Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 201 (2018) 229–235

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

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

Quantitative determination of wool in textile by near-infrared spectroscopy and multivariate models

Hui Chen [a,b], Chao Tan [a,*], Zan Lin [a,c]

[a] Key Lab of Process Analysis and Control of Sichuan Universities, Yibin University, Yibin, Sichuan 644000, China
[b] Hospital, Yibin University, Yibin, Sichuan 644000, China
[c] Department f Orthopedics, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China

ARTICLE INFO

Article history:
Received 25 March 2018
Received in revised form 29 April 2018
Accepted 3 May 2018
Available online 4 May 2018

Keywords:
Near-infrared
Textile
Wool
Partial least squares
Elastic component regression

ABSTRACT

The wool content in textiles is a key quality index and the corresponding quantitative analysis takes an important position due to common adulterations in both raw and finished textiles. Conventional methods are maybe complicated, destructive, time-consuming, environment-unfriendly. Developing a quick, easy-to-use and green alternative method is interesting. The work focuses on exploring the feasibility of combining near-infrared (NIR) spectroscopy and several partial least squares (PLS)-based algorithms and elastic component regression (ECR) algorithms for measuring wool content in textile. A total of 108 cloth samples with wool content ranging from 0% to 100% (w/w) were collected and all the compositions are really existent in the market. The dataset was divided equally into the training and test sets for developing and validating calibration models. When using local PLS, the original spectrum axis was split into 20 sub-intervals. No obvious difference of performance can be seen for the local PLS models. The ECR model is comparable or superior to the other models due its flexibility, i.e., being transition state from PCR to PLS. It seems that ECR combined with NIR technique may be a potential method for determining wool content in textile products. In addition, it might have regulatory advantages to avoid time-consuming and environmental-unfriendly chemical analysis.

1. Introduction

Wool is the textile fiber originated from sheep and other animals and mainly consists of protein together with a few percent lipids. The wool content in blended fabrics is considered as a key quality indicator of textile. The qualitative and quantitative analysis of wool content takes an important position due to common adulterations in both raw and finished textiles [1]. Also, in term of the national standard of China, all textile products should be marked by fabric type and composition on the label [2]. Conventional methods of determining wool content such as chemical dissolution and microscopic methods are complicated, destructive, time-consuming, environment-unfriendly and are easily interrupted by people [3,4]. In addition, these methods heavily depends on the experience and skill of the analyst, developing a quick, easy-to-use and green alternative method has been an expectation of many people.

In recent years, with the development of instrument techniques and chemometrics, near-infrared (NIR) spectroscopy has become a potential tool for many qualitative and quantitative tasks in many fields including the tobacco [5], food [6–10], drug [11–14], material [15], medical [16,17], and petrochemical [18] industries. NIR spectroscopy is a fast, non-destructive and environmentally friendly technique since it does not require sample pretreatment, nor the use of reagents and solvents, and not generates residues. Also, it provides the possibility of simultaneous determination of several components or parameters in a same sample by a single spectral measurement. It is widely recognized that NIR spectroscopy can serve as a good alternative to traditional chemical and instrumental methods. Applications of NIR technique in textile field have been advanced dramatically because it can maintain the integrity of samples. Unlike to mid-infrared (MIR) spectroscopy, present peaks that are characteristics of functional groups of the sample constituents, the peaks in NIR region (800–2500 nm) mainly correspond to the combinations and overtones of the sample functional groups and, and can thus be exhibited broadly and weakly. Despite having most overlapping signals, NIR still becomes one of the most used techniques. The NIR analysis of various fibers including wool has been reported [19–22].

It is well known that NIR-based method is an indirect analysis technique and depends heavily on chemometrics to develop a calibration model to relate the information of interest of a sample with its spectrum, the performance of which is also decisive. Undoubtedly, partial least squares (PLS) is the most popular multivariate tool. Often, each NIR spectrum maybe contain several hundreds even thousands of

* Corresponding author.
E-mail address: chaotan1112@163.com (C. Tan).

https://doi.org/10.1016/j.saa.2018.05.010
1386-1425/

variables. A lot of theoretical and experimental researches have shown that variable/region selection is very helpful for improving the performance of the full-spectrum PLS models. Different types of variable reduction techniques generally fall into two categories: separate variable selection and interval selection [23,24]. Generally, spectroscopists have a preference for continuous interval selection in the original space based on the interpretability of the result. In this regards, a powerful toolbox used for several interval-based PLS algorithms was developed by Nørgaard [25]. It contains interval PLS (iPLS), backward interval PLS (biPLS), Synergy interval partial least squares (siPLS), and moving window PLS, etc. The iPLS splits the whole spectrum region into a series of sub-intervals (variable-wise), constructs PLS models for each subinterval and provides the results in one plot. Similar to iPLS, taking intervals as block of information, biPLS applies iPLS to the data and the followed by backward elimination, each time eliminating the interval whose removal results in the best model. The siPLS is actually an all possible sub-interval combination tests. The mwPLS aims at seeking a spectral window with optimal information by a moving concept.

The present work was to explore the feasibility of combining NIR spectroscopy and interval-based PLS for determining wool content in textile. Full-spectrum PLS, four kinds of local PLS algorithms (iPLS, biPLS, siPLS and mwPLS) and a new algorithm called elastic component regression (ECR) were used to model. A total of 108 cloth samples with wool content ranging from 0% to 100% (w/w) were collected and all the compositions are really existent in the market. The dataset was divided equally into the training and test sets for developing and validating calibration models. When using local PLS, the original spectrum axis was split into 20 sub-intervals. No obvious difference of performance can be observed for the these models based on interval/window selection. It seems that ECR combined with NIR technique may be a potential method for determining wool content in textile products. In addition, it might have regulatory advantages to avoid time-consuming and environmental-unfriendly chemical analysis.

2. Theory and Methods

2.1. Partial Least Squares (PLS)

Partial least squares (PLS) is a classic multivariate calibration algorithms [26,27]. Similar to principal component regression (PCR), it aims to seek a few linear combinations of the original x-values, also called latent variables/factors, and to use them in the regression equation between the spectrum of samples and response variable.

Let $\mathbf{X}(n \times p)$ and $\mathbf{y}(n \times 1)$ be spectral matrix and the response, respectively, and both are mean-centered in advance. Here, n is the number of samples and p is the number of variables. The model structure can be represented as following:

$$\begin{cases} \mathbf{X} = \mathbf{TP}' + \mathbf{E} \\ \mathbf{y} = \mathbf{Uq} + \mathbf{f} \end{cases} \tag{1}$$

where, **P** and **q** are loading matrix/vector; **T** and **U** are score matrices; **E** and **f** are residuals. Since the model is linear in both loadings and scores and is thus called bilinear model. The key problem of PLS model is to extract the first several factors. Often, the first factor are extracted by maximizing the squared covariance between $\mathbf{Xw}_{pls}$ and **y**, which is involved in the following optimization:

$$\begin{cases} \max \mathbf{w}'_{pls}\mathbf{X}'\mathbf{y}'\mathbf{y}\mathbf{X}\mathbf{w}_{pls} \\ \text{s.t.}\mathbf{w}'_{pls}\mathbf{w}_{pls} = 1 \end{cases} \tag{2}$$

By stripping and repeating operation, a required number of factors can be obtained sequentially.

2.2. Elastic Component Regression (ECR)

By comparison, it can be found that the optimization problems corresponding to PLS described above and PCR (no shown here) are of the same structure and differ only in the criterion maximized. By introducing a so-called supervising factor, i.e., $\alpha \in (0,1)$, Li et al. suggested to use a general symmetric matrix expression:

$$\mathbf{H} = (1-\alpha)\mathbf{X}'\mathbf{X} + \alpha\mathbf{X}'\mathbf{y}'\mathbf{y}\mathbf{X} \tag{3}$$

The α serves to controls the information amount of **y** in **H**. If $\alpha = 0$, then $\mathbf{H} = \mathbf{X}'\mathbf{X}$.

If $\alpha = 1$, then $\mathbf{H} = \mathbf{X}'\mathbf{y}'\mathbf{y}\mathbf{X}$. The two extremes correspond to PCR and PLS, respectively. When $0 < \alpha < 1$, it can be considered as transition state from PCR to PLS. Now, the optimization problem can be written as:

$$\begin{cases} \max \mathbf{w}'\mathbf{H}\mathbf{w} \\ \text{s.t.}\mathbf{w}'\mathbf{w} = 1 \end{cases} \tag{4}$$

Such a scheme is the basis of elastic component regression (ECR) [28]. It consist of several steps when using ECR: First, a matrix, i.e., $\mathbf{H} = (1-\alpha)\mathbf{X}'\mathbf{X} + \alpha\mathbf{X}'\mathbf{y}'\mathbf{y}\mathbf{X}$ is computed; second, the eigenvector related to the largest eigenvalue of **H** is computed and denoted as **w**; third, x-score, x-loading and y-loading are computed by $\mathbf{t} = \mathbf{Hw}$, $\mathbf{p} = \mathbf{X}'\mathbf{t}/\mathbf{t}'\mathbf{t}$ and $\mathbf{r} = \mathbf{y}'\mathbf{t}/\mathbf{t}'\mathbf{t}$, respectively. Finally, both **X** and **y** are deflated by $\mathbf{X} = \mathbf{X} - \mathbf{tp}'$ and $\mathbf{y} = \mathbf{y} - \mathbf{tr}'$, respectively. Such a procedure is repeated until the required number of components is obtained. The resultant ECR model is expressed as a regression vector:

$$\mathbf{b} = \mathbf{W}(\mathbf{P}'\mathbf{W})^{-1'}\mathbf{R}' \tag{5}$$

where **W**, **P** and **R** stand for the weight, x-loading and y-loading matrix, respectively. ECR has been used to near-infrared-based calibration and exhibits good performance.

2.3. Interval Partial Least Squares (iPLS)

The interval partial least squares (iPLS) algorithm, developed by Norgaard [25], is a combination of interval selection and PLS. The basic principle of iPLS is to develop local PLS models on equidistant subintervals of the full-spectrum. Then, these local PLS models are compared with the global model based on the validation parameter, i.e., root mean squared error of cross-validation (RMSECV). One of the

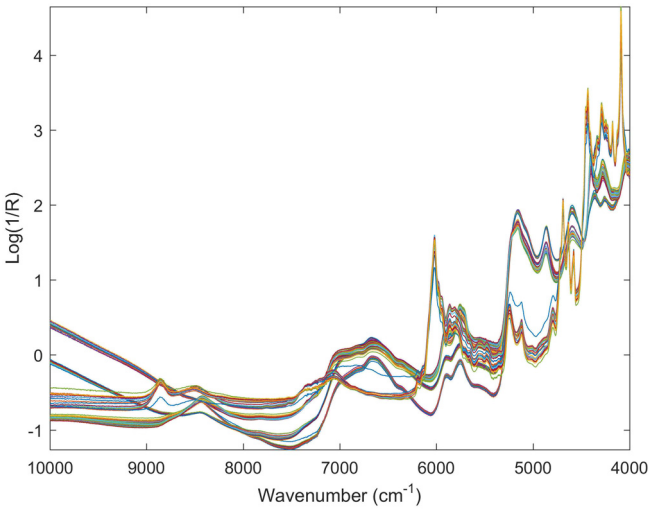

Fig. 1. Near-infrared spectra of all samples.