Food Chemistry 162 (2014) 10–15

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Analytical Methods

Determination of grain protein content by near-infrared spectrometry and multivariate calibration in barley

Chen Lin, Xue Chen, Lei Jian, Chunhai Shi, Xiaoli Jin *, Guoping Zhang *

Agronomy Department, Key Laboratory of Crop Germplasm Resource of Zhejiang Province, Zhejiang University, Hangzhou 310058, China

ARTICLE INFO

Article history:
Received 14 October 2013
Received in revised form 2 April 2014
Accepted 13 April 2014
Available online 24 April 2014

Keywords:
Grain protein content (GPC)
Barley (*Hordeum vulgare* L.)
Near-infrared spectroscopy (NIRS)
Least squares support vector machine regression (LSSVR)

ABSTRACT

Grain protein content (GPC) is an important quality determinant in barley. This research aimed to explore the relationship between GPC and diffuse reflectance spectra in barley. The results indicate that normalizing, and taking first-order derivatives can improve the class models by enhancing signal-to-noise ratio, reducing baseline and background shifts. The most accurate and stable models were obtained with derivative spectra for GPC. Three multivariate calibrations including least squares support vector machine regression (LSSVR), partial least squares (PLS), and radial basis function (RBF) neural network were adopted for development of GPC determination models. The Lin_LSSVR and RBF_LSSVR models showed higher accuracy than PLS and RBF_NN models. Thirteen spectral wavelengths were found to possess large spectrum variation and show high contribution to calibration models. From the present study, the calibration models of GPC in barley were successfully developed and could be applied to quality control in malting, feed processing, and breeding selection.

1. Introduction

Grain protein content (GPC) is an important quality determinant in cereal crops. In barley, GPC is closely associated with feed and malt quality (Clancy, Han, & Ullrich, 2003; See, Kephart, & Blake, 2002). Higher protein content is favourable for feed quality, while lower or moderate protein content is expected for malt barley. GPC affects malting quality in many ways, including yeast nutrition, haze formation in beer and enzyme activities (Clancy et al., 2003; See et al., 2002). GPC estimation by a conventional method is time consuming, labour intensive, and in addition a lot of seeds should be used. In most cases, there are limited seeds in the early generation of breeding program to be analyzed for GPC measurement (Cai et al., 2013). Obviously, it is imperative to have a simple, rapid and highly effective way of GPC analysis.

Near-infrared spectroscopy (NIRS) is a very efficient method for high-throughput screening of plant materials for their chemical characteristics. The NIR spectrum is correlated with a sample's chemical composition based on vibrational properties of the organic molecular chemical bonds and their interactions with infrared radiation (Pasquini, 2003). Recently, the NIRS technique became popular in the qualitative and quantitative analyses of biological and non-biological materials in the agriculture, food, textile, petrochemical and pharmaceutical fields (Huang et al., 2012; Li & He, 2010; Salguero-Chaparro, Baeten, Fernández-Pierna, & Peña-Rodríguez, 2013; Xu, Shi, Ye, Yan, & Yu, 2013). The spectroscopic technique has a lot of advantages comparing to the classical chemical and physical analysis, such as a short measuring time, little sample, nondestruction and the determination of multi-composition at the same time (Lammertyn, Nicolaï, Ooms, Smedt, & Baerdemaeker, 1998). Several studies have focused on applying NIRS as an alternative method for the biomass assessments, such as analysis of evaluation of nutritional value in naked oats (Bellato et al., 2011), shea tree (*Vitellaria paradoxa*) and nut fat profiles (Davrieux et al., 2010), evaluation of protein, tryptophan, and lysine in maize (Rosales, Galicia, Oviedo, Islas, & Palacios-Rojas, 2011). In short, the spectroscopic technique has been proved to be quite promising in evaluation of chemical composition in plants.

However, no report has been found so far in NIRS determination of GPC in barley. In this work, we performed a NIRS assay for barley GPC using 277 barley samples, and provided multiple options for a quick and consistent prediction of barley GPC from large populations of samples. The barley GPC determination model was developed by using linear multivariate calibration techniques (partial least squares and multiple linear regressions), so some nonlinear phenomena cannot be interpreted in the model. In this study, a more powerful nonlinear least squares support vector machine regression (LSSVR) was used for the development of the determination model for GPC in barley, and a systematic comparison

* Corresponding authors. Tel./fax: +86 (0) 57188981688 (X. Jin).
E-mail addresses: jinxl@zju.edu.cn (X. Jin), zhanggp@zju.edu.cn (G. Zhang).

http://dx.doi.org/10.1016/j.foodchem.2014.04.056
0308-8146/

among LSSVR, the artificial neural network (ANN), and PLS was made.

2. Plant materials and methods

2.1. Plant materials

Two hundred and seventy-seven barley genotypes were planted at the Huajiachi campus of Zhejiang University (Hangzhou, China, 120.0°E. 30.5°N) in the early winter of 2009. Each genotype was sown into a two-line plot, 2 m long and 0.24 m interval between lines, and 40 seeds were planted in each line. All plots were supplied with 150 kg/ha of N, including 40 kg/ha of N as compound fertilizer applied before seeding, and 110 kg/ha of N as urea supplied at two-leaf stage and booting stage, respectively with equal amount. In addition, 180 kg/ha of potassium chloride was applied prior to seeding. The experiments were arranged in a block design with two replications. In each block, the 277 barley genotypes were arranged randomly. All other agronomic managements, including weed and disease control, were the same as those applied locally. At maturity, seeds were harvested and dried, then stored at 4 °C for GPC analysis. GPC of all samples were measured, three measurements were done for each sample.

2.2. GPC measurement

Mature grains were ground in a Cyclotec 1093 sample mill (Tecator AB, Hoganas, Sweden) and passed through a 0.5 mm screen. GPC was measured using the Kjeldahl method (Kjeldahl, 1983) with three replication for each sample. Protein content is calculated by duplicating a factor of 6.25 with N content (Mariotti, Tomé, & Mirand, 2008).

2.3. NIRS measurement

The ground grains containing about 2.5 g were loaded on a circle sample cup (35 mm in diameter and 18 mm in depth) and pressed slightly to obtain similar packing density. The loading time was as short as possible to avoid excessive moisture absorption. All the samples were scanned by the NIRS monochromator, and the corresponding spectra were collected using a NIRSystems 5000 (Silver Spring, USA) instrument in reflectance mode. Thirty-two scans were performed for both the reference and each sample. Acquisition of the spectra was accomplished in the wavelength range from 1100 to 2500 nm with an interval of 2 nm by using the WinISI II (InfraSoft International, USA) software. In addition, each sample was loaded and scanned 4 times, and the average spectrum of each of the four recordings was used for NIR analysis. In modeling, all 277 samples were divided into the calibration set and the prediction set with a ratio of 2:1. To avoid bias in subset partition, all samples were first arranged in an ascending order according to their respective GPC values, and then each sample was picked out from every three samples consecutively, resulting in 92 samples of prediction set for the validation, and the remaining 185 samples formed calibration set. The statistical information of Y-value of each set was shown in Table S1. The calibration models were validated using full cross-validation (Gomez, He, & Pereira, 2006). The prediction set was also used to validate the actual prediction ability of resulted models as an external test set (Esteban-Diez, Gonzalez-Saiz, Saenz-Gonzalez, & Pizarro, 2007).

2.4. Data analysis

In order to heighten the contribution of the chemical composition to the spectral signal by reducing the systematic noise, some spectral preprocessing methods were applied comparatively. The spectra were treated with moving average smoothing that the segment size was set to 3, area normalize, spectroscopic transformation, multiplicative scatter correction (MSC), the first derivative of the calibration spectra calculated with 3 gaps of data points, linear baseline correction and standard normal variate (SNV), respectively. The pretreatments and regression algorithm named as partial least-squares regression (PLS) were carried out according to the instructions of the Unscrambler V9.5 (CAMO PROCESS AS, Oslo, Norway). The PLS is a bilinear modeling method for the relationship between a set of independent spectral variables (X) and a single dependent variable (Y). LSSVR is an interesting formulation of SVM regression (Suykens & Vanderwalle, 1999). It uses a linear set of equations to obtain the support vectors. The standard LSSVM algorithm was defined by Suykens and Vanderwalle (1999). The LSSVR was carried out based on the LSSVM toolbox of MATLAB (Version 7.8.0.347, The MathWorks. Inc US). Radial basis function neural network (RBF_NN) is a type of nonlinear neural network, which is used to solve several types of classification and regression problems. The theory of RBF_NN has been described extensively (Despagne & Massart, 1998).All calculations of RBF_NN were implemented based on the Neural Networks toolbox of MATLAB (Version 7.8.0.347, The MathWorks. Inc US).

The performance of regression models was evaluated by the following criteria. Firstly, the quality of the regression model was quantified by standard error of calibration (SEC), standard error of prediction (SEP), and the correlation coefficient (r) between the predicted and measured parameters. A model with a low SEC, a low SEP, and a high r was considered as a good model (Li & He, 2006). Secondly, the residual predictive deviation (RPD), defined as the ratio between standard deviation (SD) of the samples' reference values and SEC for NIR spectroscopy calibrations, was a good index to evaluate the quality of regression models (Arana, Jaren, & Arazuri, 2005; Fearn, 2002). A relatively high RPD value indicates that the model is able to reliably predict the chemical composition (Arana et al., 2005).

3. Results and discussion

3.1. Chemical analysis and the character of the reflectance spectra

In this study, the 277 samples were divided randomly into two groups: A training set was used to develop the calibration models (185 samples), and the remaining samples were used as testing set (92 samples). The GPC of calibration set ranged from 8.588% to 14.40% with a mean of 11.14%, while the GPC of testing set covered from 8.82% to 14.36% with a mean of 11.16% (Table S1). The range of the y values of training samples almost covered that of samples in the testing set. As regard to spectral variability, the training samples and the testing samples were evaluated by principal component analysis on spectral data. Fig. 1 shows all samples (including samples in the training set and the testing set) in the first two principal components space. The first and second component accounted for 92% and 7% of the raw spectral data, respectively. Totally, the first two components represented 99% variation of the raw spectral data (Fig. 1). All samples in the testing set distributed equably in the training samples. Thus, the division of the samples was appropriate for near infrared analysis.

In order to optimize the raw data of near infrared reflectance spectra of barley GPC, regression modeling techniques are required. Moreover, the different pre-treatment model in barley GPC was evaluated (Table S2). Seen from Table S2, preprocessing methods can generally improve the regression performance in terms of sensitivity and specificity, indicating the regression models can be improved by reducing scattering effects, baseline shifts