Vibrational Spectroscopy 52 (2010) 154–157

Contents lists available at ScienceDirect

Vibrational Spectroscopy

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

Determination of degree of polymerization of insulating paper using near infrared spectroscopy and multivariate calibration

Edmilson Oliveira dos Santos [a], Andréa Monteiro Santana Silva [b], Wallace Duarte Fragoso [b], Celio Pasquini [c], Maria Fernanda Pimentel [a,*]

[a] Departamento de Engenharia Química, Universidade Federal de Pernambuco, Av. Prof. Artur de Sá, s/n, Cidade Universitária, 50740-521 Recife, PE, Brazil
[b] Departamento de Química Fundamental, Universidade Federal de Pernambuco, 50740-540 Recife, PE, Brazil
[c] Instituto de Química, Universidade Estadual de Campinas, 13084-971 Campinas, SP, Brazil

ARTICLE INFO

Article history:
Received 17 June 2009
Accepted 11 December 2009
Available online 22 December 2009

Keywords:
Insulating paper
Electric transformers
NIR
Multivariate calibration
Degree of polymerization

ABSTRACT

A simple alternative method to determine the degree of polymerization (DP) in transformer insulating papers, using near infrared spectroscopy (NIR), was developed to reduce the costs and the analysis time of the conventional method. Seventy-five samples of Kraft, crepe and cardboard types of paper, in different stages of degradation (DP varying between 200 and 900 units) were collected from transformers over a period of 3 years. The sample set was analyzed according to the conventional method to compare with the proposed alternative spectroscopic method. The spectra were obtained by diffuse reflectance using the spectral range between 1260 and 2500 nm. Some pre-processing strategies of the spectral data were evaluated and the best results were obtained when the derivatives spectra were used (Savitzky–Golay algorithm employing a 21 point window). NIR spectroscopy combined with partial least squares regression method revealed a simple and fast technique to determine the DP of the papers. The prediction error of 83, obtained after external validation, is acceptable for evaluation of the remaining useful life of insulating paper.

1. Introduction

The main insulating system in electrical power transformers consists of a paper (solid isolator) impregnated with mineral oil [1]. The aging and the degradation of the solid insulator determines the useful life of the transformer, as the insulating paper loses its mechanical properties, becoming vulnerable to disruption when the equipment is submitted to force due to short-circuits. A transformer is considered to have reached the end of its useful life when the mechanical traction resistance of the paper is reduced to half of the value presented by a fresh paper. Many factors influence the speed of aging of an insulator, such as temperature, humidity, oxygen content, electrical stress and mechanical stress.

Insulating paper is composed of long cellulose fibers. Cellulose is a linear polymer formed by a sequence of glucose monomers. The average length of cellulose molecules can be estimated by their degree of polymerization (DP), which determines the average number of units of glucose for each cellulose chain. For a fresh insulating paper DP varies between 900 and 1400 [2].

Since 1995 a number of companies of the world-wide electric power production sector have required from the manufacturers the determination of DP for new transformers in order to evaluate the quality of the paper used and guarantee an extensive useful life. When the DP of a paper reaches approximately 250, the capacity to support electromechanical tensions is already below 50% of the capacity of fresh paper and, consequently, the transformer becomes vulnerable to failure.

The operational costs involved when a failure occurs escalate when a tear occurs. Therefore, generally, insurance is taken out against equipment failure. Insurance policy clauses include requirements to measure the degree of polymerization of the solid isolation of the transformer to anticipate the remaining useful life of the equipment. Insurance companies can refuse payment to customers who fail to observe this requirement. In Brazil, DP determination is carried out according to ABNT NBR 8148 Standard (equivalent to ASTM D-4243) [3] which is time-consuming, destructive of the sample, and demands a large amount of sample to supply the replicate measurements necessary to achieve reliable results.

These comments justify the need to develop faster and simpler analytical methods that can substitute the conventional method for DP determination of insulating paper. Baird et al., in recent papers, have used spectroscopy in the NIR region and UV–Vis region, associated with multivariate calibration (PLS), to determine the degree of polymerization in insulating papers using portable equipment especially developed for the task. They demonstrated

---

* Corresponding author. Tel.: +55 81 2126 8729; fax: +55 81 2126 7235.
E-mail address: mfp@ufpe.br (M.F. Pimentel).

0924-2031/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.vibspec.2009.12.004

the potential of NIR spectroscopy for this type of application [4–6]. However, a limited calibration set, restricted to only one type of paper, was used, with no external validation, which is essential to assure the reliability of a multivariate model.

The present work reports an additional study of the application of NIR spectroscopy to determine the degree of polymerization in insulating paper. The calibration set is composed of three different types of paper (kraft, crepe and cardboard), collected from 75 different transformers over a period of 3 years. Regression models based on partial least squares (PLS) and multiple linear regression (MLR) (after variable selection) were evaluated using various pre-processing procedures. Finally, an external validation was carried out in order to test the real quality of the models.

2. Experimental

2.1. Insulating paper samples and spectra acquisition

The sample set was composed of 75 insulating paper samples, collected from different transformers over a period of 3 years. The papers were of three types: kraft, crepe and cardboard. All samples were degreased before further handling.

The degree of polymerization (DP) of the samples was determined using the Brazilian Standard ABNT NBR 8148, equivalent to ASTM D-4243, and the values were taken as the reference for regression models [3]. The range of DP was between 200 and 900.

The samples of paper were too thin to permit using a single piece for direct diffuse reflectance measurements. Therefore, in order to acquire their NIR spectra, samples were cut in pieces with different sizes, with minimum and maximum lengths of 1 and 10 mm, respectively. The width of the papers was approximately 1, 2 and 2 mm for the kraft, crepe and cardboard papers, respectively. Cutting the paper was necessary because many samples (especially those with low degree of polymerization) were severely damaged. Cut samples were transferred to glass vials of 1.5 cm diameter and their bottoms were directly placed on the measurement window of the reflectance accessory. Two samples, randomly chosen, cut in different sizes, were employed to evaluate the effect of the paper size on the NIR spectrum. Significant differences were not observed.

Reflectance spectra of typical paper samples present a high noise in the short-wave region of the NIR spectrum (Fig. 1a). Therefore, the spectral region employed for evaluation of regression models ranged from 1200 to 2500 nm, as average of 50 scans with resolution of 8 cm^{-1}, using a Bomem MB 160 spectrometer, equipped with a diffuse reflectance accessory (SamplIR).

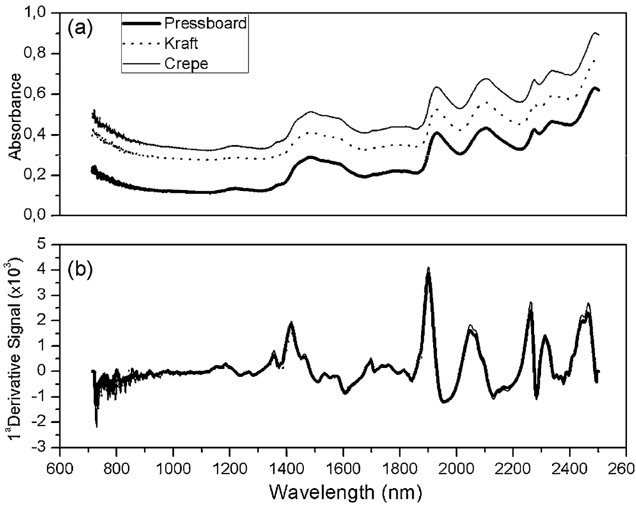

Fig. 1. (a) Raw NIR spectra and (b) first derivative spectra (Savitzky–Golay, second order polynomial and 23 points window) of samples of kraft, crepe and cardboard insulating paper.

2.2. Multivariate calibration

The spectra set obtained for the samples showed, as expected, a high scattering effect. Thus, various pre-processing techniques were evaluated, such as standard normal variate (SNV), multiplicative signal correction (MSC), first and second derivatives (Savitzy-Golay using different window sizes) and normalization, in order to minimize scatter. The models were built using partial least squares regression (PLS) and multiple linear regression (MLR). Full cross-validation (CV) was used to define the number of factors required for the PLS models.

PLS models were evaluated by using (a) the full spectral region or (b) only the variables which corresponded to the significant regression coefficients chosen by using the Jack–Knife algorithm of the Unscrambler version 9.2 software [7]. This algorithm employs cross-validation to create a number of sub-models. These sub-models were based on all remaining samples after the cross-validation segment is removed from the data set. For every sub-model, a set of model parameters (**b**-coefficients, scores, loadings and loading weights) were calculated. Variations from these sub-models were estimated so as to assess the stability of the results. In addition a total model was generated, based on all the samples. For each variable, the difference between the *b*-coefficient $\mathbf{b}_i$ in a sub-model and the $\mathbf{b}_{tot}$ for the total model was calculated. The software was able to calculate the sum of the squares of the differences in all the sub-models to get an expression of the variance of the estimated $\mathbf{b}_i$ for each variable. The significance of the estimate of $\mathbf{b}_i$ was calculated, using the t-test. Thus the resulting regression coefficients could be presented with uncertainty limits corresponding to 2 standard deviations under ideal conditions. Variables producing regression coefficients which do not cross the zero line were considered significant [8].

Variables selection for MLR models was carried out by using the successive projection algorithm (SPA). The essence of SPA consists of projection operations carried out on the calibration matrix. The SPA builds an ordered chain of variables where each element is selected in order to present the lowest colinearity (highest ortogonality) with respect to the previous ones [9–10]. The best performing chain (chosen on the basis of the lowest RMSEP achieved for a training sample set) is selected to contain the most valuable variables to be included in a MLR model. SPA-MLR models are simpler than PLS models and have shown advantageous characteristics in some applications when compared with PLS [11–12]. Recently two new implementations have been added to SPA. The first is the use of a method to eliminate variables in order to increase the parsimony of the model, removing those that do not affect the result significantly [13]. The second is the implementation of cross-validation in the SPA, eliminating the need of an additional internal validation set as demanded [14].

For the construction and evaluation of the regression models, the samples were separated into a calibration set (55 samples) and an external validation set (20 samples). The predictive ability of the models was evaluated by the root mean square error of prediction (RMSEP) obtained by using the external validation set.

All pre-processing and regression calculations were performed using the Unscrambler chemometrics software package [7]. Wavelength selection by SPA was carried out using a program written in Matlab.

3. Results and discussion

Fig. 1 shows NIR raw (Fig. 1a) and first derivative spectra (Fig. 1b) of kraft, crepe and cardboard paper samples. As can be seen,

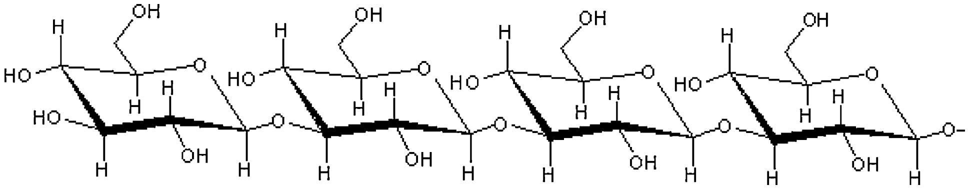

Fig. 2. Chain structure of cellulose showing the main functional groups presenting spectral absorption features in the NIR region.

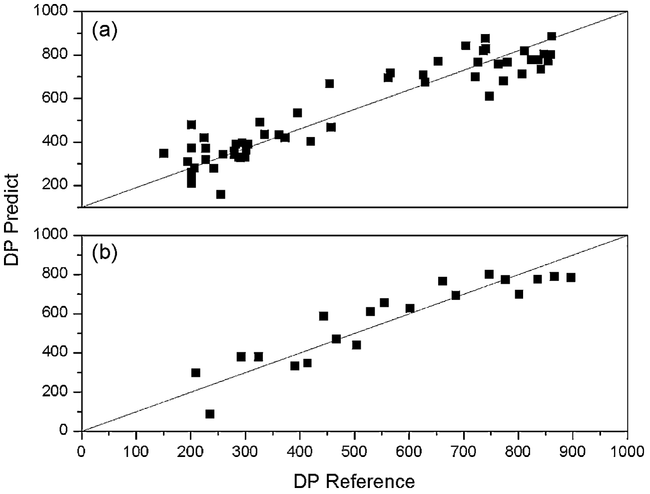

Fig. 3. Values of DP predicted by the optimized PLS model versus reference values: (a) values predicted by full cross-validation (internal validation) and (b) values predicted using an external prediction set. The line is the quadrant bisectrix which represents the agreement between predicted and reference values.

baseline shifts are minimized by using the first derivative treatment of the spectral data.

The main constituent of insulating paper is cellulose fibers, extracted from wood. The fiber is composed of different sized chains of glucose molecules of, linked by hydrogen bonds through hydroxyl groups. Cellulose is a uniform, linear polysaccharide of β-D-glucopyranose units linked in position 1:4 (Fig. 2). The linear arrangement of β-linked glucose units in cellulose presents uniform distribution of –OH groups outside each chain. When two or more cellulose chains make contact, the hydroxyl groups are ideally situated to link the chains together by action of hydrogen bonds. These linkages are considered the main factor for paper resistance.

The absorption bands in NIR infrared spectra are mainly due to overtones and combinations of stretching and bending modes of the C–H, C–C, C–O and O–H bonds found in the cellulose structure. However, bands from water are also observed. Bands in the 1200–1700 nm region are mainly due to the first and second overtones of C–H stretching vibrations. Bands between 1400 and 1600 nm may be assigned to the first overtone of O–H stretching modes of the cellulose and intramolecular hydrogen bonds as well as a result of the adsorbed water. Bands around 1900 nm are due to combinations of O–H deformation and stretching vibrations of the adsorbed water. In the region between 2000 and 2300 nm, combination bands of O–H bending and C–O stretching vibrations (cellulose) can be found.

The results obtained for PLS and SPA-MLR models, constructed with data submitted to some of the pre-processing procedures evaluated, are presented in Table 1. The number of factors of PLS models constructed by using only variables corresponding to significant regression coefficients is always lower than that for the models constructed by using all variables.

The first derivative (Savitzky–Golay) employing a second order polynomial and a 23 point width window was the pre-processing treatment resulting in the best prediction results for the external validation. In this case, using only the variables which correspond to significant regression coefficients, a root mean square error of prediction (RMSEP) of 83 for DP prediction was attained by a PLS model employing four factors. Fig. 3 shows the predicted versus reference values plot for the calibration and external validation using this model, which present determination coefficients (R) of 0.94 and 0.92, respectively.

Table 1

Results obtained for internal full cross-validation (CV) and external validation (P) for the determination of DP in insulating paper: R_C and R_P, regression coefficient for cross-validation and external validation, respectively; W, number of smoothing points; #LV, number of latent variables.

Pre-processing	RMSECV (R_C)				RMSEP (R_P)	
	All variables	#LV	Selected variables	#LV	All variables	Selected variables
PLS models						
None	94.4 (0.91)	16	151 (0.74)	5	107 (0.89)	180 (0.57)
MSC	101 (0.89)	9	109 (0.87)	5	106 (0.88)	158 (0.73)
MSC + 1st derivative SG (W23)	88 (0.92)	8	91 (0.91)	4	112 (0.88)	102 (0.88)
MSC + 1st derivative SG(W23) + normalization	79 (0.93)	9	67 (0.95)	4	100 (0.91)	92 (0.91)
SNV	89 (0.92)	13	117 (0.85)	6	120 (0.86)	122 (0.82)
1st derivative SG (W23)	82 (0.93)	9	78 (0.94)	4	87 (0.91)	83 (0.92)
1st derivative SG (W25)	81 (0.93)	9	94 (0.91)	5	101 (0.88)	93 (0.92)
Normalization	88 (0.92)	15	185 (0.56)	5	114 (0.89)	190 (0.54)
2nd derivative SG (W23)	120 (0.84)	8	88 (0.92)	5	144 (0.73)	148 (0.71)
1st derivative SG (W23) + normalization	84 (0.93)	9	75 (0.94)	4	100 (0.90)	95 (0.91)
1st derivative SG (W23) + normalization	79 (0.93)	9	67 (0.95)	4	100 (0.91)	96 (0.91)

Pre-processing	RMSECV (R_C)				RMSEP (R_P)	
	All the selected variables	#V	Significant variables	#V	All the selected variables	Significant variables
SPA-MLR models						
MSC + 1st derivative SG (W23) + Normalization	84 (0.93)	27	68 (0.95)	18	128 (0.83)	144 (0.80)
1st derivative SG (W23)	96 (0.91)	27	84 (0.93)	18	166 (0.74)	175 (0.73)
1st derivative SG (W25)	83 (0.93)	31	74 (0.94)	25	136 (0.80)	140 (0.79)

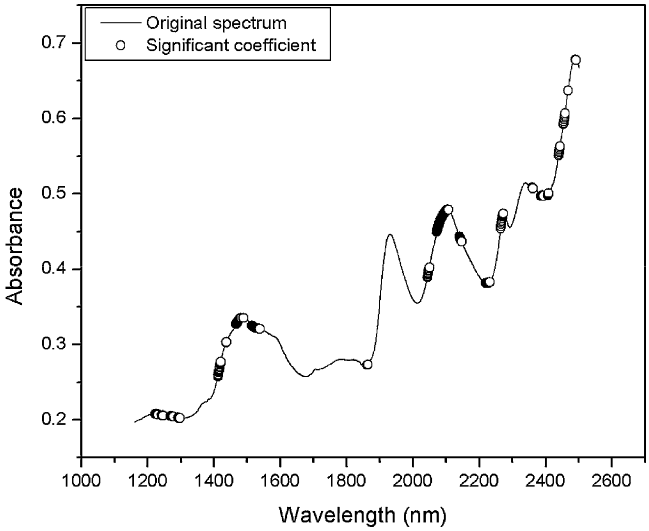

Fig. 4. Typical raw spectra of a sample of insulating paper showing the variables selected based on the significant regression coefficients.

A typical spectrum showing the variables selected to construct the optimized model is presented in Fig. 4. Notice that the absorption band attributed to the presence of adsorbed water around 1940 nm has been excluded probably due a poor correlation of the absorbance signal, due to paper moisture, to determination of the degree of polymerization of the paper. On the other hand, the variables associated with the band attributed to –OH groups of cellulose and water molecules interacting with those groups (1400–1500 nm) are relevant for the determination of the this property of insulating paper.

MLR-SPA models also produced acceptable results for the calibration, but the prediction errors in the external validation were always higher than those obtained using PLS models. Application of modified SPA, which eliminates variables in order to improve the parsimony of the model, contributes to improve the root mean square error of cross-validation (RMSECV) values. However, the RMSEP values obtained after external validation were slightly higher.

4. Conclusion

The viability of NIR spectroscopy for determination of degree of polymerization in several types of transformer paper was demonstrated. Regression models were built using samples of three different types of insulating paper (crepe, cardboard, and kraft) taken from different transformers during a period of 3 years, proving the extensive applicability of the proposed spectroscopic method.

Several data pre-treatment procedures and several different regression and variable selection algorithms were evaluated in order to search for the optimum regression model to be used for DP determination. In the present case, the models constructed by using PLS regression on the first derivative spectra and variables selected by the Jack–Knife algorithm presented superior results over MLR models made by employing variables selected by the successive projection algorithm.

The resulting spectroscopic method is fast (a typical determination of DP takes about 2 min) and requires minimum sample pre-treatment.

The repeatability of the conventional method for determination of DP has been experimentally determined as 55. Thus, the RMSEP value of 83, obtained after external validation of the optimized model, is acceptable for routine monitoring of DP in insulating paper.

Acknowledgments

The authors acknowledge Prof. Roberto Kawakami Harrop Galvão for the SPA program, CHESF and SM Controle de Qualidade Ltda for providing the paper samples and financial support by FINEP/CTPETRO, CNPq and CAPES/PROCAD. Prof. C.H. Collins is also kindly acknowledged for manuscript revision.

References

[1] R.M.N. de Morais, M.Sc. Thesis, Federal University of Rio de Janeiro, Rio de Janeiro, 1990.
[2] D.H. Shroff, A.W. Stannet, Proc. Inst. Elect. Eng. 132 (1985) 312.
[3] American Society for Testing and Materials, ASTM D-4243 Standard Test Method for Measurement of Average Viscometric Degree of Polymerization of New and Aged Electrical Papers and Boards, 1999.
[4] P.J. Baird, H. Herman, G.C. Stevens, P.N. Jarman, Non-destructive condition assessment of insulating materials in power transformers, in: Proceedings, International Symposium on Electrical Insulating Materials, Kitakyushu, Japan, (2005), p. 425.
[5] P.J. Baird, H. Herman, G.C. Stevens, P.N. Jarman, IEEE Trans. Dielect. Electr. Int. 13 (1) (2006) 309.
[6] P.J. Baird, H. Herman, G.C. Stevens, Non-destructive and in-situ analysis of insulating materials in high-voltage power transformers, in: International Conference on Solid Dielectrics, Toulouse, France, July 5–9, (2004), p. 719.
[7] The Unscrambler 9.2, CAMO ASA Nedre Vollgate 8, N-0158 Oslo, Norway, 2005.
[8] K.H. Esbensen, Multivariate Data Analysis—In Practice, 5th ed., CAMO, Oslo, 2002.
[9] R.K.H. Galvão, M.F. Pimentel, M.C.U. Araújo, T. Yoneyama, V. Visani, Anal. Chim. Acta 443 (2001) 107.
[10] M.C.U. Araújo, T.C.B. Saldanha, R.K.H. Galvão, T. Yoneyama, H.C. Chame, V. Visani, Chemom. Intell. Lab. Syst. 57 (2001) 65.
[11] M.S. Di Nezio, M.F. Pistonesi, W.D. Fragoso, M.J.C. Pontes, H.C. Goicoechea, M.C.U. de Araujo, B.S.F. Band, Microchem. J. 85 (2007) 194.
[12] M. Grünhut, M.E. Centurion, W.D. Fragoso, L.F. Almeida, M.C.U. Araujo, B.S.F. Band, Talanta 75 (2008) 950.
[13] R.K.H. Galvão, M.C.U. Araújo, W.D. Fragoso, E.C. Silva, G.E. José, S.F.C. Soares, H.M. Paiva, Chemom. Intell. Lab. Syst. 92 (2008) 83.
[14] R.K.H. Galvão, M.C.U. Araujo, E.C. da Silva, G.E. José, S.F.C. Soares, H.M. Paiva, J. Braz. Chem. Soc. 18 (2007) 1580.