Electronic identifier: 23581

Date of electronic submission:

The University of Manchester makes examined, Open Access electronic theses freely available for download and reading online via the University’s Research Explorer http://www.research.manchester.ac.uk.

This print version of my thesis is a TRUE and ACCURATE REPRESENTATION of the electronic version submitted via the University of Manchester's eThesis submission system.
The Development of Spectroscopic Methods for the Detection of Adulteration in Coconut Water

A thesis submitted to the University of Manchester for the degree of Master of Philosophy in the Faculty of Science and Engineering

2018
Paul Richardson
School of Chemistry
## Contents

Table of Abbreviations ........................................................................................................... 5

Abstract ..................................................................................................................................... 6

Declaration & Copyright Statement .......................................................................................... 7

Chapter 1: Introduction ............................................................................................................ 8

  The challenges of preserving coconut water ................................................................. 11
  Food adulteration and fraud and its detection ............................................................. 12

Spectroscopic methods .......................................................................................................... 15

  Fourier Transform Infrared (FTIR) spectroscopy ...................................................... 16
  Raman Spectroscopy ........................................................................................................ 19
  NMR spectroscopy ........................................................................................................... 22
  High performance liquid chromatography (HPLC) .................................................... 26
  Refractive index detection (RID) ................................................................................... 28

Chemometrics and multivariate analysis .............................................................................. 28

  Principal component analysis (PCA) ................................................................. 29
  Principal Component - Discriminant function analysis (PC-DFA) ...................... 29
  Partial least squares regression (PLSR) ........................................................................ 30

Aims and objectives .............................................................................................................. 30

Chapter 2: Detecting the presence of bovine milk as an adulterant in coconut water .......... 31

  Introduction ....................................................................................................................... 31
  Materials and Methods ................................................................................................. 33

  Coconut water and Stock preparation ........................................................................ 33
  FT-IR ........................................................................................................................... 33
  Raman ......................................................................................................................... 33
  HPLC-RI ...................................................................................................................... 34
Chapter 3: Detecting the dilution of coconut water and subsequent masking with sugars using Raman spectroscopy

Introduction:

Abstract

Materials and methods

Stock solution

Standard addition

Detection of adulteration

Raman spectroscopy

Data analysis

Results and discussion

Standard addition

Adulteration with solutions of single sugars

Adulteration with a mixed sugar solution

Quantitative analysis
Chapter 4: Detecting the adulteration of coconut water with a mixed sugar solution using nuclear magnetic resonance spectroscopy ............................................................. 76

Introduction: ............................................................................................................. 76

Methodology: ........................................................................................................... 77

Adulterant solution preparation: .............................................................................. 77

NMR analysis: ......................................................................................................... 77

Data analysis: .......................................................................................................... 77

Results: .................................................................................................................... 78

Sugar section: .......................................................................................................... 81

Coconut water section: ......................................................................................... 84

PLSR: ......................................................................................................................... 86

Drift in the chemical shift of malic acid signals: ................................................. 87

Chapter 5: Conclusions and future work ................................................................. 91

Appendix 1 ................................................................................................................ 98

Appendix 2 ................................................................................................................ 103

References ............................................................................................................... 104

Word count: 26,402
## Table of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATR</td>
<td>Attenuated total reflectance</td>
</tr>
<tr>
<td>AU</td>
<td>Arbitrary units</td>
</tr>
<tr>
<td>CV</td>
<td>Cross validation</td>
</tr>
<tr>
<td>CW</td>
<td>Coconut water</td>
</tr>
<tr>
<td>DF</td>
<td>Discriminant function</td>
</tr>
<tr>
<td>DNase</td>
<td>Deoxyribonuclease</td>
</tr>
<tr>
<td>DSS</td>
<td>4,4-dimethyl-4-silapentane-1-sulfonic acid</td>
</tr>
<tr>
<td>EMSC</td>
<td>Extended multiplicative scattering correction</td>
</tr>
<tr>
<td>FT-ICR MS</td>
<td>Fourier transform ion cyclotron resonance mass spectrometry</td>
</tr>
<tr>
<td>FTIR</td>
<td>Fourier transform infrared</td>
</tr>
<tr>
<td>GC</td>
<td>Gas chromatography</td>
</tr>
<tr>
<td>HFCS</td>
<td>High-fructose corn syrup</td>
</tr>
<tr>
<td>HILIC</td>
<td>Hydrophilic interaction chromatography</td>
</tr>
<tr>
<td>HPCD</td>
<td>High pressure carbon dioxide pasteurisation</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>IBS</td>
<td>Irritable bowel syndrome</td>
</tr>
<tr>
<td>IRMS</td>
<td>Isotopic ratio mass spectrometry</td>
</tr>
<tr>
<td>LOD</td>
<td>Limit of detection</td>
</tr>
<tr>
<td>LV</td>
<td>Latent variable</td>
</tr>
<tr>
<td>MIR</td>
<td>Mid-infrared</td>
</tr>
<tr>
<td>NIR</td>
<td>Near-infrared</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>OIV</td>
<td>Office Internationale des Vins (International Office of Wines)</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PC-DFA</td>
<td>Principal component-discriminant function analysis</td>
</tr>
<tr>
<td>PLSR</td>
<td>Partial least squares regression</td>
</tr>
<tr>
<td>POD</td>
<td>Peroxidase</td>
</tr>
<tr>
<td>ppm</td>
<td>Parts per million</td>
</tr>
<tr>
<td>PPO</td>
<td>Polyphenol oxidase</td>
</tr>
<tr>
<td>RF</td>
<td>Radio frequency</td>
</tr>
<tr>
<td>RID</td>
<td>Refractive index detection</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root-mean-squared error</td>
</tr>
<tr>
<td>RMSECV</td>
<td>Root-mean-squared error on the cross-validation set</td>
</tr>
<tr>
<td>RMSEP</td>
<td>Root-mean-squared error on prediction of the test set</td>
</tr>
<tr>
<td>RNase</td>
<td>Ribonuclease</td>
</tr>
<tr>
<td>RR</td>
<td>Resonance Raman</td>
</tr>
<tr>
<td>SAM</td>
<td>Standard addition method</td>
</tr>
<tr>
<td>SERS</td>
<td>Surface enhanced Raman scattering</td>
</tr>
<tr>
<td>SNIF-NMR</td>
<td>Site-specific natural isotope fractionation nuclear resonance</td>
</tr>
<tr>
<td>SORS</td>
<td>Spatially offset Raman spectroscopy</td>
</tr>
<tr>
<td>TERS</td>
<td>Tip enhanced Raman scattering</td>
</tr>
<tr>
<td>TEV</td>
<td>Total explained variance</td>
</tr>
<tr>
<td>trn</td>
<td>Training set</td>
</tr>
<tr>
<td>tst</td>
<td>Test set</td>
</tr>
<tr>
<td>UV-Vis</td>
<td>Ultraviolet-Visible light</td>
</tr>
</tbody>
</table>
Abstract

Adulteration of food, defined as the fraudulent manipulation of the contents of products for economic gain, is a practice that is both costly and potentially hazardous to the health and cultural sensitivities of consumers. It comes in many forms, the most common of which are undeclared substitution of a product with less expensive and lower quality alternatives or the addition of external products to alter organoleptic characteristics. Coconut water is a refreshing tropical beverage which has greatly risen in popularity in the last five years. Its popularity, along with its high manufacturing and retail costs, has led it to be highly susceptible to fraudulent practices such as stretching and mislabeling. Currently, the common methods of detecting adulteration often involve enzymatic reactions or time consuming and resource intensive isotope fractionation methods. Spectroscopic technologies have the potential of providing versatile and accurate measures of adulteration in an inexpensive and high-throughput manner, and could therefore be used as screening methods in a commercial setting. In this work, spectroscopic technologies are presented to provide solutions in two scenarios: the adulteration of coconut water with bovine milk, and stretching with water while using various ratios of glucose, sucrose and fructose as masking agents. Using FTIR combined with PLSR, linear quantification of bovine milk in coconut water was achieved at a concentration of ~0.3%. In comparison, HPLC-RI was unable to match this despite a complete separation of lactose, the target analyte, from sugars naturally present in coconut water. Additionally, Raman spectroscopy in combination with chemometrics was found to be a powerful tool in studying the masking of diluted coconut water. When dilution was masked by keeping the total sugar concentration constant with a single sugar, linear quantification within 5% was readily achieved. However, masking using a mixed sugar solution emulating the natural concentrations of each sugar lessened the ability of Raman spectroscopy to detect this form of masking due to the overwhelming presence of peaks attributed to sugars. Finally, NMR was chosen as a comparative method to study the masking of coconut water with a mixed sugar solution, and was able to clearly detect stretching even when individual sugar concentrations were kept constant. An adulteration related drift in the chemical shift of malic acid signals was also observed, providing potential for further research.
Declaration & Copyright Statement

The author declares that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university of other institute of learning.

The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see [http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420](http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420)), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see [http://www.library.manchester.ac.uk/about/regulations/](http://www.library.manchester.ac.uk/about/regulations/)) and in The University’s policy on Presentation of Theses.
Chapter 1: Introduction

Coconuts are a tropical fruit commonly harvested in the coastal regions of several areas including South East Asia, Sri Lanka, Brazil, and West Africa. The outside of the coconut consists of a hard shell, acting as a seed, surrounded by a softer husk. Inside the shell lies the endosperm which can be divided into two components: an initially gelatinous and translucent solid layer coating the inner walls of the coconut shell known as the meat or copra; and a clear liquid filling the cavity within known as coconut juice or coconut water. The appearance, biological compositions, and relative quantities of these components all alter during the maturation (Vigliar et al. 2006). As the coconut matures, the meat becomes more opaque and hardens while the coconut water decreases in volume, total sugar concentration, and monosaccharide to disaccharide ratio (Child & Nathanael 1950). As such, coconuts of different maturity are picked based on the components required: immature coconuts are used for their higher quantity and quality of water, while mature coconuts are used in the manufacturing of products involving the meat.

Due to their different composition, coconut meat and coconut water have different uses. The meat is generally grated, with or without the addition of water, to create coconut milk and coconut butter or oil. As copra is composed primarily of water, fat and protein (Seow & Gwee 1997), its manufactured products are therefore suited as a vegan source of fat in cooking or in the cosmetic industry. Coconut water, on the other hand, contains mainly water (~94%) and sugar (3.5-5%) (Yong et al. 2009), along with a multitude of minerals and organic compounds that contribute to its taste and pleasant aroma (Prades, Assa, et al. 2012). One should note that, while the terms coconut milk and coconut water have been used interchangeably in the past, they are fundamentally different products with different properties and uses.

Coconut water is a refreshing and nutritious drink traditionally extracted from immature (6-9 months) (Rolle 2007) coconuts and consumed in tropical countries where coconuts can be found. Due to a high mineral content and pleasant taste, it is gaining popularity worldwide as a natural isotonic beverage to compete against common sports drinks (FAO 2000). Thanks to this and its multiple perceived health benefits, the market for coconut water has exploded, with one of the major vendors reaching sales of nearly $1bn globally.
In addition to being a refreshing beverage, coconut water has been used in a number of medical contexts. Several studies have shown a positive effect of coconut water in the treatment of hypertension (Bhagya et al. 2012; Alleyne et al. 2005) and high cholesterol levels (Sandhya & Rajamohan 2008), along with multiple anti-ageing effects. Some of these include *in vivo* and *in vitro* anti-oxidant effects on DNA (Bispo et al. 2017) along with inhibition of the enzyme acetylcholinesterase responsible in part for neural degenerative diseases (Kim et al. 2008). Moreover, its high content of B vitamins and vitamin C (USDA 2016) provide more evidence for its likely health benefits. Excluding traditional medicine, its earliest documented medical use was as an ersatz intravenous rehydration drip (Campbell-Falck et al. 2000) in regions where medical saline was unavailable. This is made possible by two factors. First, the chemical composition of coconut water is somewhat similar to the chemical composition of plasma: notably while concentrations are different, it contains several important electrolytes including K\(^+\), Ca\(^{2+}\), Na\(^+\), Cl\(^-\), and Mg\(^{2+}\) (Petroianu et al. 2004), along with amino acids proline and glutamine (Eiseman 1954), among others. Table 1.1 depicts some relevant components and their concentrations. Additionally – possibly critically – the fluid is sterile within the coconut, and can therefore be used with little to no prior preparation (Campbell-Falck et al. 2000). The first instances of such a use of coconut water are rumoured to be during WWII by the British and Japanese forces in Ceylon (now Sri Lanka) and Sumatra respectively, while initial reports in the literature concerning the practice date back to Eiseman in 1954 (Eiseman 1954). Since then, several studies have confirmed the suitability of coconut water as an emergency short-term replacement intravenous drip when conventional saline solution is unavailable (Campbell-Falck et al. 2000; Petroianu et al. 2004). However, it should be noted that due to high concentrations of potassium, calcium, and magnesium, the possibility of cardiac and renal complications increases. Indeed, while restricted quantities of coconut water can be drunk with no repercussions, there have been at least two reported cases associating emergency hospital treatment for hyperkalaemia (Fagundes-Neto et al. 1993; Yartey et al. 1993) with excess consumption of coconut water. Overall, it is clear that while there is evidence of its health benefits and strong arguments for its use as rehydration fluid when
conventional options are unavailable, it is imperative that moderation and caution be exercised.

Table 1.1: Some relevant contents of young coconut water (Yong et al. 2009)

<table>
<thead>
<tr>
<th>Macroingredients</th>
<th>(g/100g) (age unspecified)</th>
<th>(g/100g) (6 months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>94.99</td>
<td>94.18</td>
</tr>
<tr>
<td>Total dry weight</td>
<td>5.01</td>
<td>5.82</td>
</tr>
<tr>
<td>Protein</td>
<td>0.72</td>
<td>0.12</td>
</tr>
<tr>
<td>Lipids</td>
<td>0.2</td>
<td>0.07</td>
</tr>
<tr>
<td>Ash</td>
<td>0.39</td>
<td>0.87</td>
</tr>
<tr>
<td>Carbohydrate (by difference)</td>
<td>3.71</td>
<td>4.76</td>
</tr>
<tr>
<td>Dietary fiber</td>
<td>1.1</td>
<td>Not detected</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sugars</th>
<th>(g/100g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sucrose</td>
<td>0.06</td>
</tr>
<tr>
<td>Glucose</td>
<td>2.61</td>
</tr>
<tr>
<td>Fructose</td>
<td>2.55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inorganic Ions</th>
<th>(mg/100g)</th>
<th>(mg/100g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calcium</td>
<td>24</td>
<td>27.35</td>
</tr>
<tr>
<td>Iron</td>
<td>0.29</td>
<td>0.02</td>
</tr>
<tr>
<td>Magnesium</td>
<td>25</td>
<td>6.40</td>
</tr>
<tr>
<td>Phosphorus</td>
<td>20</td>
<td>4.66</td>
</tr>
<tr>
<td>Potassium</td>
<td>250</td>
<td>203.70</td>
</tr>
<tr>
<td>Sodium</td>
<td>105</td>
<td>1.75</td>
</tr>
<tr>
<td>Zinc</td>
<td>0.1</td>
<td>0.07</td>
</tr>
<tr>
<td>Copper</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>Manganese</td>
<td>0.142</td>
<td>0.12</td>
</tr>
<tr>
<td>Sulfur</td>
<td></td>
<td>0.58</td>
</tr>
<tr>
<td>Aluminium</td>
<td></td>
<td>0.07</td>
</tr>
<tr>
<td>Boron</td>
<td></td>
<td>0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vitamins</th>
<th>(mg/100g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thiamine (B1)</td>
<td>0.03</td>
</tr>
<tr>
<td>Riboflavin (B2)</td>
<td>0.057</td>
</tr>
<tr>
<td>Niacin (B3)</td>
<td>0.08</td>
</tr>
<tr>
<td>Pantothenic acid (B5)</td>
<td>0.043</td>
</tr>
<tr>
<td>Pyridoxine (B6)</td>
<td>0.032</td>
</tr>
<tr>
<td>Folate</td>
<td>0.03</td>
</tr>
<tr>
<td>Ascorbic acid (C)</td>
<td>2.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phytohormones (Total)</th>
<th>(nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auxin</td>
<td>150.6</td>
</tr>
<tr>
<td>Cytokinins</td>
<td>186.46</td>
</tr>
<tr>
<td>Gibberellins</td>
<td>54.5</td>
</tr>
</tbody>
</table>

In addition to consumption, the efficacy of coconut water has been tested in several other situations. Its suitability as a storage medium for teeth awaiting re-insertion has been examined (Moura et al. 2017; Navit et al. 2017; de Souza et al. 2017) with varying levels of success, although it was consistently unable to match hospital saline solutions. Additionally, coconut water contains several phytohormones (Table 1.1) highly beneficial to plant growth.
with beneficial roles in the promotion of cell division (Schaller et al. 2014), the limitation of leaf ageing (Haberer & Kieber 2002), and the induction of plant stem and shoot growth (Dharmasiri et al. 2005).

The challenges of preserving coconut water

Coconut water, despite remaining sterile until extraction, has a labile composition. Along with susceptibility to bacterial contamination (Reddy et al. 2005), exposure to aerobic conditions activates polyphenol oxidase (PPO) and peroxidase (POD) enzymes (Matsui et al. 2008; Murasaki-Aliberti et al. 2009), both responsible for Maillard-type oxidation reactions. These result in discolouration of the liquid, a decrease in pH (from ~5 to <3), a rancid taste and smell, and an overall shelf-life of 24 h at room temperature (Rolle 2007). While the effects can be somewhat postponed through refrigeration, a major challenge in the manufacture of bottled coconut water has been to find methods to increase the shelf-life without sacrificing the delicate aroma and taste which gives the drink its popularity. Methods have therefore been developed to prevent bacterial infection and deactivate the enzymes responsible for oxidative reaction. These can largely be divided into two categories: thermal inactivation, borrowing from the principles of pasteurisation; and non-thermal inactivation which uses high pressure or filtration to extract or inactivate unwanted enzymes and microbes.

The most common thermally based sterilisation method for coconut water is pasteurisation, the practice of heating the coconut water to high temperatures for set periods of time to kill bacteria and inactivate oxidation enzymes before rapidly decreasing the temperature back to room temperature. This can either be performed at relatively low temperatures (~70-90°C) (Tan et al. 2014) for a period of 300-600 s or at high temperatures (~140°C) for a much shorter period (flash pasteurisation) of 15-60 s. This is a popular method in industry due to its relative simplicity and high throughput nature. Additives such as ascorbic acid (Campos et al. 1996) and sodium metabisulfite (Sucupira et al. 2015) can also be added to aid enzyme deactivation. Pasteurisation using microwave radiation has also been investigated and was found to achieve comparable ~10x faster at lower temperatures (92.2°C) (Matsui et al. 2008). However, the inhomogeneous heating effect of microwave radiation renders the practice very difficult to scale-up to industrial quantities. A major drawback to heat-based sterilisation, however, is significant damage to the balance of flavour inducing
compounds (Cappelletti et al. 2015), leading to a tangible negative alteration of taste, along with a detrimental effect on coconut water’s health benefits. As companies have marketed the product towards a health-conscious population, this perception can be seen as particularly damaging.

Other methods have been studied to counter the drawbacks posed by flash pasteurisation. Micro- and ultra-filtration have been fairly well researched alternatives, the former aiming to filter out microbial contaminants while the latter aiming to extract PPO and POD. Some studies have examined the suitability of such methods and observed a significant increase in shelf-life. However, most of the methodologies used either showed significant irreversible membrane resistance (Reddy et al. 2005) or used methods unsuitable for scaling-up (Jayanti et al. 2010). Also, little consideration was given for the effect on taste suitability (Nakano et al. 2011). Recently, Karmakar and De (Karmakar & De 2017) developed an up-scalable ultrafiltration method using ultra-thin hollow fibres potted in a PVC pipe, leading to an increase in shelf-life to 18 weeks while retaining acceptable taste.

Another non-thermal method studied is high pressure CO$_2$ (HPCD) pasteurisation, which involves the use of high pressure conditions to inactivate bacteria. This method can easily be up-scaled to an industrial level, remains safe and relatively inexpensive, and has been shown to successfully reduce bacterial load (Cappelletti et al. 2015) with far lower temperature conditions (35-40°C), resulting in a far weaker effect on the volatile fraction and taste when compared to thermal pasteurisation. However, HPCD requires 30 min, or 15 min when combined with high power ultrasounds (Cappelletti et al. 2014) to reach sterilisation levels achieved in 1 min by thermal pasteurisation. Similarly, while HPCD was able to inactivate oxidative enzymes in an apple juice model (Marszałek et al. 2017), it was shown to be ~40x slower than thermal methods. It is therefore not surprising that many companies have foregone taste authenticity in favour of faster manufacturing.

Food adulteration and fraud and its detection

Food adulteration is defined as the deliberate addition or substitution of products with ingredients of lesser quality for economic gain. This type of fraud can be separated into three main categories: stretching or bulking, which involves adding an inexpensive adulterant to increase the total volume or weight of a product; masking, which involves the addition of
foreign ingredients to hide the inferior quality of a product and give it a more appealing appearance, smell or taste; and mislabelling, which involves labelling products of lower quality as premium. Economically motivated adulteration and fraud of food products has likely been present since the start of food manufacture. First studied by Frederick Accum in 1820 (Accum 1820) and Dr. Arthur Hassall in 1855 (Hassal 1855), food fraud has been a constant source of scandals (Ellis et al. 2012a), especially in the last few decades with globalisation, complex international supply chains, and the introduction of easily accessible mass media and communication. While some of these were simply fraudulent adulteration with cheaper products like the infamous Horsegate scandal (Elliott 2014) in 2013 where beef products were partly substituted with horsemeat, many have had the potential for disastrous consequences to human health: the adulteration of milk and milk products with melamine (The Lancet 2009) in China in 2008; the adulteration of wine with methanol in Italy (Suro 1986) in 1986 and with diethylene glycol (a component of several anti-freeze products) in Austria in 1985 (Tagliabue 1985); and the scrapie-infected cattle feed in the UK which led the BSE and subsequent Creutzfeldt-Jakob scares of the 1990s (Cleeland 2009; Elliott 2014). While these incidents are among the more high-profile scandals due to their impact on human health and the nature of the fraud (i.e. use of a substance notoriously unsuitable for consumption), most examples of fraud involve the misrepresentation or mislabelling of products (e.g. labelling concentrated juices as "pure premium" or farmed fish as "line caught" for their premium label and cost) and adulteration with similar but less expensive products (e.g. beef replaced with pork or premium olive oil with vegetable variants). The impact of these types of fraud is three-fold: on a medical level, hidden adulterants can be carcinogenic or allergenic and lead to serious medical emergencies for certain subsets of the population, such as the presence of Sudan IV dye in chilli powder or hazelnut oil in olive oil; on a social level, unmentioned adulterants can lead people to consume foods that go against personal beliefs, like pork meat being considered unclean (kashrut or haram) and thus highly undesirable to those of Jewish and Muslim faith; and finally on an economic level, as fraud has been estimated to cost up to $40bn per year globally (Johnson 2014). To combat this costly and potentially hazardous problem, various organisations and research groups have begun to work on the discovery and application of analytical methods that could test for unwanted additives and adulterants quickly, efficiently, and non-invasively.
While there have only been few publicly known instances of adulteration of coconut water (Barlass 2015; Glotz 2017; Food Standards Agency 2015a; Sorias 2016b; Food Standards Agency 2015b), one of these incidents involved the undeclared presence of bovine milk in a coconut drink product, presumably as a whitening agent, and led to the death of two Australian children from anaphylactic shock. A robust detection method is therefore important so as to ensure such an incident is not repeated. From a more general outlook, it also has multiple properties which make it a prime target for such fraud. First and possibly most importantly, it is a premium product with an incredibly high demand. As demonstrated by its market turn-over and several celebrity endorsements, its refreshing taste and reputation as a healthier alternative to isotonic sports drinks and sodas have won over the hearts of the public even with its price being over 20% more expensive than other popular fruit juice products within the same brand (e.g. coconut water: 37 p/100 mL compared to orange and apple juice: 28.7 p/100 mL). Second, it has several costly manufacturing conditions, including transport costs from tropical regions and measures ensuring the delicate product is not damaged during processing or transit. Lowering costs through fraudulent means of adulteration and mislabelling products with the “fresh”, “pure”, “natural” or “no-added sugar” premium labels could therefore be an incredibly lucrative enterprise for groups willing to risk taking part in such fraudulent practices. Furthermore, while many examples of “common” and tropical fruit juice adulteration have been studied, the literature is incredibly scarce in that respect for coconut water.

One of the fundamental barriers to fraud detection is that, until only recently, it has been a purely reactive process. When a new instance of industrial fraud began, it was left unchallenged until its first detection; by then, huge economic losses would have already occurred and the potential health hazard has already been exposed to the population. For example, the 2008 Chinese melamine scandal only broke out in July 2008 after 16 infants were diagnosed with kidney stones. By then, the damage was such that 300,000 victims were estimated to have been affected, including 54,000 infants requiring hospital treatment. This is, however, a barrier that has historically been incredibly difficult to overcome, as adulteration can come up in a variety of forms that cannot reasonably be controlled proactively by government and third-party organisations, especially without assistance from manufacturers. However, spectroscopic methods are ideal for this task as they are generally
fast, can be tuned to very high sensitivity, and are often non-destructive. Thanks to these characteristics, they can make routine testing a relatively facile task for quality control groups and, in some cases, even be added within the manufacturing process. By creating these efficient control systems, the focus can shift towards more proactive and preventative measures.

Spectroscopic methods

Spectroscopy is a range of analytical methods that utilise different parts of the electromagnetic spectrum to induce excitation in the probed molecules, and measuring either the energy absorbed or emitted during the return to the ground state. As different kinds of excitations (electronic, vibrational or spin) require various wavelengths in order to obtain different kinds of information about the sample, each method has a specific role, advantages and setbacks. As such, the choice of spectroscopic method is an important aspect of experimental design.

The aim of this study is to identify the presence of an adulterant in samples through the use of a unique signal or abnormal ratio. The focus can also be shifted towards symptoms rather than cause; if a particular scenario (e.g. geographical provenance (Franke et al. 2008) or processing method (Liu et al. 2004)) leads to the presence of specific molecules or compounds, then the identification and quantification of these molecules/compounds via rapid screening or sampling is often a robust way of determining the occurrence of misrepresentation, adulteration or other forms of food fraud.

The ideal marker (or markers) for detection has several properties. It must be affected by the condition studied in a consistent manner, and not be affected by other factors that would falsify conclusions. Additionally, it must be readily detectable using analytical methods. In the study of food adulteration, such a marker will often be a constituent foreign to the product in question; for example, the presence of argemone oil, an inedible oil made from the Argemone Mexicana seeds, has been detected in edible mustard oil using the quantity of sanguinarine present in samples (Ghosh et al. 2005). While these ideal markers do not always exist, additional methods have been developed to deal with such situations and greatly increase the possible discerning power. In the case where no unique foreign constituents can be found for example, adulteration by using an abnormality in the ratios of certain
constituents, as is done to detect invert sugar syrup in apple juice by studying the $^{13}$C:$^{12}$C and $^2$H:$^1$H ratios (Kelly et al. 2003), could be undertaken. Additionally, the use of chemometric methods can enable the accurate detection of variance within entire spectra instead of individual peaks. This holistic approach is used in the determination of the botanical origin of honeys, where Fourier transform infrared (FTIR) spectra of test samples can be examined and compared against the previously acquired spectra of honeys of known origin using principal component analysis (Etzold & Lichtenberg-Kraag 2008). Thorough reviews characterising the use of spectroscopic methods in food adulteration by method and more recently food category have been written by Ellis et al. (Ellis et al. 2012a) and Hong et al. (Hong et al. 2017), respectively.

**Fourier Transform Infrared (FTIR) spectroscopy**

FTIR spectroscopy is a commonly used spectroscopic analytical method that uses infrared radiation to induce detectable bond vibrations in molecules. Bond vibrations are modelled on Hooke’s law of harmonic springs. As is described in Equation 1.1, the vibrational energy $\tilde{\nu}$ (represented here in wavenumbers) is equal to the square root of the force constant $k$ divided by the reduced mass $\mu$, divided by a multiple of the speed of light $c$. This reduced mass (Equation 1.2) is a manipulation of the atomic masses $m_1$ and $m_2$ of each vibrating atom to form a total inertial mass, allowing us to view the vibration as a one-body problem.

$$\tilde{\nu} = \frac{1}{2\pi c} \sqrt{\frac{k}{\mu}}$$  \hspace{1cm} \text{Equation 1.1}

$$\mu = \frac{m_1 m_2}{m_1 + m_2}$$  \hspace{1cm} \text{Equation 1.2}

These vibrations occur at set energy levels and will absorb the infrared light at wavelengths equating this energy, as depicted in Figure 1.1. An infrared spectrum is therefore constructed by comparing the intensity of light transmitted through the sample and that transmitted through a blank control measurement at each wavelength measured. The wavelengths at which light is absorbed can then be related to the presence of specific functional groups within a compound.
Unlike other spectroscopic methods like Raman and UV-Vis which use a monochromatic laser as a source, FTIR uses a polychromatic lamp which is modulated to acquire information at each measure wavelength. The light is focused using mirrors and passed through a Michaelson interferometer, where it is modulated to a single wavelength by splitting the light and reflecting each beam back to travel slightly different distances. As one of the reflecting mirrors can slide, the path length and subsequently the wavelength of the beam reflected changes. This newly recombined light is then shone on the sample after which the transmitted light is detected. The signal is detected as an interferogram, which is converted to a transmittance or absorbance spectrum using Fourier transformation.

![Jablonski diagram](image)

**Figure 1.1:** Jablonski diagram depicting a bond's potential energy well and the excitation to higher vibrational levels \(E_n\) as a result of FTIR light absorption.

Infrared light is generally divided into three categories: far-IR, mid-IR (MIR), and near-IR (NIR). Far-IR has been associated with thermal energy and is therefore used primarily in astronomy rather than chemical and biological analysis. NIR uses the 1.1-5 \(\mu\)m range, which is transparent to water and glass, making it an easy to use and convenient
method. However, the range also contains overtones and harmonics of bond vibrations, leading to weak and poorly defined peaks when compared to those collected in the MIR range.

While NIR is a very useful method, which has and continues to be used successfully within the food industry for decades, MIR can be considered to give more in-depth analysis, despite not attaining the same depth of penetration into samples and suffering more from water absorption than NIR. Generally operating in the 5-20 µm range, MIR is a very precise method that detects the fundamental molecular vibrations and can thus clearly detect the presence of various functional groups. As it excites entire molecules, very slight differences in peak wavenumbers and absorption levels can be measured and indicate relative positions of functional groups, thus creating a unique (bio)chemical or spectral fingerprint region for each compound. Furthermore, FTIR absorption follows the Beer-Lambert law, which states that signal intensity is directly proportional to concentration. These properties enable not only a relatively simple method for analysing a sample and gather a good understanding of its contents, but also the use of relative absorption levels to acquire quantitative results. Biological analysis generally focuses on specific areas: 3050-2800 cm\(^{-1}\), measuring C-H stretching vibrations in lipids; 1750-1500 cm\(^{-1}\), measuring C=O and C-N stretching and N-H bond bending in proteins and amino acids; and 1200-900 cm\(^{-1}\), measuring C-O-C and C-O bonds in sugars (Talari et al. 2017). Figure 1.2 gives an example of these regions.

Figure 1.2: Sample spectrum of bovine milk showing areas of biological importance. 1: lipid C-H bonds, 2: fatty acid C=O bonds, 3: Amide C=O bonds, 4: carbohydrate C-O & C-O-C bonds
One of FTIR’s strongest assets is its ease of use. Unlike chromatographic methods which require lengthy sample preparation and often specialised optimisation, the method requires minimal background training or sample preparation. Additionally, it gives a complete spectral picture of the sample analysed, encapsulating carbohydrates, amino acid, nucleic acid, saccharide molecules, DNA, RNA, and lipids in a high-throughput manner. As such, it has been used as an analytical tool for microbiological and food studies. For example, Nicolau et al. (Nicolaou et al. 2010) used the method to detect bacterial spoilage in milk, while Botelho et al. (Botelho et al. 2015) were able to detect the presence of several adulterants in raw milk including sodium citrate, water, and starch. It has also been used in the discriminate virgin olive oil by geographical origin (Bevilacqua et al. 2012) and honeys by botanical origin (Etzold & Lichtenberg-Kraag 2008). Finally, it has also been used to detect the sweetening with sugars of water-diluted orange juice (David I Ellis et al. 2016).

It must be noted, however, that due to the quantity of information, chemometric analysis is nearly always required to identify subtle differences in spectra. Additionally, FTIR’s sensitivity to water can lead to challenges when analysing some biological material if they have not been dehydrated, as large water peaks will often saturate detection and hide important spectral features. This is generally remedied either through air or oven drying the sample or using attenuated total reflectance (ATR) methods, a variant of FTIR that uses evanescent waves from passing the laser through a crystal to limit the effect of the water peak as depicted in Figure 1.3.

**Raman Spectroscopy**

Raman spectroscopy is a powerful analytical method that is complementary to IR. It is also a form of vibrational spectroscopy that, unlike FTIR, shines light from a monochromatic
laser to induce inelastic scattering of light which it then measures. Since the discovery of this type of scattering by C. V. Raman (Raman 1929), Raman spectroscopy has been developed to become a powerful tool for analytical chemistry.

When light is shined on a sample, it can either be reflected back, absorbed, transmitted, or scattered after interacting with molecular vibrations. This scattering is generally described as electrons absorbing a photon and hitting a virtual state to immediately return to their base state. The vast majority of interactions are elastic, indicating no net change in the photon’s energy during the interaction, and are known as Rayleigh scattering. However, a small proportion of photons ($10^{-6}$-$10^{-8}$) will gain or lose energy during these interactions by inducing a change in the molecules vibrational level, a phenomenon known as Raman scattering, as depicted by Figure 1.4.

Figure 1.4: Simplified Jablonski diagram depicting elastic (Rayleigh) scattering and inelastic (Raman) scattering in both red-shift (Stokes) and blue-shift (anti-Stokes) possibilities. FTIR absorption is also presented as a comparison.

Raman scattering can be divided into two types: Stokes shift and anti-Stokes shift. Stokes shift involves an inelastic collision where the molecules descends back to a higher
vibrational state than it was initially, thus releasing a lower energy photon, represented by a red shift. Conversely, anti-Stokes shift involves an excited electron falling back to a lower state than it occupied initially and releasing a higher energy photon, represented by a blue shift. It is worth noting that, because Stokes are far more frequent than anti-Stokes, they are used in the majority of Raman-based experiments. This difference in frequency is explained by the need for molecules exhibiting anti-Stokes shift to be in an excited state prior to interacting with the incident photon, which the Boltzmann thermal distribution model (Russell 1996) shows is less likely.

Unlike FTIR spectroscopy for which vibrational selection rules are based on the dipole moment created by a bond, Raman selection rules are based on the change in polarisability i.e. the ease with a dipole moment can be induced. In practical terms, the main parameters affecting polarisability are bond length and difference in electronegativity; the former changes the intensity of electronic interactions between the bonding pair and other electrons attached to the nuclei, thus easing movement, while the latter increases forces pulling electrons to one side of the bond. This difference in property enables FTIR and Raman spectroscopy to be complementary methods, with one method showing vibrations related to the chemical structure the other does not.

As water scatters light only weakly and thus does not show overwhelming signals, Raman has been particularly instrumental in the study of biofluids for examples, which can be examined without the need for any drying protocols (Atkins et al. 2017), and indeed, in vivo studies on several organs (Powell et al. 2017; Petersen et al. 2017; Yosef et al. 2017; Mahadevan-Jansen et al. 1998) have been performed with positive results. In the study of food samples, Raman spectroscopy’s ability to gather a full picture with little or no sample preparation has created strong incentives to develop methods to investigate bacterial spoilage (Luo & Lin 2008) and on-line quality assurance. In food adulteration, it has been used to detect the presence of vegetable oils in virgin olive oil (Dong et al. 2012), the presence of the carcinogenic Sudan-I dye in chilli powder (Haughey et al. 2015), along with the presence of artificial sugars in honey (Li et al. 2012).

Similarly to FTIR, Raman spectroscopy is able to provide a large quantity of information in a single spectrum. As such, chemometric methods are generally required to
process and analyse the data in a meaningful manner. However, inelastic scattering only manifests itself in approximately $10^6$-$10^8$ interactions, meaning the Raman effect is generally perceived as being quite weak. As such, a careful balance between achieving a high signal-to-noise ratio and ensuring the sample does not endure photo-degradation must be found. While methods such as surface-enhanced Raman scattering (SERS), tip-enhanced Raman scattering (TERS), and resonance Raman (RR) exist to enhance signals, these require more sample preparation. Finally, as some of the more powerful lasers used are in the UV-Vis range, fluorophores in a sample can be activated, limiting the effectiveness of the method.

**NMR spectroscopy**

Nuclear magnetic resonance (NMR) spectroscopy is a spectroscopic method that uses radio frequency waves to probe the chemical environment of nuclei within molecules. Originally developed in the 1940s and 1950s by the Bloch and Purcell groups at Stanford and Harvard Universities and leading to their sharing of the 1952 Nobel prize in physics, this method has become a powerful tool in analytical chemistry thanks to its sensitivity and the wealth of information it can provide.

NMR spectroscopy relies on the phenomenon known as nuclear magnetic resonance. In the absence of a magnetic field, nuclei will be in random spin-states averaging out to zero. When an external magnetic field is applied, however, nucleic spins will align in a set number of spin states either parallel or anti-parallel to the magnetic field. The number of states depends on a quantum spin number $S$, with each nucleus have $2S+1$ possible spin states. $^1$H, $^{13}$C and $^{31}$P, three commonly probed nuclei, all have a quantum spin number of $\frac{1}{2}$, meaning they have two possible states (depicted as $\alpha$ and $\beta$). As the parallel states are slightly lower in energy than the anti-parallel states, there is a slight population imbalance favouring them. Within this field, radio frequency (RF) pulses are emitted which, when hitting the correct frequency, will lead nuclei in the $\alpha$-state to be promoted to the $\beta$-state and proceed to decay back to the $\alpha$-state. This decay will create a signal that is measured and processed using a Fourier transformation.
Figure 1.5 Depiction of the nuclear magnetic resonance effect for spin \( \frac{1}{2} \) nuclei. (A): Nuclear spins in their natural state, outside a magnetic field, exhibit random spins which average to a total of 0. (B): In the presence of magnetic field \( B_0 \), nuclear spins will align with (\( \alpha \)-state) or against (\( \beta \)-state) the magnetic field, with a very small bias (1/10000) towards the lower energy \( \alpha \)-state. RF pulses can then be used to excite nuclei and bring them to the \( \beta \)-state. (C): The excited nuclei will then naturally decay back and emit detectable signals.

An NMR spectrum will plot the chemical shift \( \delta \) (Equation 1.3), a measure of the resonance frequency \( v_{\text{sample}} \) of a nucleus normalised to a reference frequency \( v_{\text{ref}} \), and the intensity at each chemical shift. For any given nucleus, the operating frequency is dependent on several factors including the nuclear mass, the strength of the external magnetic field, and the chemical environment around it.

\[
\delta = \frac{v_{\text{sample}} - v_{\text{ref}}}{v_{\text{ref}}}
\]

Electron clouds around the nucleus will spin to induce a far smaller magnetic field resisting the external magnetic field, a phenomenon known as shielding. As the strength of the shielding effect depends on the presence of this electron cloud, different chemical environments will affect the shielding and thus the chemical shift. A simple rule to follow is that electropositive or electron donating groups such as alkanes will push electrons closer to the nucleus being examined and have a shielding effect; conversely, electronegative or electron-withdrawing groups such as alcohols or amines will pull electrons away from the nucleus and have a de-shielding effect. Additionally, the alignment of nearby nuclei in different chemical environments will have a small but noticeable effect on the resonant frequency. In ethanol, for example (Figure 1.6), the hydrogen signal for \( C_1 \) would be slightly affected by the spin of \( C_2 \)'s hydrogen atoms. With both hydrogen atoms having the possibility of being in spin state \( \alpha \) or \( \beta \), a hydrogen signal would be split into the three possibilities \( \alpha \alpha \), \( \alpha \beta \), and \( \beta \beta \).
αβ/βα, and ββ, all visible on the spectrum. This is known as J-coupling. It is important to note that, as NMR is a very slow method relative to molecular movement, the spinning depicted below will lead all three hydrogen atoms on C₁ would be averaged to one environment, leading the signal to show one stronger, averaged signal.

With the wealth of chemical information available from only a small part of the full possibilities NMR offers, its popularity within various disciplines is perhaps not surprising. In medical research, in solution NMR is commonly used in toxicological and drug research (Kovacevic et al. 2016; Beckwith-Hall et al. 1998). Additionally, it has been used to monitor the composition of blood plasma and thus successfully detect precursors of coronary heart disease (Deidda et al. 2017), gestational diabetes (White et al. 2016) and nodular thyroid diseases (Wojtowicz et al. 2017), amongst others. A constant effort to expand the scope of NMR in medical research is also being produced by researchers at the University of Alberta through the creation of the Human Metabolome Database (Wishart et al. 2013) (http://www.hmdb.ca/), which has compiled data on several thousand metabolites in urine, serum and cerebrospinal fluid.

NMR has also found use in multiple food applications. While it has been useful in the study of solid food, such as the geographical provenance of beef (Jung et al. 2010), the relative ease of liquid state analysis has led it to become a powerful tool in the analysis of beverages and oils. Using ¹H, ¹³C or ³¹P NMR, it has successfully detected the adulteration of virgin olive oil with refined olive oil (Fragaki et al. 2005), hazelnut oil (Parker et al. 2014) and vegetable oils (Fang et al. 2013), as well as discriminating between various regions of origin.
Similar feats have been achieved with beverages, where coffee was differentiated from bulking materials such as barley, soybeans, or coffee husks (de Moura Ribeiro et al. 2017). Additionally, discrimination between coffee types (Arabica vs. Robusta) (Defernez et al. 2017) and manufacturing methods (Charlton et al. 2002), along with geographical origins of green and oolong tea (Lee et al. 2015; Lee et al. 2010), was possible. Finally, the study of fruit juices with NMR has revealed several possibilities. Some positive results include Soares et al. (Soares et al. 2017), who have characterised thermal degradation of passion fruit juice during sterilisation and Francini et al. (Francini et al. 2017), who have shown the discrepancies in phenolic content and anti-oxidant activity of several unpopular apple varieties. Site-specific natural isotope fractionation nuclear magnetic resonance (SNIF-NMR) is a method developed by G. J. Martin (Martin, Wood, et al. 1996) which uses deuterium NMR to examine the $^{1}H/^{2}D$ ratios of certain sites in a molecule and which has gained considerable traction in food adulteration detection. As plant physiology and metabolism have significant effect on these ratios, the common bio-molecules like sugars and ethanol of different origin can be accurately detected. As such, it has become the gold standard used by the International Office of Wines (OIV) for the chaptalisation of wines (International Office of Wines 2009; Košir et al. 2001) and has been used to authenticate the specific sugars in multiple different fruit juices to differentiate them from the common adulterants such as beet and cane sugar (Martin, Wood, et al. 1996; Gonzalez et al. 1998; Kelly et al. 2003).

Despite the wealth of information that it provides, NMR has several downsides. First and possibly most importantly, the weakness of the decay signal make it an insensitive technique relative to mass spectrometry, meaning trace components can often be missed. Additionally, the higher end NMR machines incur high initial and maintenance costs. Along with the cost of the machine, the supermagnet within generally requires liquid helium to keep it sufficiently cold, and its size limits its use to the laboratory. Finally, specialist training is required not only to use the machine and acquire spectra, but also to interpret them, adding an extra cost.

(del Coco et al. 2012; Petrakis et al. 2008).
High performance liquid chromatography (HPLC)

HPLC is one of several methods in the chromatography family. The aim of these analytical methods is generally not to characterise the constituents of a mixture directly but separation; as such, they are often coupled with other analytical methods, allowing for the identification and quantification of the compounds of interest. Some of these methods include UV-Vis spectroscopy, refractive index detection (RID), and mass spectrometry. While chromatographic methods exist and operate in several different ways, the basic principles remain constant. Separation is achieved by running a sample through a biphasic system consisting of a stationary phase and a mobile phase. These phases will have different properties and will thus attract analytes in varying proportions. As the mobile phase washes over the stationary phase, analytes more attracted to the former will move faster, while those more attracted to the latter will be slowed down, as shown in Figure 1.7. Using correct parameters can therefore induce a complete separation.

![Figure 1.7: General chromatography system. The red molecules are more attracted to the stationary phase than the blue molecules. As such, they are slowed down more and are separated from the blue molecules, leading to a later elution.](image)

HPLC uses a pressurised liquid mobile phase to pass analytes through a column lined with the stationary phase. It is a more modern version of liquid chromatography, which used gravity instead of a high pressure pump to push the mobile phase through the column. As the most common separation parameter used is octanol-water partition coefficient, most HPLC methods can be divided into two main categories: normal phase, where an organic mobile phase is passed across an aqueous stationary phase, and reversed phase, where aqueous mobile phase passes through an organic stationary phase. Other methods include size exclusion chromatography, which uses molecular size as a separation method; ion-exchange, which uses affinity between solute ions and stationary phase charged sites; and hydrophilic interaction
chromatography (HILIC), which uses both hydrophilic stationary and mobile phases to separate based on polarity.

Due to its effectiveness in the separation of organic and biological compounds, reverse phase HPLC has become a common tool in drug analysis and research. It is often used to assess purity in drug development or synthetic research. In these cases, combining it with mass spectrometry is a highly effective way to reach incredibly high sensitivity and acquire an understanding of by-products. It is also viewed as the gold standard for urine analysis in legal and medical contexts where, combined with UV spectrophotometry or mass spectrometry, it can be used to diagnose vitamin D deficiency in children (Nobilis et al. 2007) or illicit substances in athletes or the general population (Kolmonen et al. 2007; Sundström et al. 2013).

In the quest to combat food adulteration, HPLC has shown itself to be an invaluable ally. Thanks to its high sensitivity and separation power, researchers have been able to use the method to detect the tampering of a multitude of diverse products. For example, HPLC-UV and HPLC-MS have been used to detect the presence of soy proteins in meat (Castro et al. 2007) and milk powder (Jablonski et al. 2014). It has also been used to successfully determine the origins of sesame oils (Jeon et al. 2013), coffee beans (Choi et al. 2010) and various red wines (Serrano-Lourido et al. 2012), along with the authenticity of royal jelly (Garcia-Amoedo & De Almeida-Muradian 2007), vanillin (van Nederkassel et al. 2006), orange (Vaclavik et al. 2012) and pomegranate juice (Zhang et al. 2009), and organic eggs (Van Ruth et al. 2011).

As with all analytical methods, HPLC is not without its weaknesses. A major setback is the extra time and resources required for this type of analysis. Along with each experiments generally taking several minutes per run (compared to seconds for most FTIR and Raman experiments) and often requiring more complex and lengthy preparation and optimisation steps. Along with the extra time, HPLC machines incur high running costs: they often use about 2 mLmin$^{-1}$ of solvent which quickly adds up to nearly a litre in an 8 h work day, or 2-3 L in a 24 h period. Another setback is the requirement for a secondary method; while UV is an inexpensive and versatile addition that can generally give good sensitivity, it is limited by the requirement for a chromophore which must either be present or added during sample
preparation. Mass spectrometry, while arguably one of the most powerful methods, is also incredibly costly and generally requires optimisation of its own. Finally, while efforts have been made to create portable HPLC systems, these are still in development and will likely not be widespread for several more years.

**Refractive index detection (RID)**

RID is a common alternative to UV spectroscopy as a secondary detection method for HPLC. It measures the change in refractive index, or the change in angle as light enters the sample, relative to that of the mobile phase at time zero. Due to its sensitivity to change in the mobile phase, it is able to detect the presence of any and all analytes, earning it the sobriquet of “universal detector”. As such, it is most often used for the detection of molecules that would not otherwise readily be detected by UV spectroscopy, such as saccharides. Its use remains as a backup, however, due to its overall low sensitivity (Trani et al. 2017). Multiple factors including temperature, pressure, and small solvent ratio inconsistencies can affect the refractive index of the mobile phase and create large fluctuations in the baseline. As such, highly stable isocratic conditions are required to minimise such fluctuations, along with high analyte concentrations in order to achieve an acceptable signal-to-noise ratio.

**Chemometrics and multivariate analysis**

While many analytical methods have the capabilities to differentiate between incredibly similar compounds and mixtures, the similarities correlate with complex and closely resembling spectra, rendering simple visual quantitative detection near impossible with even a trained human eye. As such, a wide variety of statistical methods are employed to detect minute differences within datasets and create robust models that identify differences in analyte composition. Once these models are created, they can then be used in industry with minimal training to detect adulteration. While there are a plethora of chemometric algorithms commonly used in analytical chemistry, each with strengths and specific roles, they can generally be classified in two different ways: supervised vs. unsupervised, and qualitative vs. quantitative. The question of whether a method is supervised refers to whether any *a priori* knowledge of the samples under analysis is exploited. Supervised methods often provide a higher level correlation as they are, in effect, told about the way the data should be separated. However, as they rely on *a priori* knowledge in the form of metadata for example, its accuracy
is of paramount importance. The question of qualitative vs. quantitative models relates to, like its name implies, whether the output is qualitative or quantitative. Qualitative models are generally used to discriminate between certain criteria like geographical origin or the presence of an adulterant, while quantitative models are generally used to quantify levels of adulterant. Below is a description of the chemometric methods used throughout this thesis.

To remove artefacts and instrument-based random error, several types of pre-processing steps can be applied before chemometric methods are employed. Baseline correction (Eilers 2003), for example, can remove artefacts arising from variation in baseline intensity; it is extensively used in Raman spectroscopy, for example, where signals closer to the laser wavelength are higher intensity than those further away. Normalisation and scaling can also be used to limit the effect of fluctuations in power from the source. Normalisation uses a set signal, often from an internal standard added (Subaihi et al. 2017) at a known concentration or a solvent (Sun & Qin 2011), to provide an “anchor” for the intensity, achieved by dividing all points on a spectrum by that intensity. Scaling can be performed using various methods, some of which include dividing by the standard deviation in the dataset (autoscaling) or the variance (Pareto scaling) (Parsons et al. 2007).

**Principal component analysis (PCA)**

PCA is a qualitative multivariate analysis algorithm which aims at decreasing the dimensionality of a dataset by classifying linear correlations as components, whilst retaining the variance (Hotelling 1933). These components, or principle components (PCs) are then ranked in order of strength and the strongest correlations can be plotted against each other to greatly simplify the dataset. This is generally used as an initial analysis tool as it reliably presents correlation with little risk of false assumptions. However, being unsupervised, other methods can often provide better models.

**Principal Component - Discriminant function analysis (PC-DFA)**

PC-DFA is a supervised method based on PCA. It uses metadata to optimise the PCA model and bring data points of the same class closer to each other while separating those of a different classes (Gromski, Muhamadali, et al. 2015) (minimising within-class variance whilst maximising between-class differences). This is done using matrix calculations and the creation of eigenvalues which distort the principle components. As this is a supervised
algorithm, due care must be taken to use correct metadata. Additionally, validation steps are generally required to ensure a supervised model has been adequately created. As the model is fed the “answers” during its creation, there is a risk that the model will use noise and insignificant data to create a better looking model. This issue, known as overtraining, is problematic as the resulting model will be highly susceptible to batch-to-batch variations and will thus be unusable. Validation steps use new unknown data to test for overtraining, either by cross-validation (excluding some data from a model during creation and using it later to test the validity) or by using blind testing.

**Partial least squares regression (PLSR)**

PLSR is a supervised quantitative method that uses latent variables to create a quantitative predictive model for a dataset (Gromski, Muhamadali, *et al.* 2015). The model is generally constructed by dividing the data into a training-set, used to create the model, and a test-set, used to examine the validity of the model. The correct number of latent variables (LVs) is decided using cross-validation methods and aiming to minimise the root-mean-squared error (RMSE). Convention is to use as few latent variables as possible so as to minimise the chance of overtraining a model.

**Aims and objectives**

The aim of this work was to attempt to accurately detect the adulteration of coconut water with milk, water and sugars using the methods described above, and for the vibrational spectroscopies, used here for the first time for the analysis of coconut water. The presence of bovine milk was detected using FTIR and Raman spectroscopies, and the detection limit was found for FTIR. HPLC-RI was used as a comparative method, using lactose as a marker of bovine milk, and the limit of detection (LOD) was established. Additionally, Raman spectroscopy was used to quantify the main sugars glucose, fructose, and sucrose. It was then used to detect the dilution of coconut water with water, and the masking of this dilution with individual sugars, keeping the total amount of sugar constant, with a mixture of sugars emulating the natural concentrations found in coconut water. Finally, NMR was used as a comparative method to detect the masking of diluted coconut water with a mixture of sugars.
Chapter 2: Detecting the presence of bovine milk as an adulterant in coconut water

Introduction

Coconut water is a refreshing and nutritious drink commonly consumed throughout the world. As with other fruit juices, its premium market value and popularity make it a prime target for nefarious practices such as adulteration. One particularly hazardous practice that has been rumoured is the adulteration of coconut water with milk, so as to give a white colouration to the product (Mendick 2016). This illicit adulteration would hide enzymatic discolouration and give it a more appealing look (Figure 2.1), however it would also endanger those with food allergies and intolerances to dairy products. Evidence of this was seen in 2013, when a child died of anaphylactic shock as a result of drinking coconut drinks that had been adulterated with milk (Barlass 2015). Two further recalls of canned coconut water products having shown signs of contamination with milk (Food Standards Agency 2015b; Food Standards Agency 2015a) in 2015 further demonstrating the benefits of finding accessible and robust methods of detection of milk in these products.

![Image of various milk concentrations in coconut water](image)

Figure 2.1: visual effect of various concentrations of milk in coconut water

Dairy is not only a very common foodstuff in itself, it is also found in a multitude of both sweet and savoury products. Milk, when purchased from a shop, is an aqueous emulsion with dry constituents consisting mainly of proteins, varying concentrations of lipids and the
sugars glucose and lactose. For some, however, dairy products can be the cause of intolerance and allergic reaction. A subset of the population has a deficiency in the enzyme lactase, which splits lactose into its monosaccharide components glucose and galactose. Consumption of lactose containing products can therefore cause indigestion- and IBS-like symptoms including diarrhoea, stomach cramps, and bloating (Anon 2016). While the quantity of lactose required to cause symptoms varies on a case by case basis, it is generally recommended to avoid its consumption. Additionally, some people have an allergy to whey or casein, two proteins present in dairy (Anon 2014). The difference is important to highlight: as intolerance is an inability to efficiently metabolise a certain molecule, there is a minimum quantity required for symptoms to occur; allergies, on the other hand, are auto-immune reactions to a molecule and, in the most extreme cases, life-threatening reactions can be triggered by trace amounts.

Milk has several components that are foreign to coconut water and that can be used in detecting possible adulteration. The relatively high concentration of proteins and lipids are important differences, as is the presence of lactose. As such, there are several options as to what methods can be used for its detection. Infrared (FTIR) and Raman spectroscopy are high-throughput, simple, and robust methods of detection that can give a complete fingerprint of a sample with little or no sample preparation required. Furthermore, advances in portable versions of these technologies make them ideal for implementation in a practical setting. These methods were therefore tested in parallel to determine the better method and subsequently the limits of detection.

HPLC-RI was also used as a comparative method of detection. Unlike FTIR and Raman spectroscopy, two techniques that produce a so-called (bio)chemical or spectral fingerprint of the entire content of a particular sample, HPLC is a separation technique. Considering the aim of the study was to find the method with the best limit of detection, separating the metabolite being studied was thought to be a useful way to simplify detection. As the presence of lactose is be indicative of the presence of animal milk, it was chosen as the marker to be studied. Despite refractive index detection being known to be less sensitive (Trani et al. 2017) than UV-spectroscopy, the latter would have required an extra step to add a UV-active chromophore to lactose and would have incorporated issues of extra uncertainty and yield.
Materials and Methods

Coconut water and Stock preparation

Seven young coconuts of Costa Rican origin were purchased from a UK based online retailer. The juice was then extracted using a specialised Cocodrill® purchased from the same retailer, centrifuged (18,000g, 10 min, 4°C) and pooled together to make a consistent stock solution. The pooled stock solution was then stored at -80 °C in 45 mL aliquots until needed. Prior to use, the coconut water was thawed and separated into 1 mL aliquots, which were heat-treated using a TechneDri-Block DB-3A (Cole-Parmer, Stone, Staffordshire, UK) hot-plate for 150 s at 70 °C to simulate a pasteurisation process, recombined and rapidly cooled back down in a refrigerator at 5 °C. Heat-treated aliquots were used for a maximum of three days after being thawed. Whole cow’s milk (3.5%) was purchased from a local retailer. For the creation of sugar solutions, D-glucose anhydrous, D-sucrose (for biochemistry 99%, RNAse and DNase free), D(+)-lactose monohydrate biochemica, and D(-)-fructose 99%≤ were purchased from Fisher Scientific (Fisher Scientific Ltd. Loughborough, Leics, UK), Acros Organics (Acros Organics, Geel, Belgium), PanReac AppliChem (AppliChem GMBH, Darmstadt, Germany) and Sigma-Aldrich (Sigma-Aldrich Chemie GMBH, Stanheim, Germany) respectively.

FT-IR

Stock solutions at different dilutions (1-10%, 1% increments) of milk in coconut water were generated keeping the total volume at 1 mL. Aliquots of each stock solution were then diluted by a factor of 10 and 100 again keeping the final volume at 1mL to make analytical samples, and each sample was diluted with 500 µL water. All the samples were then vortexed for 10 s and spotted five times on a 96 well FT-IR silicon plate (sample volume: 15 µL) in a randomised order. The plate was then oven dried at 50 °C for 150 min to remove the maximum amount of water. FTIR data were collected in absorbance mode in the mid-IR region (4000-600 cm⁻¹) with 4 cm⁻¹ resolution using a Bruker Equinox 55 FTIR spectrometer (Bruker UK Limited, Coventry, UK) and OPUS 6.5 (Bruker) software.

Raman

A 10% (v/v) solution of milk in coconut water solution was made by adding 900 µL of coconut water to 100 µL of milk. An aliquot (100 µL) was taken and diluted to 1 mL with
coconut water to make a 1% (v/v) solution. This was repeated with the 1% solution, and a serial dilution down to 10^{-5}\% (v/v) was achieved. Aliquots (360 µL) of each sample were transferred into a 96 well quartz plate.

Raman spectra were collected using a Renishaw InVia Raman Spectrometer (Renishaw Plc. Old Town, Wotton-under-Edge, Gloucestershire, UK) equipped with a laser wavelength of 785 nm and an Olympus MD Plan 10 x10 long distance objective (KeyMed (Medical & Industrial Equipment) Ltd. Southend-on-Sea, Essex, UK) using a laser power of ~30 mW at the sample. A 600 l/mm grating was used and the acquisition was centred at 1500 cm^{-1}, leading to a spectral range of 286.6-2496 cm^{-1} (1015 points). The optimal depth for maximum intensity was found using a depth series acquisition, and each spectrum was analysed at optimal depth and 1, 2, and 3 µm closer using an automated system set up in-house. Six 10 s acquisitions were combined during collection for each spectrum.

**HPLC-RI**

Single sugar solutions were generated by dissolving glucose (3.15 g, 62.8 mgmL^{-1}), fructose (3.17 g, 63.4 mgmL^{-1}), sucrose (3.17 g, 63.4 mgmL^{-1}), and lactose (3.19, 63.7 mgmL^{-1}) into 50 mL aliquots of deionised water. A mixed sugar solution was prepared by combining 500 µL of each single sugar solution together and vortexing the mixture for 15 s. Analytical samples were generated by adding 100 µL of deionised water and 800 µL acetonitrile (Sigma Aldrich) to 100 µL aliquots of each stock solution.

A 20% (v/v) milk in coconut water solution was made by adding 800 µL of coconut water to 200 µL of milk. This solution was then serially diluted twice by dissolving 100 µL of solution to 1000 µL with coconut water, obtaining a 2% and a 0.2% (v/v) milk in coconut water solution. Analytical samples were generated by adding 100 µL deionised water and 800 µL acetonitrile to 100 µL aliquots of these solutions, coconut water, and milk.

Data were collected using an Agilent 1260 HPLC system and refractive index (RI) detector (Agilent Technologies LDA UK Ltd, Stockport, Cheshire, UK) and a Phenomenex Luna 5 µm Amine column (Phenomenex Inc. Macclesfield, Cheshire, UK) with dimensions 250 x 4.6 mm. The pre-mixed mobile phase consisted of 80:20 ratio of acetonitrile (HPLC grade, Fisher Scientific) to deionised water with 0.1% trifluoroacetic acid (Sigma Aldrich) added, and was ran at 1.5 mLmin^{-1}. The column and detector temperatures were set to 35°C.
and each experiment was operated for 7 min. To avoid cross-contamination, two needle washes were performed between each experimental run using an 80:20 mixture of acetonitrile and water (75 µL). No processing was performed on these data.

Analysis

FT-IR

All data analysis was carried out in Matlab R2017a (The MathWorks, Natick, MA, USA). Spectra were first separated into a 0-0.1% and a 0-1% adulteration dataset. They were subsequently scaled using extended multiplicative scattering correction (EMSC) (Martens & Stark 1991) and the CO₂ region at 2401-2270 cm⁻¹ replaced by a linear trend. Principal component-discriminant function analysis (PC-DFA) (Gromski, Muhamadali, et al. 2015) was used to obtain information about the areas of maximum variance, along with 95% confidence values. Additionally, partial least squares regression (PLSR) (Gromski, Muhamadali, et al. 2015) was used to generate a linear predictive model. This was undertaken by dividing each dataset into training (0%, 20%,…, 100%) and test sets (10%, 30%,…, 90%). The training set was used to create several models, each with an increasing number of latent variables (LVs) used and a k-fold cross-validation (k=20) system (Gromski, Muhamadali, et al. 2015) was used to test each model. The optimal model was chosen by aiming to minimise the root mean squared error on the cross-validation set (RMSECV) and the number of LVs, and the test set was then fed into it.

Raman:

Spectra were first subjected to manual cosmic ray removal in the WiRe Raman software (Renishaw), then subsequent processing and analysis was performed in Matlab R2017a. Baseline correction was performed using a partial least squares algorithm (based on an algorithm by Eilers (Eilers 2003) and created as part of an in-house analysis packet) using a smoothing parameter of 100,000 and an asymmetry parameter of 0.001. The 100% and 10% milk spectra were then removed due to insufficient signal strength compared to other spectra, as scaling would have drastically increased their noise intensity. The remaining data were auto-scaled. PC-DFA was used to analyse the processed data as was the case for the FTIR spectra.
Results and Discussion

Raman spectroscopy

Raman spectroscopy is a method that measures the inelastic scattering of light in a sample. As such, measurements can be affected by the presence of highly scattering components in a solution. The fat globules encased in casein micelles (Mather & Keenan 1975) that are present in milk are such scattering components, and their effects on the Raman signal is clear, as is shown in Figure 2.1, the samples containing a large proportion of milk have severely dampened signals compared to those with less adulteration. As the lower signal strength would likely create artefacts during scaling, these spectra were removed prior to further analysis.

![Baseline corrected average Raman spectra of each serial dilution tested. Percentage values represent the proportion of milk adulterated into coconut water.](image)

Once scaled, it becomes clear that differences between spectra are incredibly small. While this is not particularly surprising given the very small levels of adulteration being studied, it is clear that chemometric methods will be required to distinguish differences and find any concentration related trend.
Figure 2.3: Average Raman spectra after removal of $1E^5$ ppm and pure milk samples and autoscaling

PC-DFA is a commonly used and powerful chemometric method in the pursuit of qualitative analysis. It is a supervised enhancement of PCA, which uses the principal components created in an unsupervised environment and manipulates them mathematically in such a way as to minimise the variance within pre-determined classes while maximising inter-class variance. In this case, the aim was to understand the variance between levels of adulteration, and identify the spectral regions directly related to this adulteration.

Figure 2.4: PCA scores plot of coconut water samples adultered with various quantities ($1E^{-1}$-$1E^4$ ppm) of milk. The top 10 PCs were used, accounting for 54.8% of the total explained variance.
Since it is plainly visible in Figure 2.4 that the separation in PCA is insufficient to provide any useful qualitative information, the application of DFA was then performed (Figure 2.5). Immediately, we see that the addition of class data greatly improves the clustering in these data. Additionally, the presence of a trend which could be related to adulteration is hinted at in DF 2. Figure 2.5B showing DF 2 vs. DF 3 loosely separates the classes into three main groups. The coconut water control, 1E⁻¹ ppm and 1E⁰ ppm adulteration classes (black, red and blue respectively) are completely indistinguishable from each other, but show a complete separation from the 1E⁴ ppm adulteration class (brown). The 1E¹, 1E² and 1E³ ppm adulteration classes (green, blue and purple respectively) are then within this gap, with the 1E¹ ppm level of adulteration being only partly separated from the control. It is worth noting, however, that the model exhibits the 1E² ppm adulteration class as more adulterated than the 1E³ ppm adulteration class, an analysis that is obviously incorrect. Also, the low number of repetitions and lack of robust validation means the confidence values should be regarded as advisory rather than absolute.
While the DFA scores plot (Figure 2.6) has indicated the presence of a concentration-related variance, one cannot find the source of the variance without studying the loadings. The first striking feature of the loadings for DF 2 is the very low signal-to-noise ratio. The quantity of noise present severely limits our ability to determine the sources of variance between classes, and as such, the study using Raman spectroscopy was not pursued further.
An initial comparison between Raman and FTIR spectra demonstrate the stark contrast between these two methods. As is shown in Figure 2.7, milk results in a spectrum of comparable intensity to coconut water, allowing us to immediately see the large differences in constituents. Based on the peaks at 2924 cm$^{-1}$ and 2854 cm$^{-1}$ signifying the presence of triglyceride C-H$_2$ vibrations and the peak at 1747 cm$^{-1}$ signifying the fatty acid C=O ester vibrations, we can confirm the far higher fat content of milk. Additionally, the amide I band (1655 cm$^{-1}$) and amide II (1545 cm$^{-1}$) bands are indicative of protein content in milk. Based on this information alone, we can infer that FTIR will be a more effective method of detection; a comparable serial dilution was used to evaluate this inference.
Once again, PC-DFA was used to compare the extent of separation at different levels of dilutions, along with the primary sources of variance. As the PCA scores plots (not shown) in themselves add little to aid clarity, concentration data was immediately fed into the algorithm to obtain the PC-DFA scores (Figure 2.8). These indicate that both Raman and FTIR have similar levels of discrimination. For both, DF 2 shows good separation between the coconut water control sample and the $1E^4$ ppm milk samples. Additionally, it shows incomplete separation at $1E^3$ ppm and below. To aid the choice of method, the operating parameters provide interesting insight on the theoretical reproducibility of the trend. It is clear that FTIR shows a far higher signal-to-noise ratio; where it required only 5 principal components to reach $>99\%$ total explained variance, Raman used 10 to only reach 55%. As such, the 95\% confidence values showed be viewed with some scepticism. While they provide an idea of the reproducibility of the trends and the relative strength of each method, there is insufficient validation to ensure that no overtraining is happening.
Despite DFA scores for FTIR and Raman spectroscopy being comparable, examining the relevant loadings for each method shows a very different picture. While Figure 2.6 appears to be predominantly comprised of noise, Figure 2.9 gives a far clearer picture in describing the potential sources of variance. The negative peaks at 2926, 2855, 1746 and 1653 cm\textsuperscript{-1} confirm that the lipid and protein content are useful markers for the detection of milk. Along with this, the positive peaks in the 1100-900 cm\textsuperscript{-1} region, assigned to carbohydrate C-O and C-O-C vibrations, are likely due the fructose and sucrose present in coconut water and absent in milk. Overall, while both methods showed similar abilities to detect the presence of adulterated milk in coconut water, the spectral data were found to be far more accessible using IR. The clear loadings, low number of principal components required and comparable signal strength for coconut water and milk made it clear that, given the experimental parameters, FTIR would likely be the more consistent and reproducible method.
Quantification of cow’s milk in coconut water at 0-1%

Figure 2.10: Average FTIR spectra linear range from 0% (black) to 1% (red), after EMSC scaling and CO₂ effect correction. Above: full spectral range sampled. Below, left to right: enlarged relevant areas at ~2900, ~1600, and ~1100 cm⁻¹.

While previous experiments showed major differences between the spectra of pure milk and coconut water, adulteration levels below 1% (1E⁴ ppm) render the peaks unique to milk invisible to the eye (Figure 2.10). However, zooming in on certain key features shown in the 2950-2800 cm⁻¹ lipid area and the ~1100 cm⁻¹ carbohydrate area still show clear trends consistent with the loadings in Figure 2.9. Additionally, zooming highlights a significant drawback to the use of IR. Due to FTIR’s sensitivity to the presence of water, a drying procedure is required to remove as much of it as possible. However, the high sugar concentration of coconut water renders the evaporation of all the water nearly impossible, often leading to an inhomogeneous sample thickness and unpredictably high signals. While the addition of extra water and the scaling algorithm are used to mitigate this issue, it is not a perfect solution, as shown by the 0% milk spectrum being unexpectedly strong.
To create a predictive model and obtain information on the limit of detection, partial least squares regression (PLSR) was used. Using a training set consisting of even concentrations (0%, 0.2%, 0.4%,…, 1%), models were created using different numbers of latent variables (LVs), and their validity was checked using a leave-one-out cross validation model (Figure 2.11A). The model with the lowest error prediction value was chosen, aiming for the smallest number of latent variables so as to avoid overtraining. In this case, the model containing 11 latent variables contained the lowest RMSECV, indicating that it resulted in the lowest error when data points from the training set are blindly and randomly fed into it. The test set, consisting of the rest of the concentrations, was then fed into the model, as seen in Figure 2.11B. While the system shows some variance, there is a clear linear relationship between the concentrations predicted by the model and the experimental concentrations.
The statistical parameters for the PLSR model (Table 1) indicate a good fit. To examine the goodness of the fit presented in the PLSR model used, $R^2$ and $Q^2$ values were examined. $R^2$ and $Q^2$ are values between 0 and 1 which represent the proportion of variance explained by the model, where 1 represents a perfect fit. They are calculated by comparing sum of the squared deviations from the model to the sum of the squared deviations from the mean. For PLSR, $R^2$ relates to the training sets only, while $Q^2$ (CV) and $Q^2$ (tst) relate to the cross-validation and test sets respectively. The high $R^2$ and $Q^2$ values show that the model is robust, and the slightly lower $Q^2$ value for the cross-validation than the test set shows that the model has not been overtrained. It is interesting to note that the high number of LVs used indicates the possibility of non-linearity that is not visible in the model. This is likely due to reaching a level that is below the limit of detection. To confirm that the limit of detection is not at a lower order of magnitude, this experiment was repeated with a lower range.

**Table 2.2: Statistical results for PLSR model in Figure 2.11.** Trn, CV and tst represent values for the training, k-fold cross-validation (k=20) and test sets respectively.

<table>
<thead>
<tr>
<th>LVs used</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.9437</td>
</tr>
<tr>
<td>RMSE(trn)</td>
<td>0.0810</td>
</tr>
<tr>
<td>$Q^2$ (CV)</td>
<td>0.7641</td>
</tr>
<tr>
<td>RMSE(CV)</td>
<td>0.1086</td>
</tr>
<tr>
<td>$Q^2$ (tst)</td>
<td>0.8908</td>
</tr>
<tr>
<td>RMSE(tst)</td>
<td>0.0935</td>
</tr>
</tbody>
</table>
Quantification of cow’s milk in coconut water at 0-1%

It comes as no surprise that, once again, it is impossible to predict the level of adulteration by eye. Additionally, some spectra once again show abnormally high absorbance, indicating uneven sample distribution across the spot on the silicon sampling plate. However, zooming in the same areas does not show the same trends, with the 2950-2800 cm$^{-1}$ lipid area showing virtually all spectra overlapping each other for example. This is fairly consistent with the previous PC-DFA model, which produced very little separation below 0.1%. PLSR was therefore used to obtain more information on the trend.

Figure 2.12: Average FTIR spectra linear range from 0% (black) to 0.1% (pink), after EMSC scaling and CO$_2$ effect correction. Above: full spectral range sampled. Below, left to right: enlarged relevant areas at ~2900, ~1600, and ~1100 cm$^{-1}$.
Figure 2.13: PLSR model predicting the level of adulteration of milk in a linear 0-0.1% series. A - plot displaying the RMSECV for each PLS model based on the number of latent variables. B - PLSR model using 2 LVs, as it minimised both RMSECV and number of LVs.

It is immediately clear that FTIR is unable to detect any sort of trend below 0.1% concentration. Not only does the variance of the model predictions encompass a large range, there is no fit between the predictions and the experimental values (Figure 2.13B).

Overall, while the exact limit of detection cannot be explained from these current data, it can be estimated to be around 0.2-0.3% based on the linearity of the PLSR model in Figure 2.11B. Based on Figure 2.1, those groups planning to use milk as a whitening agent would likely add approximately 0.5-1%. More than 1% would likely be too white and arouse suspicion, while less than 0.5% would likely be of insufficient effect to have any practicable use. As such, FTIR spectroscopy would be able to reliably detect intentional adulteration of coconut water with milk. Additionally, it is sufficiently sensitive to detect the minimum quantities of milk required to induce intolerance responses in the majority of those with milk protein or lactose intolerance (hypolactasia). It must be emphasised, however, that FTIR spectroscopy is not a suitable method for the detection of milk contamination from negligent practice. These quantities are usually at trace level and would thus be far below the detection limit of FTIR spectroscopy. Similarly, people with milk allergies would still be susceptible, as the auto-immune reactions require far lower quantities to induce these immunological responses.
HPLC analysis

To facilitate optimisation, a model using solutions of glucose, fructose, sucrose, and lactose was used. An initial challenge faced was the low solubility of these sugars in the 80% acetonitrile mobile phase. The highest total sugar concentration was found to be 5-7 mgmL\(^{-1}\) at room temperature, requiring an overall dilution factor of 10 for coconut water. However, due to the low sensitivity of refractive index detection, a large injection volume (50 µL) was required.

![Chromatogram for mixed sugar solution](image1)

*Figure 2.15: Overlaid chromatograms for individual sugars*

Initial tests using solutions of a single sugar showed complete separation between lactose and other sugars, as shown in Figure 2.15. When overlaid, the chromatograms showed a difference of 0.3 min between the end of the detectable elution of sucrose and the beginning of the detectable elution of lactose. This is not the case of all sugars: fructose, glucose, and sucrose, for example, are not fully separated. However, as these were not components of interest, this was not considered an issue. These results were confirmed through the use a mixed solution of all four sugars, once again showing complete separation between lactose and other sugars, as shown in Figure 2.16.

![Chromatogram for mixed sugar solution](image2)

*Figure 2.15: Chromatogram for mixed sugar solution*
Outside the elution times, two more factors worthy of consideration are present. The first of these is the presence of a signal at \( \sim 2.5 \) min. This signal was consistently present in samples and in needle washes, with varying intensity and was considered to be an unavoidable artefact originating from the injection. The relatively large injection volume induced a temporary increase in pressure, which was detected when it reached the detector. While it can be treated as nothing more than a baseline fluctuation during analysis, users should be aware of this phenomenon and not mistake it for a real signal. The second factor to be considered is the presence of multiple peaks for glucose, fructose and lactose. Multiple peaks are the result of their respective \( \alpha \)- and \( \beta \)- isomeric forms, and equilibrium ratios depend on several factors including temperature and pH. If need be, more stringent pH control could therefore be applied to force the presence of only one isomer and decrease the peak width.

As the addition of acetonitrile to milk caused partial coalescence of the fat globules and their precipitation to the bottom, samples were centrifuged to remove these. The concentration of other components in milk and coconut water were considered to be insufficiently concentrated to be detectable using RI; samples were therefore approximated to being mixed sugar solutions. While adequate separation was found to still be present, the low sensitivity of RI is immediately visible. Below 20\% v/v, the lactose peak is indistinguishable from the baseline seen in the coconut water control sample. As explained before, due to the already high injection volume required to achieve a clear signal, increasing it further is not a
meaningful option. As a result, it can be said that compared to FTIR, RI lacks at least two orders of magnitude of sensitivity to achieve any meaningful detection of adulteration.

**Conclusion:**

Coconut water is refreshing and nutritious beverage with a booming popularity around the world. However, it is also susceptible to adulteration, and coconut water containing products have been recalled due to the unmentioned presence of bovine milk, a contaminant that leads to both allergic and intolerance reactions in a significant subset of the population. The ability for FT-IR and Raman spectroscopy, two common vibrational spectroscopic methods, to detect this was tested, and HPLC-RI was used to detect the presence of lactose, a defining component of bovine milk foreign to coconut water, and serve as a comparative method.

While both Raman and FT-IR demonstrated comparable detection limits using a factor of 10 serial dilution sample set, the far superior signal-to-noise ratio produced by FT-IR led to be chosen for quantitative detection. Using PLSR on a linear concentration range of 0-1% in 0.1% increments led to a good linear model with an average error prediction of 0.09% on the test set; however, a concentration range of 0-0.1% was found to be too low for quantitative detection. As such, FT-IR would be able to detect wilful adulteration of coconut water with milk, but lacks the sensitivity to detect accidental contamination at trace level. HPLC-RI, while able to successfully separate lactose from other sugars present in coconut water, was unable to concentrations of milk below 10%. It is therefore an inadequate method.
Chapter 3: Detecting the dilution of coconut water and subsequent masking with sugars using Raman spectroscopy

Introduction:

Already consisting of ~95% water, the most obvious method for the adulteration (stretching) of coconut water is dilution. Water is an inexpensive and readily available commodity that can easily increase the profit margin of a coconut water (or indeed any beverage) product. Additionally, the high proportion of water already present has the potential to render detection of dilution more difficult. However, higher levels of dilution will inevitably lead to noticeable changes in organoleptic characteristics, such as less taste and decreased sweetness, while the lower proportions of solids becomes detectable with simple methods. This has been overcome by adulterers in other fruit juices using inexpensive cane and beet sugars as masking agents, which can not only mask the dilution but also readjust sugar content to mimic its normal quantity. The addition of a sugar, such as fructose for example, can also be used by manufacturers obtaining coconut water from multiple sources as a method to normalise batch-to-batch variation arising from different species, maturities, or origin. While there is nothing intrinsically unethical or unsafe about this common practice, failure to report it on the product label renders it fraudulent. A very recent investigation accused several companies of this practice (Glotz 2017).

The current gold standard for the detection of dilution and adulteration with sugar is stable isotopic analysis, generally performed with stable natural isotope fractionation nuclear magnetic resonance (SNIF-NMR) spectroscopy (Jamin et al. 2003) and isotopic ratio mass spectrometry (IRMS) (Antolovich et al. 2001). Both of these methods detect slight but significant differences in the isotopic ratios of common organic atoms such as $^{12}\text{C}/^{13}\text{C}$, $^{15}\text{N}/^{14}\text{N}$ and $^1\text{H}/^2\text{D}$ which arise from slight differences in metabolism and origin. As a result of this, adulterant sugar from multiple sources and water can be accurately discriminated from original constituents. While effective, both of these methods require bulky and expensive equipment and time-consuming procedures. By contrast, Raman spectroscopy is an inexpensive and faster alternative that can detect subtle differences in the contents of samples rather than the isotopic differences.
The suitability of Raman spectroscopy for the detection of water and sugar adulteration was examined. The standard addition method (SAM) was first used to quantify the natural concentration of glucose, fructose, and sucrose in the coconut water stock solution. Then, the detection of adulteration with water, individual sugar solutions while keeping the total sugar content stable, and a mixed sugar solution emulating the concentrations in coconut water were tested. Additionally, the use of succinonitrile as an internal standard to minimise unwanted random error arising from laser fluctuations was examined.
Coconut water, an increasingly popular tropical drink: detection of its adulteration with sugars via Raman spectroscopy and chemometrics

Paul I. C. Richardson, Howbeer Muhamadali, David I. Ellis* and Royston Goodacre*

Manchester Institute of Biotechnology, School of Chemistry, Manchester, UK, M1 7DN

*Email: Roy.Goodacre@manchester.ac.uk  D.Ellis@manchester.ac.uk

Keywords: coconut water, food fraud, adulteration, sugars, dilution, authenticity, integrity, Raman, chemometrics
Abstract

The tropical coconut palm (*Cocos nucifera* L.) has played an important role in the mobility of humans across different geographical regions of the globe as a source of food, water, fuel and building materials. It is the perceived and promoted health benefits of coconut products, and more recently packaged coconut water in particular, that could be said to be responsible for its rapid increase in popularity and sales in the last five years. This huge increase in demand (as with other similar products labelled as pure, fresh, natural), and the short time frame in which the product is determined as optimum quality, can lead to over demand and undersupply within supply chains. This then has the potential for dysfunctionality and illicit practises, such as mislabelling, adulteration, and other forms of food fraud. Here, for the first time in fresh coconut water, we have developed Raman spectroscopy in combination with chemometrics, to explore this approach for the detection of adulteration by dilution of coconut water. Detecting the masking of dilution with individual sugars solutions and mixtures of sugars added to fresh coconut water, or direct adulteration with these sugars without prior dilution. Adulterants were added in 5% increments, from 0-100%, and a total of 103 samples and 412 Raman spectra were analysed. Partial least squares regression results of Raman spectra from both the unmasked water dilution and single sugar solution adulteration datasets demonstrated very high concentration based quantification, with $R^2$ and $Q^2$ values for cross-validation and test sets above >0.999, indicative of a robust linear model, and root-mean-squared error (RMSE) values below 2%, showing that Raman spectroscopy could accurately detect this form of adulteration to within 5%. Whilst results from the sugar mixtures adulteration dataset, despite still showing evidence of a robust model with an $R^2$ value of >0.99 and $Q^2$ values of >0.9, demonstrates a weaker predictive ability with a RMSE value on the test set of 5.2%. From these results it can be concluded that Raman spectroscopy has significant potential as an accurate analytical method for detecting adulteration/stretching, with the ability to discern small abnormalities in sugar ratios in coconut water.

Introduction

The tropical coconut palm (*Cocos nucifera* L.) has played an important role in the mobility of humans across different geographical regions of the globe as a source of food, water, and multiple other uses (Loiola *et al.* 2016), said to be unparalleled in the plant kingdom (Gunn *et al.* 2011). This success is in part due to the fruit of this tree, the coconut,
being highly mobile and able to spread by floating. The dissemination, evolution, genetic diversity and classification of coconut species continue to be studied in detail (Gunn et al. 2011; Loiola et al. 2016) and the chemical composition, properties and potential uses of the highly versatile fruit are also the subject of investigation (Prades, Dornier, et al. 2012b; Prades, Assa, et al. 2012; Yong et al. 2009). Whilst all parts of this plant have been put to an especially wide variety of uses as a source of food (Manivannan et al. 2018), drink, and multiple other uses for millennia, it is the liquid endosperm of the coconut, commonly referred to as coconut water, which is the subject of our study here. Fresh coconut water is a refreshing and nutritious drink typically extracted from immature coconuts of 6-9 months of age (Rolle 2007). This is due to the fact that as coconuts mature their composition and physicochemical properties alter (Tan et al. 2014), with the white kernel (also referred to as meat/flesh) lining the inner shell becoming opaque and hardening, while the coconut water in the centre decreases in volume and sugar content (Child & Nathanael 1950), undergoes alterations in sugar ratios, as well as increases in pH, turbidity and mineral content with a resultant loss of taste and quality (Tan et al. 2014; Vigliar et al. 2006). The nutrient content of coconut water has also been central to several studies (Santoso et al. 1996; USDA 2016) (Table 1) as have its hydration properties.

Table 3.1. Typical water, macronutrient, and sugar contents of fresh coconut water harvested at 6 months from the USDA National Nutrient Database (USDA 2016) and Santosa et al. (Santoso et al. 1996).

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>(g/100g) (6 months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>94.18</td>
</tr>
<tr>
<td>Total dry weight</td>
<td>5.82</td>
</tr>
<tr>
<td>Protein</td>
<td>0.12</td>
</tr>
<tr>
<td>Lipid</td>
<td>0.07</td>
</tr>
<tr>
<td>Ash</td>
<td>0.87</td>
</tr>
<tr>
<td>Dietary fiber</td>
<td>1.1</td>
</tr>
<tr>
<td>Sucrose</td>
<td>0.06</td>
</tr>
<tr>
<td>Glucose</td>
<td>2.61</td>
</tr>
<tr>
<td>Fructose</td>
<td>2.55</td>
</tr>
</tbody>
</table>
However, it is the perceived and promoted health benefits of coconut products (DebMandal & Mandal 2011) and packaged coconut water in particular, that could be said to be responsible for its huge increase in popularity and sales in recent years (Kaplan 2017). These health claims are wide-ranging from its effectiveness (Kalman et al. 2012), or not (Peart et al. 2017), as a sports energy drink, its vitamin B and C content (USDA 2016), potential in treatment of hypertension (Alleyne et al. 2005; Bhagya et al. 2012) and high cholesterol (Sandhya & Rajamohan 2008), through to multiple claimed medical properties including hypoglycemic and antioxidant effects (Preetha et al. 2012; Bispo et al. 2017), as well as antimicrobial (Mandal et al. 2009), antiviral, antiparasitic, antidermatophytic, hepatoprotective and immunostimulant properties (DebMandal & Mandal 2011). The relatively recent explosion in popularity of coconut water as a packaged drink in the last five years is well evidenced, with the global market for coconut water reaching $2.2 billion in 2016, up from $533 million in 2011 (Kaplan 2017). In the UK for example, 40 different brands are currently available, and it has increased in value as a product by 20 times since 2012 with consumption levels of 0.21 litres per capita, three times that of the USA (Glotz 2016b).

Due to our previous work in the area of food (Ellis et al. 2012b; Ellis et al. 2005; Goodacre et al. 1992) and drink (Nicolaou et al. 2010; David I Ellis et al. 2016; Ellis et al. 2017) authenticity and integrity (David I. Ellis et al. 2016; Ellis & Goodacre 2016), it was this relatively sudden and increased visibility on supermarket shelves which prompted us to consider this product as vulnerable to, and a potential target of, adulteration in late 2016. Our work with other fruit juices for example (David I Ellis et al. 2016) had showed that products such as these have known vulnerabilities to a variety of opportunistic behaviours and illicit practises, such as dilution with water, masking of this dilution with the addition of sugars, or straightforward sugar adulteration in a product whose label may state no-added sugars etc. Moreover, when the huge demand for a product only normally available from immature coconuts of a specific age range and sources (five countries worldwide (Glotz 2017)) has the potential to outstrip supply, then this can also lead to these supply networks becoming significantly more vulnerable to fraud (Bowman et al. 2012; Bowman et al. 2013; Ellis et al. 2015).
Several methods have been used to analyse coconut water for various purposes, these include nuclear magnetic resonance (NMR) spectroscopy with chemometrics to monitor process quality parameters, including glucose and sucrose levels (Sucupira et al. 2017), characterisation of the volatile profiles (aroma) from multiple varieties using headspace solid phase microextraction gas chromatography (HS-SPME-GC) (Prades, Assa, et al. 2012), simultaneous analysis of different classes of phytohormones with high performance liquid chromatography (HPLC) and liquid chromatography mass spectrometry (Ma et al. 2008). And monitoring the physicochemical degradation of coconut water (including sugars) using an electrospray ionisation (ESI) source combined with Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) (Costa et al. 2015). Here, for the first time, Raman spectroscopy in combination with chemometrics has been used for the analysis of coconut water to establish the feasibility of a vibrational spectroscopic approach in detecting intentional adulteration (Rodriguez-Saona et al. 2016; Everstine et al. 2013; Rodriguez-Saona & Allendorf 2011; Hu et al. 2015; Black et al. 2016). Adulteration was undertaken with deionised water (dilution) as well as the addition of incremental volumes of sucrose, fructose, and glucose and mixtures of these sugars. These sugars could be used illicitly in order to mask dilution, or added to affect the flavour of coconut water products labelled as fresh, natural, pure, no-added sugar etc. (Glotz 2017), when they are nothing of the sort, an issue observed in other fruit juices previously (David I Ellis et al. 2016; Rodriguez-Saona et al. 2001).

Materials and methods

Stock solution

Seven young coconuts of Costa Rican origin were purchased from a UK-based online retailer. The juice was then extracted using a specialised Cocodrill® purchased from the same retailer, centrifuged at 18,000g, for 10 min at 4°C and pooled together to create a consistent stock solution. This pooled stock solution was then stored at -80 °C in 45 mL aliquots until required. Prior to use, the coconut water from the stock solution was thawed at room temperature and divided into 1 mL aliquots, which were then heat-treated using a TechneDri-Block DB-3A (Cole-Parmer, Stone, Staffordshire, UK) hot-plate for 150 s at 70 °C, to emulate pasteurisation, recombined and cooled in a refrigerator at 5 °C. Heat-treated aliquots were
used for a maximum of three days after being thawed. D-glucose anhydrous, D-sucrose (for biochemistry 99%, RNAse and DNAses free) and D-(−)-fructose 99%≤ were purchased from Fisher Scientific (Fisher Scientific Ltd. Loughborough, Leics, UK), Acros Organics (Acros Organics, Geel, Belgium) and Sigma-Aldrich (Sigma-Aldrich Chemie GMBH, Stanheim, Germany) respectively.

**Standard addition**

Solutions of sucrose (5.984 g), glucose (5.990 g) and fructose (6.010 g) were each dissolved in 50 mL water (120 mg mL⁻¹). Samples were made up by adding increasing quantities (25,…, 250 µL) to 250 µL of coconut water, and filling each mixture up to 500 µL with deionised water, and unadulterated coconut water was used as a negative control. A total of 31 samples were analysed using 4 machine repetitions, resulting in a total of 124 spectra for this experiment. Sample contents are shown in Table S1 (see Appendix 1).

**Detection of adulteration**

Stock solutions matching coconut water’s total sugar concentration (USDA 2016) obtained from the standard addition method were made up by dissolving sucrose (3.171 g), glucose (3.150 g) and fructose (3.172 g) in 50 mL deionised water (63 mg mL⁻¹). Additionally, a mixed sugar solution, matching a coconut water’s naturally occurring concentration for each sugar, was made up by dissolving 1.347 g glucose, 1.220 g fructose and 0.582 g sucrose in 50 mL deionised water (26.94 mg mL⁻¹ glucose, 24.40 mg mL⁻¹ fructose, 11.64 mg mL⁻¹ sucrose, 62.98 mg mL⁻¹ total). Samples were made up by adulterating coconut water with deionised water, single sugar solutions, or the mixed sugar solution in increasing ratios (5-100%, 5% increments), with 3 samples of unadulterated coconut water being used as control samples. A total of 103 samples were analysed using 4 machine replicates, resulting in a total of 412 spectra for this experiment. Adulteration ratios are presented in Table S2 (See Appendix 1).

**Raman spectroscopy**

360 µL aliquots of each sample was transferred into a 96 well quartz sampling plate. Raman spectra were collected using a Renishaw inVia Raman spectrometer (Renishaw Plc. Old Town, Wotton-under edge, Gloucestershire, UK) equipped with a laser wavelength of
785 nm and an Olympus MD Plan 10 x10 long distance objective (KeyMed (Medical & Industrial Equipment) Ltd. Southend-on-Sea, Essex, UK). A 600 l/mm grating was used and the acquisition was centred at 1600 cm\(^{-1}\), leading to a range of 408 cm\(^{-1}\) to 2579 cm\(^{-1}\) with 1015 bins. The optimal depth for maximum intensity was found using a depth-series acquisition, and 4 machine replicates were taken by taking a spectrum of each sample at optimal depth and 1, 2, and 3 \(\mu\)m closer using an automated system set up in-house. All spectra were collected as 6 accumulations and 10 s acquisition with a laser power of \(~30\) mW at the samples.

**Data analysis**

For standard addition, spectra were first subjected to manual cosmic ray removal in the WiRe Raman software used to control the instrument. They were then imported into Matlab R2017a (The MathWorks, Natick, USA) where they were subjected to baseline correction using a partial least squares algorithm (based on an algorithm by Eilers (Eilers 2003) and created as part of an in-house analysis package) using a smoothing parameter of 10,000 and an asymmetry parameter of 0.001. The data was then transferred into Microsoft Excel 2013, where the intensity of peaks unique to glucose (1123 cm\(^{-1}\)), fructose (627 cm\(^{-1}\)) and sucrose (835 cm\(^{-1}\)) were plotted against the concentration of the sugar added. A least-squares linear fitting algorithm was used to generate a line of best fit, which was then used to calculate the concentration of each sugar in coconut water using the following equation:

\[
[S] = 2 \frac{I_0}{a}
\]

Where \(S\) (Saxberg & Kowalski 1979) is the concentration of sugar, \(I_0\) is the \(y\)-intercept of the line of best fit, and \(a\) is the slope. The values were doubled to account for the 50% aqueous dilution.

For adulteration studies, spectra were subjected to manual cosmic ray removal, imported into MatLab R2017a, and baseline-corrected as described above, after which they were auto-scaled. Principal component analysis (PCA) (Gromski, Muhamadali, *et al.* 2015) was used to reduce the dimensionality of these data into uncorrelated principal components (PCs) and describe relationships between samples. Additionally, partial least squares regression (PLSR) (Gromski, Muhamadali, *et al.* 2015) was used to generate a linear predictive model. This was
undertaken by dividing each dataset into training (0%, 10%,…, 100%) and test sets (5%, 15%,…, 95%). The training set was used to create several models, each with an increasing number of latent variables (LVs) used and a $k$-fold cross-validation (Gromski, Muhamadali, et al. 2015) system where $k=20$ was used to test each model. The optimal model was chosen by aiming to minimise the root mean squared error on the cross-validation set (RMSECV) and the number of LVs, and the test set was fed into it.

Results and discussion

Standard addition

As with all whole foods, which include fruits, vegetables, and whole grains, coconuts can vary in their total water volume and nutrient content (Manivannan et al. 2018). As well as slight natural individual variation within the same species, broader differences in coconut water can be a result of several factors such as differences between species (Prades, Dornier, et al. 2012b), their country or area of origin (Loiola et al. 2016), as well as their maturity (Santoso et al. 1996). The nature of the current study also required an accurate knowledge of the sugar content of our stock coconut water, as to mimic intentional adulteration with sugar solutions of incorrect concentrations would simply result in poor masking. The standard addition method (SAM) was therefore chosen as a method to calculate each sugar concentration. As well as serving as a proof-of-concept, in ensuring that coconut water and the changes within it can reliably be detected, it is also a simple yet effective self-validating method, removing the requirement for a second quantitative method like HPLC or GC to confirm the results. As the spiked solution is the unknown solution, any matrix effects and variances that could occur using standard concentration curves are eliminated. Standard addition functions (Westley et al. 2017; Saxberg & Kowalski 1979) by using the solution of unknown concentration as a base and adding known quantities of the analyte in question to create a model where concentration points are seen as $x, x + n, x + 2n, x + 3n…$ where $x$ is the concentration of the unknown sample and $n$ is the known quantity added. These points can then be used to create a regression model which is then shifted to place the signal at a true 0 concentration. Finally, the signal at concentration $x$ is plotted into the shifted regression model to find the true concentration.
The strongest peak of each signal was chosen as a reference to perform the standard addition calculations. As depicted in Figure S1 (see Appendix 1), these can be found at 627, 835, and 1123 cm\(^{-1}\) for fructose, sucrose, and glucose respectively. SAM assumes that the signal at 0 concentration will be 0, baseline correction was performed to minimise error. However, scaling was found to be detrimental to the model and was not performed.

Using SAM the sugar concentrations were found to be 24.39, 26.95 and 11.64 mg mL\(^{-1}\) for fructose, glucose and sucrose respectively (Figure S2 (see Appendix 1)). These concentrations were in agreement with the values reported in the literature, ranging from 24, 27, and 15 mg mL\(^{-1}\) for each sugar as reported by the Food Standards Australia New Zealand nutrient database (FSANZ 2010), to 25.5, 26.1, and 0.6 mg mL\(^{-1}\) as reported by Santoso et al. (Santoso et al. 1996). While the monosaccharide concentrations are fairly constant, a large variance in the sucrose concentration is observed, which is likely due to the different stages of maturity of the coconuts tested (Tan et al. 2014; Child & Nathanael 1950; Costa et al. 2015). Child and Nathanael (Child & Nathanael 1950) observed an increase in sucrose concentration beginning around the 7\(^{th}\) month of growth. The experimental values can therefore be confidently stated as sufficiently accurate.

Adulteration with solutions of single sugars

Dilution with water, as well as being readily detectable with Raman spectroscopy (Figure S3) and other methods (Ellis et al. 2012a), would also result in noticeable changes in organoleptic characteristics (Ellis & Goodacre 2002) if coconut water were watered down at higher concentrations. As the majority of the total dry weight of coconut water is sugar, keeping this concentration constant would be a simple way to not only mask this watered down taste (David I Ellis et al. 2016), but also often require the use of expensive and time consuming chromatography or enzymatic methods to detect the adulteration. Moreover, in order that the taste of the final product remains consistent, producers can add sugars without dilution to normalise the sweetness of coconut water and total soluble solids content as measured by Degrees Brix (°Bx) (Nicolaï et al. 2007) originating from several different processing plants.
Figure 3.1: PCA scores plot of the adulterated coconut water with increasing quantities of individual sugar solutions, using 2 PCs and achieving 90.1% total explained variance (TEV). A: PCA scores plot comparing PC 1 and PC 2. B: Loadings plot for PC 1 (black line). C: Loadings plot for PC 2 (black line). The reference peaks for fructose, glucose and sucrose (627, 1123 and 835 cm$^{-1}$ respectively) were highlighted, and their Raman spectra (red, green and blue respectively) were overlaid onto the loadings.

To obtain an understanding of the ability of Raman spectroscopy to discriminate between sugars, coconut water was replaced with individual solutions of fructose, glucose and sucrose in various levels in 5% increments (0%, 5%,..., 100%), while keeping the total sugar concentration at a constant 63 mg mL$^{-1}$. This process would allow for adulterated products to remain undetected using Brix measurements. PCA was then used to reduce the dimensionality.
of the data, and the loadings were examined for the most relevant principal components so as to understand the possible sources of the explained variance (Figure 3.1).

Even when the total sugar concentration is kept at a constant, any alteration in the concentration of individual sugars remains clearly detectable, as displayed on the PCA scores plot of all the data (Figure 1). From the PCA plot it can be observed that each sugar extends linearly in a separately distinct direction from the middle (or source) point occupied by pure unadulterated coconut water, with PC 1 separating fructose from glucose and sucrose, whilst PC 2 separates sucrose (positive side) from glucose and fructose (negative side). Additionally, the gradual colour change in the plot from black to each sugar’s respective chosen colour indicates that each trend is concentration dependent. Only the 95% and 100% data points for fructose adulteration samples were found to be outliers. Further evaluation of the spectra for these data points led to the observation of an abnormal signal in the 400-500 cm$^{-1}$ range, likely due to an as yet unexplained reduction in signal-to-noise ratios in this region. Since each sugar has a distinct spectrum, clear similarities between the PCA loadings plots (Figure 1B and 1C) and the spectra for each sugar can be found. The loadings for PC 1, for example, indicated a positive weighting for the fructose reference peak at 627 cm$^{-1}$ along with the two peaks at 822 and 871 cm$^{-1}$, while indicating a negative weighting for the glucose reference peak at 1123 cm$^{-1}$ and the two peaks at 1331 and 1363 cm$^{-1}$. The loadings for PC 2 further confirm this, as they show positive influence of the reference sucrose peak at 835 cm$^{-1}$ and the 1132 cm$^{-1}$ signal, among others. From this information, it is clear that Raman spectroscopy in combination with chemometrics can accurately detect low levels of adulteration of coconut water with a single sugar solution, even if the total concentration of sugar is kept constant.

Adulteration with a mixed sugar solution

The next step in exploring the discriminatory potential of Raman spectroscopy in detecting adulteration with sugars was to adulterate coconut water with a mixed sugar solution. To ensure comparable results, the experimental conditions and analytical protocols were kept identical. Additionally, the control dataset using unmasked dilution with water was included as a comparison. The PCA score plot (Figure 2) shows clear and regular trends relating to adulteration using solutions of single sugars. Additionally, an equivalent trend in
the scores plots is visible for the unmasked dilution with water, demonstrating that such a practice would be readily detectable with Raman spectroscopy. It is worth noting that, although the samples adulterated with varying concentrations of single sugars (glucose and sucrose) are also on the negative side of PC1 axis (Figure 3.2B), similar to that of the water-diluted samples, using PC3 scores (Figure 3.2C) these samples display distinct clusters which allows for their discrimination from one another.

Figure 3.2. PCA scores plots of the adulteration of coconut water (yellow stars) with water (grey stars), 63 mg mL⁻¹ solutions of fructose (red circles), glucose (green squares) and sucrose (blue triangles), and a mixed sugar solution (purple diamonds). The analysis was performed using 3 PCs and achieved 90.3% TEV. A: 3D scores plot, B: scores plot of PC 1 vs PC 2, C: PC 1 vs PC 3, D: PC 2 vs PC 3.

Unlike the use of individual sugars to adulterate coconut water in order to mask its dilution with water, the use of a mixed sugar solution to mask dilution while keeping individual sugar concentrations constant is significantly more difficult to detect. Compared to the PCA scores of earlier adulteration, adulteration with a mixed sugar solution shows little to no change relative to the other conditions investigated. This is predictable and even desired to
an extent; given the variance for each principal component originates from changes in the concentration of each individual sugar, a lack of change indicates that these concentrations have not altered and that the masking of dilution was successful. To examine the source of variance within the mixed sugar adulteration dataset, PCA analysis was performed on the mixed sugar dataset alone (Figure 3.3). Immediately, there are several factors present in the scores plot (Figure 3.3A) which point to a far weaker detection ability. Additionally, the total explained variance (TEV) of each principal component is far smaller; where previously 3 PCs achieved a TEV of >90%, the use of 10 PCs only achieved ~44% TEV. Although the trend is weaker, it is still present; a gradual change from black (0% adulteration) to purple (100% adulteration) can be seen (Figure 3.3A).
Figure 3.3: PCA of the adulteration of coconut water with a mixed sugar solution in 5% increments ranging from 0-100%, using 2 PCs to achieve 25.7% TEV. A: PCA scores plot comparing PC 1 and PC 2. Increasing levels of adulteration are represented by a colour change from black to purple. B: Loadings plot for PC 1, indicating the variance corresponding to each wavenumber sampled. C: Overlaid Raman spectra after baseline correction and autoscaling. For each level of adulteration, the average of four machine repetitions is presented. Once again, increasing levels of adulteration are represented by a colour change from black to purple.

In addition to the scores, the loadings plot was also examined to establish possible sources of variance within the dataset, and several areas of high variance were identified and
highlighted. The negatively weighted peaks at 835 and 1132 cm\(^{-1}\), indicative of sucrose, and the positively weighted peak at 627 cm\(^{-1}\), indicative of fructose, suggest that masking of dilution was not entirely successful (Figure 3.3B). Based on Figure 3.3A showing a decrease in scores with respect to increasing adulteration, it can be inferred that sucrose was slightly in excess and that an insufficient concentration of fructose was used. Additionally, there are two peaks with positive weighting at 967 and 1417 cm\(^{-1}\) which are not attributed to any sugar (Figure 3.3B). In Figure 3.3C, these signals can be attributed to two shoulders in the coconut water spectra, which recede with increasing adulteration. These peaks could not be readily associated with any other constituent of coconut water, though the vibrational modes associated with the peak at 967 cm\(^{-1}\) could be attributed to either \(\nu(C-O-C)\) or \(\nu(CC\ chain)\) and 1417 cm\(^{-1}\) to \(\delta(CH3)\) or \(\delta(CH2)\).

**Quantitative analysis**

To acquire a quantitative measure of detection and predictive ability, partial least squares regression (PLSR) was employed. Example plots comparing the PLSR model for the dilution and mixed sugar adulteration samples are presented in Figure 3.4, and complete statistical results for each series can be found in Table 3.2. It should be noted that as the 95% and 100% data points in the masking of dilution with fructose series were clear outliers, as also displayed in the PCA scores plot (Figure 3.1A), they were excluded from the model for this dataset only.

<table>
<thead>
<tr>
<th></th>
<th>LVs</th>
<th>(R^2)</th>
<th>(Q^2) (CV)</th>
<th>(Q^2) (tst)</th>
<th>RMSE (trn)</th>
<th>RMSE (CV)</th>
<th>RMSE (tst)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>4</td>
<td>0.9998</td>
<td>0.9972</td>
<td>0.9965</td>
<td>0.4286</td>
<td>1.3640</td>
<td>1.6743</td>
</tr>
<tr>
<td>Fructose*</td>
<td>3</td>
<td>0.9996</td>
<td>0.9967</td>
<td>0.9977</td>
<td>0.5504</td>
<td>1.3105</td>
<td>1.2323</td>
</tr>
<tr>
<td>Glucose</td>
<td>3</td>
<td>0.9997</td>
<td>0.9975</td>
<td>0.9984</td>
<td>0.5622</td>
<td>1.3020</td>
<td>1.1648</td>
</tr>
<tr>
<td>Sucrose</td>
<td>3</td>
<td>0.9997</td>
<td>0.9980</td>
<td>0.9987</td>
<td>0.5663</td>
<td>1.1551</td>
<td>1.0161</td>
</tr>
<tr>
<td>Mix</td>
<td>2</td>
<td>0.9957</td>
<td>0.9281</td>
<td>0.9671</td>
<td>2.0801</td>
<td>6.9218</td>
<td>5.2104</td>
</tr>
</tbody>
</table>

*Table 3.2: PLSR model statistical results for each dataset measured. For each dataset, \(R^2\), \(Q^2\) values for the k-fold (k=20) cross-validation and test sets, along with the RMSE values for the training, cross-validation and test sets are presented. \(R^2\) and \(Q^2\) values are measures of the total in the training, cross-validation and test sets explained by the model. * As the 95% and 100% data for the fructose dataset were found to be outliers, they were removed.*
The PLSR results of the water and single sugar solution adulteration series are largely equivalent and, most importantly, demonstrate very high concentration based discrimination abilities using Raman spectroscopy. For all four of these experiments, the $R^2$ and both the $Q^2$ values for cross-validation and test sets were above $>0.999$, indicative of a robust model. Additionally, the root-mean-squared error (RMSE) values for the test set, the optimal measure of each model’s predictive ability, all remain below 2%. While this is not an absolute measure of limit of detection, it does provide evidence that Raman could accurately detect this form of adulteration to within 5%. The PLSR results of the mixed sugar solution adulteration dataset, despite still showing evidence of a robust model with an $R^2$ value of $>0.99$ and $Q^2$ values of $>0.9$, demonstrates a weaker predictive ability with a RMSE value on the test set of 5.2%. This 4-fold increase is not surprising, as PCA demonstrated the relative weakness of the trend earlier. However, this also highlights the importance of sugar signals and their ratios in the detection of adulteration. The effect of this discrepancy in error-prediction can be visualised through the comparison of Figures 3.4A and 3.4B. The former, showing the PLSR model for the dilution series, shows very little variance with virtually no deviation from the middle blue line, indicating a very good fit between predicted and known adulteration values, with the latter showing far more variance in its cross-validation and test sets.

Figure 3.4: A) PLSR prediction model for unmasked dilution of coconut water. $Q^2$ and the RMSE for the test set are 0.9965 and 1.67% respectively. B) PLSR prediction model for dilution masked with a mixed sugar solution. $Q^2$ and the RMSE for the test set are 0.9671 and 5.21% respectively. A complete set of statistical values can be found in Table 3.2.
Concluding remarks

Coconut water is a refreshing product that has recently increased rapidly in popularity and sales, it is, quite literally, a very topical tropical drink. It is this sudden and exponential increase in demand for packaged coconut water during the last five years, and the short time period of only three months in which the source product is determined as being of optimum quality, that can itself lead to over demand and undersupply. These pressures then have the potential for dysfunctionality with supply chains making this product vulnerable (Spink et al. 2017) and a target for illicit practises such as mislabelling, substitution, misrepresentation, dilution, adulteration with sugars, and other forms of food fraud (Ellis et al. 2012a; Moore et al. 2012). In the case of coconut water, this has already been observed on multiple occasions (Glotz 2016a), from the extreme of one product being removed from shelves in the Caribbean due to it containing no coconut water at all (just chemicals and additives (Sorias 2016a)), to recent investigations in the UK by the National Food Crime Unit of the Food Standards Agency (Glotz 2017). More pertinent to our study here, this latter UK-based investigation found added sugar in 60% of samples, sugars derived from starch, sugar cane, and maize, despite the fact that the products were labelled as being pure and free from additives. This investigation led to a total of 400 tonnes of coconut water being seized and removed from the market (Glotz 2017). These cases readily illustrate the vulnerable nature of this type of product and the necessity for new methods to be able to detect these forms of adulteration, rapidly, and ideally on-site.

Here, we forward Raman spectroscopy as a potential tool for the detection of coconut water adulteration, which in combination with chemometrics successfully detected the dilution of this product, and its adulteration with solutions of single sugars to within 5%. Whilst these experiments were less successful in detecting mixtures of sugars, there is room for improvement, perhaps using laser wavelengths which are more suitable to specific types of product or type of adulteration, or closely related forms of Raman spectroscopy such as SORS (spatially offset Raman spectroscopy). SORS also has the added advantage of being able to penetrate through many types of non-metallic packing, such as bottles and cartons, to retrieve detailed chemical information, including a wide range of potential adulterants at extremely
low concentrations (Ellis et al. 2017). Whichever guise it may be in, in this study it can be concluded that Raman spectroscopy has significant potential as an accurate analytical method for the detection of adulteration in this popular product, able to discern small abnormalities in sugar ratios in coconut water while being rapid, simple to use, and with the potential for portability for on-site analyses (Ellis et al. 2015; Hargreaves 2014).

Acknowledgements

DIE and RG thank the UK ESRC and UK FSA for funding (Food fraud: a supply network integrated systems analysis (Grant number ES/M003183/1)).
Addition of an internal standard

A commonly used method to enhance predictive ability of a model is the addition of an internal standard. By adding a set amount of an external compound to all samples within a dataset, one can account for random fluctuations by 'anchoring' the spectral data to a predictable signal. This practice is used in the detection of melamine in milk, for example, to provide a stable reference intensity to compare to the relevant peak. A good internal standard is considered to be one that remains stable in the presence of the sample, and with spectral signals that do not overlap with important analyte peaks.

For this study, succinonitrile was chosen as an internal standard. The C≡N stretch vibration at 2260 cm\(^{-1}\) is in a signal-free region of the samples, and its other peaks were not found to affect the main signals. Additionally, it is a solid with very low volatility and adequate aqueous solubility. Although nitrile compounds are known to have some reactivity in water, no sensory or spectral evidence of this was observed.

![Figure 3.5: Black: Raman spectrum of aqueous succinonitrile after baseline correction. Red: Raman spectrum of unadulterated coconut water with 10.8 mg mL\(^{-1}\) succinonitrile added as an internal standard](image)

Methodology

Testing the effect of various internal standard concentrations

A concentrated solution of succinonitrile (Sigma Aldrich Ltd.) was made up by dissolving 1.581 g in 12.1 mL deionised water (131 mg mL\(^{-1}\)). Heat-treated coconut water
samples were adulterated with a mixed sugar solution analogous to the one mentioned above to create a sample set ranging from 0-100% adulteration (in 10% increments, 3 mL total volume). Each sample was then divided into five 500 µL aliquots, and each set of aliquots were spiked with varying quantities of succinonitrile solution (0, 25, 50, 75, 100 µL). Finally, the volume of each sample was equalised to 600 µL using deionised water. Raman data collection and analysis were performed in identical conditions as above (see Chapter 3, "Raman spectroscopy" and "Data analysis" sections).

**Testing the effect of internal standard with decreasing spectral quality**

A set of 1 mL samples with ratios identical to those described in the section above were made up and divided into two 500 µL sets aliquots. One set was spiked with 45 µL succinonitrile solution while the other was spiked with 45 µL deionised water. These sets were analysed with the same Raman apparatus as prior experiments (see Chapter 3, "Raman spectroscopy"), using 10 s acquisition times and an increasing number of acquisitions ranging from one to six. Data were pre-processed in an identical manner as in previous Raman experiments (see Chapter 3, "Data analysis") and analysis was undertaken using PLSR, using amended training (0%, 20%,…, 100%) and test sets (10%, 30%,…, 90%).

**Results and Discussion**

**Various succinonitrile concentrations**

To establish the correct quantity of succinonitrile to be added, an initial set of samples were spiked with various quantities of succinonitrile solution and analysed using PLSR. The $Q^2$ and root-mean-squared error of prediction on the test set were chosen as the deciding criteria (Figure 3.6).
Figure 3.6: Comparison of the predictive strength of models as a function of the quantity of succinonitrile added as an internal standard. For each model, 11 samples in a range from 0-100% adulteration (10% intervals) with a mixed sugar solution was used, which were divided into training sets (0%, 20%, ..., 100%) and test sets (10%, 30%, ..., 90%). A: $Q^2$ of the test set. B: RMSE on the test set

Little information on the optimal quantity of internal standard to add could be obtained from the results. Both the $Q^2$ and RMSE values presented in Figure 3.6 show highly erratic behaviour, and no clear relationship between predictive ability and concentration could be deduced. As such, it was concluded that the concentration of the internal standard had no demonstrable effect on the model’s predictive ability.

Varying the signal-to-noise ratio

To understand fully the effect of an internal standard on the ability to detect the masking of dilution with a mixed sugar solution, a second experiment was carried out, this time comparing two datasets, one with and one without internal standard, at various signal-to-noise ratios. This was achieved by acquiring data using varying number of spectral accumulations. Once again, $Q^2$ and RMSE for the test sets were used as criteria for the determination of efficacy.
Figure 3.7: PLSR model comparison between a dataset containing internal standard (orange) and a control without (blue) at different levels of signal-to-noise ratios, obtained by combining different numbers of 10 s spectral accumulations. PLSR models were created in a manner analogous to those in Figure 3.6. A: $Q^2$ on the test set, B: RMSE on the test set.

While the overall decreasing trend in RMSE as the signal-to-noise ratio increases is normal (Figure 3.7), the presence of succinonitrile as an internal standard did not have a meaningful positive impact on the results. At high signal-to-noise ratios (5+ accumulations) the model was marginally improved by the addition of internal standard; however, the RMSE begins to fluctuate below this threshold and leads to unpredictable results. Therefore, there is no conclusive evidence that the presence of an internal standard was beneficial to the methods’ ability to detect adulteration accurately.

Given the extensive use of internal standards (Robinson et al. 2004; Lorén et al. 2004; Aarnoutse & Westerhuis 2005) to stabilise measurements and increase predictive strength, the results presented here could be said to be unexpected. As such, it is important to analyse the experimental methodology and establish possible sources of error, so as to understand the validity of the results fully.

An examination of the experimental design should indicate that the present results are valid for the given conditions. As the internal standard was spiked using an aqueous solution, fluctuations in the concentration should be minimised to the error on the autopipettor. Furthermore, the quantity of water was consistently normalised so as to not create a dilution related artefact. While pipetting error when making up adulterated samples may be a valid concern, all datasets were created from one stock adulteration set. As such, any error should
be equally present in all experimental sets. A similar situation arises within the signal-to-noise ratio variation methodology. As all experiments were performed on one set of spiked and unspiked adulterated samples, each of them was subjected to a total of 24 accumulations of 10 s, or four times the exposure of previous experiments. The order of acquisitions was randomised so as to avoid the possibility of spectral drift, however some photodegradation may have happened. This being said, the consequences of such an effect on the results remain mitigated by the fact that both the spiked and unspiked datasets were subjected to equal laser activity and, as such, equivalent levels of photodegradation would have occurred.

Outside the sources of experimental error outlined above, another possibility that would explain the detrimental effect of an internal standard is succinonitrile simply being an inadequate internal standard. As mentioned earlier, nitrile compounds are known to react in water to form carboxylates and ammonia in acidic or basic conditions. Detectable consequences of this reaction would have been the presence of the smell of ammonia and a correlation between succinonitrile signal strength, an indicator of concentration, and the sample analysis order. Since the natural pH of coconut water is ~5.3 and neither of these signs were observed, it is likely that succinonitrile remained stable in solution; however, mass spectrometry could be used to obtain absolute evidence of this. Repeating the experiments presented would be a simple method to validate the results and confirm that they are not due to experimental and human error.
Chapter 4: Detecting the adulteration of coconut water with a mixed sugar solution using nuclear magnetic resonance spectroscopy

Introduction:

As explained earlier, coconut water is a valuable commodity that remains highly susceptible to stretching with water. As this practice on its own would have a negative impact on organoleptic characteristics, especially taste, and be readily detectable, a simple and inexpensive masking method would involve ensuring that the sugar levels remain constant. This would not only help to mask the watered down taste, but also avoid the detection of such bulking with dry weight analysis or methods unable to selectively quantify different sugars within the same mixture. When combined with chemometric methods, such as PCA and PLSR, Raman spectroscopy (Chapter 3) was found to reliably detect very small changes in the concentrations of glucose, fructose and sucrose in the samples, and could as such accurately detect and quantify dilution and masking with a single sugar solution down to below 5%. However, this detection ability was decreased when the concentrations of each individual sugar was kept constant with increasing levels of adulteration. This was found to be due to the relatively low concentration of other constituents in coconut water, leading to a lack of clear peaks in the coconut water spectra other than those attributed to sugars.

Proton nuclear magnetic resonance (\(^1\)H-NMR) was chosen as a comparative analytical technology, with the aim of obtaining a stronger predictive model. It uses series of radiofrequency pulses to alter the nuclear spin state of \(^1\)H atoms and, by measuring the frequency of proton signals (presented as chemical shift) and relaxation time, probe the chemical environment of protons. As such, a wealth of information about the functional group each proton is attached to, the concentration of each proton, along with protons connected to adjacent atoms, is provided leading to a complete picture of the samples being studied (Rodrigues et al. 2011; Heinzmann et al. 2010). However, similar to chromatography and mass spectrometry, it is a time consuming and resource-intensive method which generally requires optimisation and specific background training to operate.

To acquire data comparable to that acquired using Raman spectroscopy, a set of samples ranging from 0% to 100% adulteration with a mixed sugar solution was prepared and
analysed using an 800MHz NMR spectrometer, and the data were processed using PCA and PLSR.

**Methodology:**

**Adulterant solution preparation:**

Fructose (12.44 g), glucose (13.29 g) and sucrose (5.77 g) were dissolved in deionised water (500 mL) to make a solution emulating the concentrations of each of these sugars in coconut water (24.88, 26.58 and 11.54 mg mL$^{-1}$ respectively). To ensure that the pH and conductivity remained constant at all levels of adulteration, 1M K$_2$PO$_4$ (0.177 mL) and 1 M KH$_2$PO$_4$ (4.724 mL) was added to 45.1 mL sugar solution (pH: 5.31, conductivity: 5.69 mS m$^{-1}$). Heat-treated Coconut water (see Chapter 2) was then adulterated with this buffered solution (0-100% adulteration, 5% increments) in 2 mL samples and each aliquot was spiked with 50 µL of a 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) in D$_2$O solution (98.17 mg mL$^{-1}$) and divided into three 0.6 mL aliquots. Excluding one sample at 75% adulteration that was rendered unusable due to human error, all prepared samples were analysed using 1H-NMR spectroscopy, leading to a total of 62 spectra.

**NMR analysis:**

Proton 1D spectra were acquired at room temperature using a Bruker 800 MHz NMR AVANCE III spectrometer (Bruker UK Limited, Banner Lane, Coventry, CV4 9GH, UK) equipped with a 5mm TCI cryoprobe with temperature control unit using standard pulse parameters from the Bruker library.

**Data analysis:**

As the direct importation of NMR spectra into Matlab led to misalignment, raw spectra were first realigned through linear shifting to place the strongest DSS peak at 0.00 ppm, after which the data were truncated to include only the -0.0503-5.6004 ppm range (13277 data points per spectrum). Baseline correction was then performed using a de-trending algorithm and autoscaled, and the water signal at 4.8 ppm was removed. When required, the processed spectra were divided into two sections at 3.1962 ppm. PCA and PLSR (Gromski, Muhamadali, et al. 2015) analysis were performed in an identical fashion as in previous chapters.
To investigate the relationship between malic acid peaks chemical shift and adulteration, maxima locations for the strongest peaks in the signal at 2.7 ppm were found using the Matlab “findpeaks” command. These data were then imported into Microsoft Excel and a least squares linear regression algorithm was used to acquire a line of best fit. Due to the lack of malic acid in the 100% adulteration samples, these spectra were not included in these calculations.

Results:

As a large proportion of the NMR spectral range did not contain useful information (Figure 4.1), spectral ranges were reduced to include data points from -0.1 ppm to 5.56 ppm so as to minimise unnecessary noise and simplify subsequent data manipulation. This did exclude some small peaks in the 6-8 ppm range; however, they were weak ($10^4$ AU intensity) and most were also very broad, perhaps inferring that they were a result of complex interactions between constituents in coconut water.

Unlike Raman spectroscopy where signal intensity of different bands are generally on the same scale, NMR signals can have peaks of comparable importance with vastly different intensities. This difference in intensity of peaks generally leads to differences in the significance of variance and places an unfair importance bias on the more intense signals. In an attempt to address this, Pareto scaling (van den Berg et al. 2006) was used as a potential scaling method and compared to autoscaling, the method used in previous work. As the Pareto method scales using the variance i.e. the square root of the standard deviation used by autoscaling, it is less influenced by absolute intensity and allows overall weaker areas to be viewed with comparable weighting to the stronger areas (Gromski, Xu, et al. 2015; Parsons et al. 2007). However, giving less weight to the absolute intensity of signals also increases the influence of random variation in the form of noise and can thus have adverse effects on the data. Additionally, normalisation of the intensity of the internal standard peak at 0.0012 ppm was also tested. To obtain an accurate and meaningful comparison, a PLSR model was obtained from data after several pre-processing parameters, and the optimal method was chosen based on the root-mean-square error of prediction on the test set (RMSEP) (Appendix 2, Table 1).
The results of this experiment showed that for these data, autoscaling remained the superior pre-processing method of those tested. While smaller peaks were given a higher weighting, the effect of noise was also greatly amplified, which may be a factor leading to the weaker prediction ability. Additionally, normalisation was found to have a beneficial effect on the model when no scaling was applied, no effect when paired with autoscaling, and a detrimental effect when paired with Pareto scaling.

As stated earlier, the difference in concentration of various constituents led to signals relating to sugars to have higher intensities by orders of magnitude, resulting in a significant bias towards these signals. To mitigate this and obtain a more in-depth understanding of the signals specific to coconut water, spectra were divided into two parts: a less intense coconut water signal containing region (-0.1-3.2 ppm) and a more intense region containing sugar signals (3.2-5.6 ppm). These two spectral regions were analysed separately using PCA and
PLSR, and quantitative prediction models were compared to the results obtained from the use of the whole spectral range.
Figure 4.2: PCA of the sugar region in coconut water (3.2-5.6 ppm). A: PCA scores plot for PC 1 vs. PC 2 achieving a combined total explained variance of 88.4%. B: Loadings plot for PC 1, depicting the scores weighting for each point in the spectrum. C: Loadings plot for PC 2. D: Average spectra for each level of adulteration tested after realignment, baseline correction, autoscaling, and removal of the water signal at 4.8 ppm. Colour gradient to separate adulteration levels is identical as for the PCA scores plot.
Figure 4: Enhanced version of Figure 4 focusing on the diverse 3.2-3.3 area. A: PC 1 Loadings. B: Overlaid sugar spectra after baseline correction and adjustment. NMRA spectra for glucose (pink), fructose (green) and sucrose (blue).
The PCA scores plot for the sugar signal section (Figure 4.2A) shows a clear trend related to adulteration along a composite axis combining PC 1 and PC 2. While according to PC 2 scores there is some (14.6%) variance within each class, this variance remains orthogonal to the trend, and compared to PC 1 (73.8%) is minimal. Due to the density of signals, especially in the 3.2-4.2 region (Figure 4.3), it is difficult to establish a coherent and meaningful explanation based on the loadings. However, it is interesting to note that despite this, it is likely that even small imperfections in the concentrations of sugars in the adulterant would be readily detectable by this method.
Coconut water section:

Figure 4.4: PCA of the coconut water section of the spectra. A: PCA scores plot for PC 1 vs. PC 2 achieving a total explained variance of 89.7% with 2 PCs. B: Loadings plot for PC 1, depicting the scores weighting for each point in the spectrum. C: Loadings plot for PC 2. D: Average spectra for each level of adulteration tested after realignment, baseline correction and autoscaling. The colour gradient used to identify adulteration levels is identical to that used in the PCA scores plot.

Unlike the sugar section, the PCA scores plot for the coconut water constituent region (Figure 4.4A) showed very little variance within adulteration levels. While PC 1
4.4B) shows a clear and constant trend from positive to negative scores with increasing adulteration, PC 2 (Figure 4.4C) shows a “boomerang effect”, wherein the scores increase until the 50% adulteration mark and then begin to gradually decrease and return to their initial negative scores as the level of adulteration increases. The loadings in PC 1 show a close resemblance to the NMR spectra (Figure 4.4D), indicating that the trend is an overall decrease in intensity, with the only peaks not present being those ascribed to the internal standard at 0.00, 0.63, 1.75, 2.91, and 3.14 ppm. This effect is predictable, as virtually all the signals displayed in this region are ascribed to coconut water, and increasing the level of adulteration will consequently decrease the intensity of those peaks. The loadings of PC 2, however, show the main cause for this boomerang effect to be due to the two signals at ~2.70 and ~2.45 ppm ascribed to malic acid (Sucupira et al. 2017). Along with a decrease in intensity, a decrease in chemical shift occurs with increasing levels of adulteration. For both signals, this is characterised in the loadings by negative weighting at the more deshielded end of the spectrum in samples with lower adulteration levels, and a positive weighting when the effect of decreasing intensity in the signal becomes more impactful.
PLSR:

To obtain a quantitative predictive model that could be adequately compared with the results in Chapter 3, PLSR was employed following the same methods as previously employed above.

Table 4.1: Table comparing the PLSR results for the optimal models of the coconut water region, sugar region, and the combined range. For the coconut water (CW) region, the results of two models are presented as both seemed equally viable. LVs represents the number of latent variables used in each model, while trn, tst and CV represent the training, test and cross-validation sets respectively.

<table>
<thead>
<tr>
<th>Method</th>
<th>LVs</th>
<th>R²</th>
<th>Q² (CV)</th>
<th>Q² (tst)</th>
<th>RMSE (trn)</th>
<th>RMSE (CV)</th>
<th>RMSE (tst)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sugar region</td>
<td>3</td>
<td>0.9992</td>
<td>0.9991</td>
<td>0.9991</td>
<td>0.8863</td>
<td>0.7810</td>
<td>0.8416</td>
</tr>
<tr>
<td>CW region</td>
<td>3</td>
<td>0.9994</td>
<td>0.9989</td>
<td>0.9991</td>
<td>0.7631</td>
<td>0.8426</td>
<td>0.8515</td>
</tr>
<tr>
<td>CW region</td>
<td>5</td>
<td>0.9999</td>
<td>0.9995</td>
<td>0.9999</td>
<td>0.2993</td>
<td>0.5541</td>
<td>0.2974</td>
</tr>
<tr>
<td>Full Range</td>
<td>4</td>
<td>0.9998</td>
<td>0.9995</td>
<td>0.9996</td>
<td>0.4948</td>
<td>0.5889</td>
<td>0.5767</td>
</tr>
</tbody>
</table>

Based on the $R^2$ and $Q^2$ values for the cross-validation and test sets consistently reaching $>0.99$, the data presented demonstrated a very good fit to the linear predictive models. Additionally, the RMSE on the test sets are all within 1%, indicating that the models can quantitatively predict the level of adulteration within 5%. Two models are presented for the coconut water region, one using 3 LVs and one using 5 LVs. While the optimal model can generally be discerned from the plot comparing the RMSECV to the number of latent variables used in each model, both could be viewed as acceptable without supplementary testing. It is also interesting to compare the prediction error of the whole spectral range to that of the separated regions; while one would intuitively expect to see either an average of the two or, similarly to the PCA scores, a peak intensity based weighting leading to equivalent results to the sugar section, the combined sections seem to create a more powerful model than its individual components (with the provision that the 3 LV model is correct for the coconut water section). A possible reason for this may be a higher number of peaks that can be used to validate the model’s predictions, thus lowering random uncertainty. While increasing the number of data points used in the spectra also often increases the contribution of background
noise, in this case however, the signal-to-noise ratio in these spectra is so high that this drawback is not significant.

Comparing the prediction values of these models to those obtained using Raman spectroscopy highlights the difference in sensitivity and detection power of these two analytical methods. While Raman was only able to detect two very weak signals in coconut water other than those related to sugars, NMR was able to detect a multitude of minor organic components and accurately and linearly depict their decrease in concentration as a result of increasing adulteration. Furthermore, it showed far superior detection ability even when only utilising the peaks related to sugars.

When examining these data at a more practical level, the question of which method would be more suitable in an industrial context becomes difficult. Raman spectroscopy is inexpensive, portable, fast, and requires little training in sample preparation and analysis. Additionally, it can readily and adequately detect dilution and small changes in the concentrations of individual sugars even if the total sugar content remains constant. However, this detection ability loses effectiveness when the sugar ratios do not change, as the chemometric methods rely on alterations in very small peaks. NMR, on the hand, is a time consuming and resource intensive method that has not been made portable. However, these drawbacks are compensated by a powerful detection ability that is not hindered by masking attempts. As common industrial sweeteners such as high-fructose corn syrup and beet sugars have different fructose-glucose-sucrose ratios, these would likely be detected even at low concentrations by Raman spectroscopy. However, if a group was to invest more resources into masking dilution and use individual sources of glucose, fructose and sucrose to retain natural concentrations, Raman would likely fail and NMR would be required to provide adequate evidence of this practice.

**Drift in the chemical shift of malic acid signals:**

While the peaks present in the NMR spectra generally presented changes in intensity related to adulteration, three signals (~2.4, ~2.7, ~4.3 ppm) in the spectra presented an additional drift in chemical shift with respect to increasing sugar solution addition. All of these signals were attributed to malic acid, an organic di-carboxylic acid present in coconut water. Each signal shows a similar increase in chemical shift with increasing concentration of
coconut water that does not appear in any other signal and an increasing peak width lowering the resolution of multiplets. To obtain more information about these peaks, the chemical shift of two visible peaks in the 2.7 ppm range were plotted against the proportion of coconut water in the sample.

Figure 4.5: Average spectra depicting the three malic acid signals at A: ~2.45, B: ~2.72 and C: ~4.33 ppm after realignment, baseline correction and autoscaling. Colour gradient remains identical as in previous plots, ranging from blue (0% adulteration) to green, yellow, and red (100% adulteration). Top right: Malic acid molecule
Figure 4.6 depicts the relationship between the chemical shifts of two of the peaks in Figure 4.5B and the percentage adulteration. As the 100% adulteration levels do not contain these peaks, they were omitted.

It is clear from these plots that there is a linear relationship between the chemical shift of these malic acid peaks and the level of adulteration. The $R^2$ value of $>0.99$ indicates a strong fit, and the maximum error is approximately 5%, indicating a good consistency in the results. This calculation was not repeated with other peaks, however it is likely that similar effects would be observed.

While the evidence at hand suggests that this effect is real, it is difficult to establish its source without further experimentation. It is possible that it is an effect due to the pH set experimentally. Despite it having been kept at a constant $5.31\pm0.03$, the natural pH of coconut water, this happens to be very close to the second pKa of malic acid, estimated to be in the range of 5.0-5.2 (Skoog et al. 1999; Dawson et al. 1959). As such, it may be that the effect being observed is due to the ionisation of one of the carboxylic acid groups. However, this may also be a matrix interaction between malic acid and other components in coconut water.
If this is the case, the use of these malic acid signals may be a powerful and relatively simple quantitative detection method. Unlike peak intensity, chemical shift is an absolute measure in NMR and will remain constant given certain constant conditions e.g. pH. As such, a quantifiable and detectable change in the chemical shift of these peaks would provide as a result of adulteration would drastically simplify detection.

Conclusions:

NMR is a powerful analytical technique that is able to detect the dilution of coconut water even when masked with a solution emulating the individual concentrations of each of the sugars to within 5% adulteration. However, it remains a time consuming and resource intensive method requiring specific training and facilities. As such, despite its analytical superiority to Raman spectroscopy (Chapter 3), the difficulty of its widespread implementation means it is not a clear choice. Additionally, an upfield drift was observed in some of the peaks in the signals ascribed to malic acid with increasing levels of adulteration, with the two peaks examined at ~2.72 and ~2.70 ppm showing a consistent linear drift. Although further study is needed to fully ascertain the source of this drift and the parameters affecting it, this drift may be an effect of interactions with other constituents in coconut water. This could therefore provide a simple and absolute method to quantify the level of adulteration in coconut water.
Chapter 5: Conclusions and future work

Food fraud has been shown to be as old as the food processing and production systems themselves (Ellis et al. 2012a; Accum 1820; Hassal 1855) and, despite the best efforts of food standard authorities around the world, has grown to cost ~$40bn per year globally (Johnson 2014), said to be related to globalisation and increasingly complex, extended, and rapid distribution systems (Ellis et al. 2015). Along with the financial burden, fraud has led to hazardous conditions for the health and cultural sensitivities of consumers, with infamous incidents including the presence of melamine in milk, milk derivatives and pet food originating from China in 2007/2008 (The Lancet 2009) or the Horsemeat scandal in Europe (Elliott 2014). While not all incidents garner as much media attention, they all pose a threat to consumers and the reputations of companies found guilty. A major difficulty in combating food fraud is the multi-step and often international processing of many foods, leaving a multitude of opportunities for nefarious groups to engage in these activities; as such, it is in the best interest of not only the public and regulatory agencies, but also of retail companies to obtain robust, rapid, inexpensive and high-throughput detection screening methods so as to find and efficiently remove fraudulent products.

Coconut water is a refreshing beverage which has seen a rise in popularity in the last five years thanks to its commonly enjoyed taste and promoted health benefits as a natural alternative to isotonic sports drinks. Despite natural coconut water generally remaining sterile within its intact shell, its susceptibility to Maillard-type enzymatic oxidation and bacterial growth and delicate organoleptic properties mean that preservation is a costly challenge. Additionally, transport from the tropical areas where the fruit grows to retail outlets around the world incur yet more costs. These manufacturing costs, along with the current popularity of coconut water, mean a relatively high retail cost and a lucrative opportunity for adulteration. Some recent examples that have been discovered include the undeclared presence of dairy milk in coconut water products (Food Standards Agency 2015a; Barlass 2015; Food Standards Agency 2015b), which was found in more than occasion and led to recalls and one death as a result of anaphylactic shock, and investigations in the USA, Brazil and the UK (Glotz 2016a;
Glotz 2017; Sorias 2016a) which discovered unmentioned addition of sugars in multiple brands of coconut water with premium labels such as 'pure', 'no added sugar', etc.

While the contents (Santoso et al. 1996; Prades, Assa, et al. 2012; USDA 2016; FSANZ 2010), medical effects (Khan et al. 2003; Anurag & Rajamohan 2003; DebMandal & Mandal 2011; Sandhya & Rajamohan 2008), enzymatic reactions, and various possibilities for processing (Matsui et al. 2008; Tan et al. 2014; Prades, Dornier, et al. 2012a; Karmakar & De 2017; Cappelletti et al. 2015) of coconut water have been examined in detail, the novelty of its international popularity has led to a lack of appropriate research on detection of adulteration in this increasingly popular product. Current detection methods are generally based on enzymatic reactions or isotopic fractionation. While effective, these methods are often time consuming and resource intensive, or may lack selectivity between sugars due to their similar structures. As such, they are not suitable as rapid screening methods. Vibrational spectroscopic methods are fast, relatively inexpensive, non-destructive and have been made highly portable, allowing for high-throughput measurements to be easily implemented. Additionally, they can deliver a holistic fingerprint of a sample encompassing fatty acids, proteins, DNA, and sugars, meaning various accurate results can be acquired with a single sample. Similar studies has been performed on other juices using FTIR (David I Ellis et al. 2016; Rodriguez-Saona & Allendorf 2011), NIR (Rodriguez-Saona et al. 2001), and Raman (Hara et al. 2018) spectroscopy, however the coconut water model has not yet been explored in any substantial depth.
The studies presented with this thesis focused on FTIR and Raman spectroscopies, and their ability to detect adulteration of coconut water with milk or sugar solutions. Additionally, each example of adulteration was compared with other standard methods: HPLC-RI was tested for adulteration with milk, and NMR was tested for adulteration with sugar.

When studying the detection of bovine milk in coconut water, a comparative test between FTIR and Raman using PC-DFA found that both showed similar sensitivity, however Raman was negatively affected by the scattering properties of the fat globules in milk (Mather & Keenan 1975) and showed a far weaker signal-to-noise ratio. As such, IR was chosen to pursue further analysis. Using PLSR, the method was found to have a limit of detection of approximately 0.2-0.3%, which would allow it to detect deliberate adulteration but not contamination from negligent practice, which would have required detection in the ppm range. When the detection of lactose as a marker using HPLC-RI was tested as a comparative method, sugar separation was achieved in isocratic conditions using an amine column, but RI was found to have a detection limit of ~20%, which is insufficiently sensitive for any meaningful detection. Despite HPLC providing good separation between lactose and other sugars, RID’s low sensitivity rendered it unable to detect adulteration below 10%; further

<table>
<thead>
<tr>
<th>Method</th>
<th>Sample preparation</th>
<th>Analysis Time frame</th>
<th>Sensitivity</th>
<th>Approx. cost</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTIR</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>Chapter 2</td>
</tr>
<tr>
<td>HPLC-RI</td>
<td>+++</td>
<td>++</td>
<td>+</td>
<td>++</td>
<td>Chapter 2</td>
</tr>
<tr>
<td>Raman</td>
<td>+++++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>Chapter 3</td>
</tr>
<tr>
<td>NMR</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
<td>Chapter 4</td>
</tr>
<tr>
<td>SNIF-NMR</td>
<td>+</td>
<td>+</td>
<td>++++</td>
<td>+</td>
<td>(Martin, et al. 1996)</td>
</tr>
<tr>
<td>IRMS</td>
<td>++</td>
<td>+</td>
<td>++++</td>
<td>+</td>
<td>(Antolovich et al. 2001)</td>
</tr>
<tr>
<td>Lactose biosensors</td>
<td>++</td>
<td>++</td>
<td>+++</td>
<td>+++</td>
<td>(Yakovleva et al. 2012)</td>
</tr>
<tr>
<td>NIR</td>
<td>++++</td>
<td>+++</td>
<td>++</td>
<td>+++</td>
<td>(Rodriguez-Saona et al. 2001)</td>
</tr>
</tbody>
</table>
work combining HPLC with other methods such as MS-MS would likely allow for the detection of far lower concentrations of foreign constituents.

When studying the detection of solutions in coconut water, two examples of masking were studied: in one case, single sugars were used as an additive to ensure the overall sugar concentration remained constant, and in the other a mixed solution of fructose, glucose and sucrose emulating the natural concentrations found in the natural coconut water samples to ensure the concentration of each individual sugar remained constant. Using standard addition, the concentrations of glucose, fructose and sucrose were quantified without the requirement for secondary validation using unique peaks at 1123, 627 and 835 cm$^{-1}$ respectively, and the results were used as a basis for the adulteration solutions. When combined with PCA and PLSR, Raman spectroscopy was able readily to detect and quantify with excellent linearity unmasked stretching and masking with single sugar to within 5%, due to imbalances in the unique peaks mentioned above. However, when sugar concentrations in the adulterant solution emulated those in coconut water, the sensitivity was found to weaken to approximately 8%. This was found to be due to the low quantity and intensity of signals ascribed to other coconut water constituents; when the unadulterated coconut water spectrum was compared to the mixed sugar solution spectrum, only two small shoulders at 967 and 1416 cm$^{-1}$ were present only in coconut water. It is worth noting that the loadings explained some variance using slight differences in the signal intensities of each sugar, indicating that, despite unique peaks, the detection ability of Raman spectroscopy would be dependant in part on imperfections in the masking. Additionally, the addition of succinonitrile as an internal standard for the adulteration using mixed sugars was found to add no benefit to the model’s detection ability, instead creating more variation in the results.

Proton NMR was chosen as a comparative method to Raman spectroscopy, with the aim of comparing their respective abilities to detect stretching when masked with a mixed sugar solution. It was confirmed to be a more sensitive method than Raman spectroscopy, which allowed it to detect multiple signals unique to coconut water. Combining NMR spectroscopy with PCA and PLSR led to a quantitative model with a sensitivity of ~1% and an ability to readily detect 5% adulteration. Additionally, a regular drift in chemical shift of the signals ascribed to malic acid with increasing levels of adulteration was observed. This could
potentially be used as an absolute measure of adulteration; however, more research is needed to understand the mechanisms involved fully.

Future work

FTIR and Raman spectroscopy have both shown significant potential as accurate, non-destructive and high-throughput screening methods that could be used both by regulatory agencies and retail companies for quality control and fraud detection. However, much of the research presented here with vibrational spectroscopies is entirely novel and undertaken in a controlled laboratory environment. There, more laboratory-based experiments followed by in-field analysis of a broader range of samples would a logical progression of this form of applied research. To enhance further the detection ability of bovine milk in coconut water, for example, the addition of grated coconut meat into the coconut water to emulate coconut milk products. FTIR was found to be a powerful method to detect adulteration, however such a practice has so far only been discovered in coconut milk consumables. As stated previously (Chapter 2), coconut water consists of very little fat and protein, and bovine milk was detected using from the unnatural presence of these signals. As the addition of coconut meat would lead to the presence of natural fat and protein content, it may alter the detection ability found in FTIR. Similarly, the sugars used to mask the stretching (Chapter 3 and Chapter 4) were laboratory grade reagents. It is very unlikely that these would be used, as more common industrial sweeteners such as beet sugar, cane sugar or high-fructose corn syrup (HFCS) are far more cost effective and popular in industry. All of these have their own ratios of fructose, glucose and sucrose, meaning detection of their presence using Raman spectroscopy would depend on the induced imbalance in individual sugar concentration. While even small imbalances have been shown to be detectable and quantifiable, concrete evidence using industrial sweeteners would provide more robust evidence of Raman's suitability as a screening method. Additionally, the use of portable instruments would provide a closer estimate of the realistic detection ability along different parts of supply chains.

Along with studying industrial conditions, there are several secondary, albeit interesting, results that merit further study. While FTIR was considered to be more applicable to study due to the higher signal-to-noise ratio, Raman's insensitivity to water does provide an
advantage, as no drying procedure would be required. Varying the parameters to obtain an increased detection of milk constituents may therefore lead to better results. Different laser wavelengths (El-Abassy et al. 2012; Mendes et al. 2016), for example, may be able to overcome the scattering issues and lead to superior signals. Additionally, changing the acquisition range to exclude more of the carbohydrate area and include the C-H triglyceride vibrations at 2800-3000 cm\(^{-1}\) could allow for a stronger detection of the milk components while hiding the majority components of coconut water. Spatially offset Raman spectroscopy (SORS) has been shown to successfully detect extremely low levels of adulteration through packaging in counterfeit alcohol (Ellis et al. 2017). The use of this property would allow for quick, accurate and non-destructive measurements on products at any stage of manufacture, including those ready for retail, making it a potentially ideal route for further research.

Although standard addition using Raman spectroscopy was found to be a successful method to quantify the concentration of glucose, fructose and sucrose present in coconut water, external validation would provide more authority to these results. Similarly, further research on the drift in the malic acid chemical shifts in NMR spectroscopy may provide useful information and a simple yet absolute measure of adulteration. There are several possible causes for this effect. One of these is pH: along with being an acid and thus quite sensitive to small changes in pH, the natural pH of coconut water (~5.3) is very close to the calculated second pKa of malic acid (5.0-5.2) (Skoog et al. 1999; Dawson et al. 1959). As such, changes may be pH related despite efforts to control it. Another factor that should be tested for is the possibility of interactions with the phosphate buffer. Additionally, the drift could be as a result of changes in concentration, viscosity, or complex interactions with other constituents in coconut water; if this is the case, using the chemical shifts of these signals would likely be a good method to test for adulteration.

While groups wishing to engage in underhand adulteration practices have found creative methods to cut manufacturing costs and enhance the sensory properties of their products, families of products are generally manipulated in similar methods. In the case of fruit juices, the two other common adulteration methods alongside addition of sugar and water are substitution with less expensive fruit juices (Zhang et al. 2009; Vaclavik et al. 2012) and the passing of reconstituted juice as the premium “fresh” product. Although adulteration with
cheaper fruit juices may not be feasible with coconut water due to its delicate colour, taste and relatively low acidity (Bridges & Mattice 1939), the addition of coconut meat would likely mask this practice. Due to its extremely high water content, the sale of reconstituted coconut water marketed as fresh is likely to occur, as it provides several cost-cutting advantages especially relevant for a fruit grown in tropical regions and sold worldwide. The current methods used to detect reconstitution involve isotopic analysis, often with isotope ratio mass spectrometry (IRMS), to determine the origin of the water content (Jamin et al. 2003). Studying the subtle alterations which occur as a result of the water extraction procedure could provide a simpler detection method requiring less sample preparation; while vibrational spectroscopic methods have been shown to likely be insufficiently sensitive for this, NMR may be able to detect it (Sucupira et al. 2017).

Here, it has been shown that FTIR and Raman spectroscopies could potentially be used as inexpensive, high-throughput screening methods for the detection of economically motivated adulteration of coconut water. Through novel research into spectroscopic markers indicating the presence of bovine milk and foreign sugars, along with the interesting behaviour of some natural components, the door has now been opened to give this academic research 'real-world' significance.
Appendix 1

Coconut water, a topical tropical drink: detection of its adulteration with sugars via Raman spectroscopy and chemometrics - Supplementary information

Paul I. C. Richardson, Howbeer Muhamadali, David I. Ellis* and Royston Goodacre*

Manchester Institute of Biotechnology, School of Chemistry, Manchester, UK, M1 7DN

*Email: Roy.Goodacre@manchester.ac.uk__D.Ellis@manchester.ac.uk

Table S3: Sample contents for each sugar solution

<table>
<thead>
<tr>
<th>Coconut water (µL)</th>
<th>Sugar solution (µL)</th>
<th>Deionised water (µL)</th>
<th>Total (µL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>0</td>
<td>250</td>
<td>500</td>
</tr>
<tr>
<td>250</td>
<td>25</td>
<td>225</td>
<td>500</td>
</tr>
<tr>
<td>250</td>
<td>50</td>
<td>200</td>
<td>500</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>250</td>
<td>250</td>
<td>0</td>
<td>500</td>
</tr>
</tbody>
</table>
Table S4: Sample contents for each of the three adulterants (deionised water, individual sugar solution, mixed sugar solution)

<table>
<thead>
<tr>
<th>Coconut water (µL)</th>
<th>Adulterant (µL)</th>
<th>Adulteration (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>950</td>
<td>50</td>
<td>5</td>
</tr>
<tr>
<td>900</td>
<td>100</td>
<td>10</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0</td>
<td>1000</td>
<td>100</td>
</tr>
</tbody>
</table>
Table S3: Table of main peaks present in coconut water. The presence of these peaks in sugar spectra, along with their tentative vibrational assignments is reported.

<table>
<thead>
<tr>
<th>Peak (cm(^{-1}))</th>
<th>Present in:</th>
<th>Assignment(^{1,2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>423.3</td>
<td>Guc, Fuc</td>
<td>(CCC) Bend</td>
</tr>
<tr>
<td>519.7</td>
<td>Guc, Fuc (521), suc (524)</td>
<td>(CCO) Bend</td>
</tr>
<tr>
<td>627</td>
<td><strong>Fuc</strong></td>
<td>(CCO) Bend</td>
</tr>
<tr>
<td>706</td>
<td>Fuc</td>
<td>(CCO) Bend</td>
</tr>
<tr>
<td>821</td>
<td>Fuc</td>
<td>(C-C) Stretch</td>
</tr>
<tr>
<td>835 (shoulder)</td>
<td><strong>Suc</strong></td>
<td>(C-C) Stretch</td>
</tr>
<tr>
<td>869</td>
<td>Fuc, Suc (shoulder)</td>
<td>(C-C) Stretch</td>
</tr>
<tr>
<td>918</td>
<td>Guc, Fuc</td>
<td>(CCH) Bend</td>
</tr>
<tr>
<td>1063</td>
<td>Guc (1061), Suc (1065), Fuc (1065)</td>
<td>(C-O) Stretch</td>
</tr>
<tr>
<td>1123</td>
<td><strong>Guc</strong>, Suc (1132)</td>
<td>(COH) Bend</td>
</tr>
<tr>
<td>1265</td>
<td>Fuc, Guc, Suc</td>
<td>(CHH) Twist</td>
</tr>
<tr>
<td>1367</td>
<td>Guc, Suc, Fuc</td>
<td>(CHH) Wag</td>
</tr>
<tr>
<td>1457</td>
<td>Fuc, Suc, Guc</td>
<td>(CH(_2)) Stretch</td>
</tr>
</tbody>
</table>

**Bold**: standard addition reference peak for each sugar

Figure S3: Overlaid spectra for pure coconut water (black) and 63 mg mL$^{-1}$ solutions of fructose (red), glucose (green) and sucrose (blue). Spectra are averages of four machine replicates after least squares baseline correction and autoscaling. Reference peaks for each sugar have been highlighted.
Figure S4: Standard addition calibration results for fructose (A), glucose (B) and sucrose (C). The orange line represents the linear $I = mx+b$ regression results where $I$ is the signal intensity at each peak and $x$ is the concentration of the sugar being examined, and the green dashed line represents the “real” relationship between concentration and signal intensity excluding the effect of coconut water. The equation of best fit, $R^2$ and calculated concentration is also reported.
Appendix 2

Table 1: Table comparing various processing pre-processing methods for NMR data. Normalised refers to the normalisation of all spectra using the internal standard signal at 0.0012 ppm, while ‘no BC’ refers to models where baseline correction via a de-trending function was not performed. CV, tst and trn refer to the k-fold cross-validation set, test set and training set respectively, while LV refers to the number of latent variables used in each model.

<table>
<thead>
<tr>
<th>Method</th>
<th>LVs</th>
<th>$R^2$</th>
<th>$Q^2$ (CV)</th>
<th>$Q^2$ (tst)</th>
<th>RMSE (trn)</th>
<th>RMSE (CV)</th>
<th>RMSE (tst)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pareto not normalised</td>
<td>4</td>
<td>0.9991</td>
<td>0.9970</td>
<td>0.9980</td>
<td>0.9665</td>
<td>1.4157</td>
<td>1.2979</td>
</tr>
<tr>
<td>Pareto normalised</td>
<td>3</td>
<td>0.9974</td>
<td>0.9935</td>
<td>0.9965</td>
<td>1.6058</td>
<td>2.0807</td>
<td>1.7156</td>
</tr>
<tr>
<td>Autoscaled not normalised</td>
<td>4</td>
<td>0.9998</td>
<td>0.9995</td>
<td>0.9996</td>
<td>0.4948</td>
<td>0.5889</td>
<td>0.5767</td>
</tr>
<tr>
<td>Autoscaled normalised</td>
<td>4</td>
<td>0.9998</td>
<td>0.9995</td>
<td>0.9996</td>
<td>0.4948</td>
<td>0.5889</td>
<td>0.5767</td>
</tr>
<tr>
<td>Autoscaled no BC</td>
<td>4</td>
<td>0.9998</td>
<td>0.9995</td>
<td>0.9996</td>
<td>0.4947</td>
<td>0.5895</td>
<td>0.5776</td>
</tr>
<tr>
<td>Unscaled not normalised</td>
<td>4</td>
<td>0.9990</td>
<td>0.9979</td>
<td>0.9972</td>
<td>0.9785</td>
<td>1.1942</td>
<td>1.5258</td>
</tr>
<tr>
<td>Unscaled normalised</td>
<td>3</td>
<td>0.9988</td>
<td>0.9963</td>
<td>0.9982</td>
<td>1.1068</td>
<td>1.5799</td>
<td>1.2382</td>
</tr>
<tr>
<td>Normalised no BC</td>
<td>3</td>
<td>0.9988</td>
<td>0.9962</td>
<td>0.9971</td>
<td>1.1036</td>
<td>1.5813</td>
<td>1.2440</td>
</tr>
</tbody>
</table>
References


Johnson, R., 2014. Food Fraud and “Economically Motivated Adulteration” of Food and Food Ingredients, Washington D.C.


