Attenuated Total Reflectance-Fourier Transform Infrared spectroscopy coupled with chemometrics for the rapid detection of coconut water adulteration

Thomas A Teklemariam[1*], John Moisey[2], Jason Gotera[1]

[1] Canadian Food Inspection Agency (CFIA), Greater Toronto Area Laboratory, Midland Avenue, Toronto, ON, M1P 4R7, Canada.

[2] Canadian Food Inspection Agency CFIA, National headquarter, 1400 Merivale Road ,Ottawa, Ontario, K1A 0Y9, Canada.

* Corresponding Author:
Email: thomas.teklemariam@canada.ca

Abstract

Attenuated Total Reflectance-Fourier Transform Infrared spectroscopy (ATR-FTIR), in combination with chemometrics, was explored as a rapid method of detecting sugar adulteration in coconut water. In a simulated experiment, coconut water was substituted with binary sugars, mixed sugars, and high fructose corn syrup and discriminated using the fingerprint infrared band region between 1200-950 cm^{-1}. Principal component analysis (PCA) performed on data pre-processed by the Savitzky-Golay smoothing and gap-segment derivative, revealed data clusters discernible by the type and level of substituted sugars, enabling visual diagnosis of the similarity and anomalous features in the dataset. Statistical performance metrics following a cross-validated partial least square (PLS) regression indicated the prediction of adulterant sugars at single-digit percent substitutions. A parallel exploratory analysis of 31 different commercial coconut water samples showed a distinct PCA clustering for samples bearing the label "added sugar", suggesting the potential use of the methods to screening samples for undeclared sugar additions.

Keywords: Coconut water, Adulteration, Substitution, Sugars, Chemometrics, ATR-FTIR

1 Introduction

Fruit juices are among the top food commodities that are targeted for food fraud (Everstine et al., 2013; Dasenaki and Thomaidis, 2019). The common food fraud in fruit juices involve dilution and substitution with cheaper alternatives to maximize volume for monetary advantage (Dasenaki and Thomaidis, 2019). Although this type of adulteration is economically motivated, the practice could lead to the production of drinks that are nutritionally altered with undeclared ingredients potentially posing food safety concerns (Everstine et al., 2013; Alves et al., 2016).

Coconut (*Cocos nucifera L.*) is a tropical plant with a hard-shelled fruit that is highly prized, not only for the nutritional use of its edible white kernel and coconut water, but also for medicinal, cosmetic, and other economic applications (Yong et al., 2009; Prades, et al., 2012b). Sugars are one of the major constituents of coconut water. Depending on the coconut variety and the maturity of the nut from which it is extracted, the

sugar content may range from about 1 to 6 % (Appaiah et al., 2015, Burns et al., 2020, Prades et al., 2012b). Coconut water also contains minerals, particularly potassium, and small amounts of proteins, fats, and amino acids (Yong et al., 2009). According to recent reports, a surge in popularity and an imbalance between supply and demand (Prades et al., 2016) is making coconut water an attractive target for adulteration (Richardson et.al 2019 and references therein), where aqueous dilution is accompanied by the addition of sugars to compensate for the loss of natural sweet taste.

The production of packaged coconut water involves a series of steps from harvesting, post-harvest handling, to packaging and storage (Prades et al., 2012a). Some coconut water products are marketed as pure, natural, organic, or non-GMO, while others declare added ingredients including sugar, minerals, vitamins, antioxidants, and processing agents. All packaged products are also subjected to some kind of preservation, pasteurization, or sterilization, procedures used to increase product stability and shelf life (Burns et al., 2020). These processing steps are expected to add variations to the composition of commercially packaged products as compared to freshly extracted pure coconut water.

The traditional determination of authenticity of fruit juices involves the separation of components by chromatographic techniques and quantification by stable isotope ratio analysis (Koziet et.al, 1995; Dasenaki and Thomaidis, 2019), which is expensive and time-consuming. In recent years, there has been an increasing interest in utilizing the synergy between molecular spectroscopy and multivariate data analysis in food authenticity testing (e.g., Rodriguez-Saona and Allendorf, 2011, Rodriguez-Saona, et al., 2016). Fourier transform infrared spectroscopy (FTIR) is a popular method that requires little to no sample preparation. FTIR obtains molecular information of a substance based on absorbance by chemical bonds in the mid-infrared region (MIR). The infrared spectrum of a substance is shaped by the characteristic bonds and functional groups of molecules that are present. The spectral bands represent many of the general stretching, bending, and wagging motions performed by molecules when it absorbs infrared energy (Larkin, 2018). Since the spectrum maps a broad range of specific information about the substance's structure and chemical environment, it can be regarded as a chemical fingerprint.

FTIR spectroscopy has been used to detect sugar adulteration in apple juice (Sivakesava et al., 2001) and in honey (Rios-Corripio et.al, 2011). A recent study by Richardson et al., (2019) showed the successful use of Raman spectroscopy for the detection of sugar adulteration in coconut water. Although Raman and FTIR are both forms of molecular vibrational spectroscopy, the two methods harvest signals based on different principles, i.e., the scattering of narrow-band monochromatic light versus the absorption of broadband infrared light when it comes into contact with a material, respectively. This difference provides each method with unique strengths and limitations that is suited for different applications, making these techniques

complementary in nature (Hashimoto et al., 2019). Since water strongly absorbs infrared and interferes with FTIR analysis, the water-inactive Raman is usually the preferred method for the analysis of aqueous samples. Nevertheless, FTIR is a widely used, simple, cost-effective, and versatile method with the option for different sampling accessories that facilitate the quick collection of spectra from various materials. The Attenuated Total Reflectance (ATR)-FTIR sampling technique, in particular, has gained the most traction in recent years due to its combined advantages of being a simple, non-destructive, and inexpensive method that provides fast data acquisition and high-quality reproducible spectra in different fields of research (e.g. Ramer and Lendl, 2013). Although, the sensitivity range of ATR-FTIR has some limitation for trace-level analysis, it is suited for the analysis of sugars in coconut water, which are present at the lower, single-digit percent concentrations.

The adulteration of coconut water with an aqueous sugar solution is expected to alter the chemical environment and the ratio of the different sugars in the product. The ATR-FTIR spectrum can serve as a molecular fingerprint for identification of the resulting changes, provided that interference from the intense OH-stretching and bending bands originating from the water matrix that dominates the IR spectrum are isolated. This study explores the combined application of ATR-FTIR spectroscopy and chemometrics as a rapid method to detect sugar adulteration in coconut water based on a selected fingerprint spectra region that is characteristic for sugars, but not affected by water absorption bands.

2 Experimental

2.1 Reagents and preparation of standard solutions

Analytical grade of anhydrous D-Fructose (99%), D(+)- Sucrose for biochemistry (DNAse, RNAse and Protease free) and Dextrose (D-Glucose) were obtained from Alfa Aesar, ACROS and Fisher, respectively. High Fructose Corn Syrup HFCS-42 (42% fructose) and HFCS-55 (55% fructose), were kindly supplied by Ingridion Canada Corporation. A total of 20 young coconut fruits were purchased from a local supermarket for the study. The fruits were trimmed (the husk removed and the remaining fruit wrapped with plastic) when purchased. The fruits originated from Southeast Asia but as the products are purchased at retail, age and postharvest storage conditions were unknown. For the adulteration study, coconut water samples extracted from the 10 out of the 20 fruits were combined and cleared by centrifugation 10,000 g at 4 °C for 20 minutes and used as a stock coconut water solution. The solution, which had a pH of 6.5 (Mettler-Toledo pH meter) and Brix value of 7.4° (Fisher Brand HDR-P1 series Digital Handheld Refractometer), was divided into smaller aliquots and stored at -20° C until used for analysis. Thirty-one different commercially packaged coconut water products (brands and some with the same brand but different product lines) were also

obtained from retail in 5 replicate per product. These samples together with coconut water samples extracted from the remaining 10 young fruits were used to assemble an ATR-FTIR spectral library.

2.2 Simulated coconut water adulteration

2.2.1 The standard addition method

The method used for determining the background concentration of sugars in the coconut water stock solution and the preparation of adulterated coconut water via substitution of different concentration-matched sugar solutions (briefly outlined below), is essentially as described in Richardson et al. (2019). First the natural concentration of the different sugars in the stock fresh coconut water sample were determined by the standard addition method (SAM). Samples for SAM were prepared by spiking 250 µl aliquots of coconut water with various volumes of sugar solution ranging from 25 µl to 250 µl, at 25 µl step increments, and filling each sample to a final volume of 500 µl with deionized water. A total of 35 spectra were collected for SAM from unadulterated coconut water (control) and adulterated coconut water at 10 different levels for each sugar adulterant (sucrose, fructose, glucose). The intensity values of distinct bands at 1057 cm^{-1}, 1061cm^{-1}, and 1080 cm^{-1}, identified from the overlaid spectra of solutions made from the analytical grade sucrose, fructose and glucose in deionized water (Fig 1), were used for the SAM calibration based on the least-square linear regression (Bader, 1980). The concentrations obtained by the SAM for the coconut water stock solution were 24.7, 33.5, 31.7 mg ml^{-1} for sucrose, fructose, and glucose, respectively.

2.2.2 Substitution of coconut water with different sugar solutions

For the adulteration detection experiment, individual stock solutions of analytical grade sucrose, glucose and fructose were prepared at a concentration of 89.9 mg ml^{-1} to match the total sugar concentration of the coconut water stock solution, as determined by SAM (section 2.2.1 above). For mixed sugars, a solution of total sugars corresponding to this level was prepared by dissolving 1.24 g glucose + 1.66 g fructose + 1.58 g sucrose, in 50 ml of distilled water, keeping the ratio for the 3 sugars approximately consistent with the composition of the fresh coconut water. For the adulteration experiments using HFCS-42 and HFCS-55, stock solutions approximately matching that of the coconut water's total sugar content (7.4 ^{o}Bx) were made by dilution of the original solutions with deionized water. The adulteration was made by substituting the coconut water with deionized water and the above concentration-matched solutions of the binary sugars, mixed sugars, HFCS-42 and HFCS-55 at levels ranging from 5–100% at 5% step increments. A total of 130 spectra were collected for the experiment corresponding to 10 individual spectra for pure coconut water, plus 20 spectra each for the adulterated samples at 20 different levels. The samples were measured the same day as

they were prepared and the leftover thawed coconut water was discarded. A fresh-frozen aliquot was used per day.

For the ATR-FTIR spectral database, individual spectra were recorded from the coconut water extract obtained from each of the 10 young coconut fruit (10 spectra) and each of the 5 packages of the 31 commercially packaged products (155 spectra). The relevant product information declared on the packaging for each product was recorded as variables for the product's spectral database (Table S1).

2.3 Spectroscopic measurements

Thermo Nicolet™ iS50 FTIR Spectrometer equipped with Thermo Scientific™ Smart™ iTX Accessory fitted with a diamond crystal and Omnic software was used for spectral acquisition. A drop of the sample was placed on the ATR crystal using a disposable bulb-top Pasteur-pipette. ATR-FTIR spectra were recorded in wavenumber range between 500-4000 cm^{-1}, averaging 10 scans per sample using a resolution of 4 cm^{-1}. This single spectrum acquisition process takes less than a minute to complete.

2.4 Data analysis

All collected spectral data were converted and exported as comma-separated value (CSV) files. Data merging and analysis were performed using RStudio: integrated development for R, a language and environment for statistical computing (RStudioTeam, 2020). In order to remove unwanted signals, to reduce noise, and to improve the signals of interest, the spectral data was subjected to a series of established preprocessing methods. First, the spectral range of interest that was considered for analysis was set between 950-1200 cm^{-1} and the spectra that lie outside of this specified range were removed. Baseline correction was done using "*spmblp*" function in the "*spftir* " package that is based on a local polynomial fitting (Valenzuela and Rodriguez-Llamazares, 2016). The "*spmblp*" function was run with the following options: degree of polynomial = 2, rep (iteration) = 100, tolerance (between iterations) = 0.001. The "*prospectr* " package (Stevens and Ramirez-Lopez, 2020) was used to convert the baseline-corrected data into a hyperspec object prior to performing smoothing and generating first and higher level derivative spectra applying the gap-segment derivative (gapDer algorithm). The gapDer algorithm performs first Savitzky-Golay smoothing under a given segment size (s), followed by a derivative (Stevens and Ramirez-Lopez, 2020). The gapDer transformed derivative spectra were used in subsequent data dimensionality reduction for exploratory data analysis and predictive chemometric modeling using principal component analysis (PCA) and partial least square (PLS) regression. PCA was performed using the "*prcomp*" function in R and plots are generated using "*ggplot2"*

(Wickham, 2016) the autopilot function from "*ggfortify* "package, an extension to "*ggplot2"* (Tang and, Horikoshi, 2016) and "*plotly"* package (Sievert, 2020). PLS was performed using the PLS package (Mevik and Wehrens., 2007).

3 Results and discussion

3.1 ATR-FTIR Spectra of adulterated coconut water

The ATR-FTIR spectra of coconut water and the aqueous solutions of different sugars used for substitution are shown in Fig 1. Dictated by the basic chemical structure of sugars with aldehyde, ketone, and multiple OH functional groups, characteristic absorption bands are expected to show at (a) the 3700 – 2700 cm^{-1} band region, which is associated with the stretching of the OH and the C-H groups; (b) the 1700 – 1600 cm^{-1} band region corresponding to the stretching vibrations of the ketone (C=O) of fructose and the aldehyde (CH=O) of glucose coinciding with the bending vibration of the OH group from water; and (c) the fingerprint region between 1200 - 800 cm^{-1}, which represents the most complex part of the IR spectrum, is associated with the C-C, C-O, O-C-H, and C-O-H functional groups (Tipson, 1968; Grube et al., 2002). In the first 2 broad regions, the intense water absorption bands fully overlap with the spectra of sugars making this region unavailable for direct analysis without going through complicated spectral manipulations to isolate the signal. The fingerprint band region centered around 1200-950 cm^{-1}(Fig. 1A & B) is, however, free of interference and shows distinct peaks that are characteristic for sugars (Fig 1B). These bands are generated by the stretching, bending, and wagging vibrations associated with C-C, C-O, O-C-H, and C-O-H functional groups (Grube et al., 2002). Since the infrared spectra of carbohydrates shows a greater contrast in this band region (Max and Chapados, 2007), there is increased potential for the direct application of the ATR-FTIR method to characterize sugars in coconut water and likely other similar aqueous samples.

3.2 Derivative spectroscopy

The selection of spectral data transformed as a gap-segment derivative spectrum (gapDer) was based on preliminary visual exploratory data analysis for data pattern and variability following different preprocessing methods. Derivative spectra are well known to improve spectral data quality by removing baseline offset to help separate overlapping peaks and sharpen spectral features. However, since digital derivatization increases noise (Fig 2 A & C), the process is usually coupled with smoothing (Fig 2 B & D). These steps are conveniently integrated into gapDer algorithm, which performs first a smoothing under a given segment size(s), followed by a derivative (Stevens and Ramirez-Lopez, 2020). The gapDer algorithm was run by setting the option for filter length (w) i.e, the spacing between points over which the derivative is computed as 11,

smoothing segment size (s), i.e., the range over which the points are averaged as 10, and the order of derivative (m), (1 to 4 derivatives were explored).

3.3 PCA of authentic and adulterated coconut water

Infrared spectra are high-dimensional data consisting of a number of correlated variables (wavenumbers), which makes it amenable to multivariate statistical analysis. PCA is one of the most commonly used methods for dimensionality reduction and extraction of dominant features in the dataset. PCA accomplishes this via the orthogonal projection of the whole dataset onto subspaces of a few uncorrelated variables, the principal components (PCs). The PCs represent reduced data but also preserve most of the information contained in the original dataset as eigenvectors and eigenvalues, which accounts for their direction and magnitude. The 1st PC of PCA summarizes most of the variability in the data with each of the succeeding orthogonal components accounting for the remaining variance in decreasing order (Jolliffe and Cadima, 2016).

The PCA projection of the gap-segment 1st derivative adulteration data is shown in Fig 3. The first 2 principal components (PC1 and PC2) explained the greatest portion (93 %) of the variance (Fig 3A). Three PCs were sufficient to explain >95% of the total variance in the dataset (Fig 3A & B). On the PC1 vs PC2 plot, the fresh unadulterated coconut water projected data are located at the center of the plot, and the substituted coconut water, particularly with the binary sugar solutions, coalesces in narrow clusters. The clusters diverge away from the center in different directions, with the highest % substituted data being the farthest from the center. These patterns are consistent with the ATR-FTIR based study reported for honey adulteration (Rios-Corripio et.al, 2011) and Raman based study of coconut water (Richardson et.al, 2019). The Mixed sugars and HFCS-42 and HFCS-55 substituted data clusters were also evident, although these clusters are less separated and positioned closer to the center of the PC1 vs PC2 plot of the 1st and 2nd orders of derivatives (Fig 3 A, C). These clusters are, however, better resolved on the PC1 vs PC3 plot (Fig 3B & D). This is also the case for the higher-order derivatives. For example, on the PC1 vs PC3 plot, there was a close overlap between fructose and mixed sugar data of 1st order derivative (Fig 3B) and between HFCS-42 and HFCS-55 data of the 2nd order derivative (Fig 3D). The mixed sugar data is, however, better resolved on PC1 vs PC2 and PC1 vs PC3 plot on the 3rd order derivative as well as PC1 vs PC3 of the 4th order derivative (Fig S1 A, B, D). It was evident that the reorientation of the data on the different PCs when using the higher-order derivatives was able to uncover patterns that were otherwise less obvious on the plots derived from the 1st or 2nd order derivatives. The overlap between HFCS-42 and HFCS-55 data clusters may reflect their similarity in composition with 42:58 vs 55:45 difference in fructose: glucose ratios, respectively. Notably, both HFCS-42 and HFCS-55 are well discriminated on PCA plots from the related disaccharide-sucrose, which has a 50:50 glucose and fructose ratio.

3.4 PCA analysis of commercially packaged coconut water spectral data

The PCA analysis of the 31 different products (Table S1) is shown in Fig 4A. The PC1 vs PC2 plot revealed data clusters, with different degrees of separation, together explaining over 90% of the variability in this dataset. As expected, "between brand" differences are shown by the spread of the data across the PC1 vs PC2 plot relative to the "within brand" data points, which in most cases, are tightly clustered together (Fig S2). Yet, there are some obvious "between brand" features that were revealed by the PC plots. For example, when the PCA scores are labeled by the variable "added sugar" or "no added sugar", as declared on the packaging label of each product, at least 3 broad clusters could be identified (Fig 4A). The data cluster at the bottom (symbol "X") corresponds to all the reference coconut water samples that were freshly extracted with no added sugar. Those samples declared to have "added sugar" were congregated in one direction along the PC1 axis (blue squares), while the remaining variously processed samples are in the middle part (pink circles). The samples with "added sugar" labels had higher sugar levels with values ranging between 7.8 - 8.7 ^{o}Bx, compared to 4.5 - 7.1 ^{o}Bx for those samples with "no added sugar" (Table S1). Note that samples that make no explicit mention of added sugar status were considered as having no added sugar for the analysis. Additional clustering features were also revealed by using a combination of different PCs. A clearly isolated cluster on the PC1 vs PC3 plot (Fig S3A) correlated with the coconut water samples that contain "added vitamin C". When the PC1 vs PC2 plot was labeled by the variable "organic" or "not-organic", a cluster was evident for the organic products (Fig S3B), which may be useful for fraud investigation in the organic food sector, where labeling non-organic products as premium-priced organic is a common problem (Miller, 2019). However, more data is needed to confirm whether these observed additional clustering patterns are the result of pure coincidence or in fact, some features associated with these variables are picked by the method, which is targeted for the sugar analysis. In general, the above analysis illustrates how a visual summary of patterns or correlations, not necessarily causation, between variables generated by PCA can be used for quick initial screening and flagging of suspected samples during fraud-related investigations.

3.5 Prediction using PCA

The PCs computed by PCA for one data set (active individuals) could be applied to new supplementary data that did not take part in the PCA computation to help visualize their relationship (e.g., Abdi et.al., 2010, Wold et.al., 1987). Here, for this PCA-based prediction, the spectral data collected from the 31 different coconut water samples was considered as new supplementary individuals. The factor scores obtained by positioning these new observations on the PCA space are then projected onto the PCs drawn from the simulated coconut water adulteration data (the active individuals) using *factoMiner* and *factoextra* R package (Husson et.al,

2017; Kassambara and Mundt, 2020). Fig 4B shows the plots of the first two principal components and the projection of this supplementary data layer (shown as open & closed circles). The supplementary dataset represented the diversity of commercially packaged coconut water samples in the market (Table S1). Although the broad cluster of the data along the coordinate predicted by PCA may reflect this diversity, a dominant trend was also observed. All the coconut water products that are declared as having "added sugar" (indicated as filled circles), were grouped away from the center along the sucrose cluster arm (Fig 4B). Data regarding the type of sugar added to the packaged coconut water products is not available. However, the behavior of the projected supplementary data suggests that the added sugar has spectral properties that are more similar to sucrose than to the other sugars used in the adulteration model. This can be expected considering that sucrose and syrups are the most commonly used sweeteners in the food and beverage industry (Pepin et al., 2019). Although the nutritional and health risk associated with sugars mainly originates from its excess intake, there is also an increasing concern with respect to the type of sugar consumed. This has been particularly the case for artificial sweeteners whose risk-free status remains controversial (e.g., Bruyère et al., 2015). It is, therefore, important to improve the ability of the analysis to provide details on the nature of sugar adulteration by adding more experimental data using different sugar adulterants (artificial sweeteners, various types of syrups), in addition to the variety of packaged coconut water samples that are available in the market.

3.6 PLS regression modeling of coconut water adulteration

PLS is an algorithm that is considered an extension of PCA in which both the predictor and response data are considered. PLS is suited when the matrix of predictors has more variables than observations and there is inter-correlation between the x-variables, which is true for spectroscopic data. The PLS algorithm looks for the fundamental relationship between the x and y matrices via the decomposition of both the x and y data into a set of scores or latent variables (LVs), similar to PCA. However, in this case, the scores for both the x and y data are selected in order to maximize the correlation between the scores for both the x and y variables (Abdi, 2010).

Prior to building a predictive model using PLS, the preprocessed spectral data for each adulterant was randomly split into a train: test ratio of 80:20. The training set is used for building the calibration model parameters. PLS was run with cross-validation (CV) to evaluate model performance, where the training set is divided into k- subsets (folds). Each time, 1 subset is taken out and the model runs on the rest of the dataset. Model performance is then evaluated using the left out sets and the process was repeated for all the remaining k-folds. There are different variations of the CV technique. The leave-one-out cross-validation (LOO-CV)

was used in this study where, each time only one data point is left out during model training and repeated until all the training data is given a chance to be held out (Fushiki, 2011). Model accuracy i.e., the agreement between reference and PLS predicted values (Table 1, Eq1) was determined by calculating the mean square error (RMSE) of calibration (RMSEC), cross-validation (RMSECV) and prediction (RMSEP) and the corresponding coefficients of determination, R^2_{ca}, R^2_{cv} and R^2_{test}, respectively (Table 1, Eq 2). Selecting the correct dimensionality is critical for building a PLS model with good predictive performance (Faber and Rajkó, 2007). The smallest RMSECV and highest R^2_{cv} were used as a guide for selecting the optimum number of components. The PLS model predicted vs measured values are shown in Fig 5. Only one or two LVs were sufficient for > 90% of explained variance for glucose, fructose, and sucrose, while 3 to 4 LVs were needed to explain > 70% of the variance for substitution by mixed sugar HFCS-42 and HFCS-55 (Table 1, Fig S4). The small RMSE, the closeness these values in the calibration and prediction set, and high R^2 for binary sugar adulterants indicate a more accurate model, relative to the values obtained for mixed sugars, HFCS-42, and HFCS-55 (Table 1). This is also consistent with the less separation observed on the PCA plot for these adulterants (Fig 3).

The actual versus predicted levels of adulteration obtained by the optimized PLS model for the binary sugars, mixed sugars and the high fructose corn syrups (Fig 5, Table1) indicate slopes > 0.9 and R^2 values are close to the 1 to 1 theoretical line of prediction for both the calibration and prediction sets. The loading spectra for the different substituted sugars (Fig S4) indicates the degree of influence, positive or negative correlation of the variables i.e., the different wavenumbers, to each of the LV's and the corresponding explained variance. The PLS scores are simply a linear combination of the original data weighed by these loadings. Inspection of the loading spectra wavenumbers will help the identification of influential variables for each of the LVs. The computed limit of detection (LOD) was based on the IUPAC definition (Eq 3 and 4) (Ostra et. al., 2008). The <10% LOD values for the different sugar adulterants (Table 1) indicates that the method can be used to detect any practical level of substitution, which is expected to be significantly higher, provided that the motivation for substitution is maximizing volume to achieve a measurable financial return.

4.0 Conclusion

In this study, ATR-FTIR based fingerprint spectra and multivariate data analysis were used to profile sugars in coconut water. The aim was to evaluate if the system can be utilized as a quick screening method for detecting coconut water adulteration by substitution with different sugar solutions. The direct use of ATR-FTIR spectra to discriminate between different sugars in coconut water was made possible by the presence of an interference-free fingerprint spectral band region between 1200-950 cm^{-1}, which originates

from the vibrational modes that are characteristic to sugars. Simplicity, speed of analysis, and low cost are important attributes for a method dealing with food adulteration, which usually involves screening a large number of samples. While quick spectral data acquisition was achieved by the ATR-FTIR method, PCA-based transformation and classification facilitated easy visualization and detection of dominant and anomalous features in the dataset. An optimized predictive model was developed using PLS, which according to the model performance parameters, could reliably predict the adulterants in the coconut water at < 10% substitution. The potential for incidents of adulteration is expected to rise with increasing trade volume and complexity of the food supply chain. This ATR-FTIR spectral fingerprint-based method is faster and cheaper than the traditional chemical techniques and an alternative to the Raman-based method for the detection of sugar adulteration in coconut water. Additionally, it can potentially be applied to other aqueous food matrices in which dilution and adulteration with sugars are suspected. Implementation of the method in routine testing laboratory require assembling a database of coconut water samples. The acquisition of ATR-FTIR spectrum, using the setup in this study takes less than a minute to complete for each sample. The subsequent preprocessing and analysis of the large volume spectral data involves a series of steps that can take time. However, the data handling could be made as rapid as the spectral acquisition by streamlining the process using the chosen software application. In this study, RStudio and interactive Shiny application (Cheng et al., 2021) were used to efficiently preprocess and analyze data and generate visualization plots.

Acknowledgement

The authors would like to thank Beata Kolakowski, David Awmack, from Canadian Food Inspection Agency, and the three anonymous reviewers for their insightful suggestions and careful reading of the manuscript. Thanks to Todd Marrow, Director, Greater Toronto Area Laboratory, Canadian Food Inspection Agency, for his encouragement and support of the project. This work was carried out with funding from CFIA Short Term Project (PROJECT ID:000229).

References

Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression). *WIREs Computational Statistics*, *2*(1), 97–106. https://doi.org/10.1002/wics.51

Alves, R. C., Barroso, M. F., González-García, M. B., Oliveira, M. B. P. P., & Delerue-Matos, C. (2016). New Trends in Food Allergens Detection: Toward Biosensing Strategies. *Critical Reviews in Food Science and Nutrition*, *56*(14), 2304–2319. https://doi.org/10.1080/10408398.2013.831026

Appaiah, P., Sunil, L., Kumar, P. K. P., & Krishna, A. G. G. (2015). Physico-chemical characteristics and stability aspects of coconut water and kernel at different stages of maturity. *Journal of Food Science and Technology*, *52*(8), 5196–5203. PubMed. https://doi.org/10.1007/s13197-014-1559-4

Bader, M. (1980). A systematic approach to standard addition methods in instrumental analysis. *Journal of Chemical Education*, *57*(10), 703. https://doi.org/10.1021/ed057p703

Bruyère, O., Ahmed, S. H., Atlan, C., Belegaud, J., Bortolotti, M., Canivenc-Lavier, M.-C., Charrière, S., Girardet, J.-P., Houdart, S., Kalonji, E., Nadaud, P., Rajas, F., Slama, G., & Margaritis, I. (2015). Review of the nutritional benefits and risks related to intense sweeteners. *Archives of Public Health*, *73*(1), 41. https://doi.org/10.1186/s13690-015-0092-x

Burns, D. T., Johnston, E.-L., & Walker, M. J. (2020). Authenticity and the Potability of Coconut Water—A Critical Review. *Journal of AOAC INTERNATIONAL*, *103*(3), 800–806. https://doi.org/10.1093/jaocint/qsz008

Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2021). shiny: Web Application Framework for R. R package version 1.6.0. https://cran.r-project.org/package=shiny

Dasenaki, M. E., & Thomaidis, N. S. (2019). Quality and Authenticity Control of Fruit Juices-A Review. *Molecules (Basel, Switzerland)*, *24*(6), 1014. PubMed. https://doi.org/10.3390/molecules24061014

Everstine, K., Spink, J., & Kennedy, S. (2013). Economically motivated adulteration (EMA) of food: Common characteristics of EMA incidents. *Journal of Food Protection*, *76*(4), 723–735. https://doi.org/10.4315/0362-028X.JFP-12-399

Faber, N. M., & Rajkó, R. (2007). How to avoid over-fitting in multivariate calibration—The conventional validation approach and an alternative. *Papers Presented at the 10th International Conference on Chemometrics in Analytical Chemistry*, *595*(1), 98–106. https://doi.org/10.1016/j.aca.2007.05.030

Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation. *Statistics and Computing*, *21*(2), 137–146. https://doi.org/10.1007/s11222-009-9153-8

Grube, M., Bekers, M., Upite, D., & Kaminska, E. (2002). Infrared spectra of some fructans. *Journal of Spectroscopy*, *16*. https://doi.org/10.1155/2002/637587

Hashimoto, K., Badarla, V. R., Kawai, A., & Ideguchi, T. (2019). Complementary vibrational spectroscopy. *Nature Communications*, *10*(1), 4411. https://doi.org/10.1038/s41467-019-12442-9

Husson, F., Josse, J., Lê, S., & Mazet, J. (2017). FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R. *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R*. http://factominer.free.fr

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses* (1.0.7) [R]. http://www.sthda.com/english/rpkgs/factoextra

Koziet, J., Rossmann, A., Martin, G. J., & Johnson, P. (1995). Determination of the oxygen-18 and deuterium content of fruit and vegetable juice water. An European inter-laboratory comparison study. *Analytica Chimica Acta*, *302*(1), 29–37. https://doi.org/10.1016/0003-2670(94)00424-K

Larkin, P. J. (2018). Chapter 1—Introduction: Infrared and Raman Spectroscopy. In P. J. Larkin (Ed.), *Infrared and Raman Spectroscopy (Second Edition)* (pp. 1–5). Elsevier. https://doi.org/10.1016/B978-0-12-804162-8.00001-X

Max, J.-J., & Chapados, C. (2007). Glucose and fructose hydrates in aqueous solution by IR spectroscopy. *The Journal of Physical Chemistry. A*, *111*(14), 2679–2689. https://doi.org/10.1021/jp066882r

Mevik, B.-H., & Wehrens, R. (2007). The **pls** Package: Principal Component and Partial Least Squares Regression in *R*. *Journal of Statistical Software*, *18*(2). https://doi.org/10.18637/jss.v018.i02

Miller, H. I. (2019). Buying "Organic" to Get "Authenticity"? Or Safer and More Nutritious Food? Think Again. And Again. *Missouri Medicine*, *116*(1), 8–11. PubMed. PMCID: PMC6390794

Ostra, M., Ubide, C., Vidal, M., & Zuriarrain, J. (2008). Detection limit estimator for multivariate calibration by an extension of the IUPAC recommendations for univariate methods. *The Analyst*, *133*, 532–539. https://doi.org/10.1039/b716965p

Pepin, A., Stanhope, K. L., & Imbeault, P. (2019). Are Fruit Juices Healthier Than Sugar-Sweetened Beverages? A Review. *Nutrients*, *11*(5), 1006. https://doi.org/10.3390/nu11051006

Prades, A., Dornier, M., Diop, N., & Pain, J.-P. (2012a). Coconut water preservation and processing: A review. *Fruits*, *67*(3), 157–171. https://doi.org/10.1051/fruits/2012009

Prades, A., Dornier, M., Diop, N., & Pain, J.-P. (2012b). Coconut water uses, composition and properties: A review. *Fruits*, *67*(2), 87–107. Cambridge Core. https://doi.org/10.1051/fruits/2012002

Prades, A., Salum, U. N., & Pioch, D. (2016). New era for the coconut sector. What prospects for research? *OCL*, *23*(6). https://doi.org/10.1051/ocl/2016048

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Ramer, G., & Lendl, B. (2013). Attenuated total reflection fourier transform infrared spectroscopy. In *Encyclopedia of analytical chemistry*. American Cancer Society. https://doi.org/10.1002/9780470027318.a9287

Richardson, P. I. C., Muhamadali, H., Ellis, D. I., & Goodacre, R. (2019). Rapid quantification of the adulteration of fresh coconut water by dilution and sugars using Raman spectroscopy and chemometrics. *Food Chemistry*, *272*, 157–164. https://doi.org/10.1016/j.foodchem.2018.08.038

Rios-Corripio, M. A., Rojas-López*, M., & Delgado-Macuil, R. (2012). Analysis of adulteration in honey with standard sugar solutions and syrups using attenuated total reflectance-Fourier transform infrared spectroscopy and multivariate methods. *CyTA - Journal of Food*, *10*(2), 119–122. https://doi.org/10.1080/19476337.2011.596576

Rodriguez-Saona, L. E., Giusti, M. M., & Shotts, M. (2016). Advances in Infrared Spectroscopy for Food Authenticity Testing. In G. Downey (Ed.), *Advances in Food Authenticity Testing* (pp. 71–116). Woodhead Publishing. https://doi.org/10.1016/B978-0-08-100220-9.00004-7

Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny*. https://doi.org/10.1201/9780429447273

Sivakesava, S., Irudayaraj, J., & Korach, R. (2001). Detection of adulteration in apple juice using mid infrared spectroscopy. *Applied Engineering in Agriculture*, *17*. https://doi.org/10.13031/2013.6825

Stevens A, Ramirez-Lopez L (2020). An introduction to the prospectr package. R package version 0.2.0. https://antoinestevens.github.io/prospectr/

Tang, Y., Horikoshi, M., & Li, W. (2016). ggfortify: Unified Interface to Visualize Statistical Results of Popular R Packages. *R J*. https://doi.org/10.32614/rj-2016-060

Tipson, R. S. (1968). *Infrared Spectroscopy of Carbohydrates: A Review of the Literature*. https://doi.org/10.6028/nbs.mono.110

Valenzuela, C., & Rodriguez-Llamazares, S. (2016). *spftir: Pre-Processing and Analysis of Mid-Infrared Spectral Region* (0.1.0) [Computer software]. https://www.rdocumentation.org/packages/spftir

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists*, *2*(1), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9

Yong, H. J. W., Ge, L., Ng, Y. F., & Tan, S. N. (2009). The Chemical Composition and Biological Properties of Coconut (Cocos nucifera L.) Water. *Molecules*, *14*(12), Article 12. https://doi.org/10.3390/molecules14125144

Table 1. Statistical parameters and figures of merit for PLS model. RMSE (root mean squared error) of calibration (RMSEC), cross-validation (RMSECV), prediction (RMSEP), y_{pred} is the predicted value, y_{obs} is observed value, y_{avg} is the average of the observed value, systemic error of the model indicated as Bias; coefficients of determination (R^2) were calculated for calibration (R_{cal}) cross-validation (R_{cv}) and test (R_{test}) datasets. Limit of detection (LOD), S_{blank} is the standard deviation of the blank, which in this case is estimated from the standard deviation of the calibration line (Sy/x).

Figures of merit (FOM)		**Glucose**	**Fructose**	**Sucrose**	**Mixed**	**HFCS-42**	**HFCS-55**
Latent variables		1	2	1	3	3	4
$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_{pred} - y_{obs})^2}{n-1}}$	RMSEC (%v/v)	1.269	1.070	1.441	0.886	3.322	0.892
	RMSECV(%v/v)	1.364	1.705	1.648	2.600	5.629	3.054
	RMSEP (%v/v)	1.672	0.327	2.056	3.870	5.441	1.676
$R^2 = \frac{\sum_{i=1}^{n}(y_{pred} - y_{obs})^2}{\sum_{i=1}^{n}(y_{pred} - y_{avg})^2}$	R^2_{cal}	0.998	0.998	0.998	0.999	0.993	0.999
	R^2_{cv}	0.998	0.996	0.997	0.992	0.960	0.989
	R^2_{test}	0.997	1.000	0.981	0.980	0.992	0.996
$Bias = \frac{\sum_{i=1}^{n}(y_{pred} - y_{avg})}{n}$		-0.133	0.125	-0.051	0.128	-0.774	0.127
$LOD = 3.3 \times \frac{S_{blank}}{b}$, where s_{blank} is estimated from $S_{y/x} = \sqrt{\frac{\sum_{i}^{n}(y_{obs} - y_{pred})^2}{n-2}}$		4.6	2.7	5.5	4.6	4.5	2.2

Figure Captions

Figure 1. The raw ATR-FTIR spectra of fresh coconut water (CW) and six different sugar solutions used for adulteration experiment (A) and the selected band region for the analysis of sugars (B). The spectral bands of standard glucose, fructose, and sucrose solutions that were used for calibration using standard addition method are shown as G, F, S respectively.

Figure 2. The 1st and 2nd derivative spectra (A, C and corresponding gap-segment derivative spectra (B, D) performed following the Savitzky-Golay smoothing using the gapDer algorithm.

Figure 3. PCA plots based on gap-segment derivative spectra of pure and adulterated coconut water (CW) via substitution with different sugar solutions at incremental levels. PC1 vs PC2 and PC1 vs PC3 plots were shown for the 1st (A, B) and 2nd (C, D) derivatives. The increasing size of the symbols shows the increasing level substitution by each sugar solution (0 - 100% at 5% increments).

Fig 4 A=PCA plot of ATR-FTIR spectra (950 -1200cm-1) for the freshly extracted reference coconut water (CW) and 31 commercially packaged products with (Y) or without (N) "added sugar" claim. B= PCA model plot of experimentally adulterated CW (0-100% level of substitution) and the projection of supplementary data (packaged coconut water samples collected from the market) using the co-ordinates predicted by the PCA model. The supplementary data with "no added sugar" and "added sugar" claim are shown as open and closed circles, respectively.

Figure 5. Plots of measured vs PLS predicted values of sugar adulterants in coconut water following 80:20 train-test data split and leave-one-out cross-validation (LOO-CV). Values shown for calibration (calib), cross-validation (cvalid) and test sets. Dashed lines shows the theoretical 1:1 relationship.

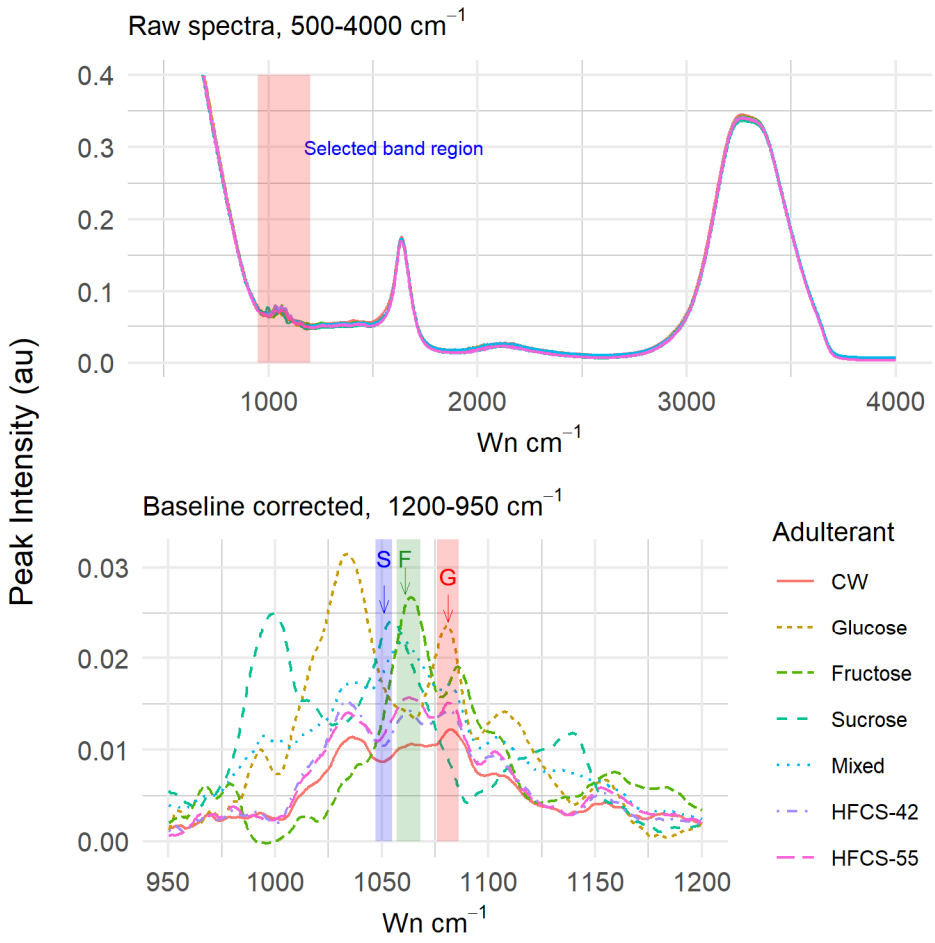

Figure 1. The raw ATR-FTIR spectra of fresh coconut water (CW) and six different sugar solutions used for adulteration experiment (A) and the selected band region used for analysis of sugars that is not affected by infrared absorption of water (B). The intensity bands of standard glucose, fructose and sucrose solution that were used for calibration using standard addition method are shown as G, F, S respectively.

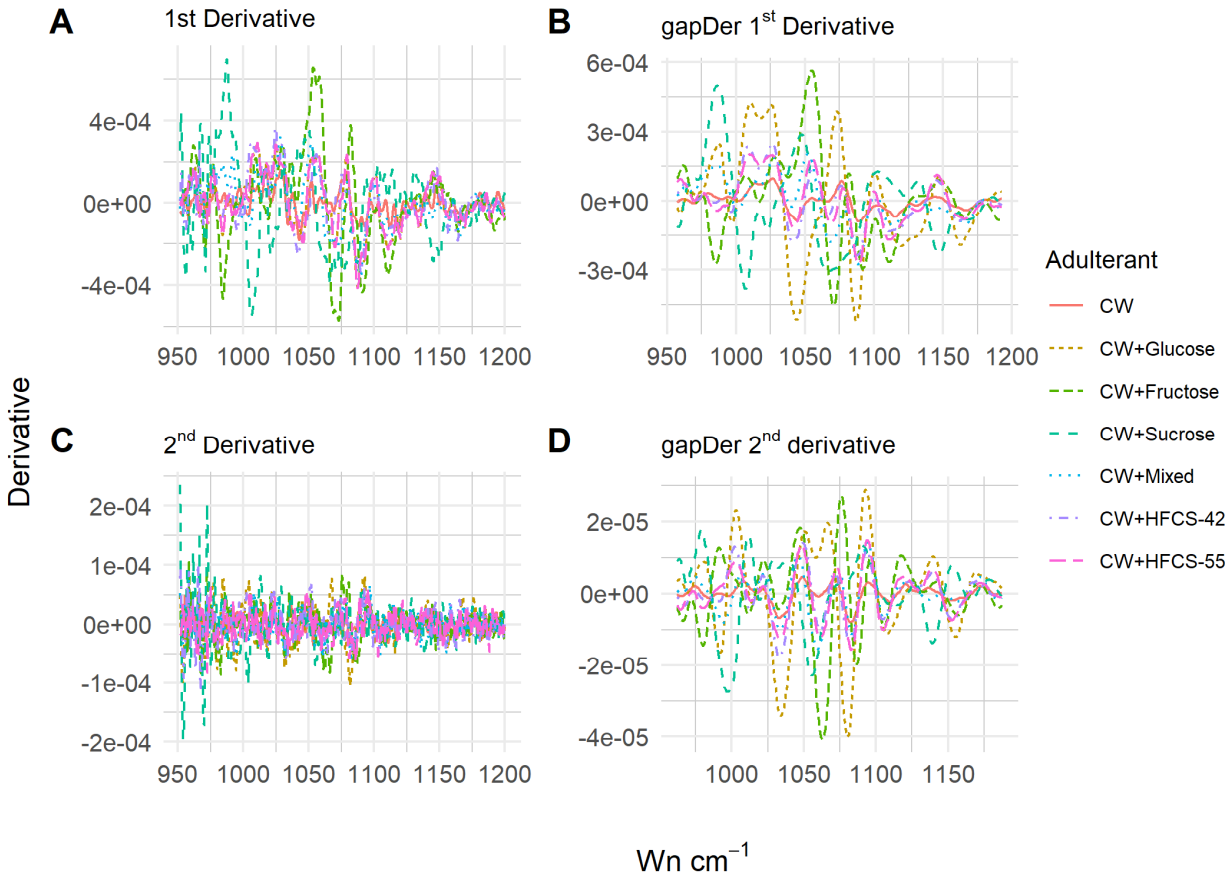

Figure 2. The 1st and 2nd derivative spectra and corresponding gap-segment derivative spectra performed following the Savitzky-Golay smoothing using the gapDer algorithm.

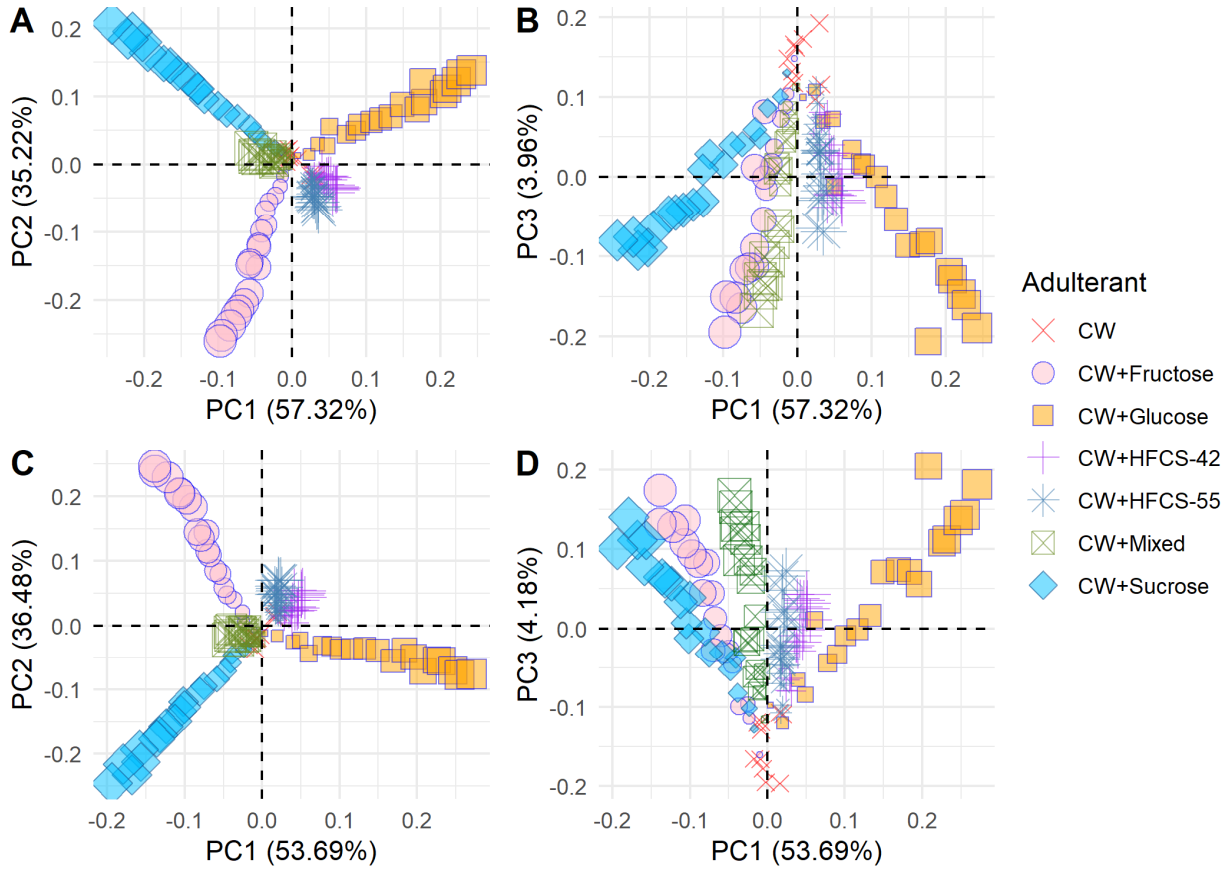

Figure 3. PCA plots based on gap-segment derivative spectra of pure and adulterated coconut water (CW) via substitution with different sugar solutions at incremental levels. PC1 vs PC2 and PC1 vs PC3 plots were shown for the 1st (A and B) and 2nd (C and D) derivatives. The increasing size of the symbols shows the increasing level substitution by each sugar solution (0 - 100% at 5% increments).

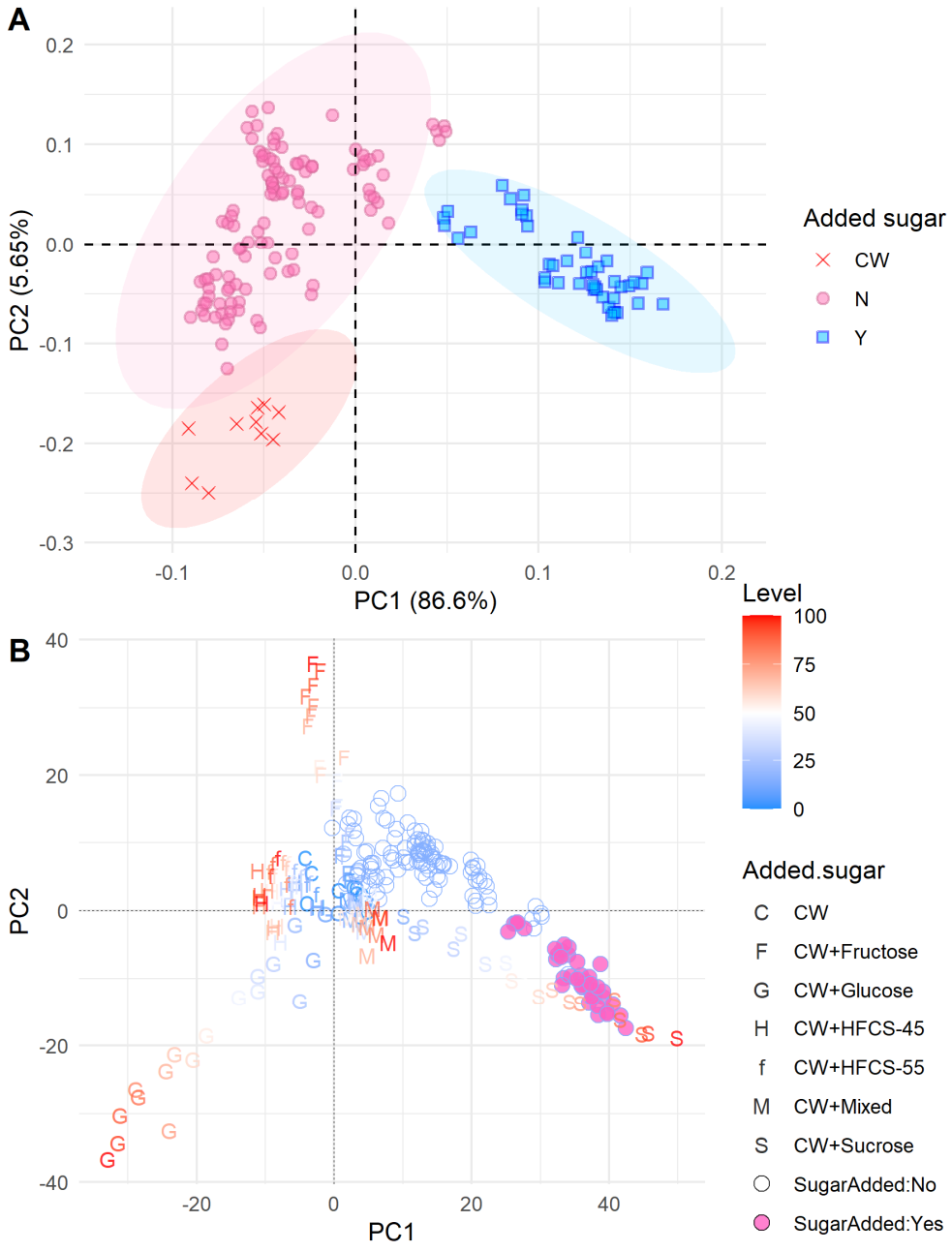

Fig 4 A=PCA plot of ATR-FTIR spectra (950 -1200cm-1) for the freshly extracted reference coconut water (CW) and 31 commercially packaged products with (Y) or without (N) added sugar label. B= PCA model plot of experimentally adulterated CW (0-100% level of substitution with different sugary solutions at 5% increments) and the projection of supplementary data (packaged coconut water samples collected from the market) using the coordinates predicted by the PCA model. The supplementary data with "no added sugar" and "added sugar" claim are shown as open and closed circles, respectively.

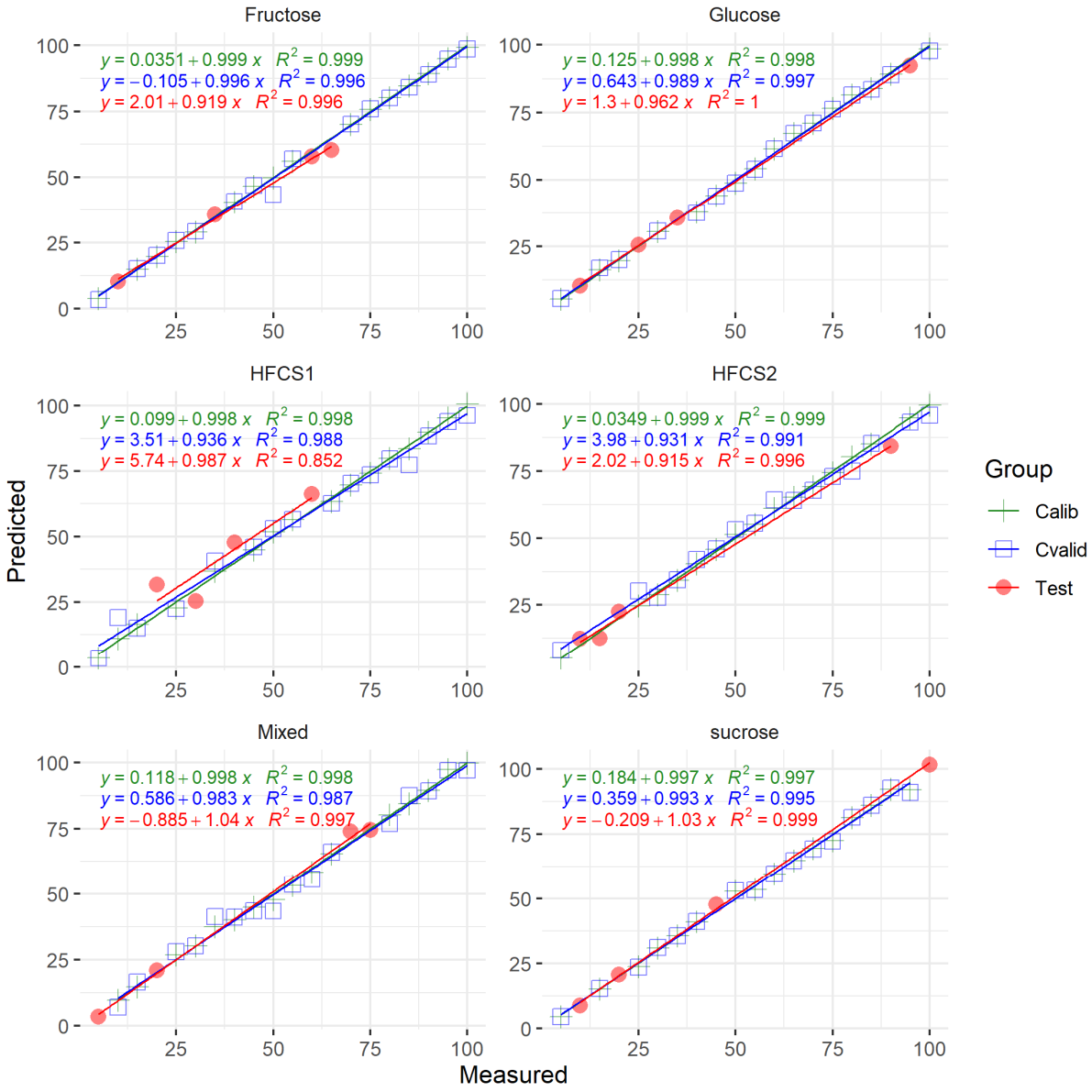

Figure 5. Plots of measured vs PLS predicted values of sugar adulterants in coconut water following 80:20 train-test data split and leave-one-out cross-validation (LOO-CV). Values shown for calibration (calib), cross-validation (cvalid) and test sets.

Table S1. List of commercially packaged coconut water samples and ingredients. Codes A-S represents the different brands and superscripts following the brand code indicate the same brand but different product lines. The presence of a particular ingredient (Y) and absence (N) is based on the information given on the product label. SMS = sodium metabisulphite, SMC = sodium carboxymethyl cellulose; MA= malic acid; CW= fresh coconut water extracted from fruit.

Product	Brix (°bx)	Mixed	Organic	Origin	Pulp	Added Sugar	Gelling	Anti-oxidant	SMBS	SCMC	Flavour	Sulphite	Vit.C	Pressed	MA
CW	7.4	N	N	Unknown	N	N	N	N	N	N	N	N	N	N	N
A	5.0	N	Y	Viet Nam	N	N	N	N	N	N	N	N	N	N	N
B	5.1	N	N	Thailand	Y	N	$Y_{(E418)}$	$Y_{(E223)}$	N	N	N	N	N	N	N
$C^{1,2,3}$	5.2^1, 5.7^2, 7.1^3	$N^{1,2}$,,$Pineapple^3$	N	Thailand	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$Y^{1,2,3}$	$N^{1,2}$,Y^3	$N^{1,2,3}$
$D^{1,2,3}$	5.9^1,8.5^2,6.4^3	$N^{1,2,3}$	$N^{1,2,3}$	Thailand	$N^{1,2,3}$	$N^{1,3}Y^2$,	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,3}Y^2$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$
$E^{1,2,3,4}$	5.0^1, 6.1^2,5.1^3,5.1^4	$N^{1,3,4}$,$Aloe\ vera^2$	$N^{1,2,4}Y^3$	Philippines	Y^1,$N^{2,3,4}$	$N^{1,2,3,4}$	$N^{1,2,3,4}$	$N^{1,2,3,4}$	$N^{1,2,3,4}$	$N^{1,2,3,4}$	$N^{1,2,3,4}$	$N^{1,2,3,4}$	$N^{1,2,3,4}$	$N^{1,2,3,4}$	$N^{1,2,3,4}$
$F^{1,2,3}$	4.8^1,5.4^2,8.0^3	$N^{1,2,3}$	$N^{1,2,3}$	Thailand	$N^{1,2}Y^3$	$N^{1,2}Y^3$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2}Y^3$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$
$G^{1,2}$	8.5	N	N	Thailand	Y	Y	N	N	N	N	N	N	N	N	N
H	6.0	N	N	Thailand	N	N	N	N	N	N	N	N	N	N	N
I	7.2	N	N	Philippines	Y	Y	N	N	N	Y	Y	Y	N	N	N
$J^{1,2,3}$	4.5^1,8.7^2,5.7^3	$N^{1,2,3}$	$Y^1N^{2,3}$	Thailand	$N^{1,3}Y^2$	$N^{1,3}Y^2$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$	$N^{1,2,3}$
K	8.4	N	N	Thailand	Y	Y	N	N	N	N	N	N	N	N	N
$L^{1,2}$	5.5^1, 5.6^2	$N^{1,2}$	Y^1, N^2	Philippines	$N^{1,2}$	$N^{1,2}$	$N^{1,2}$	$N^{1,2}$	$N^{1,2}$	$N^{1,2}$	$N^{1,2}$	$N^{1,2}$	$N^{1,2}$	$N^{1,2}$	$N^{1,2}$
M	5.3	N	N	NewZealand	N	N	N	N	N	N	N	N	N	N	N
N	5.5	N	N	Indonesia	N	N	N	N	N	N	N	N	N	N	N
O	7.8	N	N	Thailand	Y	Y	N	N	Y	N	N	Y	N	N	N
P	5.3	N	N	Thailand	N	N	N	N	N	N	N	N	N	N	N
Q	8.7	N	N	Thailand	Y	Y	N	N	Y	Y	N	Y	N	N	Y
R	8.6	N	N	Thailand	Y	Y	N	N	N	N	N	Y	N	N	N
S	6.2	N	N	Thailand	N	N	N	N	N	N	N	N	N	N	N

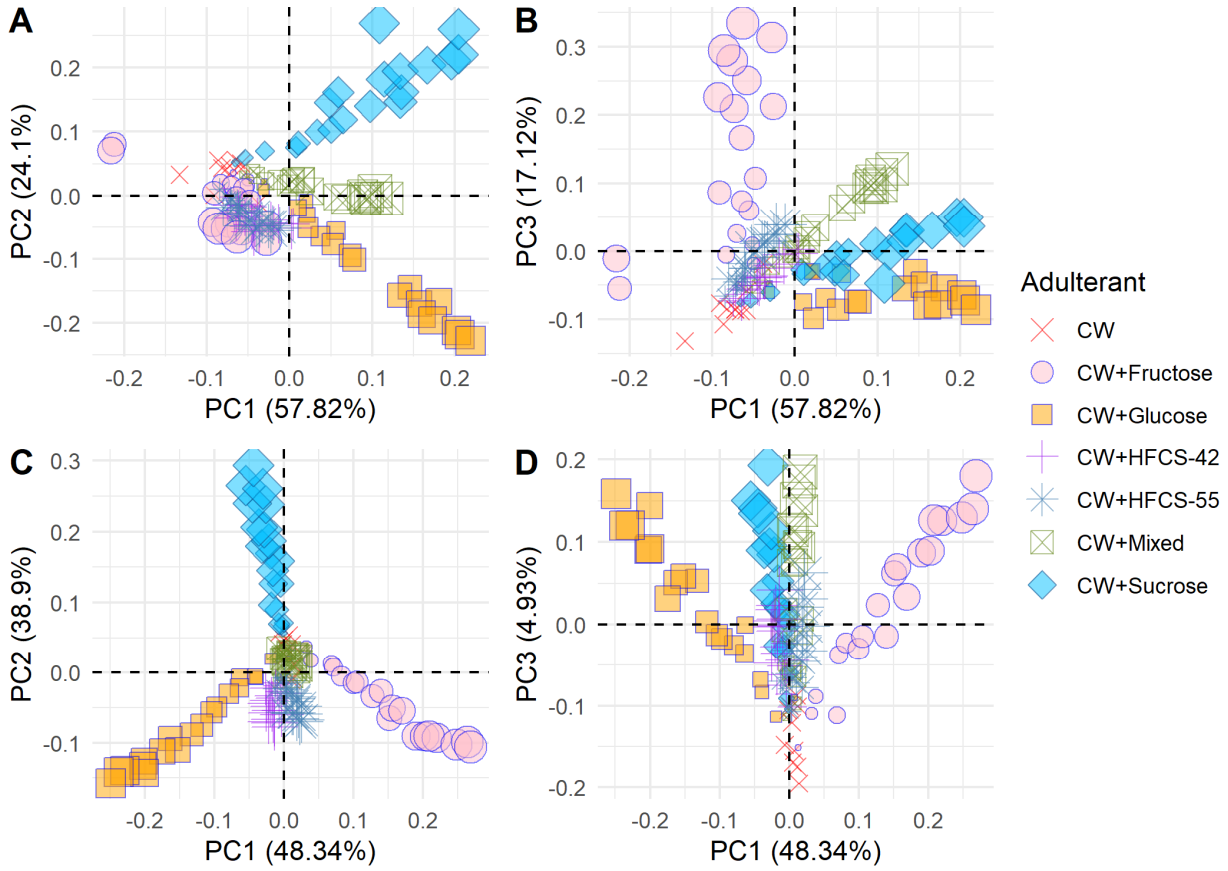

Figure S1. PCA score plots based on gap-segment derivative spectra of pure and adulterated coconut water (CW) via substitution with different sugar solutions at incremental levels. PC1 vs PC2 and PC1 vs PC3 plots were shown for the 3rd (A and B) and 4th (C and D) derivatives. The increasing size of the symbols shows the increasing level substitution by each sugar solution (0 - 100% at 5% increments).

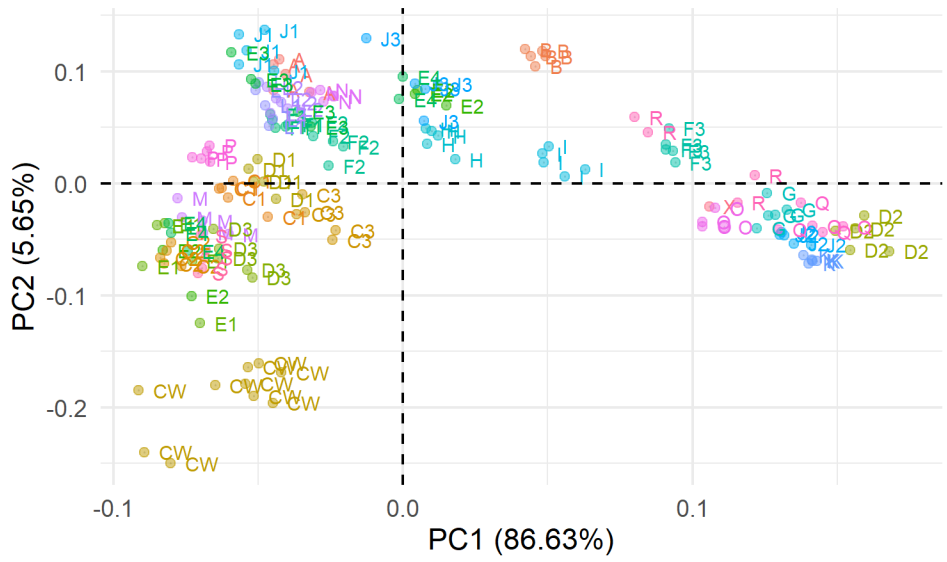

Figure S2. PCA plots for reference coconut water extracted from young coconut fruit (CW) and 31 different commercial coconut water samples (coded A to S, see table S1). Each brand is represented by 5 replicate data. The same character code followed by a number (1, 2, 3, or 4) indicates the same brand but a different product line.

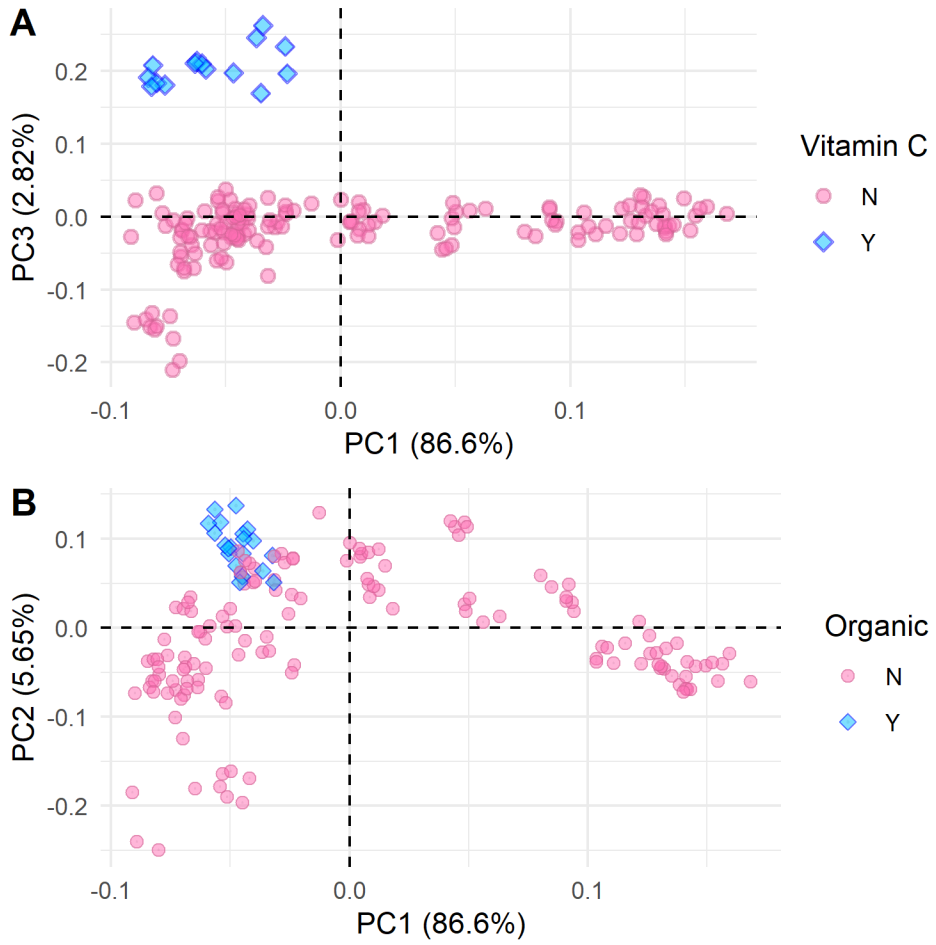

Figure S3. PCA score plots based on gap-segment derivative ATR-FTIR spectra (950-1200 cm^{-1}) for the freshly extracted reference coconut water and the 31 different products and product lines of packaged coconut water samples labeled by A= the presence(Y) or absence of (N) of added vitamin C and B= whether the product is organic (Y) or not organic (N).

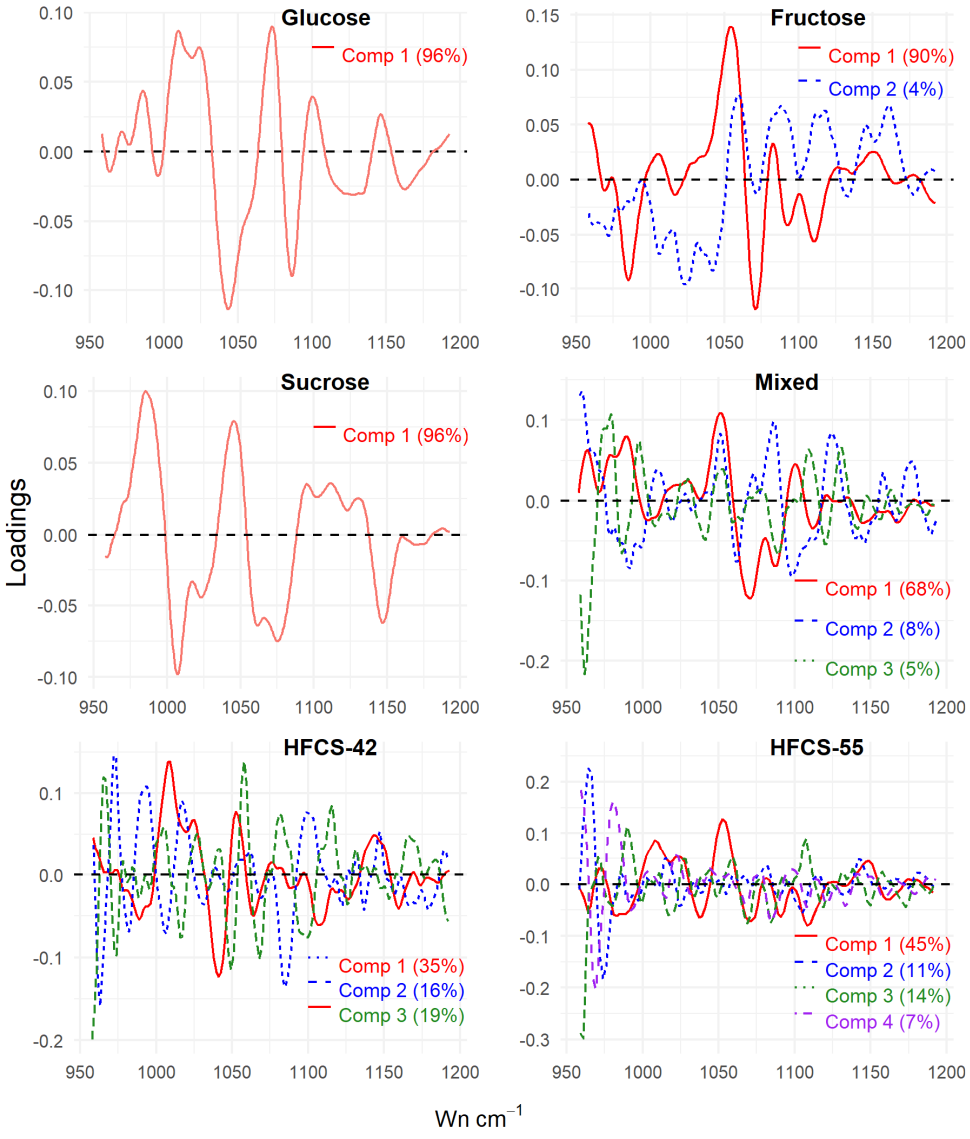

Figure S4. PLS loading plot of the latent variables and variance explained by each component.