

Coal analysis by diffuse reflectance near-infrared spectroscopy: Hierarchical cluster and linear discriminant analysis

M.T. Bona, J.M. Andrés*

Instituto de Carboquímica, CSIC, Procesos Químicos, Miguel Luesma Castán, n. 4, 50018 Zaragoza, Spain

Received 22 November 2006; received in revised form 19 January 2007; accepted 22 January 2007

Available online 30 January 2007

Abstract

An extensive study was carried out in coal samples coming from several origins trying to establish a relationship between nine coal properties (moisture (%), ash (%), volatile matter (%), fixed carbon (%), heating value (kcal/kg), carbon (%), hydrogen (%), nitrogen (%) and sulphur (%)) and the corresponding near-infrared spectral data. This research was developed by applying both quantitative (partial least squares regression, PLS) and qualitative multivariate analysis techniques (hierarchical cluster analysis, HCA; linear discriminant analysis, LDA), to determine a methodology able to estimate property values for a new coal sample. For that, it was necessary to define homogeneous clusters, whose calibration equations could be obtained with accuracy and precision levels comparable to those provided by commercial online analysers and, study the discrimination level between these groups of samples attending only to the instrumental variables. These two steps were performed in three different situations depending on the variables used for the pattern recognition: property values, spectral data (principal component analysis, PCA) or a combination of both. The results indicated that it was the last situation what offered the best results in both two steps previously described, with the added benefit of outlier detection and removal.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Near-infrared spectroscopy (NIR); Coal analysis; Partial least squares regression (PLS); Hierarchical cluster analysis (HCA); Linear discriminant analysis (LDA)

1. Introduction

The demand for electricity at the lowest possible cost has led to the development for online analysers able to provide fast and representative information about coal feed in a power plant [1–3]. Real time information on coal quality helps efficient management of coal stockpiles and improves plant performance. Thus, traditional methods consisted of laboratory analysis using standard methods are substituted by several techniques, some of them, based on nuclear, capacitance, microwave or ultrasonic energy [4–6]. These techniques are characterized for relying on the inorganic coal constituents to further extrapolate for the calculation of the other coal properties. Really useful for coal batches of the same origin, the usual practice of blending coals to fulfil environmental regulations [7] prevents its widespread use in coal burning power plants.

In order to find an alternative technique, several research works present Fourier transform diffuse reflectance infrared spectroscopy (DRIFT) as a proper technique able to characterise scattering samples with very little preparation and a high signal-to-noise ratio (SNR) in a non-destructive way. These advantages were detailed by Fuller and Griffiths [8] when they used this optical technique to the measurement of powder samples, subsequently applied in the analysis of coal blends [9,10]. Other applications for quantification of functional groups followed [11] and the full characterization of thermal coals was attempted by applying multivariate calibration techniques to the mid infrared spectra [12–16] with relative success.

Advances in data acquisition and processing computer packages allowed the use of DRIFT-near infrared spectroscopy [17,18] to characterise coal samples and, lately, modified DRIFT-near IR spectra have been correlated with coal properties such as moisture, volatile matter, oxygen content, maximum fluidity temperature and solidification temperature, by the application of multivariate calibration [19].

Once demonstrated the adequacy of the technique to the coal analysis, the object of this study is the establishment of a method

* Corresponding author. Tel.: +34 976 73 39 77; fax: +34 976 73 33 18.
E-mail address: jmandres@carbon.icb.csic.es (J.M. Andrés).

able to estimate coal property values (moisture, ash, volatile matter, fixed carbon, heating value, carbon, hydrogen, nitrogen and sulphur) for an unknown sample attending only to its DRIFT-near IR spectral data.

For that, a wide and diverse coal sample set was analysed and correlated to near-infrared spectral variables [20], obtaining relatively good precision. In the same paper, the application of this technique to a more homogeneous set of samples, referred to its origin, showed that most of the properties could be determined with better reproducibility analysis results. Nevertheless, because of the origin of the coal is sometimes unknown, coal samples were grouped into six clusters following the ASTM D388 reference norm [21] for coal classification covering from lignite B to anthracite coal samples. The improvement of the prediction error compared with the obtained for the whole set of samples was proved [22]; however, the clusters could not be differentiated in a reliable way, neither by soft independent modelling of class analogy (SIMCA) nor linear discriminant analysis.

As a consequence of that, there is a major interest in establishing other coal sample groups based on different variables leading to an adequate calibration equation accompanied by a proper differentiation between the clusters when spectral data were used.

In this way, three situations have been studied depending on the independent variables used in the hierarchical clustering process: laboratory coal analysis values, spectroscopic near-IR variables or a combination of both.

Once established the clusters and defined the calibration equations for all the coal properties, coal samples are classified attending only to its spectral data to study the capability of the system to discriminate properly among the groups. An adequate classification process would be the first stage in a feasible estimation of coal properties for a new coal sample.

2. Experimental

2.1. Coal samples

One hundred and forty-two samples of coal from different suppliers were used. Some of them are raw coal samples coming from mines and other from power stations. All of them follow the specifications for pulverized coal burners, attending to a controlled particle size [23], typically: 98% < 300 μm ; 95% < 150 μm ; 75% < 75 μm . They were analysed for moisture, ash, volatile matter (VM), fixed carbon (FC), heating value (HV), carbon, hydrogen, nitrogen and sulphur by the usual ISO/ASTM standard methods [24–29]. The statistical results, minimum, maximum, mean property value and standard deviation of the full set of samples analyses are presented in a previous work [20].

2.2. Data acquisition and multivariate statistical analysis

NIR spectra were acquired by using an ATI Mattson Infinity Series FTIR spectrometer equipped with a tungsten-halogen source, quartz beamsplitter and an InGaAs detector. The sample

was placed in a diffuse reflectance sample cup and the instrument was controlled by the WinFIRST 3.5 software package by Mattson Instruments. Spectra obtained were the result of co-adding 32 scans over the range 1100–2500 nm performed at 1 cm^{-1} of digital resolution.

All the spectra were acquired in absorbance mode, viz. $\log(1/\text{reflectance})$, subsequently pre-processed mathematically in order to reduce extraneous effects such as differences in surface roughness or particle size [30] and to study the correlation between the modified spectral data and the coal properties in the calibration models.

Pre-treatment spectra and multivariate techniques were performed with The Unscrambler v9.1, from CAMO, whereas hierarchical cluster analysis (HCA) and linear discriminant analysis (LDA) were implemented using SPSS v12.0 for Windows.

3. Results and discussion

A wide and varied group of samples were considered for the study, with significant differences both in analytical and spectral values [20]. These differences were the basis of the constitution of homogeneous clusters with the aim of improving the calibration and prediction errors for each coal property after the application of the appropriate calibration model.

In that way, the analysis of a new coal sample would be divided in three steps: spectrum acquisition; sample classification based on the obtained spectrum and coal property determination by applying to the spectrum the calibration method established for that group. The fully establishment of the analytical method would be achieved by studying the following stages: the definition of the groups, the determination of appropriate calibration methods for every property and group previously defined and finally, the development of a successful assignment methodology able to classify each sample with high sensitivity and specificity to its corresponding group.

In this paper, the definition of the groups examines three situations depending on the variables used in the hierarchical clustering process [31]: laboratory coal property values (Situation A), DRIFT-near spectral data (Situation B) or a combination of both (Situation C). These three situations were processed similarly by using the Ward's minimum variance algorithm [32] and the squared Euclidean distance as centroid-linkage measurement to create the clusters.

Independently, each situation presented the corresponding coal distribution dendrogram plot and, consequently, calibration models were applied by partial least squares regression (PLSR). To decide the final number of clusters the reduction of the hydrogen content prediction error [20] determined by the root mean squared error of cross-validation (RMSECV) was followed along the clustering process.

Once defined the clusters, a supervised linear discriminant analysis [33,34] was applied to evaluate the discriminating capacity of the spectral variables in the clusters differentiation. The development of this methodology requires a reduction of the dimensionality as LDA cannot deal with the high number of variables provided by the spectra. Thus, two possibilities were considered for each situation: (i) the scores obtained by apply-

Download English Version:

<https://daneshyari.com/en/article/1245862>

Download Persian Version:

<https://daneshyari.com/article/1245862>

[Daneshyari.com](https://daneshyari.com)