

# Détecter les nouvelles pollutions de l'air

Aurore Guillevic

Université de Lorraine, CNRS, Inria, LORIA, Nancy, France

Cigognes, Ramonchamp, 24 octobre 2023



<https://members.loria.fr/AGuillevic/files/talks/23-cigognes.pdf>

[https://members.loria.fr/AGuillevic/files/talks/sac\\_a\\_dos.py](https://members.loria.fr/AGuillevic/files/talks/sac_a_dos.py)

# Parcours

<https://members.loria.fr/AGuillevic/>

- 2005 : bac Scientifique (maths, physique-chimie, SVT) spécialité maths

# Parcours

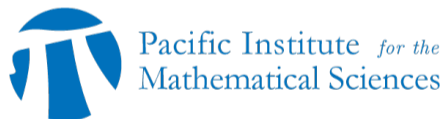
<https://members.loria.fr/AGuillevic/>

- 2005 : bac Scientifique (maths, physique-chimie, SVT) spécialité maths
- 2005–2008 : licence de mathématiques et informatique  
Université de Bretagne Sud, Lorient, Vannes
- 2008–2010 : master de mathématiques appliquées à la cryptographie  
Sciences des codes secrets

# Parcours

Ensuite on est rémunéré, d'abord comme stagiaire puis en CDD :

- 2010 : 6 mois de stage chez Thales Communications, Colombes, Hauts-de-Seine
- 2010–2013 : thèse de doctorat en alternance (CIFRE), Thales et École Normale Supérieure, Paris
- 2014, 2015, 2016 Post-doctorat, dont 8 mois au Canada



# Un algorithme pour détecter la pollution de l'air

Collaboration en 2018–2021 avec des physiciens de Zurich en Suisse







Détecter de nouvelles pollutions de l'air

1. Un peu de physique et chimie
2. Un peu d'algorithmique





# L'air

l'Atmosphère (air sec, sans la vapeur d'eau ) est composé de

- 78% 
- 21% 
- 0,9% 
- 0,041% 
- autres gaz à l'état de traces

# L'air

l'Atmosphère (air sec, sans la vapeur d'eau ) est composé de

- 78%  diazote  $N_2$
- 21%  dioxygène  $O_2$
- 0,9%  argon Ar
- 0,041%  dioxyde de carbone  $CO_2$
- autres gaz à l'état de traces

# Des atomes, des molécules

Hydrogène, Carbone, Azote, Oxygène, Fluor, Phosphore, Soufre, Chlore



H



C



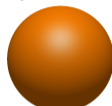
N



O



F



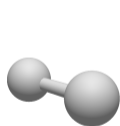
P



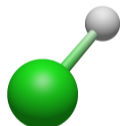
S



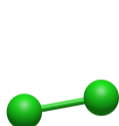
Cl



H<sub>2</sub>



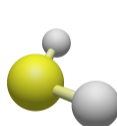
HCl



Cl<sub>2</sub>



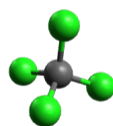
H<sub>2</sub>O



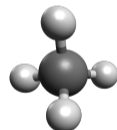
H<sub>2</sub>S



NH<sub>3</sub>



CCl<sub>4</sub>



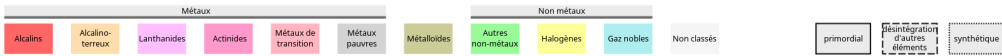
CH<sub>4</sub>



# Tableau périodique des éléments chimiques

Gruppe	1	II A	III A	IV A	V A	VI A	VII A	VIII	IX	X	XI	II B	III B	IV B	V B	VI B	VII B	18
1	Hydrogène <b>1</b> H 1,007975																	Hélium <b>2</b> He 4,002602
2	Lithium <b>3</b> Li 6,9395	Béryllium <b>4</b> Be 9,0121831											Bore <b>5</b> B 10,8135	Carbone <b>6</b> C 12,0106	Azote <b>7</b> N 14,006855	Oxygène <b>8</b> O 15,9994	Fluor <b>9</b> F 18,99840316	Néon <b>10</b> Ne 20,1797(6)
3	Sodium <b>11</b> Na 22,98976928	Magnésium <b>12</b> Mg 24,3055											Aluminium <b>13</b> Al 26,9815385	Silicium <b>14</b> Si 28,085(1)	Phosphore <b>15</b> P 30,97376200	Soufre <b>16</b> S 32,0675	Chlore <b>17</b> Cl 35,4515	Argon <b>18</b> Ar 39,948(1)
4	Potassium <b>19</b> K 39,0983(1)	Calcium <b>20</b> Ca 40,078(4)	Scandium <b>21</b> Sc 44,955908(5)	Titane <b>22</b> Ti 47,867(1)	Vanadium <b>23</b> V 50,9415(1)	Chrome <b>24</b> Cr 51,9961(6)	Manganèse <b>25</b> Mn 54,938044	Fer <b>26</b> Fe 55,845(2)	Cobalt <b>27</b> Co 58,933194	Nickel <b>28</b> Ni 58,6934(4)	Cuivre <b>29</b> Cu 63,546(3)	Zinc <b>30</b> Zn 65,38(2)	Gallium <b>31</b> Ga 69,723(1)	Germanium <b>32</b> Ge 72,630(8)	Arsenic <b>33</b> As 74,921595	Sélénium <b>34</b> Se 78,971(8)	Brome <b>35</b> Br 79,904	Krypton <b>36</b> Kr 83,798(2)
5	Rubidium <b>37</b> Rb 85,4678(3)	Strontium <b>38</b> Sr 87,62(1)	Yttrium <b>39</b> Y 88,90584	Zirconium <b>40</b> Zr 91,224(2)	Niobium <b>41</b> Nb 92,90637	Molybdène <b>42</b> Mo 95,95(1)	Technétium <b>43</b> Tc [98]	Ruthénium <b>44</b> Ru 101,07(2)	Rhodium <b>45</b> Rh 102,90550	Palladium <b>46</b> Pd 106,42(1)	Argent <b>47</b> Ag 107,8682(2)	Cadmium <b>48</b> Cd 112,414(4)	Indium <b>49</b> In 114,818(1)	Étain <b>50</b> Sn 118,710(7)	Antimoine <b>51</b> Sb 121,760(1)	Tellure <b>52</b> Te 127,60(3)	Iode <b>53</b> I 126,90447	Xénon <b>54</b> Xe 131,293(6)
6	Césium <b>55</b> Cs 132,905452	Baryum <b>56</b> Ba 137,327(7)	Lanthanides <b>57-71</b>	Hafnium <b>72</b> Hf 178,49(2)	Tantale <b>73</b> Ta 180,94788	Tungstène <b>74</b> W 183,84(1)	Rhénium <b>75</b> Re 186,207(1)	Osmium <b>76</b> Os 190,23(3)	Iridium <b>77</b> Ir 192,217(3)	Platine <b>78</b> Pt 195,084(6)	Or <b>79</b> Au 196,966569	Mercury <b>80</b> Hg 200,592(3)	Thallium <b>81</b> Tl 204,3835	Plomb <b>82</b> Pb 207,2(1)	Bismuth <b>83</b> Bi 208,98040	Polonium <b>84</b> Po [209]	Astate <b>85</b> At [210]	Radon <b>86</b> Rn [222]
7	Francium <b>87</b> Fr [223]	Radium <b>88</b> Ra [226]	Actinides <b>89-103</b>	Rutherfordium <b>104</b> Rf [267]	Dubnium <b>105</b> Db [268]	Seaborgium <b>106</b> Sg [269]	Bohrium <b>107</b> Bh [270]	Hassium <b>108</b> Hs [277]	Méitnérium <b>109</b> Mt [278]	Darmstadtium <b>110</b> Ds [281]	Roentgenium <b>111</b> Rg [282]	Copernicium <b>112</b> Cn [285]	Nihonium <b>113</b> Nh [286]	Flerovium <b>114</b> Fl [289]	Moscovium <b>115</b> Mc [289]	Livermorium <b>116</b> Lv [293]	Tennessee <b>117</b> Ts [294]	Oganesson <b>118</b> Og [294]
				Lanthane <b>57</b> La 138,90547	Cérium <b>58</b> Ce 140,116(1)	Praséodyme <b>59</b> Pr 140,90766	Néodyme <b>60</b> Nd 144,242(3)	Prométhium <b>61</b> Pm [145]	Samarium <b>62</b> Sm 150,36(2)	Europium <b>63</b> Eu 151,964(1)	Gadolinium <b>64</b> Gd 157,25(3)	Terbium <b>65</b> Tb 158,92535	Dysprosium <b>66</b> Dy 162,500(1)	Holmium <b>67</b> Ho 164,93033	Erbium <b>68</b> Er 167,259(3)	Thulium <b>69</b> Tm 168,93422	Ytterbium <b>70</b> Yb 173,045	Lucétium <b>71</b> Lu 174,9668
				Actinium <b>89</b> Ac [227]	Thorium <b>90</b> Th 232,0377	Protactinium <b>91</b> Pa 231,03588	Uranium <b>92</b> U 238,02891	Neptunium <b>93</b> Np [237]	Plutonium <b>94</b> Pu [244]	Américium <b>95</b> Am [243]	Curium <b>96</b> Cm [247]	Berkélium <b>97</b> Bk [247]	Californium <b>98</b> Cf [251]	Einsteinium <b>99</b> Es [252]	Fermium <b>100</b> Fm [257]	Mendelevium <b>101</b> Md [258]	Nobelium <b>102</b> No [259]	Lawrencium <b>103</b> Lr [266]

← nom de l'élément (**gaz**, **liquide** ou **solide** à 0°C et 101,3 kPa)  
 ← numéro atomique  
 ← symbole chimique  
 ← masse atomique relative [ou celle de l'isotope le plus stable]  
 © [ CIAAW "Atomic Weights 2013" + rev. 2015 ]



Par Scaler, Michka B - Travail personnel, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=8985780>

# La pollution de l'air

Combustion incomplète : Monoxyde de Carbone CO

NO<sub>x</sub> (NO, NO<sub>2</sub>) oxydes d'azote

particules fines pm 10, pm 2.5 taille inférieure à 10 microns, 2,5 microns

abrasion des freins, usure des pneus sur la route

suies de fours et cheminées

Pollutions très dangereuses : les dioxynes (avec du Benzène)

Usines et sites "Seveso" : catastrophe industrielle en Italie, 1976

- Lubrizol à Rouen en 2019
- Beyrouth en 2020 (explosion du port)
- Leverkusen en 2021, incendie d'entrepôts chimiques



# Détecter la pollution

## Mesure de concentrations de molécules connues

Airparif, AtMO Grand Est *surveillance de certaines pollutions*

pollution due au trafic routier et à l'activité humaine : Ozone O<sub>3</sub>, dioxyne d'azote NO<sub>2</sub>, dioxyde de soufre SO<sub>2</sub>, particules fines

## Détections de nouvelles molécules connues

Les nouvelles molécules connues (sur catalogue) sont recherchées dans l'air pour détecter leurs émissions par l'industrie

*surveillance des suspects* réseau international NOAA <https://www.noaa.gov/>

## Recherche en aveugle de pollutions inconnues

search for unknown unknowns

cas des catastrophes industrielles

Doit-on évacuer la population ?

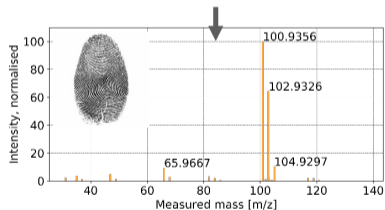
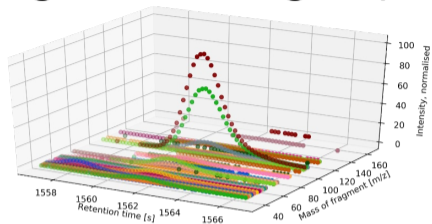
Les pompiers peuvent-ils intervenir, avec quel équipement ?

Production agricole contaminée ?

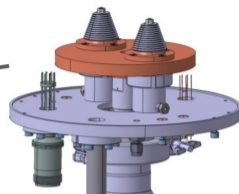
# How to search for unknown unknowns?



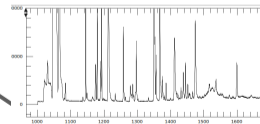
# Target screening : Aprecon – GC – ToF-MS



**Electron impact (EI)**  
**Time-of-Flight**  
**Mass spectrometer**  
**Tofwerk AG**

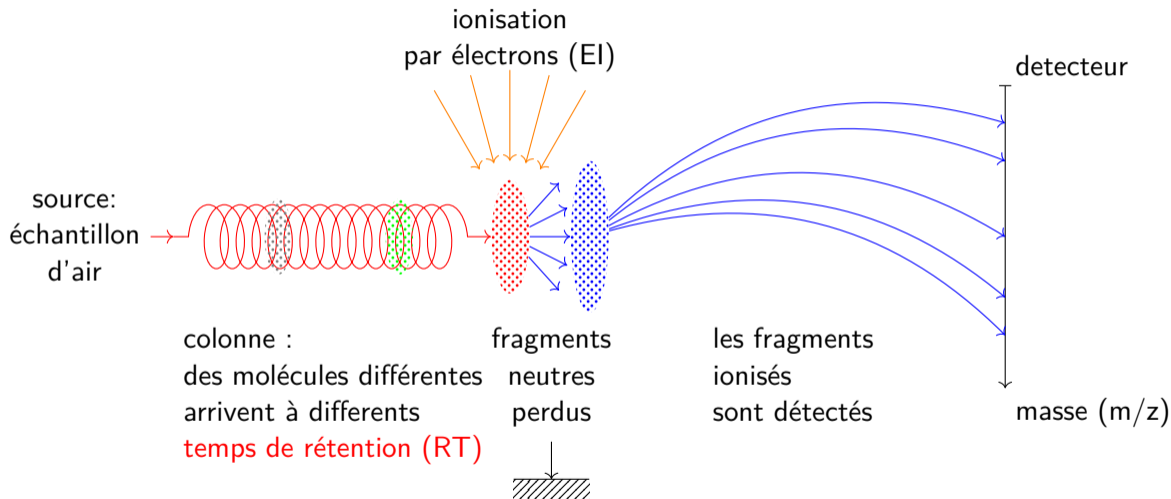


**Pre-concentration**  
**(APRECON)**

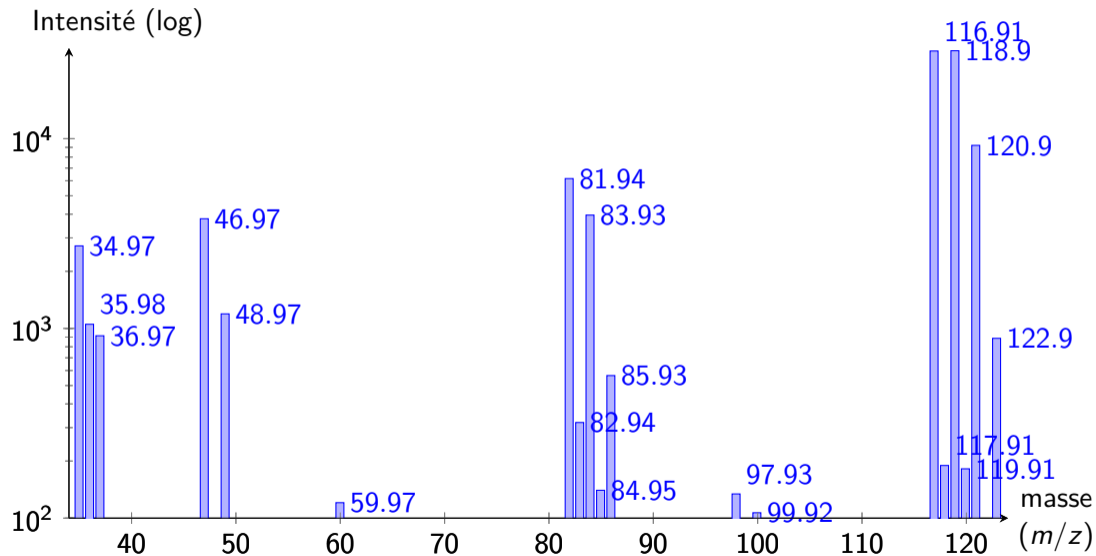


**Gas Chromatography**  
**GasPro column**

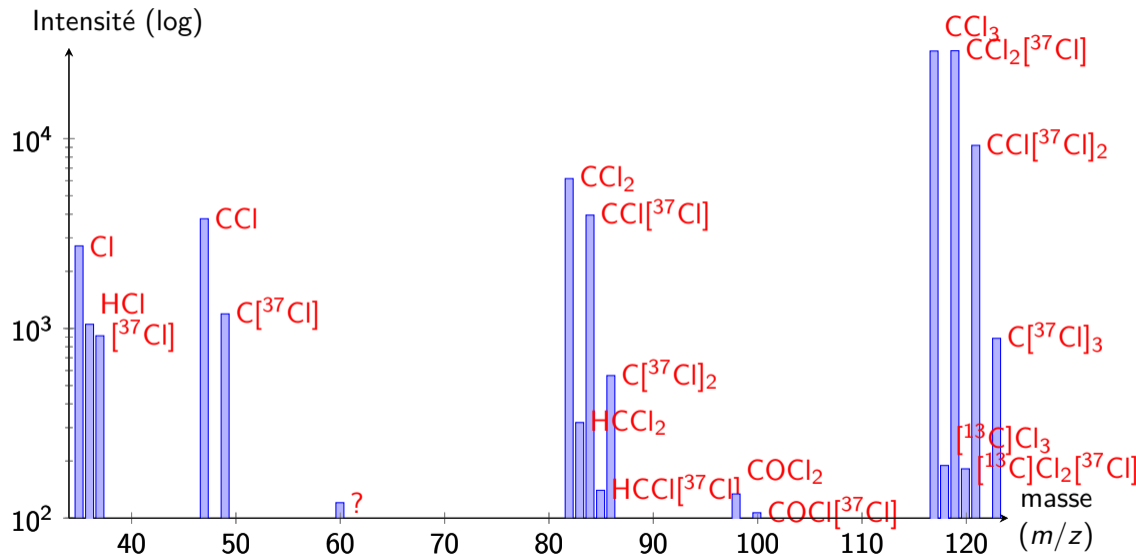
# Spectromètre de masse à temps de vol et ionisation par électrons



# Données d'entrée : un spectre de masse



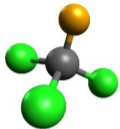
## But : annoter la figure



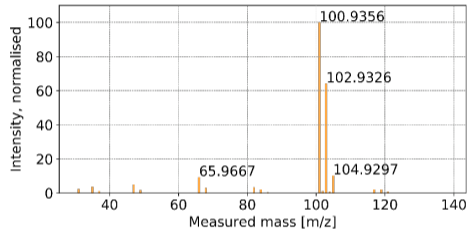


# Target vs non-target screening

Target screening:



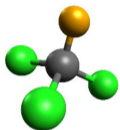
CFC-11  
CFCl3



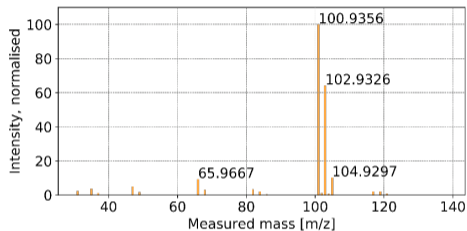
Instrumental fingerprint

# Target vs non-target screening

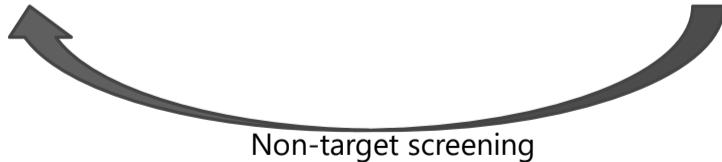
Target screening:



CFC-11  
CFCl3



Instrumental fingerprint



# Workflow: Knapsack algorithm



Weight:  $100.936 \pm 0.0002$  g/mol  
(U = 2 ppm)

Which **atoms** can be packed together to match the measured masses,  $\pm u$ ?

9 atoms: H, C, N, O, S, F, Cl, Br, I

## Un exemple : trouver la molécule inconnue

Données d'entrée :

- masse détectée :  $\approx 16$  m/z
- liste des masses IUPAC des atomes arrondi : Hydrogène H 1, Bore B 11, Carbone C 12, Azote N 14, Oxygène O 16.

Quelles possibilités ? **Combinatoire**

H	1.0078250319
B	11.00930536
C	12.
N	14.0030740074
O	15.9949146223
F	18.99840316
P	30.973762
S	31.97207073
Cl	34.96885271
Br	78.9183376
I	126.9044719

## Un exemple : trouver la molécule inconnue

Données d'entrée :

- masse détectée :  $\approx 16$  m/z
- liste des masses IUPAC des atomes arrondi : Hydrogène H 1, Bore B 11, Carbone C 12, Azote N 14, Oxygène O 16.

Quelles possibilités ? **Combinatoire**

- $O \approx 16$  m/z
- $NH_2 \approx 14 + 2 = 16$  m/z
- $CH_4 \approx 12 + 4 = 16$  m/z
- $BH_5 \approx 11 + 5 = 16$  m/z
- $H_{16} \approx 16$  m/z

Mais c'est "en théorie" !


H	1.0078250319
B	11.00930536
C	12.
N	14.0030740074
O	15.9949146223
F	18.99840316
P	30.973762
S	31.97207073
Cl	34.96885271
Br	78.9183376
I	126.9044719

## Informations supplémentaires

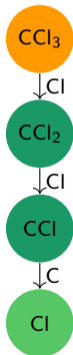
- Les atomes ont une **valence** : une propension à former des molécules ou pas  
BH<sub>5</sub> et H<sub>16</sub> sont chimiquement impossibles
- Les éléments ont des **isotopes** : connaissez-vous le Carbone 14 ?
- Les spectres de masse sont issus de la fragmentation de molécules identiques :  
théorie des graphes

Hiérarchiser les informations

# Algorithmes de graphes pour $\text{CCl}_4$

 tout seul,  Maximal,  Nœud,  feuille.

Fragmentation  
graph of  $\text{CCl}_4$

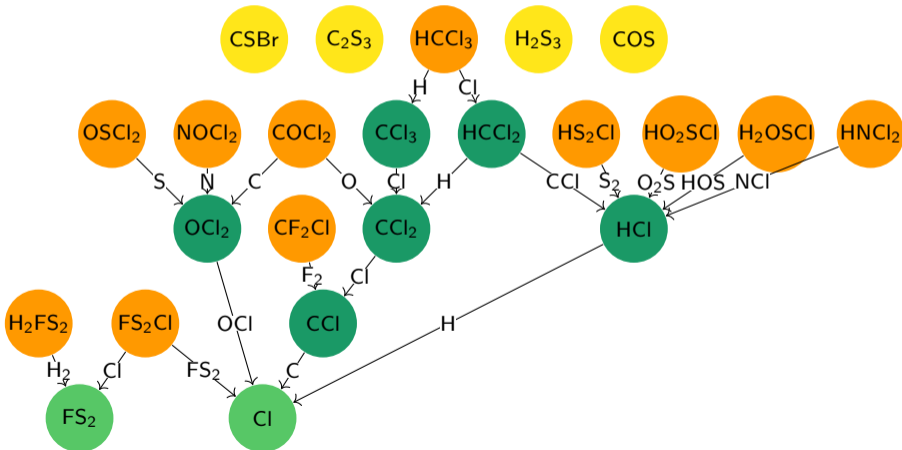
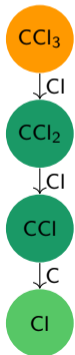


# Algorithmes de graphes pour $\text{CCl}_4$

● tout seul, 
 ● Maximal, 
 ● Nœud, 
 ● feuille.

Pseudo-fragmentation graph of knapsack fragments w.r.t. partial order

Fragmentation graph of  $\text{CCl}_4$





## Un peu de programmation Python

`https://gitlab.inria.fr/guillevi/alpinac/`

`https://members.loria.fr/AGuillevic/files/talks/knapsack.py`

# Aspects du métier en recherche

Faire de la recherche :

- Trouver une question
- Réfléchir
- Tester
- Calculer
- Rédiger
- Publier

Notre article :

<https://dx.doi.org/10.1186/s13321-021-00544-w>

Guillevic et al. *J Cheminform* (2021) 13:78  
<https://doi.org/10.1186/s13321-021-00544-w>

Journal of Cheminformatics

## RESEARCH ARTICLE

Open Access



# Automated fragment formula annotation for electron ionisation, high resolution mass spectrometry: application to atmospheric measurements of halocarbons

Myriam Guillevic<sup>1\*</sup>, Aurore Guillevic<sup>2</sup>, Martin K. Vollmer<sup>1</sup>, Paul Schlauri<sup>1</sup>, Matthias Hill<sup>1</sup>, Lukas Emmenegger<sup>1</sup> and Stefan Reimann<sup>1</sup>

### Abstract

**Background:** Non-target screening consists in searching a sample for all present substances, suspected or unknown, with very little prior knowledge about the sample. This approach has been introduced more than a decade ago in the field of water analysis, together with dedicated compound identification tools, but is still very scarce for indoor and atmospheric trace gas measurements, despite the clear need for a better understanding of the atmospheric trace gas composition. For a systematic detection of emerging trace gases in the atmosphere, a new and powerful analytical method is gas chromatography (GC) of pre-concentrated samples, followed by electron ionisation, high resolution mass spectrometry (EI-HRMS). In this work, we present data analysis tools to enable automated fragment formula annotation for unknown compounds measured by GC-EI-HRMS.

**Results:** Based on co-eluting mass/charge fragments, we developed an innovative data analysis method to reliably reconstruct the chemical formulae of the fragments, using efficient combinatorics and graph theory. The method does not require the presence of the molecular ion, which is absent in ~40% of EI spectra. Our method has been trained and validated on >50 halocarbons and hydrocarbons, with 3–20 atoms and molar masses of 30–330 g mol<sup>-1</sup>, measured with a mass resolution of approx. 3500. For >90% of the compounds, more than 90% of the annotated fragment formulae are correct. Cases of wrong identification can be attributed to the scarcity of detected fragments per compound or the lack of isotopic constraint (no minor isotopocule detected).

**Conclusions:** Our method enables to reconstruct most probable chemical formulae independently from spectral databases. Therefore, it demonstrates the suitability of EI-HRMS data for non-target analysis and paves the way for the identification of substances for which no EI mass spectrum is registered in databases. We illustrate the performances of our method for atmospheric trace gases and suggest that it may be well suited for many other types of samples. The L-GPL licenced Python code is released under the name ALPINAC for ALgorithmic Process for Identification of Non-targeted Atmospheric Compounds.

# Aspects du métier en recherche

- travailler à plusieurs, souvent à distance
- co-auteurs
- articles en anglais
- conférences en Europe et dans le monde (US, Asie)
- Enseignement
- encadrement de stagiaires et doctorants
- médiation (comme aujourd'hui)
- administratif (comités de spécialistes, jury)
- chercher des financements
- déplacements, séjours à l'étranger



New Delhi, Inde, décembre 2017

Questions ?