PyC2MC: an open-source software solution for visualization and treatment of high-resolution mass spectrometry data

Maxime SUEUR[1,3], Julien F. MAILLARD[1,3], Oscar LACROIX-ANDRIVET[1,2,3], Christopher P. RÜGER[3,4], Pierre GIUSTI[1,2,3], Hélène LAVANANT[1], Carlos AFONSO[1,3]

End User School
13/12/2022

1 Normandie Univ, UNIROUEN, INSA Rouen, CNRS, COBRA, 76000 Rouen, France.
2 TotalEnergies OneTech, TotalEnergies Research & Technology Gonfreville, BP 27, 76700 Harfleur, France
3 International Joint Laboratory - iC2MC: Complex Matrices Molecular Characterization, TRTG, BP 27, 76700 Harfleur, France.
4 Joint Mass Spectrometry Centre, Chair of Analytical Chemistry, University of Rostock, 18059 Rostock, Germany; Interdisciplinary Faculty, Department Life, Light & Matter (LL&M), University of Rostock, 18051 Rostock, Germany.

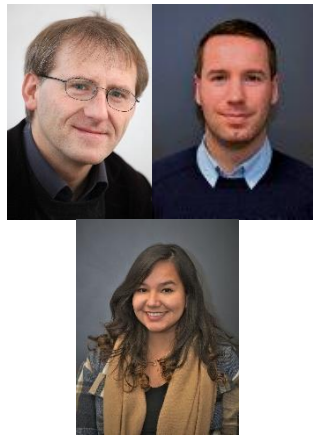
About me

- Position: Ph.D. student (since 02/2021)
- Location: COBRA laboratory (near Rouen, France)
- Research subject: Thermal analysis and Ion Mobility coupled to high-resolution mass Spectrometry for organic Aerosol Characterization (TIMSAC)
- Members:

University of Rouen Normandie

C.AFONSO
H.LAVANANT
M.SUEUR

iC2MC
INTERNATIONAL – COMPLEX MATRICES MOLECULAR CHARACTERIZATION



University of Rostock

R.ZIMMERMAN
C.RÜGER
S.VESGA

iC2MC joint lab

Academic collaborations

Universidad de Oviedo

Universität Rostock

Universidad Industrial de Santander

R. Rodgers

National High Magnetic Field Laboratory

FFI Future Fuels Institute – Florida State University

Florida State University 1851

iC2MC

INTERNATIONAL – COMPLEX MATRICES MOLECULAR CHARACTERIZATION

DP&Pol

Saft

TotalEnergies

Power

F&Lab

TotalEnergies

P. Giusti

Upstream

Sustainability

e2s UPPA energy environment solutions

IPREM Institut des sciences analytiques et de physico-chimie pour l'environnement et les matériaux

cnrs

Université de Pau et des Pays de l'Adour

Pr. B. Bouyssiere

Université de Rouen Normandie

Laboratoire Cobra

INSA

cnrs

Pr. C. Afonso

- **>35 people in 2022**
  - 7 from TotalEnergies
  - 8 Post-Docs
  - 11 on-going PhDs

Interests

- Biofuels
- Oils
- Dissolve organic matters
- Extraterrestrial matters
- …

Instrument manufacturers partnerships

Waters – THE SCIENCE OF WHAT'S POSSIBLE.®

BRUKER

Agilent Technologies

SHIMADZU – Excellence in Science

VUV ANALYTICS

Main analytical technics

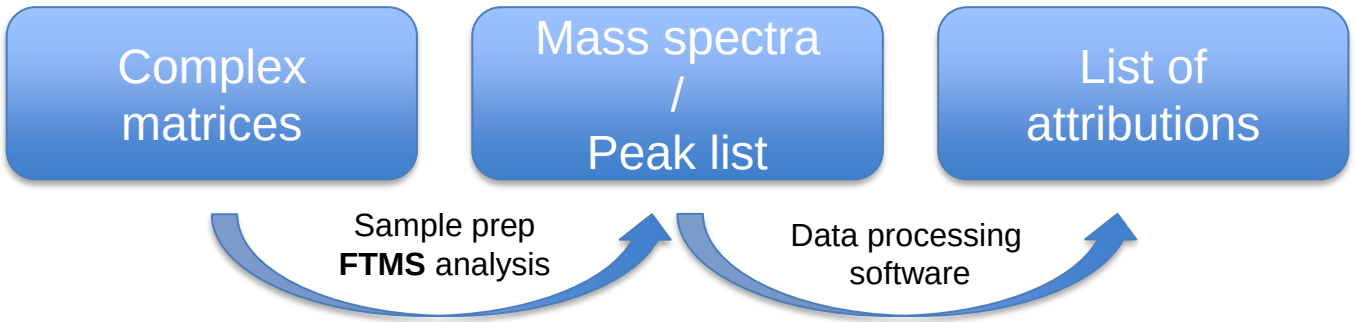
- **Mass spectrometry**
- Ion mobility
- GC/LC/GPC/SFC

Complex matrices molecular characterization

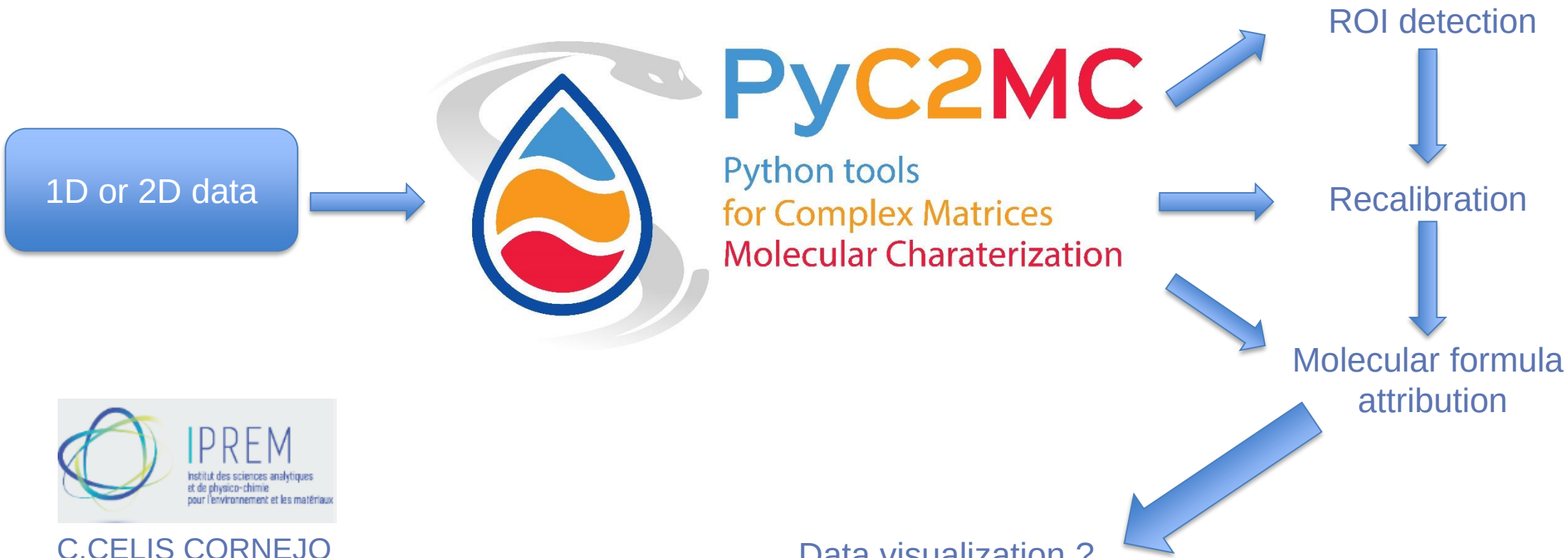
- Complex matrices:
  - Encountered in almost all research domains ➔ From biology to environmental sciences.
  - Humongous quantity of information at the molecular level

- Accessing complex matrices molecular information:
  - High-end instrumentation is required: FTICR / Orbitrap mass spectrometers that can achieve very high resolving power (>10^5) and mass accuracy. (<1ppm)
  - These two factors allow for the resolution of isobaric compounds and attribution of molecular formulas using constructors' software or self-elaborated routines (ex: Data Analysis, Composer, Peak-by-Peak)

- Python tools for complex matrices molecular characterization or PyC2MC



PyC2MC



IPREM
Institut des sciences analytiques et de physico-chimie pour l'environnement et les matériaux

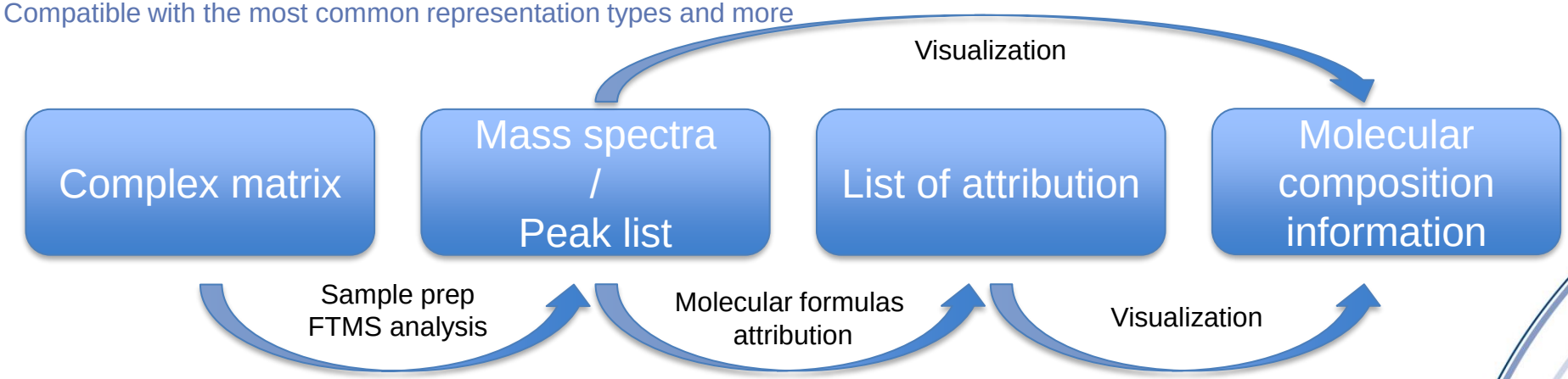
C.CELIS CORNEJO
G.SALVATO VALLVERDU
C.RÜGER
B.BOUYSSIERE
R.RODGERS

Complex molecular data visualization

- Visualizing the molecular information:
  - Generic data treatment software:
    - Microsoft Excel
    - OriginLab Origin

    } Easy to use but low efficiency
  - R, Python or Matlab based self-written software:
    - DEIMoS, (*S.M. Colby et al.*)
    - Constellation, (*D.R. Letourneau et al.*)
    - PyKrev, (*E. Kitson et al.*)

    } Several minor drawbacks
  - PyC2MC viewer:
    - User-friendly
    - Open-source
    - Able to handle voluminous data
    - Compatible with the most common representation types and more



PyC2MC Viewer

Peak list
Attribution list

PyC2MC

Python tools
for Complex Matrices
Molecular Charaterization

Viewer Edition

Data processing

Primary visualization

Samples comparison

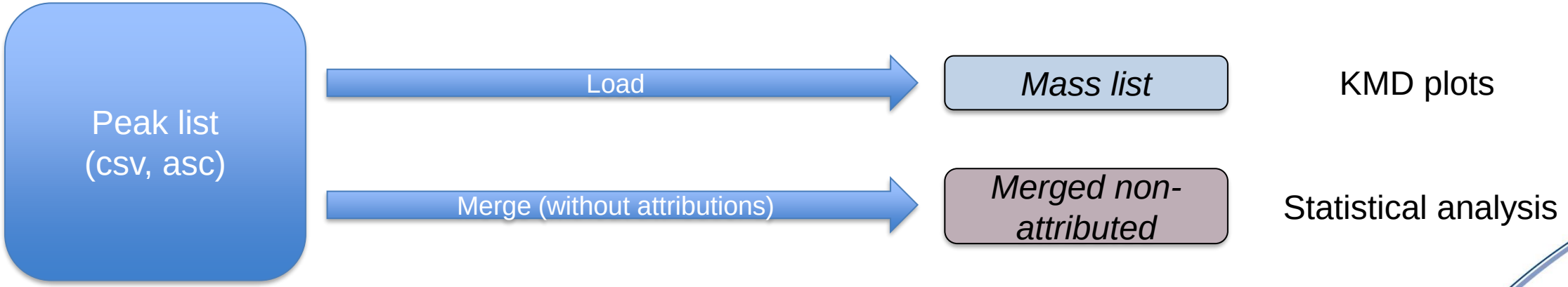
Statistical analysis

Laboratoire Cobra

M.SUEUR
J.MAILLARD
O.LACROIX-ANDRIVET

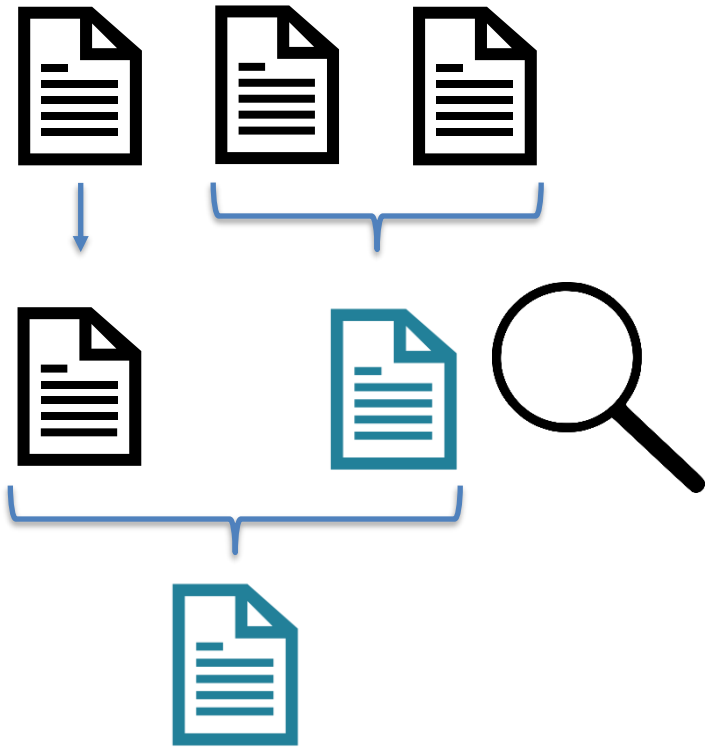
Peak lists as input file

- Compatible constructor software: DataAnalysis (mass list export in *.asc*), Xcalibur (*.csv*)
- File architecture: *m/z* ratio, intensity in an *.asc* file
- Data processing option: merge (without attribution)



Merge (without attributions)

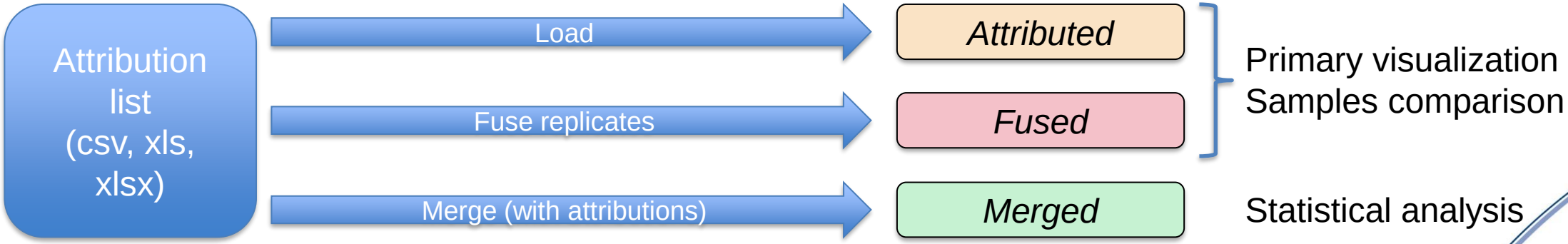
Example: using pandas.merge_asof() and a tolerance (default Tol = 0,1mDa)



m/z ratio	Count	Name of file 1	Name of file 2
If corresponding ratio within the tolerance	3	Intensity value	Intensity value
If exclusive to file 1	1	Intensity value	0 filled
If exclusive to file 2	1	0 filled	Intensity value

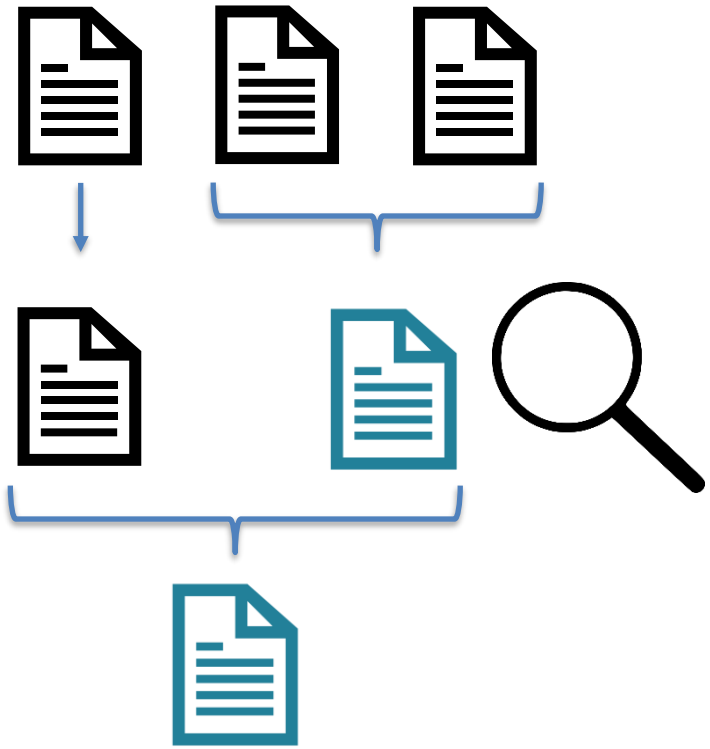
Attribution lists as input file

- Compatible constructor software: DataAnalysis ('Export for Van Krevelen', *.csv*), Xcalibur (*.csv*), PetroOrg (*.csv*, *.xls*) or CERES processing (*.xlsx*)
- File architecture: *m/z* ratio, intensity, error (in ppm) and molecular formula in a *.csv* file
- Data processing options:
  - Merge (with attribution)
  - Fuse replicates



Merge (with attributions)

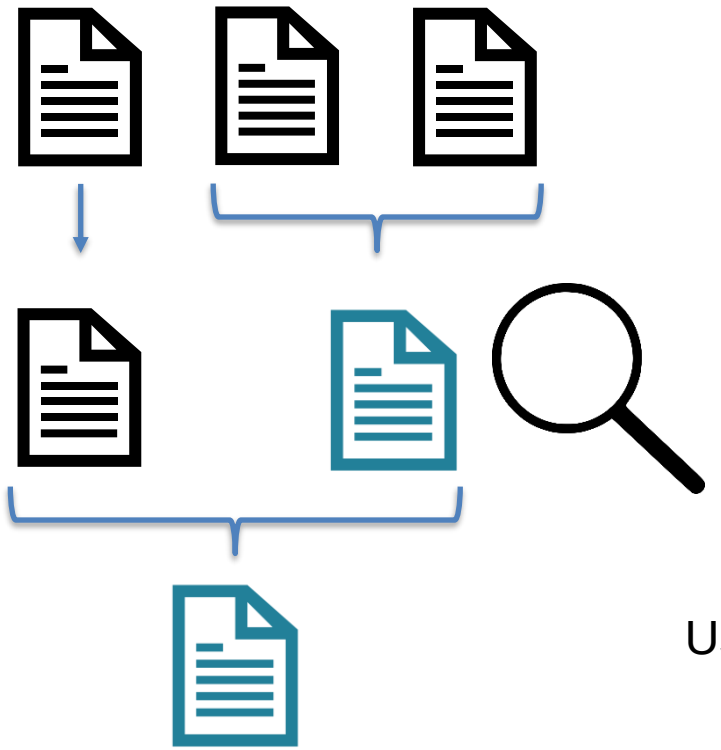
Example: using pandas.merge () on molecular formula



Molecular formula	Count	Name of file 1	Name of file 2
If corresponding formula	3	Intensity value	Intensity value
If exclusive to file 1	1	Intensity value	0 filled
If exclusive to file 2	1	0 filled	Intensity value

Fuse replicates

Example: using pandas.merge () on molecular formula

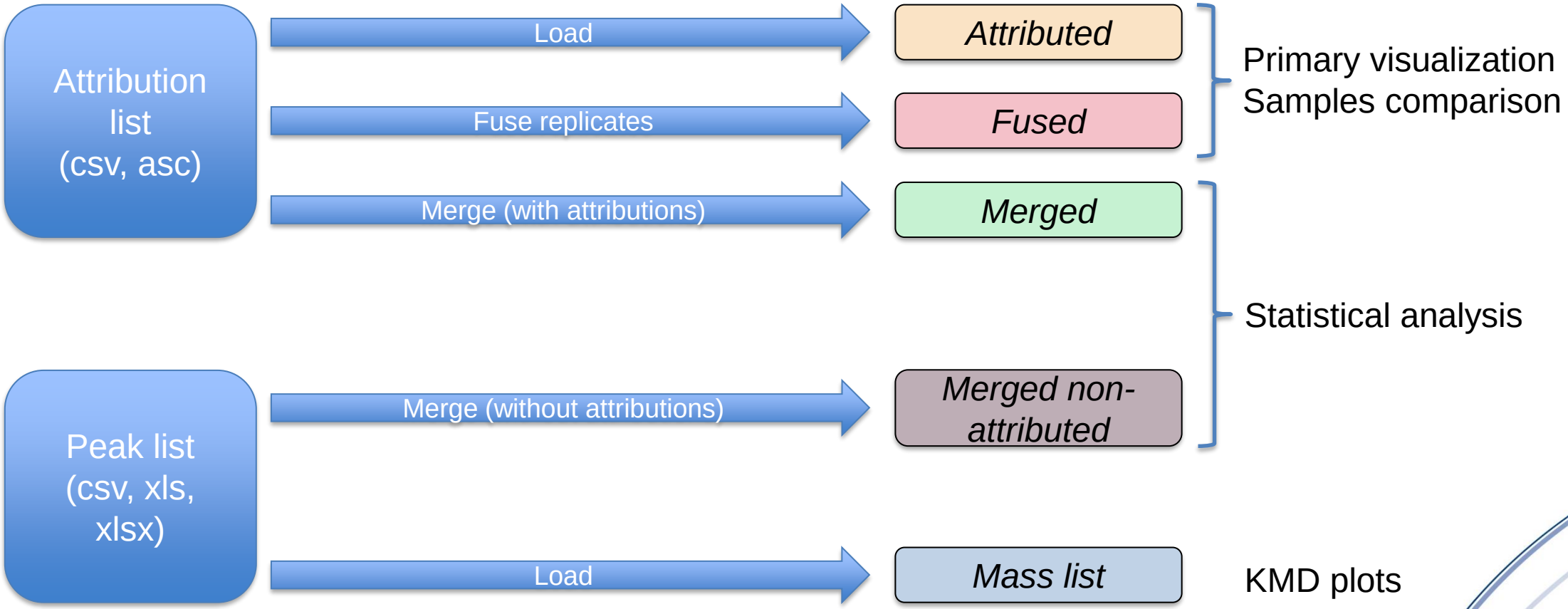


Molecular formula	Count	Name of file 1	Name of file 2	Name of file 3
If corresponding formula	3	Intensity value	Intensity value	Intensity value
Common to file 1 & 3	2	Intensity value	0 filled	Intensity value
Common to file 2 & 3	2	0 filled	Intensity value	Intensity value

User input : X = minimum occurrence of each attribution
with $1 \leq X \leq$ nb sample

Calculates a mean intensity for each peak ➔ Files is treated as a normal attribution list but with statistical information (mean, std. dev.)

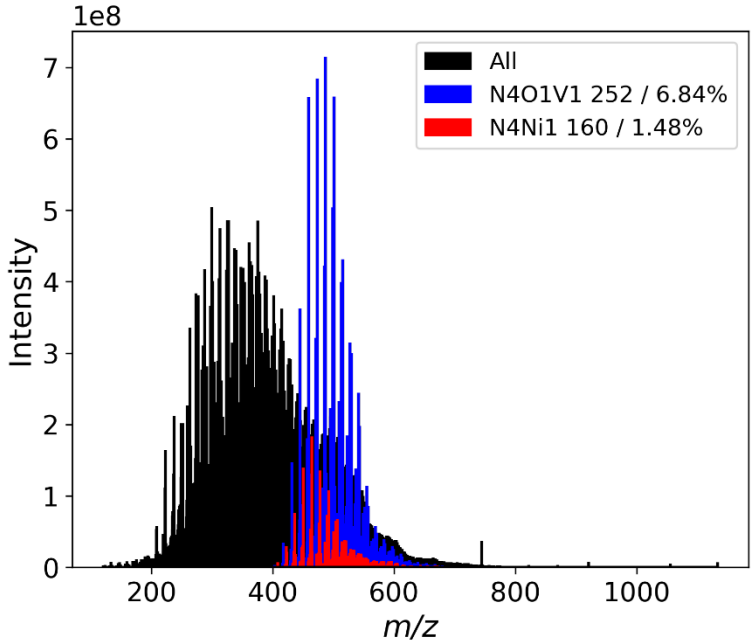
Input and data processing summary

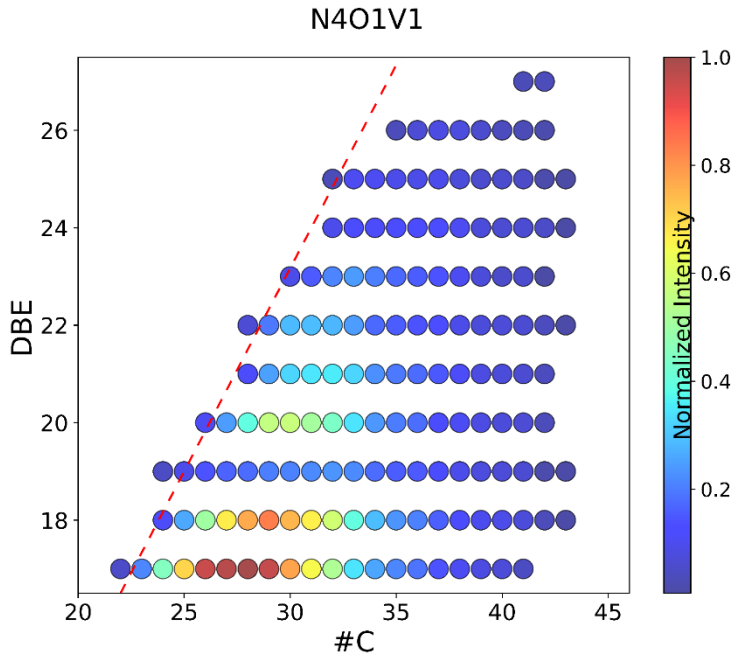


Primary visualization

- Compatible items: *Attributed* *Fused*

- Dataset:
  - Plastic pyrolysis oil analyzed by APCI(+) (*C. MASE et al*)
- Primary visualization :





Also:
- Van Krevelen plots (using environmental science variables as color code or not)
- Kroll plots
- Kendrick diagrams

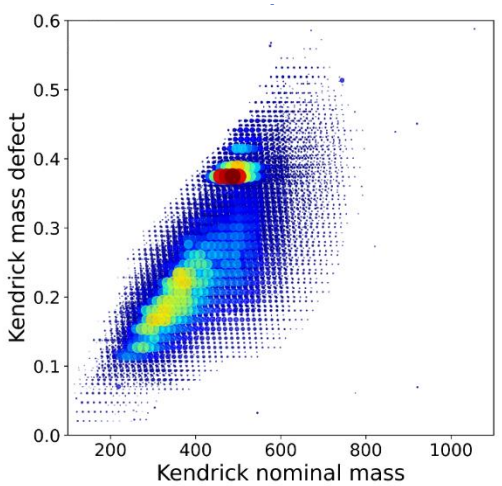
KMD plots and series extraction

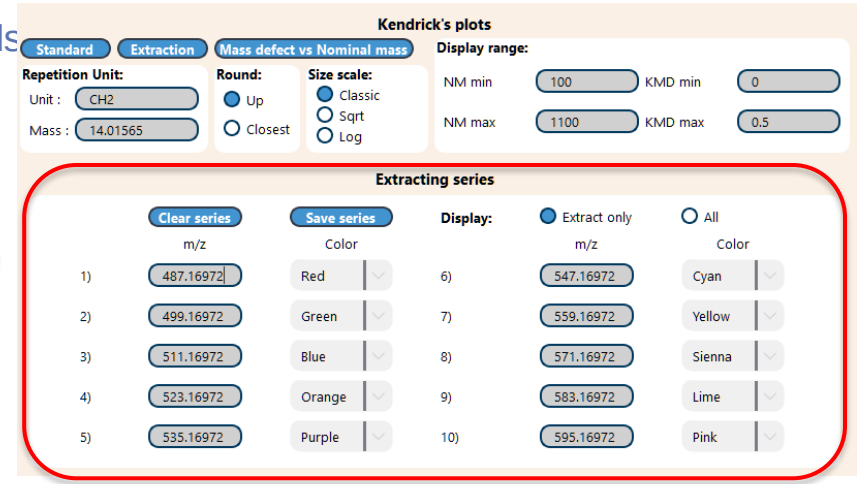
- Compatible items: *Attributed* *Fused* *Mass list*

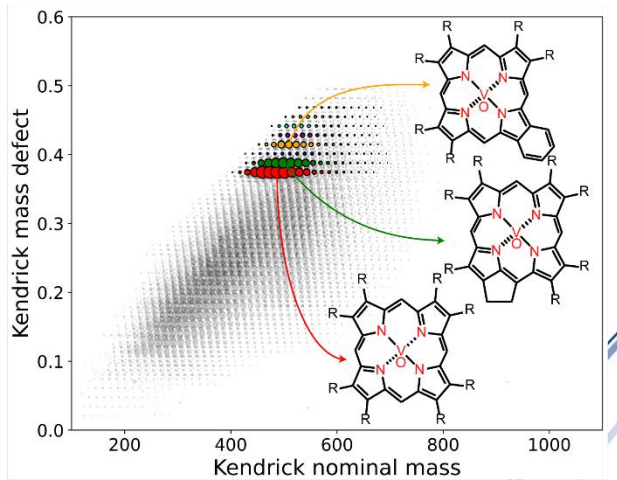
- Dataset: Shipping fuels and corresponding emissions analyzed by ET-MALDI FTICR MS (*M. SUEUR et al.*)
- KMD plots as an attribution tool:
  - KMD vs NKM calculation:

1. $KMD = Nominal\ Kendrick\ mass - Kendrick\ mass$
2. $Kendrick\ mass\ = observed\ mass\ \times \frac{Repetition\ pattern\ nominal\ mass}{Repetition\ pattern\ exact\ mass}$
3. $Kendrick\ nominal\ mass = rounded\ Kendrick\ mass$

Highlighted series of species for manual attribution (*.csv)





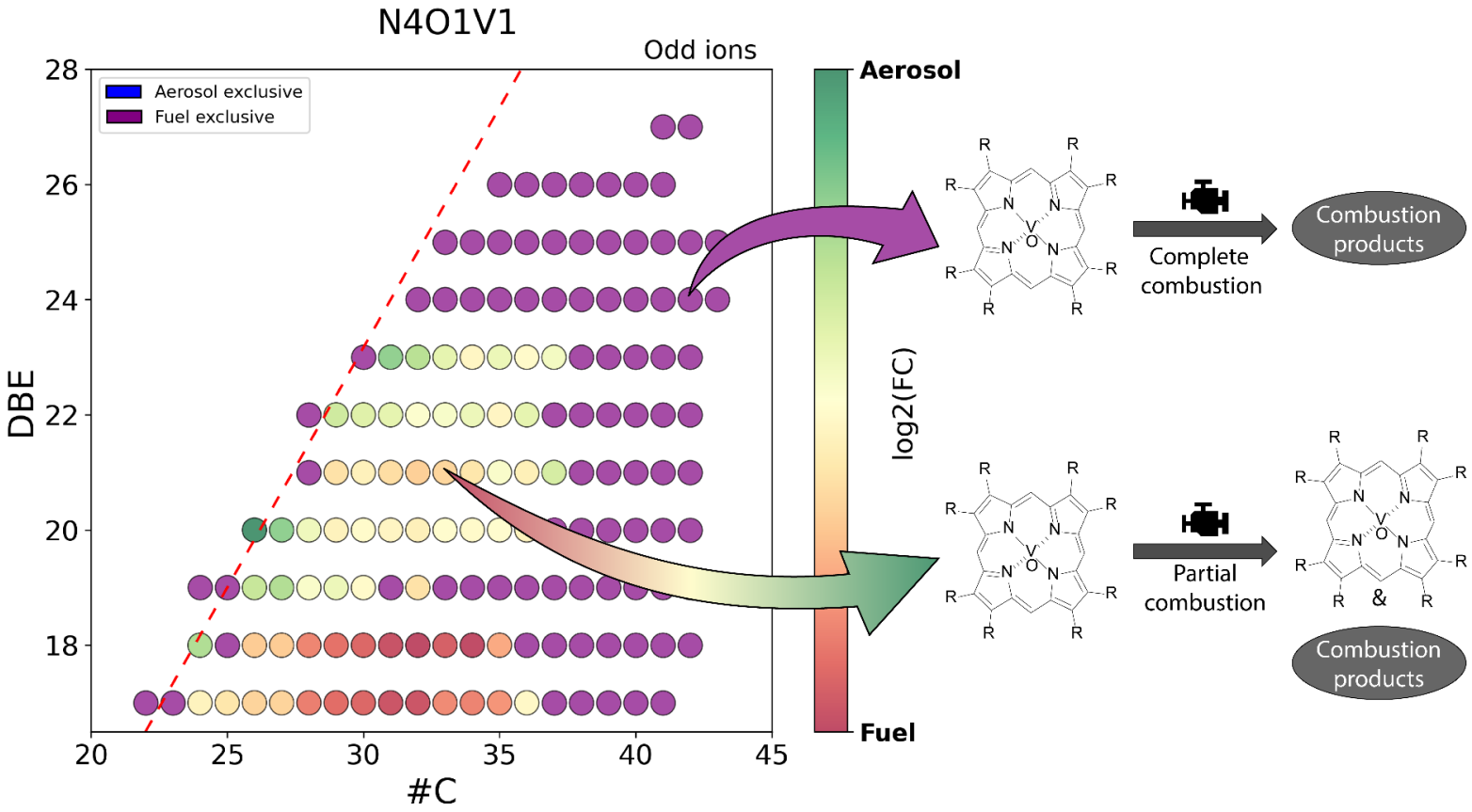


Sample comparison

- Compatible items:

Attributed *Fused*

$$FC = \frac{Peak\ intensity\ in\ sample\ 2}{Peak\ intensity\ in\ sample\ 1}$$ ➔ $\log_2$(FC) is used as the color coding variable.



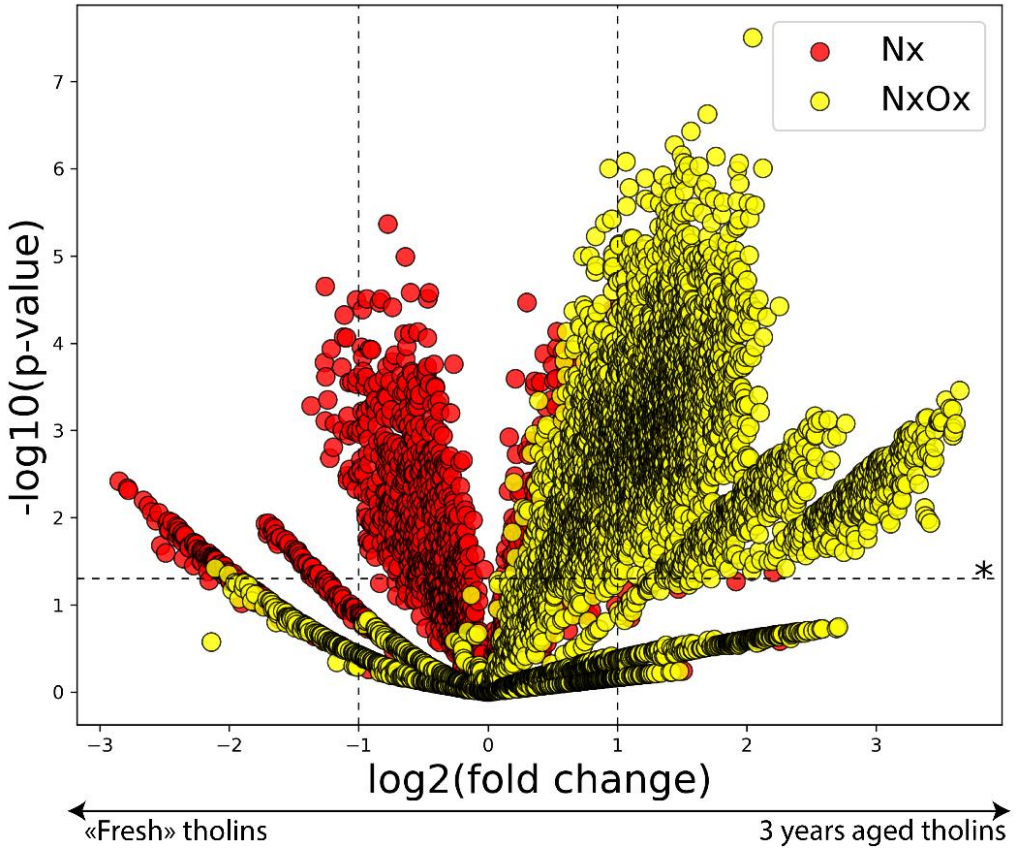
$\log_2(FC) > 0$: Higher intensity in particles than in the fuel

$\log_2(FC) = 0$: No differences

$\log_2(FC) < 0$: Higher intensity in the fuel than in particles

Statistical analysis

- Compatible items: *Merged* *Merged non-attributed*
- Dataset: Aerosol samples mimicking Titan's atmosphere (Tholins) (*J.MAILLARD et al*)



Also:
- Venn diagrams
- Volcano plots

Conclusion

- Software is available on GitHub (https://github.com/iC2MC/PyC2MC_viewer)

- 2 versions :
  - Open code, to be executed with a Python interpreter
  - Executable file (.exe)

- Bug report or suggestions: maxime.sueur1@univ-rouen.fr

- Preprint paper: “PyC2MC: an open-source software solution for visualization and treatment of high-resolution mass spectrometry data” on ChemRxiv

Acknowledgements

Universität Rostock

iC2MC
INTERNATIONAL – COMPLEX MATRICES MOLECULAR CHARACTERIZATION

TotalEnergies

ED 508 NC
Normande de Chimie

Laboratoire Cobra
Chimie organique, Bioorganique, Réactivité et Analyse.

UNIVERSITÉ DE ROUEN NORMANDIE

EU FT-ICR MS

UNION EUROPÉENNE
Fonds européen de développement régional