*Clays and Clay Minerals*, Vol. 60, No. 6, 655–665, 2012.

# DETERMINATION OF THE PREDOMINANT MINERALS IN SEDIMENTARY ROCKS BY CHEMOMETRIC ANALYSIS OF INFRARED SPECTRA

MICHAL RITZ[1,*], LENKA VACULÍKOVÁ[2], EVA PLEVOVÁ[2], DALIBOR MATÝSEK[3], AND JIŘÍ MALIŠ[3]

[1] VŠB-Technical University Ostrava, Regional Materials Science and Technology Centre, 17. listopadu 15, 708 33 Ostrava-Poruba, Czech Republic

[2] Institute of Geonics of the AS CR, Institute of clean technologies for mining and utilization of raw materials for energy use, Studentská 1768, 708 00 Ostrava-Poruba, Czech Republic

[3] VŠB-Technical University Ostrava, Institute of clean technologies for mining and utilization of raw materials for energy use, 17. listopadu 15, 708 33 Ostrava-Poruba, Czech Republic

**Abstract**—The objective of the present study was to determine the predominant minerals in sedimentary rocks using Fourier-transform infrared (FTIR) spectroscopy and chemometric analysis. The chemometric analysis was performed on three types of sedimentary rock samples (claystones, clay slates, and sandstones), each with different predominant mineral components. Chemometric models were created to determine the major minerals of the rock samples studied – chlorite, muscovite, albite, and quartz. The FTIR spectra were obtained in transmission mode from pressed pellets of KBr-sample mixtures or by diffuse reflectance from hand-packed mixtures of samples with KBr. Spectral regions measured were 4000–3000 and 1300–400 $cm^{-1}$, which contained important spectral information for the creation of the chemometric models. Principal component analysis was used in the chemometric method, with calibration models being created by a partial least-squares regression method. The mean relative error, standard error of prediction, and relative standard deviation were calculated for the assessment of accuracy, precision, and reproducibility. The value of the mean relative error was 15–20% for most of the calibration models; the value of the standard error of prediction was up to 6 w/w % for most of the calibration models. The values of the standard relative deviation ranged from ~2 to 8% for calibration models based on diffuse reflectance spectra whereas calibration models based on transmission spectra had values of relative standard deviation of ~15–20%.

**Key Words**—Albite, Chemometry, Chlorite, Infrared Spectroscopy, Muscovite, Quartz, Sedimentary Rocks.

## INTRODUCTION

The type and amount of minerals present in rocks have a significant influence on the behavior and properties of the rocks as well as on the whole rock massif. A detailed qualitative and quantitative mineral analysis is therefore essential for the characterization of the properties of the rocks in geological, geochemical, and geomechanical studies.

Several existing conventional analytical methods can be used to examine the mineral composition of rocks: optical microscopy, electron microscopy, X-ray diffraction (XRD), FTIR spectroscopy, Raman spectroscopy, thermal gravimetric/differential thermal analysis (TG/DTA), and bulk chemical analysis (Kodama *et al*., 1989; Chipera and Bish, 2001; Środoń, 2002; Vogt *et al*., 2002). Unfortunately, the determination of minerals (especially clay minerals) in rocks by these methods is rather complicated and often inaccurate. The main analytical difficulties are related to the variable chemical composition and common structural anomalies of clay minerals. The clay minerals occur in the form of mixtures with various ratios of the particular clay minerals.

Current FTIR spectroscopy represents a fast, reliable, and efficient tool for phase analysis. Combined with multivariate calibration by chemometric methods, FTIR spectroscopy is a very useful tool for quantitative phase analysis. The theory of chemometric methods has been described by many authors (*e.g.* Fredericks *et al.*, 1985; Geladi and Kowalski, 1986; Lorber and Kowalski, 1988; Martens and Naes, 1989). Numerical methods, such as multiple linear regression (MLR), principal component regression (PCR), or partial least-squares (PLS) regression, are being used increasingly to circumvent the problems posed by the presence of interferences, spectral overlap, or major matrix effects (Luis *et al*., 2004). Application of the chemometric method to FTIR spectroscopic analysis has been described by many authors (*e.g.* Fuller *et al.*, 1988; Haaland and Thomas, 1988; Iñón *et al.*, 2003; Armenta *et al.*, 2007; Breen *et al.*, 2008). In recent years, chemometric methods have also been applied to other analytical techniques such as voltammetric or chromatographic methods (Moneeb, 2006; Al-Degs *et al.*, 2008; Wagieh *et al.*, 2010; Zapata-Urzúa *et al.*, 2010).

The spectroscopic applications of multivariate calibration tend to use the full spectrum. This approach

* E-mail address of corresponding author:
michal.ritz@vsb.cz
DOI: 10.1346/CCMN.2012.0600609

provides a more accurate description of the model than a single measurement at a specific wavelength or wavenumber. However, full-spectrum applications pose some problems: (1) some of the information gathered can be redundant; and (2) the measured signal for some wavelengths may be noisy or nonlinear (Luis *et al.*, 2004). The most useful methods for a chemometric approach in FTIR spectroscopy are principal component regression (PCR) and partial least-squares (PLS) regression (Hasegawa, 2002). An important feature of PCR is the fact that only spectral information is used for the generation of basis factors. This is not a problem when the concentration information is absolutely accurate. In practice, however, the concentration matrix contains error or noise. Another problem of PCR is that colinearity of absorbance data will make the calibration unstable (Hasegawa, 2002). To solve this problem, a more stable chemometric method is required, a method which takes both (absorbance and concentration) matrices into account simultaneously. This method is PLS regression in which absorbance and concentration matrices are used in a complementary fashion to yield a stable calibration.

PLS regression works with two matrices, **X** and **Y**. The **X** matrix contains the independent X variables, *e.g.* the spectral data (in absorbance units). The **Y** matrix consists of the dependent Y variables, *e.g.* quantitative (concentration) data. The NIPALS (Nonlinear Iterative Partial Least Squares) algorithm is that used most often in the creation of PLS models which can be considered as consisting of outer relations (**X** matrix and **Y** matrix individually) and an inner relation (linking both matrices) (Geladi and Kowalski, 1986). The outer relations for the **X** and **Y** matrices can be written as follows:

$$\mathbf{X} = \Sigma_h \mathbf{t}_h \mathbf{p}_h + \mathbf{E}_X = \mathbf{TP} + \mathbf{E} \quad (1)$$

$$\mathbf{Y} = \Sigma_h \mathbf{u}_h \mathbf{q}_h + \mathbf{E}_Y = \mathbf{UQ} + \mathbf{F} \quad (2)$$

where $\mathbf{t}_h$ and $\mathbf{u}_h$ are score vectors, $\mathbf{p}_h$ and $\mathbf{q}_h$ are loading vectors, **T** and **U** are score matrices, **P** and **Q** are loading matrices, and **E** and **F** are matrices of residuals. A graphical representation of equations 1 and 2 (Figure 1) was described by Geladi and Kowalski (1986) in which $n$ is the number of samples, $m$ is the number of independent variables (*e.g.* absorbance values), $a$ is the number of factors, and $p$ is the number of dependent variables (concentration values). One attempts to describe the **Y** matrix as well as possible and hence to make the **F** matrix as small as possible, and, at the same time, develop the inner relation between the **X** and **Y** matrices (Geladi and Kowalski, 1986). Because **X** and **Y** are summations of the inner product of the loading vector ($\mathbf{p}_h$ and $\mathbf{q}_h$) and a score vector ($\mathbf{t}_h$ or $\mathbf{u}_h$), decomposition to yield each inner product is performed step-by-step with an iteration process. Although the spectra scores and concentration scores are defined in the individual space, their changes after normalization should be identical to each other, if the absorbance change is linear with concentration change (Hasegawa, 2002). Thus, making the maximum correlation between $\mathbf{t}_h$ and $\mathbf{u}_h$ is an important starting concept for PLS.

The PLS regression can be categorized into two procedures: PLS1 (sometimes referred to as standard PLS) and PLS2 (sometimes referred to as global PLS). PLS1 employs information from only one chemical constituent to make the calibration. Thus, only one column concentration vector (**y**) taken from a concentration matrix (**Y**) is used for the calibration in PLS1. The basic procedure in this calibration is the same as in PLS2, except vector $\mathbf{q}_h$ becomes a scalar ($q_h$). The scalar ($q_h$) produces a further difference during the calibration (Hasegawa, 2002). Note that the slopes of the regression of $\mathbf{t}_h$ and $\mathbf{u}_h$ are both equal to 1. As a result, PLS1 is summarized by the next two equations:

$$\mathbf{X} = \Sigma_h \mathbf{t}_h \mathbf{p}_h + \mathbf{E} \quad (3)$$

$$\mathbf{y} = \Sigma_h \mathbf{t}_h \mathbf{q}_h + \mathbf{F} \quad (4)$$

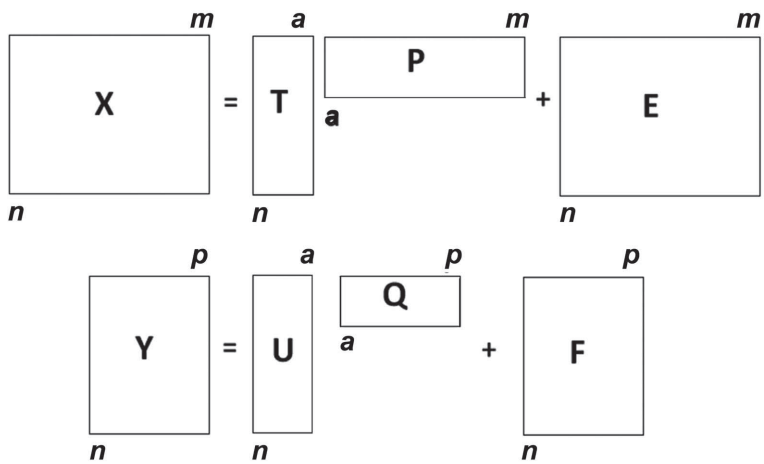


Figure 1. Graphical representation of PLS (after Geladi and Kowalski, 1986).

PLS1 clearly uses the vector $\mathbf{t}_h$ as a common score vector for both spectra and concentration modeling. Therefore, the inner relation is not necessary. The concept of the common score is particularly reasonable when the intrinsic concept of PLS, that the correlation between $\mathbf{t}_h$ and $\mathbf{u}_h$ should be maximized, is taken into account. PLS2 uses two or more chemical constituents simultaneously and comprises relatively complicated calculation procedures. PLS2 regression is carried out using information from more than one constituent concentration. A disadvantage of PLS2 is that some constituents are calibrated simultaneously, even if their appropriate numbers of factors are different from each other (Martens and Naes, 1989). This sometimes causes a calibration error for the component with a different number of basis factors.

PLS modeling is complex. Detailed description is beyond the scope of this paper, but its mathematical algorithm can be found in many publications (*e.g.* Geladi and Kowalski, 1986; Wold *et al.*, 2001; Hasegawa, 2002).

The aim of the present study was to determine the main minerals (chlorite, muscovite, albite, and quartz) in sedimentary rocks by FTIR spectroscopy combined with partial least-squares (PLS1) regression.

## EXPERIMENTAL

### *Samples*

One hundred and thirty-five samples of sedimentary rocks (claystones, clay slates, and sandstones) were used here. One hundred and twenty-five samples (S1−S125) were used as a calibration set (Table 1) and a further ten (CS1−CS10) were used as control samples. In the calibration set were 32 samples of claystones (S1−S32), 46 samples of clay slates (S33−S78), and 47 samples of sandstones (S79−S125). Three control samples (CS1−CS3) were claystones, three (CS4−CS6) were clay slates, and four (CS7−CS10) were sandstones. All the samples were obtained from the collection at VŠB-Technical University, Ostrava, in the Czech Republic.

The claystone samples were taken from several lower-to-middle Cretaceous strata belonging to the Silesian unit of the Moravian-Silesian Beskydy Mountains (Hradiště and Lhota formations, Godula and Mazak formations, Štramberk area). The samples of clay slates came from the Kyjovice layers that are stratigraphically adherent to the Lower Carboniferous period of the Moravian-Silesian area. Sandstone samples were taken from several Upper Carboniferous layers in the Czech part of the Upper-Silesian coal basin.

Every sample of tested rock was ground to fine powder using a Pulverisette 5 instrument (Fritsch GmbH, Germany). Powder samples were homogenized by careful stirring.

The XRD patterns and FTIR spectra of the rock samples indicate similar mineralogical compositions. Chlorite, muscovite, quartz, and albite are the most common components of all of the samples. For this reason the research was focused on quantitative analysis of those minerals. The clay minerals were represented mainly by chlorite with smaller amounts of kaolinite and a small amount of illite in some samples.

Albite was observed in most of the rock samples, usually in association with quartz and muscovite. Other types of feldspar (orthoclase, microcline) appear rarely and in small amounts only. Besides the above-mentioned minerals, carbonates (calcite, siderite, and ankerite) were also detected in some rock samples. The presence of pyrite, hematite, and rutile in trace amounts was confirmed by XRD analysis. Peak positions in the FTIR spectra of the rock samples were assigned to vibrational modes of the various constituents according to the literature (Couty and Velde, 1986; Russell *et al.*, 1994; Madejová and Komadel, 2001; Vaculíková and Plevová, 2005). More detailed information about mineral identification on the basis FTIR spectra and chemometric methods was given by Ritz *et al.* (2010).

### *Powder XRD analysis*

The amounts of the various minerals present in the samples of sedimentary rocks were determined by X-ray powder diffraction (XRD) using the Rietveld method (Rietveld, 1969), which is a quantitative technique for crystal-structure analysis from powder diffraction data (*e.g.* Brindley, 1980; Bish and Howard, 1988; Hillier, 2000; Chipera and Bish, 2001). The theoretical XRD pattern was calculated on the basis of structural data (*e.g.* crystal symmetry, unit-cell parameters, atomic coordinates, and occupancy) of the minerals that were present. The theoretical XRD pattern was subsequently compared with the measured diffraction pattern using multidimensional regression.

Ground (to ~5 μm) and homogenized samples were placed in glass capillaries. Powder diffraction measurements were carried out on a fully-automated diffractometer ID3003 (Rich Seifert-FPM, Germany) under the following conditions: CoKα radiation, Fe filter, 2θ-θ goniometer geometry, step mode with 0.05º2θ steps, 3 s measurement time per step, and with digital processing of the resultant data. For measurement and qualitative evaluation, the software packages *RayfleX* and *RayfleX Autoquan* (GE Sensing & Inspection Technologies, USA) were used.

### *FTIR measurements*

For FTIR spectroscopic analysis, samples were mixed with spectral grade KBr prior to analysis. For diffuse reflectance (DRIFT) spectral analysis, ~5−10 mg of sample was ground with approximately 400 mg of dried KBr. This mixture was placed as a manually compacted powder into the cup of the DRIFT accessory of a Nexus

Table 1. List of calibration set of minerals.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| S1 | 8.2 | 40.3 | 3.0 | 44.9 | S64 | 13.9 | 33.8 | 16.4 | 35.9 |
| S2 | 8.2 | 42.1 | 3.3 | 39.5 | S65 | 10.8 | 9.9 | 27.0 | 52.3 |
| S3 | 8.9 | 39.0 | 2.6 | 46.2 | S66 | 9.9 | 42.0 | 13.5 | 34.3 |
| S4 | 3.5 | 38.2 | 3.1 | 39.1 | S67 | 17.3 | 27.3 | 17.3 | 32.7 |
| S5 | 5.4 | 8.0 | 3.5 | 56.4 | S68 | 17.6 | 34.7 | 13.2 | 34.5 |
| S6 | 9.7 | 23.5 | 2.5 | 64.3 | S69 | 21.0 | 44.9 | 3.0 | 27.7 |
| S7 | 10.5 | 35.1 | 1.8 | 52.6 | S70 | 22.5 | 45.9 | 3.0 | 25.9 |
| S8 | 7.9 | 48.8 | 9.2 | 34.1 | S71 | 18.9 | 32.5 | 15.3 | 33.4 |
| S9 | 6.2 | 65.5 | 8.4 | 20.0 | S72 | 16.9 | 23.6 | 13.0 | 40.6 |
| S10 | 7.8 | 53.3 | 4.7 | 29.4 | S73 | 21.6 | 31.9 | 12.6 | 34.0 |
| S11 | 8.8 | 56.0 | 5.6 | 36.1 | S74 | 19.5 | 35.0 | 16.2 | 25.0 |
| S12 | 5.1 | 52.1 | 4.0 | 33.2 | S75 | 14.7 | 31.8 | 18.3 | 31.0 |
| S13 | <1.0 | 2.2 | <1.0 | 48.7 | S76 | 16.6 | 31.3 | 16.7 | 33.7 |
| S14 | 3.8 | 9.7 | <1.0 | 57.0 | S77 | 20.5 | 30.2 | 17.0 | 31.3 |
| S15 | <1.0 | 16.5 | <1.0 | 30.6 | S78 | 19.0 | 29.7 | 15.8 | 35.5 |
| S16 | 3.4 | 17.7 | 1.5 | 19.1 | S79 | 3.1 | 10.3 | 4.4 | 76.4 |
| S17 | <1.0 | 1.6 | <1.0 | 31.4 | S80 | <1.0 | <1.0 | 7.9 | 63.4 |
| S18 | 5.0 | 12.4 | <1.0 | 15.7 | S81 | 2.5 | 7.1 | 8.6 | 73.0 |
| S19 | 2.8 | 17.5 | <1.0 | 37.2 | S82 | 9.4 | 7.9 | 8.0 | 68.3 |
| S20 | 3.8 | 10.5 | <1.0 | 13.5 | S83 | 8.0 | 17.2 | 4.6 | 60.6 |
| S21 | 5.5 | 14.4 | <1.0 | 27.3 | S84 | <1.0 | 13.9 | 1.9 | 69.5 |
| S22 | 5.8 | 14.9 | 2.3 | 41.6 | S85 | 3.0 | 9.4 | 14.6 | 73.1 |
| S23 | <1.0 | 10.9 | <1.0 | 23.1 | S86 | 2.1 | 13.3 | 18.5 | 66.1 |
| S24 | <1.0 | 9.1 | <1.0 | 23.5 | S87 | 1.8 | 4.8 | 6.6 | 86.1 |
| S25 | 6.0 | 10.1 | 1.8 | 22.1 | S88 | <1.0 | 36.7 | <1.0 | 54.3 |
| S26 | <1.0 | 19.0 | 2.6 | 26.6 | S89 | 6.2 | 28.8 | 6.7 | 49.7 |
| S27 | <1.0 | 10.7 | 1.1 | 25.2 | S90 | 3.4 | 5.9 | 13.4 | 60.8 |
| S28 | <1.0 | 28.4 | <1.0 | 28.4 | S91 | 5.0 | 11.2 | 21.4 | 57.8 |
| S29 | <1.0 | 13.4 | <1.0 | 34.5 | S92 | 2.0 | 6.7 | 13.8 | 74.4 |
| S30 | 1.6 | 22.6 | <1.0 | 61.4 | S93 | 1.4 | 6.9 | 14.9 | 75.6 |
| S31 | <1.0 | 9.3 | 1.7 | 34.0 | S94 | 1.8 | 8.2 | 14.1 | 75.4 |
| S32 | 5.4 | 15.4 | <1.0 | 23.9 | S95 | 2.9 | 10.8 | 14.6 | 69.4 |
| S33 | 22.2 | 39.5 | 18.9 | 19.5 | S96 | 11.7 | 36.2 | 12.4 | 35.5 |
| S34 | 17.6 | 32.4 | 19.8 | 30.1 | S97 | 3.2 | 5.4 | <1.0 | 75.8 |
| S35 | 15.6 | 27.8 | 16.8 | 39.8 | S98 | 1.9 | 6.2 | <1.0 | 81.2 |
| S36 | 9.1 | 21.6 | 15.6 | 53.8 | S99 | 2.6 | 6.1 | <1.0 | 82.5 |
| S37 | <1.0 | <1.0 | 1.8 | 69.6 | S100 | 2.2 | 11.4 | <1.0 | 71.9 |
| S38 | <1.0 | 34.1 | <1.0 | 29.2 | S101 | 4.6 | 10.2 | <1.0 | 78.7 |
| S39 | <1.0 | 24.1 | <1.0 | 6.5 | S102 | 2.0 | 10.0 | <1.0 | 82.1 |
| S40 | 10.4 | 23.2 | 18.0 | 48.4 | S103 | 8.0 | 10.6 | <1.0 | 72.5 |
| S41 | 28.1 | 41.6 | 9.1 | 18.9 | S104 | 2.6 | 6.9 | <1.0 | 82.5 |
| S42 | 17.2 | 31.1 | 17.9 | 26.3 | S105 | 3.6 | 8.1 | <1.0 | 82.3 |
| S43 | 17.2 | 53.8 | 3.7 | 25.2 | S106 | 4.4 | 7.2 | <1.0 | 83.9 |
| S44 | 22.7 | 25.1 | 21.4 | 30.3 | S107 | 5.1 | 23.8 | <1.0 | 38.2 |
| S45 | 17.6 | 37.2 | 8.9 | 22.0 | S108 | 4.3 | 19.3 | <1.0 | 70.3 |
| S46 | <1.0 | 71.9 | 2.9 | 5.0 | S109 | 3.3 | 6.3 | <1.0 | 88.1 |
| S47 | 17.0 | 29.2 | 14.6 | 25.8 | S110 | 2.3 | 7.1 | <1.0 | 89.6 |
| S48 | 11.6 | 47.6 | 3.8 | 36.9 | S111 | 3.9 | 19.3 | <1.0 | 75.4 |
| S49 | 29.6 | 43.3 | 7.0 | 20.2 | S112 | 6.5 | 10.6 | 5.1 | 44.1 |
| S50 | 19.2 | 33.7 | 19.0 | 28.1 | S113 | 13.4 | 9.2 | 15.1 | 50.3 |
| S51 | 54.6 | 12.5 | 2.5 | 27.4 | S114 | 1.7 | 4.5 | 8.0 | 81.5 |
| S52 | 3.3 | 38.1 | 11.2 | 43.0 | S115 | 4.2 | 7.8 | 17.1 | 67.5 |
| S53 | 9.6 | 18.6 | 41.5 | 30.4 | S116 | 6.9 | 14.3 | 6.5 | 24.3 |
| S54 | 12.3 | 25.0 | 2.3 | 49.5 | S117 | 8.0 | 25.4 | 14.3 | 52.3 |
| S55 | 20.2 | 33.3 | 13.4 | 32.5 | S118 | 5.7 | 7.0 | 23.0 | 60.1 |
| S56 | 18.0 | 25.4 | 14.7 | 36.9 | S119 | 2.5 | 4.7 | 12.8 | 62.9 |
| S57 | 22.8 | 43.9 | 10.6 | 21.1 | S120 | 10.8 | 30.4 | 13.2 | 35.5 |
| S58 | 25.5 | 42.5 | 2.7 | 27.5 | S121 | 3.3 | 10.5 | 2.3 | 77.4 |
| S59 | 19.8 | 50.4 | 1.4 | 29.4 | S122 | 4.6 | 8.8 | 21.4 | 64.1 |
| S60 | 19.6 | 30.4 | 13.2 | 33.1 | S123 | 3.9 | 7.0 | 16.0 | 73.1 |
| S61 | 20.9 | 31.1 | 12.9 | 35.2 | S124 | 4.4 | 8.2 | 1.1 | 86.3 |
| S62 | 23.4 | 36.5 | 12.3 | 27.8 | S125 | 2.0 | 12.0 | 0.6 | 76.6 |
| S63 | 22.8 | 34.0 | 10.9 | 32.4 | | | | | |

470 spectrometer (ThermoScientific, USA) with a deuterated TriGlycine sulfate (DTGS) detector. The measurement parameters were as follows: spectral region 4000–400 $cm^{-1}$; spectral resolution 8 $cm^{-1}$; 128 scans; Happ-Genzel apodization. The spectra were measured in Kubelka-Munk units, and a two-points linear baseline was used. Freshly dried KBr was used for the background measurement. Every sample was prepared and measured 3–5 times. The mean FTIR spectrum of every sample was calculated and subsequently used for the creation of chemometric models.

For transmission FTIR spectroscopic analysis, 0.5 mg of sample was ground with 200 mg of dried KBr, then pressed into a pellet for 30 s under vacuum. The FTIR spectra were collected using an Avatar 320 FTIR spectrometer (ThermoScientific, USA) with DTGS detector. The measurement parameters were as follows: spectral region 4000–400 $cm^{-1}$; spectral resolution 8 $cm^{-1}$; 64 scans; Happ-Genzel apodization. The spectra were measured in absorbance units, and a two-point linear baseline was used. An empty sample compartment was used for background measurement. Every sample was prepared and measured only once. The spectrum of a pure KBr pellet was subtracted from sample spectra. The background-corrected FTIR spectra were used subsequently for the creation of chemometric models.

### *Chemometric analysis*

Chemometric analysis was performed using *The Unscrambler* 9.7 software package (CAMO Software AS, Norway). The PCA and PLS regressions were used as representative chemometric methods. PCA was used for preparatory data analysis; outlier spectra were detected using score plot and influence plot, and important spectral regions were specified by line loading plot in PCA. The PLS1 method was employed to create chemometric models for the determination of minerals (chlorite, muscovite, albite, and quartz) in slate.

Multiplicative scatter correction (MSC) was performed for DRIFT spectra. Scatter effects may be caused by an optical pathlength that varies and by pressure variations in DRIFT spectra. These effects generally consist of both the so-called multiplicative effect as well as an additive effect. The MSC method was used to compensate for both multiplicative and additive effects.

The transmission and DRIFT IR spectra of samples from the calibration set were used to build PLS models and simultaneously for their validation. The number of optimal PLS parameters was determined by statistical comparison of PRESS (predicted residual error sum square) values (*e.g.* Esbensen, 2006) as a function of numbers of factors. The model validation was performed by the cross-validation (CV) method. The segmented cross-validation was performed as the validation method for calibration models. The size of the cross-validation segment was two samples.

## RESULTS AND DISCUSSION

### *Preparatory data analysis*

The FTIR spectra of all samples in the calibration set were transformed to data matrices. One data matrix was prepared from transmission FTIR spectra of KBr pellets; the next data matrix was prepared from DRIFT spectra. Principal component analysis (PCA) was performed for detecting the outlier spectra in the calibration set and for selection of the important spectral regions. The following plots were prepared: the score plot of the first two principal components, the influence plot of the first three principal components, and the loading plot for the first principal component.

The score plots and the influence plots (Figure 2) were used to detect outlier spectra and to check data homogeneity. The data used were homogenous; no outlier spectra were found in either of the data matrices.

The line-loading plots were used for the selection of important spectral regions. This type of loading plot looks like a spectrum. The important spectral regions, therefore, have the character of spectral bands (Figure 3). The important spectral regions determined by loading plots of both data matrices were 4000–3000 $cm^{-1}$ and 1300–400 $cm^{-1}$. The spectral bands present at the spectral region 4000–3000 $cm^{-1}$ belonged to the stretching band of structural hydroxyl groups (3630 $cm^{-1}$) and to the stretching band of water (3310 $cm^{-1}$). The most significant spectral bands in the region 1300–400 $cm^{-1}$ could be assigned to the following vibrations: Si–O stretching vibrations of quartz (1160 and 1090 $cm^{-1}$), Si–O stretching vibration (1030 $cm^{-1}$), deformation vibration of Al–Fe–OH (875 $cm^{-1}$), Si–O stretching vibrations of quartz (800 and 780 $cm^{-1}$), deformation vibration of Si–O (695 $cm^{-1}$), and the deformation vibrations of Al–O–Si and Si–O–Si (530 and 475 $cm^{-1}$, respectively). The assignment of spectral bands was according to Madejová and Komadel (2001).

### *Creation of PLS1 models*

The data matrices mentioned previously and the contents of the minerals in the calibration samples (Table 1) were used for the creation of the PLS models. Half of the detection limit (0.5 w/w %) was used if the value was below the detection limit (1.0 w/w %). The PLS1 method was used: a separate calibration model was created for determination of each mineral. The predicted *vs*. measured plot (Figure 4) was created for every PLS model. The important parameters (Table 2) of the PLS models are: slope, offset, $R^2$ (multiple correlation coefficient), $Q2$ (cross-validated multiple correlation coefficient), *RMSEC* (root mean squared error of calibration), *RMSECV* (root mean squared error of cross-validation), $Y$ variance explained, and number of PLS factors. Slope, offset, and multiple correlation coefficients are regression parameters that reflect the

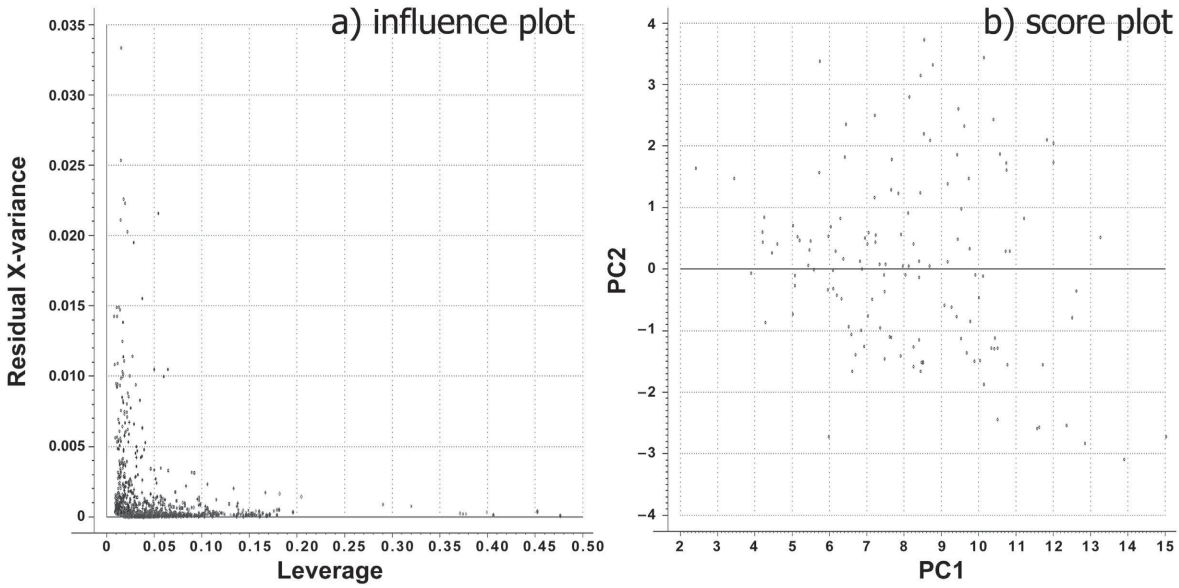


Figure 2. Influence plot (a) and score plot (b) of PCA of DRIFT spectra.

linear dependence between data predicted by the PLS model and data from the reference method (XRD analysis) or cross-validated regression, respectively. The calibration error of the PLS model was expressed by *RMSEC*:

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{n}(c_{i,\mathrm{cal,pred}} - c_{i,\mathrm{refer}})^2}{n}} \qquad (5)$$

where $c_{i,\mathrm{cal,pred}}$ is the value of the mineral content of the $i^{\mathrm{th}}$ calibration sample predicted from the PLS model, $c_{i,\mathrm{refer}}$ is the value of the mineral content of the $i^{\mathrm{th}}$ calibration sample obtained by the reference method (XRD analysis), and $n$ is the number of samples in the calibration set. The value of the validation error of the model was expressed by *RMSECV*, analogous to *RMSEC*:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n}(c_{i,\mathrm{val,pred}} - c_{i,\mathrm{refer}})^2}{n}} \qquad (6)$$

where $c_{i,\mathrm{val,pred}}$ is the value of the mineral content of the $i^{\mathrm{th}}$ validation sample predicted by the PLS model. '*Y*

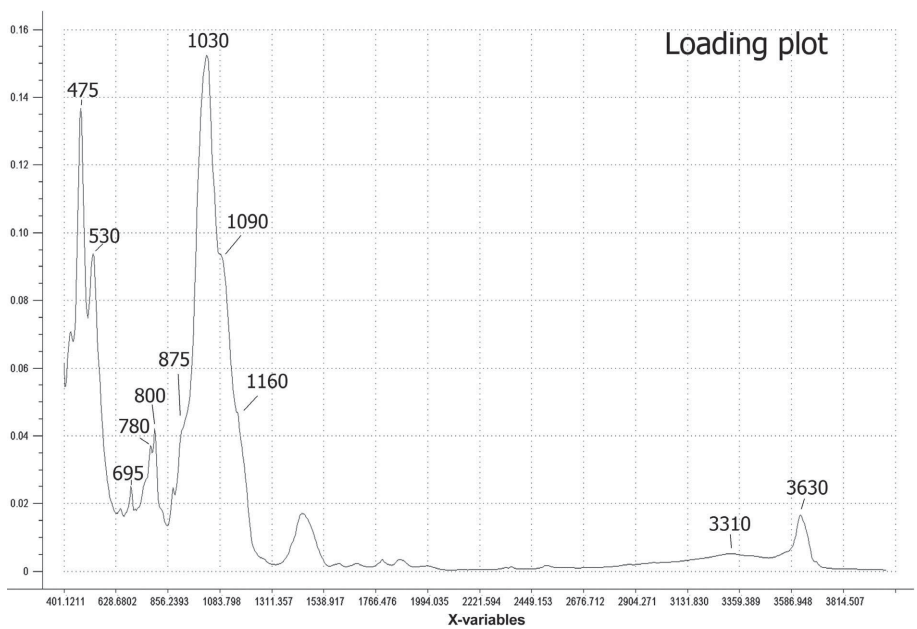


Figure 3. Line loading plot of PCA of transmission spectra.

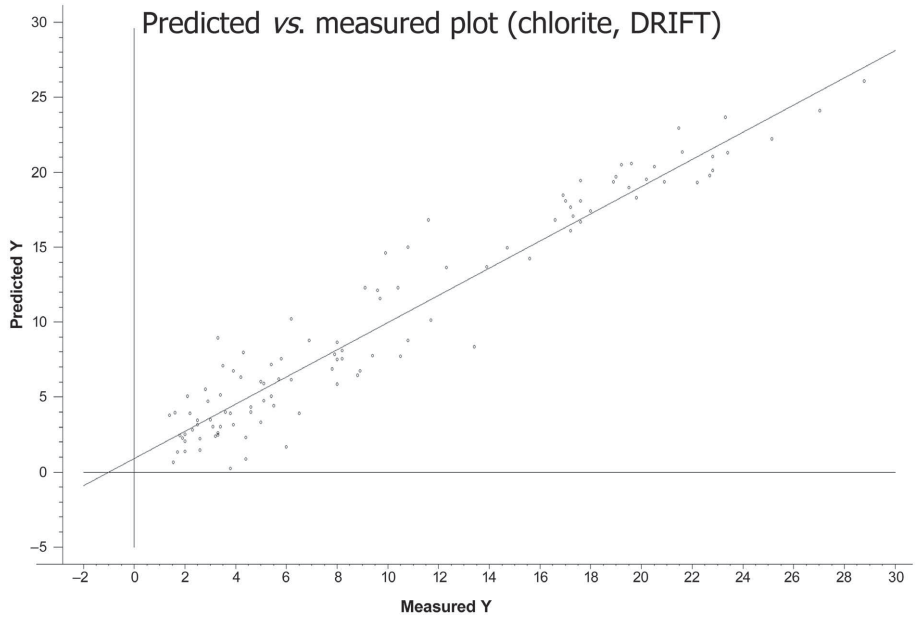


Figure 4. Predicted *vs*. measured plot (chlorite PLS model; DRIFT spectra).

variance explained' is the parameter that specifies the percentage of variability of the system which is described by the number of PLS factors used.

The regression parameters obtained showed acceptable values. The values of slope ranged from 0.7877 to 0.9287. The best value of slope had a model for chlorite (DRIFT) and the worst had a model for muscovite (DRIFT). The multiple regression coefficients of all models reached very good values. Calibration and validation errors for models of chlorite and albite were up to 4 w/w %; calibration and validation errors of the remaining models were ~2−3 times greater.

*Analysis of control samples (accuracy and precision)*

The predictive ability of the PLS models was tested by analysis of the ten control samples (Table 3). They were prepared as KBr pellets for FTIR transmission analysis, and were also analyzed by the DRIFT technique (each spectrum was the average of three independent DRIFT measurements). All of these spectra were used for prediction of the mineral content by the PLS models that were created.

Predictions of the mineral content in control samples were tested for statistical significance with the reference values (results of XRD analysis) of the control samples. The testing methods included the F-test, t-test (Student's test), and paired comparison (Meloun and Militký, 2004). All of these methods showed statistical significance between the predicted and reference values for the control samples.

The predictive ability of chemometric models can be described using several validation diagnostics. The parameters used for validation (Table 4) were bias, standard error of prediction (SEP), and mean relative error (RE). Bias and SEP parameters were used according to Esbensen (2006); RE parameters were

Table 2. Parameters of PLS models.

| Mineral | Method | Slope | Offset | $R^2$ | $Q^2$ | *RMSEC* (% w/w) | *RMSECV* (% w/w) | *Y* variance explained | Number of PLS factors |
|---|---|---|---|---|---|---|---|---|---|
| Chlorite | DRIFT | 0.9287 | 0.74 | 0.9712 | 0.9619 | 2.28 | 2.58 | 93.0% | 3 |
| | KBr pellet | 0.8357 | 1.61 | 0.9285 | 0.9099 | 3.58 | 4.01 | 91.5% | 3 |
| Muscovite | DRIFT | 0.7877 | 5.00 | 0.9385 | 0.9234 | 6.92 | 7.72 | 91.5% | 3 |
| | KBr pellet | 0.8396 | 3.18 | 0.9355 | 0.9262 | 7.05 | 8.83 | 88.5% | 3 |
| Albite | DRIFT | 0.8337 | 1.73 | 0.9456 | 0.9292 | 2.97 | 3.59 | 89.0% | 3 |
| | KBr pellet | 0.8379 | 1.42 | 0.9293 | 0.9107 | 3.39 | 4.24 | 88.5% | 3 |
| Quartz | DRIFT | 0.8531 | 6.80 | 0.9740 | 0.9663 | 8.23 | 9.38 | 90.0% | 2 |
| | KBr pellet | 0.8579 | 4.23 | 0.9661 | 0.9564 | 10.20 | 12.03 | 90.0% | 2 |

Table 3. List of control samples and results of their analysis.

| Sample | Chlorite (w/w %) | | | Muscovite (w/w %) | | | Albite (w/w %) | | | Quartz (w/w %) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | XRD | – PLS – | | XRD | – PLS – | | XRD | – PLS – | | XRD | – PLS – | |
| | | DRIFT | Pellets | | DRIFT | Pellets | | DRIFT | Pellets | | DRIFT | Pellets |
| CS01 | 12.1 | 12.4 | 11.7 | 20.7 | 22.4 | 26.3 | 20.1 | 12.2 | 13.5 | 48.1 | 47.8 | 45.7 |
| CS02 | 5.5 | 7.0 | 4.4 | 12.1 | 12.2 | 15.8 | 6.7 | 4.0 | 6.6 | 30.0 | 33.2 | 43.8 |
| CS03 | 2.3 | 2.8 | 0.8 | 8.2 | 6.4 | 7.9 | 17.4 | 11.1 | 12.6 | 38.0 | 37.5 | 36.7 |
| CS04 | 3.2 | 5.1 | 3.7 | 36.7 | 30.2 | 39.7 | 22.6 | 17.4 | 18.6 | 69.2 | 80.3 | 86.5 |
| CS05 | 5.0 | 6.5 | 7.9 | 37.4 | 37.1 | 30.6 | 4.5 | 3.5 | 4.9 | 62.9 | 62.9 | 55.0 |
| CS06 | 14.8 | 8.6 | 10.8 | 40.8 | 37.6 | 32.9 | 10.7 | 10.9 | 12.9 | 52.9 | 42.0 | 38.2 |
| CS07 | 19.1 | 20.1 | 20.5 | 36.5 | 37.2 | 35.9 | 14.3 | 13.2 | 13.9 | 58.3 | 46.1 | 54.6 |
| CS08 | 24.3 | 19.1 | 20.4 | 30.9 | 36.8 | 36.2 | 5.9 | 3.6 | 7.5 | 48.9 | 42.2 | 43.9 |
| CS09 | 11.7 | 8.7 | 14.0 | 30.0 | 42.5 | 29.5 | 14.5 | 13.7 | 13.9 | 29.4 | 26.7 | 36.6 |
| CS10 | 21.0 | 20.1 | 14.6 | 27.2 | 32.8 | 22.6 | 12.0 | 11.6 | 9.8 | 39.8 | 37.7 | 36.7 |

created for the purpose of this study:

$$\text{bias} = \frac{\sum_{i=1}^{n}(c_{i,\text{pred}} - c_{i,\text{refer}})}{n} \quad (7)$$

$$\text{SEP} = \sqrt{\frac{\sum_{i=1}^{n}(c_{i,\text{pred}} - c_{i,\text{refer}})^2}{n}} \quad (8)$$

$$\text{RE} = \frac{\sum_{i=1}^{n}\left(\frac{|c_{i,\text{pred}} - c_{i,\text{refer}}|}{c_{i,\text{refer}}}\right)}{n} \cdot 100 \quad (9)$$

where $c_{i,\text{pred}}$ is the value of the mineral content of the $i^{\text{th}}$ control sample predicted from the PLS model. Note that the precision and accuracy of results obtained from PLS models cannot be better than the precision and accuracy of the reference method (in the present case, the Rietveld method of XRD analysis).

Bias represents the average difference between the predicted values and the reference values for control samples. Bias is a commonly used measure of the accuracy of a chemometric model. Bias is also used to check any systematic differences observed between the average values of the control samples and the validation samples (Esbensen, 2006). Another way to express the accuracy of a chemometric model is the mean relative error. The standard error of prediction (SEP) expresses the precision of the predicted results.

The bias of most of the PLS models was negative. Only two models had positive values for bias (muscovite and quartz; both for DRIFT spectra). The absolute values of bias ranged from 0.1 to 2.8. The values of RE ranged around 20%; only DRIFT models for the prediction of quartz had values of RE ~10% (see Table 3). For most of the predicted minerals, parameter RE showed similar values between models obtained from different spectral techniques. Only models for quartz showed differences between the RE parameters from the transmission technique and RE parameters from the DRIFT spectra. The RE values (the parameter representing the accuracy of the method) for all of the models created were very similar to the values of accuracy for results of XRD analyses cited in the literature (*e.g.* Moore and Reynolds, 1997). (In current quantitative phase analysis of rocks, acceptable RE values are ~20−25%). Values of the RE parameters of control samples achieved by the reference quantitative method (XRD analysis) were between 10 and 20%. The best values for the precision parameter (SEP) were PLS models for the prediction of chlorite (3.2 w/w %); a model created for the prediction of the amount of albite from transmission spectra showed a very similar value (3.3 w/w %). The remaining models showed values of up to 7 w/w %. All models had very similar values for SEP from DRIFT and transmission spectra. The precision of analysis of control samples by the reference quantitative method (XRD analysis) in the present study was up to 10 w/w %.

*Reproducibility*

Two control samples (CS02 and CS07) were used to estimate the precision of the PLS models that were created. Each of the two above-mentioned control

Table 4. Parameters of predicted ability of PLS models.

| Parameter | — Chlorite — | | — Muscovite — | | — Albite — | | — Quartz — | |
|---|---|---|---|---|---|---|---|---|
| | DRIFT | Pellets | DRIFT | Pellets | DRIFT | Pellets | DRIFT | Pellets |
| bias (w/w %) | −1.0 | −1.0 | 1.7 | −0.7 | −2.8 | −1.5 | −2.1 | 0.1 |
| SEP (w/w %) | 3.2 | 3.2 | 5.7 | 5.5 | 4.0 | 3.3 | 7.0 | 6.9 |
| RE (%) | 24.9 | 26.3 | 16.2 | 14.6 | 21.7 | 18.1 | 9.8 | 16.9 |

samples had different levels of mineral content. Sample CS02 contained less of the minerals determined than did sample CS07. Ten KBr pellets were prepared from each sample and their FTIR spectra were measured over a period of 1 month. Ten FTIR spectra were also obtained by the DRIFT technique; each spectrum was calculated as the mean of three independent DRIFT measurements (including homogenization and grinding with KBr) during the same period. All of these spectra were used in the PLS models to predict the amounts of minerals. The reproducibility was expressed by the relative standard deviation (RSD), values of which (Table 5) were calculated from the results that were obtained:

$$\mathrm{RSD} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(x_P - x_i)^2}{n-1}}}{x_P} \cdot 100 \qquad (10)$$

where $x_i$ is the predicted value of the mineral content of the $i^{\mathrm{th}}$ analysis of reproducibility and $x_P$ is the mean value of the predicted mineral content.

As expected, poorer reproducibility was obtained from samples with small mineral contents.

A comparison of reproducibility of chemometric models obtained by the transmission *vs*. the DRIFT techniques revealed that the DRIFT technique yielded significantly better reproducibility even though DRIFT is not primarily intended for quantitative applications. For the transmission model, the RSD values were ~15% for sample CS07 and ~20% for sample CS02. For DRIFT spectra models, values of the RSD for sample CS02 were within the range 4–10%; for sample CS07 the values were within the range 2–6%. The differences in reproducibility were probably caused by the sample mass differences in the transmission (0.5 mg) and DRIFT (5–10 mg) preparations. The relatively small amount used in the transmission method is at the limit of accuracy of the analytical balance, thus elevating the weighing error. The use of larger sample weights could lead to values of absorbance which were too high and these spectral bands cannot be used for chemometric analysis.

## CONCLUSIONS

Principal component analysis (PCA) and especially PLS regression were used here as chemometric methods. PCA was used to detect outliers and to select important spectral regions. No outliers were detected. The infrared regions 4000–3000 $\mathrm{cm}^{-1}$ and 1300–400 $\mathrm{cm}^{-1}$ were selected for the creation of calibration models by the PLS regression technique. All PLS models that were created also showed a significant correlation between real and predicted mineral contents. For a set of control samples, values of mean relative error (RE) of ~15% were achieved in most of the PLS models that were created. The reproducibility of the PLS models was evaluated for two control samples. The values of the RSD ranged from 2 to 20%. The better values of the RSD were achieved for the control sample with larger amounts of minerals determined as expected. Surprisingly, the significantly better values of the RSD were achieved for the PLS models based on the DRIFT technique.

Important advantages of the chemometric FTIR technique over the more popular quantitative phase analysis method (Rietveld technique of XRD analysis), including shorter time for analysis, better availability of instrumentation, and simplicity. First, the time of chemometric analysis of FTIR spectra is much shorter than Reitveld analysis time for XRD. The creation of calibration models is rather time-consuming, of course, but subsequent analysis is very quick. Second, FTIR spectrometers are present in a significantly larger number of laboratories than XRD. Finally, chemometric processing of results from FTIR spectra is considerably simpler than Rietveld treatment of diffraction data.

Determination of mineral constituents of rocks by means of FTIR spectroscopy together with chemometric analysis is simple and reliable. This method is also

Table 5. Reproducibility – list of results.

| | Chlorite | | Muscovite | |
|---|---|---|---|---|
| | CS02 (6.2 w/w %) RSD (%) | CS07 (19.1 w/w %) RSD (%) | CS02 (10.4 w/w %) RSD (%) | CS07 (36.5 w/w %) RSD (%) |
| DRIFT | 8.0 | 1.6 | 10.0 | 3.1 |
| Pellets | 20.6 | 13.1 | 17.3 | 12.8 |

| | Albite | | Quartz | |
|---|---|---|---|---|
| | CS02 (7.4 w/w %) RSD (%) | CS07 (14.3 w/w %) RSD (%) | CS02 (30.0 w/w %) RSD (%) | CS07 (58.3 w/w %) RSD (%) |
| DRIFT | 7.4 | 5.7 | 4.2 | 2.7 |
| Pellets | 19.0 | 13.1 | 22.8 | 15.9 |

feasible without knowledge of the characteristic absorption bands of the minerals in the FTIR spectra or exact knowledge of all the mineral components in the rocks analyzed. The present study has shown that FTIR spectroscopy in conjunction with the PLS regression method provided an acceptable alternative to the more commonly used methods for quantitative phase analysis – the Rietveld technique of XRD analysis.

## ACKNOWLEDGMENTS

The present study was carried out under project no. CZ.1.05/2.1.00/01.0040 of the 'Regional Materials Science and Technology Centre' within the framework of the operation program 'Research and Development for Innovations' financed by structural funds and by the state budget of the Czech Republic.

The study was carried out in conjunction with the Project Institute of Clean Technologies for Mining and Utilization of Raw Materials for Energy Use, reg. no. CZ.1.05/2.1.00/03.0082, supported by the 'Research and Development for Innovations Operational Programme' which is financed by structural funds of the European Union and by means of the state budget of the Czech Republic.

The authors thank George Laynr for assistance with the English.

## REFERENCES

Al-Degs, Y.S., El-Sheikh, A.H., Al-Ghouti, M.A., Hemmateenajed, B., and Walker, G.M. (2008) Solid-phase extraction and simultaneous determination of trace amounts of sulphonated and azo sulphonate dyes using microemulsion-modified-zeolite and multivariate calibration. *Talanta*, **75**, 904–915.

Armenta, S., Garrigues, S., and de la Guardia, M (2007) Determination of edible oil parameters by near infrared spectroscopy. *Analytica Chimica Acta*, **596**, 330–337.

Bish, D.L. and Howard, S.A. (1988) Quantitative phase analysis using the Rietveld method *Journal of Applied Crystallography*, **21**, 86–91.

Breen, C., Cleeg, F., Herron, M.M., Hild, G.P., Hillier, S., Hughes, T.L., Jones, T.G.J., Matteson, A., and Yarwood, J. (2008) Bulk mineralogical characterization of oilfield reservoir rocks and sandstones using Diffuse Reflectance Infrared Fourier Transform Spectroscopy and Partial Least Square analysis. *Journal of Petroleum Science and Engineering*, **60**, 1–17.

Brindley, G.W. (1980) *Crystal Structures of Clay Minerals and their X-ray Identification*. Monograph **5**, Mineralogical Society, London.

Couty, R. and Velde, B. (1986) Pressure-induced band splitting in infrared spectra of sanidine and albite. *American Mineralogist*, **71**, 99–104.

Chipera, S.J. and Bish, D.L. (2001) Baseline studies of the Clay Minerals Society source clays: Powder X-ray diffraction analyses. *Clays and Clay Minerals*, **49**, 398–409.

Esbensen, K.H. (2006) *Multivariate Data Analysis in Practice*, fifth edition. Camo, Oslo.

Fredericks, P.M., Lee, J.B., Sborn, P.R., and Swinkels, D.A.J. (1985) Material characterization using factor analysis of FT-IR spectra. Part 1: Results. *Applied Spectroscopy*, **39**, 303–316.

Fuller, M.P., Ritter, G.L., and Draper, C.S. (1988) Partial least-squares quantitative analysis of infrared spectroscopic data. Part I: Algorithm implementation. *Applied Spectroscopy*, **42**, 217–236.

Geladi, P. and Kowalski, B.R. (1986) Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, **185**, 1–17.

Haaland, D.M. and Thomas, E.V. (1988) Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction information. *Analytical Chemistry*, **60**, 1202–1208.

Hasegawa T. (2002) *Principal Component Regression and Partial Least Squares Modeling*. Pp. 2293–2312 in: *Handbook of Vibrational Spectroscopy* (J.M. Chalmers and P.R. Griffits, editors). Vol. **3**, John Wiley & Sons, Chichester, UK.

Hillier, S. (2000) Accurate quantitative analysis of clay and other minerals in sandstones by XRD: comparison of a Rietveld and a reference intensity ratio (RIR) method and the importance of sample preparation. *Clay Minerals*, **35**, 291–302.

Iñón, F.A., Garrigues, J.M., Garrigues, S., Molina, A., and de la Guardia, M. (2003) Selection of calibration set samples in determination of olive oil acidity by partial least squares-attenuated total reflectance-Fourier transform infrared spectroscopy. *Analytica Chimica Acta*, **489**, 59–75.

Kodama, H., Kotlyar, L.S., and Ripmeester, J.A. (1989) Quantification of crystalline and noncrystalline material in ground kaolinite by X-ray powder diffraction, infrared, solid-state nuclear magnetic resonance, and chemical-dissolution analyses. *Clays and Clay Minerals*, **37**, 364–370.

Lorber, A. and Kowalski, B.R. (1988) A note on the use of the partial least-squares method for multivariate calibration. *Applied Spectroscopy*, **42**, 1572–1574.

Luis, M.L., Fraga, J.M.G., Jimenéz, A.I., Jimenéz, F., Hernandéz, O., and Arias, J.J. (2004) Application of PLS regression to fluorimetric data for the determination of furosemide and triamterene in pharmaceutical preparations and triamterene in urine. *Talanta*, **62**, 307–316.

Madejová, J. and Komadel, P. (2001) Baseline study of the Clay Minerals Society source clays: Infrared spectroscopy. *Clays and Clay Minerals*, **49**, 410–432.

Martens, H. and Naes T. (1989) *Multivariate Calibration*. John Wiley & Sons, New York.

Meloun, M. and Militký, J. (2004) *Statistical Analysis of Experimental Data* (in Czech). Second edition, Academia, Praha, Czech Republic.

Moneeb, M.S. (2006) Polarographic chemometric determination of zinc and nickel in aqueous samples. *Talanta*, **70**, 1035–1043.

Moore, D.M. and Reynolds, R.C. (1997) *X-ray Diffraction and the Identification and Analysis of Clay Minerals*. Oxford University Press, New York.

Rietveld, H.M. (1969) A profile refinement method for nuclear and magnetic structures. *Journal of Applied Crystallography*, **2**, 65–71.

Ritz, M., Vaculíková, L., and Plevová, E. (2010) Identification of clay minerals by infrared spectroscopy and discriminant analysis. *Applied Spectroscopy*, **64**, 1379–1387.

Russel, J.D. and Fraser, A.R. (1994) Infrared Methods. Pp. 11–67 in: *Clay Mineralogy: Spectroscopic and Chemical Determinative Methods* (M.J. Wilson, editor). Chapman & Hall, London.

Środoń, J. (2002) Quantitative mineralogy of sedimentary rocks with emphasis on clays and with applications to K-Ar dating. *Mineralogical Magazine*, **66**, 677–687.

Vogt, C., Lauterjung J., and Fischer, R.X. (2002) Investigation of the clay fraction (<2 μm) of the Clay Minerals Society Reference Clays. *Clays and Clay Minerals*, **50**, 388–400.

Vaculíková, L. and Plevová, E. (2005) Identification of Clay Minerals and micas in sedimentary rocks. *Acta Geodynamica et Geomaterialia*, **2**, 163–171.

Wagieh, N.E., Hegazy, M.A., Abdelkawy, M., and Abdelaleem, E.A. (2010) Quantitative determination of oxybutynin hydrochloride by spectrophotometry, chemometry and HPLC in presence of its degradation and additives in different pharmaceutical dosage forms. *Talanta*, **80**, 2007–2015.

Wold, S., Sjöström, M., and Eriksson, L. (2001) PLS-regression: a basic tool of chemometrics *Chemometrics and Intelligent Laboratory Systems*, **58**, 109–130.

Zapata-Urzúa, C., Pérez-Ortiz, M., Bravo, M., Olivieri, A.C., and Álvarez-Lueje, A. (2010) Simultaneous voltammetric determination of levodopa, carbidopa and benserazide in pharmaceuticals using multivariate calibration. *Talanta*, **82**, 962–968.

(*Received 5 March 2012; revised 8 December 2012; Ms. 668; AE: S. Wold*)