



Development of NIR calibration models to assess year-to-year variation in total non-structural carbohydrates in grasses using PLSR

Nisha Shetty^{a,*}, René Gislum^a, Anne Mette Dahl Jensen^b, Birte Boelt^a

^a Aarhus University, Science and Technology, Department of Agroecology, Research Centre Flakkebjerg, DK-4200, Slagelse, Denmark

^b University of Copenhagen, Faculty of Life Sciences, Forest and Landscape, Rolighedsvej 23, DK-1958, Frederiksberg C, Denmark

ARTICLE INFO

Article history:

Received 4 July 2011

Received in revised form 3 November 2011

Accepted 9 November 2011

Available online 20 November 2011

Keywords:

NIR

PLSR

iPLS

TNC

Grasses

Fructan

ABSTRACT

Near-infrared (NIR) spectroscopy was used in combination with chemometrics to quantify total non-structural carbohydrates (TNC) in grass samples in order to overcome year-to-year variation. A total of 1103 above-ground plant and root samples were collected from different field and pot experiments and with various experimental designs in the period from 2001 to 2005. A calibration model was developed using partial least squares regression (PLSR). The calibration model on a large data set spanning five years demonstrated that quantification of TNC using NIR spectroscopy was possible with an acceptable low root mean square of prediction error (RMSEP) of 1.30. However, for some years the estimated RMSEP was too optimistic as year-to-year variation for new years was not included in the model. Interval partial least squares (iPLS) regression was applied to remove non-relevant spectral regions and in order to improve model performance, but still it was not possible to avoid year-to-year variation using iPLS, however iPLS simplified the interpretation of the regression model. The best option was to expand the database with samples from a new year, to include these samples in the calibration model and to apply this on the remaining samples from the future year.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Total non-structural carbohydrates (TNC) include water soluble carbohydrates (WSC) as well as starch and fructans. In cool season grasses, WSC consist mainly of the simple sugars: glucose, fructose and sucrose, whereas fructans are storage carbohydrates based on a long chain of fructose polymers. Fructans are considered an alternative to starch for storing carbohydrates when sucrose concentration is above a certain threshold level, and recently fructans are suggested to provide protection against environmental stresses such as drought [1]. Fructans are gaining increased attention due to their potential for enhancing the forage quality of plants for ruminants [2] and involvement regrowth after defoliation [3]. The increased focus on TNC has resulted in an increased desire to measure WSC and fructans. Although quantification of WSC is easier to perform than detailed structural analysis of the fructans, it is still impractical to characterise a large number of samples. Therefore there is a need for rapid methods for quantification of TNC. Development of an innovative method for quantification and/or screening of samples for TNC will be of interest in breeding and grass testing programmes. The ability to measure fructans in large numbers of samples rapidly will facilitate the

identification of grasses with higher forage quality through a higher and more stable concentration of fructan during the growing season.

Near-infrared (NIR) spectroscopy has been used for the screening of crossbred material from cereal breeding programmes since the late 1970s, and the ease and speed of the technique have made it possible to perform thousands of quality tests in the time between harvesting of one generation and sowing of the next [4].

Despite the successful implementation of NIR within cereal breeding programmes, NIR has gained limited attention as a method for quality testing in grass breeding. It is not that the use of NIR for prediction of quality parameters in grass has not been investigated. Numerous studies have been performed and are thoroughly described in Andrés et al. [5], as well as in newer publications such as Feuerstein and Paul [6] describing the use of on-line NIR on a plot harvester, while the use of NIR for determination of quality parameters in dry and grinded grass samples was discussed by Gislum et al. [7]. However, most of these studies have been proof-of-concept studies made on a small number of samples (<100) not representing variation between e.g. years, cultivars, species or locations. An exception to the small number sample study was made by Shetty and Gislum [8], who developed PLSR models using 1459 samples covering a large variation. Even though some of these studies have demonstrated good correlations between quality parameters and NIR, they have not resulted in the development of methods that are in use in the industry.

In order to make robust models (capable of performing well when applied to the new data) that can handle the multivariate nature of

* Corresponding author. Tel.: +45 89156000; fax: +45 89156082.
E-mail address: nisha.shetty@agrsci.dk (N. Shetty).

the samples and the highly collinear NIR spectroscopic data it is essential to apply appropriate data analytical techniques such as the chemometric methods principal component analysis (PCA) [9], partial least squares regression (PLSR) [10], and interval PLS (iPLS) [11]. Here, PLSR will be used in the pursuit of a robust method. Furthermore it will be investigated whether variable selection can improve the predictive ability (or robustness) of the models as well as to perceive on which vibrational bands in the spectra, information on TNC can be found. For this purpose iPLS which searches for a spectral interval that is particularly informative with respect to the parameter under consideration, was used. This method often leads to improvements of prediction ability over standard full-spectrum PLSR models [11–13]. This is due to the high degree of redundancy in an NIR spectrum. Furthermore, there may be large parts of the spectra which do not correlate to the constituent of interest, and thus this part of the spectra may add noise, or inconsistencies to the model. Making a model on only the relevant part of the spectra can thus lead to superior regression models.

PLSR was used to develop NIR calibration model to assess year-to-year variation in TNC in grasses, when the model is applied on future samples. Even though variation between different species could also be the topic, the present study focuses on the year-to-year variation. The high year-to-year variation in the present study was due to location, species and climate.

2. Materials and methods

2.1. Materials

A total of 1103 above-ground plant and root samples were collected from different field and pot experiments from 2001 to 2005 (abbreviated from 1 to 5). Plants including roots were dug up from the soil; plants were separated from the roots at soil surface and then roots were hand washed. The sample set consisted of red fescue (*Festuca rubra* L.), perennial ryegrass (*Lolium perenne* L.), smooth stalked meadow grass (*Poa pratensis* L.) and orchard grass (*Dactylis glomerata* L.) samples. Plant and root samples were weighed and freeze-dried and weighed again. Afterwards the samples were ground using a Cyclotec 1093 sample mill (Foss, Hillerød, Denmark) to pass through a 1 mm screen prior to determination of carbohydrate concentration on aliquots of the samples.

2.2. Chemical analysis

TNC were extracted using a 0.1 M sodium acetate buffer and hydrolysed with 0.037 M sulphuric acid. Quantification of TNC was done using a coupled enzymatic assay procedure [14]. Results are presented in percentage on a dry matter basis.

The reproducibility of the measurement of TNC (\mathbf{y} data) was estimated as the standard deviation of differences (SDD). SDD was calculated on five measurements of four samples, three samples with different TNC concentrations and one sample with the same extract (Table 1).

$$SDD = \sqrt{\frac{\sum (d_i - d_m)^2}{(n-1)}}$$

where d_i = difference in \mathbf{y} between five replicate measurements of sample i , d_m = mean value of all replicate differences ($\sum d_i/n$) and n = number of samples.

The repeatability of the \mathbf{y} data was calculated as the standard deviation (SD) on one sample which was measured in almost each run ($n=48$).

$$SD = \sqrt{\frac{\sum (y_i - y_m)^2}{(n-1)}}$$

Table 1

Mean and standard deviation of difference (SDD) of TNC for five measurements of four samples (A, B, C and D) listed in order of TNC. In samples A, B and C five sub-samples from the three samples were weighed and analysed. In sample D five sub-samples were measured from the same extract.

Sample	Mean TNC	SDD TNC
C	3.47	1.84
B	15.00	3.47
A	17.98	4.18
D	26.57	6.07

where y_i is the y -value of replicate i , y_m is the average y value of all replicates and n is the number of replicates.

SD and SDD were used to evaluate the performance of the PLSR models.

2.3. NIR measurements

Reflectance spectra of dried and ground plant and root samples were obtained using a QFA-Flex 400 FT-NIR instrument (Q-interline, Roskilde, Denmark). The samples were packed as uniformly as possible in glass vials (height 6 cm, diameter 2.6 cm) and measured using a rotating sample device. The sample was rotated at three rounds per minute. The measuring sample window at the rotating sample device had a diameter of 6 mm and the analysis surface was $\approx 510 \text{ mm}^2$. Spectra were collected at every 2 nm in the NIR region from 1100 to 2498 nm. One spectrum was obtained for each sample as an average of 64 sub-scans. The spectra were reported as $\log(1/R)$. Using this procedure, approximately 20 samples could be analysed per hour.

2.4. Multivariate data analysis

PCA was performed as an explorative data analysis in order to obtain a first look at the structure of the data, to identify outliers and to delineate classes. The PLSR method was used to derive calibration models. PLSR models were obtained on raw and pre-processed NIR spectra. Pre-processing included Savitzky–Golay first derivative (1 d) and second derivative (2 d) [15] averaging over 5 points using a second order polynomial and multiplicative signal correction (MSC) [16]. Root mean square error of cross-validation (RMSECV) plotted against the number of PLSR components using different pre-processing methods is used to select the optimum pre-processing method and optimum number of components in the PLSR model. The optimum number of PLSR components was chosen as the first local minimum in the smooth declining RMSECV curve or the point where this curve flattened. Random cross-validation with 10 segments and 9 iterations was used.

The performance of the PLSR models were evaluated using the root mean square error of prediction (RMSEP), standard error of performance (SEP) and bias. The relationship between RMSEP, SEP and bias is: $\text{RMSEP}^2 \approx \text{SEP}^2 + \text{Bias}^2$, if bias is ≈ 0 then $\text{RMSEP}^2 \approx \text{SEP}^2$ [17]. Initially, the obtained model was validated using 20% of the available data. The test set was selected as every fifth sample after sorting according to TNC to ensure that samples from the whole concentration range were represented in the test set. Models were also validated using one of the years as test set in a loop in order to evaluate how year-to-year variations affected the model.

All data analyses were carried out using MATLAB version 7.9. (R2009b) (The Mathworks, Inc., Natick, MA, USA) along with the PLS toolbox 5.5.2 (Eigenvector Research, Inc., Manson, WA, USA).

3. Results and discussion

The SDD of TNC ranged from 1.84% TNC to 4.18% TNC depending on the mean TNC concentration when sub-samples were analysed from the same sample (Table 1). When a repeated measurement of

Download English Version:

<https://daneshyari.com/en/article/1179935>

Download Persian Version:

<https://daneshyari.com/article/1179935>

[Daneshyari.com](https://daneshyari.com)