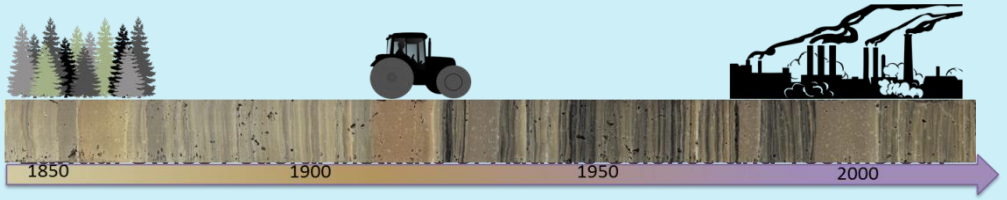
Prioritizing relevant nontarget contaminants using statistical analysis of LC-HRMS data of lake sediments

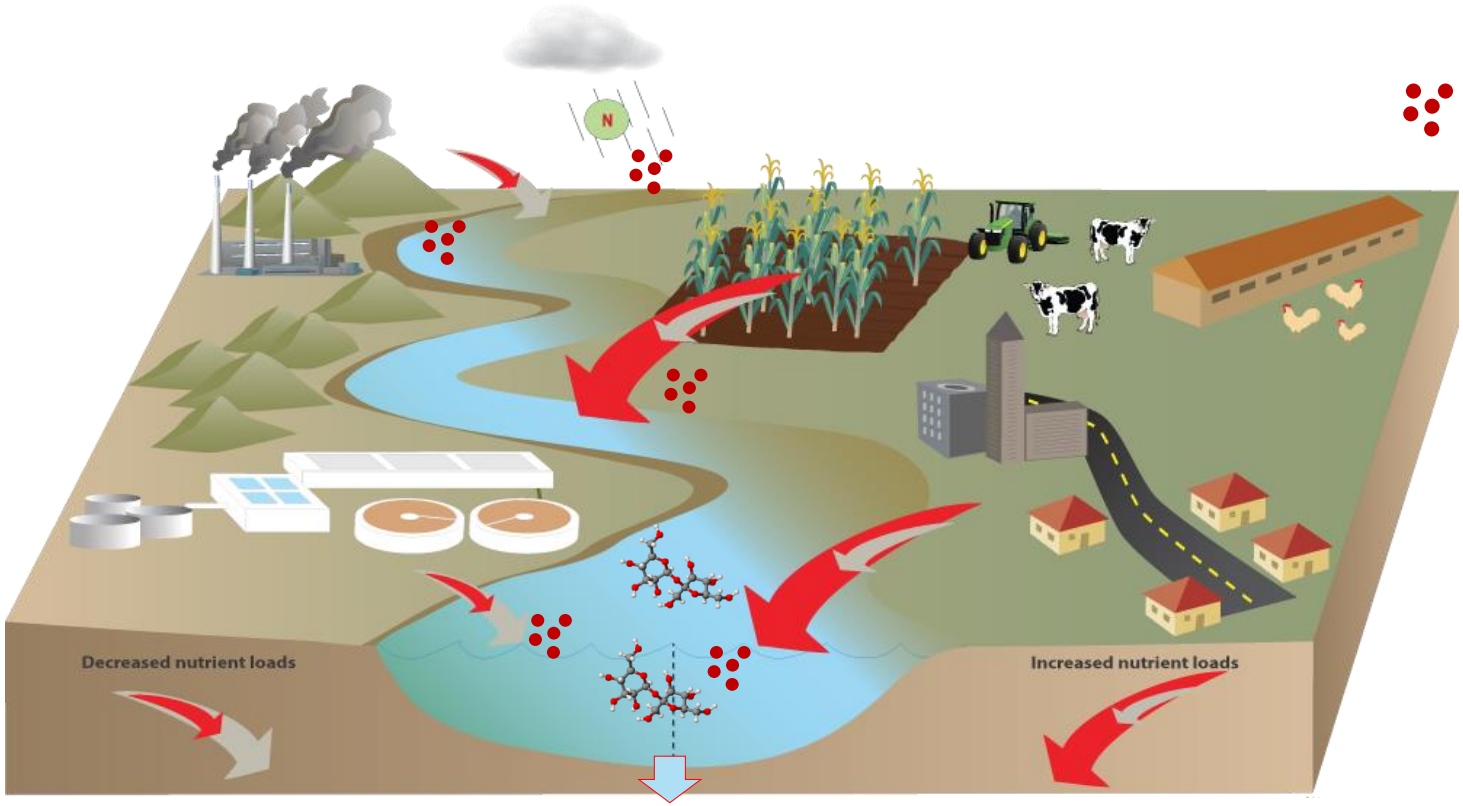
Aurea C. Chiaia-Hernandez, B. F. Günthardt, Martin P. Frey and J. Hollender



© Nathalie Dubois

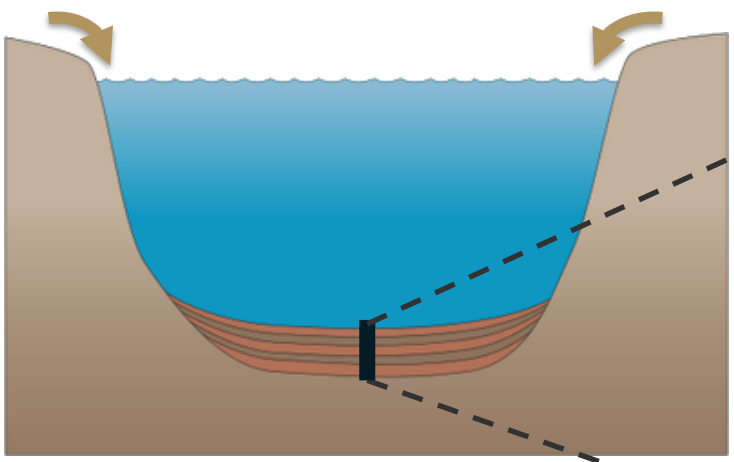
Lakes are Sensitive Barometers

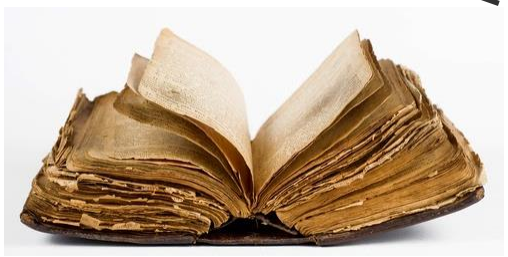
- They respond to changes around them
  - Climate change (e.g. changes in rainfall)
  - Catchment change (e.g. caused by people)

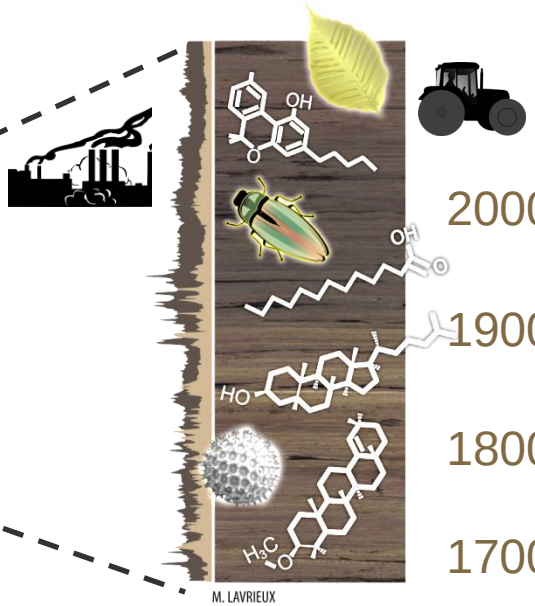


Jane Hawkey, IAN image library

Sediments = History Books of the Environment







2000

1900

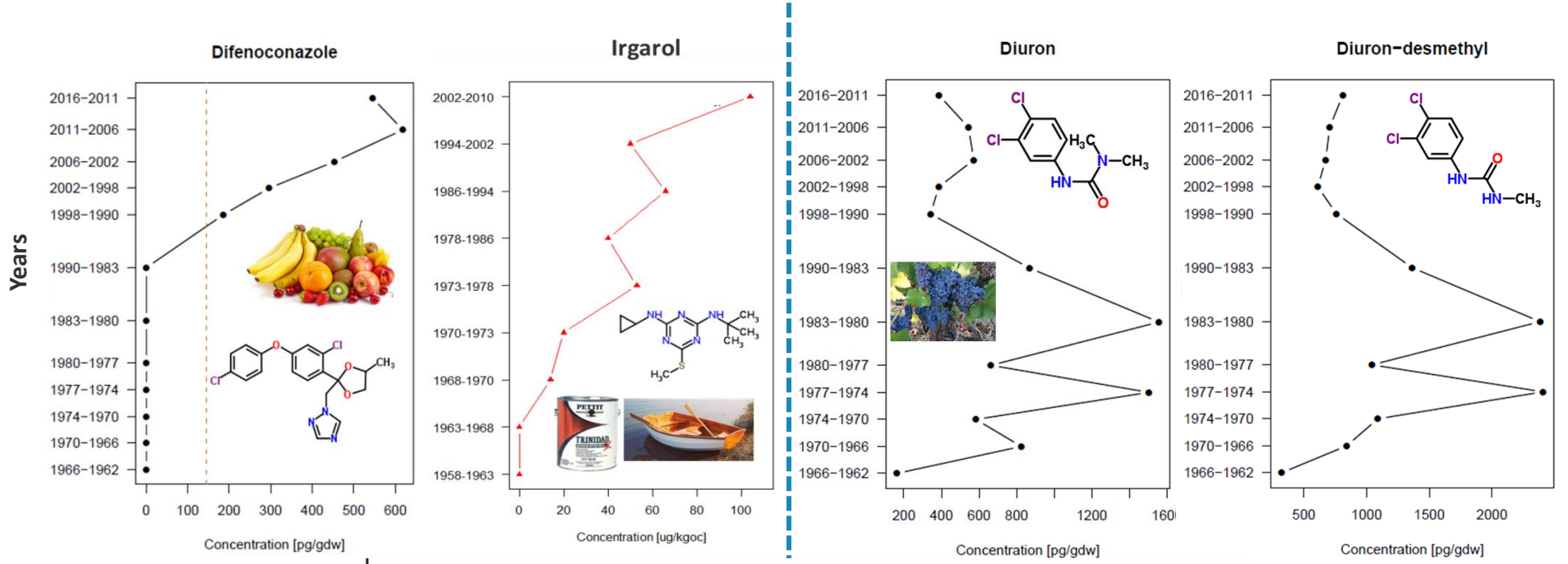
1800

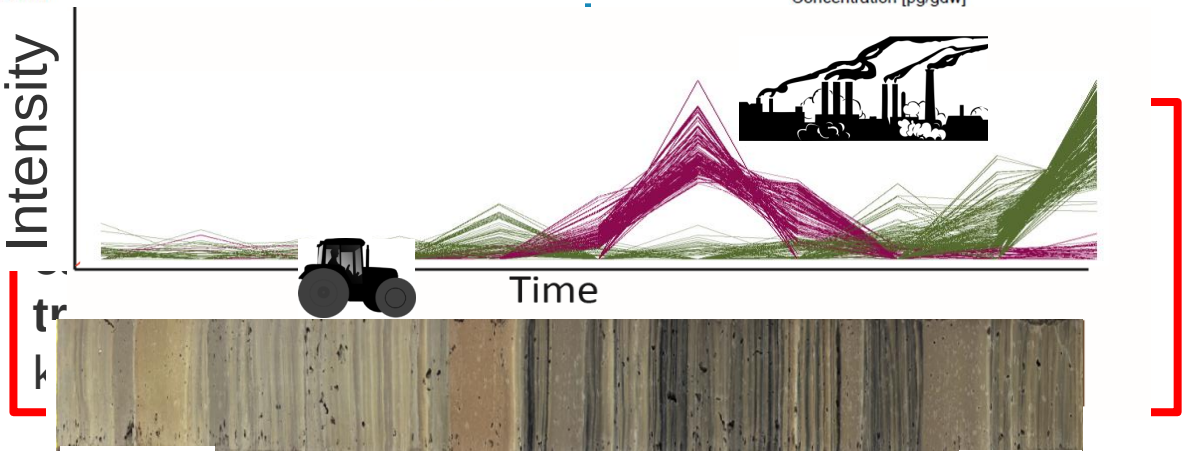
1700

Matrix

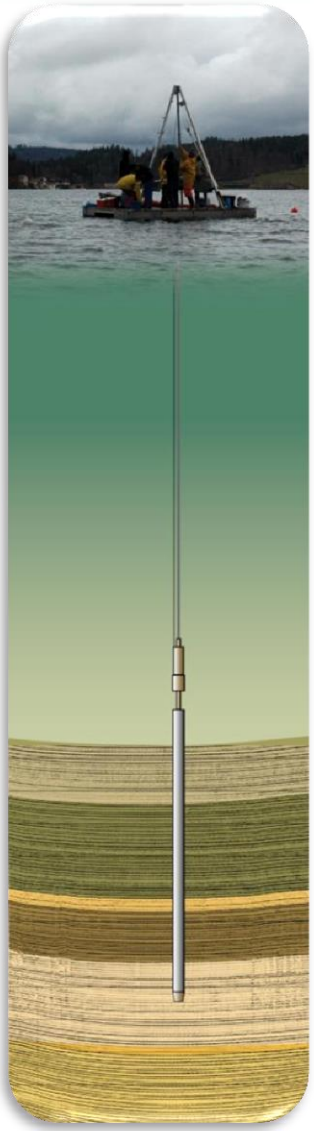
Slide modified from N. Dubois, Eawag

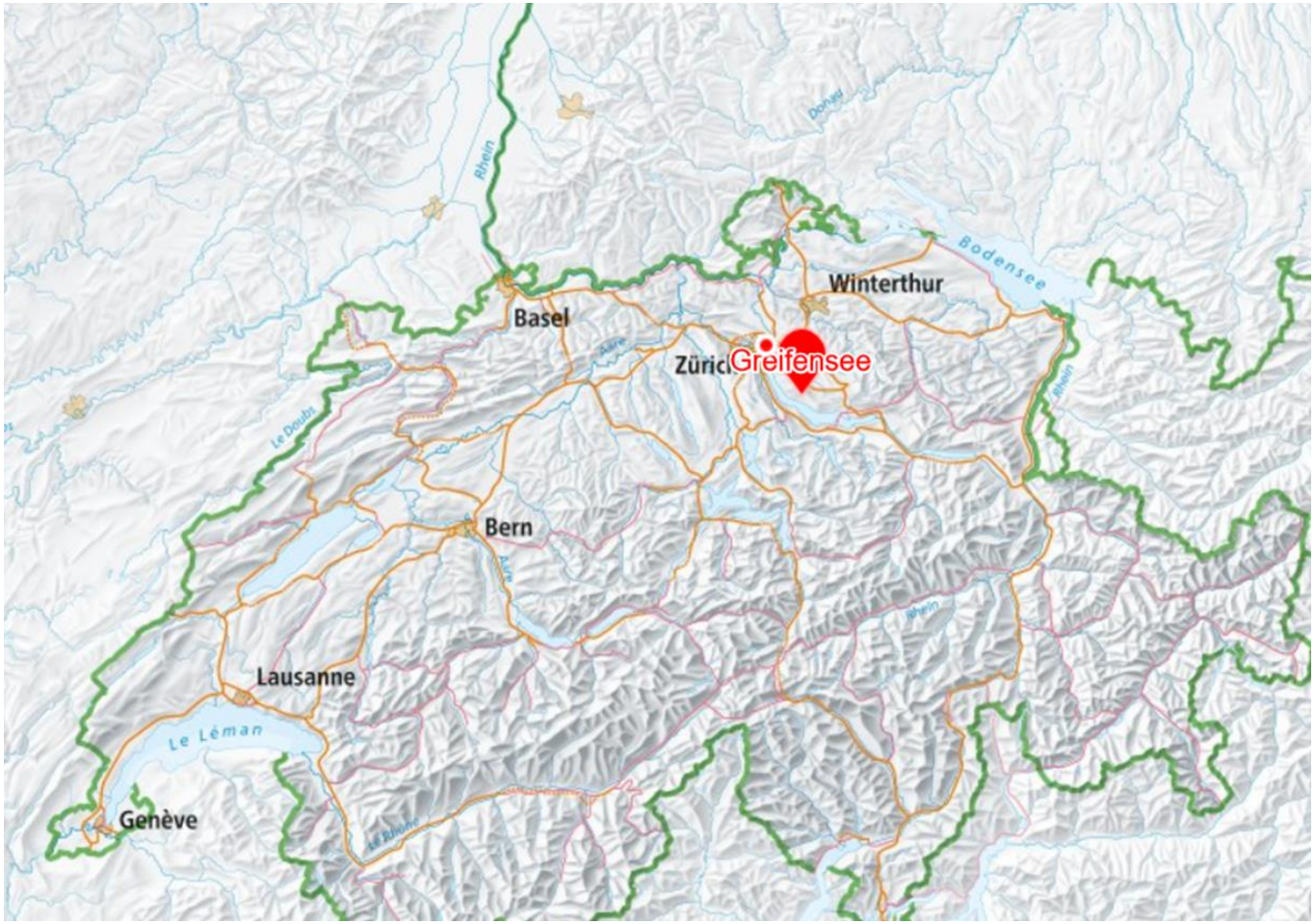
Target and Suspect Screening by HRMS – Known Compounds





1. Collecting Lake Sediments

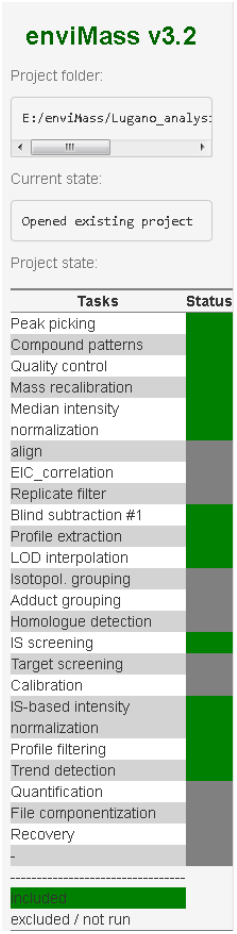


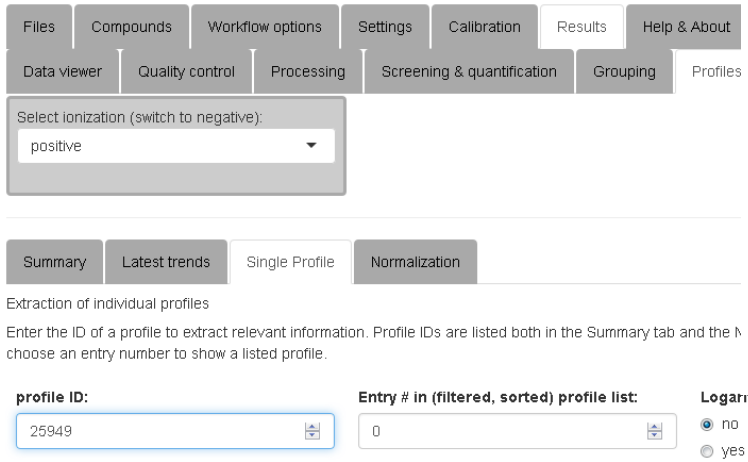


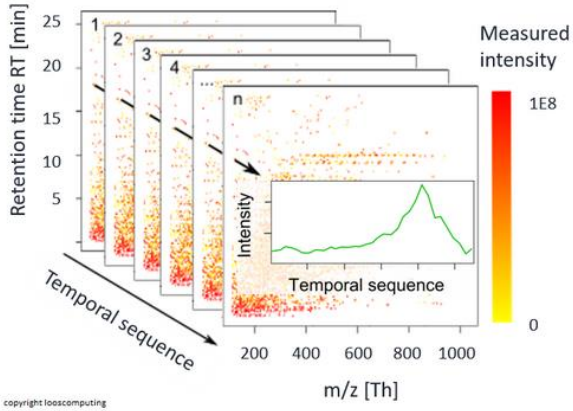
© Typhaine Guillemot

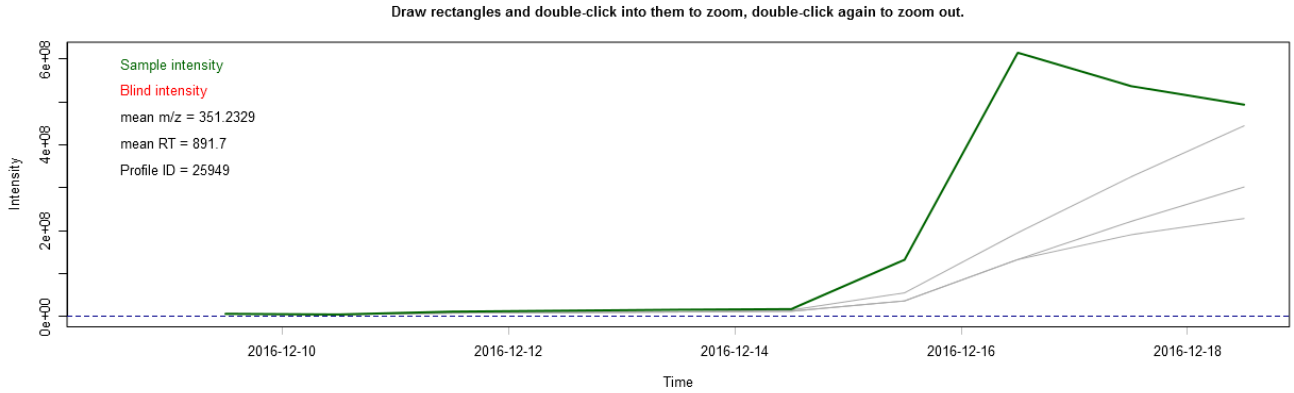
https://map.geo.admin.ch/

2. Profile Detection- enviMass









~ 1800

2016

Loos, M. enviMass
http://www.looscomputing.ch/eng/enviMass/overview.htm (accessed 14/06/2017).

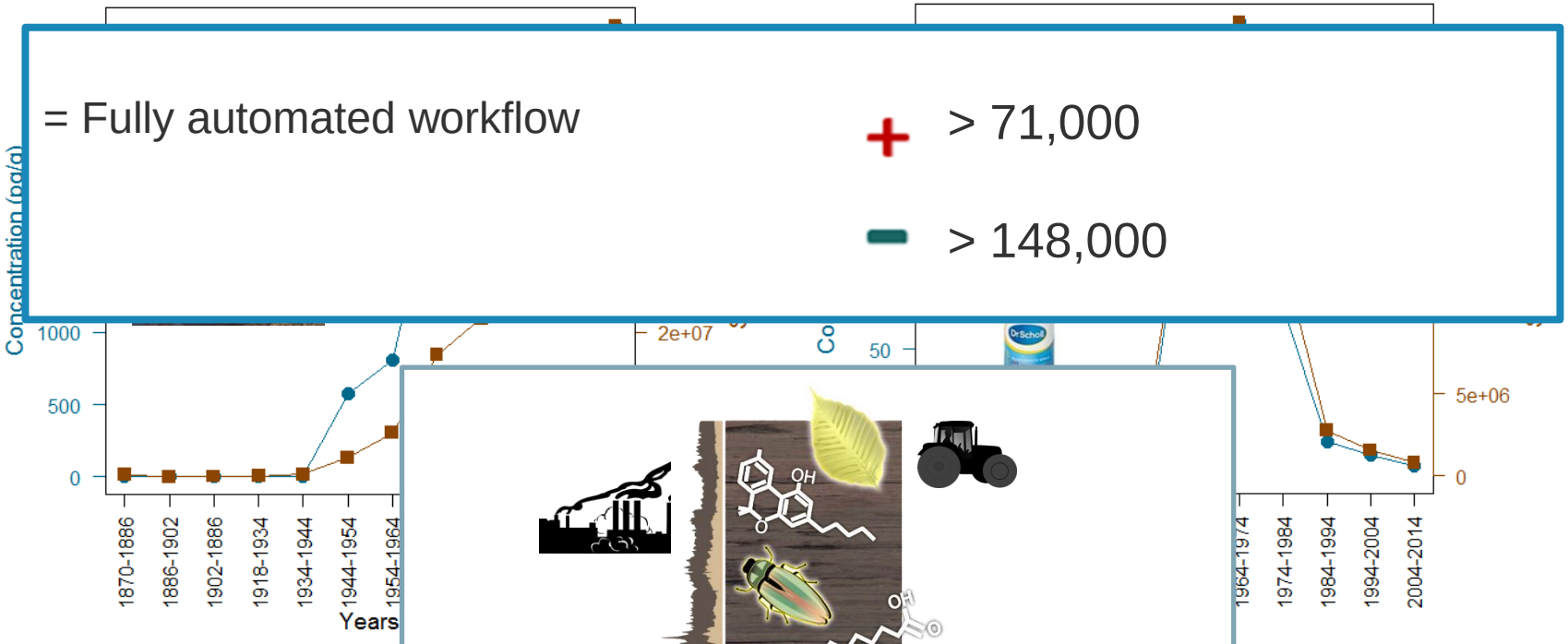
Profile Detection (enviMass) vs. Concentration Profiles

Irgarol **Dichlorophene**

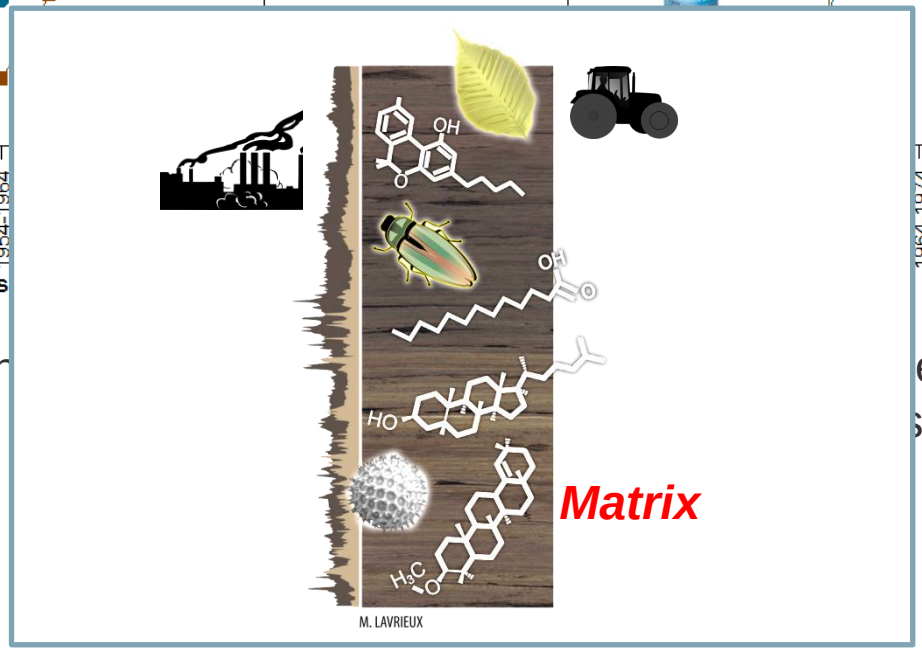
= Fully automated workflow

\+ > 71,000

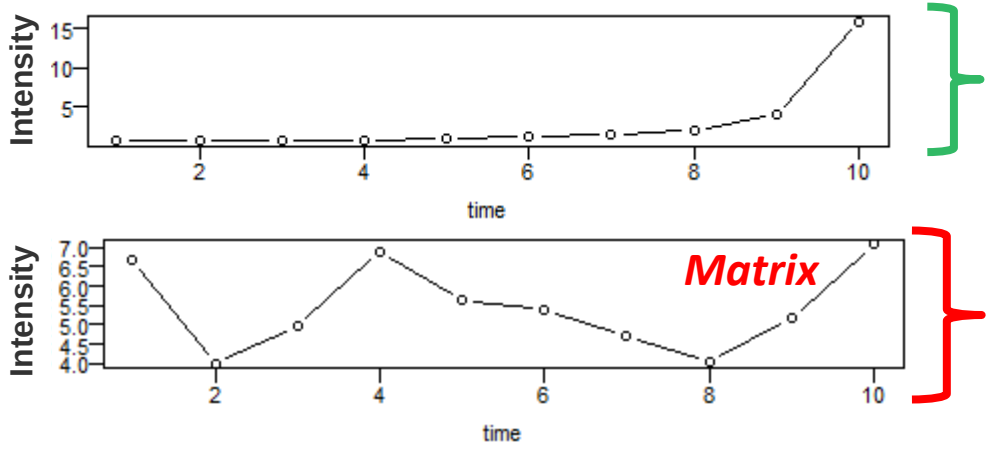
− > 148,000



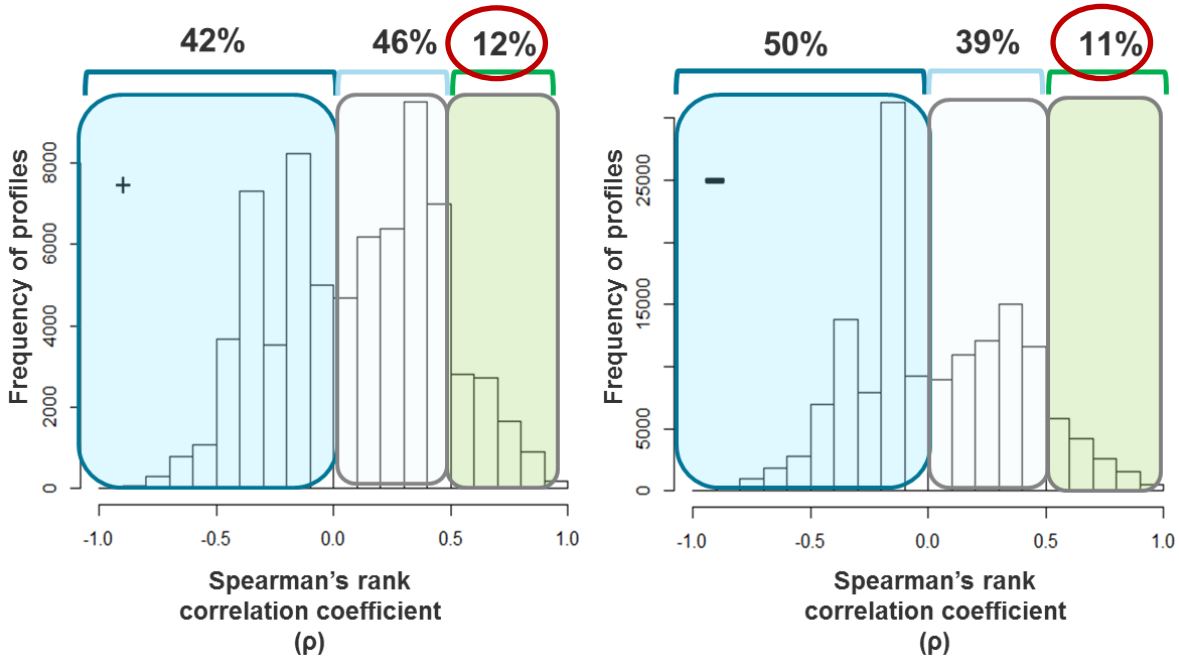
- Measured con ed using ss



3. Trend Characterization: Spearman's Rank Correlation

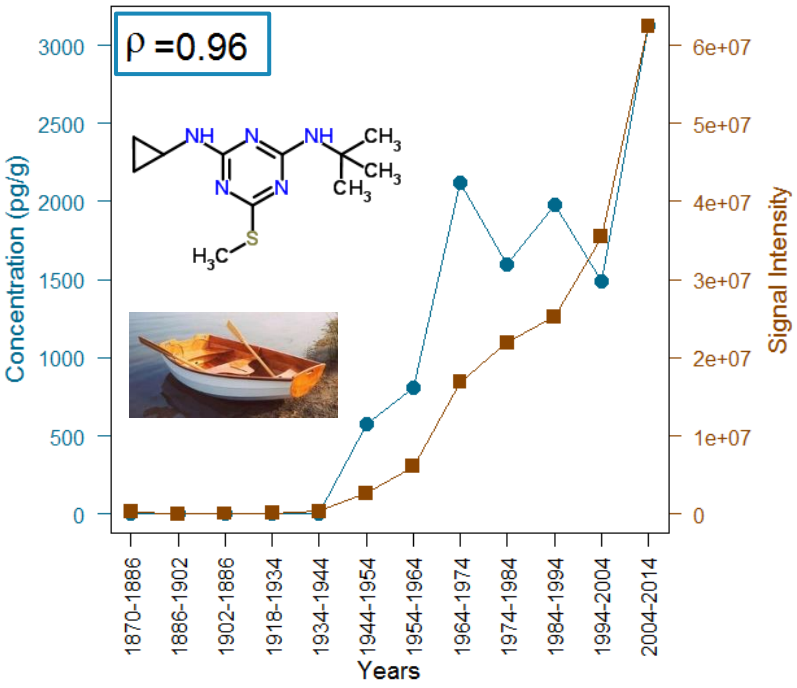


	Pearson (*r*)	Spearman (ρ)
1	0.66	0.95
2	0.00	0.10

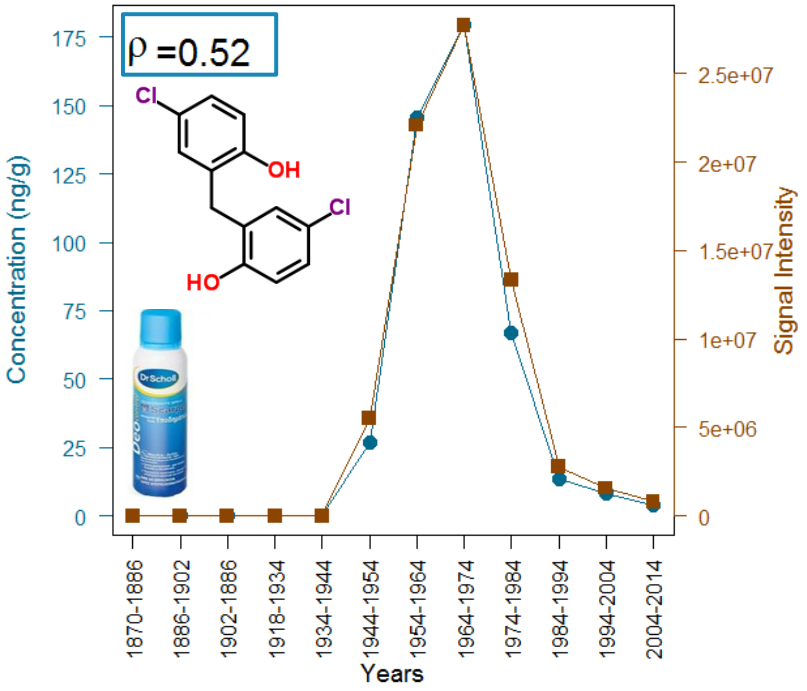


Trend characterization 1: Spearman's rank correlation

Irgarol



Dichlorophene



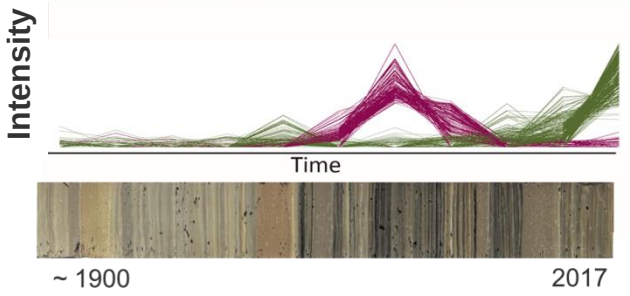
Measured concentration

Profile extracted using enviMass

4. Trend Characterization: Hierarchical Cluster Analysis

Cluster analysis: Group samples that are similar

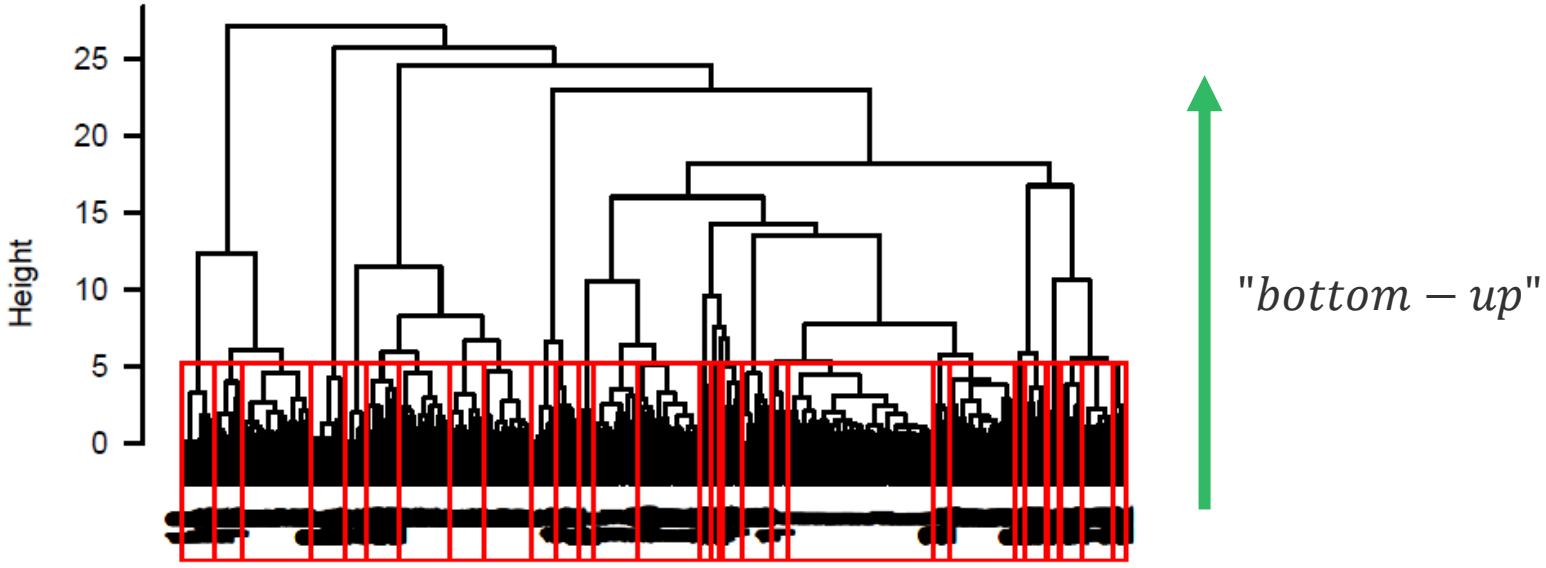
- Intensity paterns in sediments



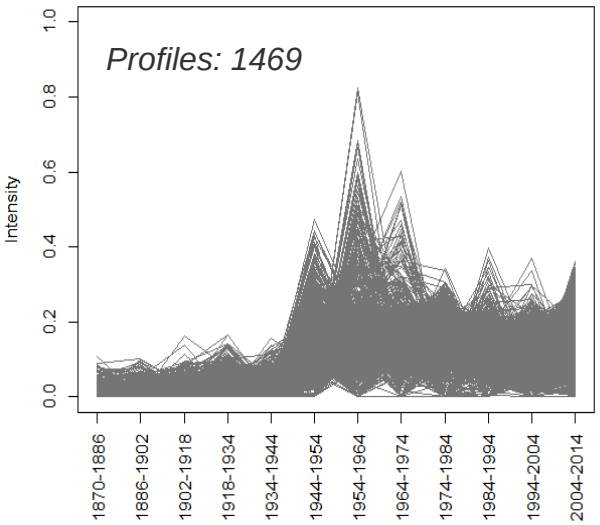
Agglomerative hierarchical clustering

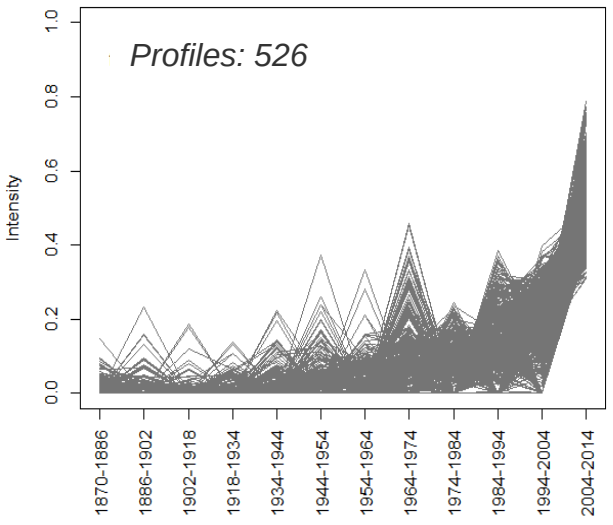
- Classification
- Information retrieval

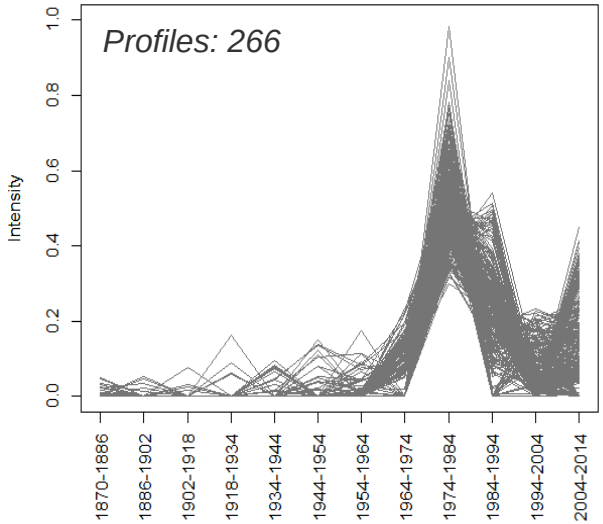
Dendogram

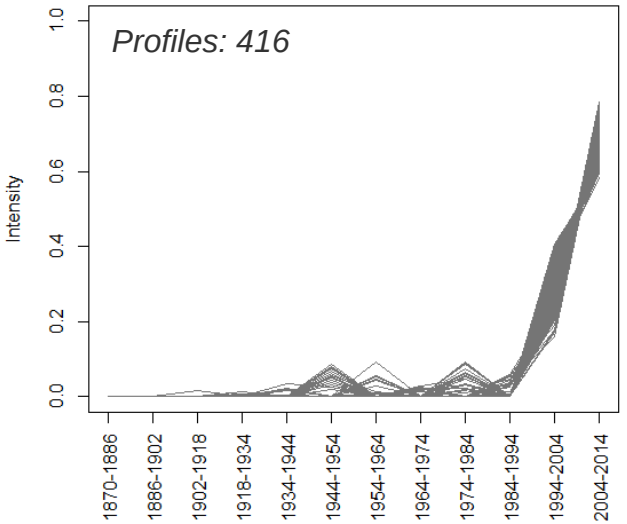


Hierarchical Clustering- Greifensee

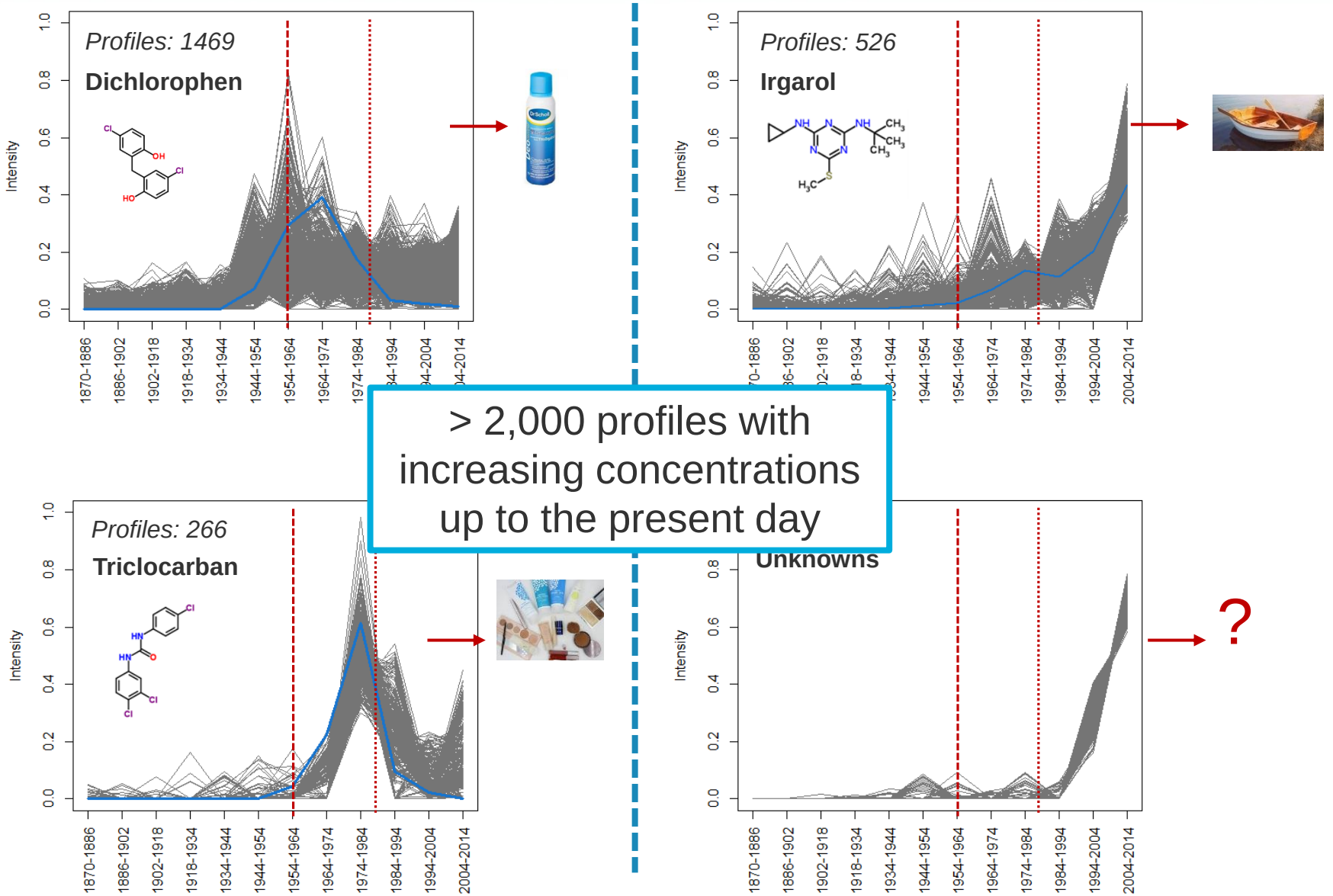




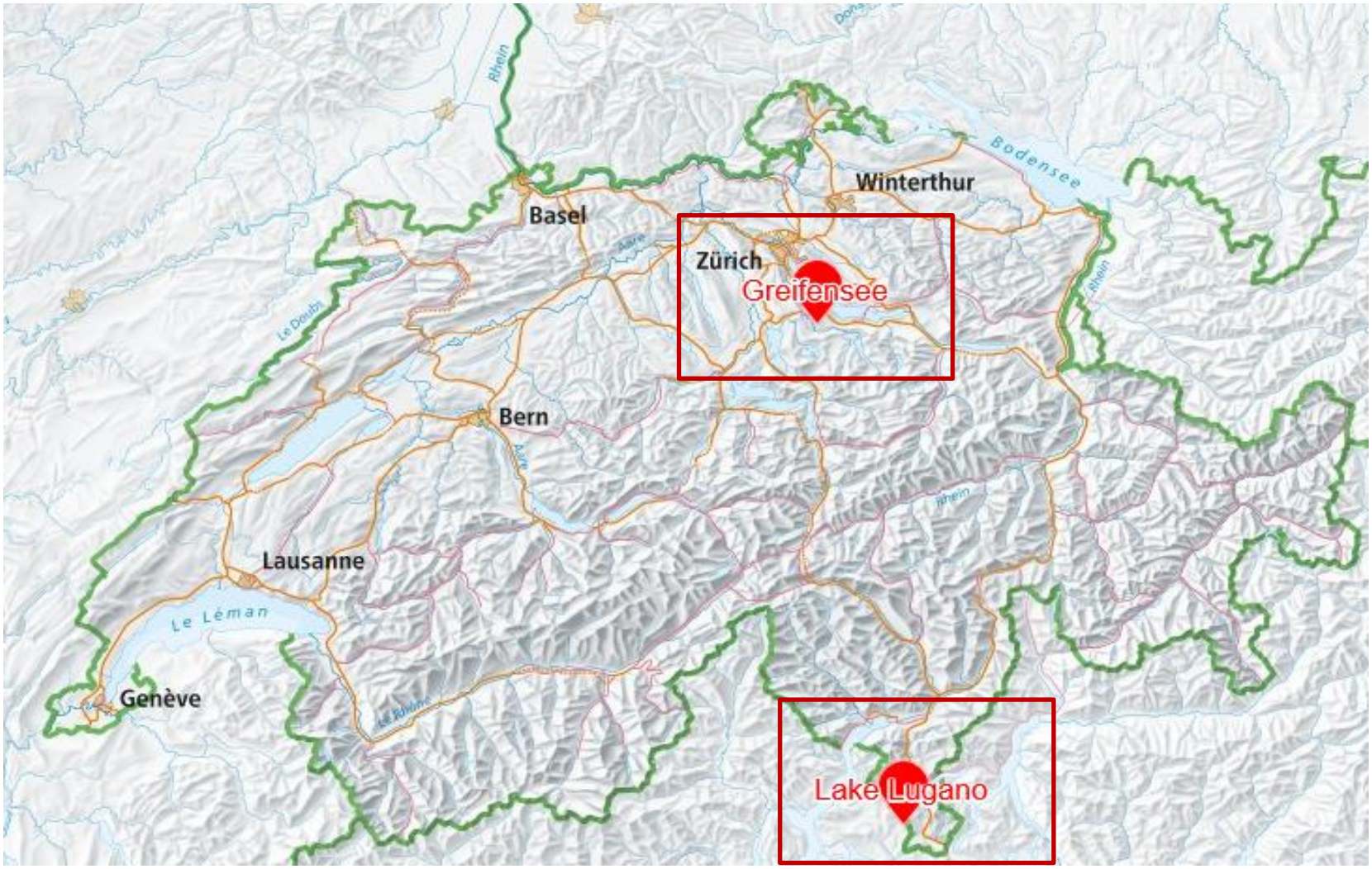




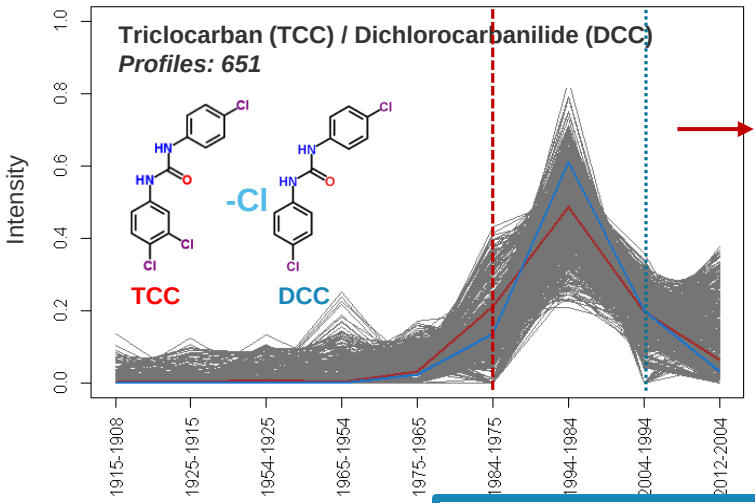
Hierarchical Clustering- Greifensee

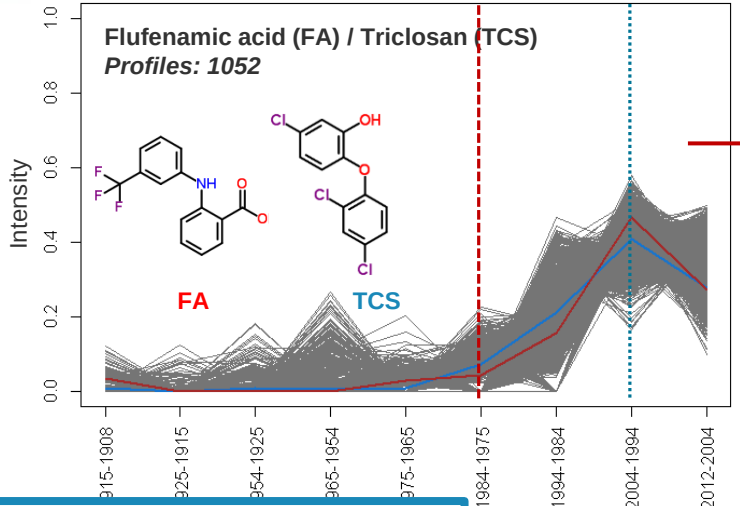


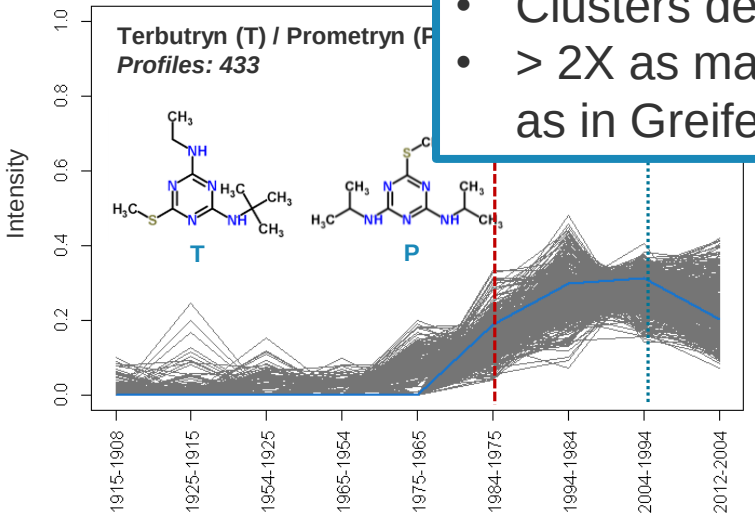
Hierarchical Clustering- Lake Lugano

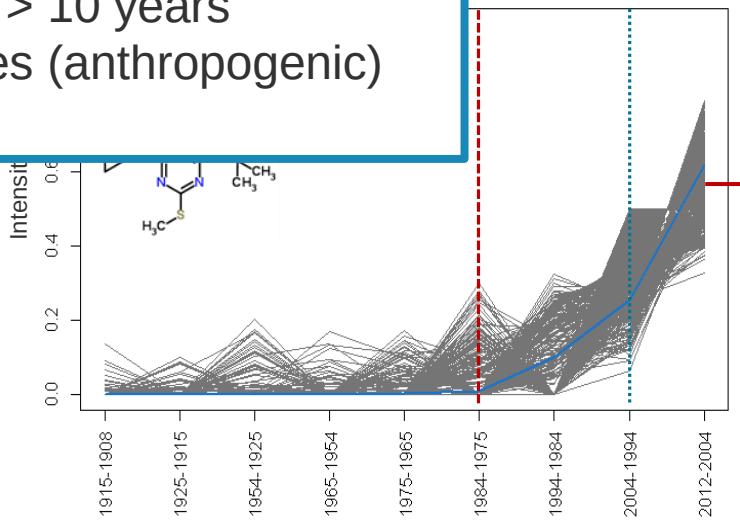


Hierarchical Clustering- Lake Lugano



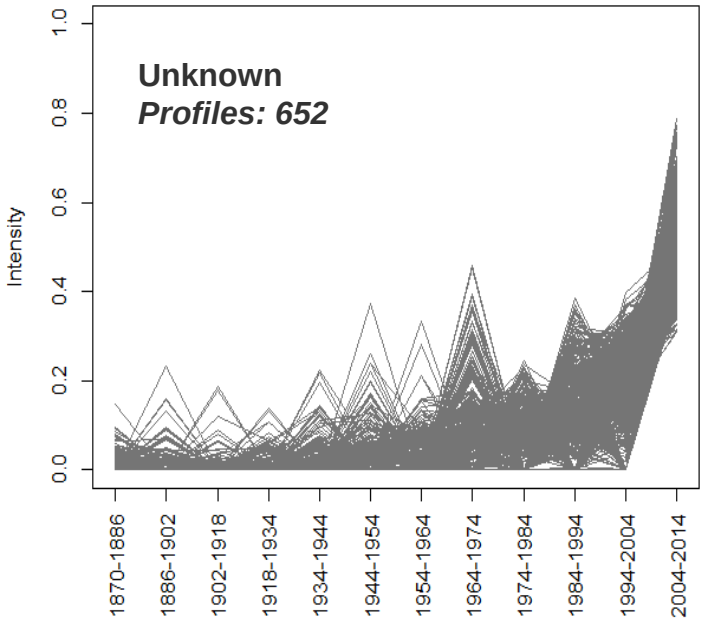


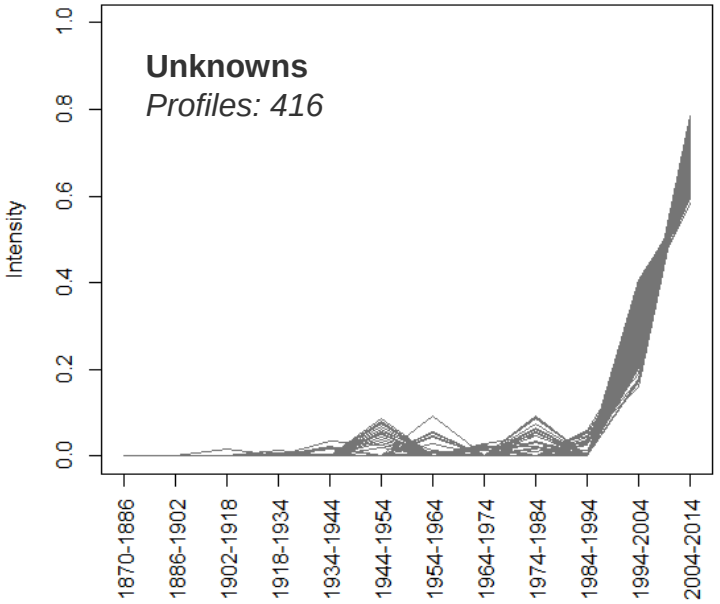




- Clusters delayed by > 10 years
- > 2X as many profiles (anthropogenic) as in Greifensee

Prioritization of Relevant Nontarget Contaminants





List of candidates

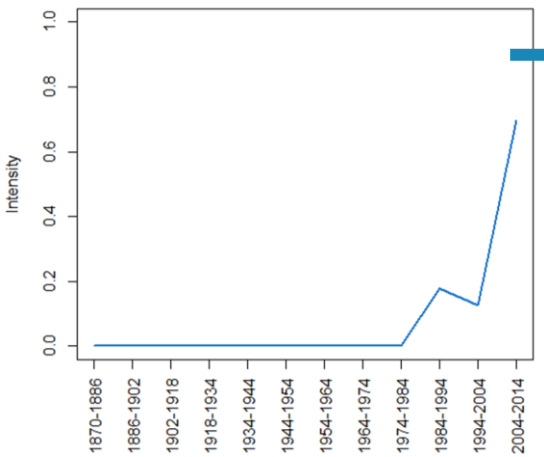
Re-measured at different collision energies (HCD) of 15, 35, and 55, 75 and 90 %

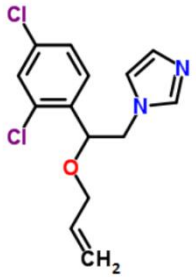
prioritized masses **(+)**

- **>10^6**
- **very characteristic isotopic pattern**

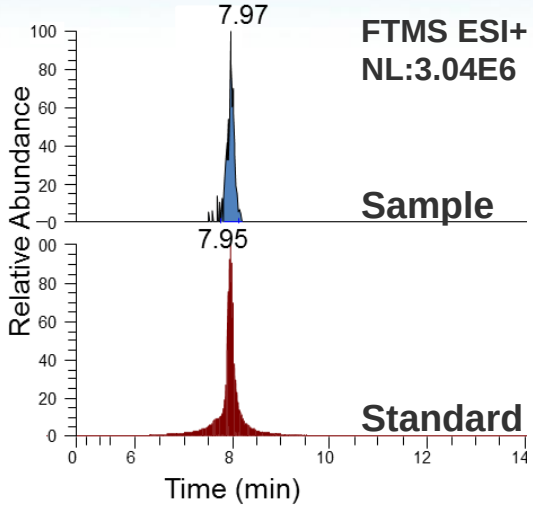
Match to different DB
MetFrag 2.3 (R version)

Identifying Nontarget Contaminants

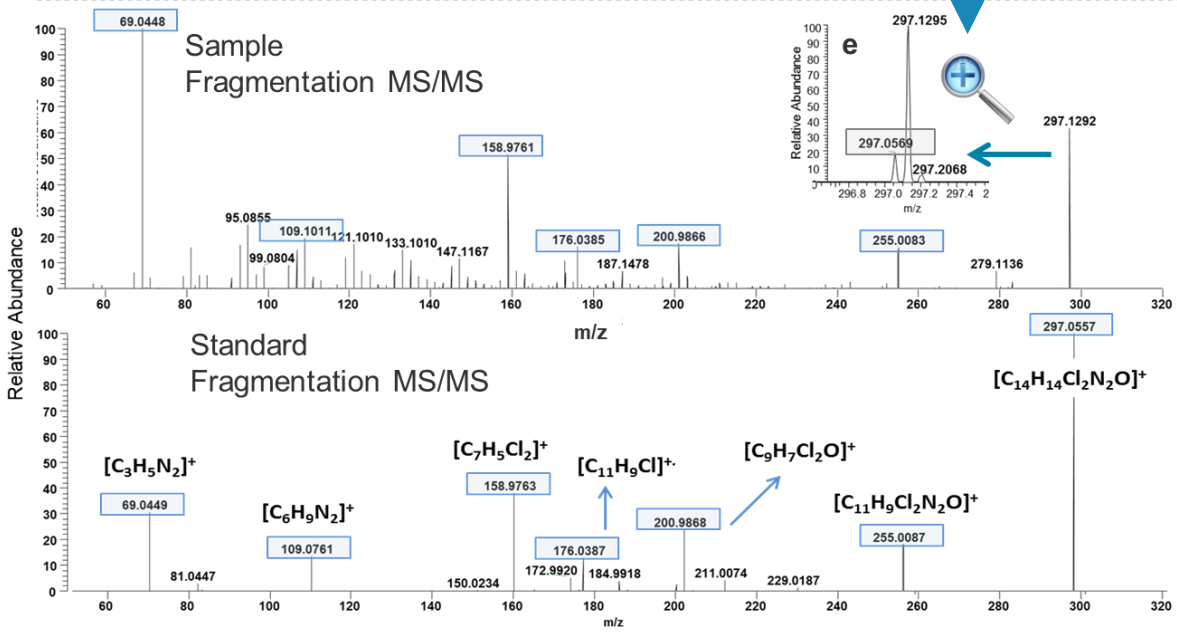




$C_{14}H_{14}Cl_2N_2O$
m/z:297.0549
*log K_{ow}: 3.44 (± 0.35)



Imazalil



- < 50 kg/year in agriculture
- Veterinary product (enilconazole)

- **"post-harvest fungicide in citrus fruit "**

Conclusion:

- Spearman's rank correlation coefficient help to identify substances with increasing trends over time
- Hierarchical clustering is a very useful method to evaluate the contamination and distinguish compounds with different trend patterns- *Transfer to other matrices*
- Prioritization of non-targets using statistical tools is promising to reduce matrix interferences and focus on relevant contaminants
- Sediments are useful to identify chemicals with unrecognized input pathways

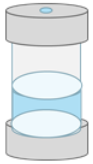
Thank You!

Adi Müller (eawag)
Martin Loos (eawag)
Emma Schymanski (eawag)
Heinz Singer (eawag)
Uchem Department-eawag
SURF Department-eawag

Funding:
Federal Office for the Environment (BAFU)
Academic Transition Grant (Eawag)

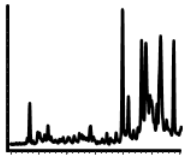
Analytical Methods

Preservation and Storage



Extraction and Enrichment

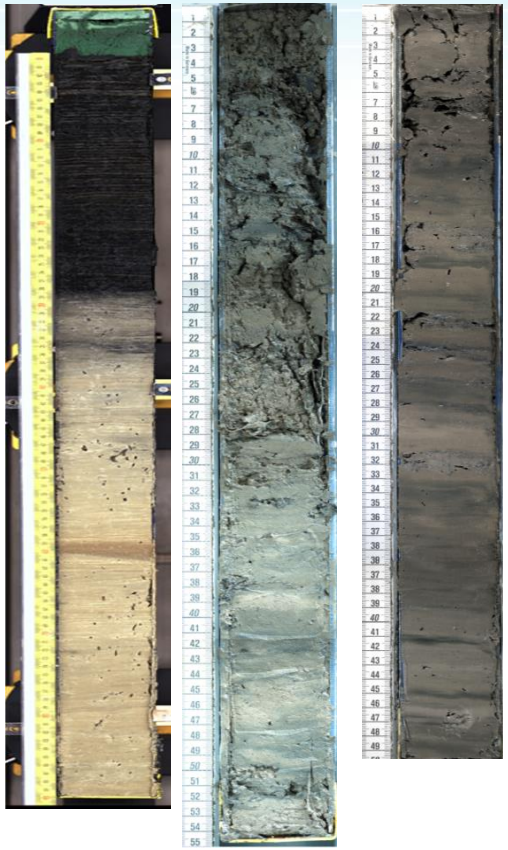
- In-cell clean up (Florisil)*



Chromatography

Orbitrap-MS
HR Chromatogram

1. Target
2. Suspect
3. Non-target

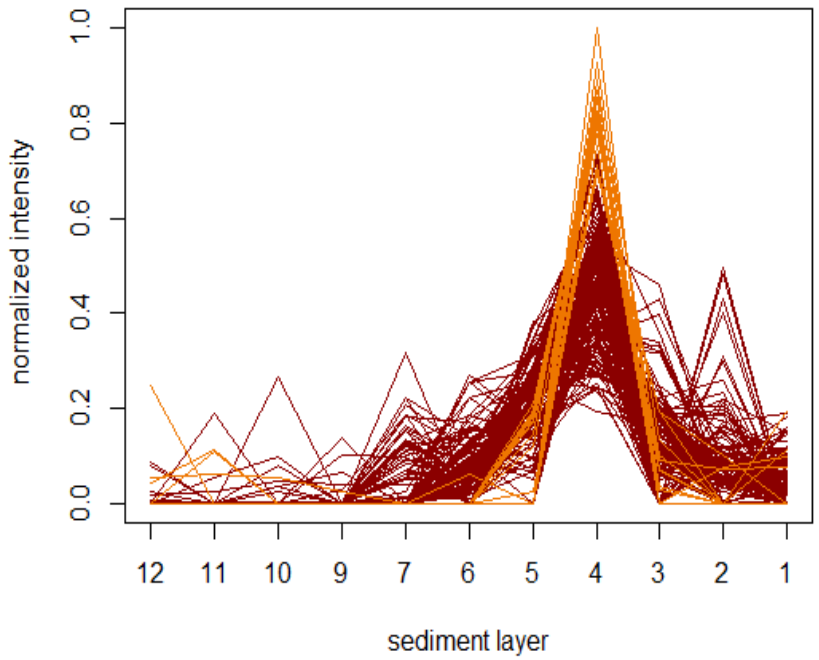


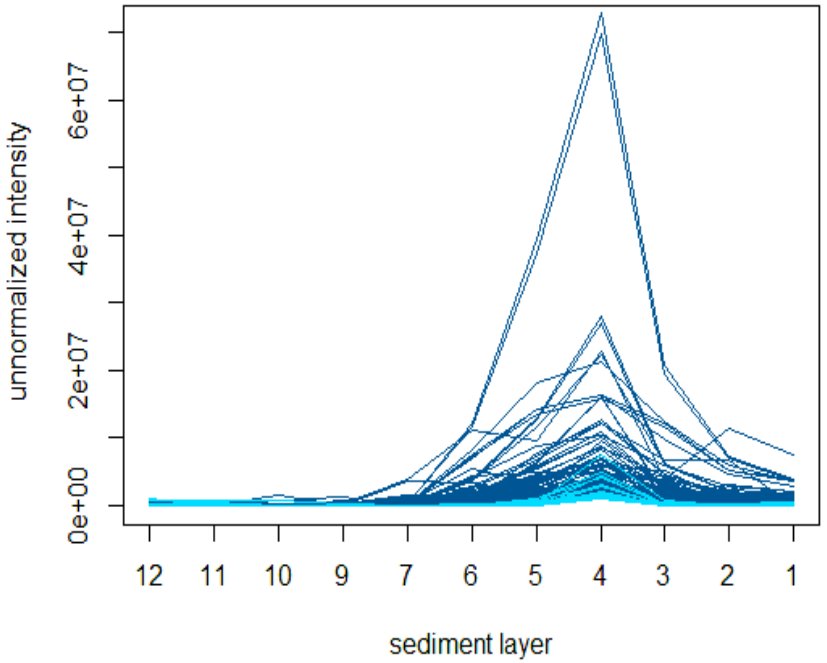
- Generic method
- Non-compound class specific

Chiaia-Hernandez et al. *ES&T* 2013, 47(2) pp. 976-986
*Pintado-Herrera et al. *Chemosphere* **2014,** *95*, 478-485

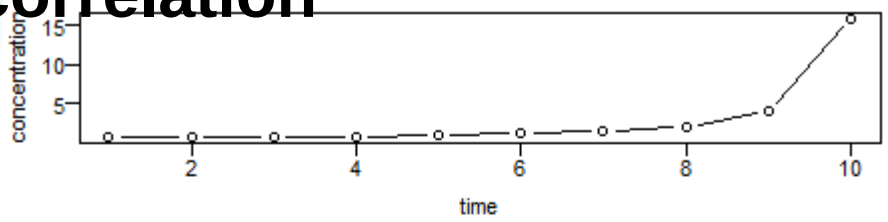
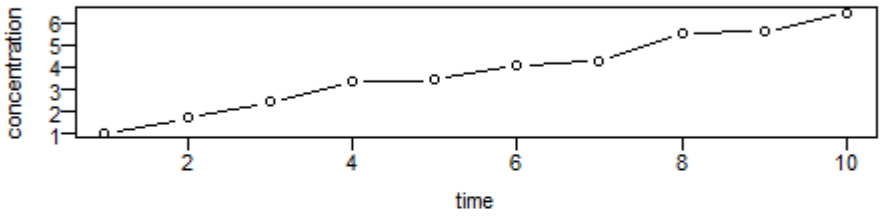
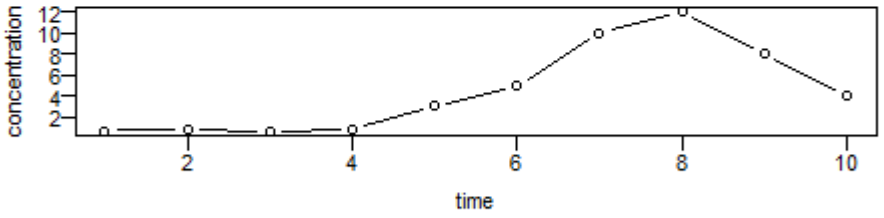
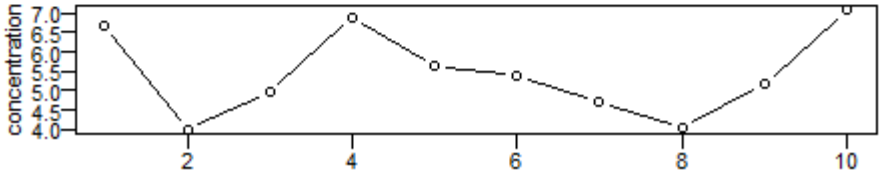
Appendix:

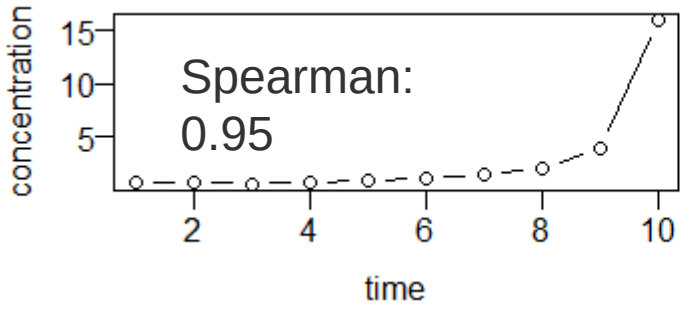
Normalization cluster analysis:

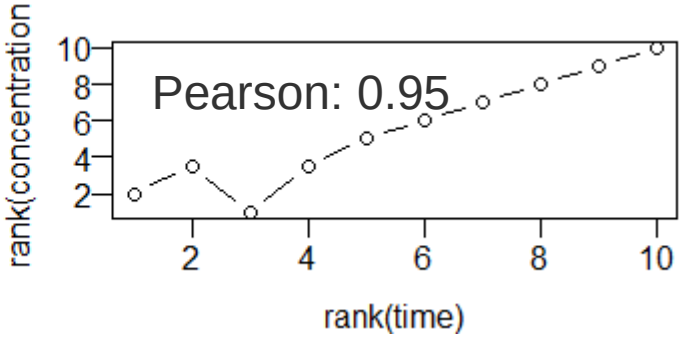




Trend characterization 1: Spearman's rank correlation

			Spearman	Pearson
1		interested	0.95	0.66
2		interested	1.00	0.99
3		not interested	0.80	0.73
4		not interested	0.10	0.00



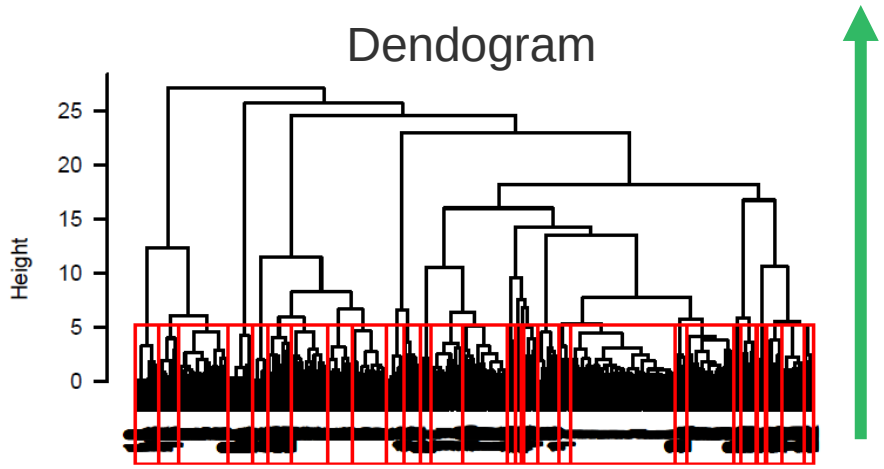


Cluster Analysis

- **Cluster Analysis (or Data segmentation)**
  - **Goals:** to group a collection of objects into subsets or "clusters“
  - Different types of cluster analysis

Hierarchical Clustering

- Classification and information retrieval
- Seeks to build a hierarchy of clusters
- Two basic paradigms: **Agglomerative (bottom up)** and divisive (top-down)



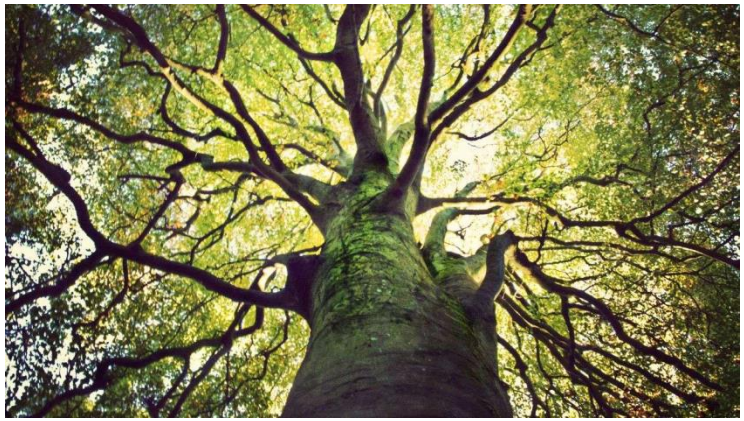
$"bottom - up"$

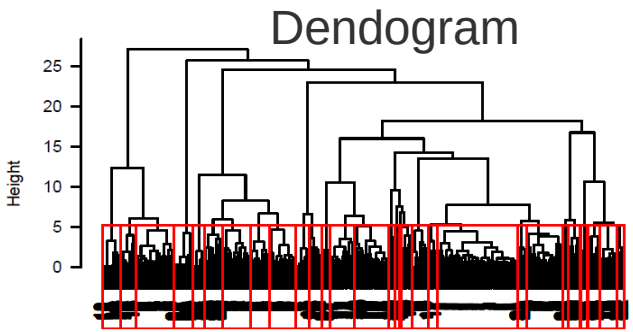
Hierarchical cluster analysis

Four main steps:

I) Normalization (over the sum)
II) Distance matrix (Euclidian distance)
III) **Hierarchical clustering (Ward's minimum variance method)**
IV) **Define number of clusters (some knowledge)**

In R
library(cluster)

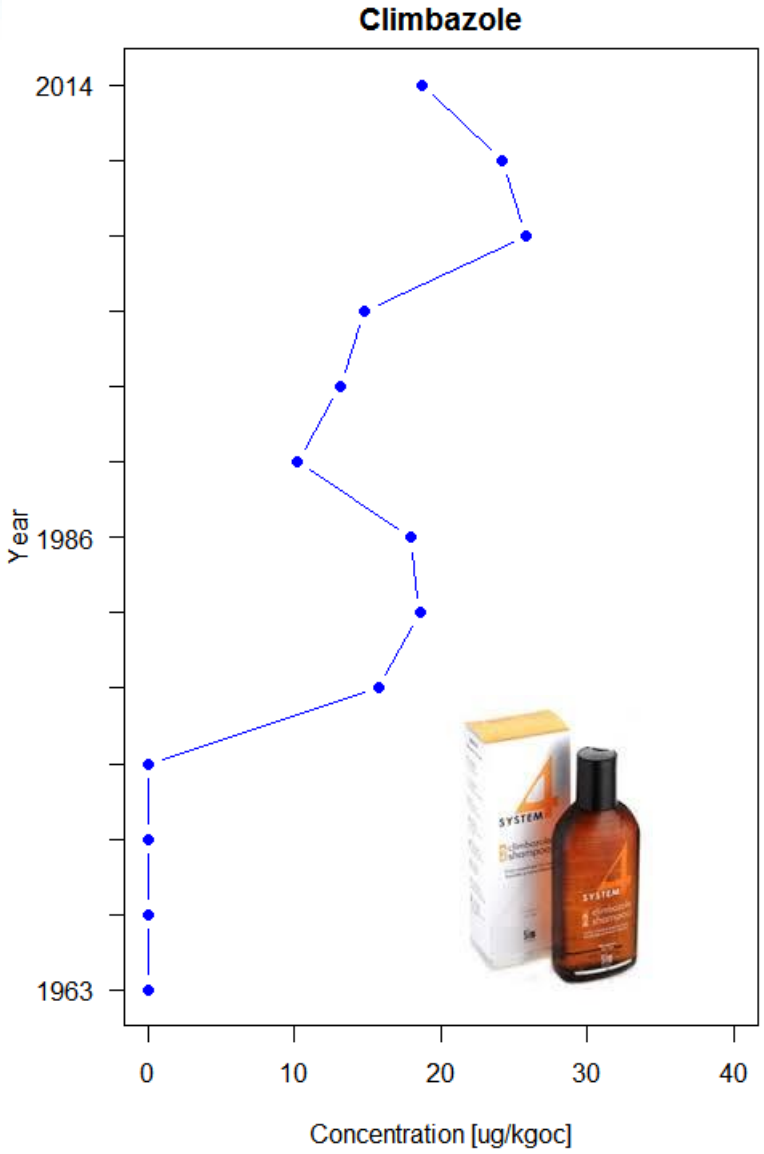


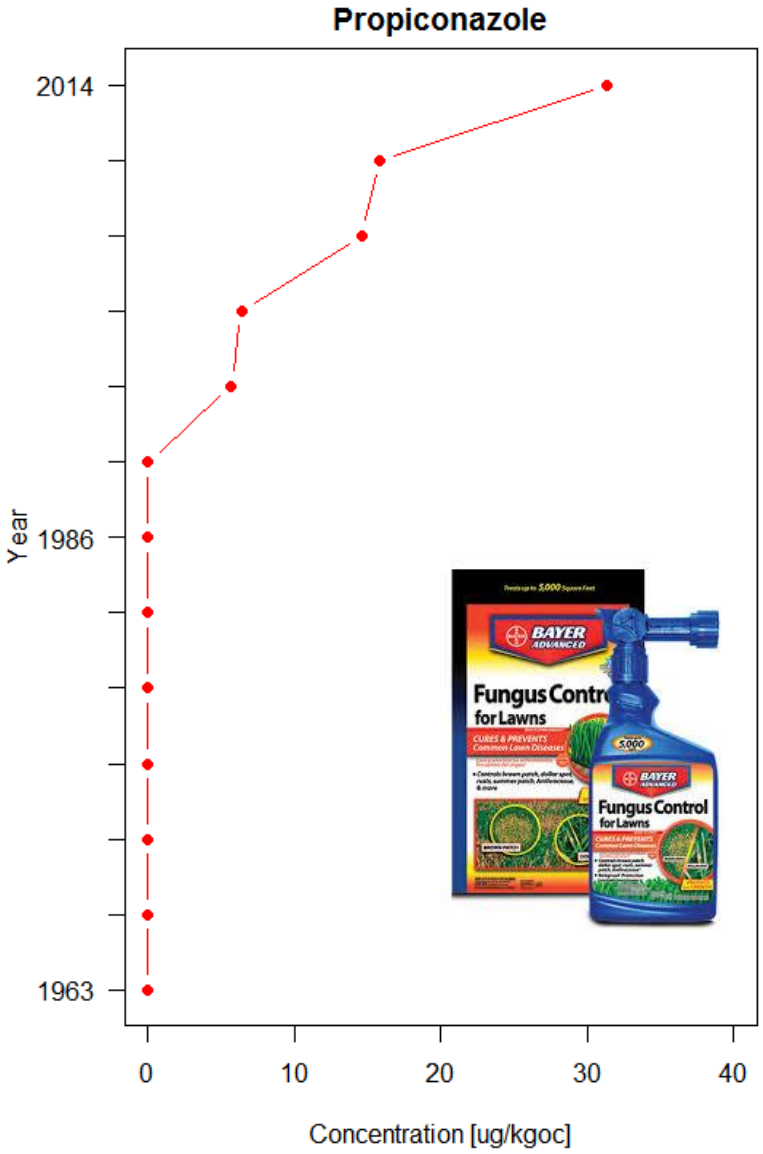


Function within the package:
Daisy: Distance matrix
Arguments - methods
euclidian= root sum-of-squares of differences

Function within the package:
Hclust
Arguments - methods
ward.D2 = Ward's minimum variance method

Different Temporal Trends-Target Analysis





Hierarchical cluster analysis

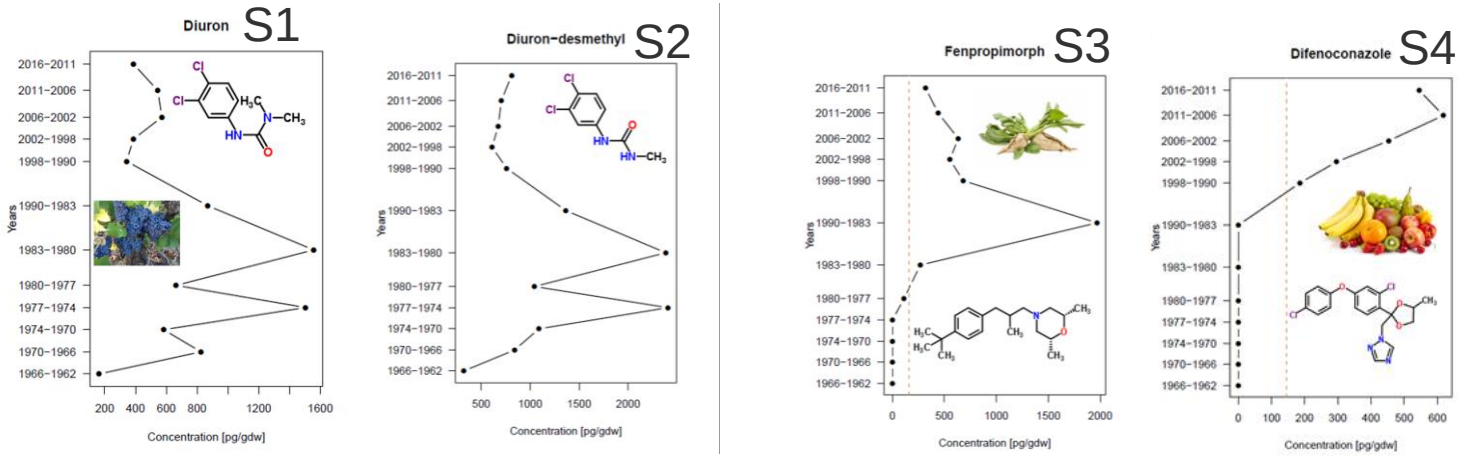
Four main steps:

I) Normalization (over the sum)

	S1	S2	S3	S4	……
S1	**0**	0.2	1.6	1.7	
S2	0.2	**0**	1.4	1.8	
S3	1.6	1.4	**0**	1.5	
S4	1.7	1.8	1.5	**0**	

hod)

- Euc
  aut
  - [



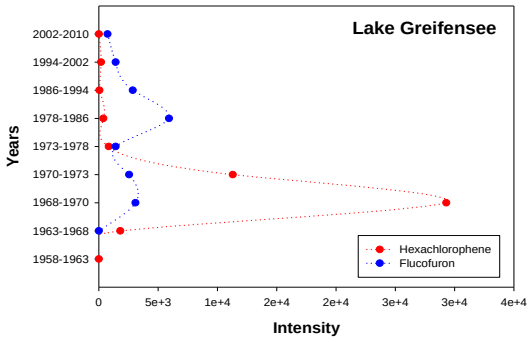
$$d(\mathbf{p},\mathbf{q}) = d(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2 + \cdots + (q_n-p_n)^2}$$

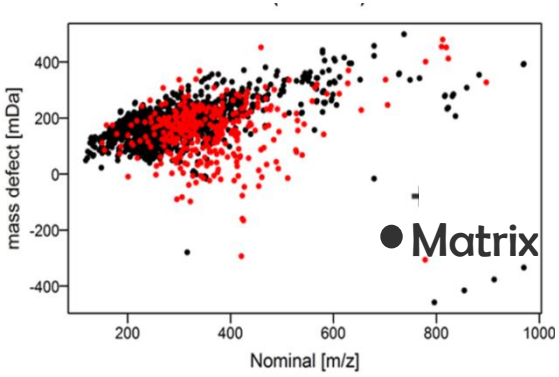
$$= \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}.$$

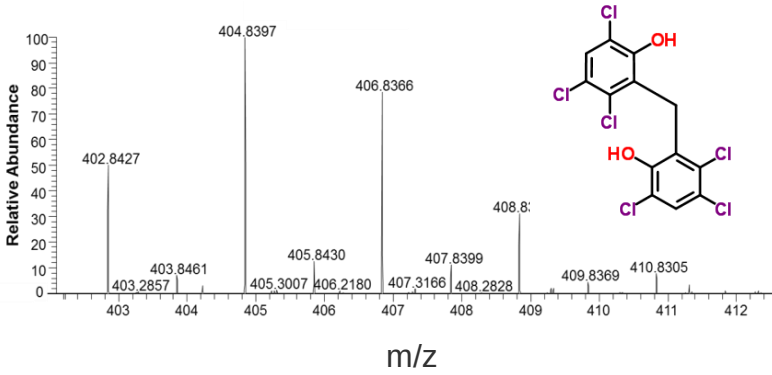
Challenges in Nontarget Screening - Unknown Compounds

- Different tools and approaches have been developed in recent years for structure elucidation
  - *In silico* fragmentation and including information on exposure, chromatographic retention, toxicity prediction

- **Pre-selection of relevant features**
  - Masses with increased identification probabilities
    - Highest intensities
    - Specific mass defects
    - Based on isotopic patterns (e.g. Cl, Br)







Chiaia-Hernandez et al. *Anal. Bioanal. Chem.* 2014, 406 (28), pp 7323-7335