector distance could be distances relevant to specific compounds or sets of compounds. Another possibility would be to take other elements of evidence than GCxGC-MS results (like those stemming from dogs for example) into account, either by using of an appropriate more complex vector distance, or by introducing evidence supporting (or against) the null hypothesis in the a priori probabilities ($P(H_0) \neq P(H_1)$).

5. Conclusions

Considering the important amount of compounds that need to be monitored for human hand odor fingerprinting, the development and implementation of statistical tools is essential to perform a proper identification. In this study, 600 compounds were used to characterize the hand odor samples. Bayesian classifiers based on a distance between chromatograms

defined on a subsets of selected compounds and calibrated on 75 subjects of both genders, all ages and various phototypes, proved to be very efficient with AUCs as higher than 98% on a fully independent test set of 44 subjects. These first results are promising considering that the compound selection is purely statistical, so that the analysis of the chemical relevance of the selected compounds, which will be the subject of a future paper, is likely to lead to a better normalization of the compound abundances or to use specific ratios of these abundances. Finally, the principle of the method extends easily to more complex distances possibly involving other elements of evidence, and is hence likely to be suitable for forensic applications.

References

- (1) V. Cuzuel, G. Cognon, I. Rivals, C. Sauleau, F. Heulard, D. Thiébaut, J. Vial, Origin, analytical characterization and use of human odor in forensics, *J. Forensic Sci.* 62 (2017) 330–350. doi:10.1111/1556-4029.13394.
- (2) L.J. Colón-Crespo, D. Herrera-Hernández, H. Holness, K.G. Furton, Determination of VOC marker combinations for the classification of individuals by gender and race/ethnicity, *Forensic Sci. Int.* (2016). doi:10.1016/j.forsciint.2016.09.011.
- (3) L.E. DeGreeff, K.G. Furton, Collection and identification of human remains volatiles by non-contact, dynamic airflow sampling and SPME-GC/MS using various sorbent materials., *Anal. Bioanal. Chem.* 401 (2011) 1295–307. doi:10.1007/s00216-011-5167-0.
- (4) D.J. Penn, E. Oberzaucher, K. Grammer, G. Fischer, H. a Soini, D. Wiesler, M. V Novotny, S.J. Dixon, Y. Xu, R.G. Brereton, Individual and gender fingerprints in human body odour., *J. R. Soc. Interface.* 4 (2007) 331–40. doi:10.1098/rsif.2006.0182.
- (5) M. Gallagher, C.J. Wysocki, J.J. Leyden, A.I. Spielman, X. Sun, G. Preti, Analyses of volatile organic compounds from human skin, *Br. J. Dermatol.* (2008) 780–791. doi:10.1111/j.1365-2133.2008.08748.x.
- (6) J.G. Logan, M. a. Birkett, S.J. Clark, S. Powers, N.J. Seal, L.J. Wadhams, a. J. Mordue, J. a. Pickett, Identification of human-derived volatile chemicals that interfere with attraction of *Aedes aegypti* mosquitoes, *J. Chem. Ecol.* 34 (2008) 308–322. doi:10.1007/s10886-008-9436-0.
- (7) A.M. Curran, C.F. Ramirez, A. a Schoon, K.G. Furton, The frequency of occurrence and discriminatory power of compounds found in human scent across a population determined

by SPME-GC/MS., *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* 846 (2007) 86–97. doi:10.1016/j.jchromb.2006.08.039.

(8) J.S. Brown, P.A. Prada, A.M. Curran, K.G. Furton, Applicability of emanating volatile organic compounds from various forensic specimens for individual differentiation., *Forensic Sci. Int.* 226 (2013) 173–182. doi:10.1016/j.forsciint.2013.01.008.

(9) L. Rust, K.D. Nizio, S.L. Forbes, The influence of ageing and surface type on the odour profile of blood-detection dog training aids, *Anal. Bioanal. Chem.* (2016). doi:10.1007/s00216-016-9748-9.

(10) J.O. Berger, T. Sellke, Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence, *J. Am. Stat. Assoc.* 82 (1987) 112–139. doi:10.1080/01621459.1987.10478397.

(11) C.G.G. Aitken, F. Taroni, *Statistics and the Evaluation of Evidence Second Edition*, John Wiley & Sons Ltd, 2004.

(12) A. Barcaru, G. Vivò-Truyols, Use of Bayesian statistics for pairwise comparison of megavariable data sets: extracting meaningful differences between GCxGC-MS chromatograms using Jensen-Shannon divergence, *Analytical Chemistry* 88 (2016), 2096–2014.

(13) A. Biedermann, F. Taroni, Bayesian networks for evaluating forensic DNA profiling evidence: a review and guide to literature. *Forensic Sci Int Genet* 6(2) (2012) 147–57.

(14) N. M. Haan, S. J. Godsill, Bayesian models for DNA sequencing. *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*; 2002 May 13-17; Orlando, FL. Piscataway, NJ:IEEE Xplore, 2011.

(15) T.B. Fitzpatrick, The validity and practicality of sun-reactive skin types I through VI., *Arch. Dermatol.* 124 (1988) 869–71.

(16) V. Cuzuel, E. Portas, G. Cognon, I. Rivals, F. Heulard, D. Thiébaud, J. Vial, Sampling method development and optimization in view of human hand odor analysis by thermal desorption coupled with gas chromatography and mass spectrometry., *Anal. Bioanal. Chem.* 409 (2017) 5113–5124. doi:10.1007/s00216-017-0458-8.

(17) V. Cuzuel, G. Cognon, I. Rivals, F. Heulard, D. Thiébaud, J. Vial, Human odor and forensics. Optimization of a comprehensive gas chromatography method based on orthogonality: how not to choose between criteria., *J. Chromatogr. A.* accepted (2017). doi:10.1016/j.chroma.2017.08.060.

(18) V. Cuzuel, Développement d'une stratégie de caractérisation chimique de la signature odorante d'individus par l'analyse chimiométrique de données issues de méthodes sépara-

tives bidimensionnelles., Thèse de doctorat de l'Université Pierre et Marie Curie, ED388.
<https://hal.archives-ouvertes.fr/tel-01667972>.

(19) H. Jeffreys, *Theory of probability*, Oxford University Press, Oxford, 1939.

(20) C. Robert, *The Bayesian Choice: from Decision-Theoretic Motivations to Computational Implementation*, Springer-Verlag, New York, 2001.

(21) H. E. Daniels, The relation between measures of correlation in the universe of sample permutation, *Biometrika*, Volume 33, Issue 2 (1944), 129-135, <https://doi.org/10.1093/biomet/33.2.129>.

(22) A.J. Hanley, J.B. McNeil, The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve, *Radiology*. 143 (1982) 29–36. doi:10.1148/radiology.143.1.7063747.

Figure Captions

Figure 1. GC×GC-MS chromatogram of a hand odor sample of a 25-year-old man.

Figure 2. PCA on the correlation matrix of the normalized intensities of the 600 compounds. The numbers refer to the identity of the subjects inside a set.

Figure 3. Bayesian identification using the Spearman correlation based distance on the normalized intensities of all 600 compounds. a) Histogram of the distances between chromatograms of the same subject (341 couples under H_0), and Gaussian mixture fit; b) Histogram of the distances between chromatograms of different subjects (33 329 couples under H_1), and Gaussian mixture fit; c) Fits and posterior probability of H_0 .

Figure 4. Bayesian identification using the correlation based distance on the binarized intensities of 82 selected compounds only. a) Histogram of the distances between chromatograms of the same subject (H_0), and Gaussian mixture fit; b) Histogram of the distances between chromatograms of different subjects (H_1), and Gaussian mixture fit; c) Fits and posterior probability of H_0 .

Figure 5. Bayesian identification using the Spearman correlation based distance on the normalized intensities of 146 selected compounds only. a) Histogram of the distances between chromatograms of the same subject (H_0), and Gaussian mixture fit; b) Histogram of the distances between chromatograms of different subjects (H_1), and Gaussian mixture fit; c) Fits and posterior probability of H_0 .

Figures

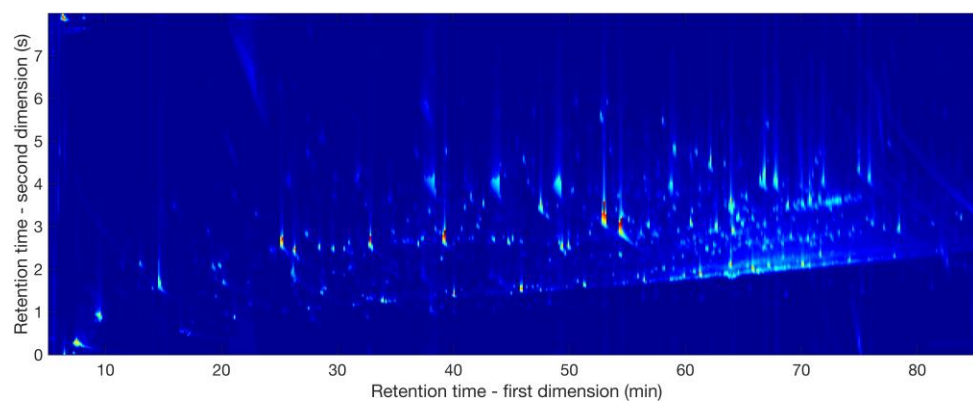


Figure 1. GCxGC-MS chromatogram of a hand odor sample of a 25-year-old man.

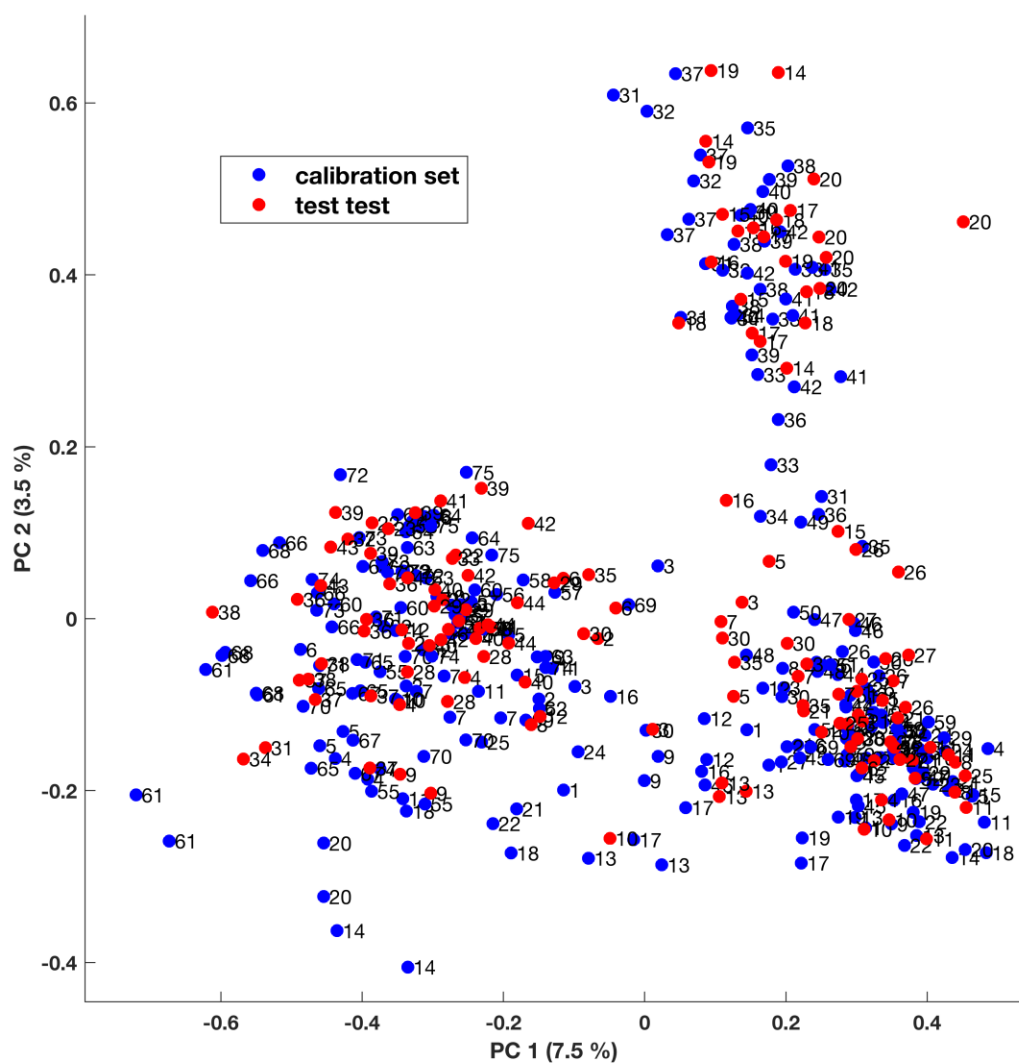


Figure 2. PCA on the correlation matrix of the normalized intensities of the 600 compounds. The numbers refer to the identity of the subjects inside a set.

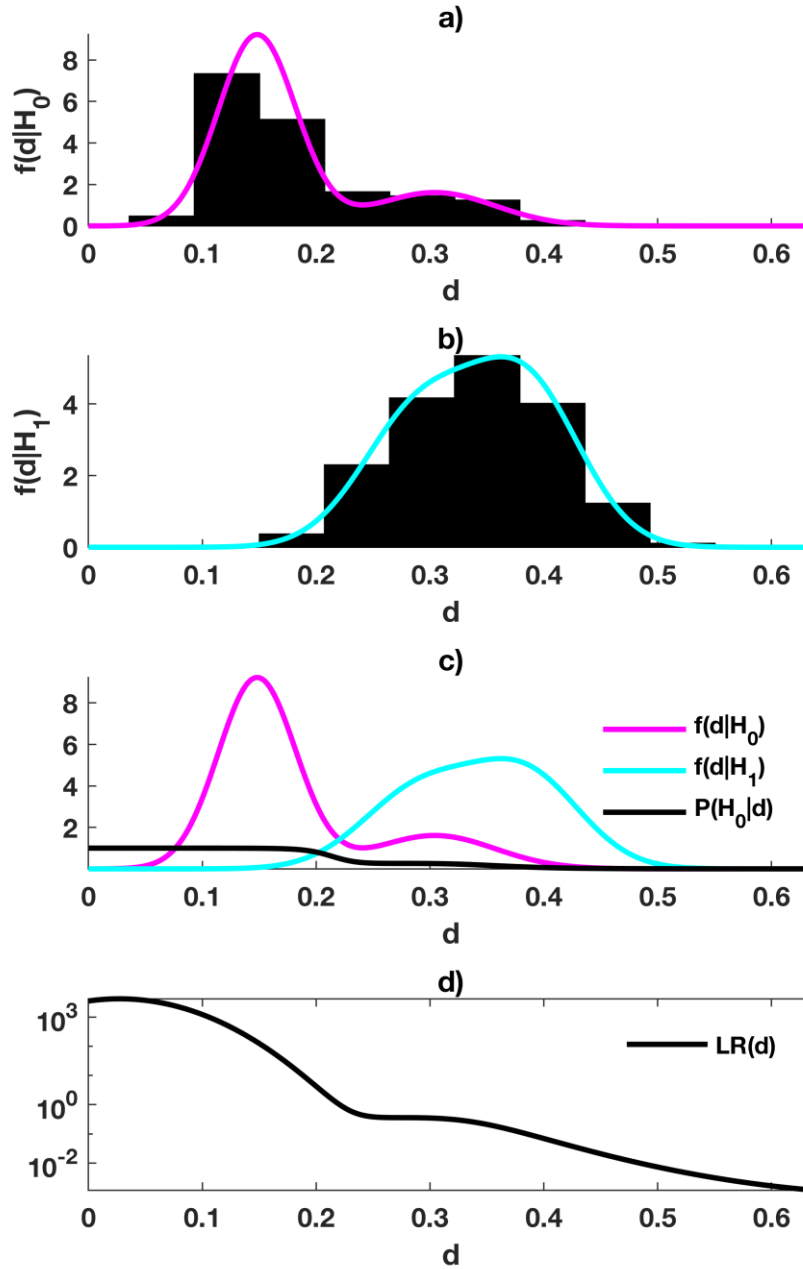


Figure 3. Bayesian identification using the Spearman correlation based distance on the normalized intensities of all 600 compounds. a) Histogram of the distances between chromatograms of the same subject (341 couples under H_0), and Gaussian mixture fit; b) Histogram of the distances between chromatograms of different subjects (33 329 couples under H_1), and Gaussian mixture fit; c) Fits and posterior probability of H_0 ; d) Likelihood ratio.

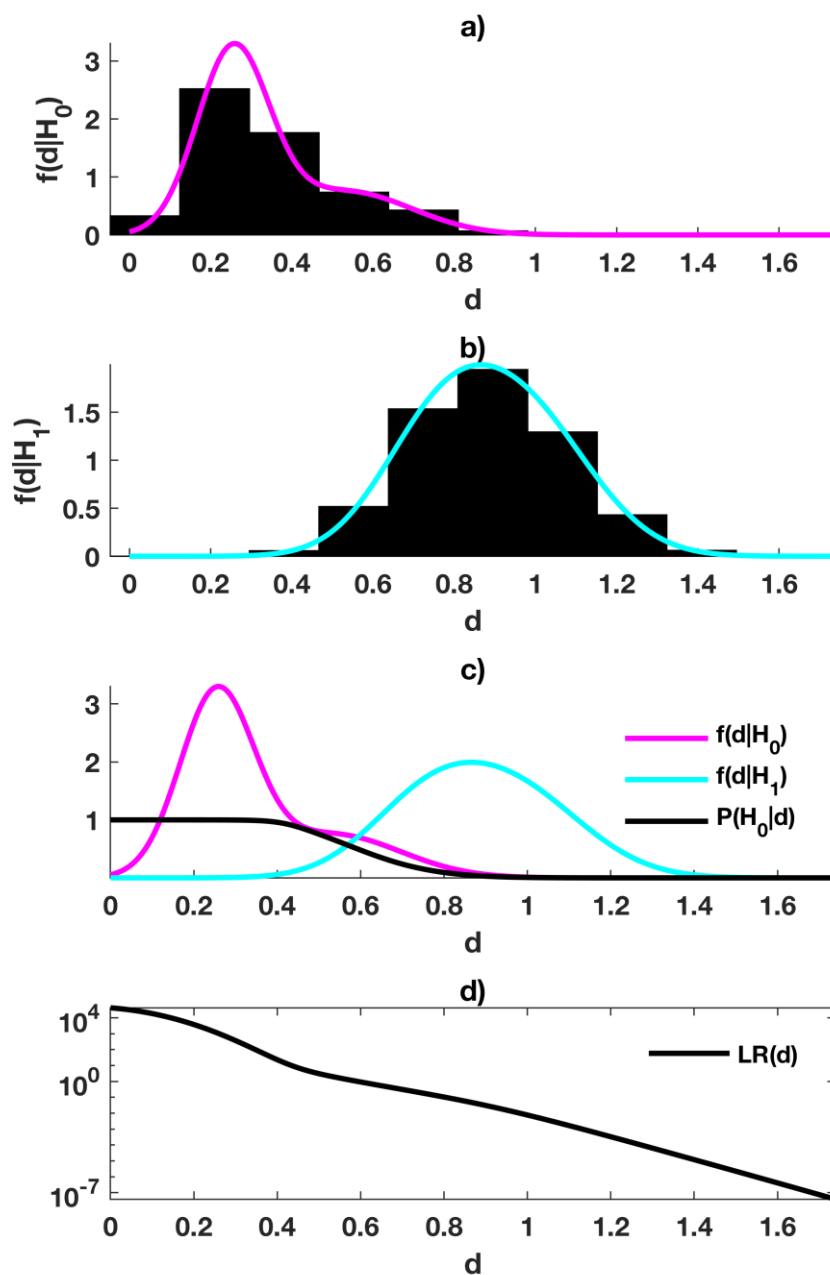


Figure 4. Bayesian identification using the correlation based distance on the binarized intensities of 82 selected compounds only. a) Histogram of the distances between chromatograms of the same subject (H_0), and Gaussian mixture fit; b) Histogram of the distances between chromatograms of different subjects (H_1), and Gaussian mixture fit; c) Fits and posterior probability of H_0 ; d) Likelihood ratio.

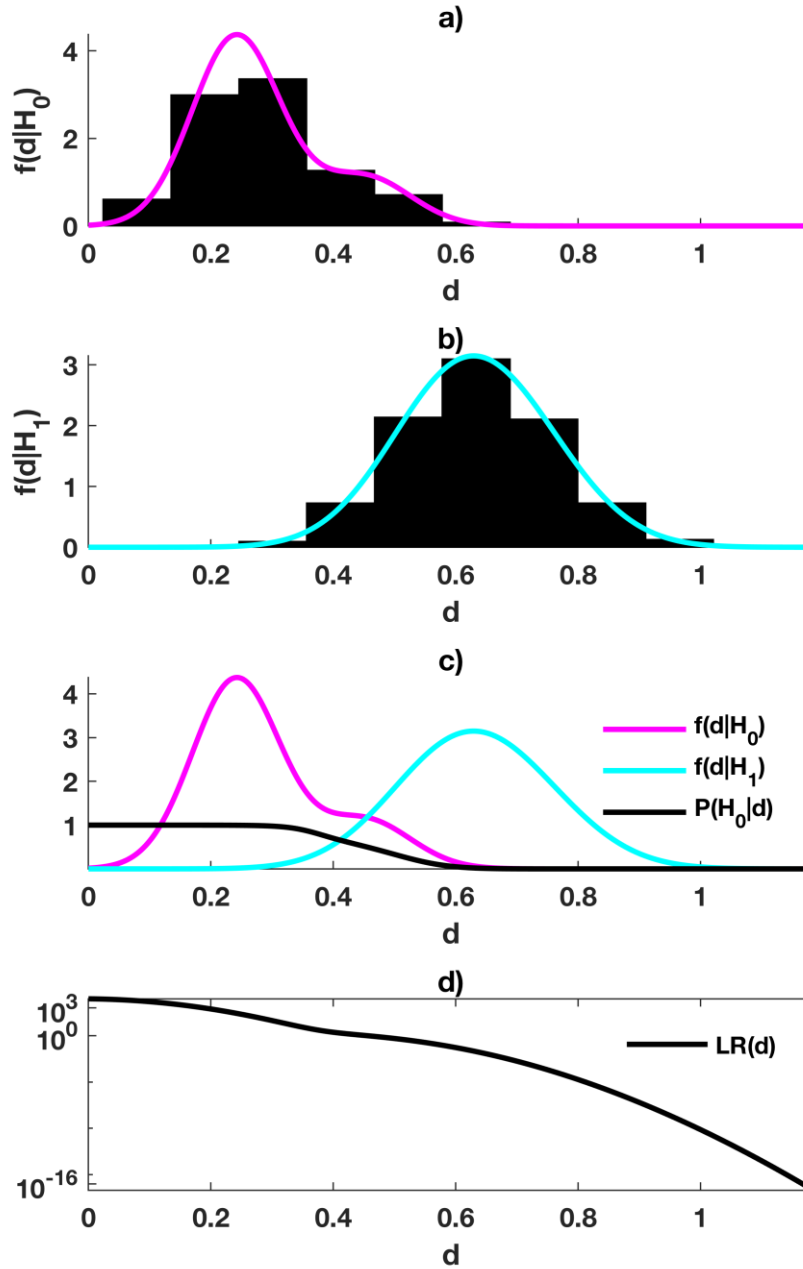


Figure 5. Bayesian identification using the Spearman correlation based distance on the normalized intensities of 146 selected compounds only. a) Histogram of the distances between chromatograms of the same subject (H_0), and Gaussian mixture fit; b) Histogram of the distances between chromatograms of different subjects (H_1), and Gaussian mixture fit; c) Fits and posterior probability of H_0 ; d) Likelihood ratio.

Tables

Table 1. Performance without compound selection, expressed as AUCs: the first percentage corresponds to the calibration set, the second one to the test set.

distance \ intensities	Euclidean	Pearson's	Spearman's
normalized	62.7% / 64.6%	74.6% / 74.7%	92.4% / 93.6%
binarized	88.4% / 91.6%	89.6% / 91.7%	

Table 2. Performance obtained with compound selection. The first figure corresponds to the threshold π on the test $-\log_{10}(p\text{-value})$ of Wilcoxon or Fisher's test, the second one to the number of selected compounds, the first percentage to the AUC on the calibration set, and the second one to the AUC on the test set.

distance \ intensities	Euclidean	Pearson's	Spearman's
normalized	$\pi=12$, 61 comp. 76.2% / 73.9%	$\pi=13$, 54 comp. 78.1% / 75.2%	$\pi=7$, 146 comp. 97.5% / 98.2%
binarized	$\pi=18$, 82 comp. 93.1% / 94.8%	$\pi=18$, 82 comp. 97.4% / 98.1%	