Development and validation of near-infrared spectroscopy models for predicting nitrogen and carbon contents in rapeseed tissues (Brassica napus)

Sophie Rolland, Retailleau Lucile, Françoise Leprince, Elise Alix, Aurélien Carrillo, Solenn Guichard, Bernard Moulin, Tolleneare Alina, Christophe Tardy, Françoise Le Cahérec, et al.

Development and validation of near-infrared spectroscopy models for predicting nitrogen and carbon contents in rapeseed tissues (*Brassica napus*)

Sophie ROLLAND[1], Lucile RETAILLEAU[1], Françoise LEPRINCE[1], Elise ALIX[1], Aurélien CARILLO[1], Solenn GUICHARD[1], Bernard MOULIN[1], Alina TOLLENAERE[1], Christophe TARDY[2], Françoise LE CAHÉREC[1], Nathalie NESI[1], Anne LAPERCHE[1]

This project was supported by the RAPSODYN project

CONTEXT

Primary metabolites play a crucial role in plant developmental processes such as yield elaboration or stress response. As a consequence, determining the nitrogen (N) and carbon (C) contents throughout the plant cycle is of high importance to unravel and compare the performance of rapeseed varieties. Equations are already available and extensively used to predict grain quality traits, but few references exist to predict N or C content from powder samples. However, in the context of genetic studies, huge number of samples is generally required acquired along the whole crop cycle and conventional methods for N and C analyses show limitations due to low throughput. The aim of this study was to develop a low-cost and high-throughput tool to assess the N and C contents of thousands of samples differing in terms of organs (stem, root, leaves …) in a short time. The use of near-infrared spectroscopy (NIRS) calibration models was therefore questioned to address this issue.

PIPELINE

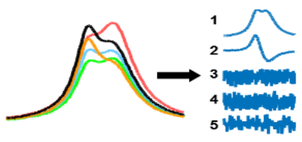
1. Data acquisition

Reference data ✚ Spectra data

2. Cleaning and partitioning datasets

→ Outliers detection
Calibration & validation datasets

3. Pretreatments

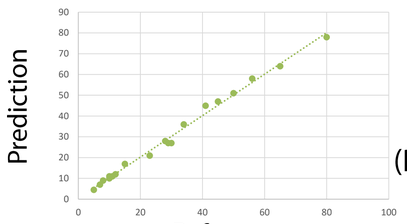


Selecting spectral ranges
Data prepocessing

4. Creation of a model

→ PLS regression

5. Validation of prediction model



Statistic performance indicators (RMSEP, R^2, RPD)

MATERIEL AND METHODS

❑ Highly diversified samples

Around 2800 samples of **63 winter oilseed rape genotypes** from 3 crop seasons were selected to be representative of the diversity of our experimental designs :

- **Growth conditions**

Field
(3 years, ~1400 samples)

Semi-controlled conditions (SCC)
water-stress (2 years, ~400 samples)

Semi-controlled conditions (SCC)
N-stress (1 year, ~1000 samples)

- **Organs** : leaf, stem (+/- branching), tap root, root, flower, pod

- **Developmental stages**

❑ Sample preparation

Samples were dried for two days in an oven (70°C) and grinding using the Tissue Lyser II (Qiagen®) in order to obtain a fine and homogeneous powder.

❑ Dumas combustion for C and N reference analyses

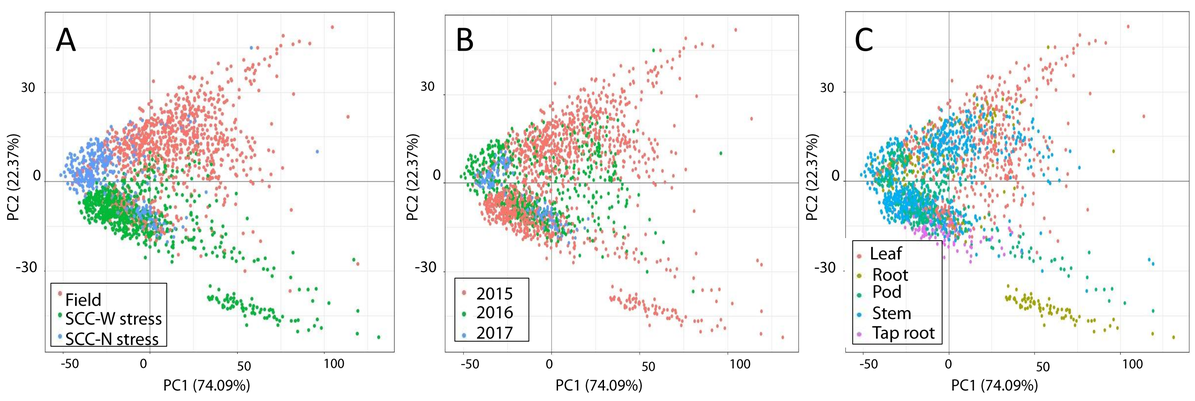
Nitrogen and carbon contents were determined by Dumas combustion method with an automated CN analyser (Vario Micro, Elementar®) from 5 to 10 mg of dry samples.

❑ Spectra data acquisition and development of prediction models

Spectra acquisition was performed using a near-infrared analyser MPA (Bruker®). Spectra were collected in reflectance mode from 4000 to 10000 cm^{-1} with an 16 cm^{-1} optical resolution and obtained as an average of 64 scans.
Calibration models were performed using the software OPUS 8.1 (Bruker®). The data were randomly subdivised into a set of calibration and a set of validation (50/50). The quality of each model was estimated by statistic parameters : coefficient of determination (R^2), root mean square error of prediction (RMSEP), ratio of performance to deviation (RPD), and error pourcentage between prediction and reference data.

RESULTS

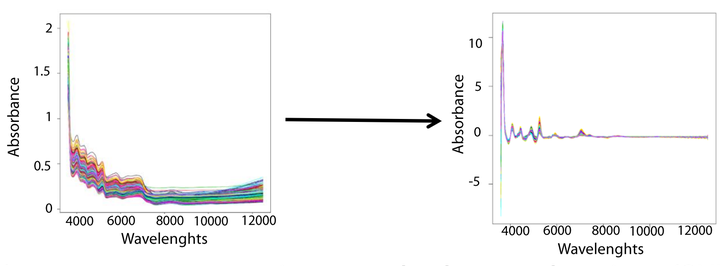
1- Diversified samples for a continuuous cloud



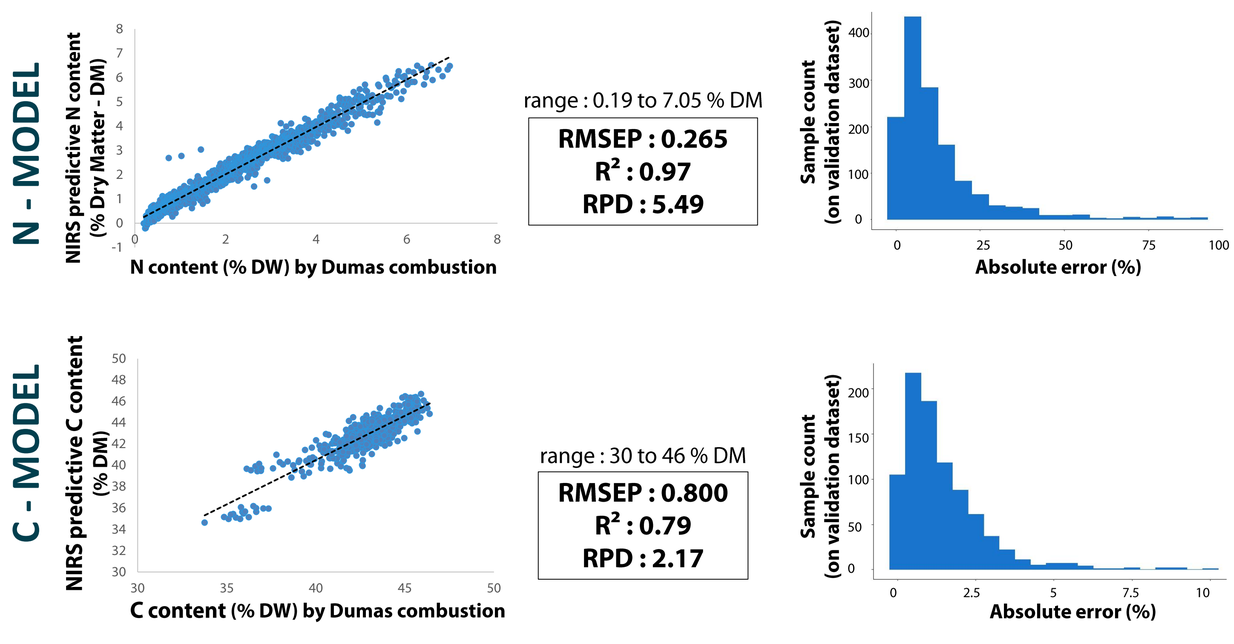
The spectra data were analyzed by PCA (above) and 96% of the sample set diversity was captured by the two first axes (PC1, PC2). In addition, a structuration of all spectra data was shown according to the growth conditions (A), crop season (B) or organ type (C). However, the high number of samples allowed to avoid major clusters. There was no effect of the genotype on the PCA (data not shown).

2- Spectra smoothing

In order to reduce baseline drift and spectra deformation due to physical parameters (granulometry, temperature …), a pretreatment using standard normal variate (SNV) and first derivate was applied to the spectra dataset.



3- Predictive models



N and C predictive models were established using a partial least square (PLS) algorithm regression. The predictions for the N model were very high (R^2>0.9, RPD>> 3) with less than 10% of error for 60% of the samples. The least relevant predictions were for the poorest samples in nitrogen (%N < 1% DM).

According to the statistical parameters, C model seemed less efficient. However, 98% of the samples were predicted with an error less than 5%. In the validation dataset, the absolute error prediction did not exceed 10%, which allowed validating the model.

CONCLUSIONS

We validated the prediction of N and C contents on grinded samples of rapeseed by NIRS measurements. This method is accurate, rapid and low-cost. When introducing new sample sets, it is necessary to implement the models by integrating 20 to 30 additional spectra. This allows erasing the variability that could be due to the environmental variations of the culture or preparation of the samples (grinding, temperature ...).

1 IGEPP, INRA, Agrocampus Ouest, Univ. Rennes, 35650 Le Rheu, France

2 Bruker France SAS Division Optics, 77447 Marne la Vallée, France

www.rennes.inra.fr