Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 149 (2015) 271–278

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

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

Rapid analysis of diesel fuel properties by near infrared reflectance spectra

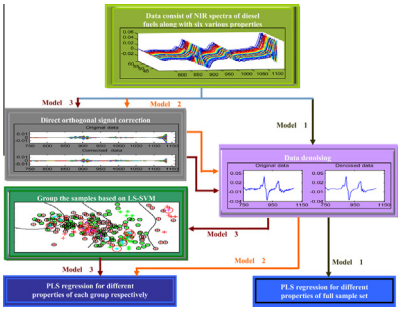
Fei Feng, Qiongshui Wu *, Libo Zeng

Electronic Information School, Wuhan University, Wuhan 430072, Hubei, China

HIGHLIGHTS

- Applied PLS with NIR spectra for multi-parameter analysis of fuel properties.
- Models for LS-SVM grouped samples had higher precision and smaller prediction error.
- Models based on DOS correction yielded a considerable error reduction.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 12 December 2014
Received in revised form 20 March 2015
Accepted 21 April 2015
Available online 4 May 2015

Keywords:
Near infrared reflectance spectra
Partial least square regression
Support vector machine
Direct orthogonal signal correction

ABSTRACT

In this study, based on near infrared reflectance spectra (NIRS) of 441 samples from four diesel groups (−10# diesel, −20# diesel, −35# diesel, and inferior diesel), three spectral analysis models were established by using partial least square (PLS) regression for the six diesel properties (i.e., boiling point, cetane number, density, freezing temperature, total aromatics, and viscosity) respectively. In model 1, all the samples were processed as a whole; in model 2 and model 3, samples were firstly classified into four groups by least square support vector machine (LS-SVM), and then partial least square regression models were applied to each group and each property. The main difference between model 2 and model 3 was that the latter used the direct orthogonal signal correction (DOSC), which helped to get rid of the non-relevant variation in the spectra. Comparing these three models, two results could be concluded: (1) models for grouped samples had higher precision and smaller prediction error; (2) models with DOSC after LS-SVM classification yielded a considerable error reduction compared to models without DOSC.

Introduction

Diesel is one of the main power sources of vehicles. Using poor quality or inferior fuel could damage the engines and cause air pollutants. Rapid analysis of diesel fuel properties is of great interest. Currently, the main detection methods include Gas Chromatography (GC), Nuclear Magnetic Resonance, etc [1,2]. However, the GC requires complicated operations and has lower efficiency, and the nuclear method has potential dangerousness. It is important to find an alternative rapid and safe approach for multi-parameter analysis of fuel properties.

Recently, near infrared reflectance spectra (NIRS) has become an effective method for rapid and real-time analysis of fuel properties [3–13]. Organic functional groups that contain hydrogen can form broad absorption bands within near infrared wavelength range. Diesel fuel mainly consists of organic materials such as cetane, which contains a large amount of functional groups.

* Corresponding author. Tel.: +86 18986016327.
E-mail address: qswu@whu.edu.cn (Q. Wu).

http://dx.doi.org/10.1016/j.saa.2015.04.095
1386-1425/

Therefore, it is appropriate to apply NIRS to rapid analysis of diesel fuel properties. Considerable progress has been reported over the last decade [4–12], e.g., Felicio measured the composition of fuel by NIRS [10]; Coope predicted density, viscosity, and boiling point at 50% recovery of diesel fuel using NIRS [11]; Dehui Wu tried online measurement for diesel fuel with NIRS [12].

The NIR spectra analysis has two aspects, the quantitative and qualitative analysis. For quantitative analysis, the primary method is partial least square (PLS) regression [10,14–16]. Support vector machine (SVM) has been gradually applied to near-infrared spectra in qualitative analysis in recent years [17–19].

In this paper, as shown in Fig. 1, three models were established to analyze the six properties (i.e., boiling point, freezing temperature, density, viscosity, cetane number, and total aromatics). Each property plays its specific role on the performance of vehicle engines. The boiling point is the main factor that affects the atomization and evaporation [2], the freezing temperature is an important indicator to ensure the smooth supply of fuel in an engine [1,2], and the density and viscosity affect the diesel fuel flowing properties in pipeline [1]. Cetane number is another important properties, based on the cetane number, fuel can be divided into different brands [2]. The aromatics increase the sediment in engine and aromatic hydrocarbons produce toxic substances [2].

In this study, rapid analysis of diesel fuel properties was carried out with NIR spectra. Since models based on individual classified samples had better prediction ability than that based on a full set [13], we applied least square support vector machine (LS-SVM) to the spectra, which had high classification accuracy. Then, different PLS regression methods were built for different groups. To improve the prediction precision, we introduced direct orthogonal signal correction (DOSC) [20,21], models with and without DOSC were statistically compared. The advantage of this method used in this paper was that it could be done online without any sample pre-treatment. Besides, properties could be simultaneously analyzed, and the whole process only takes seconds.

Data and methods

Data sets

Data used in this paper was collected by Southwest Research Institute in a project sponsored by the U.S. Army. Eigenvector Research Incorporated provides 784 samples in *Near Infrared Spectra of Diesel Fuels* [23]. Some of the properties are not measured on some samples, in this case, samples which have missing values were removed. As a result, 441 samples were remained from four groups, the −10# group (190 samples), the −20# group (134 samples), the −35# group (62 samples) and the inferior group (55 samples).

The spectra of full sample was plot in Fig. 2, the *x*-axis was wavelength, ranged between 750 nm and 1550 nm with a resolution of 2 nm. The *y*-axis was cetane number of each sample and the z-axis was absorbance.

Fig. 3 presented four typical spectra of each group. The absorbance features appeared at the similar position in the spectra curve.

Pre-treatment methods

Since the raw data contains irrelevant information, data pre-treatment process is very necessary and important [22]. Pre-treatment methods were listed as follows:

(1) *Smoothing:* To eliminate or reduce the random noise, a Savitzky–Golay filter was designed and applied to the raw data. The filter coefficients were determined by an unweighted linear least square regression and a polynomial model of specified degree.
(2) *Multiplication scatter correction (MSC):* MSC is capable of reducing scattering effects which caused by sample size. Spectrum can be calibrated using the average spectrum, therefore, they had the same scatter level.
(3) *Derivative:* Derivative can correct the baseline drift and reduce the flat background interference. A Savitzky–Golay differentiation was applied for noise reduction. Specifically, the first derivation of the spectra was obtained by taking a window length of 21 and 4 order polynomial fitting.

Direct orthogonal signal correction

The pre-treatment method (1) to (3) are just for NIRS data itself without any dispose of the property data. In this case, direct orthogonal signal correction was used to remove the original data that had little relation with the property.

Here, we adopted the Westerhuis proposed DOSC algorithm which has a very strict orthogonal constraints. Assume that the spectral matrix is X and the property matrix is Y.

(1) The first step of DOSC is to decompose Y, $\widehat{Y}$ is the projection of Y onto X, and the residual part M is orthogonal to X. Thus,

$$\begin{cases} \widehat{Y} = P_X Y = X\left(\left(X^T\right)^{-1}\right)^T Y \\ W = Y - \widehat{Y} \end{cases} \quad (1)$$

(2) Set $Z = X - \widehat{Y}\left(\left(\widehat{Y}\right)^{-1} X\right)$, which is orthogonal to both Y and $\widehat{Y}$. Apply principle component analysis (PCA) to matrix Z and the principal component scores, matrix t, can be obtained. The weights matrix $w_\perp$ equals $X^{-1}t$, then the score vector $t_\perp$ can be expressed as $t_\perp = Xw_\perp$, the loading vector is

$$p_\perp = X^T t_\perp / (t_\perp^T t_\perp) \quad (2)$$

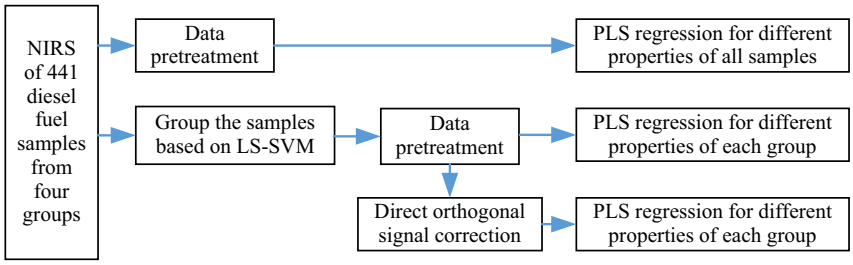


Fig. 1. Three models for rapid analysis of diesel fuel properties.