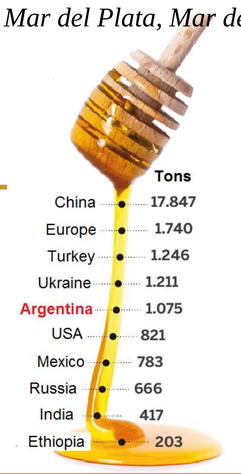
Is vibrational spectroscopy an adequate tool for assessing the geographical origin of honey?

Tito Damiani[1], Chiara Dall'Asta[1], Inés Aubone[3],Vincent Baeten[2], Sandra Fuselli[3,4], Rosa M. Alonso Salces[3,5], Quentin Arnould[2], Lisa Plasman[2], Juan Antonio Fernández Pierna[2]

1-Department of Food and Drug, University of Parma, Parma, Italy
2-Food and Feed Quality Unit, Valorisation of Agricultural Products Department, Walloon Agricultural Research Centre (CRA-W), Gembloux, Belgium
3 – Grupo Investigación Microbiología Aplicada (GIMA), Centro de Investigación en Abejas Sociales, Facultad de Ciencias Exactas y Naturales, Univ. Nac. de Mar del Plata, Mar del Plata, Argentina
4 - Comisión de Investigaciones Científicas (CIC), La Plata, Argentina
5- Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina

j.fernandez@cra.wallonie.be; foodfeedquality@cra.wallonie.be

Honey is the natural substance produced by *Apis mellifera* bees from the plant nectar or excretions of plant-sucking insects. In the supply and distribution chain, **mono-geographic** honeys usually command a premium price. Argentina is on the main honey producers in the world and methods to authenticate specific regions are needed.



Main honey exporting countries (2009)

Purpose: Investigation of vibrational spectroscopy and chemometrics as fast and reliable tool for honey geographical discrimination.

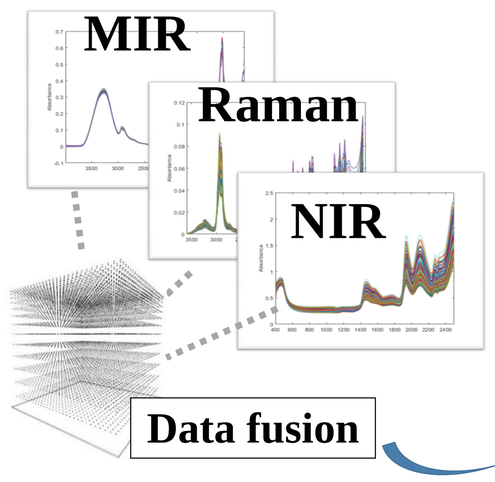
Sampling

Multifloral honey ($n = 502$) collected among three honey-producer provinces of Argentina

Sampling repeated over four harvesting seasons (2014, 2015, 2016 and 2017)

Instrumental Analysis

Fingerprinting through three spectroscopic techniques:



Low-level (LL-DF): sample-wise concatenation of data blocks
Mid-level (ML-DF): relevant features extraction from each data source and concatenation into single array
High-level (HL-DF): classification models computation from each data source and combination of the individual responses

Data Treatment

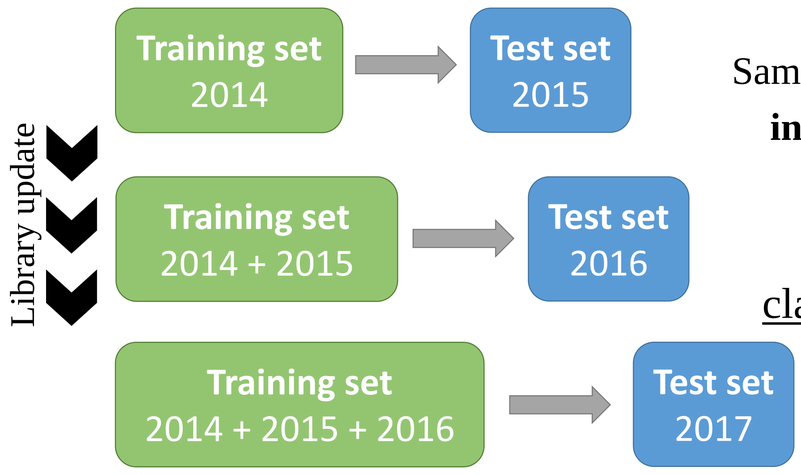
- Performance comparison between spectroscopic techniques (ROC curves)
- Binary models:

Buenos Aires vs Catamarca | Buenos Aires vs Misiones | Catamarca vs Misiones

Prediction ability evaluation and validation

Validation Strategy

The following Validation approach was adopted:

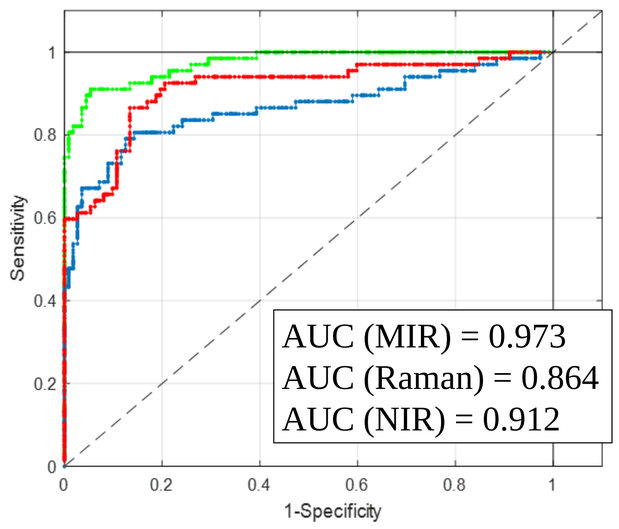


Samples with confirmed provenance **included in the calibration set**

Model re-computation and classification boundaries update

Results

Spectroscopic techniques comparison



The Area Under the Curve (AUC) is a summary performance parameter of the model.

Larger AUC = Better performance

Model built on MIR data provided superior discrimination ability

Model validation (MIR data)

Predicted Harvest	Correct classification rate [%]		
	Buenos Aires vs Catamarca	Buenos Aires vs Misiones	Catamarca vs Misiones
2015	79	86	89
2016	84	100	92
2017	86	100	94

Performance ↓

When more samples were included in the data library, larger biological variability was covered

Enlarging the data library with more samples accounting for more variability can improve the model prediction ability

Data Fusion (DF)

Predicted Harvest	Correct classification rate [%]		
	Buenos Aires vs Catamarca	Buenos Aires vs Misiones	Catamarca vs Misiones
Low-Level DF	81	98	91
Mid-Level DF	78	98	81
High-Level DF	86	98	98

Fusion of MIR data with Raman and NIR did not improve the model performance

To take advantage of data fusion, the merged data must bring **complementary information** about the samples.
This is not the case of the present study, where similar information seems to be provided by the investigated techniques.