*CELLULOSE CHEMISTRY AND TECHNOLOGY*

# NEW ALTERNATIVE TO DETERMINE THE YIELD OF ETHYLENE OXIDE SUBSTITUTION IN HYDROXYETHYL CELLULOSE REACTION BY NEAR-INFRARED SPECTROSCOPY AND PARTIAL LEAST SQUARES REGRESSION

M. M. LÓPEZ-TAPIA,[*,**] C. C. LÓPEZ-SUERO[*] and R. MARÍN-SÁNCHEZ[**]

[*]*Engineering and Chemical Sciences Department, Universidad Iberoamericana AC,*
*Prolongación Paseo de la Reforma 880, Lomas de Santa Fe, 01219, Mexico City, Mexico*
[**]*Research, Development and Technological Innovation Center of Grupo Petroquímico Beta, Periférico Sur 4194, PB2, 01900, Mexico City, Mexico*
✉ *Corresponding author: C. C. López-Suero, carolina.lopez@ibero.mx*

*Received* May 19, 2017

The yield of substitution of ethylene oxide in the hydroxyethyl cellulose reaction is a crucial parameter disclosed by molar substitution and is closely related to the performance of the polymer. A simple, fast and green method for determining the molar substitution of hydroxyethyl cellulose has been developed in this study, using near infrared spectroscopy (NIRS). One of the major advantages of this method is saving time for the quality control procedures in chemical industry – at least 7 hours compared to the reference gas chromatography method. This technique eliminates chemical reactions, the need to manipulate hazardous materials, and minimizes waste generation and energy consumption. NIRS spectra were collected in the range from 5708.26 to 5553.98 cm$^{-1}$ for samples with molar substitution from 1.18 to 2.14. Partial least squares regression was used, providing a RMSEC (root mean square error of calibration) of 0.00363 and a RMSEP (root mean square error of prediction) of 0.0171.

***Keywords***: ethylene oxide efficiency, hydroxyethyl cellulose (HEC), molar substitution (MS), near infrared spectroscopy (NIRS), partial least squares (PLS) regression

## INTRODUCTION

Water soluble hydroxyethyl cellulose is manufactured by reacting alkali cellulose with ethylene oxide at elevated temperatures and pressures in a mixture of organic solvents and water.[1] This reaction, having a large amount of solvent as compared to the cellulose content, has a very low yield of substitution; the ethylene oxide reacts not only with the cellulose, but also with the available water in the medium.

This is the reason why, for obtaining a specific level of substitution or molar substitution (MS), ethylene oxide has to be added in larger amounts than the desired MS.

The average number of moles of substituent groups per anhydroglucose unit is known as molar substitution (MS).[2]

The ethylene oxide (EO) yield is calculated in terms of MS as follows:

$$EO\ efficiency = \frac{Real\ MS}{Loaded\ MS} \qquad (1)$$

where Real MS = MS obtained by the analytical procedure; Loaded MS = charge of EO moles/charge of AGU moles (anhydroglucose unit, dry basis). Note: for the calculation, the charges within the reaction are considered.

Vesa Myllymaki and Reijo Aksela reported ethylene oxide efficiencies of 50 to 70% in alkali catalyzed hydroxyethyl ether formation.[3] Other authors reported even lesser efficiencies in the homogeneous etherification reaction – of 20 to 30%.[4]

This parameter, the MS, controls the solubility and compatibility properties of the product. Low substitutions, MS 0.05-0.5, are for products soluble only in aqueous alkali; while MS ≥ 1.5 produces water soluble hydroxyethyl cellulose (HEC).[5] HEC is a non-ionic polymer that is water soluble and is a rheology modifier used in multiple applications. Low MS products have been used as sizing agents in the textile industry and as raw material for packaging film.[1] Most commercial HECs have their MS between 1.5 and 4, and are used as thickeners in cosmetic preparations,[6] but also as thickeners, stabilizers and suspending agents in water-borne paints;[7] in construction, helping to reduce the amount of water in formulations, improving the adhesion, strength and plasticity of the materials;[8] in oil field applications, used for fluid loss control in water-based drilling fluids.[9] Furthermore, high molecular weight HEC is used in the pharmaceutical industry as an excipient to provide a swellable diffusion barrier in controlled release applications.

The MS, determined by analytical procedures, is the parameter that depicts the efficiency of ethylene oxide substitution, it is crucial to guide and determine the best manufacturing conditions of the cellulose ether reactions, as well as to identify the product's fields of application, which are related to this parameter. The official method for determining the MS is the gas chromatography technique, which involves a complicated process of reaction as pretreatment. Simple and rapid analytical techniques are always the best option for quality control departments in the chemical industry. Hence, it is necessary to implement a method that allows obtaining rapid and accurate results of MS.

In 1971, Svante Wold invented the word 'chemometrics', the art of extracting relevant chemical information from data reproduced in chemical experiments.[10] Chemometrics is the bridge between connecting the state of a chemical system to the measurements of the system. It has become an essential part in modern chemical and biomedical industries. Near infrared spectroscopy is a simple and fast technique, which has already been used in many analytical applications. As mentioned by Yongfu Diao *et al.*, to optimally develop and use HEC, more detailed information on its complex structural and functional properties is required. Consequently, there is a growing urgency to develop novel analytical methods for the characterization of HEC.[11]

The ASTM D 4794-94[12] is the official method currently used to determine hydroxyethoxyl content in cellulose ether products from which the MS can be calculated. The procedure involves a reaction developed in Reacti-Vials, and has a condition for acceptance: the content should be discarded if the loss is greater than 25 mg at the end of the reaction, the problem is the hermeticity of the required Reacti-Vials and Mininert valves, because air tightness fails usually two of three times. The reactants used in the gas chromatography (GC) technique involve hazardous materials, whereas with the NIRS technique, the risk of causing damage to the analyst and the environment is eliminated. "It is better to prevent waste than to treat or clean up waste after it is formed",[13] as mentioned in the first of the Twelve Principles of Green Chemistry. Another of the principles pursued by the implementation of the NIRS alternative is the elimination of unnecessary substances, minimized energy consumption and reduced potential for chemical accidents.

Zhou (2005) reported the measurement of the MS in HEC by NMR, in combination with elemental analysis to confirm the results,[4] but although it represents an alternative to the official GC method, it is a complex and expensive technology.

NIRS is a faster technique, compared to the other ones, it is non-destructive[14] and does not require any reaction. Moreover, it is a reliable technique, even in the hands of relatively untrained personnel.[15] NIRS has been widely used for quantifying components, as described in a study aiming to evaluate the feasibility of using multipoint NIRS for in-line moisture content quantification during the freeze-drying process.[16] In another field of application, it is a possible alternative to GC for the quantitative analysis of fatty acids in forages;[17] in addition, the protein content of *A. auricula* was successfully determined using this technique.[18] Likewise, in food chemistry, it was found that NIRS could be used as an easy, rapid tool to quantitatively predict free amino acids and sugars in Chinese rice wine without sophisticated methods[19-20] and to determine the quality factors in food products.[21-22] It has also been used in combination with other techniques for the characterization and evaluation of petroleum with satisfactory results.[23] A combination of NIRS and refractometry demonstrated to be an accurate and reliable method to determine the concentration of alcohol and extract in different beer samples.[24]

Another important application is the in-time on-process ability of high scanning speed and high sensitivity to catch rapid changes in hydrolysis reactions, allowing studying the kinetic behaviors.[25] This can be a substitute for some costly and time-consuming conventional wet chemistry techniques, the data generated can be readily used for efficient risk assessments and effective decision making programs for environmental safety.[26]

No studies about the molar substitution (or hydroxyethyl content) by NIRS on HEC were found in the literature. Recently, an article was published about NIRS for the determination of the hydroxypropyl content in a different molecule, the hydroxypropyl methylcellulose (HPMC),[27] exploring a range from approx. 5.5 to 10%; while in the present study the MS range of the model represents the hydroxyethoxyl substitution range from 35 to 50%, besides, the reference techniques used to validate the NIRS method are different.

As mentioned before, the most developed methods in NIRS use partial least squares (PLS) as the regression technique for the calibration of the models: the advantages of NIR spectroscopy and the PLS method lie in their abilities to develop multivariate regression models with good correlation coefficients and low predictive errors.[28]

PLS regression is a recent technique that generalizes and combines features from principal component analysis and multiple regression. It is particularly useful when we need to predict a set of dependent variables from a large set of independent variables (*i.e.* predictors). It originated in the social sciences (specifically economy),[29] but became popular in chemometrics (*i.e.* computational chemistry) in part due to Herman's son Svante[30] and in sensory evaluation.[31]

PLS-regression, PLSR, is the PLS approach in its simplest, and in chemistry and technology, the most used form is the two-block one. PLSR is a method for relating two data matrices, X and Y, by a linear multivariate model, but goes beyond traditional regression in that it also models the structure of X and Y. PLSR derives its usefulness from its ability to analyze data with many noisy, collinear and even incomplete variables in both X and Y. PLSR has the desirable property that the precision of the model parameters improves with the increasing number of relevant variables and observations.[32]

The PLS method is popular in the industries that collect correlated predictor data, for example, for multivariate calibration in analytical chemistry; for spectroscopy in chemometrics; and for quantitative structure activity relationships in drug design. The PLS method extracts orthogonal linear combinations of predictors, known as factors, from the predictor data, which explain variance in both the predictors and the response(s).[33] PLS is considered especially useful for constructing prediction equations, when there are many explanatory variables and comparatively little sample data.[34]

## EXPERIMENTAL

### Apparatus and materials

NIR spectra were collected in a Thermo Scientific Antares II Near IR Analyzer, equipped with the Thermo Result 3 SP5 Build 25 Integration software for data collection and the TQ Analyst 8.4.259 software for data treatment using PLS.

The GC analysis, used as reference, was performed on a Bruker Scion 456-GC, with a TCD detector, using Compass CDS 3.0 version 3.0.0.68 – Bruker GC Chromatography data system.

### MS determination by GC

It is important to mention the methodology of GC to understand the reason for developing an alternative to it that does not involve the complexity found in this reference method, the ASTM D 4794-94 (reapproved, 2009).[12] The reaction of hydrogen iodide with the cellulose ether liberates a mol of iodoethane per each mol of hydroxyethoxyl substituted on the cellulose chain. The iodoethane is extracted *in situ* with 1,2-xylene and quantified by GC, using an internal standard technique. The determination involves the following procedure: drying the sample for 1 h at 105 °C, weighing 60 to 80 mg of the specimen in a clean 5-mL Reacti-Vial (already at constant weight); adding 2 mL of a standard solution of toluene and 1,2 xylene, and adding 2 mL of hydrogen iodide. The vial must be capped tightly and weighed, shaken and placed in a heating block at 180 ± 5 °C for 2 h. After removing from heat and cooling the vial, the specimen will separate into two layers, it has to be reweighed to determine any loss due to leakage. Vial content should be discarded if the loss is greater than 25 mg. The analysis is done by injecting 1.0 µL of the upper layer of the specimen and using the data of known weights of toluene and sample for the preparation to perform the calculations.

The final data obtained by the method represents the hydroxyethoxyl content in the cellulose ether, this result should be corrected for ash content and, then, it can be converted to MS by the following formula:

$$MS = \frac{162 \; x \; H}{6100 - (44 \; x \; H)} \quad (2)$$

where H = % hydroxyethoxyl (ash-free basis); 162 = molecular weight of anhydro cellulose unit; 44 = molecular weight of ethylene oxide; 6100 = factorized number for transforming hydroxyethoxyl group to ethylene oxide.

### Calibration model in NIR

MS determination in HEC was performed on a total of 27 samples, from which 18 were selected for the calibration set and the rest were used as the validation set. The models developed were constructed with different conditions, including the spectral range, number of PLS factors and data treatment processes. The PLS technique was selected according to the criteria of having a large number of components, variable spectral baselines, and a sample matrix with unknown materials or impurities.[35]

### Sample treatment for NIRS

Although the water content specification for the HEC product used as sample was 5% max., it was subjected to a drying process of 1 h at 105 °C before processing, and then sieved through an ASTM 40 mesh.

## RESULTS AND DISCUSSION

### GC analysis

The samples were analyzed by gas chromatography by the method for Determination of Hydroxyethoxyl Substitution in Cellulose Ether Products[12] and by applying the formula for molar substitution, as already described. Each sample was run in duplicate and each Reacti-Vial content was injected 20 times. To ensure the correctness of the results, the standard deviation (SD) was applied to the data and the control limits were calculated by the mean ± SD. The data found beyond the established values for the control were discarded; the rest of the results were used to calculate the mean. Finally, the MS was calculated by the mean of the duplicates. The average SD found for the results by this method was 0.0204.

### Calibration model development

To build the NIRS method for MS determination, nine samples were randomly selected for the validation set and 18 samples were used as the calibration set. Calibration models by partial least squares (PLS) were built and used to predict the MS in the validation set. These models were constructed with varying conditions, including the use of the original spectra, applying first and second derivatives; testing with different spectral ranges, number of PLS factors, and data pretreatment processes. Performance was judged by the values of coefficient $R^2$, PRESS (predictive residual sum of squares), RMSECV (root mean squared error of cross-validation), RMSEC (root mean square error of calibration), and RMSEP (root mean square error of prediction). The selected range of molar substitution was from 1.18 to 2.14, which covers the grade of material under study – low viscosity hydroxyethyl cellulose.

The development of the ideal model in NIR consisted, in the first place, in selecting the predictive analytical technique. As mentioned before, PLS was selected, and in addition, the standards were selected and analyzed by the reference technique and then measured by NIRS; the data were studied within different regions of the NIR spectra, and after a long search for the best correlation, the region was defined in the range of 5708.26 to 5553.98 cm$^{-1}$.

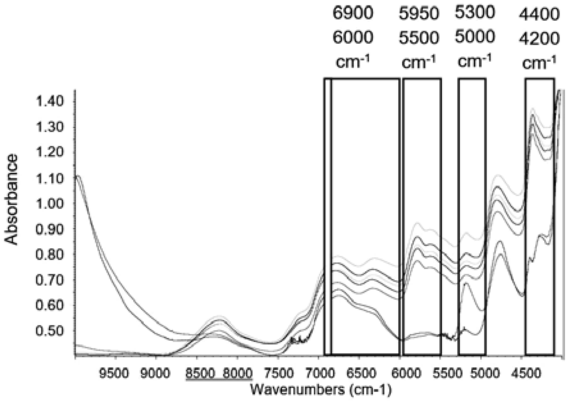


Figure 1: Main regions identified for representing different absorbances between HEC and cellulose used as raw material

The main regions studied for constructing the model were selected according to the literature[36] and the study of the NIR spectra differences found between the cellulose samples used as raw material for the HEC and the hydroxyethyl cellulose samples. Figure 1 shows two sets of spectral patterns, one corresponding to the raw cellulose material (the ones below, near to the “x” axis), and the other for HEC – the product; this last one is situated in the upper position, where larger absorbance is shown compared to that in the raw cellulose pattern.

In Figure 1, there are some rectangles, indicating the zones where the major differences were found between the raw material and the final product. These regions were explored in order to find the one with the best sensitivity for identifying the substitution of cellulose. In the region from 4400 to 4200 $cm^{-1}$, the wavenumber of 4400 $cm^{-1}$ corresponds to O-H/C-O from glucose, the peak at 4386 $cm^{-1}$ represents C-H starch, those at 4307, 4292, 4283 and 4252 $cm^{-1}$ are ascribed to C-H groups, all corresponding to polysaccharides. Nevertheless, in this region, poor correspondence was found in the calibration between the results for MS by NIR and GC.

The region from 5300 to 5000 $cm^{-1}$, shown in Figure 1, was discarded because it represents the water content in the samples. This was observed by comparing the samples before and after the drying process for cellulose and HEC. This is the reason for establishing 1 h drying of the sample at 105 °C before the analysis, even though the product had a water content specification of maximum 5%. Another pretreatment implemented was to pass the samples through a 40-ASTM mesh: it is preferable to have a regular size in the samples for the study, because particle size can interfere with the NIR measurement.[14] The above-mentioned lines were corroborated with findings in the literature: in this region, there are different O-H bonds all related to water: the wavenumbers 5200, 5155 and 5150 $cm^{-1}$ identify correspond to the O-H assigned to molecular water; that of 5181 $cm^{-1}$ is ascribed to O-H stretching and HOH bending in polysaccharides; that of 5102 $cm^{-1}$ identifies polymeric O-H for polysaccharides. The rest of the wavenumbers up to 5000 $cm^{-1}$ correspond to amino groups. This region, from 5300 to 5000 $cm^{-1}$, was eliminated from the study for searching the best area of calibration for MS.

In the region from 5950 to 5500 $cm^{-1}$, the methylene functional group is identified in different structures. From the literature, the main ones related to the functional group in the study were as follows: 5675 $cm^{-1}$ and 5666 $cm^{-1}$ for C-H methylene corresponding to hydrocarbons; C-H methyl aromatic associated (ArCH3) to 5650 $cm^{-1}$ and C-H methylene (CH2) corresponding to cellulose at 5618 $cm^{-1}$. Thus, this region shows the best possibilities for developing the calibration model.

The last region identifying differences in the NIR spectra of cellulose and hydroxyethyl cellulose was from 6900 to 6000 $cm^{-1}$: the region from 6897 to 6471 $cm^{-1}$ is mainly related to amines, the wavenumbers of 6330 and 6319 $cm^{-1}$ are related to O-H in alcohols and water. Therefore, this region was also left out from the study.

The method was calibrated and validated with more samples. The selected wavenumber range was the one that provided the best correlation. In this study, several spectral regions were tested, including some narrow regions with only a few spectral data points (2 nm/one point), but often the correlations were of even less than 0.60, only the selected regions showed positive correlation between the reference values and those calculated by the model in NIRS.

Partial least squares (PLS) calibration used by TQ Analyst software is based on the PLS1 algorithm. It is a statistical approach to quantitative analysis. The PLS1 algorithm examines the specified region or regions of the calibration spectra to determine the areas varying statistically as a function of component concentration.[35]

With linear regression, the best straight line, known as regression curve, is located, which can be used for determining the concentration of the samples. Nevertheless, by using PLS, a regression of the registers of the sample is done instead of the optical data that are directly produced by NIR instruments. The PLS method is formally defined as the method of predictive regression of two blocks based on latent variables.

The responses of a NIR instrument to different wavenumbers are usually highly collinear (intercorrelated), which is thus a problem in an ordinary multiple lineal regression, because it assumes each variable as unique. For being useful, NIR calibration has to handle the phenomena of multicollinearity. In the regression methods that use estimated latent variables, the collinearity is seen as a stabilizing advantage rather than as a problem.[16]

Multivariate calibration can be viewed as a two-stage procedure. First, the model is constructed using training samples, for which the predictor and predictand variables are known or measured. Next, the model is validated by comparing the predictions against the reference values for samples that were not used for model building. The result of this validation step is an estimate of the average prediction uncertainty, often denoted as root mean squared error of prediction (RMSEP).[36]

Different regions were studied using PLS with MSC (multiplicative signal correction). MSC is an important method for the correction of scattered light, which is used to modify additive and multiplicative effects in spectral data.[35]

Figure 2 displays a plot of RMSECV of the developed method in NIRS *vs*. PLS factors, indicating that a calibration model built with 10 PLS factors performs best in terms of the value of RMSECV. The RMSECV was 0.10378 for this model.

The best estimate of the future performance of a calibration model is the RMSECV, which is calculated in cross-validation by iteratively removing (and replacing) each sample of the data set. The optimal number of factors is selected considering the first minimum in the plot, or choosing the global minimum if it yields significantly lower RMSECV. For this method, different numbers of factors were studied, 10 factors were selected according to the mentioned criteria.

When the PLS method is calibrated, all of the relevant concentration information and spectral information in the analyzed region or regions of the calibrated standards is condensed into a set of factors. Each factor represents an independent source of variation in the data. The factors are ranked by the amount of variance they describe. The first factor describes most of the variation in the calibration standards. Each additional factor describes most of the remaining variation. Therefore, the first factor contains the most common information in the data. The other factors contain information that is more specific, representing small variations in the data, which are often important to the analysis.[35]

Figure 3 (A) illustrates the performance of the selected calibration model with the calibration data, having a RMSEC of 0.00363 and a correlation coefficient of 0.9999. For the validation data, represented in Figure 3 (B), this model has a RMSEP of 0.0156. The models, as can be seen in the MS correlation plot, illustrate good correspondence between the results obtained by the reference GC method and the ones obtained by the development of the calibration in NIRS and PLS regression for both sets of data, calibration and validation.

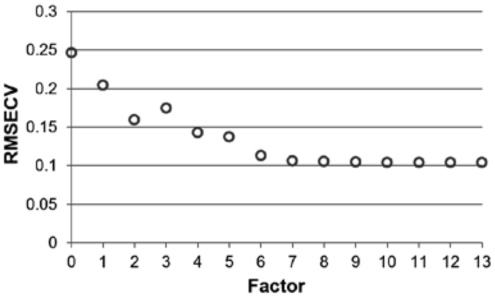
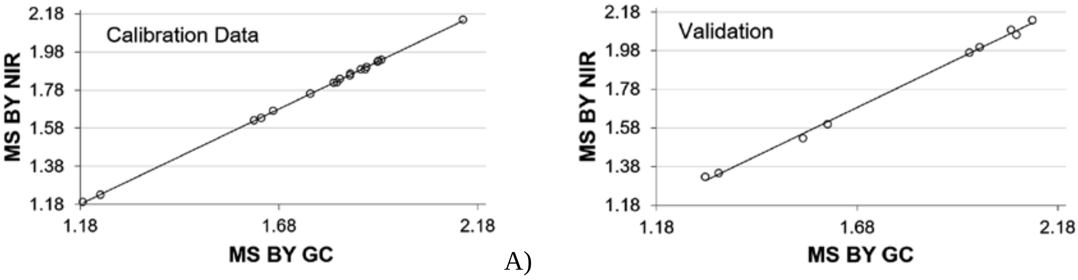


Figure 2: RMSECV (root mean squared error of cross-validation) *vs*. PLS factors in the range of 5708.26-5553.98 $cm^{-1}$



A) B)

Figure 3: MS correlation plots for (A) the calibration data set and (B) the validation data set, by PLS, with 10 factors in the spectral range of 5708.26-5553.98 $cm^{-1}$

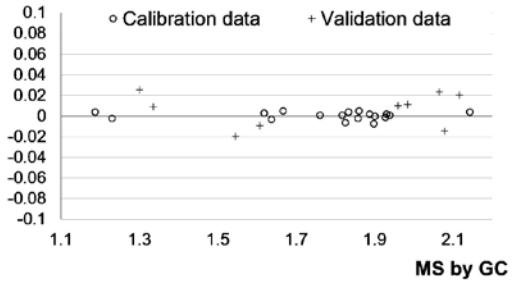


Figure 4: Difference plot for calibration (o) and validation (+) data sets by PLS with 10 factors in spectral range of 5708.26 to 5553.98 cm$^{-1}$

PLS is not a method, it is a set of methods with associated algorithms. To mention some of them: the algorithm Path-PLS, the algorithm PLS of principal component analysis (NIPALS-PCA) or the algorithm PLS for canonic correction analysis (NIPALS-CCA). Something in common for the PLS methods is that all of them have an algorithm associated with a uniform format. The PLS methods proceed in a less intuitive manner compared to classical statistical methods, are not formulated in terms to be optimized algebraically, and no criteria of maximization or minimization are presented. Generally, a model is expressed in a way for identifying components and equations between components. Instead of developing an analytical solution, the development is done until reaching a solution through a series of sequential repetitive steps for obtaining a stable approximation.

Figure 4 represents the difference found in the results between the data obtained by the GC technique and the results calculated by applying the calibration model of the NIR technique. Only three of nine validation points display a difference of 0.0203, 0.0232 and 0.0247, respectively, while the rest of the data had a difference lower than 0.0193, three of them – below 0.01. Prediction errors are calculated for the samples as the difference between prediction and reference values.[37]

Most of the data shown in Figure 4 report minor differences from the standard deviation of the reference method itself; the SD for the chromatography method is 0.02, compared with the results of the difference between the two techniques, NIR represents an equivalent variation against the reference method.

## CONCLUSION

The implementation of the method for determining the MS in HEC in the range from 1.18 to 2.14, using NIR spectroscopy in the spectral range of 5708.26-5553.98 cm$^{-1}$, exhibited excellent correlation with the data obtained by the GC method based on ASTM D 4794-94[12] used as reference for building the calibration model.

The NIR spectroscopy technique only requires a few minutes for measuring and obtaining the MS for a range of 1.18-2.14, with the most important advantages of simplicity, no need for any reaction for the determination, rapidity – saving between 7 and 12 h of the analyst's labor, and providing reliable results, without involving any technical difficulties due to the hermeticity of the required vials. These advantages are translated into a simple manner for obtaining critical data for the industrial reaction process of cellulose ether, having the specifications desired for specific applications.

The implementation of this new alternative provides a tool for Green Chemistry by eliminating the use of chemicals and the risk for humans and the environment represented by the GC methodology.

A new, faster, greener and accurate method for determining molar substitution in hydroxyethyl cellulose has been demonstrated. The method is suitable for an MS range from 1.18 to 2.14; nevertheless, it can be adjusted by constructing a new calibration for higher MS; also, the ideal spectral zone can be identified.

***ACKNOWLEDGMENT***: We thank Grupo Petroquímico Beta for providing funding, samples and resources for this study.

## REFERENCES

[1] J. N. BeMiller and R. L. Whistler, "Industrial Gums: Polysaccharides and Their Derivatives", 3rd ed., Academic Press Inc., USA, 1993, pp. 649-672.

[2] O. Ololade, "Natural Polymers: Industry Techniques and Applications", Springer, Switzerland, 2016, p. 228.

[3] V. Myllymaki, R. Aksela and A. Kemlr, “Method for Preparing a Cellulose Ether”, US 2007/0112185 A1, 2007.

[4] Q. Zhou, L. Zhang, M. Li, X. Wu and G. Cheng, *Polym. Bull.*, **53**, 243 (2005).

[5] H. F. Mark, “Encyclopedia of Polymer Science and Technology”, concise, 3rd ed., John Wiley & Sons, USA, 2007, p. 172.

[6] S. Dumitriu, “Polymeric Biomaterials”, revised and expanded, 2nd ed., CRC Press, New York, 2001.

[7] A. A. Tracton, “Coatings Materials and Surface Coatings”, CRC Press, Boca Ratón, FL, 2006.

[8] F. Pacheco-Torgal, V. Ivanov, N. Karak and H. Jonkers, “Biopolymers and Biotech Admixtures for Eco-efficient Construction Materials”, Woodhead Publishing, USA, 2016, p. 231.

[9] J. Fink, “Petroleum Engineer’s Guide to Oil Field Chemicals and Fluids”, Gulf Professional Publishing, USA, 2012.

[10] C. Márquez Ruíz, “Modelo de regresión PLS”, Universidad de Sevilla, Spain, 2017.

[11] Y. Diao, M. Song, Y. Zhang, L.-Y. Shi, Y. Lv *et al*., *Carbohyd. Polym.*, **169**, 92 (2017).

[12] ASTM D 4794-94, “Standard Test Method for Determination of Ethoxyl or Hydroxyethoxyl Substitution in Cellulose Ether Products by Gas Chromatography”, reapproved in 2009.

[13] C. Jiménez-González and D. J. C. Constable, “Green Chemistry and Engineering. A Practical Design Approach”, John Wiley & Sons, 2011, pp. 17-39.

[14] X. P. Maldague, “Theory and Practice of Infrared Technology for Nondestructive Testing”, John Wiley & Sons, USA, 2001.

[15] D. A. Burns and E. W. Ciurczak, “Handbook of Near Infrared Analysis”, 3rd ed., CRC Press, Boca Ratón FL, 2007.

[16] A. Kauppinen, M. Toiviainen, O. Korhonen, J. Aaltonen, K. Järvinen *et al*., *Anal. Chem.*, **85**, 2377 (2013).

[17] J. G. Foster, W. M. Clapham and J. M. Fedders, *J. Agric. Food Chem.*, **54**, 9 (2006).

[18] L. Fei, H. Yong and G. Sun, *J. Agric. Food Chem.*, **57**, 4520 (2009).

[19] S. Fei, N. Xiaoying, Y. Danting, Y. Yibin, L. Bobin *et al*., *J. Agric. Food Chem.*, **58**, 9809 (2010).

[20] Q. Ouyang, Q. Chen and J. Zhao, *Spectrochim. Acta, Part A*, **154**, 42 (2016).

[21] S. Li, X. Zhang, Y. Shan, D. Su, Q. Ma *et al*., *Food Chem.*, **218**, 231 (2017).

[22] A. Guelpa, F. Marini, A. du Plessis, R. Slabbert and M. Manley, *Food Control*, **73**, 1388 (2017).

[23] J. T. C. Rocha, L. M. S. L. Oliveira, J. C. M. Dias, U. B. Pinto, M. de L. S. P. Marquez *et al*., *Energ. Fuels*, **30**, 1 (2015).

[24] S. Castritius, A. Kron, T. Schafer, M. Radle and D. H. Diedrich, *J. Agric. Food Chem.*, **58**, 12634 (2010).

[25] Q. Zhai, C. Zhou, S. Zhao, C. Peng and Y. Han, *Ind. Eng. Chem. Res.*, **53**, 13598 (2014).

[26] B. Singh, D. F. Malley, A. Farenhorst and P. Williams, *J. Agric. Food Chem.*, **60**, 9948 (2012).

[27] Q. Guo, L. Nie, L. Li and H. Zang, *Spectrochim. Acta, Part A*, **177**, 158 (2017).

[28] X. Xiao, Y. Hou, J. Du, Y. Liu, L. Dong *et al*., *J. Agric. Food Chem.*, **60**, 7830 (2012).

[29] H. Wold, “Estimation of Principal Components and Related Models by Iterative Least Squares. Multivariate Analysis”, Academic Press, New York, 1966, pp. 391-420.

[30] P. Geladi and B. R. Kowalski, *Anal. Chim. Acta*, **185**, 1 (1986).

[31] H. Martens and T. Naes, “Multivariate Calibration”, John Wiley & Sons, New York, 1989.

[32] S. Wold, M. Sjöström and L. Eriksson, *Chemometr. Intell. Lab.*, **58**, 109 (2001).

[33] S. Wold, “PLS for Multivariate Linear Modeling QSAR: Chemometric Methods in Molecular Design. Methods and Principles in Medicinal Chemistry”, Verlag-Chemie, 1994.

[34] A. Hoskuldsson, *J. Chemometr.*, **2**, 211 (1988).

[35] TQ Analyst User’s Guide, Thermo Fisher Scientific Inc., 2007.

[36] J. Workman and Jr. L. Weyer, “Practical Guide to Interpretive Near Infrared Spectroscopy”, CRC Press, USA, 2008.

[37] N. M. Faber, X. H. Song and P. K. Hopke, *Trends Anal. Chem.*, **22**, 5 (2003).