Chemometrics and Intelligent Laboratory Systems 107 (2011) 44–49

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Rapid identification and assay of crude oils based on moving-window correlation coefficient and near infrared spectral library

Xiao-Li Chu *, Yu-Peng Xu, Song-Bai Tian, Jing Wang, Wan-Zhen Lu

Research Institute of Petroleum Processing, Beijing 100083, PR China

ARTICLE INFO

Article history:
Received 24 June 2010
Received in revised form 9 January 2011
Accepted 14 January 2011
Available online 22 January 2011

Keywords:
Moving window
Correlation coefficient
Crude oil assay
Pattern recognition
Spectral searching

ABSTRACT

Based on near infrared spectroscopy, a method for rapid identification of crude oil type, named moving window correlation coefficient method, was proposed. This new spectral searching method can distinguish highly similar crude oil spectra. The main parameters of the moving window correlation coefficient method were discussed and selected in this study. Combined with crude oil assay database, detailed crude oil assay of unknown crude oils can be rapidly provided by this method. With the expansion of the NIR spectral library and assay database, this strategy would play more and more important role in making business decisions, and in planning, controlling and optimizing operations of refineries.

1. Introduction

Crude oil assay plays a very important role in petroleum exploitation, trading and processing. A complete set of conventional crude oil assay programs have been well-developed mainly based on ASTM methods, which can provide accurate and detailed information about the physicochemical properties of crude oil across the boiling range. However, those methods are time-consuming, tedious, costly, and labor-intensive. Crude oils must be sent in large volumes to the laboratory capable of performing the required tests. Physical distillation must be conducted for the determination of the distillation yield profiles (weight and volume %), and the narrow distillation cuts are collected for further testing. Distillation can take 2–4 days for completion and complete testing of various properties for each distillation cut may require 2–6 weeks [1]. Obviously, newer analytical methodologies are required to provide timely data for rapid decision making.

As a fast and viable alternative, spectroscopic methods such as nuclear magnetic resonance (NMR), Raman, infrared (IR), and near infrared (NIR) have been developed [2–5], since the corresponding spectra reflect the complete molecular composition of the petroleum and petroleum products. In recent years, there have been many successful applications of NIR spectroscopy combined with chemometric calibration modeling techniques for rapidly determining some important chemical and physical parameters of crude oil, such as specific gravity (API gravity) [6,7], true boiling point (TBP) curve [7,8], and SARA (saturates, aromatics, resin, and asphaltenes) components [9]. There are also researches in predicting the properties of residual fractions by using NIR spectra of crude oils. The predicted properties of residual fractions include density, sulfur content, pour point, and carbon residue [10]. Those studies have demonstrated that the NIR absorbance spectrum contains enough information to be successfully employed to characterize crude oil.

Unlike gasoline or diesel analysis, crude oil assay may involve hundreds of chemical and physical parameters for a whole sample and its subfractions. Therefore, development and maintenance of such large number of regression models built by traditional multivariate calibration methods such as partial least squares (PLS) one by one are infeasible. The aim of this study was to develop a new method to provide a complete crude oil assay data based on the NIR spectral library and crude oil assay database by pattern recognition techniques.

The solution proposed includes the following steps: (1) constructing a crude oil NIR spectral library and corresponding detailed crude oil assay database; (2) for an unknown crude oil sample, search its identical sample in the crude oil NIR spectral library using its NIR spectrum; and (3) detailed crude oil assay report of unknown crude oil can be rapidly obtained from the crude oil assay database according to the NIR recognition results. If there is no identical crude oil with the unknown crude oil in the library, a nearest crude oil type, matching degree and corresponding crude oil assay can be given, which can provide important information for making business decisions, and for planning, controlling, and optimizing operations in refineries.

Traditional pattern recognition methods, such as distance or correlation coefficient etc. [11,12], cannot easily identify crude oil samples because some of crude oil NIR spectra are highly similar. Therefore, a new spectra searching method, named moving window

* Corresponding author.
E-mail address: cxlyuli@sina.com (X.-L. Chu).

0169-7439/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2011.01.006

correlation coefficient method, was proposed in this paper, which can distinguish highly similar crude oils accurately.

2. Moving window correlation coefficient

The traditional correlation coefficient method is often used to compare the similarity of two spectra by all spectra variables or selected spectral regions, which can be expressed as follows:

$$R_{ij} = \frac{\sum_{k=1}^{n} (x_{ik} - \bar{x}_i)\left(x_{jk} - \bar{x}_j\right)}{\sqrt{\sum_{k=1}^{n} (x_{ik} - \bar{x}_i)^2 \sum_{k=1}^{n} \left(x_{jk} - \bar{x}_j\right)^2}} \quad (1)$$

where, $\bar{x}_i$ is the mean value of total absorbance of i spectra, $\bar{x}_j$ is the mean value of total absorbance of j spectra, and n is the number of wavelength channels. The correlation coefficient expresses the degree that, on an average, two spectral variables change correspondingly. A correlation coefficient of 1 implies that the two spectra have identical characteristics. The higher the correlation coefficient, the more similar the spectra.

The concept of moving window has been used for several chemometric methods, such as moving window average smoothing, moving window partial least squares method, and moving window two-dimensional correlation spectroscopy [13–16]. The moving-window average smoothing method is the classic and the simplest smoothing method, which can be utilized to enhance the signal-to-noise ratio for analytical signals. Through this treatment, the influence of noise can be reduced because the signals are often distributed normally on both positive and negative sides [13]. The moving window partial least squares method is a new wavelength interval selection method, which builds a series of PLS models in a window that moves over the whole spectral region and then locates useful spectral intervals in terms of the least complexity of PLS models reaching a desired error level [14,15]. Moving window two-dimensional correlation spectroscopy is the combination of the moving window concept with two-dimensional correlation spectroscopy to facilitate the analysis of complex datasets, which splits a large data matrix into smaller and simpler subsets and to analyze them instead of computing overall correlation. Complicated spectral intensity changes along the perturbation direction can be easily discriminated, and their characteristic perturbation points are precisely found [16].

The way of the moving window correlation coefficient method proposed by this study is to select a spectral window that starts at the kth spectral channel and ends at the $(k+w-1)$th spectral channel (the window width is w), moving the spectral window successively along the equally spectral data to construct a series of moving window (a total of $n-w+1$ windows), and then calculating corresponding correlation coefficients in each moving window using the traditional correlation coefficient formula.

It should be noted that the moving window correlation coefficients of the first $(w-1)/2$ and the last $(w-1)/2$ data cannot be calculated in the process. A vector of correlation coefficients with $n-w$ elements can be obtained by this method. If the two spectra compared are identical, the elements of the correlation coefficient vector in all moving windows are all equal to 1. If there is a difference in only a certain wavelength region between two spectra, the moving correlation coefficients of this wavelength region would be decreased obviously. Compared with traditional correlation coefficient scalar obtained by the full spectral region, this method can provide more difference information between two very similar spectra with very subtle differences, which can give more accurate identification results.

For the moving window correlation coefficient method, the window width is a very critical parameter. A narrow moving window would be useful to distinguish tiny differences between two spectra, but the risk is that the repetitive spectra of the same crude oil collected in different dates may not be accurately identified because of spectral measurement errors and instrumental variations. A wide moving window could reduce the influence of external test conditions such as temperature and humidity, but two different samples with very subtle differences may not be effectively distinguished. Thus, the window width needs to be optimized in the practical applications according to the differences of samples and the spectra measurement conditions. The guidelines for choosing a right value for this parameter will be discussed in Section 4.1.

3. Experimental

3.1. Crude oil samples and assay database

From 1998 to 2010, 265 crude oil samples were collected from our crude oil evaluation group in the Research Institute of Petroleum (RIPP) of SINOPEC. Those samples originate from more than 200 crude oil fields distributed within China and around the world. In our crude oil evaluation group, the detailed assay data of each crude oil collected were determined by traditional analytical methods according to conventional assay strategy. The crude oil assay database has been constructed based on the above samples and the corresponding assay data using Excel Worksheet.

3.2. Near infrared spectroscopy

Near infrared spectroscopy measurements were carried out at room temperature on a Thermo Antais II FT-NIR spectrometer. The spectra resolution was 8 cm^{-1} and 64 scans were accumulated over the range of 3500 to 10,000 cm^{-1} for each spectrum. A quartz

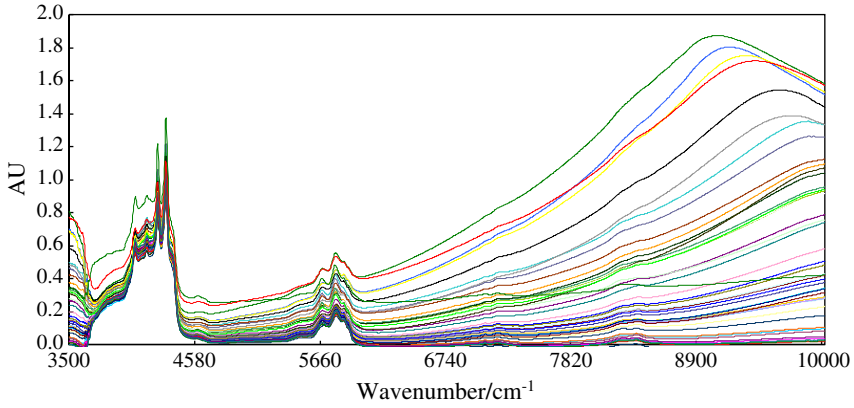

Fig. 1. Raw NIR absorbance spectra of crude oils in the range of 3500 to 10,000 cm^{-1}.