

DETECTION AND QUANTIFICATION OF STICKINESS ON COTTON SAMPLES USING NEAR INFRARED HYPERSPECTRAL IMAGES

L.S. Severino^a, B.F. Leite^b, F. F. Gambarra-Neto^c, J. B. Araújo^a, and E. P Medeiros^a

^a *Empresa Brasileira de Pesquisa Agropecuária, Embrapa Algodão, Laboratório de Tecnologia Química Avançada, Campina Grande, Brasil. E-mail: liv.severino@embrapa.br.*

^b *Universidade Estadual da Paraíba, Departamento de Química Industrial, Campina Grande, Brasil.*

^c *Universidade Federal da Paraíba, Departamento de Agronomia, Areia, Brasil.*

Stickiness is a kind of contamination of cotton fibres caused by sugars excreted by insect pest such as whitefly (*Bemisia argentifolii*) and aphid (*Aphis gossypii*). Stickiness causes severe losses to cotton growers and textile industry because it is hard to detect, and the problem is revealed only after the fibre is being processed in textile machinery. The objective of this study was to develop a non-destructive technique for detecting cotton stickiness using near-infrared hyperspectral (HSI-NIR) images. For the study, the sugars that cause stickiness and other sugars that occur naturally in cotton fibres were applied over clean cotton fibre at controlled rate. Images were acquired with HSI-NIR cameras, and the data was processed to develop methods for identification of the contamination spots. The method was successful to identify the pixels in the image that were contaminated with sugars, and it discriminated between the glucose that composes the cellulose fibres and the sugars that were introduced to the sample simulating stickiness (melezitose, trahalose, glucose, fructose, and sucrose). The statistical procedures were able to estimate the sugar concentration in the spots, and it was very repeatable when the simulations were made with several replications. The method was able to detect the presence of sugar applied both as water solution and in crystal form. The sensibility of the method is quite lower than the threshold for classifying a cotton sample as sticky (0.3% m/m).

INTRODUCTION

Stickiness is a kind of contamination of cotton fibre caused by infestation of sucking insect pests. The most common insects causing cotton stickiness are whitefly (*Bemisia argentifolii* (whitefly) and aphids (*Aphis gossypii*). Those insects excrete a concentrated solution of sugars (honeydew) that melts and spreads through the fibres, damaging the quality of final product and industrial machinery (HEQUET & ABIDI, 2002). Cotton stickiness causes losses that reach millions of dollars to the world textile industry every year. There are several methods for detecting and quantifying honeydew, but all of them are time consuming and labour demanding, and sometimes they are unreliable depending on the sugar present in the sample. Detecting and quantifying cotton stickiness is difficult because the chemical structure of the honeydew' sugars (melezitose and trahalose) is very similar to the sugars that occur naturally in the fibres (glucose and fructose) and to cellulose, which is the basic structure of cotton fibre (Table I).

Detecting cotton stickiness using images is an interesting option because the test can be applied to a large number of samples without increasing the costs or spending time and resources for laboratorial analysis. The option was made for Near Infrared Hyperspectral Imaging (NIR-HSI) considering that it is an analytical technique able to discriminate materials based on chemical composition and that can add both spatial and spectral data from a sample since each pixel from the image is connected with an individual spectrum. The

purpose of this study was to evaluate the ability of NIR-HSI for detecting and quantifying stickiness in cotton fibres as a non-destructive, fast, and low-cost procedure.

Table I - Chemical structure of honeydew (melezitose and trehalose), physiological sugars (glucose and fructose) and cellulose.

Compound	Composition	Structure
Cellulose	Polymer $\leq 99,0\%$ Glucose	
Fructose	Monomer carbohydrate	
Glucose	Monomer carbohydrate	
Melezitose	Trimer carbohydrate: two molecules of glucose and one fructose.	
Sucrose	Dimer carbohydrate: one molecule of glucose and one fructose.	
Trehalose	Dimer carbohydrate: one molecule of glucose and one fructose.	

MATERIAL AND METHODS

Sample Preparation

The following sugars were dissolved in water to simulate the insects excrements: Sigma Aldrich Trehalose (> 98,5%), Melezitose (> 99,0%), Glucose (> 99,5%), Sucrose (> 99,5%), and Merck Fructose (> 99,0%). The concentration of each sugar was calculated by mass as described in Hequet & Abidi (2006). In order to obtain treatments contrasting for both sugars composition and total sugar concentration, the 5 types of sugars were arranged in 26 combinations. The combinations were planned providing that every solution had a different concentration of each sugar (i.e., there was not two sugars with the same concentration in the same solution), and that the total sugar content varied from 0,075% to 0,833% (m/m). An additional treatment consisted in the sugars applied in the crystal form: 50 mg of each sugar were mixed together in a mortar and pistil without dissolving in water. Control treatments were prepared applying water replacing the sugar solutions and with cotton samples without any treatment.

Regular cotton was used to prepare 2 g samples flatten to 15 x 5 cm. Each cotton sample was contaminated with 40 drops of 5 μ L of each sugar solution. Ten samples were prepared repeating the solution 0.356% (m/m) in order to measure the repeatability of the method.

Image acquisition and data processing

Hyperspectral image analysis has as its main objective to acquire pictures containing specific information about the composing compounds on the surface measured (Amigo et al., 2015). Using a correlation of individual spectrum for each pixel requires very specific parameter settings in order to collect the best set of information. NIR-HSI images were acquired using a Specim® hyperspectral camera (SisuChema SWNIR, Finland) on reflection mode with 50 mm lens, 1.6 ms exposition time, and 6.25 nm of spectral resolution in wavelength region between 1000 and 2500 nm at a spatial resolution of 10 nm with $150 \mu\text{m}^2$ pixel size. The cotton samples were analysed over a Teflon plate, which presents no activity in NIR and can be easily removed afterwards. Data was recorded using the software ChemaDAQ and processed with the hyperspectral image analyzer software Evince 2.7.5 (Umbio, Umea, Sweden), which was used to avoid further requirements on programming while executing image spectral and spatial processing.

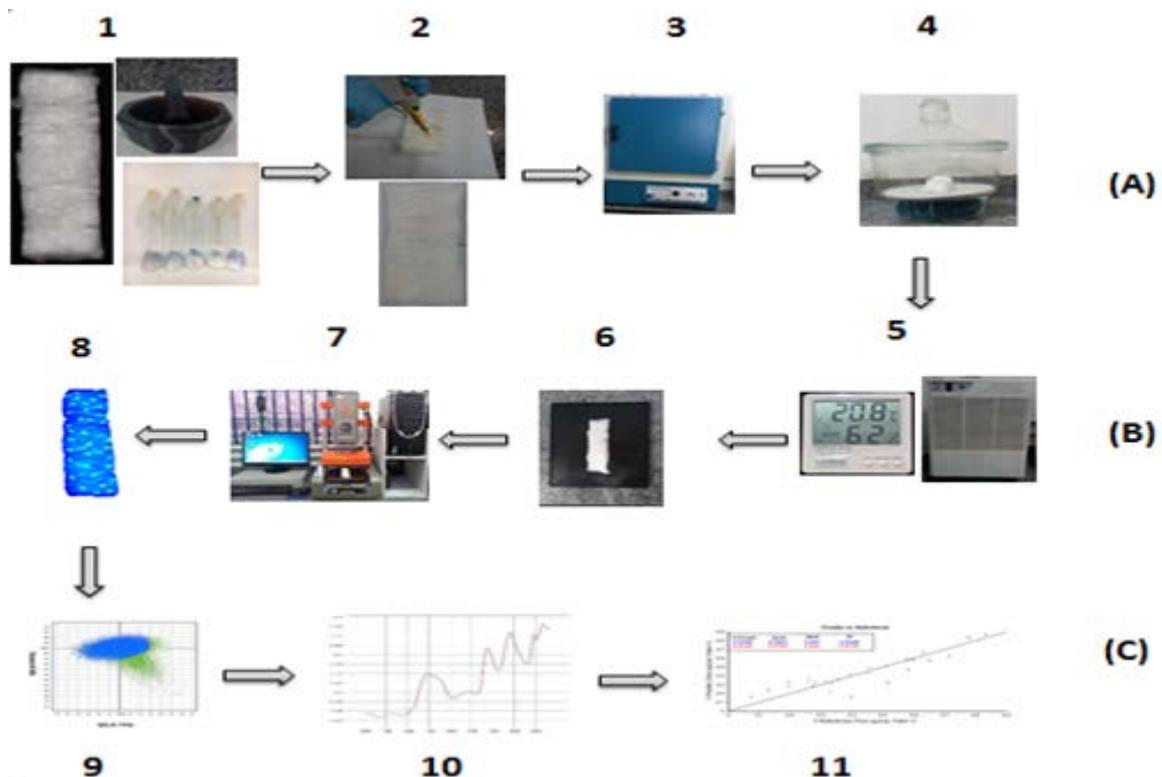


Figure 1 - Schematic representation of the experimental workflow on synthetic sticky cotton samples imaging and statistical analysis. 1 – cotton samples were prepared for contamination with a combination of sugars at controlled concentrations; 2 – 40 drops of 5 μL sugar solutions were applied in the cotton sample; 3 and 4 - the samples were oven dried and stored in desiccator; 5 – samples were acclimated to standard temperature and humidity; 6 – samples placed in Teflon plate to acquire data; 7 - NIR-HSI image was acquired; 8 - spatial and spectral processing was performed; 9 – Principal Components Analysis clustered pixels of clean cotton and contaminated spots; 10 - medium spectra was recorded, and 11 – Partial Least Square regression was calculated.

For each image the following data processing was done: (1) standard normal variate (SNV) was applied to lower the light scatter effect; (2) complete background removal to keep only information about the sample; (3) variables between 900 and 993.75 nm and between 2356.25 and 2500nm were excluded because they are spectrum noisy that hinder the visualization of the spots contaminated with sugars.

Principal Component Analysis (PCA) is automatically generated when editing data in Evince allowing a statistical evaluation about differentiation between compounds. Because of this, the resulting PCA for each sample can be described in terms of different components (PCs) that maximize the clustering capacity of the data set. Once the Principal Components are established, pixels belonging to the sugar spots can have their medium spectra recorded and quantitative information inferred from NIR can be achieved. The medium spectra from the experimental planning were exported to the multivariate statistical software The Unscrambler X.3 (CAMO, Oslo, Norway) to create a Partial Least Square linear regression for total sugar percentage. The complete experimental workflow is summarized in Figure 1.

RESULTS AND DISCUSSION

Visualization of sugar spots

The illustration of the results (Figure 2) is presented in the samples with minimum (A), medium (B), and maximum (C) sugar content and in the sample contaminated with sugar in crystal form (D). The sugar spots applied as solution cannot be visualized in the contaminated cotton (in the left) but the sugar in crystal form had a dark colour and were visible. After data analysis (in the right) the sugar spots become clear and the contamination spots are easily visualised. It was confirmed that the contaminations made with sugar crystals are equally identified by the technique.

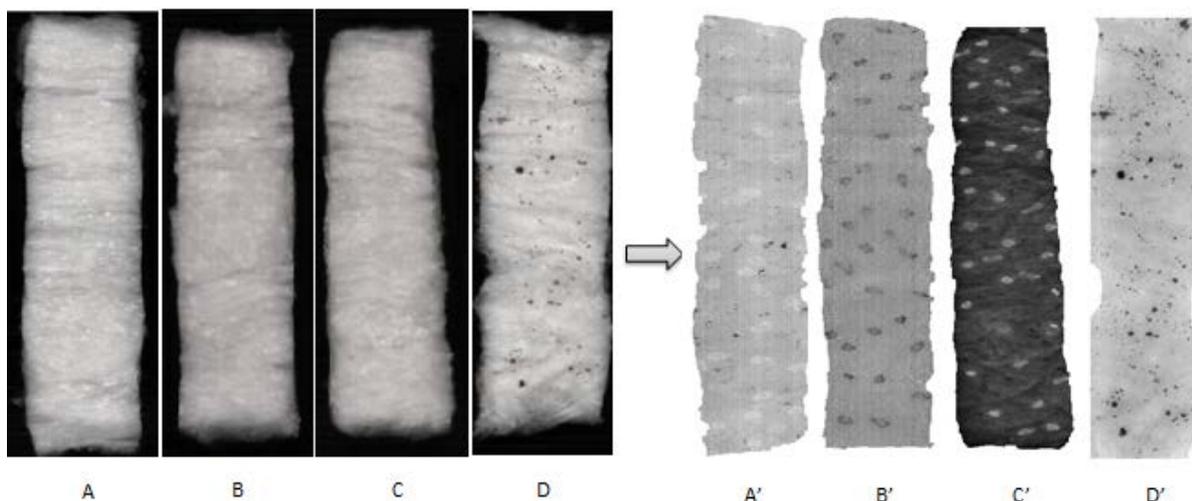


Figure 2 - Cotton samples with minimum (A), medium (B), and maximum (C) sugar content contaminated with sugar in crystal form (D) before (left) and after image analysis.

Exploratory Analysis

Exploratory data analysis on NIR hyperspectral images displays PCAs profiles confirms the ability to discriminate between clean and contaminated pixels and samples contaminated by both sugar solutions and sugar crystals are equally detected (Figure 3). The discriminating power of the method increased as the sugar solution was more concentrated (comparing Figure 3B to 3K) but it still efficient even with very low sugar concentration.

In addition to it is possible to observe which the Explained Variance (EV) and number of Principal Components are considerable different from what is usually seen in NIR-HSI analysis. The number of components mainly increases with declining in total sugar percentage, as can be noticed comparing, again, PCAs B and K in Figure 4, represented by PC 6 vs. PC 4 and PC 3 vs. PC 2, respectively. Although the results expose high number of PCs, the sum of their EVs (eg. EV_{PC1} to EV_{PC6} , in B) is $\geq 99.0\%$.

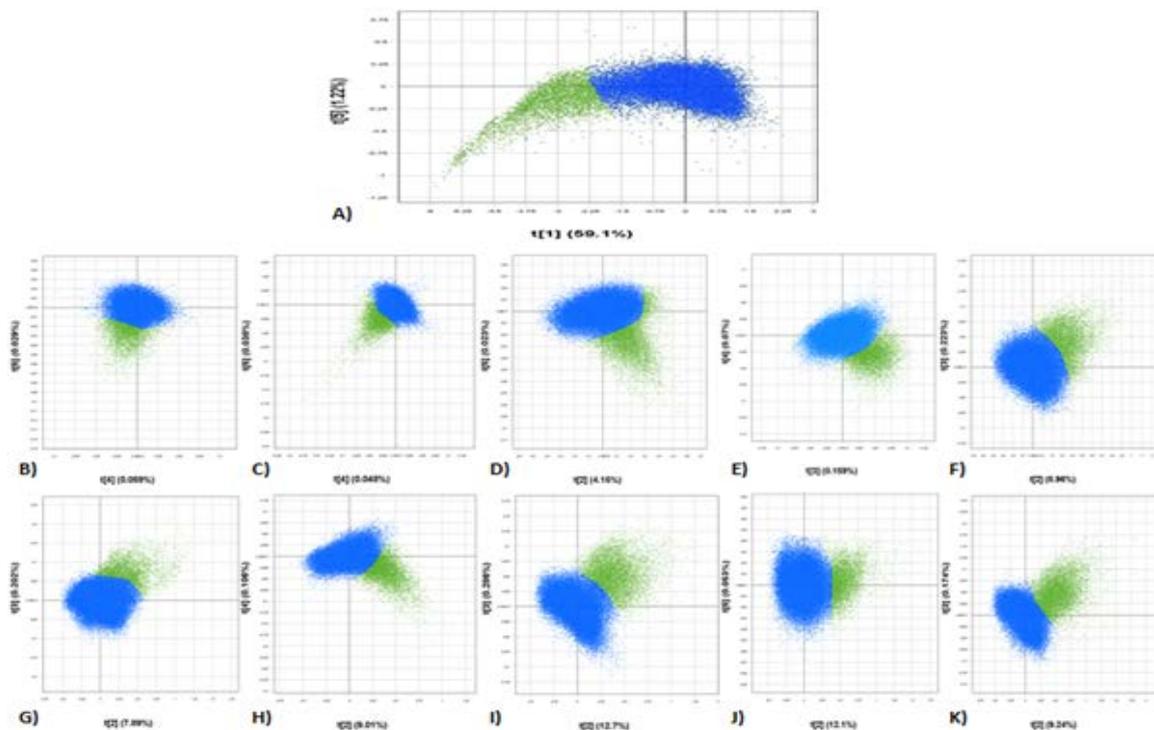


Figure 3 - PCAs' scores maps by category for commercial cotton samples contaminated with sugars. Green dots are sugar-contaminated pixels and blue dots are clean cotton; A - sample contaminated with sugar crystals; B to K - cotton sample contaminated with sugar solutions increasing from 0.197 (B) to 0.833 % (m/m).

A clustering analysis with Principal Components Analysis considering the 41 samples with varying levels of sugar contamination illustrate the capacity to discriminate stickiness in cotton (Figure 4). For illustration, it was assumed that the threshold contamination to be considered sticky cotton was 0.3%, as suggested by Abidi & Hequet (2006). The analysis found a cluster with clean cotton (blue squares in the top left), a cluster with low sugar contamination (red circles in the centre), and a cluster with high sugar contamination (green triangles in the bottom right). There were minor overlapping and the pattern agrees between sugar concentration and calculated values.

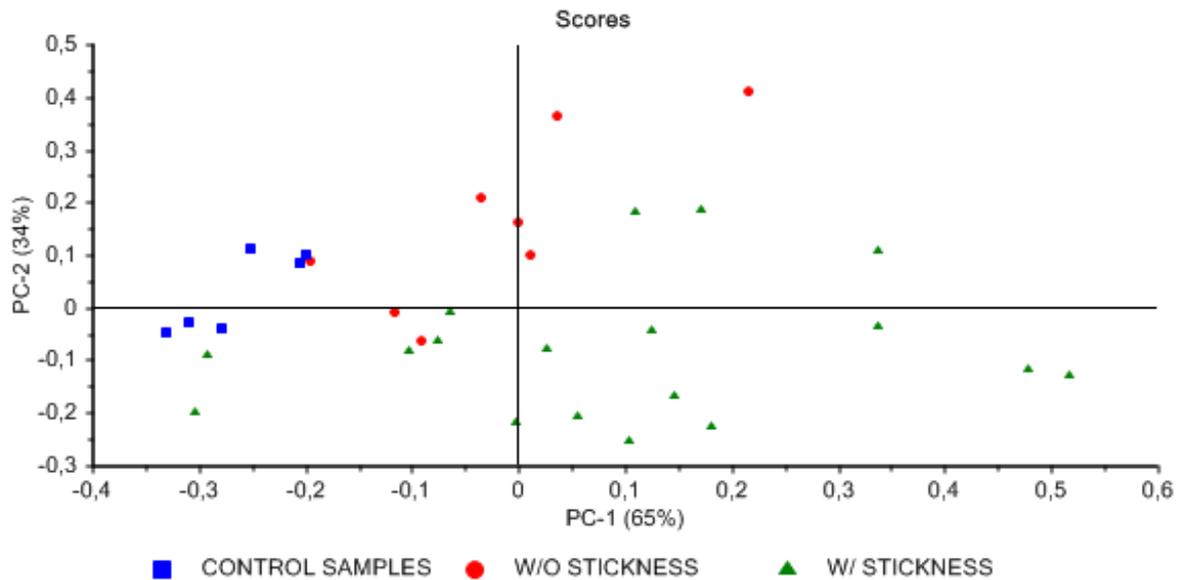


Figure 4 - Clustering analysis calculated with Principal Components Analysis discriminating cotton samples with varying levels of sugar contamination: clean cotton (blue squares), low sugar contamination (<0.3%, red circles), and high sugar contamination (> 0.3%, green triangles).

All spectra from the contaminated pixels (green dots presented in Figure 3) were treated with Savitzky-Golay derivative with a second order polynomial and a three smoothing points window to detect slight differences between samples. This procedure revealed that the spectral profiles (Figure 5) are composed of five main bands corresponding to pure cellulose and the other sugars. Vibration bands in 1400 nm and 1900 nm refer to R-OH structures; vibrations in the peaks 2050, 2200, and 2300 nm related to OH combinations, CC, and CH₂, respectively. Differentiation occurs in terms of pseudo-absorbance increasing that is directly related to the total sugar content. Because it relies on the number of chemical entities involved, a mathematical correlation was designed using a linear regression to estimate the sugar content in the contamination spots.

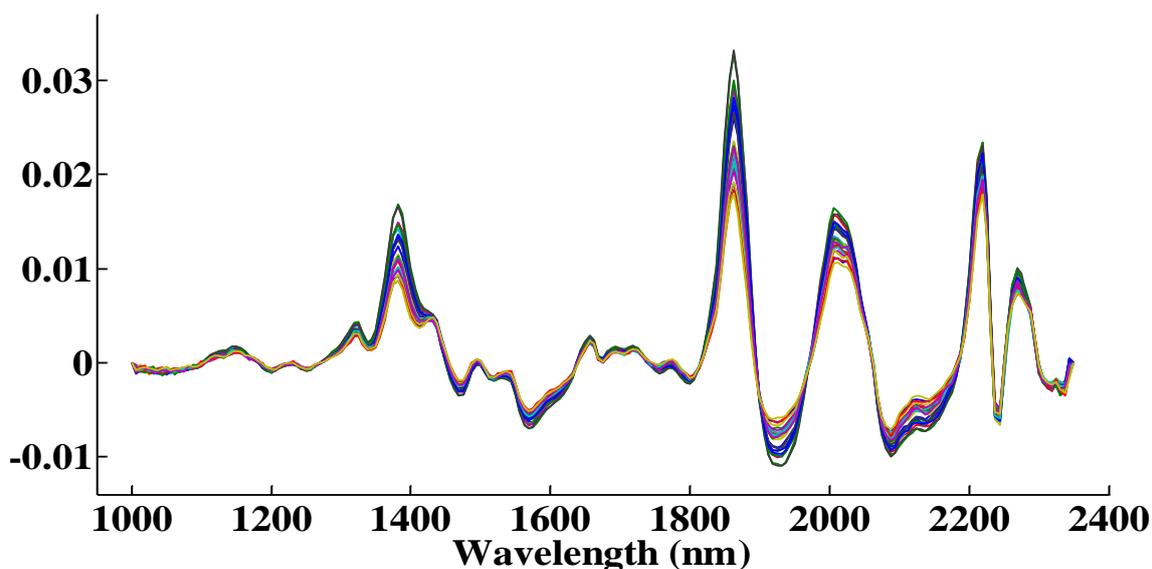


Figure 5 - Medium spectra subjected to Savitzky-Golay derivative of cotton samples. The differentiation is calculated using the points in which the spectra are contrasting and the vibrations are correlated with the sugar content.

A Partial Least Square Linear Regression was developed to estimate the total sugar percentage in each sample. The regression resulted in great accuracy and explained 84% of the variance (Figure 7). R-PLS is constituted of 27 samples varying total sugar percentage and the control with clean cotton. The method was able to detect sugar contaminations quite below the threshold of 0.3% (m/m) which is considered for classifying the sample as sticky cotton.

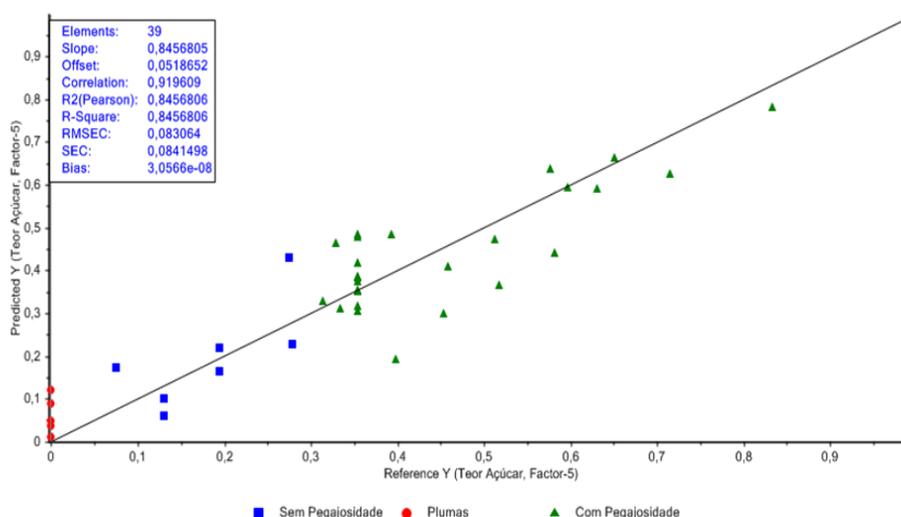


Figure 7. R-PLS of medium spectra of contaminated regions on samples and its corresponding total sugar percentage. Blue squares stands for control samples; red circles represents samples without stickiness and green triangles, sticky samples. The threshold for stickiness is a concentration of total sugar $\geq 0,3\%$ (m/m)

Reproducibility and crystal/solution form

The variation in spectra of the control treatment (clean cotton), which were measured in three different samples, was as low as 0.01 u.a. among peak crests of same vibrational bands. This is evidence that the technique has a great reproducibility among samples with equal composition. The control treatments in which water was applied replacing the sugar solutions did not influence the results, and the estimation of sugar content was reliable.

The treatment with medium sugar content (0.356% m/m) was applied to ten cotton sample, and the results confirmed that the variation among measurements were very low (0.02 u.a.). This result confirms the reproducibility of the method regardless of the problems associated with a soft matrix such as cotton.

CONCLUSIONS

- 1) Analysis of Principal Components in hyperspectral NIR images allows to identify the pixel pattern of sugars that cause stickiness in cotton linter;
- 2) The spectral profile of the samples in association with the weights of the regression model revealed that the most relevant information for the detection and quantification of total sugars in cotton linter are in the specific vibration bands corresponding to R-OH (1400 and 1900 nm) and in the combinations of OH (2100 nm);
- 3) Considering the lower threshold for stickiness in 0.3% of total sugars established in the literature, the HSI-NIR with R-PLS technique allowed to determine the percentage of total sugars with high sensitivity;
- 4) 4) The analysis of sugar content in cotton has low cost, requires simple sample preparation, and the data acquisition is quick (30 s) and non destructive;
- 5) The system is feasible to be adapted for automatic or portable devices for use in industry and field conditions.

REFERENCES

- Amigo, J. M.; Babamoradi, H.; Elcoroaristizabel, S. 2015. Hyperspectral image. A tutorial. *Analytica Chimica Acta* 896:34-51.
- Hequet, F. E.; Abidi, N. 2002. Processing Sticky Cotton: Implications of Trehalulose in Residue Build-up. *J. Cotton Sci.*, 06: 77- 90.
- Hequet, F. E.; Abidi, N. 2006. Sticky cotton: measurements and fiber processing. Texas University Press. 14-24.