GC×GC-MS AND BAYESIAN TESTING IN FORENSICS: TOWARDS THE IDENTIFICATION OF SUSPECTS THROUGH THEIR ODOR

Isabelle RIVALS¹, Vincent CUZUEL², Guillaume COGNON², Roman Leconte, Didier THIEBAUT³, Charles SAULEAU², Jérôme VIAL³

¹Équipe de Statistique Appliquée, UMRS 1158, ESPCI Paris, France
²Institut de Recherche Criminelle de la Gendarmerie Nationale, Cergy-Pontoise, France
³UMR CBI 8231 - Laboratoire Sciences Analytiques, Bioanalytiques et Miniaturisation – ESPCI Paris – PSL Research University

- GCxGC 2018, Riva Del Garda, Italy
Popularization of the techniques used by the police

Criminals are more attentive and cautious!

Human odor

GCxGC-ISCC Riva del garda 2018
USE OF TRAINED DOGS

- Sufficient for identification of a person

- Limited probative value in courts of justice

- Need for corroborative evidence by analytical tools:
  - **Support the information** provided by dogs
  - **Probative value** to evidence in courts of justice
OBJECTIVES

Develop a global strategy to characterize the olfactory fingerprints of individuals using **analytical and statistical tools**

- Volatile compounds **at trace levels**: preconcentration step required
- **Complex mixtures**: multidimensional separation (GC×GC-MS)

Question to be answered

- Is the comparison of an “odor” reference chromatogram to a chromatogram obtained using an odor sample from a suspect (crime scene...) sufficient to prove that the odor belongs to the same person?
GLOBAL STRATEGY

SAMPLING/PANEL → SEPARATION AND DETECTION → DATA PROCESSING

P(A|B) = P(B|A) P(A) / P(B)

GCxGC-ISCC Riva del Garda 2018
Preconcentration and Analysis: Purge and Trap - GC×GC

**Thermodesorption coupled with GC×GC-MS**

VSP4000, Action Europe (Sausheim, France)

Direct sampling

Indirect sampling

Sample temperature = 190°C

Purge flow = 20 mL/min

Purge time = 20 min

Split = 0 mL/min

DB1MS-DB1701

2°C/min – modulation 8 s

Desorption optimization DC:

- **synthetic mixture** of human odor (80 compounds)
- **full factorial design**

Sample temperature = 190°C

Purge flow = 20 mL/min

Purge time = 20 min

Split = 0 mL/min

---

**CHROMATOGRAM OF A REAL SAMPLE**

- Shimadzu GC×GC/MS Q2010Plus
- Gradient: 2.5°C/min 40°C → 250°C

- Nonanal
- Decanal
- 5-hepten-2-one, 6-methyl
- 1,7-octanediol, 3,7-dimethyl
- 5,9-undecadien-2-one, 6,10-dimethyl (E)
- Phenol, p-tert-butyl
- Ethanol, 2-phenoxy
- Ethanol, 2-phenoxy
- Primary and secondary odor

1st dimension: DB1-MS (apolar)

**GC×GC-ISCC Riva del Garda 2018**
COMPARISON OF REAL SAMPLES?

- Complex samples
- Comparison is not trivial
- A lot of data to process

Need for an automated data processing to extract relevant information
Need for a panel of persons to evaluate the strategy

GCxGC-ISCC Riva del Garda 2018
Panel of 119 persons

<table>
<thead>
<tr>
<th>gender</th>
<th>age (years)</th>
<th>phototype</th>
</tr>
</thead>
<tbody>
<tr>
<td>total</td>
<td>♂ 10-23</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>♀ 24-36</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>♂ 37-81</td>
<td>3</td>
</tr>
<tr>
<td>119</td>
<td>61</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>58</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>15</td>
</tr>
</tbody>
</table>

• Phototype 1 - skin is sun sensitive and does not burn
• Phototype 2 - intermediate skin
• Phototype 3 – well tanning skin
• 4 direct samplings of hands/person (Sorb-star®)
  • 15 minutes
  • Blank (sampling room)
• TD* - GC×GC-MS**
• 3 chromatograms/person


**Cuzuel et al., Human odor and forensics. Optimization of a comprehensive gas chromatography method based on orthogonality: how not to choose between criteria., 2017, Journal of Chromatography A GCxGC-ISCC Riva del garda 2018
DATA PROCESSING / BAYESIAN APPROACH

GCxGC-ISCC Riva del Garda 2018
**Chromatographic Data Processing With Matlab**

1. **Export** incompatible format (mzXML)
2. **Import** of data to Matlab
3. **Conversion** of files
4. **Pre-treatment**
   - Baseline correction
   - Selection of investigated zones
5. **Detection of peaks**
   - Detection of local maxima
   - Extraction of associated informations (alkanes, $^1t$, $^2t$, LRI, MS spectrum, name)
6. **Transfer of data in the libraries**
   - Import to NIST and **own** library (3 persons) (>600 compounds)
   - **Library update** (known odor compounds*)
   - Identification and peak assignation
7. **Treatment using statistics**
   - Frequentist Approach
   - **Bayesian Approach**

---

Chromatogram ↔ 1 vector corresponding to 600 compounds peak intensity


GCxGC-ISCC Riva del Garda 2018
Bayesian Approach (a posteriori)

: the two chromatograms are obtained from the same person
: the two chromatograms are NOT obtained from the same person

D represent the observed data (the two chromatograms), Bayes formula gives:

\[ P(H_0 \mid D) = \frac{f(D \mid H_0)P(H_0)}{f(D \mid H_0)P(H_0) + f(D \mid H_1)P(H_1)} \]

Protocole:

Definition of a distance \textit{d} between 2 chromatograms \((D \equiv d)\)

Panel of chromatograms of individuals (119 persons sampled 4 times) splitted in independent calibration and test groups

Calibration group \(\Rightarrow\) estimation of distributions of \textit{d} for couples of chromatograms from the same person \(f(d \mid H_0)\) and from different persons \(f(d \mid H_1)\)

Test group \(\Rightarrow\) estimation of performance (AUC, sensitivity, spécificity)
**Bayesian Approach: Choice of Distance between Chrom**

**Estimation of the statistical likelihood**

Options:

- $d$: distances between 600-vectors of intensities:
  - a) euclidian distance
  - b) $1$ – Pearson correlation coefficient
  - c) $1$ – Spearman correlation coefficient
- Intensities normalized / binarized ($b=c$)

Calibration group (260 chromatograms / 75 persons)
- 341 couples of chromatograms for $H_0$ (same person)
- 33 329 couples de chromatograms for $H_1$ ( différents persons)

- **Histograms** of $d$ values for $H_0$ and $H_1$

Adjustment of histograms using several **gaussian curves**

- $f(d \mid H_0)$ and $f(d \mid H_1)$
• Probabilities \textit{a priori} : $P(H_0) = P(H_1) = 0.5$

\[ P(H_0 \mid d) = \frac{f(d \mid H_0)P(H_0)}{f(d \mid H_0)P(H_0) + f(d \mid H_1)P(H_1)} \]

• Fictitious examples of statistical likelihood

Bayesian Approach: Expected Results

Distance between chromatograms

Distance between chromatograms

GCxGC-ISCC Riva del garda 2018
# Bayesian Approach: Results Using 600 Compounds

(%AUC calibration / %AUC test)

<table>
<thead>
<tr>
<th>Distance</th>
<th>Euclidian</th>
<th>$1 - \rho_{\text{Pearson}}$</th>
<th>$1 - \rho_{\text{Spearman}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normalized</td>
<td>62.7% / 64.6%</td>
<td>74.6% / 74.7%</td>
<td>92.4% / 93.6%</td>
</tr>
<tr>
<td>Binarized</td>
<td>88.4% / 91.6%</td>
<td>89.6% / 91.7%</td>
<td></td>
</tr>
</tbody>
</table>

N.B. using the test group, there are 173 / 9 418 couples for $H_0 / H_1$ respectively.

2 modes!

GCxGC-ISCC Riva del Garda 2018
Discriminating compounds for $H_0$ and $H_1$: those which intensity differences $|\Delta i|$ are significantly lower for $H_0$ than $H_1$.

Quantification: $p$-value using unilateral Fisher test (binarized intensities) or Wilcoxon (normalized intensities) on $|\Delta i|$

Examples:

**Bayesian Approach: Discriminating Compounds**

- Octanoic acid, octyl ester: $pF = 8e^{-48}$
- Benzene, 1-ethyl-3-methyl-: $pW = 1e^{-31}$

GCxGC-ISC Riva del Garda 2018
**Bayesian Approach: Results using Discriminating Compounds**

(%AUC calibration / %AUC test)

Binarized

Normalized

Test bayésien : AUC 97.4% (cali.) / 98.1% (test)

Test bayésien : AUC 97.5% (cali.) / 98.2% (test)

GCxGC-ISCC Riva del Garda 2018
Bayesian Approach: Results using Discriminating Compounds

(\%AUC calibration / \%AUC test)

- **Threshold \( \pi \) value** \(-\log_{10}(p)\) of Fisher test (binarized intensities) or Wilcoxon (normalized intensities): optimized value obtained using cross validation (K=3) on calibration group

<table>
<thead>
<tr>
<th>distance intensities</th>
<th>euclidian</th>
<th>( 1 - \rho_{\text{Pearson}} )</th>
<th>( 1 - \rho_{\text{Spearman}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>normalized</td>
<td>( \pi = 12 / 61 ) comp.</td>
<td>( \pi = 13 / 54 ) comp.</td>
<td>( \pi = 7 / 146 ) comp.</td>
</tr>
<tr>
<td></td>
<td>76.2% / 73.9%</td>
<td>78.1% / 75.2%</td>
<td>97.5% / 98.2%</td>
</tr>
<tr>
<td>binarized</td>
<td>( \pi = 18 / 82 ) comp.</td>
<td>( \pi = 18 / 82 ) comp.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>93.1% / 94.8%</td>
<td>97.4% / 98.1%</td>
<td></td>
</tr>
</tbody>
</table>

GCxGC-ISCC Riva del Garda 2018
Discussion
(%AUC calibration / %AUC test)

Performances

<table>
<thead>
<tr>
<th>intensities</th>
<th>AUC</th>
<th>sensitivity</th>
<th>specificity</th>
<th>nb. compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>binarized</td>
<td>97.4% / 98.1%</td>
<td>89.4% / 90.0%</td>
<td>94.9% / 92.5%</td>
<td>82</td>
</tr>
<tr>
<td>normalized</td>
<td>97.5% / 98.2%</td>
<td>89.1% / 85.9%</td>
<td>93.7% / 95.0%</td>
<td>146</td>
</tr>
</tbody>
</table>

Adequate distance ⇒ quantitative exploitation of compounds intensities despite the analytical variability

Selection ⇒ second modes of $f(d|H_0)$ et $f(d|H_1)$ are strongly decreased ⇒ better results

Binarized : more parsimonious (82/146 compounds to be used)

57 common compounds for both classifiers

Nota bene

same direct samples
no pollution by other odors
CONCLUSION AND PERSPECTIVES

Direct/non direct sampling procedures for human (hand) odor analyses

Comprehensive GC×GC-MS method and data (ToF)

Validation of procedures in the field with dog handlers

Large Panel of individuals to test the model

Storage of samples: standardized procedure

Data processing in progress for real application

✓ Different samples (direct or not...) and sampling conditions

✓ Study of discriminating compounds

✓ Normalization on discriminating compounds, more complex distance...

The final answer to the question must be YES or NO not 98.2%

GCxGC-ISCC Riva del Garda 2018
THANK YOU FOR YOUR ATTENTION!