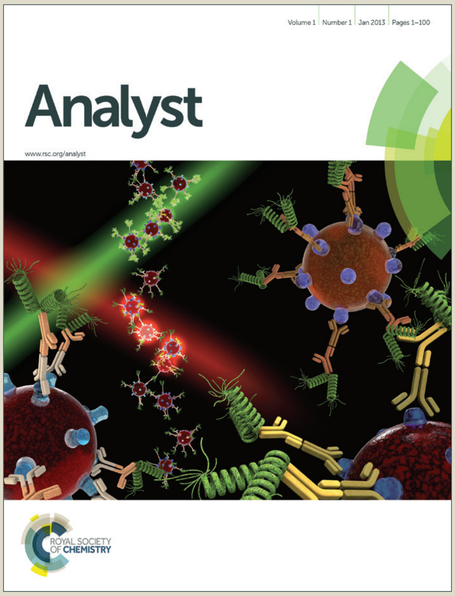
Analyst

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/analyst

Comparing Near-Infrared Conventional Diffuse Reflectance Spectroscopy and Hyperspectral Imaging for Determination of Bulk Properties of Solid Samples by Multivariate Regression: Determination of Mooney Viscosity and Plasticity Indexes of Natural Rubber

Carlos Juliano da Silva*[1] and Celio Pasquini[1]

[1]Chemistry Institute, University of Campinas – UNICAMP, Caixa Postal: 6154, CEP: 13084-971, Cidade Universitária, Campinas , SP, Brazil. Phone: 55 (19) 35213136.

* Corresponding author e-mail: julianocjs@gmail.com

ABSTRACT

Conventional reflectance spectroscopy (NIRS) and hyperspectral imaging (HI) in the Near-Infrared region (1000 – 2500 nm) are evaluated and compared, using, as case study, the determination of relevant properties related to the quality of natural rubber. Mooney viscosity (MV) and plasticity indexes (PI) (PI_0 - original plasticity, PI_{30} - plasticity after accelerated aging, and PRI - the plasticity retention index after accelerated aging), of rubber were determined using multivariate regression models. Two hundreds and eighty six samples of rubber were measured by conventional and hyperspectral near-infrared imaging reflectance instruments in the range 1000 - 2500 nm. The sample set was split into a regression (n = 191) and external validation (n = 95) sub-sets. Three instruments were employed for data acquisition: a line scanning hyperspectral camera, and two conventional FT-NIR spectrometers. Sample heterogeneity was evaluated using hyperspectral images obtained with resolution of 150 x 150 μm and Principal Component Analysis. The probed sample area (5 cm^2; 24,000 pixels) to achieve representativeness was found to be equivalent to the average of 6 spectra for a 1 cm diameter probing circular window of one FT-NIR instrument. The other spectrophotometer can probe the whole sample in only one measurement. The results show that the rubber properties can be determined with very similar accuracy and precision by Partial Least Square (PLS) regression models regardless HI-NIR or conventional FT-NIR produce the spectral data sets. The best Root Mean Square Errors of Prediction (RMSEP) of external validation for MV, PI_0, PI_{30}, and PRI were 4.3, 1.8, 3.4, and 5.3 %, respectively. Though the quantitative results provided by the three instruments can be considered equivalent, the hyperspectral imaging instrument presents a number of advantages, being about 6 times faster than conventional bulk spectrometers, producing robust spectral data by ensuring sample representativeness, and minimizing the effect of the presence of contaminants.

Keywords*:* Near-infrared hyperspectral image, diffuse reflectance, determination of viscosity, determination of plasticity, natural rubber, multivariate regression

INTRODUCTION

Currently, analytical chemists have two main types of instruments available to obtain spectral information based on reflectance measurements in the near-infrared (NIR) region aiming, for example, to determine bulk properties or bulk composition of solid samples by means of multivariate regression techniques. Conventional spectrophotometers comprise instruments capable to obtain spectra by probing relatively large areas of the solid sample. The area probed, required to produce a sample representative single spectrum, can vary from less than one to several squared centimeters, depending on sample homogeneity. Because the correlation with bulk properties is seeking, it has been demonstrated by myriad of works found in the literature that the spectral data obtained by these instruments can supply the necessary information regarding quantitative multivariate techniques, such as Partial Least Square (PLS) regression[1–4].

On the other hand, hyperspectral image instruments (hyperspectral cameras) operating in the near-infrared region, though more expensive, has becoming more and more common in the laboratories. This type of instrument is capable to add information by probing a sample with spatial resolution, recording a whole spectrum for each spatial resolution element (pixel). This additional dimension imparted to the spectral data is employed to approach analytical issues related with analyte distribution or properties dependent on spatial distribution of chemical species such as the homogeneity of pharmaceutical samples[5,6], potential of seeds germination[7,8], and identification and localization of anomalies and impurities present in several types of samples matrices[9,10].

Considering the application of spectral information to construct multivariate models aiming to determine bulk properties of samples, apparently, at first sight, there should have no significant gain in obtaining a hyperspectral NIR image of the sample. In principle, the average spectrum obtained in a conventional instrument should be representative and correlates better with the magnitude of bulk properties than any individual pixel spectrum. Nevertheless, some advantages of spectral imaging over conventional reflectance spectroscopy may be foreseen. The homogeneity of the sample can be tested through image data, and the presence of impurities degrading the determination of bulk properties can be detected and its impact minimized, before producing average spectra to feed a multivariate model or to predict a given sample property. Furthermore, line and/or plane hyperspectral cameras can attain representative measurements faster than the majority of the conventional NIR spectrometers. These potential advantages of image over conventional NIR instruments have not been evaluated yet, considering the use of the spectral information for determination of bulk physical-chemical properties of heterogeneous rubber and rubber products.

In order to compare the image and conventional approach to spectral measurement aiming at determination of bulk properties of solid samples, this work employs natural rubber as test material.

The first step of the industrial processing cycle of several rubber products based on natural rubber consist in coagulate the latex collected in the field, washing, ground and heated to reduce the water content resulting in a raw product named “brown crepe”. The quality of this product is mostly attested by its rheological properties: viscosity and plasticity. Viscosity of the natural rubber is assessed by a well-defined protocol and instrumentation and named Mooney viscosity (MV). The MV corresponds to the shearing torque required to spin a rotor embedded in the rubber sample, heated inside a cylindrical cavity. The procedure to determine this parameter is described by the ASTM test method D1646-07[11]. The routine procedure is carried out in a heavy, complex mechanical instrument demanding constant and expensive maintenance. One MV determination requires typically about 25 minutes. When the product is intended to tires manufacturing, MV values should be in the range 75 – 80.

Plasticity parameters refer to the characteristics of the natural rubber of deforming under the effect of an external mechanical force. The Wallace plasticity, determined in an instrument with the same name (Wallace plastimeter), is a measurement of the thickness variation of a small disc, about 3 mm thick, 2 cm diameter, of rubber, cut from a sheet obtained by passing the raw brown crepe 10 times through heated cylinders of a calender. Three plasticity parameters are used to specify the rubber quality. The initial plasticity index (PI_0) is measured in the original sample of natural rubber. The plasticity index after accelerated aging (PI_{30}) is obtained by carrying on the same measurement after the sample has been heated for 30 minutes at 140 ^{o}C. Finally, the plasticity recovery index (PRI) is obtained by the ratio PI_{30} / PI_0, and expressed in parts percent. The pertinent standard methods describing the procedures for the determination of the plasticity indexes of natural rubber are found in the ASTM D3194-04[12]. Determination of the three rheological parameters is time consuming and requires careful and usually expensive, maintenance of the Wallace plastimeter.

Near infrared spectroscopy has been used for analysis of natural rubber, mainly raw latex, and several works can be found elsewhere aiming at determination of apparent viscosity[13], dry material and solids content[13,14]. On the other hand, it was possible to find only one work attempting to employ NIR spectroscopy and multivariate calibration to predict the MV of natural rubber[15]. In that work, conventional measurements of the sample reflectance in the NIR spectral region and Partial Least Square (PLS) regression are employed to produce models capable to predict MV with a root mean square error of external validation (RMSEP) of about 4 MV units. Regarding the determination of plasticity indexes using NIR and multivariate regression, no work could be found in the literature.

The main objective of this work is to compare the relative performance of imaging and conventional NIR instruments to obtain spectral information required to construct multivariate regression models to determine bulk properties of solid samples. natural rubber is taken as test material to carry out this evaluation and, at the same time, to evaluate multivariate models to predict the entire set of rheological parameters required to characterize the quality of this industrial product.

EXPERIMENTAL

Samples

Two hundred and eighty six samples of natural rubber, collected during a two months period, were provided by Braslátex Ltda., Bálsamo – SP – Brazil. The samples were preprocessed in a calender as recommended by the standard method for determination of plasticity indexes[12], and cut to produce rubber pieces about 4 cm wide, 6 cm long and 3 mm thick. All samples were analyzed by the ASTM recommended methods for their Mooney viscosity (MV)[11] and plasticity indexes (PI_0, PI_{30} and PRI)[12], and have been sent to the Chemistry Institute, UNICAMP, for spectral data acquisition, after a time interval never exceeding 96 hours.

The original sample set was randomly split into a calibration and external validation set containing 191 and 95 samples, respectively, in order to obtain representative values of the rheological parameters modeled in both sets. Table I depicts the relevant information of the sample sets.

Instrumentation

Three near-infrared instruments were employed in this study. Two are conventional Fourier transform interferometric based instruments (FT-NIR), thereafter referred as NIR-FT(1) and NIR-FT(2). NIR-FT(1) (Bomem, MB164) has been equipped with a reflectance accessory (Powder NIR) that presents a circular probing window of 1 cm diameter on what the rubber sheet is placed in order to measure its reflectance spectrum. NIR-FT(2) (AIT, Diamond 20) was also equipped with a reflectance accessory. However, this accessory has a wider circular window, effectively probing about 15 cm^2 of the sample surface. Therefore, each reflectance measurements made in this instrument represents the whole rubber sample. Both instruments were set up to work in the spectral region 1000 – 2500 nm with resolution of 8 cm^{-1}. However, for the NIR-FT(2) instrument the useful spectral range is from 1000 – 2200 nm. The reflectance spectra were obtained as an average of 50 scans of the each probed area of the samples. The time necessary to obtain one spectrum using any equipment is about 1 min. Samples were pressed against the probing windows of the instruments by a PTFE cylinder 10 cm diameter and 5 cm high.

The third instrument is a line scan hyperspectral camera (SisuChema, SWIR) operating in the same spectral region of the conventional instruments. The pixel size (spatial resolution) was set to 150 x 150 µm. The spectral resolution was of 10 nm and 256 spectral channels are acquired per pixel. The samples of rubber sheets were placed on the instrument tray, which moves under the illuminated line of the instrument at a speed that allows for reading a 40 cm long sample in about 1 min. A typical rubber sample produces an image of 250 x 400 pixels covering the whole sample, and is acquired in about 15 s.

Data processing

Spectral data pretreatment and multivariate analysis and regression algorithms were processed using the Unscrambler 10.3 chemometric software (CAMO, Norway).

RESULTS AND DISCUSSION

Evaluation of sample heterogeneity through multivariate analysis of hyperspectral images

The intra-sample heterogeneity was evaluated taking the second derivative of the spectra set representing the whole hyperspectral image (100,000 pixels) of a typical 4 x 6 cm rubber sample. This pretreatment removes most of the data variability related to radiation scattering promoted by the irregular surface of the sample. The pretreated spectra were submitted to a Principal Component Analysis (PCA). The results show that the first two principal components explain 38 and 17% of data variability, respectively. Scores values assume a normal distribution centered at zero, because the spectra data has been mean centered before PCA. Other pretreatment such as Multiplicative Signal Correction (MSC) and Standard Normal Variate (SNV) aiming to minimize the same scattering effects were evaluated, with similar results.

To evaluate the sample heterogeneity the averages of the scores values obtained by several randomly selected sub-sampling areas equivalent to 1000, 2000, 5000, and 24,000 pixels, were compared with the average of the scores for the 100,000 pixels, representing the whole sample. Figure 1 shows the average values of score for the first and second principal components as a function of the number pixels included in the subsample. As can be noted, the scores averages show a wide dispersion of values around the expected mean value (zero) when the subsampled area is small, equivalent to 1000 and 2000 pixel, for example. The average values converge to the expected zero value, showing a narrow dispersion only when the subsampled area is equivalent to about 24,000 pixels, or ~5 cm^2. The same behavior of the average values was observed for 5 other rubber samples randomly chosen and submitted to the same procedure. It means that, in order to represent adequately the sample, and avoid a severe effect of sample intra-heterogeneity, average spectra must be obtained by probing an area of at least 5 cm^2 of the sample.

Evaluation of spectral representativeness for conventional instruments

When the sample spectra is obtained by using the conventional NIR-FT(1) equipment (the most usual configuration), an area of only 0.85 cm^2 is probed through its 1.0 cm diameter circular window, where the reflectance measurements are made.

Therefore, and according to the results for the sample heterogeneity showing above, this spectrum may not be representative of the sample, and its correlation with the value of a bulk property measured by a reference method can be in risk.

To evaluate the minimum number of spectra necessary to produce a representative average, six spectra of 103 samples of natural rubber were obtained in six different locations of the same sample sheets. The data were employed to produce subsets of averages of 2, 4, 5, and 6 spectra, representing each sample. Spectral data of 69 samples has been treated by second derivative and partial least square (PLS) regression models for MV were generated employing the average spectra calculated with different numbers of individual spectra. In addition, models were constructed by using only one spectrum representing each sample.

Furthermore, data obtained from the hyperspectral camera were also employed to construct PLS regression models for MV. In each case, the spectra corresponding to the number of pixels (24,000) equivalent to the total area probed by the conventional NIR-FT(1) spectrometer was averaged. The same pretreatment (2nd derivative) was employed. All models were validated by an external set containing spectra of 34 samples.

Figure 2 shows the behavior of the root mean square errors of prediction (RMSEP) (external validation) as a function of the number of spectra and number of pixels included for average. Observe that the number of combinations possible of the six original spectra depends of the number of spectra included in the average. Of course, it was possible to construct only one model using the average of the six spectra obtained for each sample. The spectra employed for average (or to be included in the calibration set when only individual spectra were employed) were taken randomly. For comparison, the results (RMSEP) for the average of 4,000 pixels and 24,000 pixels registered for each rubber sample images was included in the Figure 2. These number of pixels is equivalent to one measurement and to the average of six measurements of the same sample obtained by the NIR-FT(1), respectively.

The RMSEP values converge to a stable and low value only for the models constructed by employing averaged spectra sets calculated by using at least 5 individual spectra obtained in different portions of the same sample or when the equivalent number of pixels of the hyperspectral image is employed (20,000). These results also confirm the heterogeneity of the rubber samples and are in agreement with the results found for the image analysis of sample heterogeneity reported above.

Therefore, in order to be sure to work with spectra sets truly representing the rubber samples, averages of 6 spectra were taken whenever the equipment NIR-FT(1) was used, and averages of the spectra of at least 24,000 pixels, when the image system was employed.

It is relevant, at this point, to observe that the hyperspectral image analysis of typical samples designated to bulk analysis and multivariate calibration is very useful and save time in defining the number of measurements necessary to achieve representativeness of heterogeneous samples measured in conventional instruments. In the present case, the analysis of intra-heterogeneity by means of image analysis was carried out in about 3 working-hours, while the use of conventional instrument required about 32 working-hours to achieve the same result.

Evaluation of multivariate models

Once the necessary number of measurements to achieve a representative average was established, PLS models were developed to predict the rheological properties of prevulcanized rubber by using data produced by the three instruments. The average spectra of 191 samples were employed as calibration sets to produce PLS regression models for MV, PI_0, PI_{30} and PRI. It must be emphasized that the NIR-FT(2) instrument has a large probing area that allows access the whole sample surface with only one measurement. Image data refer to the average of 100,000 pixels, equivalent to one measurement taken by the NIR-FT(2) spectrophotometer. During model development, some samples were found as outliers and eliminated from the calibration set, based on their residues and leverages values, which exceeded the usual threshold limits[14]. The final number of samples employed for each instrument and bulk property was: FT-NIR(1), 179, 181, 182, 179; FT-NIR(2), 182, 184, 183, 182, and NIR-HI, 185, 183, 186, 183, respectively for MV, P_0, P_{30}, and PRI.

Separate sets of spectra obtained from additional 95 samples were employed for external validation of the models. All spectra were pretreated by second derivative (Savitz-Golay, 33 and 13 points window for NIR-FT(1 and 2), and image camera, respectively; 2nd degree polynomial). All 95 samples were always employed for external validation as no outlier has been detected in this group.

The PLS regressions were developed using full cross-validation to help establishing the optimal number of factors to model the rheological properties of rubber. By using cross-validation, the Unscrambler software can perform a Jack-knife algorithm that evaluates the significance of the regression coefficients of PLS models[15]. The variables generating significant coefficients are marked by the program and new PLS models were constructed by employing only these variables. The new models do not improve significantly the RMSEP of external validation. However, all models had their optimal number of factors, required to achieve the lowest RMSEP, reduced at least by one unit, being more parsimonious, and certainly contributing for the model robustness.

Table II summarizes the principal results obtained by the optimized PLS models developed to quantify the MV, PI_0, PI_{30} and PRI in natural rubber. The RMSEP obtained for the external validation were within or close to the repeatability of the standard methods employed in the rubber industry to determine the rheological parameters of the natural rubber. References 10 and 11 mention absolute repeatability of 3.4 and 6.9 %, and absolute reproducibility of 5.4 and 24.3% , for MV and PRI, respectively. The coefficients of determination (R^2) achieved for external validation are in the range 0.269 – 0.760. The lower values are associated to the PRI and P30 determination, and are in agreement with the relative narrow range of references values employed for model construction and validation and the repeatability of the reference methods[11,12]. Therefore, the results obtained for PRI and P30 must be considered only as semi-quantitative. On the other hand, the estimated errors of the models for PRI and P30 were close to the reference method repeatability and were judged sufficient to be employed in routine analysis of natural rubber.

The repeatability of PLS models have been evaluated for two instruments by employing 6 measurements of each of 3 samples of preprocessed rubber. The averages of the estimated standard deviation for the instruments, and for MV, P_0, P_{30} and PRI, respectively, are: HI-NIR (1.4, 0.9, 1.0, and 1.5 %); FT-NIR(1) (1.6, 1.0, 0.9, and 2.2 %).

As shown in Table III, there is no significant difference in the performance of the models. Apparently, the poorer spectral resolution of the imaging instrument does not impart any degrading effect in the performance of the multivariate models. It has been reported in the literature that, for reflectance of solid samples containing wide spectral absorption bands, the resolution is not an issue[16].

Up to this point, it is possible to affirm that, after proper evaluation of sample heterogeneity, the image approach would offer no major advantage over conventional spectrophotometers for determination of bulk properties. On the other hand, it is important to mention that an image instrument is faster and can assist considerably the definition of the measurement protocol for conventional spectrophotometers aiming to obtain representative spectral data for multivariate regression, as demonstrated above.

Interpretation of the models

The rheological parameters of preprocessed natural rubber is greatly dependent on the chemical characteristics of the main natural polymer, cis-1,4-polyisoprene, whose chemical structure is depicted in Figure 3, representing about 96% of the dry material found in latex, and on the effect of processing parameters, such as drying temperature and mechanical treatment. The relative importance of the variables and their respective regression coefficients should reflect this dependence.

Throughout the development of the multivariate models showed in this work a small difference in performance was observed when the spectra set was pre-processed by MSC or second derivative. On the other hand, the interpretation of the relative importance of the variables is greatly facilitated if the original spectral (absorbance) dimension is preserved. Therefore, the following discussion was based on the behavior of the regression coefficients for the models produced using the spectra treated only by MSC. The interpretation made by using the second derivative results are in agreement with that presented below for the spectra treated by MSC. Figure 5 shows a typical spectrum of natural rubber sample and the relevant regression coefficients for prediction of the four rheological parameters. Looking at the regression coefficient values of the variables selected by the Jack-knife algorithm[16] for the validated models, is possible to interpret their relevance according to the expected changes in the chemical composition and their relationship with rheological parameters.

Considering the Mooney Viscosity (MV), it is well known that preprocessing the raw material by mechanical mastication in presence of oxygen cleaves the isoprene chain and generates hydroperoxide radicals, which posteriorly can react with double bonds of the main chain, producing lateral chains (branches)[17]. The overall effect is the reduction of the average length of the isoprene chain, resulting in lower values of MV. The resulted structural changes can be observed in the NIR reflectance spectrum by the increasing absorption around 1441 nm. This wavelength is associated to the combination band of the 1st overtone of stretching with on the plane bending of the $-CH_2$ groups present in the branched aliphatic chains [$RC(CH_3)_3$) or $RCH(CH_3)_2$], as described in page 243 of reference [18]. The regression coefficients observed in the region around 1411 nm agree with this fact.

Another evidence of main chain cleavage and branching can be noticed around the wavelength 2270 nm, usually attributed to the strong absorption of the $-CH_3$ functional groups, as described in page 27 of reference 17. Considering the aspects related above, cleavage and branching of the main chain can result in a significant increase in the number of terminal $-CH_3$ groups already detected in the NIR spectrum of the pre-processed rubber, producing significant regression coefficients whose values agree with the expected change (reduction) of MV.

In the region of 1396 nm it is observed the absorption of methyl groups associated to branched aliphatic chains [$RC(CH_3)_3$) or $RCH(CH_3)_2$], due the combination of the 1st overtone with the angular bending of $- CH_3$, as described in page 243 of reference 17. This fact is also in agreement with the observed behavior of the regression coefficients close to this spectral region.

The MV can be also associated with the amount of cross-link between the isoprene polymeric chains. Cross-link is favored by the presence of nitrogen containing compounds such as proteins, peptides and aminoacids[19]. The spectral region 1990 to 2180 nm corresponds to absorption bands associated to the combination of vibrational modes of these type of compounds. Significant regression coefficients to the prediction of MV were also found in this spectral region.

Finally, the spectral range 1630 – 1637 nm shows absorption bands associated to C – H of vinyl groups. The band at 1630 corresponds to the isoprene vinyl C – H group [$-CH{=}C(CH_3){-}CH{=}CH-$], as described in page 249 of reference 17. The values of regression coefficients experimentally observed in this region are directly correlated to MV, corroborating the hypothesis of great number of isoprene vinyl C – H groups present in naturally long chains.

Considering the initial plasticity (P_0), it can be shown that it is highly correlated to MV. Therefore, the regression coefficients of the PLS model for this property reflects, with great similarity, the changes in the same chemical characteristic as for

MV. In fact, the same variables reported above, with similar relative values of regression coefficients, were found relevant for determining P_0 in rubber samples.

In the case of P_{30} and PRI, other chemical characteristics become relevant. The resistance of the rubber to thermal degradation is highly dependent on the presence of the anti-oxidants in the raw material. The amines are, among several other compounds found in latex, the ones showing anti-oxidant characteristics and present in relative high concentration. Not coincidently, the significant regression coefficients associated to the determination of P_{30} and PRI are found in the spectral region (2000 – 2180 nm) where the amine functional groups show well know absorptions in the near-infrared[18].

Of course, a full interpretation of the regression models should consider other effects reflected in the NIR spectra due secondary and tertiary effects at microscopic and even macroscopic level[20]. On the other hand, the evidences above serve to demonstrate unequivocally the correlation between the spectral information employed at the modeling stage with the rheological properties of preprocessed rubber.

Effect of the presence of contaminants

Samples collected from industrial processes are often subject to contamination caused by contact of the products with impurities found in the manufacturing ambient. Vegetable materials (for example, leaves, and bark fragments), normally present in a latex processing plant, can contaminate rubber samples. Diffuse reflectance measurements of solid samples can be severely affected by the presence of contaminants or impurities on the surface of the sample exposed to the probing area of the instrument. Significant errors can result by using inaccurate spectra, altered by the presence of contaminants, to predict bulk properties through multivariate regression models.

However, the effect of impurities has not been previously evaluated or compared for conventional and imaging instruments. To simulate the presence of impurities, fragments of dry leaves and barks were placed on the surface of several samples of preprocessed rubber and their reflectance spectra were measured. Figure 4 shows some typical spectra obtained by the three instruments in the presence and absence of contaminant. The resulting spectra were employed to predict the rheological parameters. The results obtained in the presence and absence of impurities can be observed in Table III and IV.

Table III shows that spectra obtained in the presence of impurities by both NIR-FT spectrophotometers predict the rheological properties with large errors. However, imaging results show very low differences between the results obtained in the presence or absence of impurities, as shown in Table IV.

The example of the small fragment of vegetable leaf adhered to the rubber sample surface can be used to explain why these results were obtained. The leaf fragment occupies about 0.65 cm^2 of the 22 cm^2 total surface area of a typical rubber sample. For spectra obtained using conventional instruments, the effect is notable because, for example, the 0.65 cm^2 accounted by the impurity will eventually represent one of the six spectra averaged when the NIR-FT(1) instrument is employed. The effect is reduced in the NIR-FT(2) spectrophotometer. However, the presence of the contaminant also affects considerably the reflectance spectra and the predicted results. On the other hand, imaging data is averaged over 100,000 spectra collected from the total sample surface, from which only about 3000 are affected by the presence of the impurity. Therefore, the effect is attenuated after averaging, as the majority of the spectra are representative of the sample, not of the contaminant.

Of course, previous analysis of the six spectra obtained by the NIR-FT(1) instrument can detect and exclude the outlier spectra affected by contamination, before averaging, considering that one very distinct spectrum will be obtained when most of the probing area of the instrument is occupied by the contaminant. This procedure, however, is not feasible if the impurity is randomly distributed on the sample surface, as the individual spectra will appear very similar one each other. In addition, the use of the procedure is not possible for the NIR-FT(2) instrument, because the whole sample is probed by only one measurement. Furthermore, an evaluation of the average spectrum representative of the sample could be made to detect any significant difference relative to the spectra set employed in the modeling stage. However, if a significant difference is detected, the analysis is lost.

The results shown above claim a superior performance for the imaging instrument over conventional ones in the presence of contaminants. Nevertheless, imaging results can be further improved because a multivariate pre-treatment by using Principal Component Analysis (PCA) can identify those spectra affected by the presence of the contaminant either localized or dispersed on the sample surface. In addition, the effect of these spectra can be prevented by removing them from the data set before calculating the average spectrum, as shown below.

Figures 6A-B depict chemical images based of the scores values for each pixel obtained for the first and second principal components after a PCA analysis of the whole spectra set of a typical rubber sample containing a small piece of leaf on its surface. The 1st and 2nd PC explain 56 and 12% of the data variability, respectively. Figure 6C shows the score plots for the 1st and 2nd PCs and the grouping pattern defined by the spectra of the sample and those of the contaminant. Figure 6D shows the distribution of the scores values as bar graphs for the first PC, which reflect the presence of two type of spectra (scores) population in the data set. The user or an automatic software can easily identify a threshold value to be employed as criteria to remove the spectra whose score values are lower (or higher) than the limit value, largely releasing the spectra set representing the sample from the effect of the contaminant. Figures 6E-F show the score images after removing the contaminated spectra. Figure 6G allows observing the score plot for the first and second PC and Figure 6H shows the similarity of the distribution of the score values for the sample after removal of the contaminated spectra.

In the present case, the level of impurities tested does not disturb significantly the average spectra calculate either by using the complete set of spectra or by using the spectra sub-set resulted from removing the spectra associated with contaminants. Therefore, as shown in Table IV, the results are not significantly different in the presence or absence of contaminants, though a tendency towards lower error values can be observed. However, for higher levels of contaminants, the effect could be serious and the proposed procedure can help to minimize its impact on the predicted results. Further experiments can determine the threshold value of the ratio between the number of spectra of the sample and contaminants beyond which removal of the contaminant spectra becomes necessary in order the models can produce reliable results. Anyway, in the present case, it will necessary to preserve the minimum number (20,000) spectra of the sample matrix to achieve representativeness.

The proposed procedure does not depend on the type of contaminant present in the sample, as long it have a spectrum significantly different of that of the sample. Even mixtures of more than one contaminants and a unknown contaminant can, probably, be tolerated.

CONCLUSION

Contrary to the common sense, which suggests that imaging techniques can find useful and justified application by approaching only analytical issues associated with spatial localization of chemical species, this work demonstrates that NIR hyperspectral imaging techniques present several advantages over the conventional diffuse reflectance measurements of solid samples, even when the determination of bulk composition and/or physical-chemical characteristics are sought.

The main advantage came out from the higher speed (up to 6 times) with which the representativeness of the spectral data employed in multivariate regression can be obtained using hyperspectral cameras, such as the line-scan instrument employed in this work, and from the possibility of identify and minimize the effect of contaminants, which can affect significantly the results of bulk properties predicted by multivariate models. Of course, this advantage is obtained only when conventional or FT-based spectrophotometers are supplied with reflectance sampling devices capable to probe a small sample area.

The main disadvantage of using imaging instruments is its current high cost, when compared with the conventional FT based instruments. However, the cost of image instruments is decreasing over the years and it could be anticipate that the cost/benefit ratio of using this type of instrument can reach attractive values in the near future. In addition, imaging instruments show promising for in/on line monitoring of industrial processes, also aiming at determination of bulk properties.

On the other hand, imaging data are, in any instance, relevant to aid the development of spectral data acquisition protocol for conventional NIR spectrophotometers. The information extracted from images can positively guide the data acquisition by evaluating the sample intra-heterogeneity. Therefore, the area of the sample to be probed by conventional spectrophotometers to attain the necessary representativeness can be promptly established.

At the same time this work demonstrates the advantages of using imaging instruments for determinations of bulk properties, the potential of NIR reflectance measurement to determine the whole set of rheological properties characterizing the pre-processed natural rubber has been also demonstrated for the first time. The Mooney viscosity and Wallace plasticity indexes (P_0, P_{30} and PRI) can be determined with accuracy level demanded to attest the quality of the natural rubber, employing spectral information supplied by any of the NIR instruments used in this work, allowing for cost reduction and expeditious quality control in the rubber industry.

ACKNOWLEDGEMENTS

This is a contribution of the INCTAA (FAPESP, proc. no. 2008/57808-1 and CNPq, proc. no. 573894/2008-6). C.J.S. is grateful to CNPq for the PhD fellowship (proc. No. 130688/2012-6). The authors are grateful to BRASLÁTEX and Mr. Benildes Prince for providing the rubber samples.

REFERENCES

1. H. J. S. Shenk and B. G. Osborne, *J. Near Infrared Spectrosc.*, 2006, **114**, 93–101.
2. G. Fox and M. Manley, *J. Agric. Food Chem.*, 2009, **57**, 5647–57.
3. A. F. C. Pereira, M. J. C. Pontes, F. F. G. Neto, S. R. B. Santos, R. K. H. Galvão, and M. C. U. Araújo, *Food Res. Int.*, 2008, **41**, 341–348.
4. I. Alig, B. Steinhoff, and D. Lellinger, *Meas. Sci. Technol.*, 2010, **21**, 1–19.
5. M. L. Hamad, C. D. Ellison, M. A. Khan, and R. C. Lyon, *J. Pharm. Sci.*, 2007, **96**, 3390–3401.
6. J. G. Rosas and M. Blanco, *J. Pharm. Biomed. Anal.*, 2012, **70**, 691–699.
7. C. M. McGoverin, P. Engelbrecht, P. Geladi, and M. Manley, *Anal. Bioanal. Chem.*, 2011, **401**, 2283–9.

8. M. Arngren, P. W. Hansen, B. Eriksen, J. Larsen, and R. Larsen, *J. Agric. Food Chem.*, 2011, **59**, 11385–94.

9. P. Vermeulen, J. a Fernández Pierna, H. P. van Egmond, J. Zegers, P. Dardenne, and V. Baeten, *Anal. Bioanal. Chem.*, 2013, **405**, 7765–72.

10. J. a. Fernández Pierna, P. Vermeulen, O. Amand, a. Tossens, P. Dardenne, and V. Baeten, *Chemom. Intell. Lab. Syst.*, 2012, **117**, 233–239.

11. ASTM D1646 07, *Standard Test Methods for Rubber—Viscosity, Stress Relaxation, and Pre-Vulcanization Characteristics (Mooney Viscometer)*, 2007, vol. 07.

12. ASTM D3194 04, *Standard Test Method for Rubber From Natural Sources - Plasticity Retention Index*, 2004, vol. 04.

13. P. Sirisomboon, R. Chowbankrang, and P. Williams, *Appl. Spectrosc.*, 2012, **66**, 595–599.

14. P. Sirisomboon, M. Deeprommit, W. Suchaiboonsiri, and W. Lertsri, *J. Near Infrared Spectrosc.*, 2013, **21**, 269.

15. C. Pasquini, F. C. Figueiredo, and B. Prince, *Spectrosc. Lett.*, 2005, **38**, 741–748.

16. H. Martens and M. Martens, *Food Qual. Prefer.*, 2000, **11**, 5–16.

17. G. M. Bristow, *J. Polym. Sci.*, 1962, **62**, 168–171.

18. J. Workman and J. L. Weyer, *Practical Guide to Interpretative Near-Infrared Spectroscopy*, Taylor & Francis Group, 1th edn., 2007, vol. 22.

19. P. W. Allen and G. M. Bristow, *J. Appl. Polym. Sci.*, 1963, **7**, 603–615.

20. P. Williams and K. Norris, *Near Infrared Technology in the Agricultural and Food Industries*, American Association of Cereal Chemists, Inc., Sant Paul, 2nd edn., 2001.

Captions for figures

Figure 1. Effect of the number of pixels on the average values of scores for the first (A) and second (B) principal component resulting from a PCA carried out on the whole spectra set of a typical rubber sample measured by the NIR hyperspectral camera.

Figure 2. Effect of the number of spectra (measurements) take for average on the RMSEP of PLS models constructed aiming at the determination of Mooney viscosity.

Figure 3. Chemical formula of the cis-1,4-polyisoprene rubber, the predominant species in latex.

Figure 4. Spectra obtained in the presence and absence of contaminant (vegetable leaf piece) on the rubber sample by the HI-NIR (A), FT-NIR(1) (B), and FT-NIR(2) (C) instruments.

Figure 5. Regression coefficients of the PLS models. Typical spectrum of the natural rubber (A), regression coefficients for MV (B), regression coefficients for PI_0 (C), regression coefficients for PI_{30} (D) and regression coefficients for PRI (E).

Figure 6. Chemical images based on score values obtained by PCA of the whole spectral data set of a rubber sample containing a small impurity (leaf piece) for the first (A) and second principal component (B); plot of the scores values for the first and second principal component after analysis carried out on the whole spectral data set (C); distribution of score values for the first principal component of the whole data set (D); E, F, G, and H, the same significance as A, B, C, and D, after the scores with values above 0.02 had been removed in D and the remaining data had been submitted to a new PCA, releasing the data set from most of the contaminated pixels.

Table I. Characteristics of the sample set of preprocessed rubber employed in this work.

		Rheological Parameter			
		MV	**Po**	**P30**	**PRI (%)**
Number of Samples	**CV**	191	191	191	191
	P	95	95	95	95
Minimum value	**CV**	54.1	21.3	10.3	41.4
	P	56.6	21.7	12.0	41.9
Maximum value	**CV**	97.8	48.3	31.3	79.3
	P	96.1	45.3	29.3	72.8
Average	**CV**	79.5	33.5	21.0	62.5
	P	79.6	33.4	20.4	61.1
Standard Deviation	**CV**	9.6	5.3	4.5	7.4
	P	9.5	5.2	4.5	7.1

CV – used for cross-validation of the models P – used for external validation (prediction)

Table II. Main results obtained by the PLS models constructed to determine the rheological characteristics of preprocessed rubber with spectral data produced by the three types of NIR instruments.

Rheological parameter	**Cross Validation**			**External Validation**		
	NIR-FT(1)	**NIR-FT(2)**	**NIR-HI**	**NIR-FT(1)**	**NIR-FT(2)**	**NIR-HI**
	RMSECV R^2 Bias PCs	RMSECV R^2 Bias PCs	RMSECV R^2 Bias PCs	RMSEP R^2 Bias PCs	RMSEP R^2 Bias PCs	RMSEP R^2 Bias PCs
MV	4.4 0.787 0.0 7	4.7 0.745 0.0 8	4.7 0.753 0.0 7	4.8 0.742 1.0 7	5.3 0.686 0.2 8	4.6 0.760 0.4 7
Po	2.7 0.737 0.0 6	2.8 0.692 0.0 8	2.6 0.741 0.0 6	2.9 0.649 0.0 6	3.4 0.555 0.2 8	2.9 0.672 0.3 6
P30	2.70 0.627 0.0 6	3.1 0.521 0.0 6	3.0 0.544 0.0 6	3.0 0.552 -0.2 6	3.7 0.336 0.2 6	3.2 0.506 0.4 6
PRI (%)	5.3 0.403 0.0 8	5.6 0.326 0.0 6	5.0 0.370 0.0 7	5.3 0.400 0.1 8	5.8 0.286 -0.8 6	5.7 0.269 1.2 7

Table III. Results obtained for the prediction of rheological parameters of rubber samples employing spectra obtained by the conventional NIR spectrophotometers with and without the presence of impurities.

		FT-NIR (1)							**FT-NIR (2)**						
Impurity	**RP**	**Sample without contaminant**			**Sample with contaminant**			**AMD**	**Sample without contaminant**			**Sample with contaminant**			**AMD**
		1	2	3	1	2	3		1	2	3	1	2	3	
Leaf	**MV**	65.1	66.4	77.1	10.8	18.4	25.4	51.3	70.2	66.3	82.1	47.1	59.8	65.4	15.4
	Po	24.8	25.7	31.9	19.5	22.2	24.8	5.3	27.8	28.4	37.2	14.9	21.0	25.1	10.8
	P30	17.0	21.3	20.3	-14.4	-5.4	-8.8	29.1	22.8	23.3	26.9	33.9	22.9	36.2	6.9
	PRI (%)	60.5	71.0	60.6	-15.2	-3.0	-4.6	71.6	68.2	71.7	69.4	29.9	53.2	31.2	30.9
Bark	**MV**	65.1	66.4	77.1	21.7	24.3	31.7	43.6	70.2	66.3	82.1	31.4	30.8	43.5	37.6
	Po	24.8	25.7	31.9	15.6	26.0	21.3	6.7	27.8	28.4	37.2	4.6	8.0	17.7	21.0
	P30	17.0	21.3	20.3	-3.4	-4.4	-0.3	22.2	22.8	23.3	26.9	37.6	40.7	35.2	13.5
	PRI (%)	60.5	71.0	60.6	24.9	17.1	25.7	41.5	68.2	71.7	69.4	42.1	51.0	45.6	22.8

RP = Rheological Parameter AMD = Absolute Mean difference

Table IV. Results obtained for the prediction of rheological parameters of rubber samples employing spectra obtained by the NIR hyperspectral imaging instrument with and without the presence of impurities

		HI-NIR										
	RP	**Sample without contaminant**			**Sample with contaminant**			**AMD**	**Sample with contaminant removed**			**AMD**
		1	2	3	1	2	3		1	2	3	
Leaf	**MV**	64.8	62.4	77.6	67.2	64.0	74.1	2.5	68.1	64.7	73.5	3.2
	Po	27.7	28.1	36.5	31.6	30.3	37.5	2.4	30.7	29.3	35.4	1.8
	P30	18.5	20.9	20.7	16.4	17.0	19.0	2.6	18.7	19.4	20.6	0.6
	PRI (%)	54.9	59.9	53.5	55.2	54.7	51.5	2.5	58.9	58.0	53.8	2.1
Bark	**MV**	64.8	62.4	77.6	66.6	67.8	76.9	2.6	66.1	64.7	75.0	2.1
	Po	27.7	28.1	36.5	32.8	32.8	39.9	4.4	30.5	30.7	37.6	2.2
	P30	18.5	20.9	20.7	17.2	17.2	17.7	2.7	16.7	17.4	18.1	2.6
	PRI (%)	54.9	59.9	53.5	68.4	68.3	63.5	10.6	61.9	65.0	59.3	6.0

RP = Rheological Parameter AMD = Absolute Mean difference

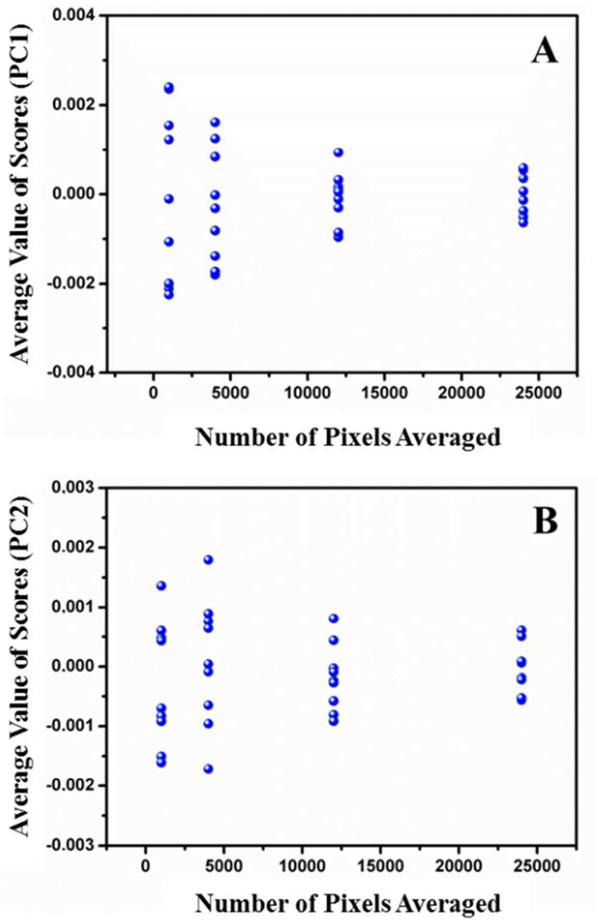


Figure 1. Effect of the number of pixels on the average values of scores for the first (A) and second (B) principal component resulting from a PCA carried out on the whole spectra set of a typical rubber sample measured by the NIR hyperspectral camera.

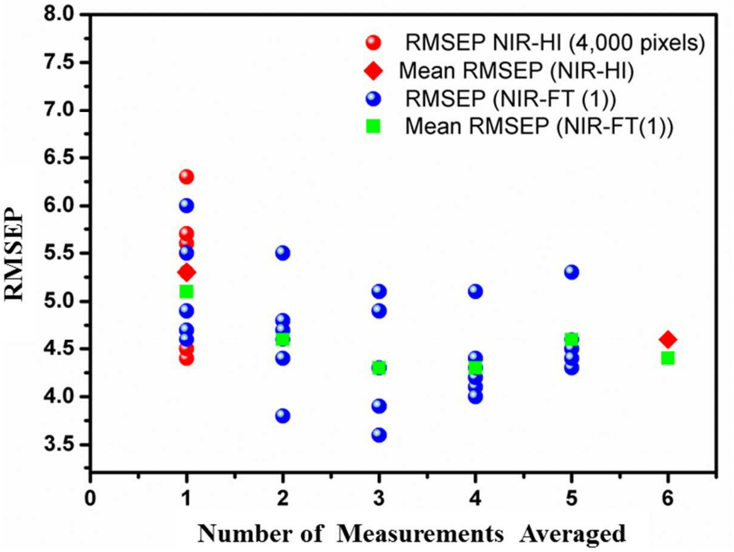


Figure 2. Effect of the number of spectra (measurements) take for average on the RMSEP of PLS models constructed aiming at the determination of Mooney viscosity.

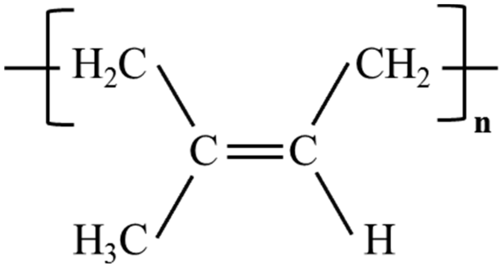


Figure 3. Chemical formula of the cis-1,4-polyisoprene rubber, the predominant species in latex.

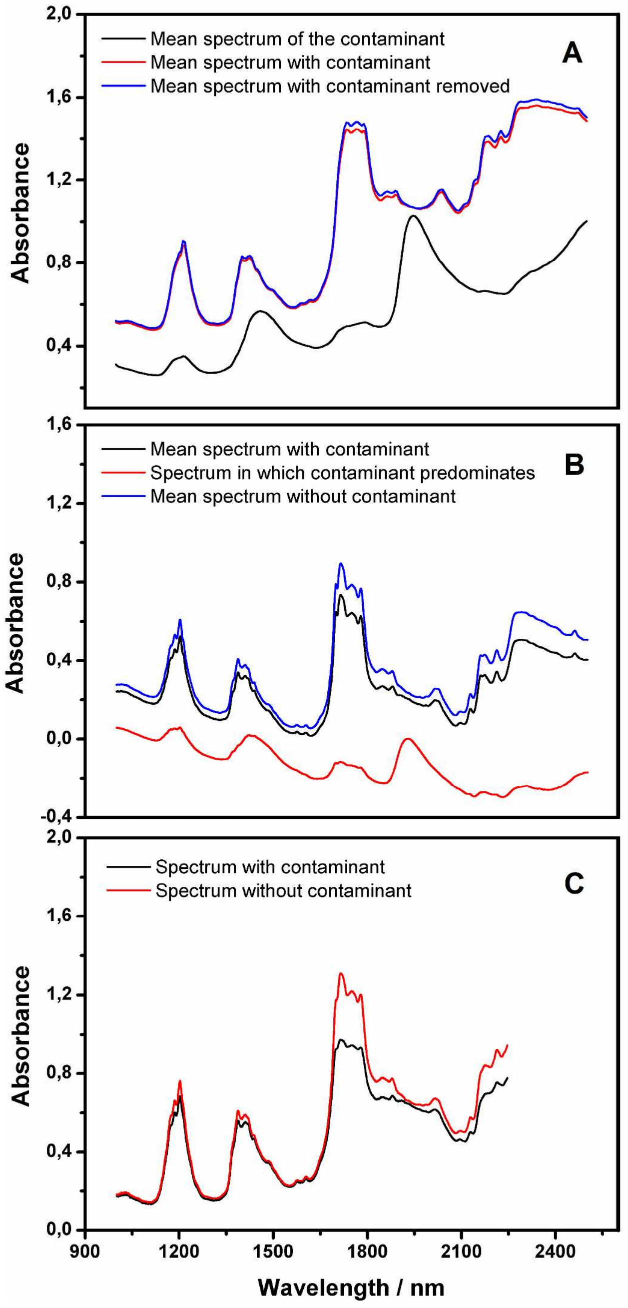


Figure 4. Spectra obtained in the presence and absence of contaminant (vegetable leaf piece) on the rubber sample by the HI-NIR (A), FT-NIR(1) (B), and FT-NIR(2) (C) instruments.

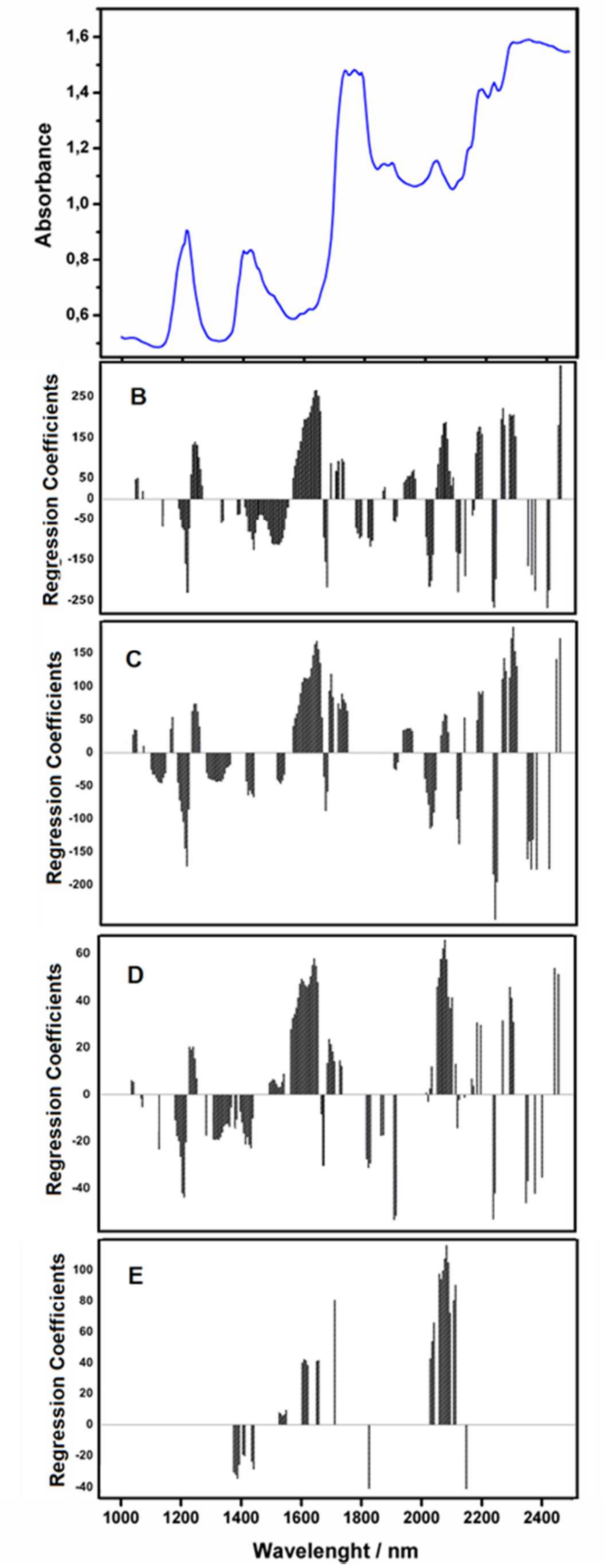


Figure 5. Regression coefficients of the PLS models. Typical spectrum of the natural rubber (A), regression coefficients for MV (B), regression coefficients for PI_0 (C), regression coefficients for PI_{30} (D) and regression coefficients for PRI (E).

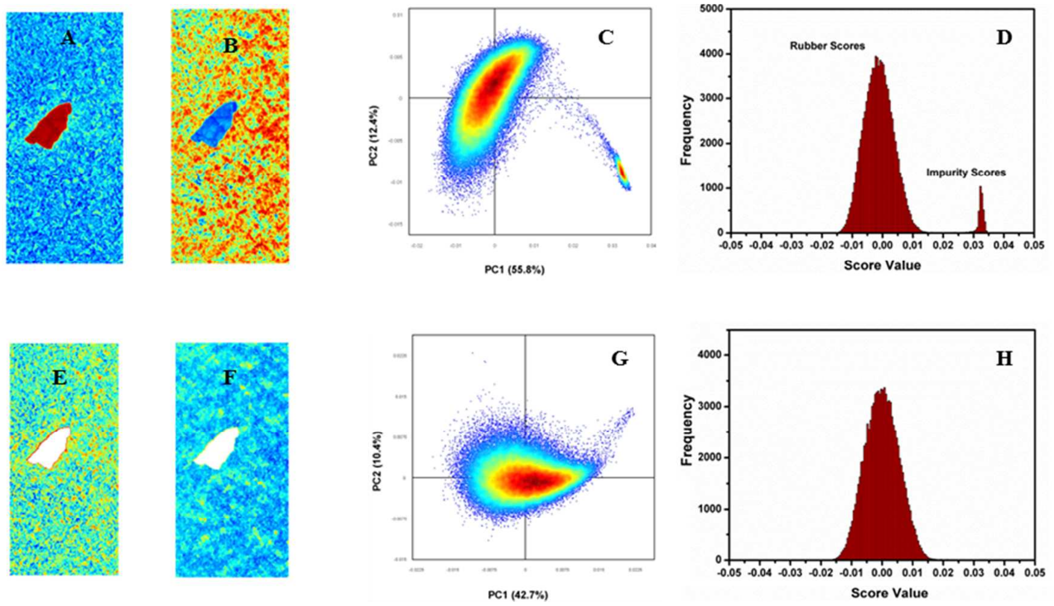


Figure 6. Chemical images based on score values obtained by PCA of the whole spectral data set of a rubber sample containing a small impurity (leaf piece) for the first (A) and second principal component (B); plot of the scores values for the first and second principal component after analysis carried out on the whole spectral data set (C); distribution of score values for the first principal component of the whole data set (D); E, F, G, and H, the same significance as A, B, C, and D, after the scores with values above 0.02 had been removed in D and the remaining data had been submitted to a new PCA, releasing the data set from most of the contaminated pixels.