



Automated identification of illicit substances by chemiometrics. Application to FTIR and RAMAN analysis.

Vesna Simic, Jacques-Henri Sublemontier, Frédérick Carrel, Loïc Paillat, Arthur Gourain

► To cite this version:

Vesna Simic, Jacques-Henri Sublemontier, Frédérick Carrel, Loïc Paillat, Arthur Gourain. Automated identification of illicit substances by chemiometrics. Application to FTIR and RAMAN analysis.. The 5th international conference on research and innovation, May 2022, Lille, France. , 2022. cea-03963564

HAL Id: cea-03963564

<https://cea.hal.science/cea-03963564>

Submitted on 30 Jan 2023

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.

Simic V.^{1*}, Sublemontier J. H.¹, Boudergui K.¹, Carrel F.¹, Paillat L.², Gourain A.²¹ Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France;² LCPP, Division Intervention et Enquête sur Site, Intervention spécialisée et Développement Capacitaire, Paris, France

1. Context and motivation

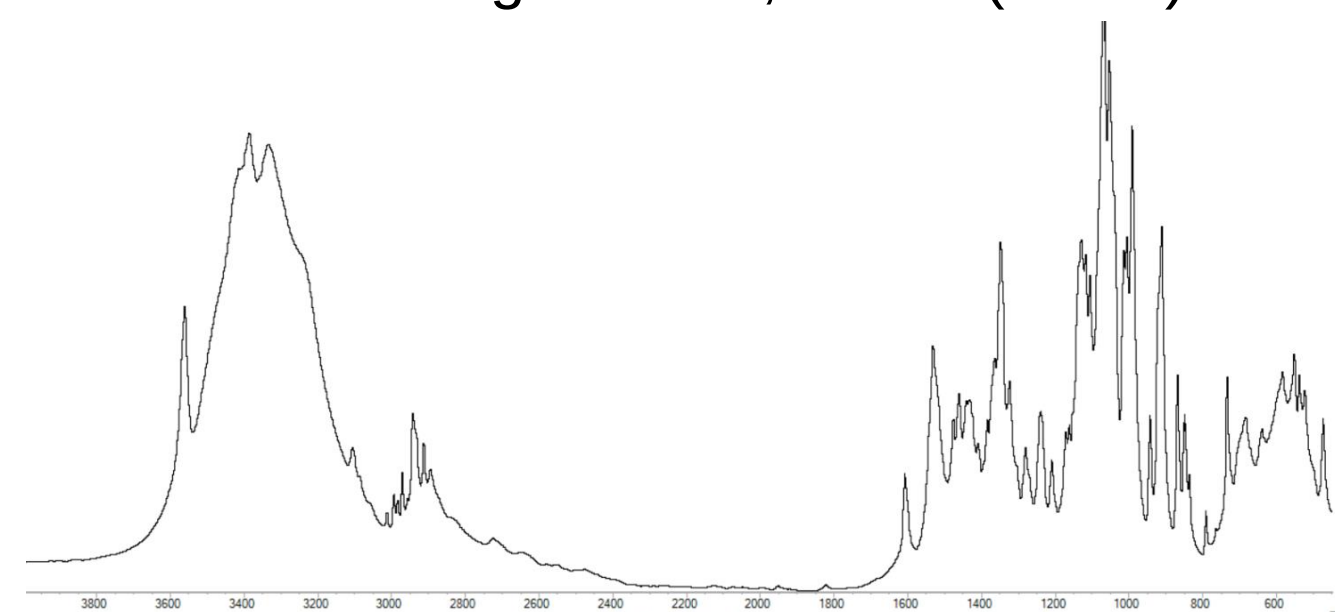
→ First steps of field investigation and analysis are largely based on robust and handheld analytical techniques such as FTIR or Raman spectrometry, etc. (fig. 1).

→ Removal of doubts is performed by comparing unknown spectra with reference spectra (database) and if necessary, cross-referencing the result of different analytical tools

→ **Automated techniques currently in use** are based on **spectral library searching with matching algorithms**.

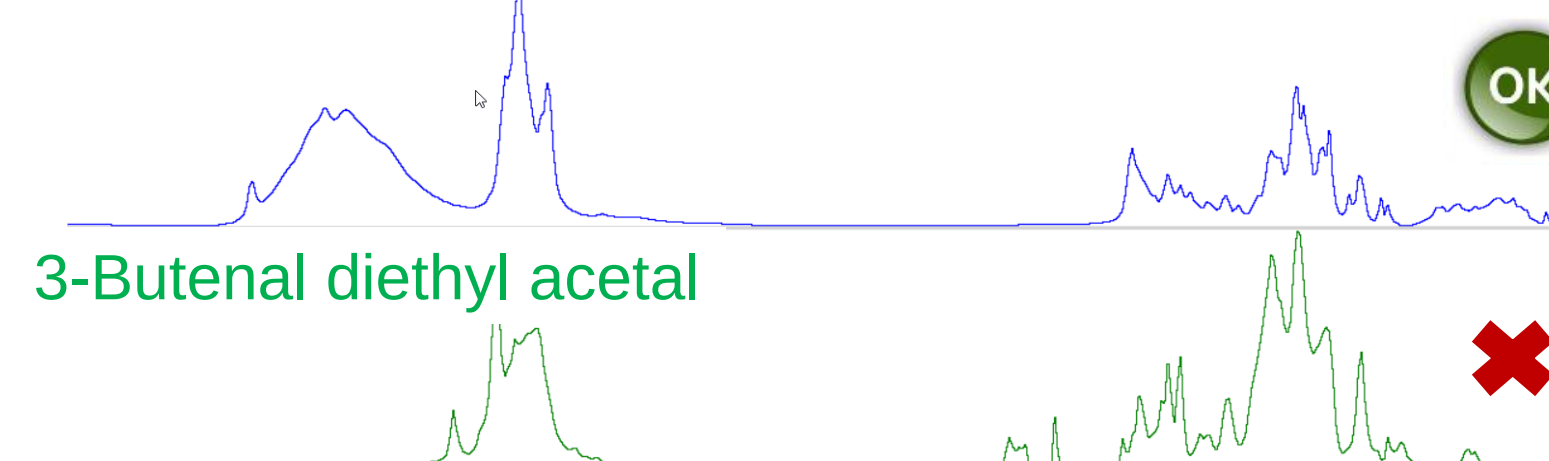
→ **These methods are limited** when searching target substances (drugs, explosives, chemical agent) in mixtures and even more if these substances are present in low concentrations.

Mixture of cane sugar and 2,4-DNT (88/22)



Correlation
Euclidean Distance
Absolute Value
1st Deriv Absolute Value
Least Squares
1st Deriv Least Squares
1st Deriv Correlation
Example of spectral searching result

Sucrose



(66/33)

OK

3-Butenal diethyl acetal

X

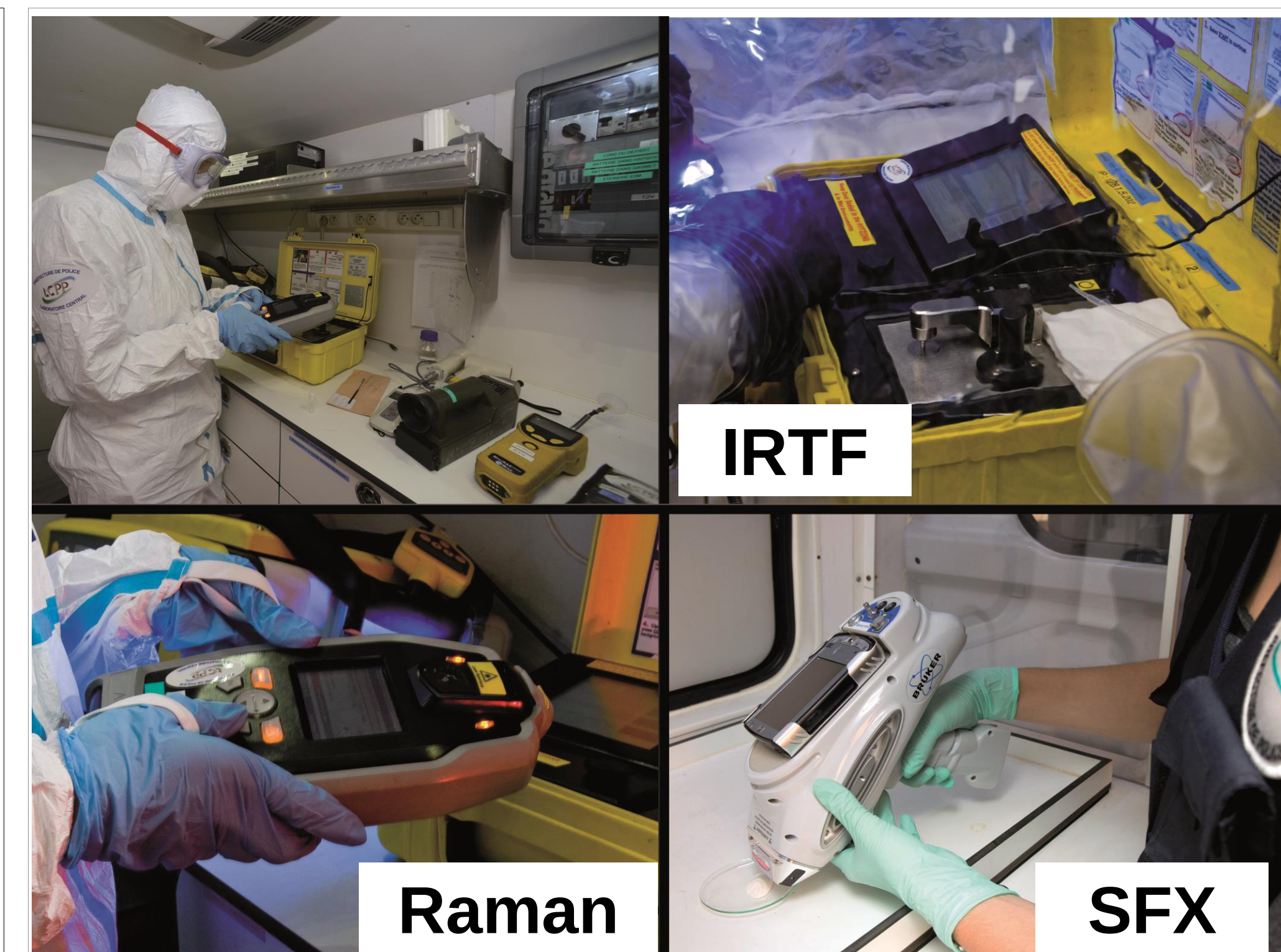


Figure 1 : Équipements analytiques du laboratoire mobile du Laboratoire central

© préfecture de police

We propose hereafter a new approach based on training data generation (synthetic mixtures) and data separation algorithms to automatically identify threat signatures in unknown spectra

2. Methodology and main achievements

Unknown spectra analysis

Automated Threat Identification



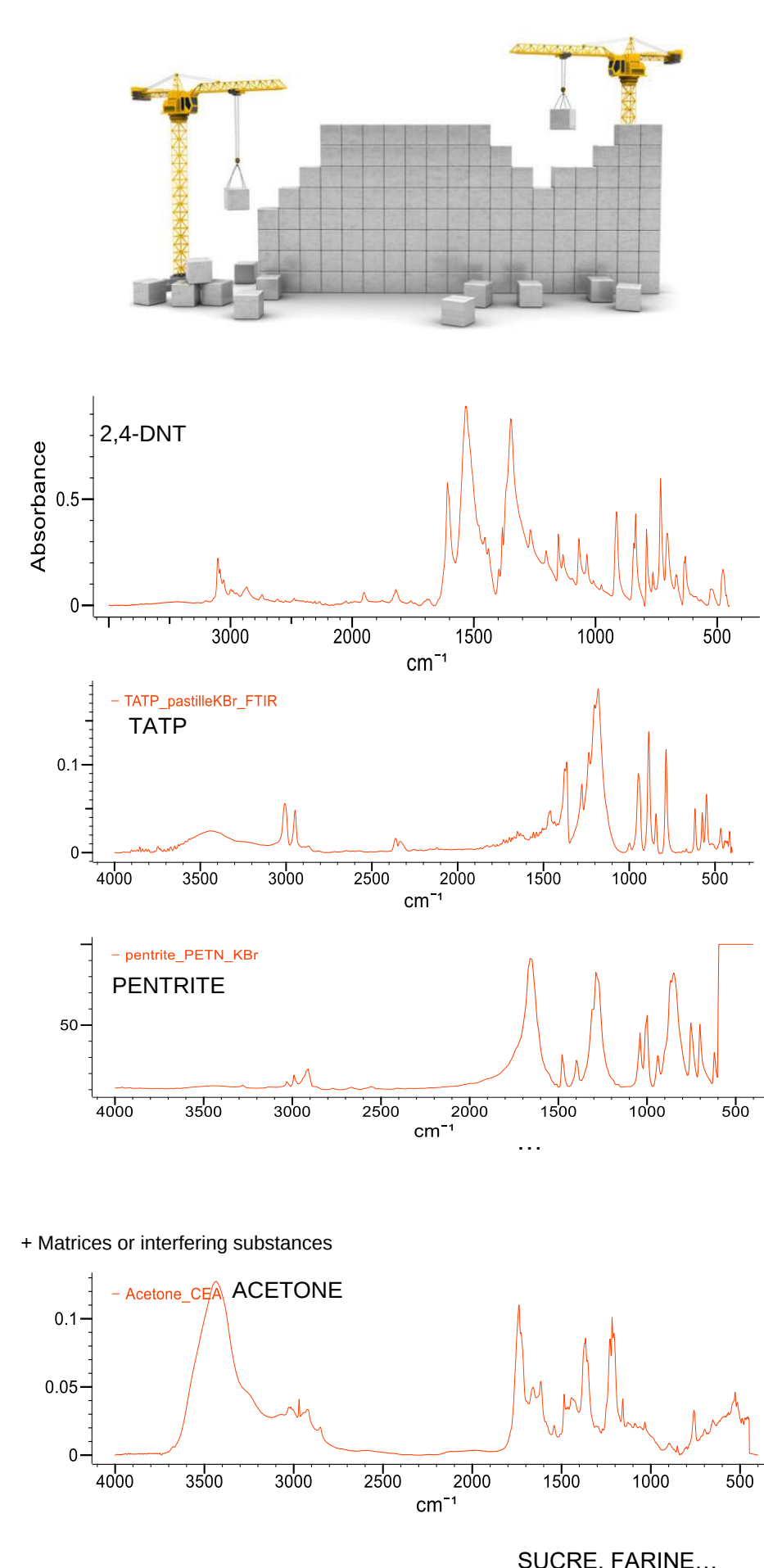
Compounds library

Synthetic mixtures generation

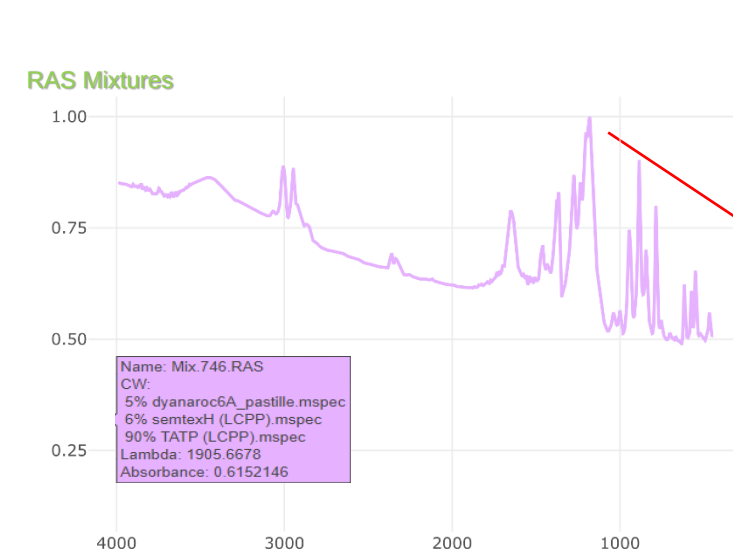
Features extraction

Models fitting

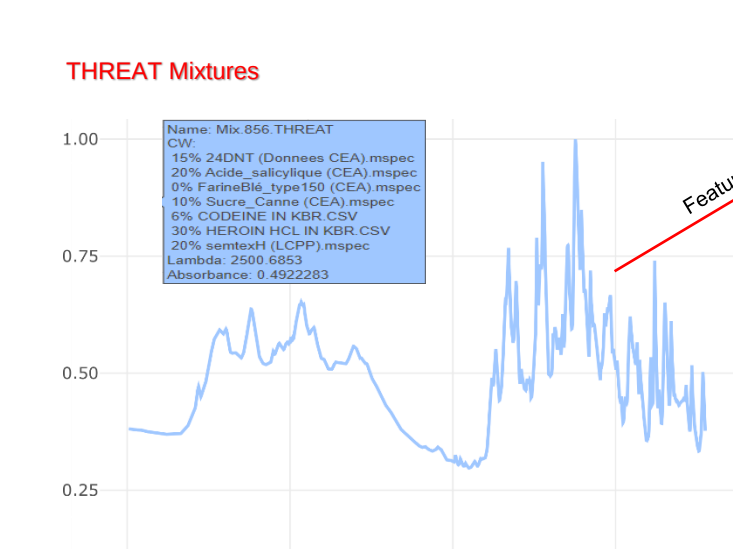
Models assessment



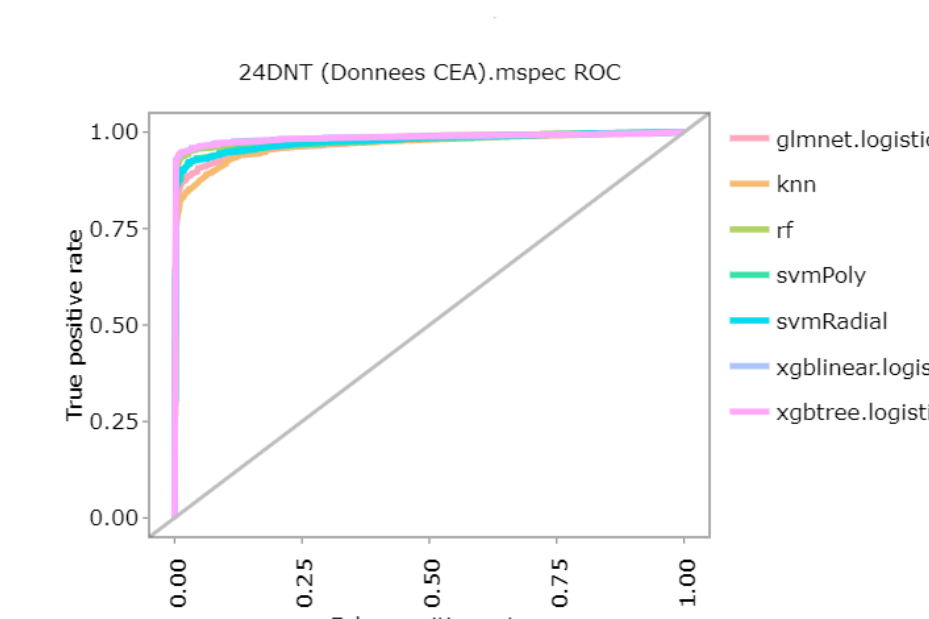
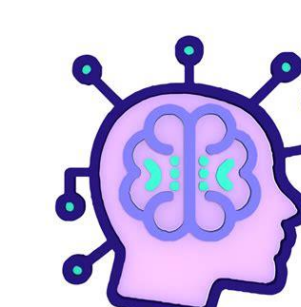
Example of mixture without 2,4-DNT



Example of mixture containing 2,4-DNT



0.86 0.14 0.98
-0.43 0.01 0.55
0.9 -0.6 0.95



		Condition	
		THREAT	RAS
Prédiction	THREAT	TP	FP
	RAS	FN	TN

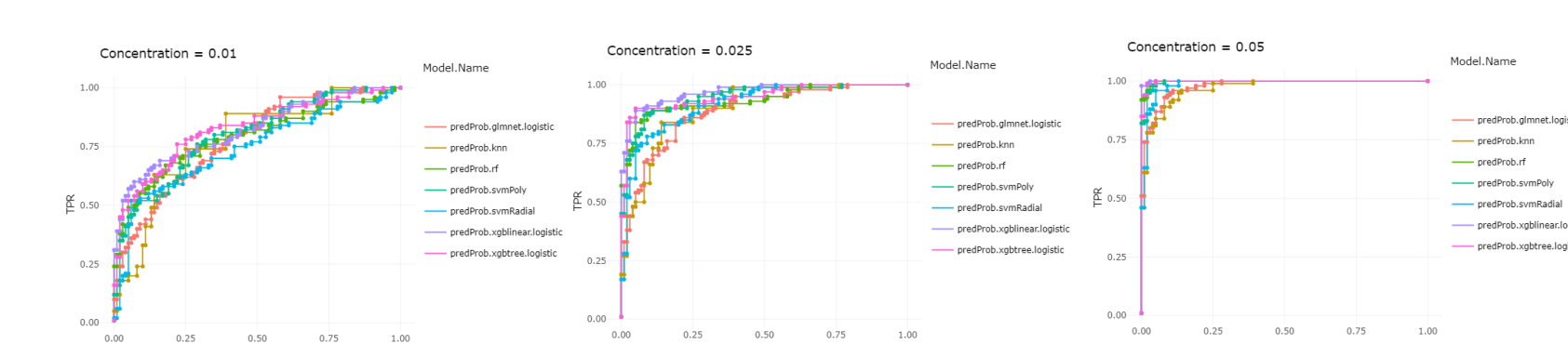
$$TPR = \frac{TP}{TP + FN}$$
$$FPR = \frac{FP}{FP + TN}$$

Complexity of synthetic mixes

1. Compounds.per.mix.expected = 4
2. Abundance.profile.expected = 2
3. Target.threat.presence.probability = 0.4
4. Number.of.mixtures = 1000

Diversity of models (extendable)

- ✓ SVM
- ✓ kNN
- ✓ GLMs
- ✓ Random Forest
- ✓ Gradient boosting



3. Conclusions and Perspectives

The new tool under development allows the :

- generation of complex mixtures parametrized by 3 variables and a noise.
- training of different models
- performance assessments of the different supervised learning algorithms
 - on synthetic mixtures (200 inputs min), complex mixtures (quaternary mixtures on average)
 - at different concentrations of the target threat
 - on real spectra (pure threats/interfering substances and unknown mixtures)

These new algorithms are compatible with different analytical techniques that could be combined to improve decision making process.

CONTACTS | vesna.simic@cea.fr;
jacques-henri.sublemontier@cea.fr