Talanta 131 (2015) 236–242

Contents lists available at ScienceDirect

Talanta

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

Chemometric optimization of the robustness of the near infrared spectroscopic method in wheat quality control

Milica Pojić [a,*], Dušan Rakić [b], Živorad Lazić [c]

[a] Institute of Food Technology, University of Novi Sad, Bulevar cara Lazara 1, 21000 Novi Sad, Serbia
[b] Faculty of Technology, University of Novi Sad, Bulevar cara Lazara 1, 21000 Novi Sad, Serbia
[c] BASF Catalysts LLC, Technical Center, 108 Briarcliff Rd. Gordon, GA 31031, USA

ARTICLE INFO

Article history:
Received 14 March 2014
Received in revised form
21 July 2014
Accepted 21 July 2014
Available online 30 July 2014

Keywords:
Near infrared spectroscopy
Screening design
Fractional factorial design
Optimization
Propagation of error
Wheat quality control

ABSTRACT

A chemometric approach was applied for the optimization of the robustness of the NIRS method for wheat quality control. Due to the high number of experimental ($n=6$) and response variables to be studied ($n=7$) the optimization experiment was divided into two stages: screening stage in order to evaluate which of the considered variables were significant, and optimization stage to optimize the identified factors in the previously selected experimental domain. The significant variables were identified by using fractional factorial experimental design, whilst Box-Wilson rotatable central composite design (CCRD) was run to obtain the optimal values for the significant variables. The measured responses included: moisture, protein and wet gluten content, Zeleny sedimentation value and deformation energy. In order to achieve the minimal variation in responses, the optimal factor settings were found by minimizing the propagation of error (POE). The simultaneous optimization of factors was conducted by desirability function. The highest desirability of 87.63% was accomplished by setting up experimental conditions as follows: 19.9 °C for sample temperature, 19.3 °C for ambient temperature and 240 V for instrument voltage.

1. Introduction

The optimization in analytical chemistry is performed in order to discover the operational conditions at which the observed procedure provides the most reliable response, and therefore to improve the performance of an analytical system [1]. In this regard, it is closely associated with the determination of method robustness. The traditional optimization techniques which included the monitoring of the influence of one factor at a time (OVAT) have been overcome by application of chemometrics techniques involving the design of experiments (DOE). DOE techniques imply the reduction of the number of experiments to be performed, the development of mathematical models enabling the assessment of the relevance of the factor effects being studied, their interactions and statistical significance [1–7]. Among them, the most frequently used technique for optimization of analytical methods and understanding of system performance is response surface methodology (RSM) [1,8,9]. The multivariable optimization of analytical methods has been widely performed especially for extraction methods, spectroanalytical methods, chromatographic methods, capillary electrophoresis, electroanalytical methods and thermogravimetry [6].

The robustness of the near infrared spectroscopic method (NIRS) has not be extensively studied due to the fact that it is considered as a built-in method characteristic accomplished by employment of so-called repeatability file (REP file) during the development of the NIRS calibration model. REP file is commonly introduced during calibration model development so that it contains spectral information on the uncontrolled variations that may occur in routine operation which must be minimized and hence improve the model robustness (i.e., temperature variations, small differences between cups, different levels of cup wear, and differences between instruments). Adequately designed and structured REP file enables the minimization of the uncontrolled variations affecting NIRS analysis. However, the REP file concept is a special feature only available in WinISI calibration software – software by Infrasoft International, LLC [10–14]. However, the background of calibration development procedure is not commonly available to the average end-user of the NIRS method, who nevertheless must demonstrate its suitability for intended purpose. The robustness test of the NIRS method for its application in analyzing wheat samples examines the potential sources of variability in responses (analytical and spectral). To achieve this, a number of factors related to the operating procedure (operational factors) and ambient conditions (ambient factors) are

* Corresponding author. Tel.: +381 21 485 3782; fax: +381 21 450 730.
E-mail address: milica.pojic@fins.uns.ac.rs (M. Pojić).

http://dx.doi.org/10.1016/j.talanta.2014.07.059
0039-9140/

examined in an experimental design, in an intervals that slightly exceed the variation that can be expected in a routine use of the method [7,15]. By examining the impact of various factors (operational and ambient) using screening Plackett–Burman factorial experimental design, Pojić et al. [16] identified the factors with the most influencing impact on analytical and spectral responses of the NIRS method as being applied for wheat moisture, protein and wet gluten content predictions. Starting from the importance of reliable results of wheat quality in trade, where, for example, the premiums of 1.30–1.50$ are paid for increments of 0.1% in protein content per ton [17], as well as the growing importance of ensuring of confident measurement results, the objectives of the present study were defined: to (i) indicate the most important factors that might affect the variations in NIRS responses when determining the selected wheat quality parameters and to (ii) use the response surface methodology to find the optimal factor settings which exhibit the minimum variability in responses around the specified target values.

2. Materials and methods

The chosen experimental design was performed in two stages. The initial stage involved a *screening design* performed to select the most important factors that might affect variations in the NIRS responses. The subsequent stage involved the performance of response surface methodology with previously identified significant factors, to obtain the least square adequate model to test the robustness by the propagation of error (POE) method, hereinafter called *optimization design* [1,7,18]. The POE method was used to find the factor settings that minimize the variation in responses [19,20]. Both screening and optimization designs were setup and processed using software Design-Expert 9, trial version (StatEase, Inc., Minneapolis, USA).

2.1. Screening design

Two-level fractional factorial design (minimum run resolution V, $n=22$ runs) was set up in a way to vary six experimental factors: number of subsamples in the NIRS measurement (A), ambient temperature (B), sample temperature (C), ambient air humidity (D), instrument voltage (E) and lamp aging (F). Experimental factors and their levels in screening design are given in Table 1. The levels were selected on the basis of the NIR systems specification as well as variations that are most likely to happen when the method is transferred between different laboratories, different instruments or over time. All experiments were carried out in duplicate and averaged. The significant effects were identified by graphical interpretation of the estimated effects and analyzing half-normal probability plot, where the significant effects were identified as outliers from the straight line (the non-significant effects were lying on a straight line passing through zero) (these data not shown). Contribution of main effects and their interactions is calculated as ratio of sum of squares for observed effect and total sum of squares.

Table 1
Factors and their levels investigated during the screening design.

Factor	Low value (-1)	High value (+1)
Number of subsamples (*A*)	5	15
Sample temperature (*B*) (°C)	5	35
Ambient temperature (*C*) (°C)	10	30
Ambient humidity (*D*) (%)	40	80
Instrument voltage (*E*) (V)	200	240
Lamp aging (*F*)	New	Old

2.2. Optimization design

The next step was to determine the optimal factor settings to achieve the minimal variation in responses. For this purpose, the response surface methodology (RSM) using a Box-Wilson rotatable central composite design (CCRD) was adopted. The significant factors identified in the previous phase comprised: sample temperature (A), ambient temperature (B) and instrument voltage (C). A CCRD comprised of a two-level full factorial design (2^f experiments), a star design ($2f$ experiments) and six center points, thus requiring 20 experiments to examine the impact of f factors ($N=2^f+2f+6$). The points of the full factorial design were set at -1 and $+1$ factor levels, whilst those of star design were set at 0, $-\alpha$ and $+\alpha$, where α level was 1.682 due to three factors selected [8]. The central points were set at the factors level 0. Experimental factors and their levels in optimization design are given in Table 2. The measurements of all responses were carried out in duplicate and averaged.

A second-degree polynomial (quadratic) model was used to describe the relation between the response(s) and the three factors under consideration:

$$Y = b + \sum_{i=1}^{3} a_i X_i + \sum_{i=1}^{3} a_{ii} X_{ii}^2 + \sum_{i=1}^{3} \sum_{j=i+1}^{3} a_{ij} X_i X_j$$

Where: Y is the modeled response(s) for obtained response(s) (protein content, moisture content, wet gluten content, Zeleny sedimentation value and deformation energy), b_0 is an intercept; X_i is the factor and a_i is the corresponding coefficient; X_{ii} is the quadratic factor; a_{ii} is the quadratic coefficient; X_{ij} is the two-factor interaction; and a_{ij} is the two-factor interaction coefficient. The statistical significance of the terms in the regression equation was determined by analysis of variance (ANOVA) for each response.

The adequacy of the model was evaluated by coefficient of determination (R^2) and model p-value. A subsequent step in optimization design comprised the determination of the optimal factor settings derived from response surfaces built with the design results. The optimal factor settings were specified to achieve the minimal variation in responses, which was conducted by minimizing the propagation of error (POE). POE is defined as the amount of variation transmitted to the response and it was derived from variability originating from control factors and the normal process variation obtained from ANOVA [9,21]:

$$\mathrm{POE} = \sqrt{\sigma_Y^2}$$

$$\sigma_Y = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\right)^2 \sigma_{x_i}^2 + \sigma_{resid}^2$$

The simultaneous optimization was conducted by desirability function.

2.3. Experimental conditions

The sample temperature was adjusted by using chamber convection thermostat (TMA Bodalec & Bodalec, Oborovo, Croatia), whilst the ambient temperature was adjusted by using air

Table 2
Experimental factors and their levels in optimization design.

Factor	Lowest value ($-\alpha$)	Low value (-1)	Central value (0)	High value ($+1$)	Highest value ($+\alpha$)
Sample temperature (*A*) (°C)	−5	5	20	35	45
Ambient temperature (*B*) (°C)	3	10	20	30	37
Instrument voltage (*C*) (V)	186	200	220	240	254