MACHINE LEARNING METHODS FOR SUGAR QUANTIFICATION IN GRAPES BASED ON NEAR-INFRARED HYPERSPECTRAL IMAGING

Julien BOYER[1], Jordane POULAIN[1], Dr Sylvie ROUSSEL[1], Nicolas SAURIN[2], Maxime RYCKWAERT[3], Ryad BENDOULA[3], Carole FEILHES[4] and Eric SERRANO[4]

[1]Ondalys, Clapiers, FRANCE - sroussel@ondalys.fr - [2]ITAP, Univ Montpellier, INRAE, Institut Agro, Montpellier, France - [3]UE Pech Rouge, Univ Montpellier, INRAE, Gruissan, France - [4]IFV Sud-Ouest, V'innopôle, Peyrole, France

1. Context

- **Objective :** grape bunch maturity prediction directly in vineyards
- **Database** from VINIoT project
  - 131 hyperspectral images
  - 2 red grape varieties (Fer Servadou N. and Syrah N.)
- **HSI Instrument :** SPECIM IQ (SPECIM - Konica Minolta)

2. Why using Hyperspectral Imaging[1] for quantification?

- Better representativity of the sample
- Knowledge about spatial distribution
- Hand-held instruments for easy use in the field

Constraints

- Lower signal to noise ratio compared to a classical spectrometer
- Reference value on the whole sample (No local reference value on each pixel)
- Measurements directly in the field

➔ high heterogeneity (weather, shadows, non-uniform background)

Solutions

- Average of pixels

➔ higher signal to noise ratio, representativity of the whole grape

- Spectral preprocessing
- Segmentation of the Region Of Interest (ROI) to remove background

3. Methodology & Results

- Segmentation:
  - Choice of some heterogenous images + manual selection of grapes
  - Raw normalization to attenuate the effect of non-uniform lightning
  - Soft Independent Modeling of Class Analogy (SIMCA[2]) ➔ $D = \sqrt{T2^2_{reduced} + Q^2_{reduced}}$
  - Segmentation on Image D with Particle Size Analysis (MIA_Toolbox® – Eigenvector Research Inc.)
- Average of the Region Of Interest (ROI) for each image
- Regression model: comparison of PLSR[3] and SVMR[4]
  - ➔ SVMR is significantly better at a risk of 5% (bootstrap testing)
- Application on the whole image to visualize spatial distribution

Segmentation

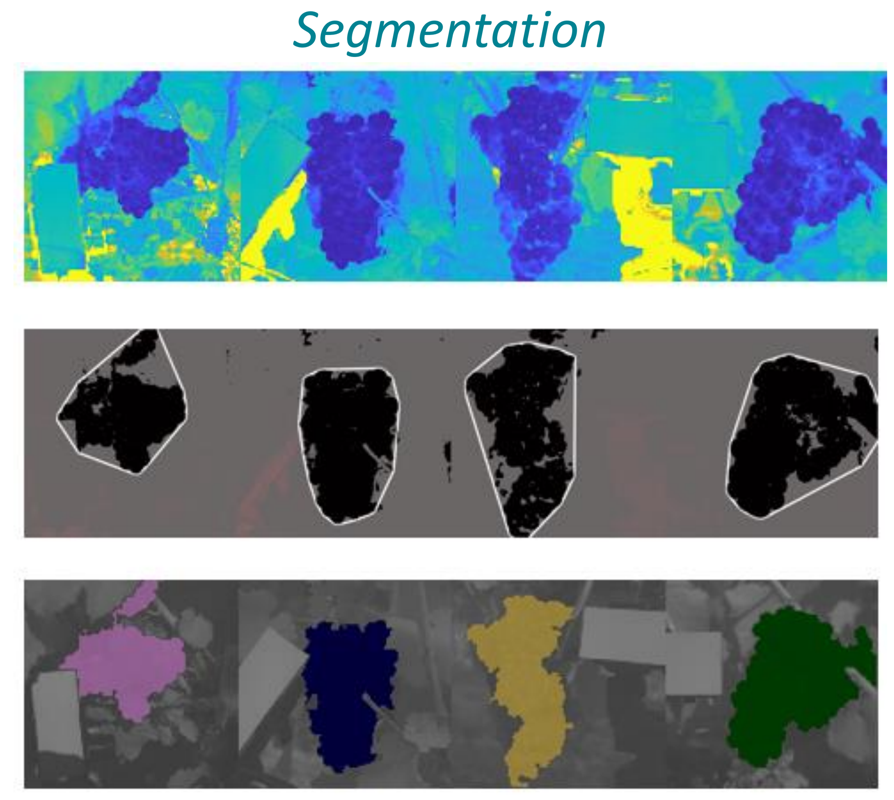


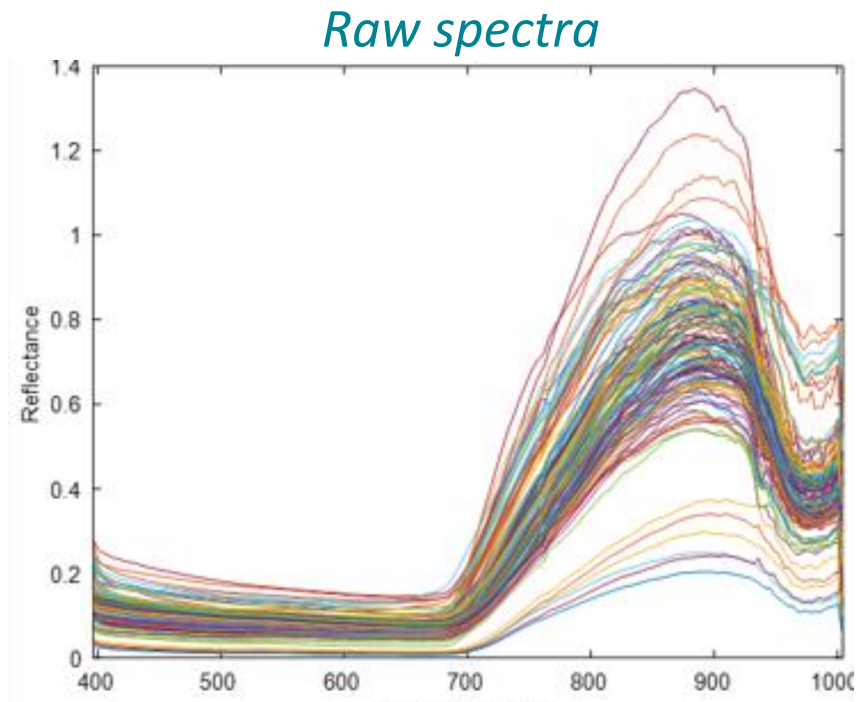
Image D

Thresholding

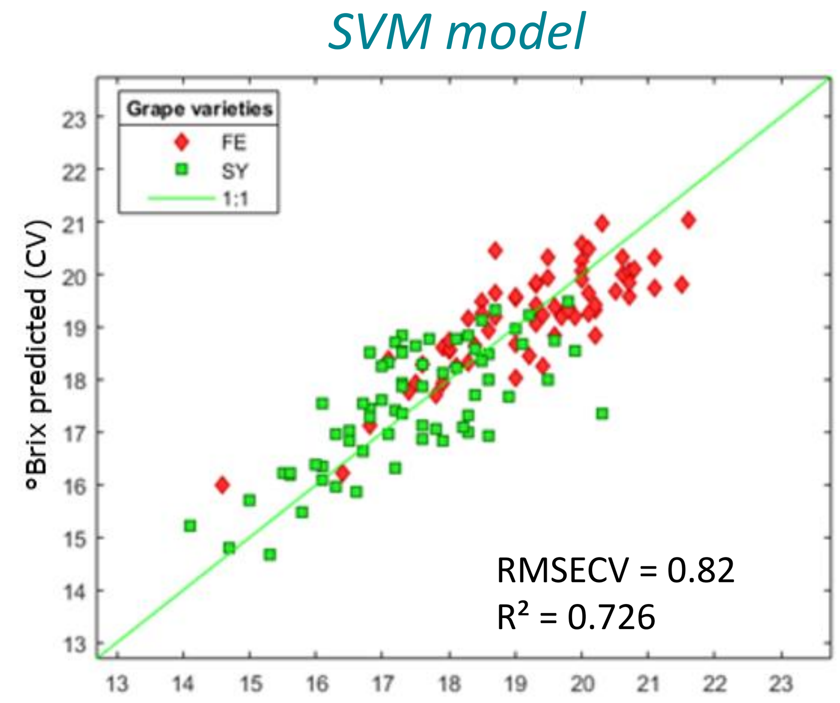
Final segmentation

MIA Toolbox® (Eigenvector Research Inc.)

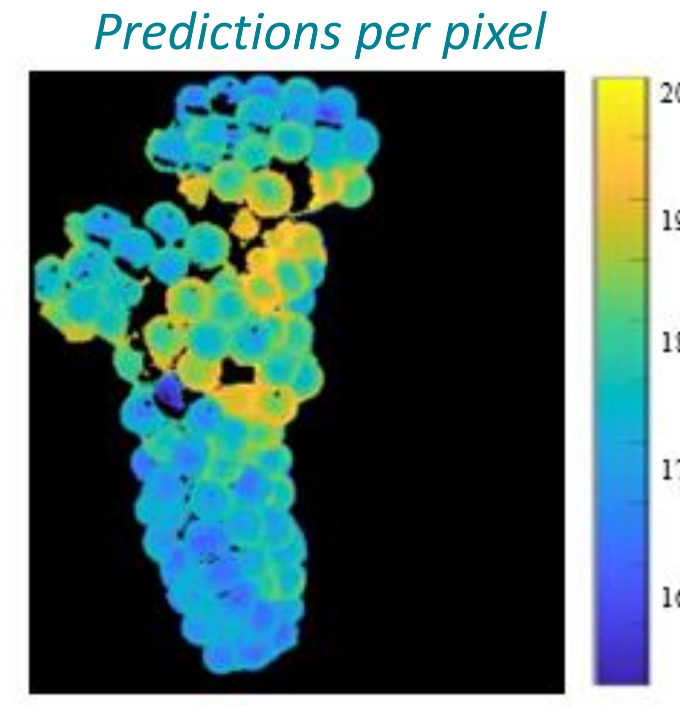
Raw spectra



SVM model



Predictions per pixel



Conclusions

- **Good prediction model performances** despite direct measurements in vineyards
- **Spatial distribution of sugar content** in grape bunches

ACKNOWLEDGMENTS

This HSI data analysis study was carried out by Ondalys in collaboration with INRAE (UMR ITAP, Team COMIC and UE Pech Rouge) and IFV (IFV South West) as part of the VINIoT project, an Interreg SUDOE project funded by the European Regional Council.

REFERENCES

[1] Amigo, José Manuel, éd. Hyperspectral imaging. Data handling in science and technology 32. Amsterdam, Netherlands; Cambridge, MA, United States: Elsevier, 2020.

[2] S. Wold, M. Sjostrom, SIMCA: A method for analyzing chemical data in terms of similarity and analogy. In Kowalski, B.R., ed., Chemometrics Theory and Application, American Chemical Society Symposium Series 52, 1977, p. 243-282.

[3] C. Cortes, V. Vapnik, Support-vector networks. Mach. Learn. 20, 273–297 (1995).

[4] H. Wold. Estimation of principal components and related models by iterative least squares. In P.R. Krishnaiaah (Ed.). Multivariate Analysis. 1966. pp.391-420.

www.ondalys.fr
contact@ondalys.fr