HélioSPIR 2025

MACHINE LEARNING INTERPRETABILITY METHODS APPLIED TO CALIBRATION MODELS DEVELOPED ON NEAR INFRARED SPECTROSCOPIC DATA

Dr. Astrid MALECHAUX, Jordane POULAIN, Dr. Sylvie ROUSSEL

amalechaux@ondalys.fr

Ondalys – CLAPIERS, FRANCE

Introduction

Growing interest for Machine Learning models

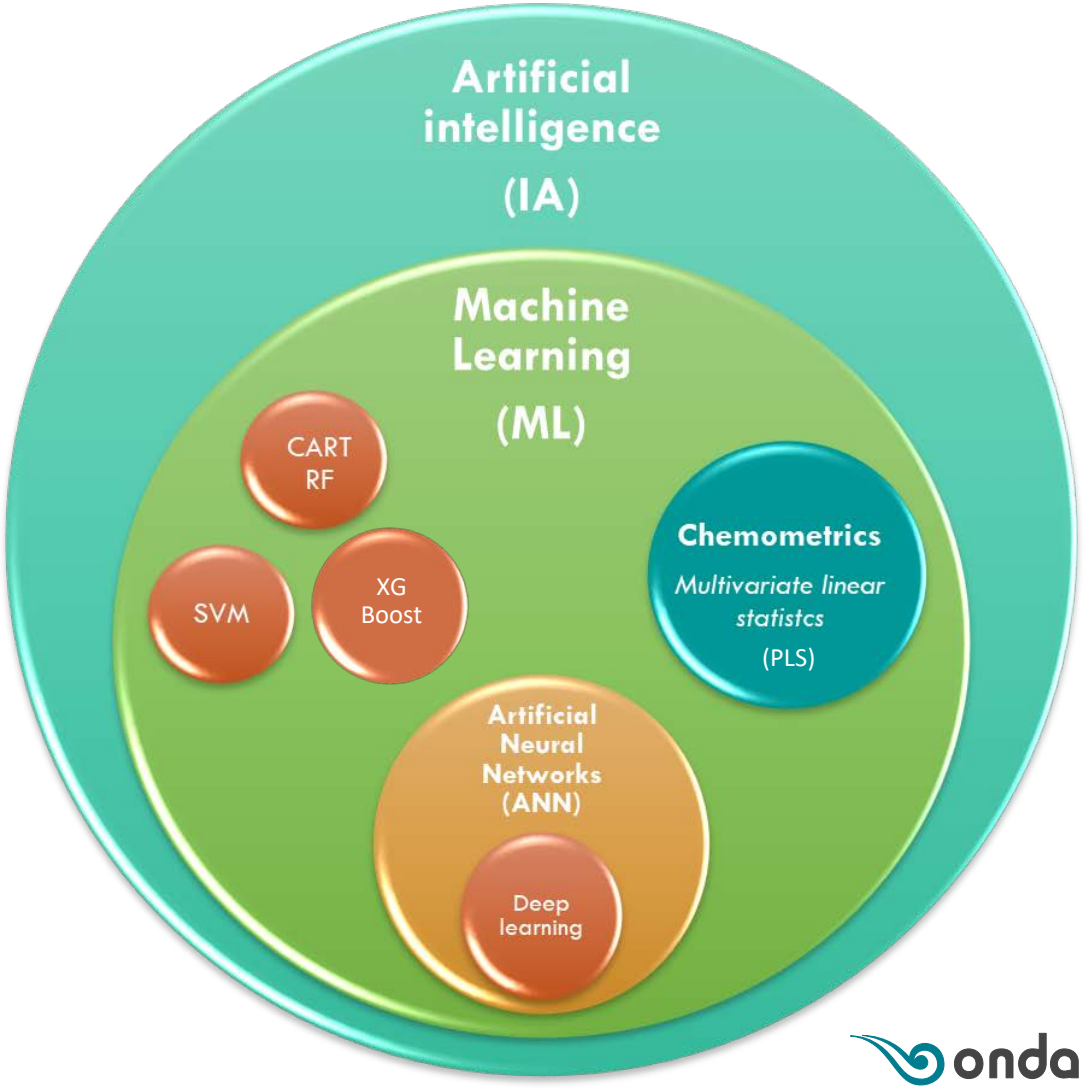
> ☺ Gain of performance: able to model more complex relationships (non-linearity, variability, …)

> ☹ Loss of interpretability: « black-box » models

> ☹ Increased risk of over-fitting

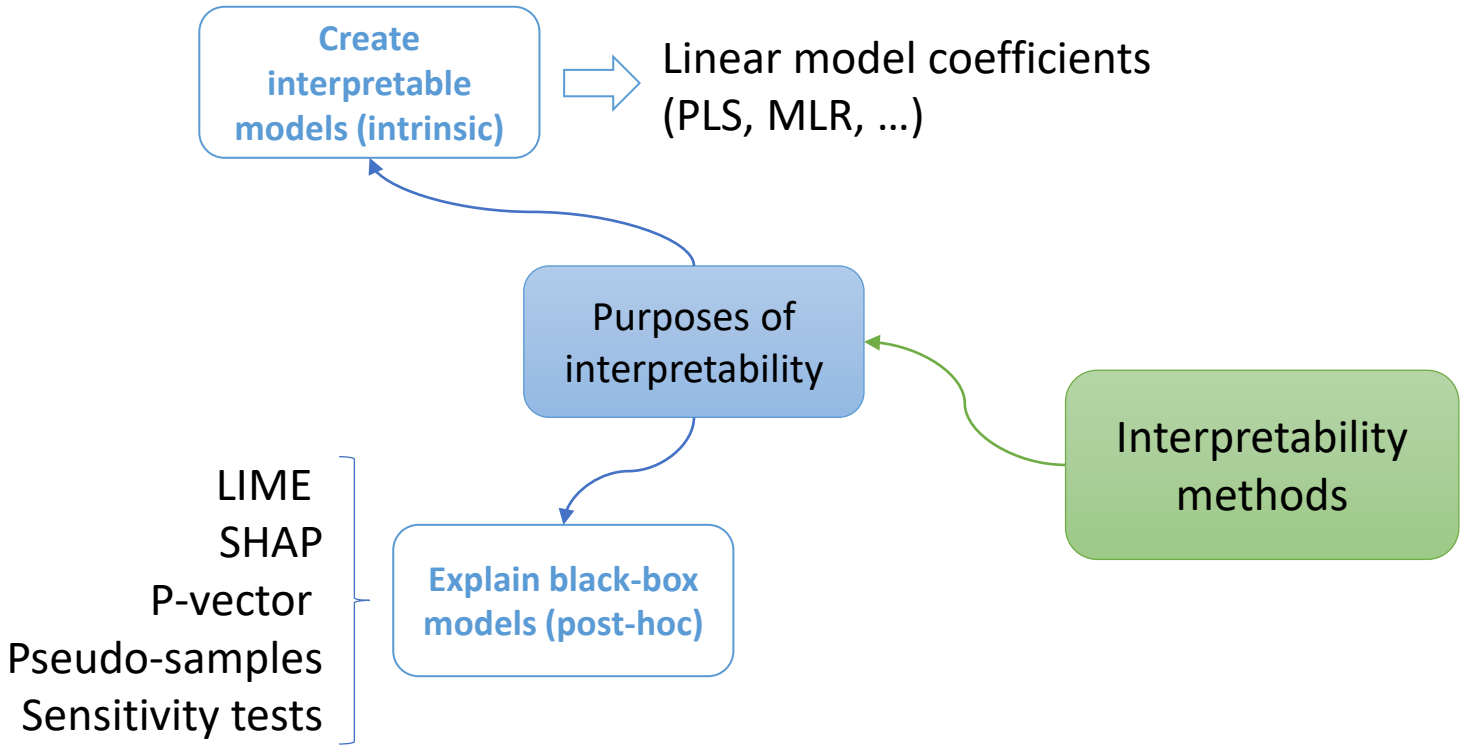
➔ Need for interpretability tools

1. Explain models:
   - Understand which variables are important to obtain the predictions
   - Check if the model « makes sense » chemically
2. Diagnose overfitting:
   - Help to avoid overfitting during the optimization of model hyperparameters



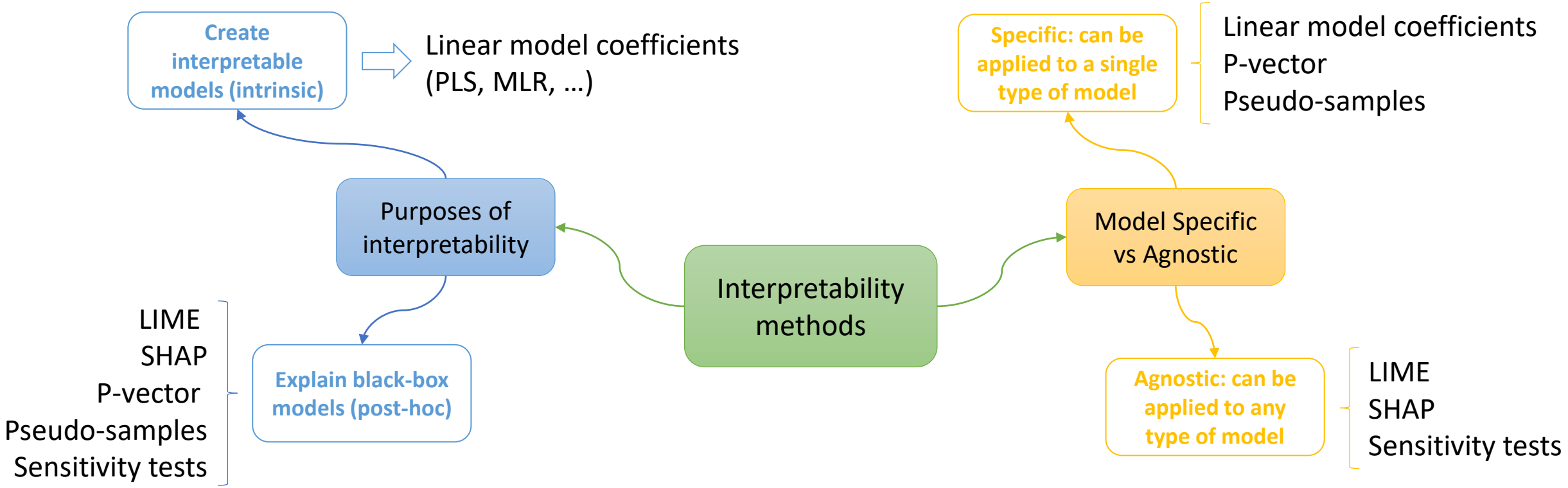
ondalys

Introduction



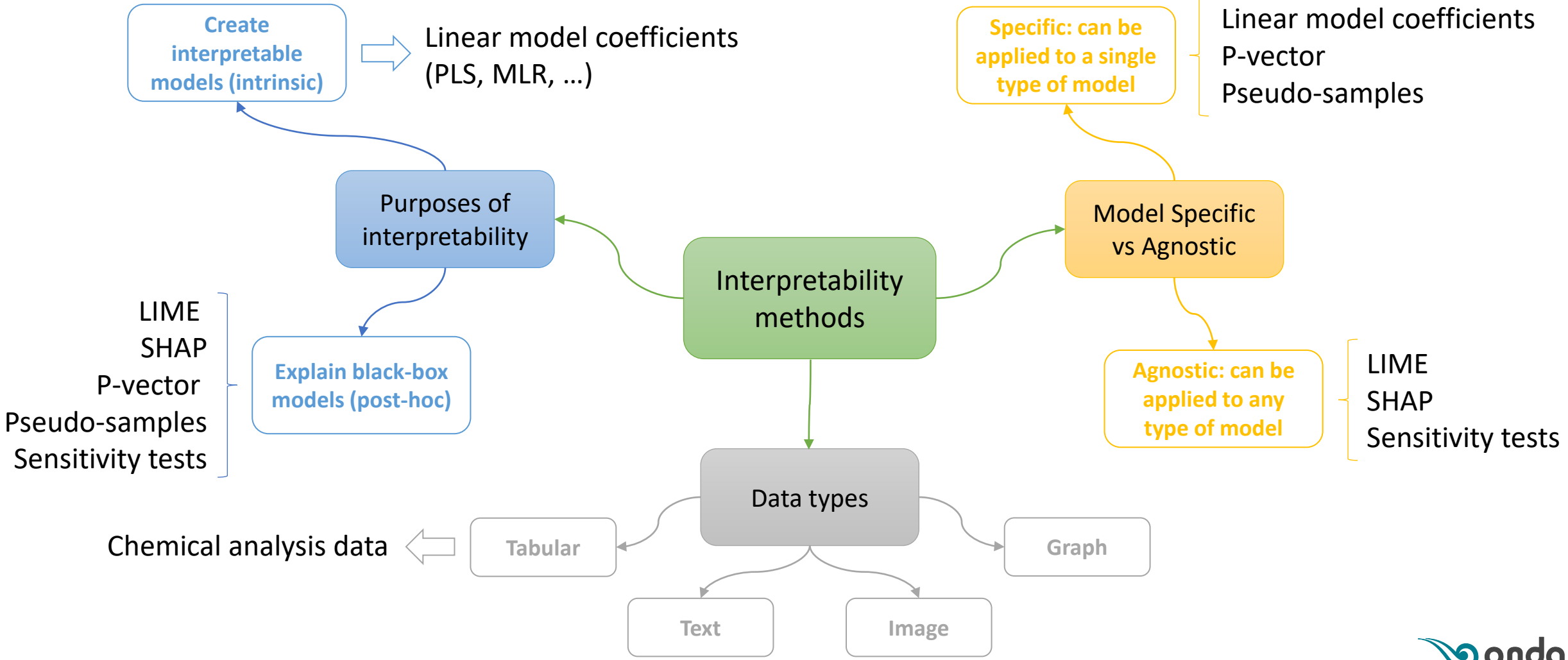
Adapted from: Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy 2021, 23, 18

Introduction



Adapted from: Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy 2021, 23, 18

Introduction



Adapted from: Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy 2021, 23, 18

Interpretability methods

- LIME[1]: Local interpretable model-agnostic explanations
  - > Explains the prediction of individual samples by fitting a surrogate interpretable model (ex: LASSO regression)
  - > Generates « perturbed samples » and computes their prediction by the black-box model, then trains the local interpretable model on the « perturbed samples » weighted by their proximity to the explained sample
  - > Available in Matlab, Python, R, …

- SHAP[2]: Shapley additive explanation
  - > Explains the prediction of individual samples by combining Shapley values from game theory (average contribution of each « player », or variable, to the total « gain », or difference from the average prediction) with local model explanations
  - > Available in PLS_Toolbox®, Matlab, Python, R, Julia, …

[1] M. T. Ribeiro, S. Singh, C. Guestrin, "Why should I trust you?" Explaining the predictions of any classifier, Proceedings of the 22nd ACM SIGKDD, 2016, 1135-1144.
[2] S. M. Lundberg, S. Lee, A Unified Approach to Interpreting Model Predictions, Advances in Neural Information Processing Systems, 2017, 30, 4765-4774.

Interpretability methods

- Pseudo-samples[1]
  - > Approximates kernel-based model coefficients by predicting a matrix of dummy samples, for which all variables except for one have their value set to 0 (the non-null variable takes a value in the range of spectral intensity)
  - > Computable in Matlab, Python, R, Julia, …
- Sensitivity tests
  - > Compares the predictions obtained with different perturbations of the original data one variable at a time, such as the difference of prediction obtained when the intensity of each variable is:
    - increased or decreased by 1%[2]
    - increased by 1% of its standard-deviation[3]
    - replaced by 0
  - > Available in PLS_Toolbox® and computable in Matlab, Python, R, Julia, …

[1] G. J. Postma, P. W. T. Krooshof, L. M. C. Buydens, Opening the kernel of kernel partial least squares and support vector machines, Analytica Chimica Acta, 2011, 705(1-2), 123-134.
[2] D. B. Funk, Instrumentation considerations for robust near infrared applications, Proceedings of the 9th ICNIRS, 2000, 171-176.
[3] https://www.wiki.eigenvector.com/index.php?title=Tools_ModelRobustness

Application – dataset presentation

Tecator dataset*

***Source :** http://lib.stat.cmu.edu/datasets/tecator

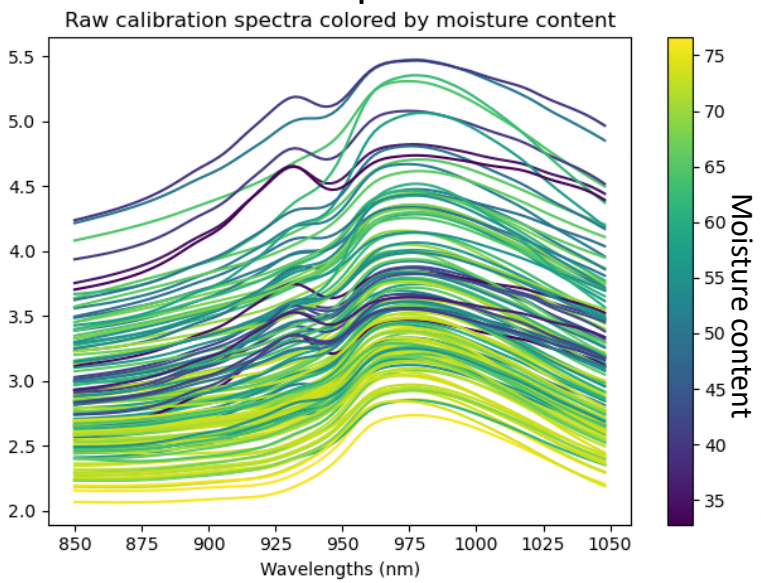
> Data description

- Near Infrared spectroscopic data on raw meat (FOSS Tecator Infratec Food and Feed Analyzer)
- 3 quantitative responses: moisture content, fat content, protein content

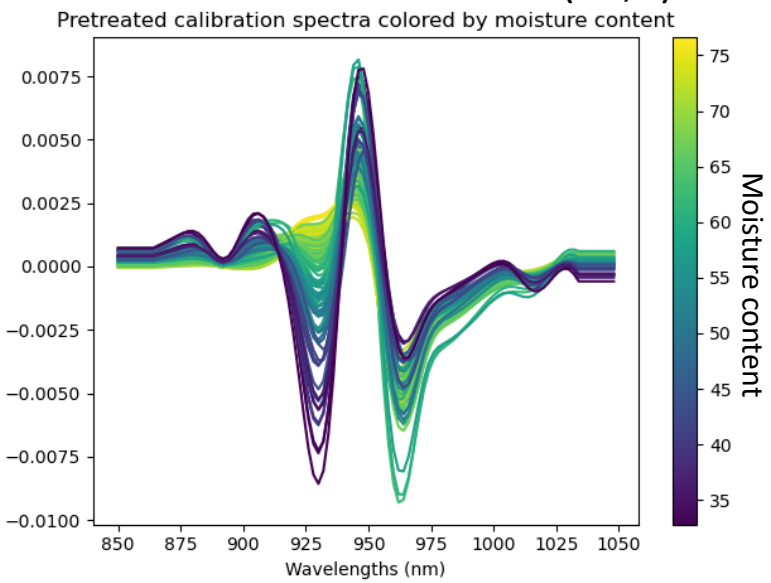
> Effect of preprocessing – example for moisture

- 2nd derivative corrects for baseline variations and enhances peaks
- SNV (Standard Normal Variate) corrects for multiplicative effects and enhances the gradient of spectral intensity as a function of moisture but distributes the information over the different wavelengths

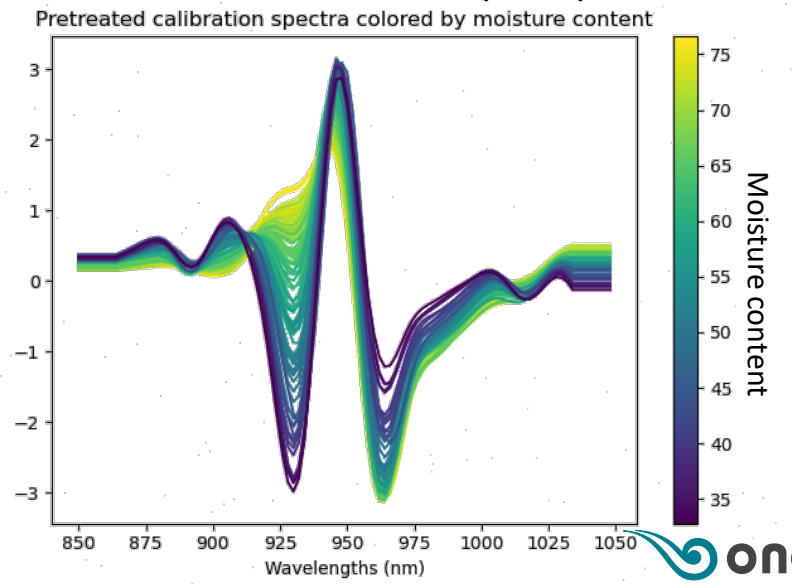
Raw spectra



After SG 2nd derivative (15,2)



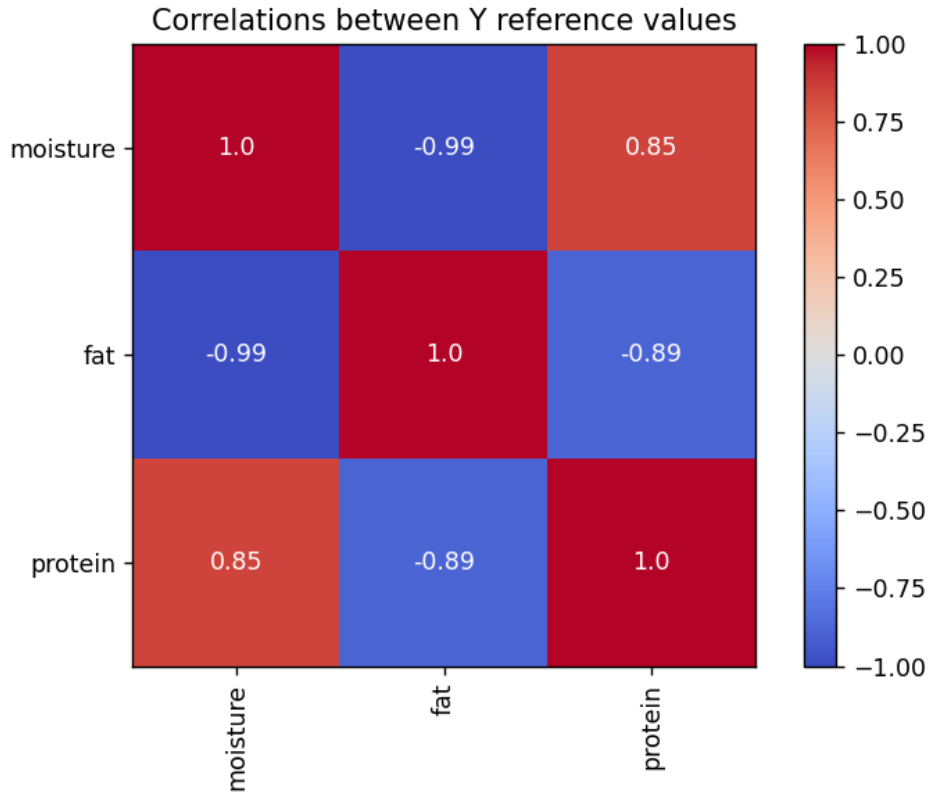
After SG 2nd derivative (15,2) + SNV

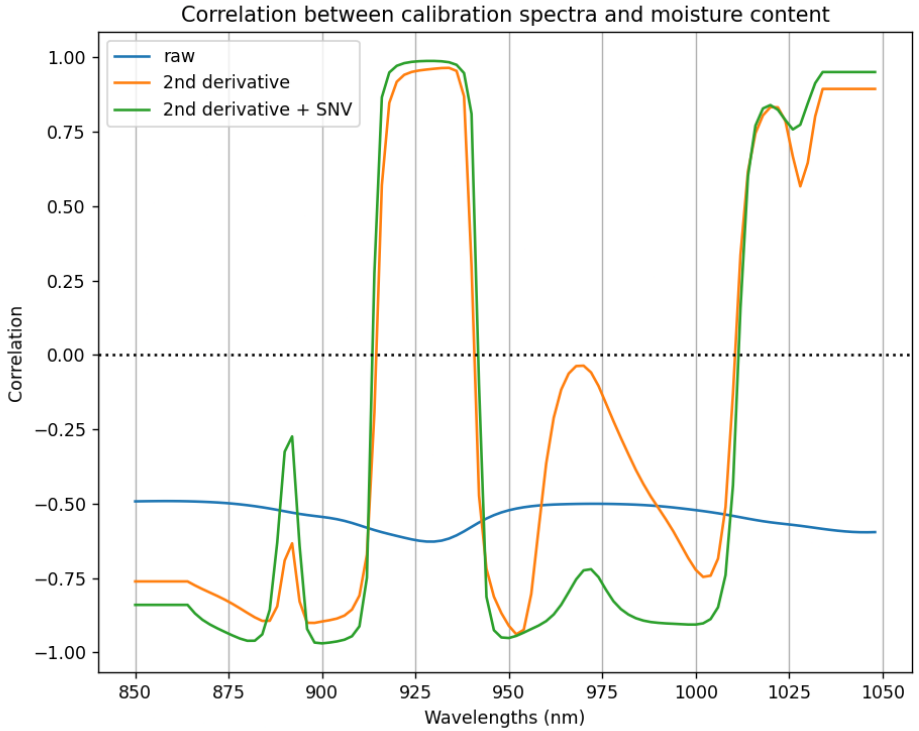


Application – dataset presentation

Tecator dataset: responses visualization

- Strong correlations between the 3 responses, especially between fat and moisture
- Preprocessing increases the correlations between spectra and moisture content, but not only for the water band (SNV, correlation with other responses)

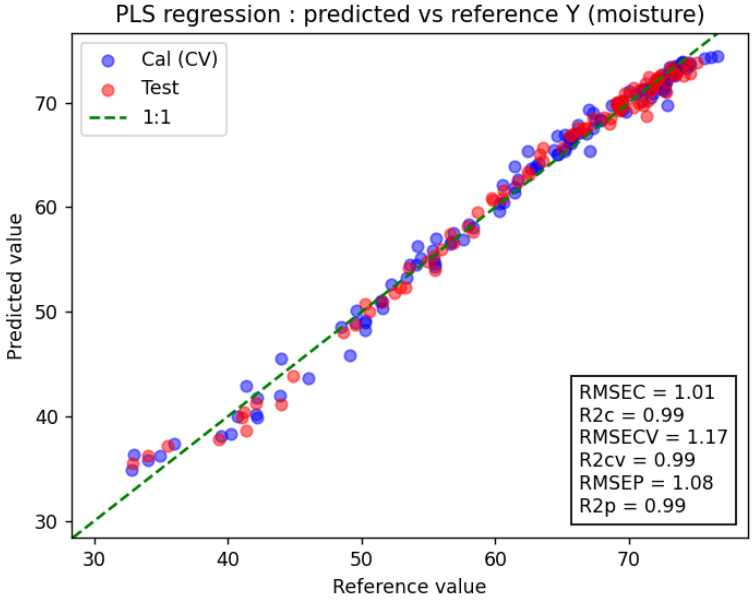




Application – regression models

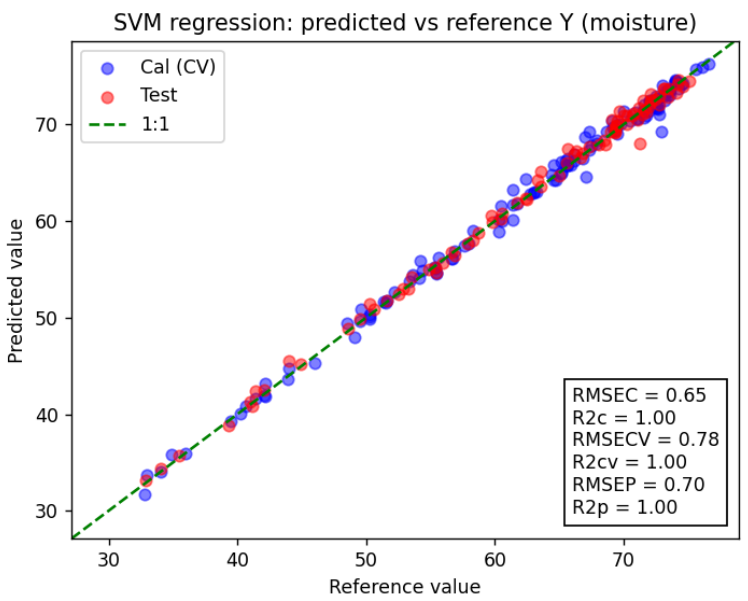
Tecator dataset: prediction of moisture content

- 130 samples in calibration set / 86 samples in test set
- Model optimization by cross-validation (KFold, 5 groups)
- Spectral preprocessing: Savitzky-Golay 2nd derivative + SNV+ mean center



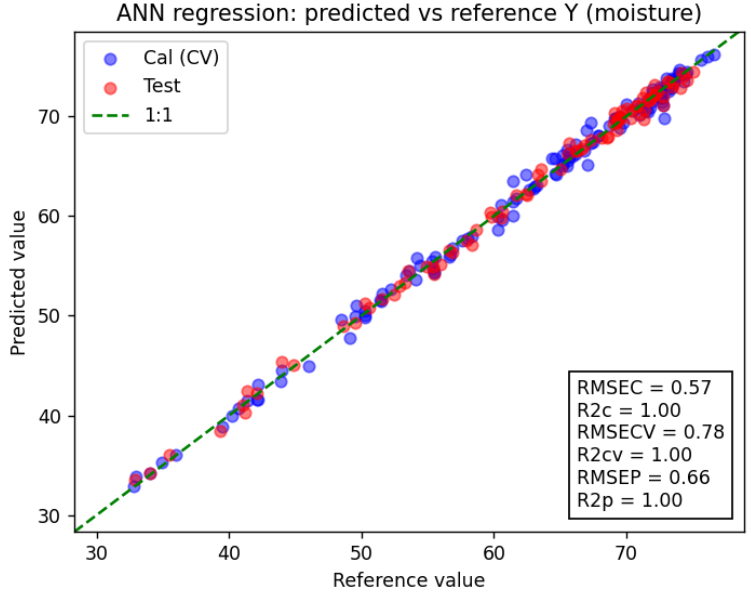
PLS parameters:

- 8 latent variables



SVM parameters:

- RBF kernel
- Epsilon = 0.1
- Cost = 4000
- Gamma = 0.004



ANN parameters:

- lbfgs solver
- tanh activation function
- 1 hidden layer
- 2 neurons per layer
- Learning rate = 0.01

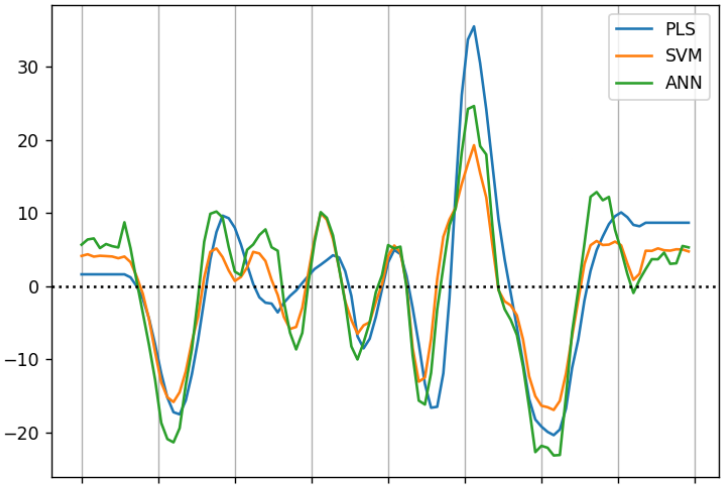
Application – interpretability criteria for moisture content

Criterion using SHAP values

> Method based on the computation of SHAP values for each model

> Different amplitude but similar shape to PLS coefficients for PLS, SVM and ANN models

Approximation of coefficients with a method based on SHAP values

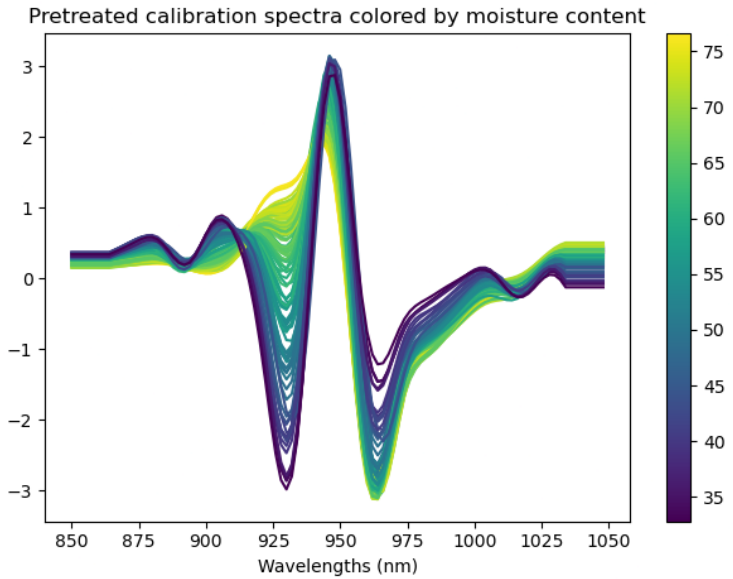


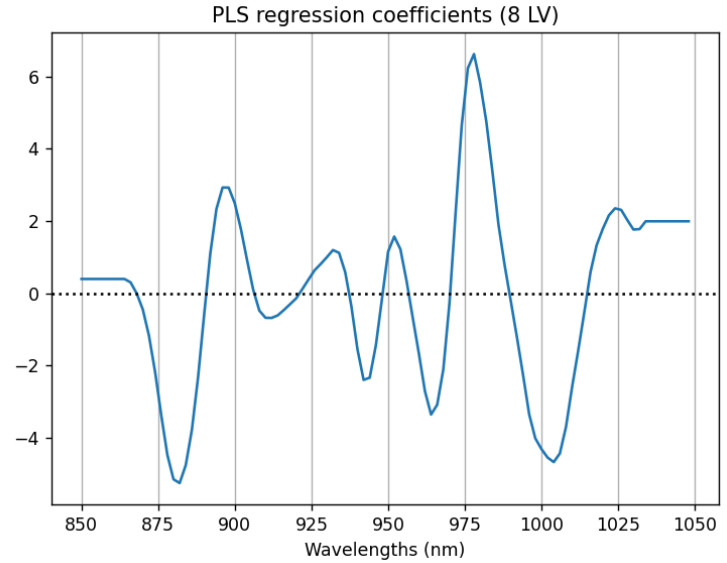
SHAP values criterion

☺ Similarity with PLS regression coefficients

☺ Seems applicable to spectroscopic data despite correlations between variables

☹ Computation time can be long



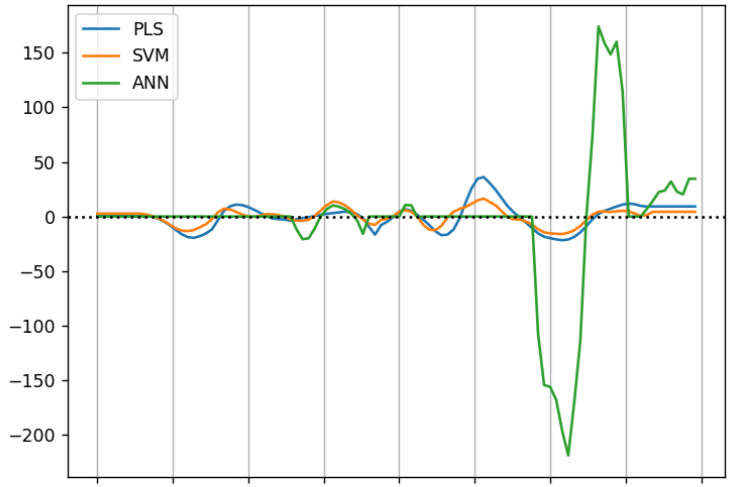


Application – interpretability criteria for moisture content

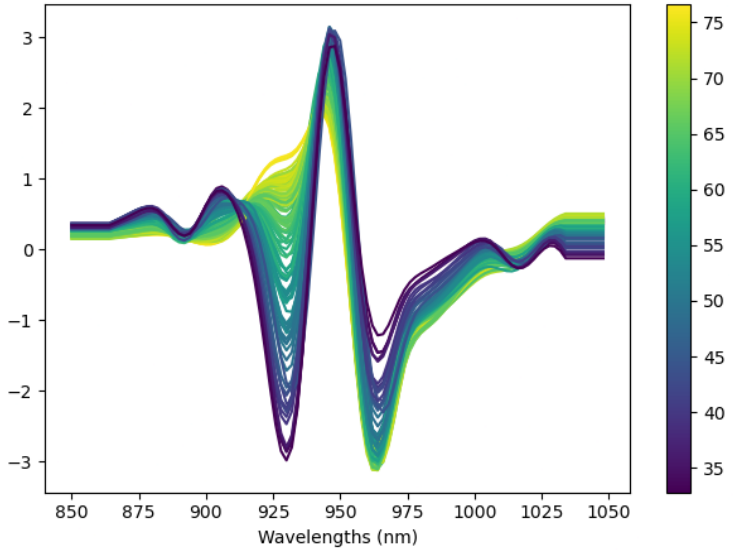
Criterion using pseudo-samples

- Method based on the computation of pseudo-samples predictions for each model
- Different amplitude but similar shape to PLS coefficients for PLS and SVM models, but inconsistent profile for ANN

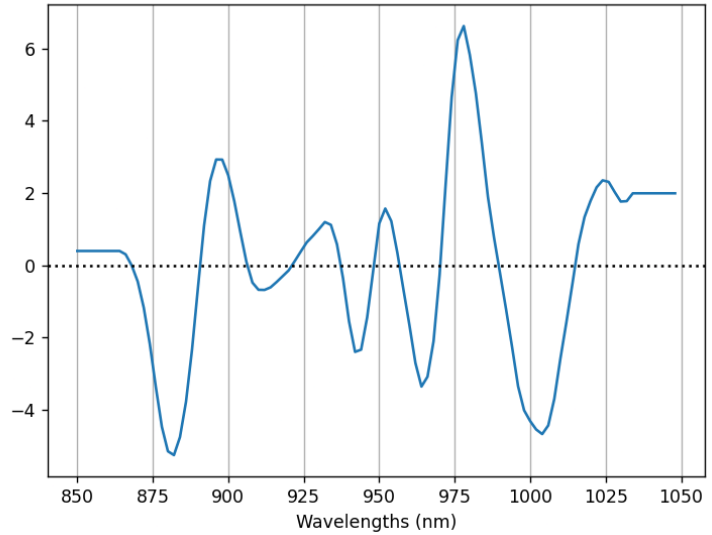
Approximation of coefficients with a method based on pseudo-samples



Pretreated calibration spectra colored by moisture content



PLS regression coefficients (8 LV)



Application – interpretability criteria for moisture content

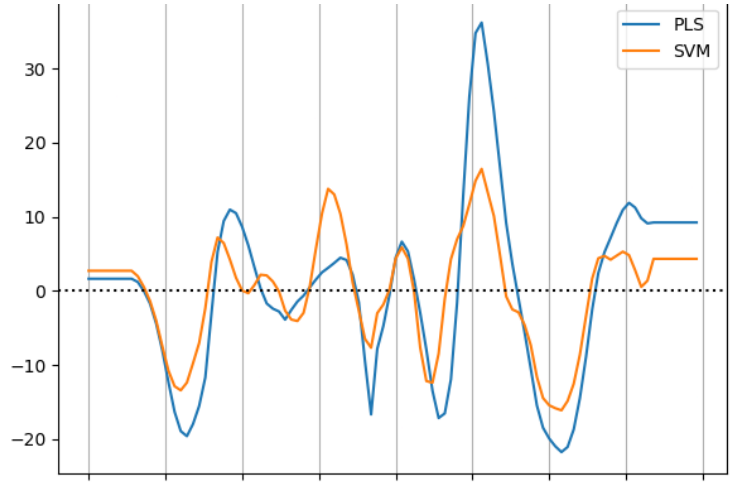
Criterion using pseudo-samples

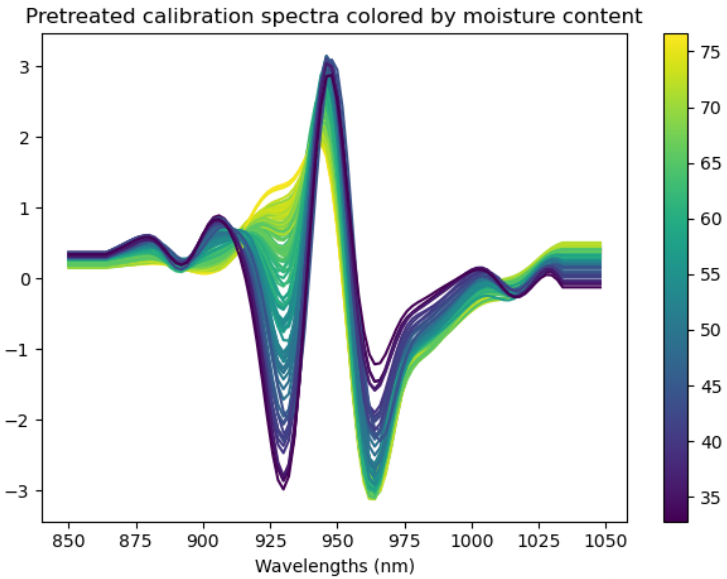
- Method based on the computation of pseudo-samples predictions for each model
- Different amplitude but similar shape to PLS coefficients for PLS and SVM models

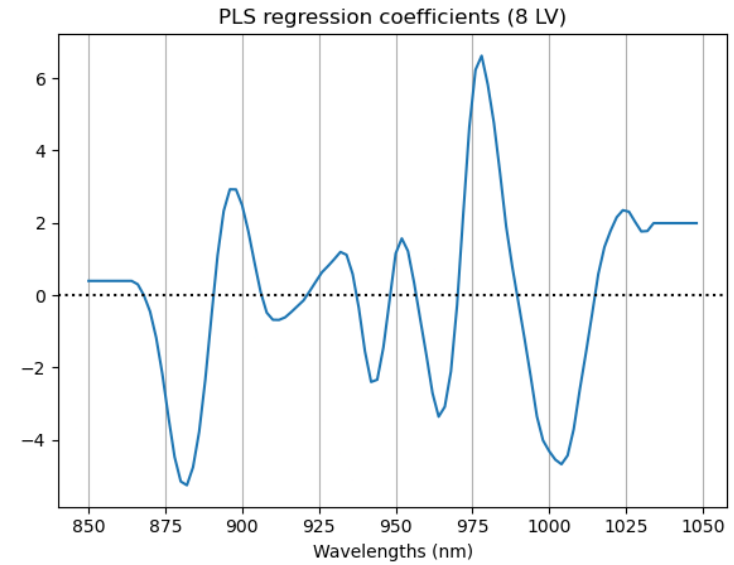
Pseudo-samples criterion

☺ Similarity with PLS regression coefficients
☺ Seems applicable to spectroscopic data despite correlations between variables
☹ Not adapted for ANN

Approximation of coefficients with a method based on pseudo-samples







Application – interpretability criteria for moisture content

Criterion using sensitivity test

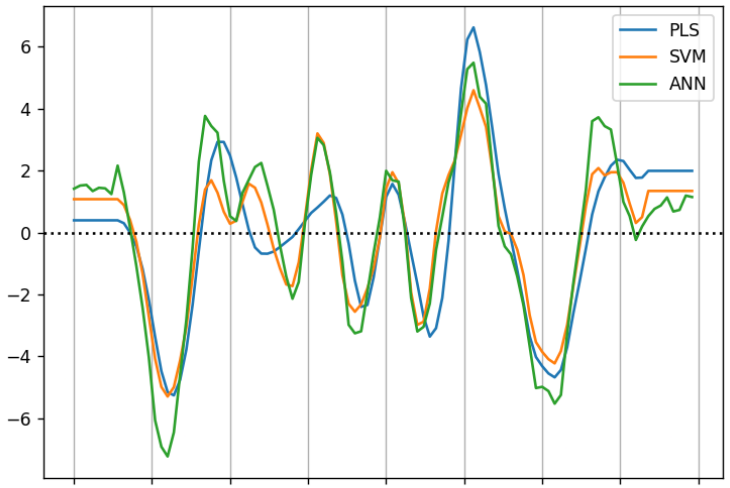
> Method based on the computation of a sensitivity test around the mean spectrum for each model

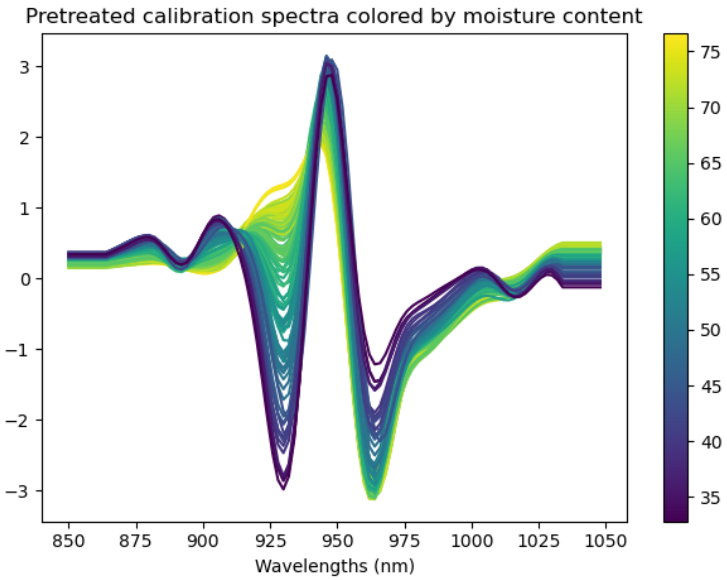
> Very similar to PLS coefficients for PLS model, and consistent shape and amplitude for SVM and ANN

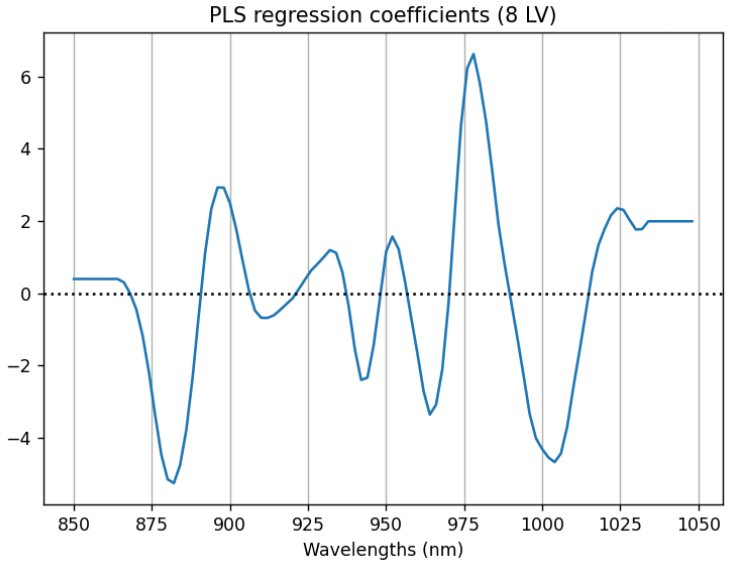
Sensitivity test criterion

☺ Similarity with PLS regression coefficients
☺ Seems applicable to spectroscopic data despite correlations between variables
☺ Short computation time

Approximation of coefficients with a method based on a sensitivity test







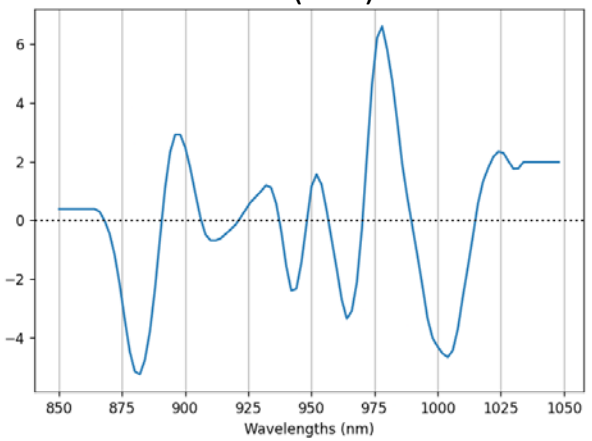
Overfitting detection

Tecator dataset: prediction of moisture content

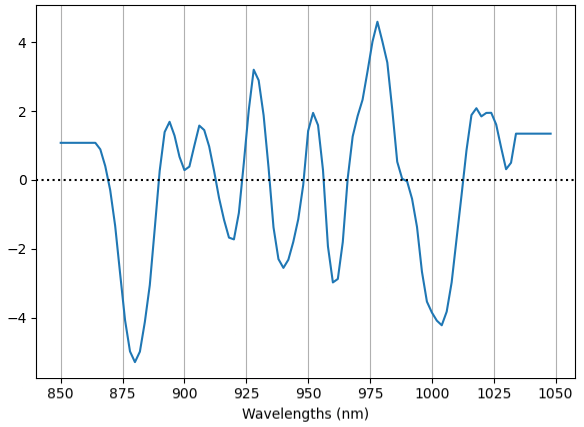
- Comparison between optimized models (top) and models of greater complexity (bottom)
- Example with sensitivity test method

+ overfitting -

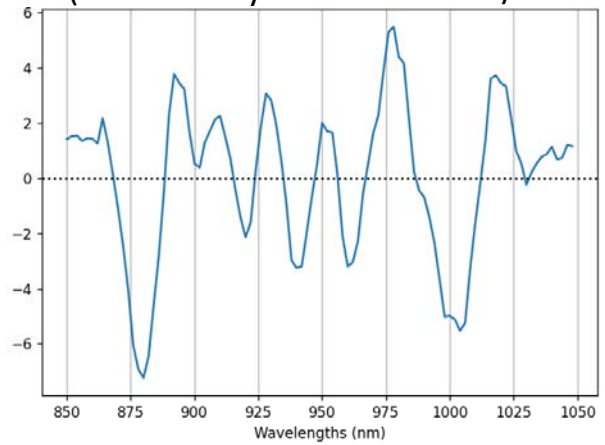
PLS (8 LV)



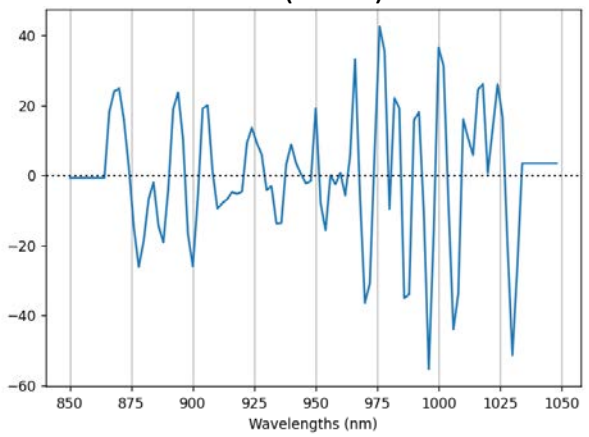
SVM (ε=0.1 / C=4000 / γ=0.004)



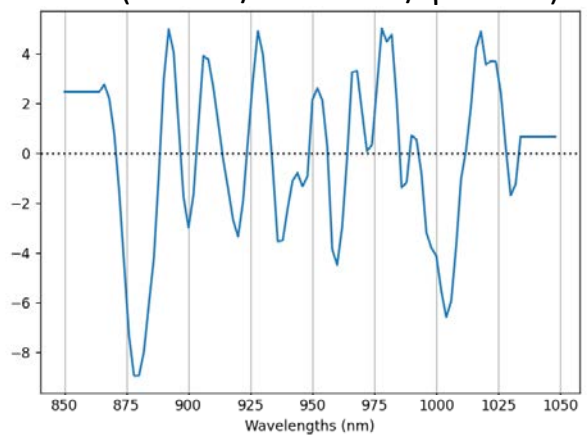
ANN (1 hidden layer of 2 neurons / α=0.01)



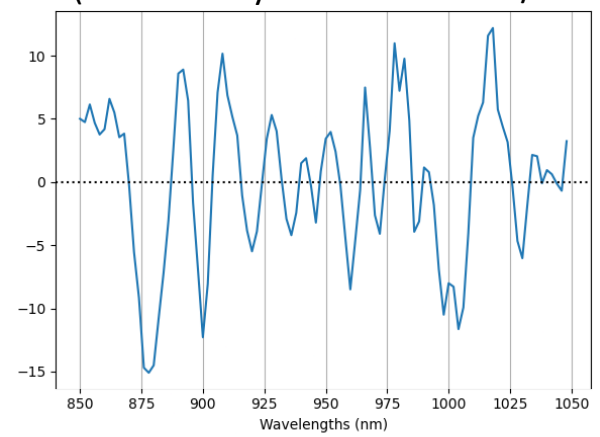
PLS (18 LV)



SVM (ε=0.05 / C=10000 / γ=0.006)



ANN (2 hidden layers of 3 neurons / α=0.005)



Conclusions

➔ Explainable AI / Interpretability tools for Machine Learning

1. Explain models: understand model structure by estimating model coefficients
2. Diagnose overfitting: help to optimize hyperparameters and avoid overfitted ML models

Conclusions

➔ Explainable AI / Interpretability tools for Machine Learning

1. **Explain models:** understand model structure by estimating model coefficients
2. **Diagnose overfitting:** help to optimize hyperparameters and avoid overfitted ML models

- The different methods result in good approximations of the PLS regression coefficients
  - > ☺ Can be used to explain Machine Learning models applied to spectroscopic data
  - > ☺ Can be used to diagnose overfitting by checking the amount of noise
  - > ☹ Longer computation time for the method based on SHAP values
  - > ☹ Method based on pseudo-sample predictions is not applicable to ANN
  - > ☺ Sensitivity tests are easy and fast to compute, and applicable to all models tested

Conclusions

➔ Explainable AI / Interpretability tools for Machine Learning

1. **Explain models:** understand model structure by estimating model coefficients
2. **Diagnose overfitting:** help to optimize hyperparameters and avoid overfitted ML models

- The different methods result in good approximations of the PLS regression coefficients
  - > ☺ Can be used to explain Machine Learning models applied to spectroscopic data
  - > ☺ Can be used to diagnose overfitting by checking the amount of noise
  - > ☹ Longer computation time for method based on SHAP values
  - > ☹ Method based on pseudo-sample predictions is not applicable to ANN
  - > ☺ Sensitivity tests are easy and fast to compute, and applicable to all models tested

- Interpretability methods can be computed with various software/languages
  - > PLS_Toolbox, Matlab, Python, R, Julia, …

Thank you for your attention!
Any questions?

Discover our Machine Learning services

R&D Services
- Feasibility studies
- Model development
- Model transfer

Software
- PLS_Toolbox®
- SOLO®

Training / Coaching

Training in advanced methods of Machine Learning
- Open-courses: 14-15 Oct. 2025
- In-house training sessions

ondalys

Tel : +33 (0)4 67 67 97 87 - www.ondalys.fr