La révolution NIRS dans l'agriculture et l'agro-alimentaire n’a pas fini de nous émerveiller

---

Quand Heliospir m’inspire !!

Vincent Baeten

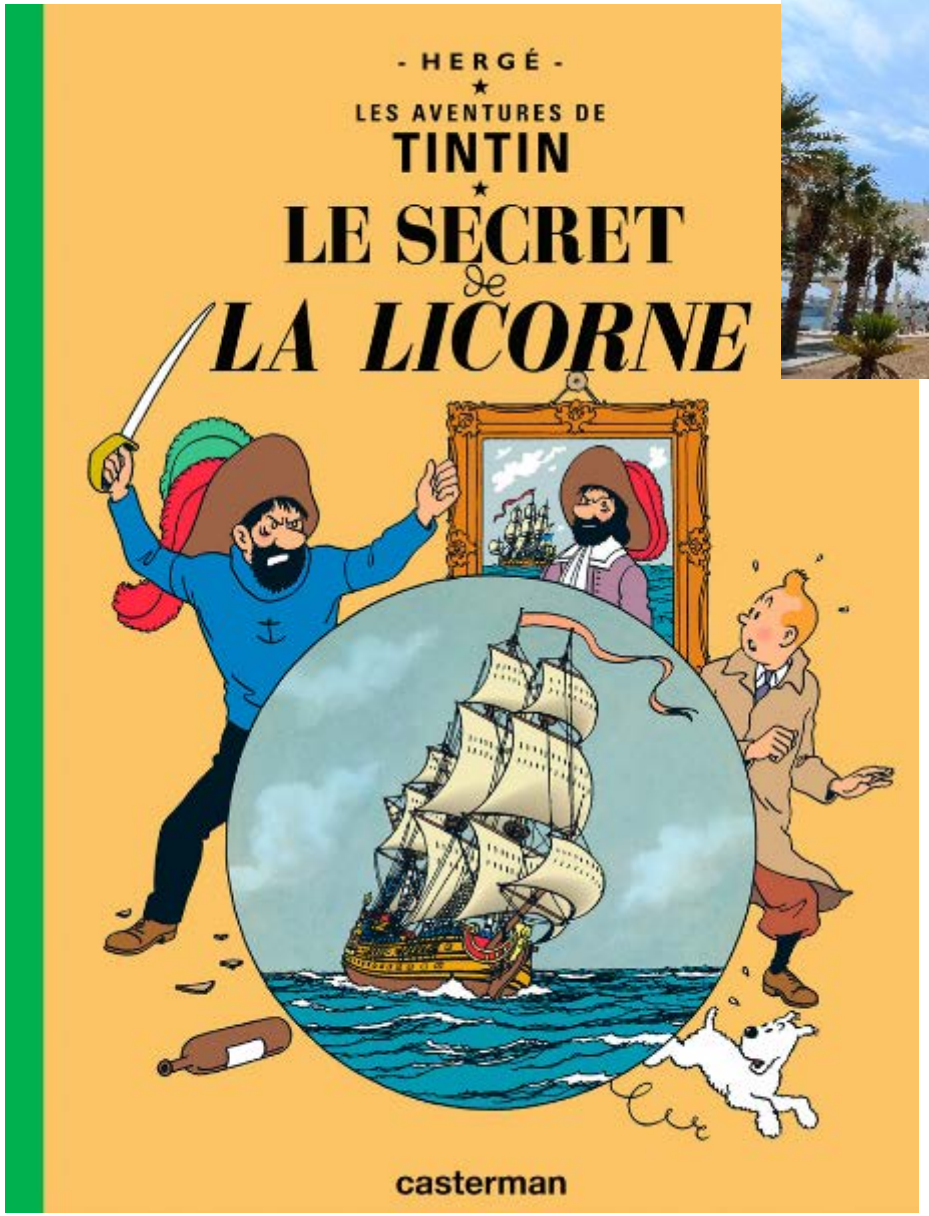
v.baeten@cra.wallonie.be

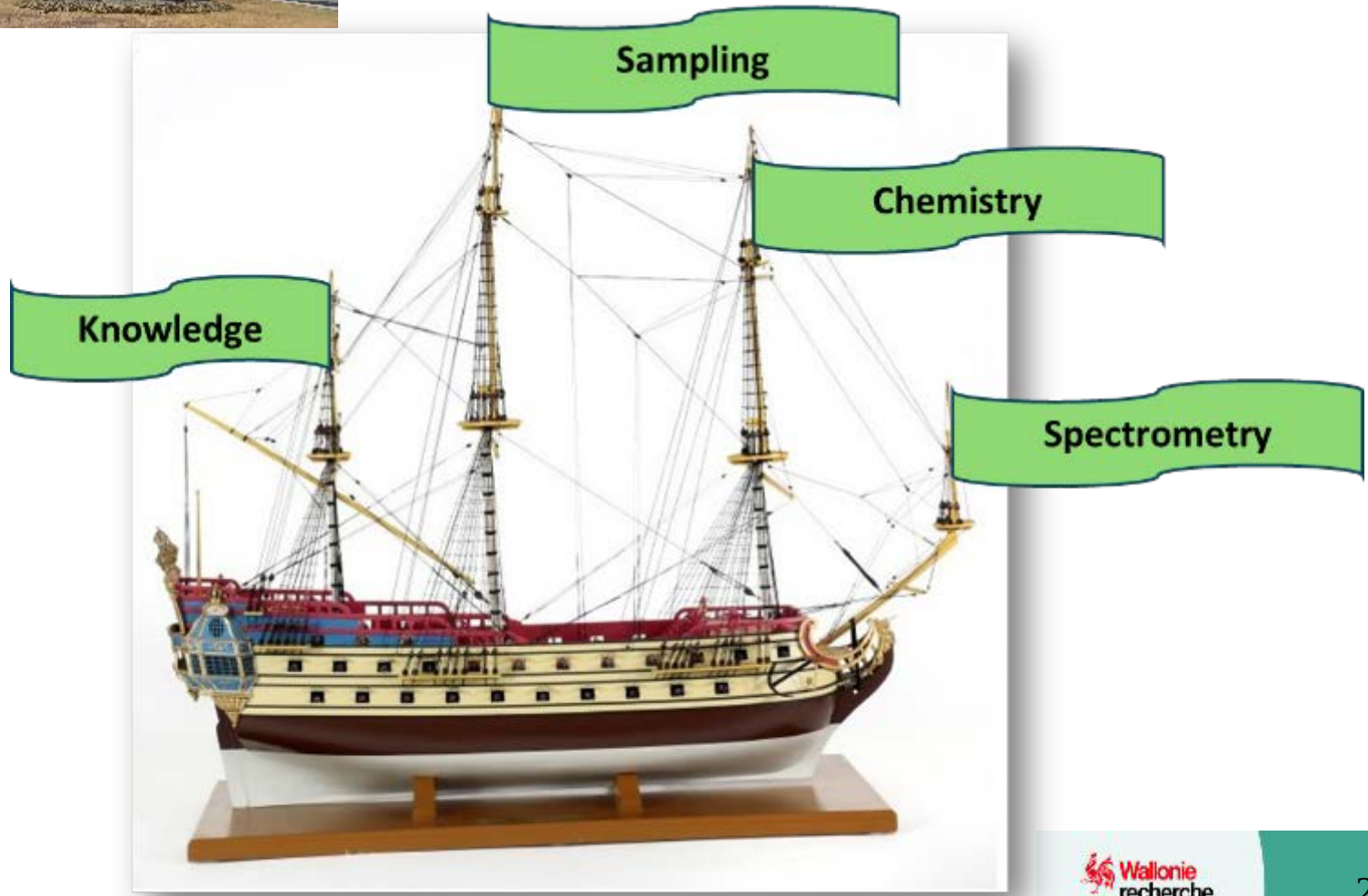
Quality and Authentication of Products Unit

Knowledge and valorization of agricultural products Department
Walloon Agricultural Research Centre – CRA-W
Gembloux, Belgium

Hélio SPIR 20 ans

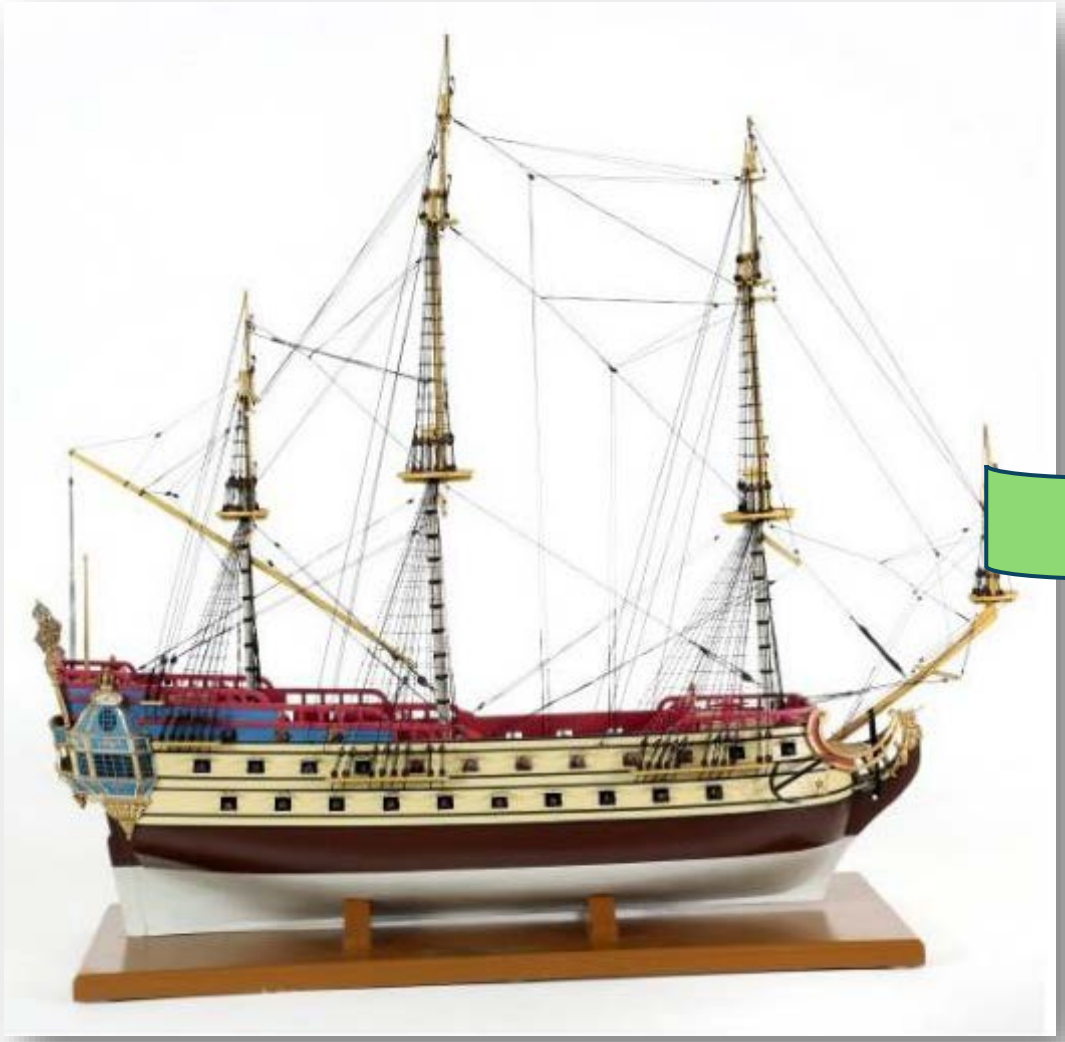
Wallonie recherche CRA-W





Hélio SPIR 20 ans

Wallonie recherche CRA-W

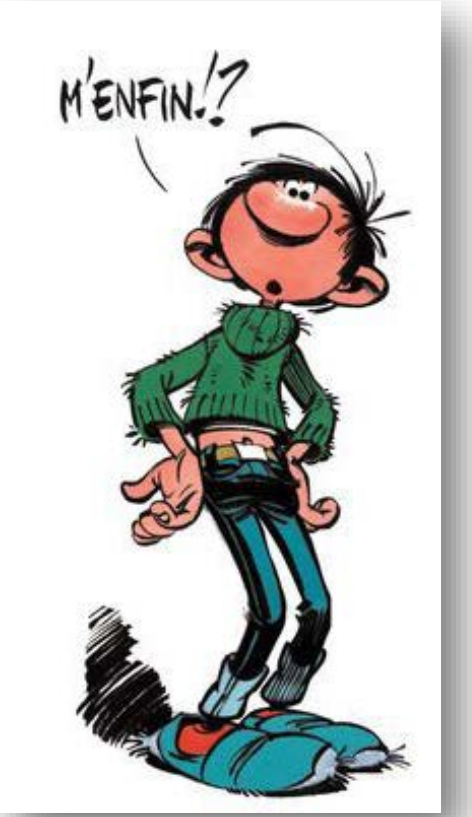


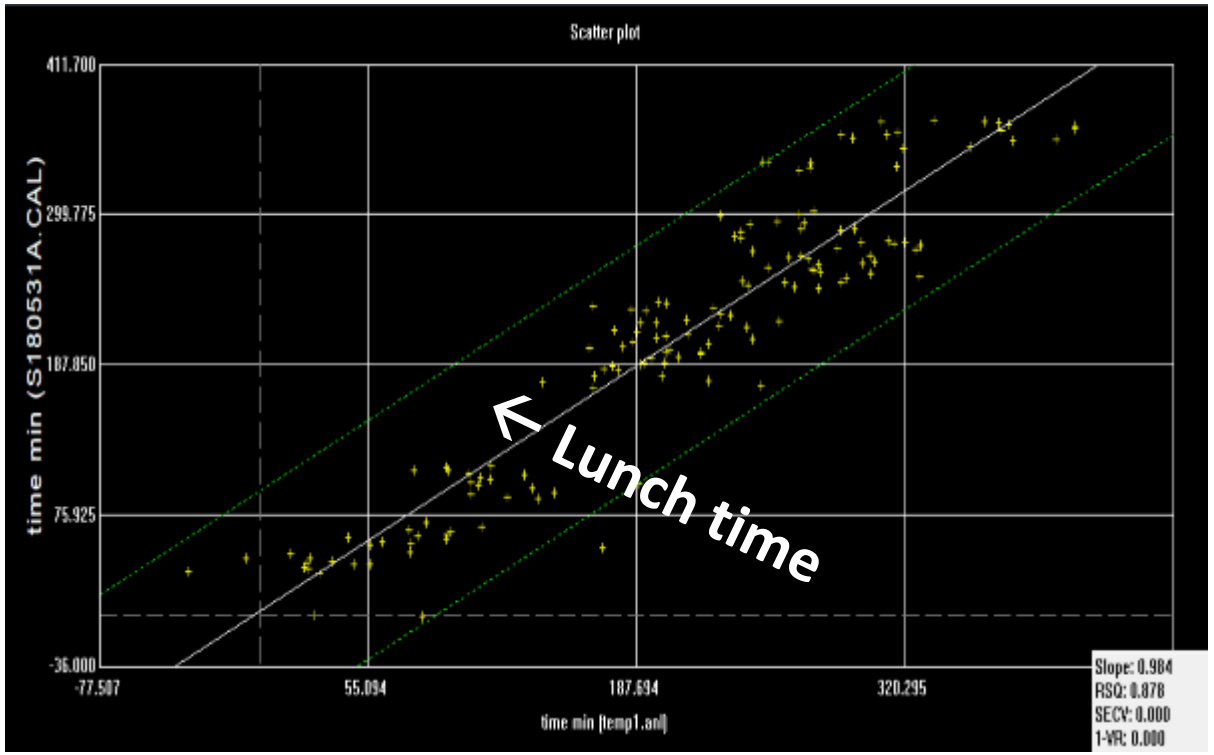
Spectrometry

- **Spectrometry** *instead of Spectroscopy*
- Decades of data … something **change**?
- What about **technology**?

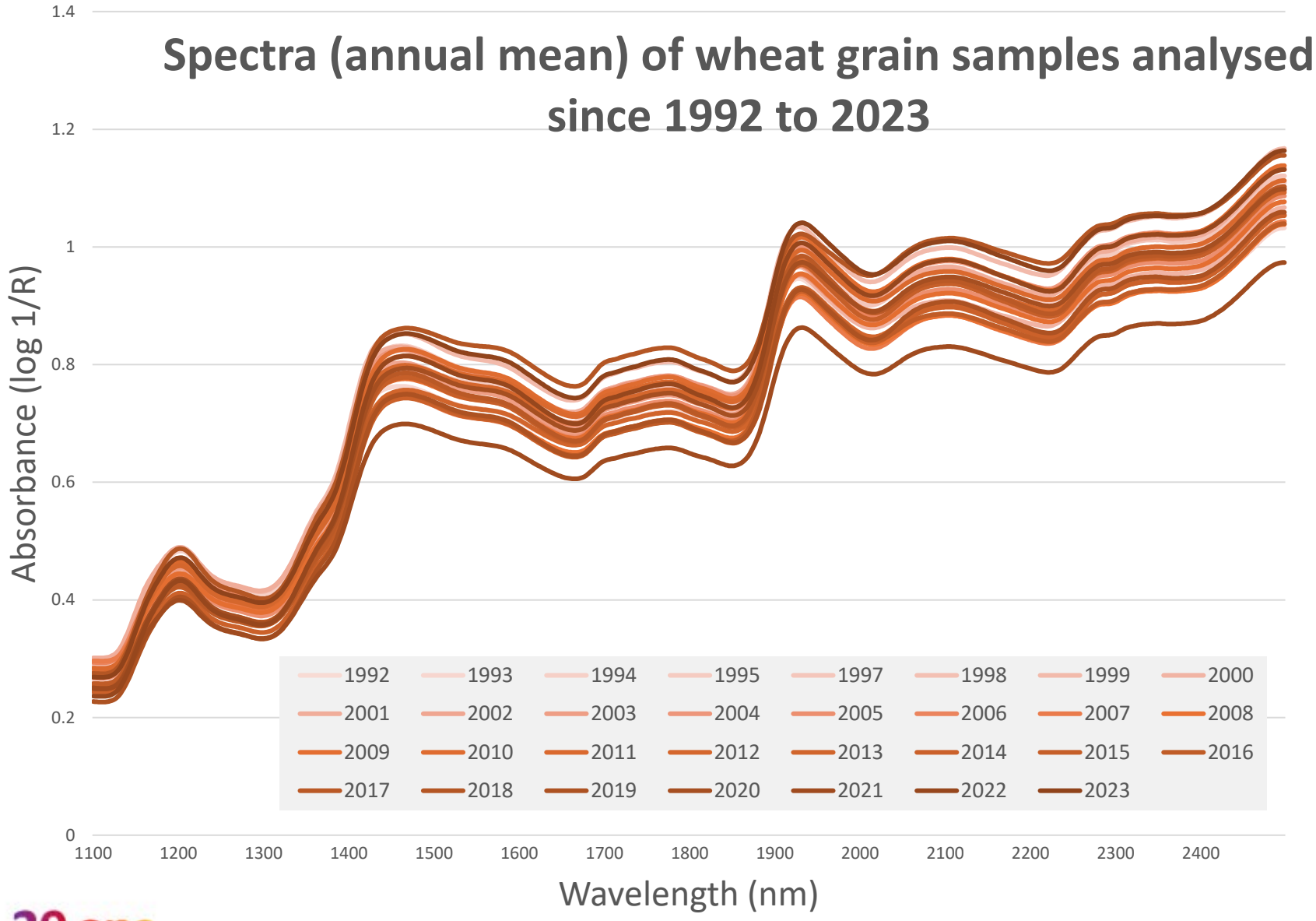
« *Gaston Lagaffe* » study

- Daily work analysis of 146 samples of wheat cereals (2021)
- Parameter : minutes of analysis (from 0 to 371 minutes)
- XDS instrument – quarter cup
- ISO17025 protocol

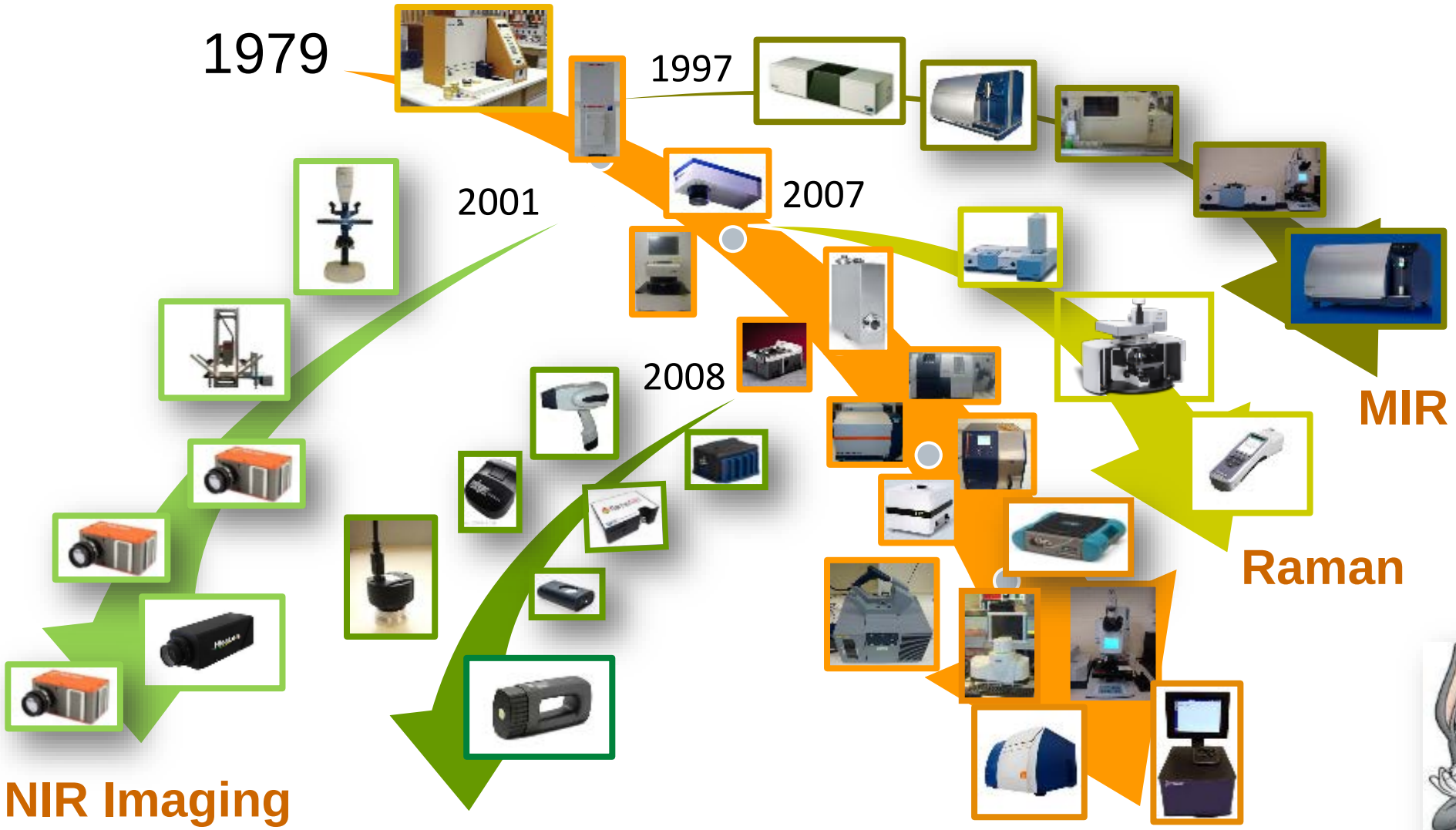




Decades of data … something change?



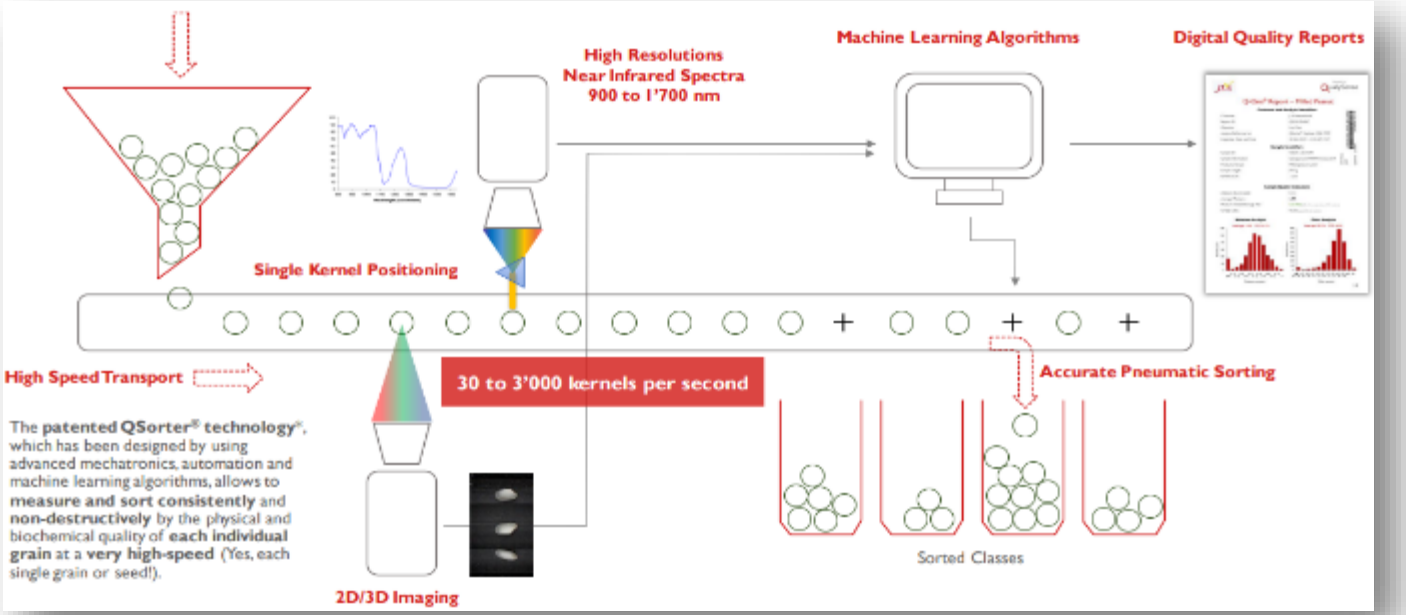
45 years of spectroscopy instrumentations - what's new?

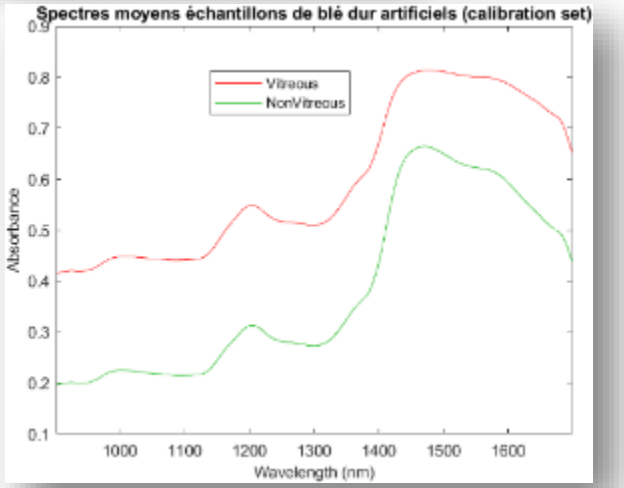


Instrumentations - what's new?

NEW TREND

Qsorter – kernel sorting (installed at CRA-W in 2022 – VALCERWAL project)

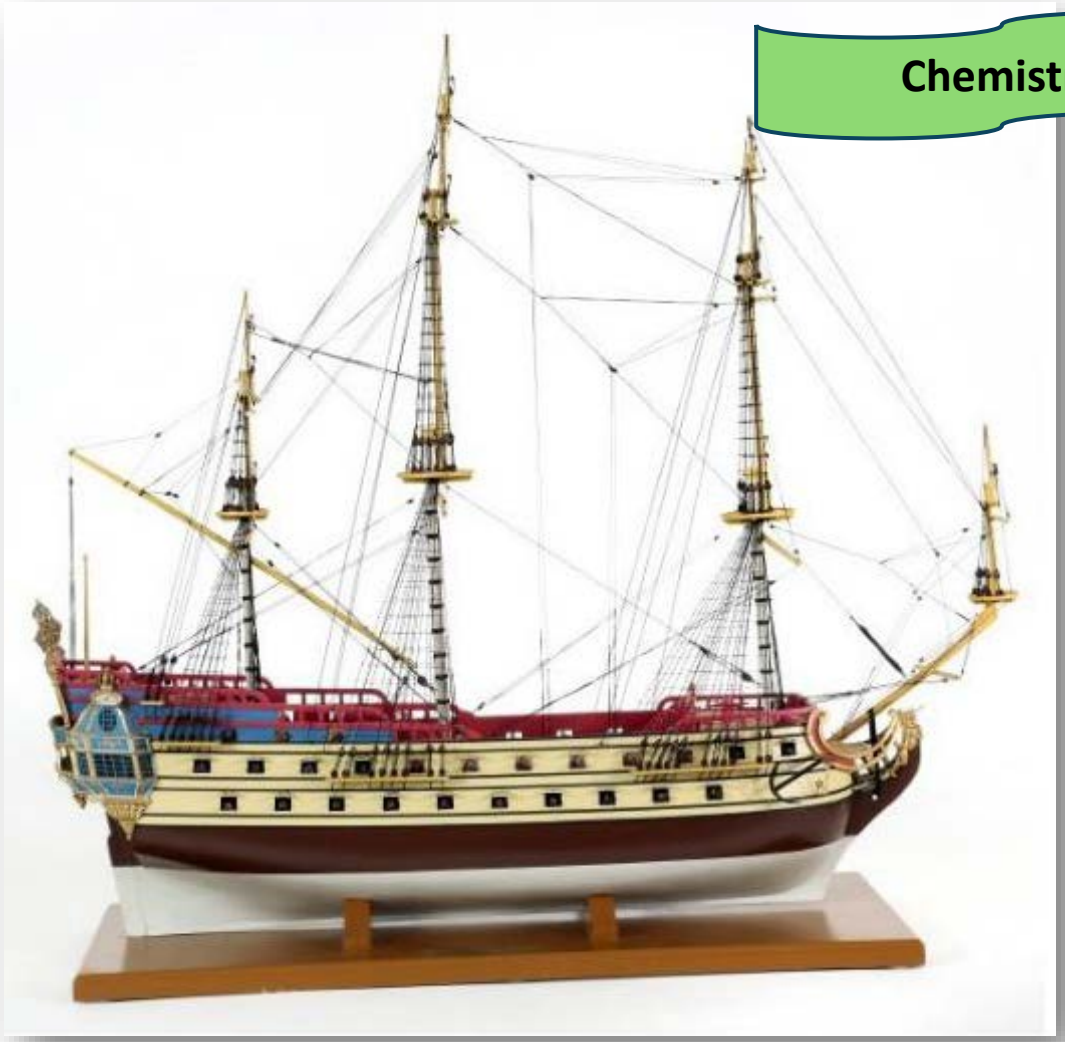




(Contacts : Bruno Godin, b.godin@cra.wallonie.be & Pierre-Yves Werrie, p.werrie@cra.wallonie.be)

Hélio SPIR 20 ans

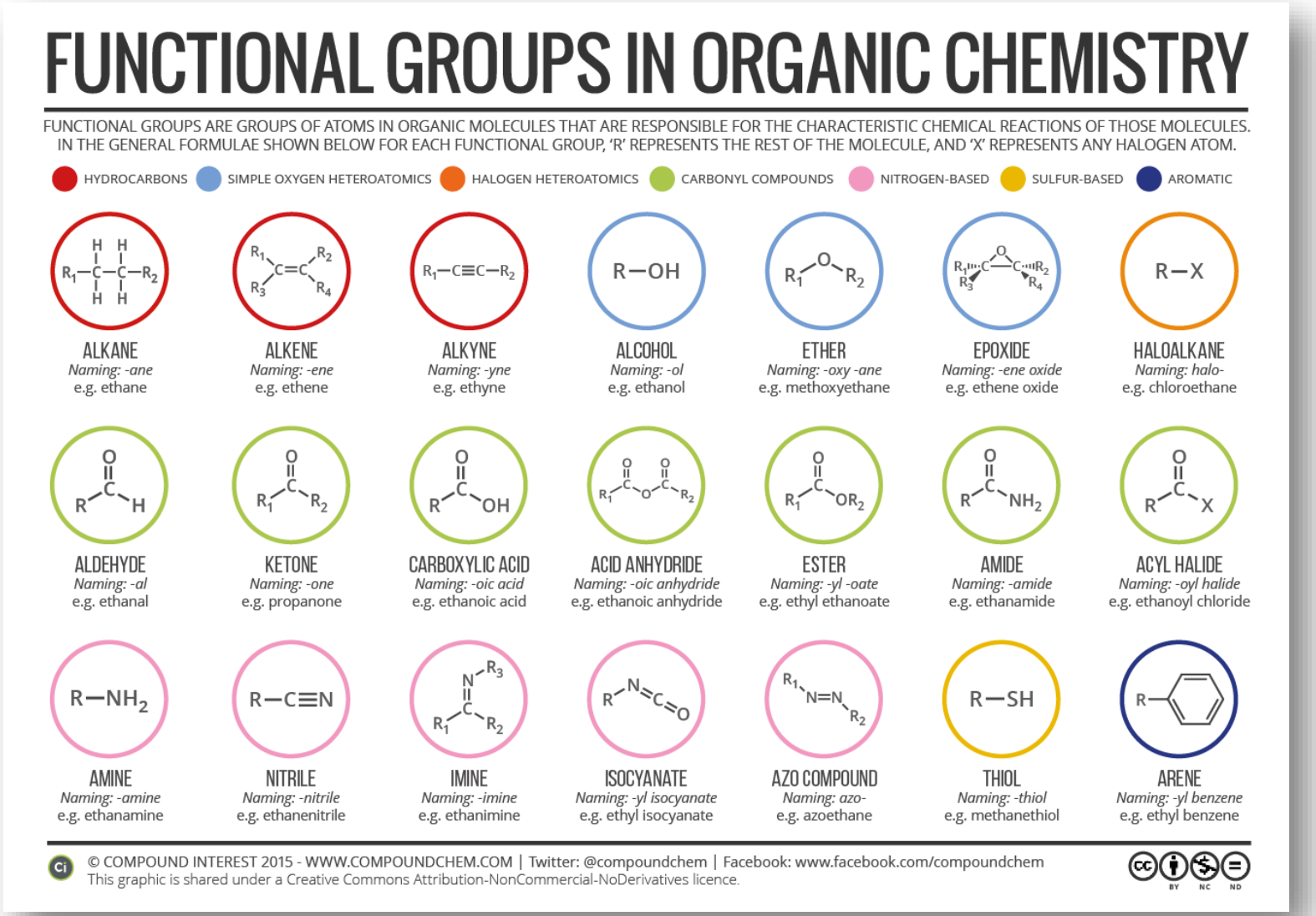
Wallonie recherche CRA-W



Chemistry

- ❑ Still the **basics!**
- ❑ NIR is a **Green** analytical technology
- ❑ We are doing **a good job**
- ❑ Is NIR always **the best**?

Our own experience and observation !



FUNCTIONAL GROUPS IN ORGANIC CHEMISTRY

FUNCTIONAL GROUPS ARE GROUPS OF ATOMS IN ORGANIC MOLECULES THAT ARE RESPONSIBLE FOR THE CHARACTERISTIC CHEMICAL REACTIONS OF THOSE MOLECULES. IN THE GENERAL FORMULAE SHOWN BELOW FOR EACH FUNCTIONAL GROUP, 'R' REPRESENTS THE REST OF THE MOLECULE, AND 'X' REPRESENTS ANY HALOGEN ATOM.

HYDROCARBONS · SIMPLE OXYGEN HETEROATOMICS · HALOGEN HETEROATOMICS · CARBONYL COMPOUNDS · NITROGEN-BASED · SULFUR-BASED · AROMATIC

ALKANE *Naming: -ane* e.g. ethane	**ALKENE** *Naming: -ene* e.g. ethene	**ALKYNE** *Naming: -yne* e.g. ethyne	**ALCOHOL** *Naming: -ol* e.g. ethanol	**ETHER** *Naming: -oxy -ane* e.g. methoxyethane	**EPOXIDE** *Naming: -ene oxide* e.g. ethene oxide	**HALOALKANE** *Naming: halo-* e.g. chloroethane
ALDEHYDE *Naming: -al* e.g. ethanal	**KETONE** *Naming: -one* e.g. propanone	**CARBOXYLIC ACID** *Naming: -oic acid* e.g. ethanoic acid	**ACID ANHYDRIDE** *Naming: -oic anhydride* e.g. ethanoic anhydride	**ESTER** *Naming: -yl -oate* e.g. ethyl ethanoate	**AMIDE** *Naming: -amide* e.g. ethanamide	**ACYL HALIDE** *Naming: -oyl halide* e.g. ethanoyl chloride
AMINE *Naming: -amine* e.g. ethanamine	**NITRILE** *Naming: -nitrile* e.g. ethanenitrile	**IMINE** *Naming: -imine* e.g. ethanimine	**ISOCYANATE** *Naming: -yl isocyanate* e.g. ethyl isocyanate	**AZO COMPOUND** *Naming: azo-* e.g. azoethane	**THIOL** *Naming: -thiol* e.g. methanethiol	**ARENE** *Naming: -yl benzene* e.g. ethyl benzene

Oil and Fat Classification by Selected Bands of Near-Infrared Spectroscopy

PIERRE HOURANT, VINCENT BAETEN, MARIA T. MORALES, MARC MEURENS, and RAMON APARICIO*

One hundred and four edible oil and fat samples from 18 different sources, either vegetable (Brazil nut, coconut, corn, sunflower, walnut, virgin olive, peanut, palm, canola, soybean, sunflower) or animal (tallow and hydrogenated fish), have been analyzed by high-performance gas chromatography (HPGC) and near-infrared spectroscopy (NIRS). Fatty acids were quantified by HPGC. The near-infrared spectral features of the most noteworthy bands were studied and discussed to design a filter-type NIR instrument. An arborescent structure, based on stepwise linear discriminant analysis (SLDA), was built to classify the samples according to their sources. Seven discriminant functions permitted a successive discrimination of saturated fats, corn, soybean, sunflower, canola, peanut, high oleic sunflower, and virgin olive oils. The discriminant functions were based on the absorbance values, between three and five, from the 1710–1800 and 2100–2400 nm regions. Chemical explanations are given in support of the selected wavelengths. The arborescent structure was then checked with a test set, and 90% of the samples were correctly classified.

Index Headings: Oils and fats; Near-infrared spectroscopy; Authentication; Statistics.

INTRODUCTION

In the literature, there is not complete agreement on the definition for characterization of a food product, and confusions about the authentication term are widespread. Characterization of a food product could be defined as

electron ionization mass spectrometry,[9] nuclear magnetic resonance,[10] and ultraviolet,[11] infrared,[12–18] and Raman[19,20] spectroscopies also render good results in combination with chemometrics algorithms.[3]

Concerning the last group of spectroscopic techniques, some papers have been published setting out the possibilities offered to detect adulteration or purity of edible oils and fats by near-infrared (NIR) spectroscopy,[21–26] by mid-infrared spectroscopy,[27–30] and by Raman spectroscopy.[31–33] From a mathematical viewpoint, only recently have chemometrics procedures been used in the authentication or characterization of edible fats and oils. Most of these procedures are nonsupervised, and they are applied to the whole spectra where noise or random data can have such a pronounced influence on the results that they can make them nonreproducible.[3]

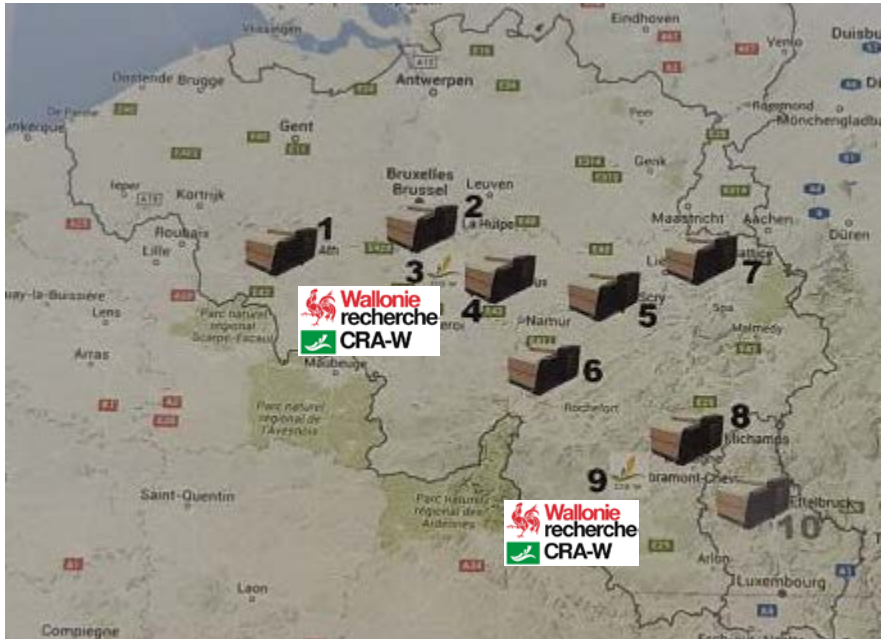
The aim of the present study is to demonstrate that the absorbances at few wavelengths are sufficient for the discrimination of several sources of oils and fats, the results being supported by chemical explanations. To reach this objective, the work combined the advantages of NIR spectroscopy (reliability, speed, absence of solvents, etc.) with a supervised statistical procedure [stepwise linear discriminant analysis (SLDA)] to classify a large set of different kinds of fats and oils. The statistical procedure was used to select the wavelengths of the spectrum with

TABLE III. Assignments of the major near-infrared absorption bands of edible oil spectrum.

Region	Wavelength (nm)	Molecule	Group	Vibration
A	1090–1180	$-CH_2$	C–H	2nd overtone
(1100–1300 nm)	1100–1200	$-CH_3$	C–H	2nd overtone
	1150–1260	–CH=CH–	C–H	2nd overtone
B	1350–1430	$-CH_2$	C–H	Combination
(1300–1600 nm)	1360–1420	$-CH_3$	C–H	Combination
	1390–1450	H_2O	O–H	1st overtone
C		$-CH_2$	C–H	1st overtone
(1600–1850 nm)	1650–1780	$-CH_3$	C–H	1st overtone
		–CH=CH–	C–H	1st overtone
D	1880–1930	H_2O	O–H	Combination
(1850–2050 nm)	2010–2020	–CH=CH–	C–H	Combination
E	2100–2200	–CH=CH–	C–H	Combination
(2050–2230 nm)				
F	2240–2360	$-CH_3$	C–H	Combination
(2230–2500 nm)	2290–2470	$-CH_2$	C–H	Combination

Our own experience and observation !

Reagent/chemical	Reference values	NIR values
Reagent	☹	☺ (95 % reduction)
Time / sample / technician	**4 hours**	**5 minutes**
Samples / day/ technician	**5-10**	**100**



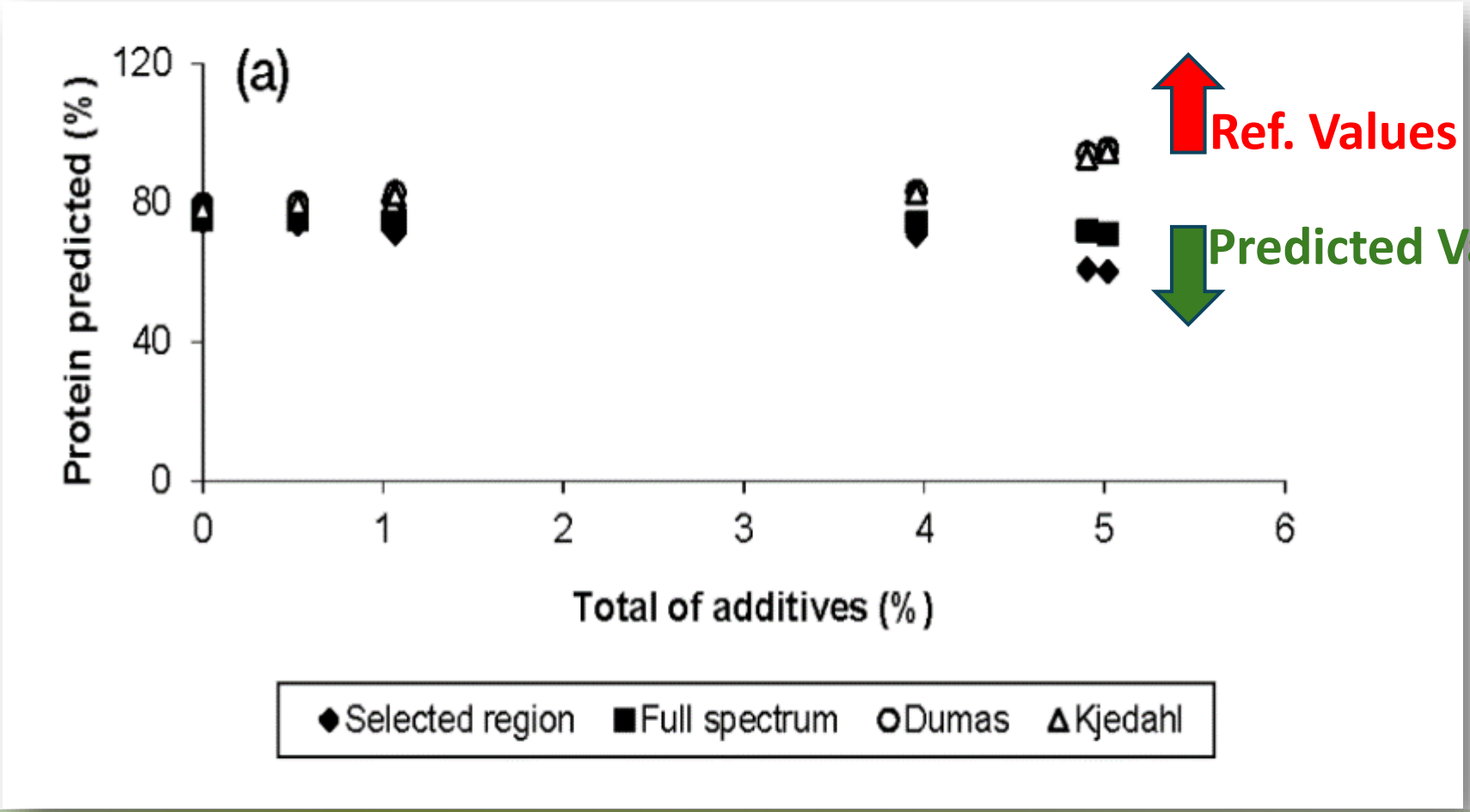
requasud
réseau de laboratoires wallons

ISO 17025 & ISO 17043

Hélio SPIR 20 ans

*(Source: Armenta et al. Chapter 1 - **Green Analytical Chemistry: concepts, evolution, and recent developments**, Editor(s): Emanuela Gionfriddo, Green Approaches for Chemical Analysis, Elsevier, 2023, pp. 1-37)*

Our own experience and observation !



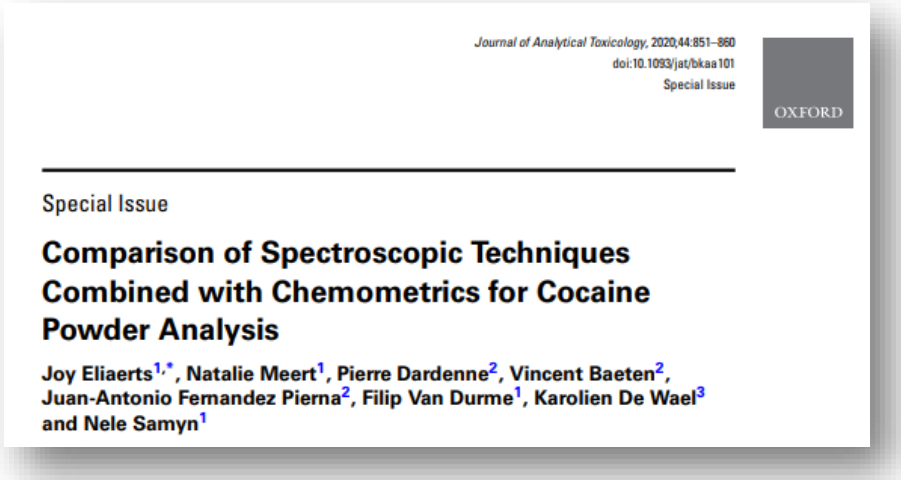
Ref. Values (Kjeldahl, Dumas)

Predicted Values (NIRS)

Classical regression models help in adulteration detection

Abbas et al. (2013) ***Detection of melamine in feed ingredients by near infrared spectroscopy and chemometrics****. Journal of Near Infrared Spectroscopy, 21(3), 183-194.*

Our own experience and observation !



- natural, illicit drug
- stimulant
- high prevalence: Belgium 'the cocaine capital of Europe'war on drugs
- pure or mixed with cutting agents a highly variable matrix

SVM classification

TECHNIQUES	MIR	Raman	NIR
accuracy	99.8% (1 FN)	99.6% (2 FN)	**97.7% (4 FP)**

SVM quantification

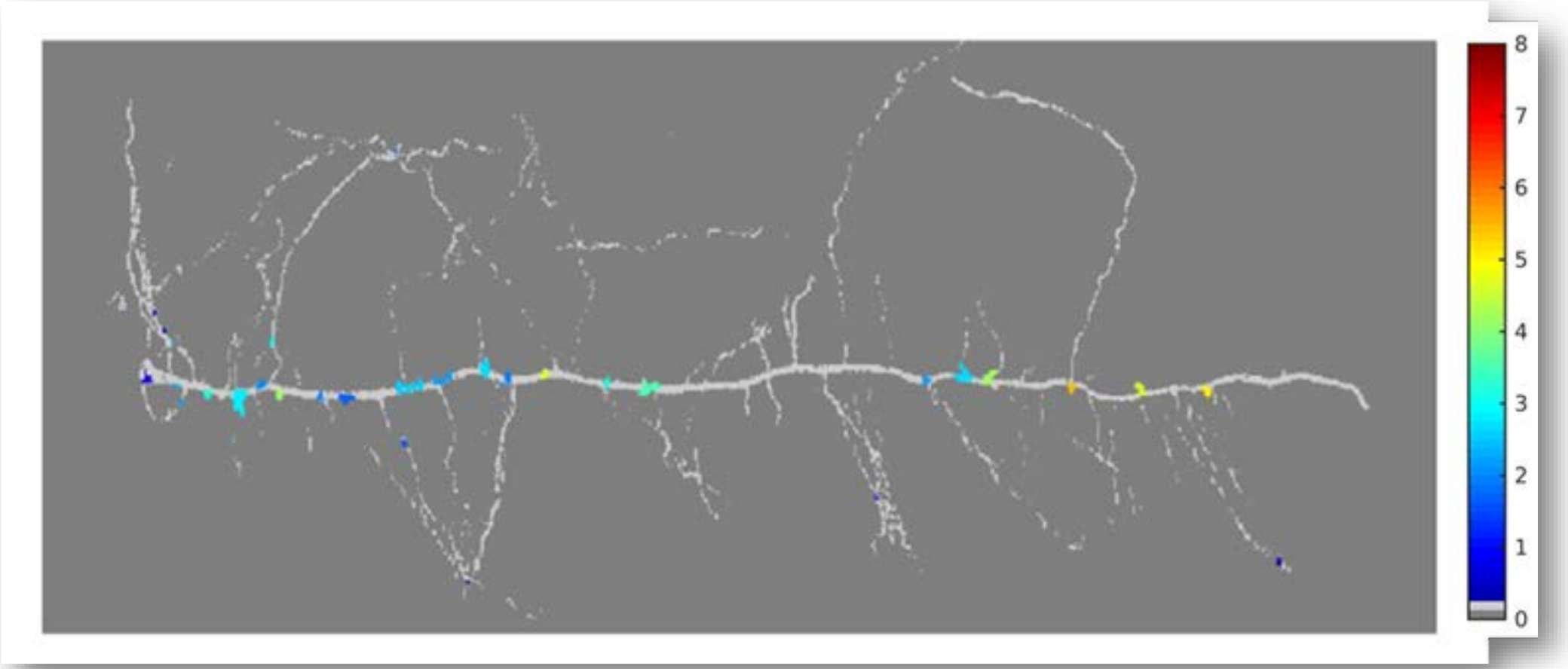
TECHNIQUES	CROSS-VALIDATION		
	RMSECV	R^2cv	bias
MIR	6.76	0.90	0.30
Raman	6.79	0.90	0.66
NIR	**3.79**	0.97	-0.26

NIR most precise (intra-day SD 0.2 and interday SD 3.2)

Our own experience and observation !

To use HSI to make the impossible alive :
new perspectives !!!

Eylenbosch et al. (2018). ***Quantification of leghaemoglobin content in pea nodules based on near infrared hyperspectral imaging spectroscopy and chemometrics****. J. Spectral Imaging, 7(a9), pp.1–10.*



Sampling

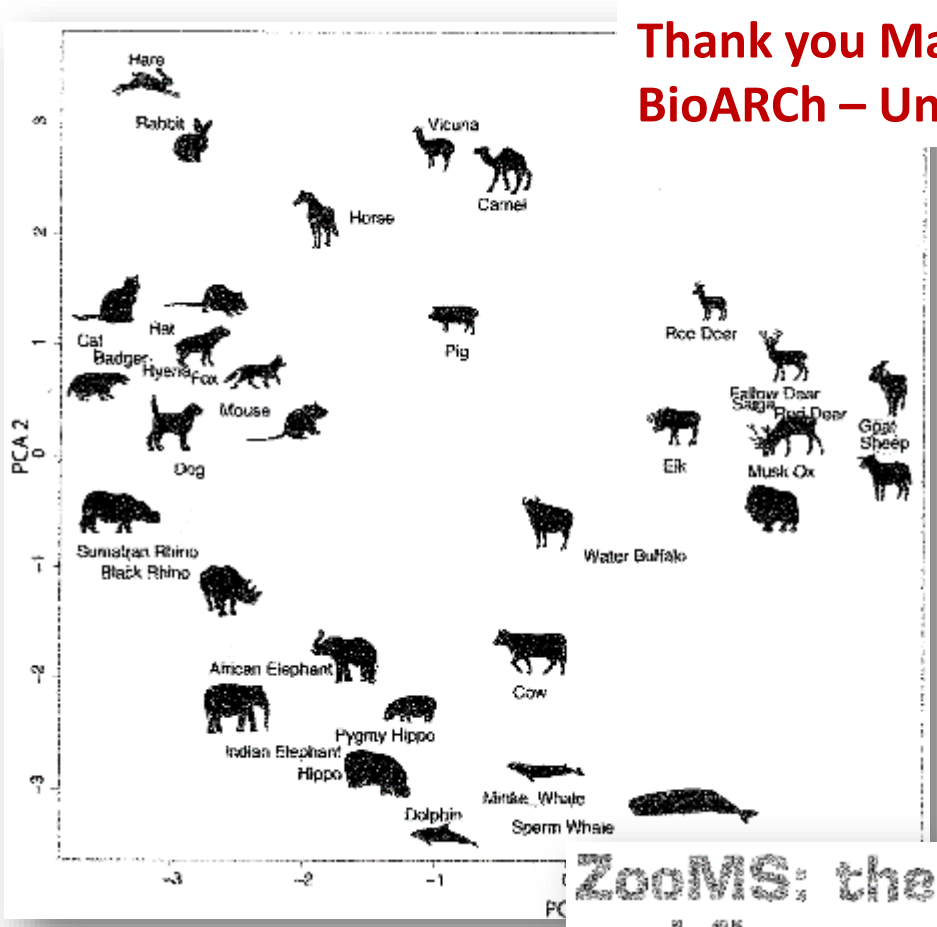
- Non-destructive spectral sensors are (for the time being) **the best** to address sampling issues.
- We must change the way we approach the quality and safety of the food chain - **paradigm shift**.
- Can be **NIR better than** reference method in protein content determination?

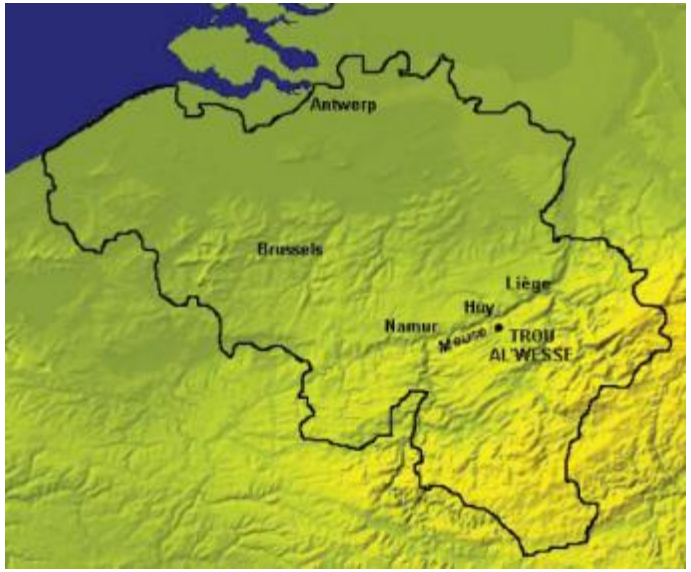
Cross fecundation … pivotal in science!

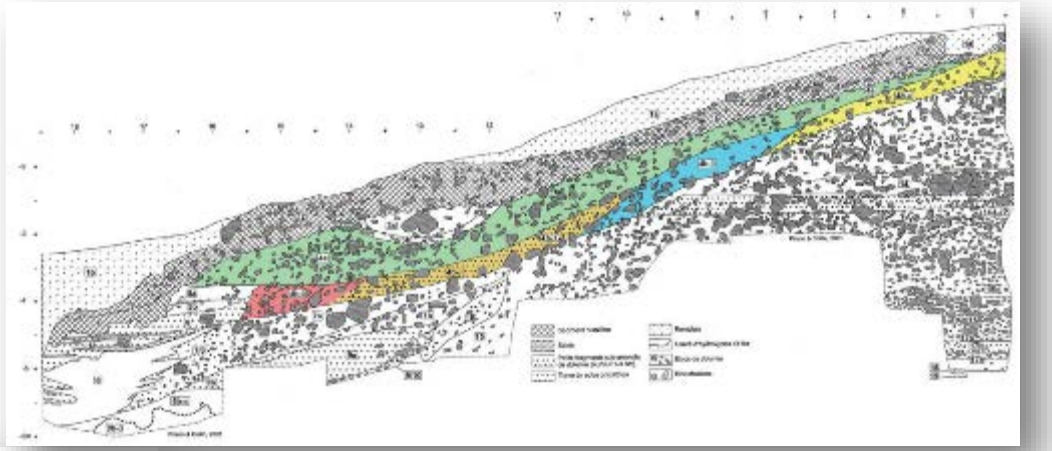
2010

Safeed-pap book : conclusions of the project

Thank you Matthew Collins
BioARCh – Uni. Of York







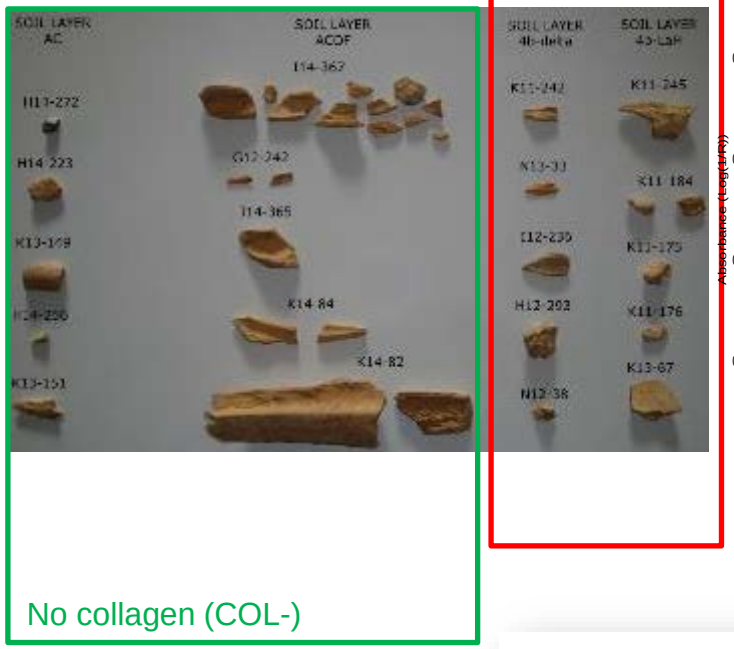
ZooMS: the collagen barcode and fingerprints

Matthew Collins,[a] Mike Buckley,[a] Helen R. Grundy,[b] Jane Thomas-Oates,[a] Julie Wilson[a] and Nienke van Doorn[a]
[a]BioARCh, Departments of Biology, Archaeology and Chemistry, the University of York, York YO10 5DD, UK
[b]The DEFRA Food and Environmental Research Agency (FERA), Sand Hutton, York YO41 1LZ, UK

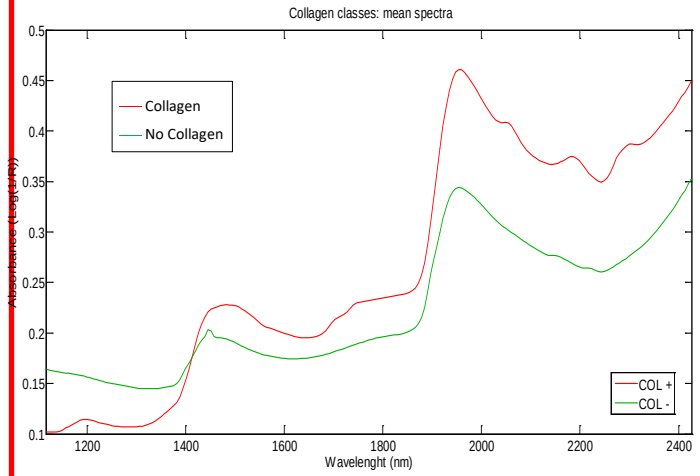
Cross fecundation … pivotal in science!

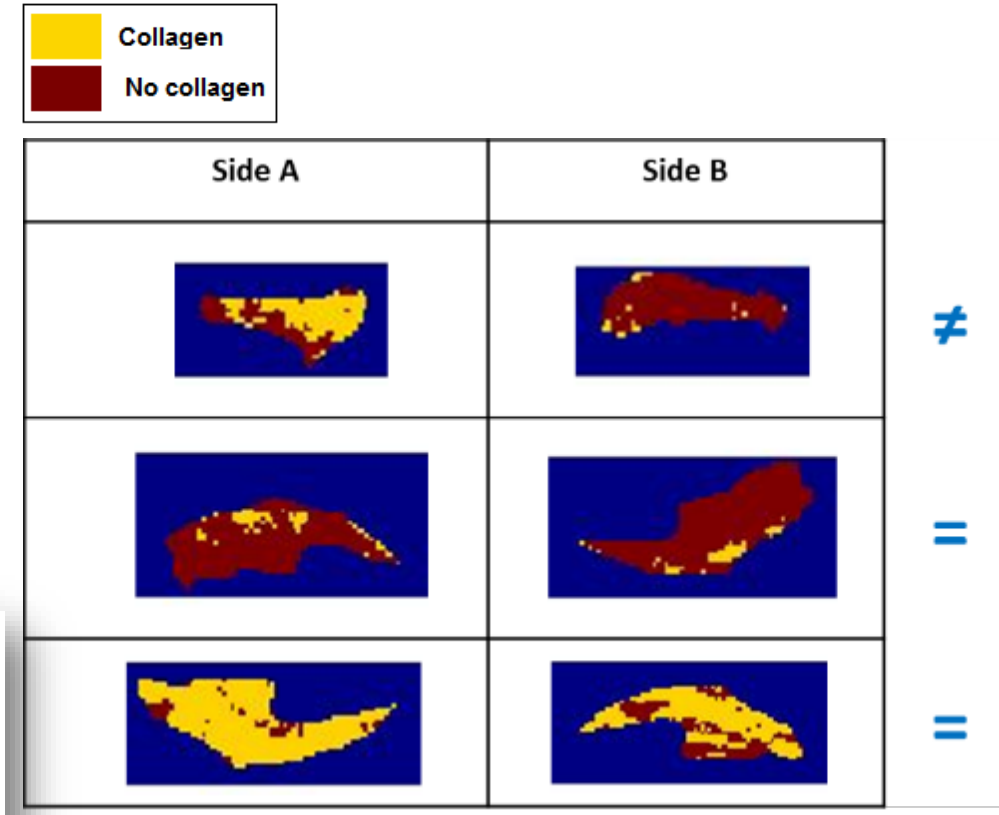
NIR is the best !

Collagen (COL+)



No collagen (COL-)





Talanta 125 (2014) 381–388

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Analysis of collagen preservation in bones recovered in archaeological contexts using NIR Hyperspectral Imaging

Damien Vincke [a], Rebecca Miller [b], Édith Stassart [b], Marcel Otte [b], Pierre Dardenne [a], Matthew Collins [c], Keith Wilkinson [d], John Stewart [e], Vincent Baeten [a], Juan Antonio Fernández Pierna [a,*]

Hélio SPIR 20 ans

Wallonie recherche CRA-W

Detection of adulteration at industrial level

STEP 1
Loading of 5 tons of soybean meal

STEP 2
Loading of 5 tons of soybean meal
+
"Massive" adulteration

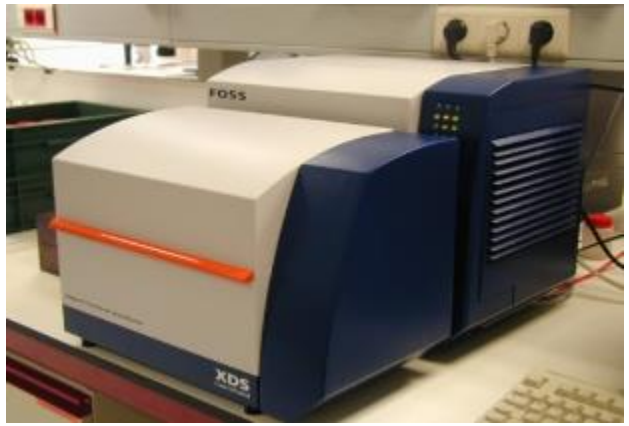
STEP 3
Loading of 5 tons of soybean meal

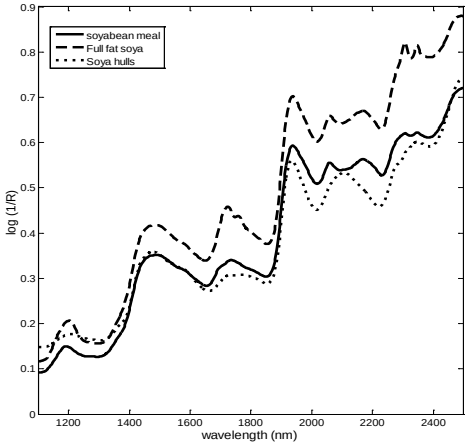
STEP 4
Loading of 5 tons of soybean meal
+
"Smooth" adulteration

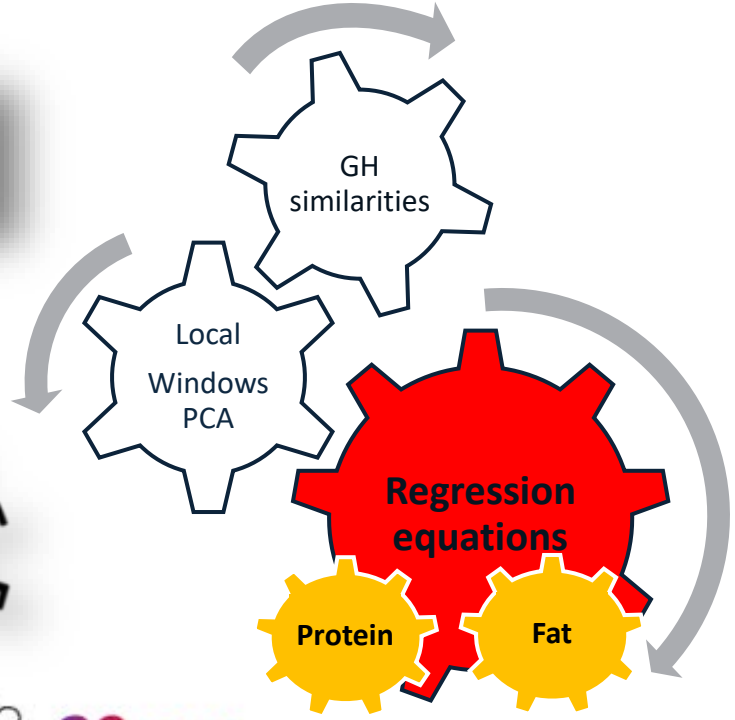
STEP 5
Loading the remaining part of the soybean meal

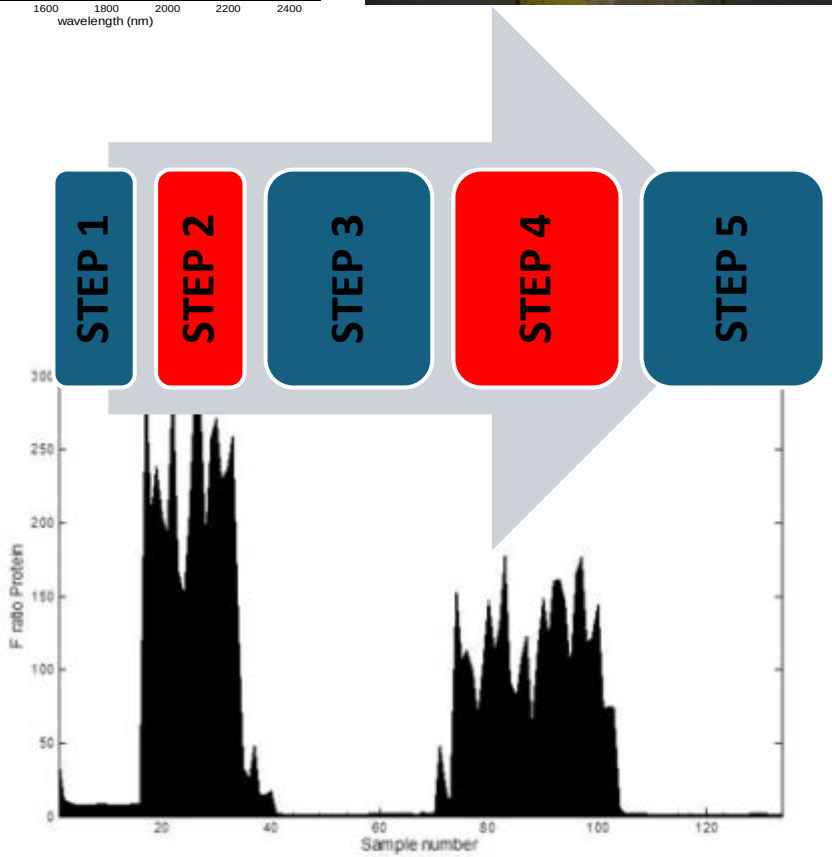
Fernández et al. (2015). NIR fingerprint screening for early control of non-conformity at feed mills. Food Chemistry, 189, art. no. 16461, pp. 2-12.

How to address better sampling issue? → generalize the use of NIR

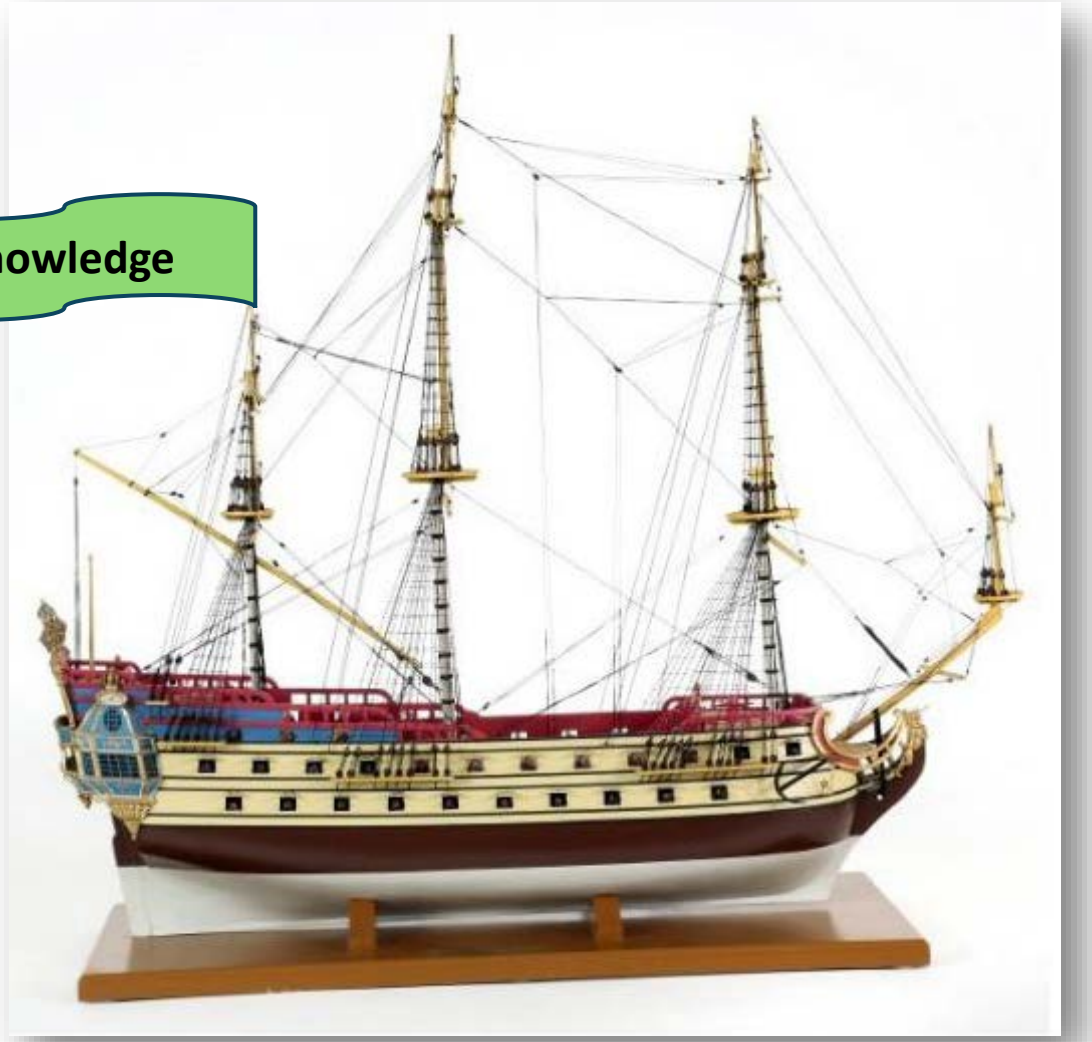








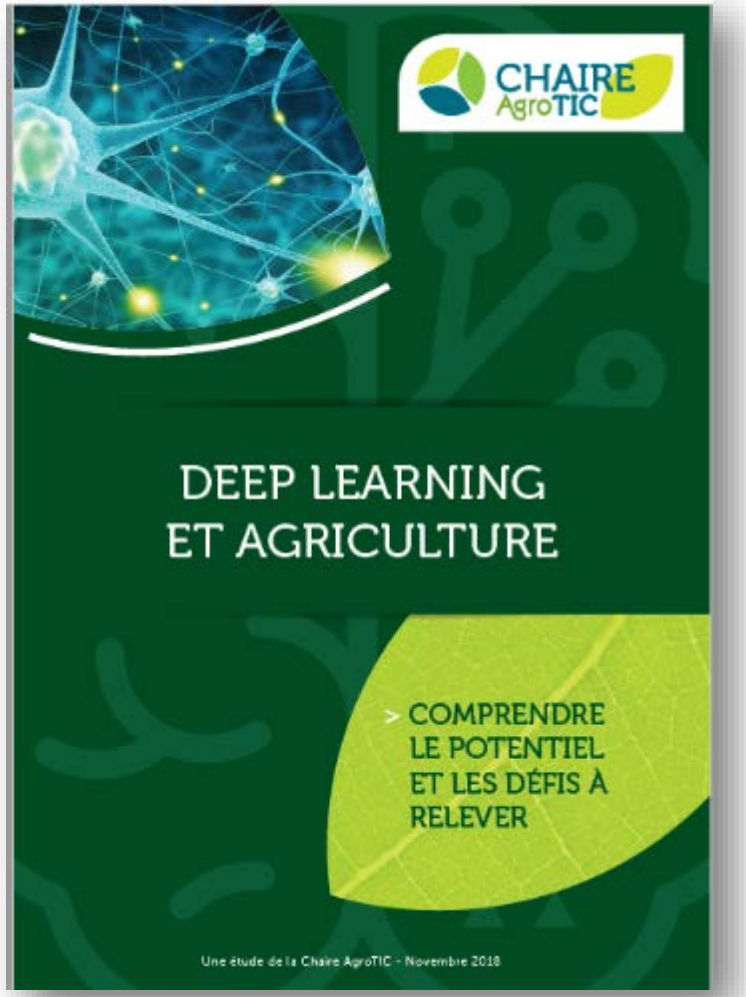
Protein determination

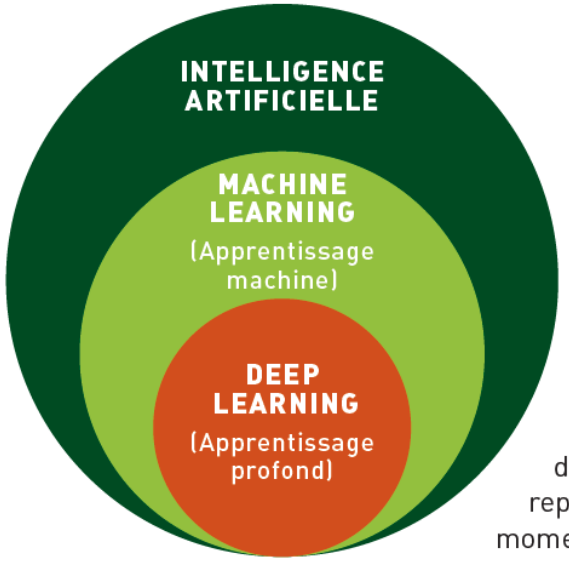


Knowledge

- **Artificial intelligence** could lead us to approach our handling of spectra in a different way
- Are NIR spectra just **2D graphs**?
- Can **GPT-4** do my job?

Artificial Intelligence – and so what?





L'IA, UN VASTE DOMAINE SCIENTIFIQUE

Notion apparue dès les années 50, l'intelligence artificielle peut-être définie comme l' « ensemble de techniques permettant à des machines d'accomplir des tâches et de résoudre des problèmes normalement réservés aux humains et à certains animaux » (1). Il s'agit donc de techniques qui visent à reproduire, imiter, simuler l'intelligence, ou en tout cas les capacités que l'on peut associer à ce terme : capacité à percevoir son environnement, à interagir avec lui, à communiquer au moyen d'un langage, à mettre en œuvre un raisonnement et à résoudre des problèmes, à planifier des actions afin d'atteindre un but, à représenter la connaissance, à appliquer cette connaissance au moment opportun...

"This is not a pipe" (René Magritte, 1929)

"Ceci n'est pas un spectre" (Vincent, 2024)

« The complexicity of NIR data »

Can GPT-4 do my job?

Crédit: Wikipédia

GitHub Copilot
Crédit: Wikipédia

codium ai
Crédit: codium.ai

Gemini
Crédit: Wikipédia

?

Essor of Generative AI → Are they good at my job? →

Crédit: Deep Dream Generator

Meta AI
Crédit: L'usine digitale

Crédit: Wikipédia

Crédit: Wikipédia

Crédit: Wikipédia

Crédit: Zapier

→ Challenge 2024 of
Pierre Dardenne & CRA-W →

For the 2024 challenge, only 1 parameter has to be predicted. The calibration data set is a public data set of 19 036 spectra of soil samples collected in whole Europe. The data set is usually called the LUCAS dataset. https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data#tabs-0-description=1

The dried and ground samples were scanned on a near infrared spectrometer from 400 to 2498 nm.

The Y parameter to be calibrated is the content of Total Organic Carbon (TOC) expressed in % of dried soil.

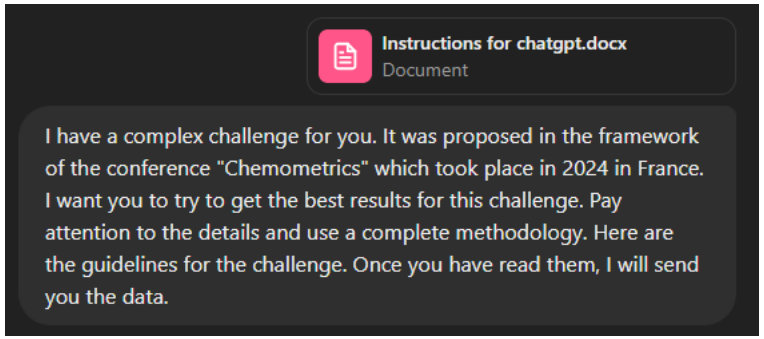
The 2 test sets are soil samples scanned on two other instruments. These 2 instruments are NOT represented in the CAL set.

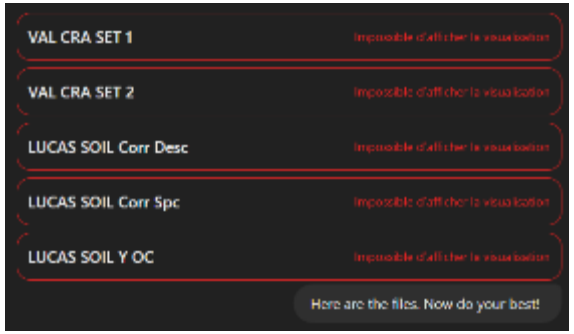
The goal is to predict accurately the Y values (only COT) of the 2 validation sets.

The criterion to classify the results will be the RMSEP of the 2 merged sets.

Can GPT-4 do my job?

- Challenge guidelines
- Calibration set (19 036 samples)
- Validation sets (226 and 229 samples)





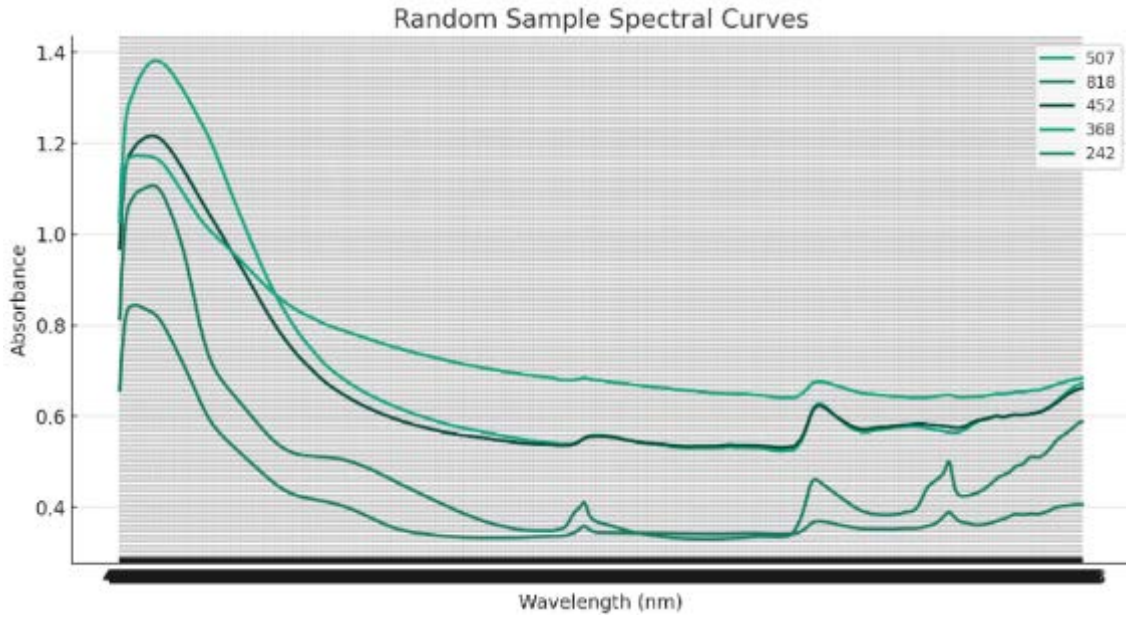
- Methodology proposed:

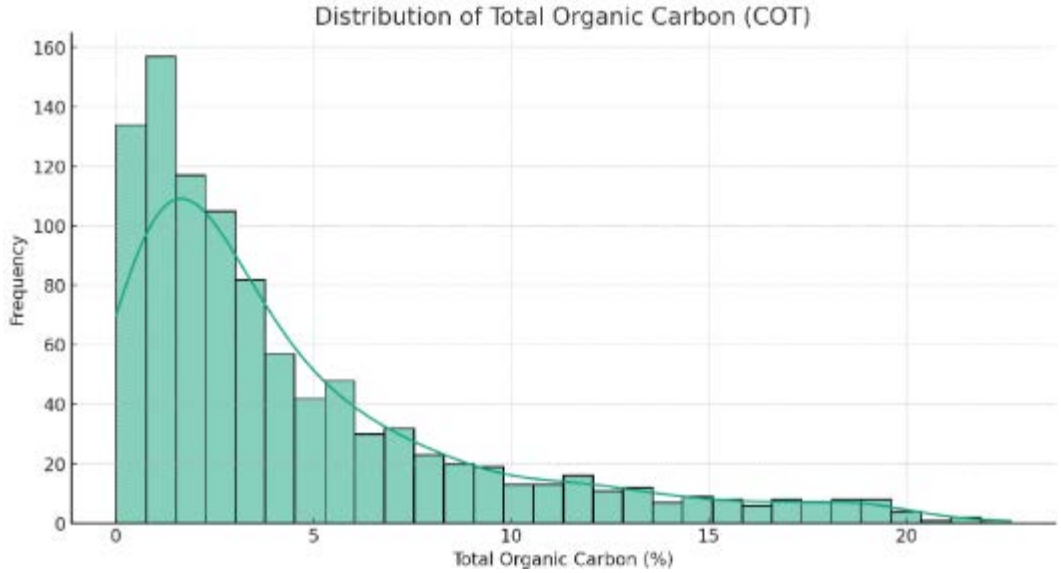
1. Data exploration and Preprocessing (check structure, NA, outliers TOC + spectra, normalize or standardize spectra)
2. Spectral Preprocessing (Savitzky-Golay, MSC, SNV)
3. Feature Selection or Extraction (PCA or PLS to reduce dimensionality)
4. Model Development (PLS, SVR, RF regression)
5. Model Validation and Optimization (on new dataset with RMSE/R^2)
6. Prediction and Evaluation (Apply final model on validation set)
7. Documentation and Justification

- Data too heavy: I'm currently unable to directly manipulate the data due to technical issues

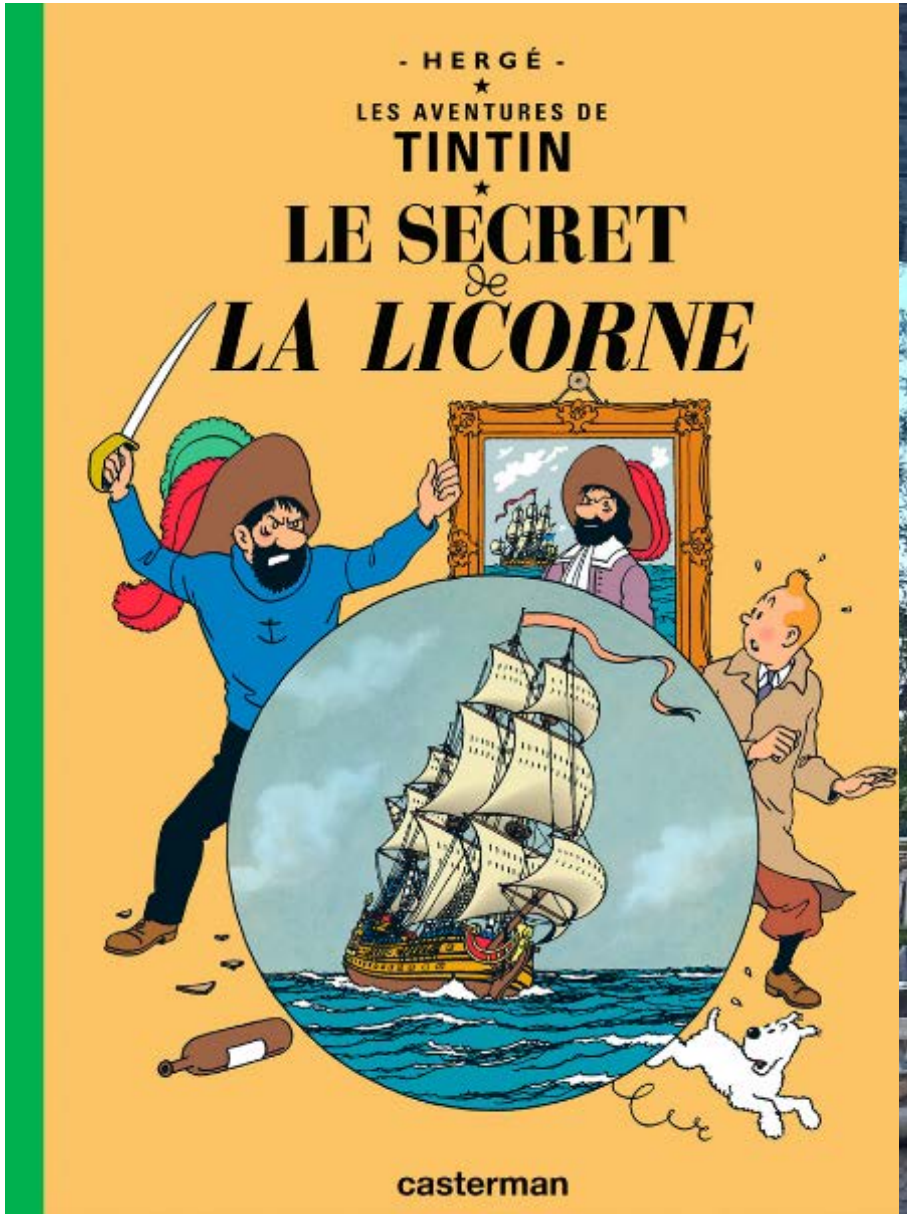
Can GPT-4 do my job?

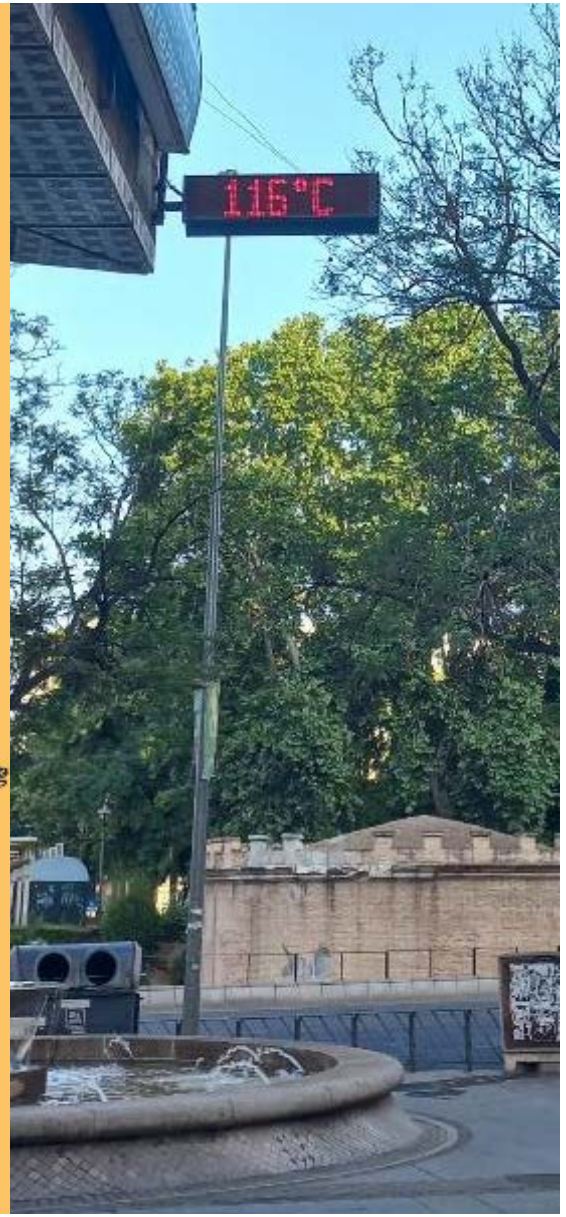
- All the steps at once = too much → New prompt = **Write the methodology then proceed one step at a time (ask me if you can move on to the next step).**
- Steps carried out. Plots generated. Model = PLS.



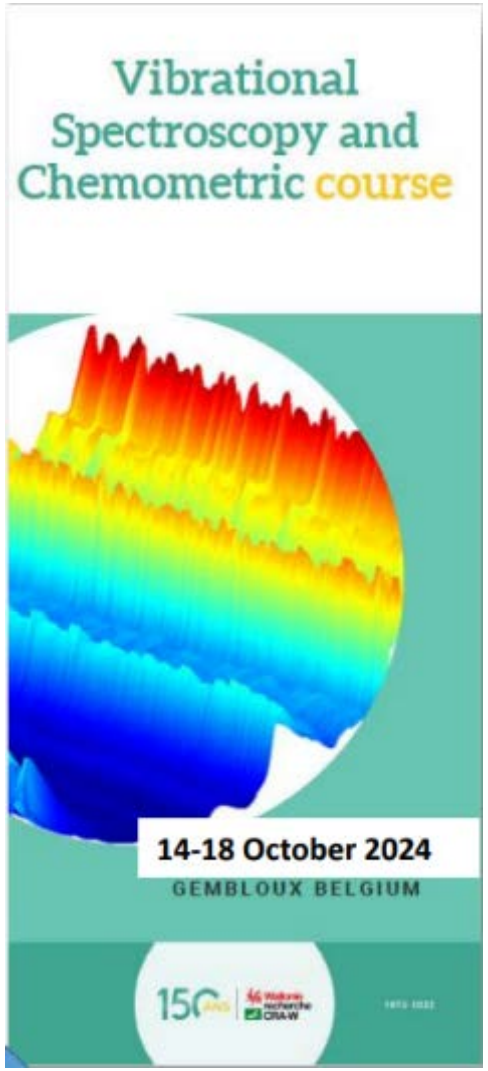


- RPD = 1,13 RMSECV = 4.12% < RMSEP participants (2,26% to 1,63%)





- ❑ What next ? In 2044?
- ❑ Heliospir in Belgium?



Hélio SPIR 20 ans

Merci !

Wallonie recherche CRA-W