Determination of water quality using NIR spectroscopy and aquaphotomics

by

Judi Psarrakis

Thesis presented in fulfilment of the requirements for the degree of

Master of Science in Food Science

at

Stellenbosch University

Department of Food Science, Faculty of AgriSciences

Supervisor: Dr PJ Williams

Co-supervisor: Prof GO Sigge

Co-supervisor: Dr C Lamprecht

March 2020

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: March 2020

Summary

Water is one of the world's most crucial resources as it is an essential component of all life forms. Although water covers 70% of the world's surface, freshwater is still a rare find. Freshwater makes up only 3% of the water found on earth, but two-thirds are either unavailable for use or are in the form of glaciers. Water scarcity and quality degradation present a major challenge to both developed and developing countries, as it poses a risk to both the environment and human health. With the increasing global population, water resources have been exposed to a variety of pollutants as a result of anthropogenic activities. This study aimed to investigate the use of near-infrared spectroscopy combined with the aquaphotomics approach as a screening method for water quality.

NIR spectroscopy combined with the aquaphotomics approach was used to differentiate between spring water from different sources and types of bottled water. Due to the variation in mineral content of the three sources, it was possible to differentiate between the sources. This resulted in differences in the water spectral patterns of the different sources. NIR spectroscopy combined with the aquaphotomics approach could distinguish between the three spring water sources.

With the use of the aquaphotomics approach, it was possible to distinguish between mineral and spring water. The two water types produced two completely different water spectral patterns, indicating that each water type had a different configuration of water species.

The water spectral pattern of river water filtered using different filter media was investigated. The effect of different filtration material can be monitored with aquaphotomics since the filtration process alters the hydrogen bonding and water molecular species within the water. With the use of the aquaphotomics approach, it is easy to track changes in water with respect to changes in the water clustering. Furthermore, river water monitored over a ten-day period, indicated that the water molecular species distribution changed each day. This was due to changes in the physico-chemical parameters of the river water.

NIR spectroscopy combined with the aquaphotomics approach can be used to monitor water changes. As the technique is able to differentiate between water types and the source of origin. This method can also to monitor variations in the water spectral patterns due to changes in the water spectral pattern as a result of filtration treatment and changes in water quality.

Opsomming

Water is een van die belangrikste middele ter wêreld, aangesien dit 'n noodsaaklike komponent van alle lewensvorme is. Alhoewel water 70% van die wêreld se oppervlak bedek, is vars water steeds 'n seldsame hulpbron. Varswater vorm slegs 3% van die water wat op aarde aangetref word, maar twee derdes is óf nie beskikbaar vir gebruik nie óf in die vorm van gletsers. Water skaarsheid en kwaliteit agteruitgang verteenwoordig 'n groot uitdaging vir beide ontwikkelde en ontwikkelende lande, soos dit 'n risiko vir beide die omgewing en menslike gesondheid inhou. Met die snellende groei in die wêreldbevolking is waterbronne blootgestel aan 'n verskeidenheid besoedelende stowwe as gevolg van antropogene aktiwiteite. Hierdie studie het ten doel gehad om die gebruik van naby-infrarooi spektroskopie, gekombineer met die “aquaphotomics” benadering, as 'n metode vir waterkwaliteit bepaling toe te pas.

NIR spektroskopie gekombineer met die “aquaphotomics” benadering is gebruik om te onderskei tussen fonteinwater uit verskillende bronne en tiepes gebottelde water. Die variasie in mineraalinhoud van die drie bronne, het dit moontlik gemaak om tussen die bronne te kan onderskei. Dit het gelei tot verskille in die water spektrale patrone van die verskillende bronne. NIR-spektroskopie, gekombineer met die “aquaphotomics” benadering kan onderskei tussen die drie waterbronne.

Met die gebruik van die “aquaphotomics” benadering was dit moontlik om tussen mineraal- en fonteinwater te onderskei. Die twee tipes water het twee heeltemal verskillende water spektrale patrone geproduseer, wat aandui dat elke tipe water 'n ander rangskinking van water spesies het.

Die water spektrale patroon van rivierwater wat met verskillende fmedia gefiltreer is, is ondersoek. Die effek van verskillende filtrasiemateriaal kan met “aquaphotomics” gemonitor word, aangesien die filtrasieproses die waterstofbinding en waterklasse verander in die water. Met die gebruik van die “aquaphotomics” benadering, is dit maklik om veranderinge in water op te spoor weens die veranderinge in die water spesies. Rivier water was ook gemonitor oor 'n tydperk van tien dae en het aangedui dat die verspreiding van water spesies verander elke dag. Dit was as gevolg van die veranderinge in die fisies-chemiese eienskappe van die rivierwater.

NIR-spektroskopie gekombineer met die “aquaphotomics” benadering kan gebruik word om veranderings in water te monitor. Aangesien die tegniek in staat is om tussen watertiepes en die oorsprong daarvan te kan onderskei. Hierdie metode kan ook

veranderings in die water spektrale patrone monitor as gevolg van veranderinge in die filtrasiebehandeling en veranderinge in die waterkwaliteit.

Acknowledgements

I would like to thank my supervisors, Dr Williams, Prof Sigge and Dr Lamprecht, for the continued support throughout my research. Their patience, guidance, motivation and knowledge were greatly appreciated throughout the my research and writing of my thesis.

The following people and institutions also need to be acknowledged as they played a key role in the completion of this thesis:

Ms Nina Muller and Dr Erika Moelich for supplying bottled water samples.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF.

The 3rd Aquaphotomics International Symposium Organizing Committee for providing me with a travel grant to attend the 3rd Aquaphotomics International Symposium 2018 held in Awaji, Hyogo, Japan.

All the staff and post-graduate students at the Department of Food Science, Stellenbosch University for providing a friendly working environment and support when needed.

To my writing buddies, Nicola, Anika, Angel, Daphne and Celeste, thanks for the support, laughter, late nights and memories. Without you, I would not have made it this far.

To Carissa, thank you for encouraging me to never give up and telling me it will be worth it in the end.

A special thanks to my parents, Jana & Hennie Gous and Aris & Nicolene Psarrakis, who backed every decision I made and providing me with all the support I needed.

Table of Contents

Declaration i

Summary ii

Opsomming iii

Acknowledgements v

Table of Contents vi

List of Figures ix

List of Tables xiv

List of abbreviations xv

Chapter 1 1

Introduction 1

1.1 References 3

Chapter 2 6

Literature review 6

2.1 Introduction 6

2.2 Water resources 8

2.3 Water quality 8

2.3.1 Drinking and irrigation water quality 10

2.4 The structure of liquid water 10

2.5 Near Infrared spectroscopy 13

2.5.1 Instrumentation 14

2.6 Aquaphotomics 15

2.7 Near-infrared spectral data analysis 21

2.7.1 Pre-processing 21

2.7.2 Multivariate data analysis 23

2.8 Conclusion 25

2.9 References 25

Chapter 3 34

Materials and methods....34
3.1 Influence of temperature....34
3.1.1 Samples....34
3.1.2 Spectral acquisition....34
3.2 Spring water from different sources....34
3.2.1 Samples....34
3.2.2 Spectral acquisition....35
3.3 Different types of bottled water....35
3.3.1 Samples....35
3.3.2 Spectral acquisition....35
3.4 River water and different filtration materials....35
3.4.1 Samples and sample preparation....35
3.4.2 Spectral acquisition....36
3.5 Physico-chemical analyses....36
3.6 NIR Instrumentation....37
3.7 Spectral data analysis....38
3.8 References....39
Chapter 4....41
Results and Discussion....41
4.1 The effect of temperature....41
4.1.1 Spectral analysis....41
4.1.2 Principal component analysis (PCA)....42
4.1.3 Conclusion....44
4.2 Spring water from different sources....45
4.2.1 Spectral analysis....45
4.2.2 Multivariate data analysis....47
4.2.3 Aquagrams....56
4.2.4 Conclusion....58

4.3 Different types of bottled water 59
4.3.1 Spectral analysis 59
4.3.2 Multivariate data analysis 61
4.3.3 Aquagram 70
4.3.4 Conclusion 71
4.4 River water and different filtration materials 71
4.4.1 Filtered river water 72
4.4.2 Changes in river water over time 83
4.4.3 Conclusion 92
4.5 References 92
Chapter 5 94
General discussion and conclusion 94
5.1 References 96

This thesis is presented in the format prescribed by the Department of Food Science at Stellenbosch University. The language, style and referencing format used are in accordance with the requirements of the International Journal of Food Science and Technology. This thesis represents a compilation of manuscripts where each chapter and sub-chapters are individual entities and some repetition between chapters has, therefore, been unavoidable.

List of Figures

Figure 2.1 Water clusters with different numbers of water molecules and conformational structures (Segarra-Martí *et al.*, 2012). 11
Figure 2.2 Schematic diagram of an FT-NIR instrument with a polarisation interferometer (BÜCHI Labortechnik AG, 2018). 14
Figure 2.3 Aquagram of pure honey and high fructose corn syrup (Bázár *et al.*, 2016). ... 17
Figure 3.1 Basic schematic of the columns constructed with the different filtration material. 36
Figure 4.1 Mean spectra obtained at the three different temperatures for the first overtone of water (1300 – 1600 nm). 41
Figure 4.2 Second derivative spectra (calculated with a Savitzky-Golay filter using 2nd order polynomial and 21 points) in the first overtone of water (1300 – 1600 nm) of the water scanned at the three different temperatures. 42
Figure 4.3 Principal component score plot (PC1 (98 %) vs. PC2 (2 %)) of the water scanned at the three different temperatures. 43
Figure 4.4 First principal component (PC 1) loadings plot for the water scanned at the three different temperatures. 43
Figure 4.5 Second principal component (PC 2) loadings plot for the water scanned at the three different temperatures. 44
Figure 4.6 Mean (unprocessed) spectra of spring water from three different sources. 45
Figure 4.7 Difference spectra (Savitzky-Golay smoothing with 2nd order polynomial and 21 points) of the three different sources. 46
Figure 4.8 Second derivative (calculated with a Savitzky-Golay filter using 2nd order polynomial and 21 points) spectra in the first overtone of water (1300 – 1600 nm) of the three different spring water sources. 46
Figure 4.9 Principal component analysis score plot (PC1 (62 %) vs. PC2 (26 %)) of spring water sourced from three different sources. 47
Figure 4.10 Principal component analysis score plot (PC2 (26 %) vs. PC3 (7 %)) of spring water sourced from three different sources. 48
Figure 4.11 Principal component analysis score plot (PC3 (7 %) vs. PC4 (3 %)) of spring water sourced from three different sources. 48
Figure 4.12 PCA loadings line plot for PC 1(62%) 49
Figure 4.13 PCA loadings line plot for PC 2 (26%) 49
Figure 4.14 PCA loadings line plot for PC 3 (7%) 50

Figure 4.15 PCA loadings line plot for PC 4 (3%). 50
Figure 4.16 3D PLS-DA score plot (LV1 (34.23 %) vs LV2 (26.89 %) vs LV3 (26.10 %)) of source A (red), B (green) and C (blue).....52
Figure 4.17 PLS-DA score plot (LV1 (34.23 %) vs LV2 (26.89 %)) of source A (red), B (green) and C (blue).....52
Figure 4.18 PLS-DA prediction score plot of source A (red) showing the predicted objects. 53
Figure 4.19 PLS-DA prediction score plot of source B (green) showing the predicted objects 53
Figure 4.20 PLS-DA prediction score plot of source C (blue) showing the predicted objects 54
Figure 4.21 PLS-DA regression vector of source A in the 1300 – 1600 nm wavelength range. 55
Figure 4.22 PLS-DA regression vector of source B in the 1300 – 1600 nm wavelength range. 55
Figure 4.23 PLS-DA regression vector of source C in the 1300 – 1600 nm wavelength range 56
Figure 4.24 Aquagrams visualising the spectral patterns of source A (blue), B (red) and C (green). 57
Figure 4.25 Mean spectra of the different brands of bottled water samples. 59
Figure 4.26 Difference spectra (Savitzky-Golay smoothing with 2nd order polynomial and 21 points) of the different brands of bottled water samples.....60
Figure 4.27 Second derivative (calculated with a Savitzky-Golay filter using 2nd order polynomial and 21 points) spectra in the first overtone of water (1300 – 1600 nm) of the different bottled water samples. 60
Figure 4.28 Principal component analysis score plot (PC1 (68 %) vs. PC2 (20 %)) of bottled water. 61
Figure 4.29 Principal component analysis score plot (PC2 (20 %) vs. PC3 (7 %)) of bottled water. 62
Figure 4.30 Principal component analysis score plot (PC3 (7 %) vs. PC4 (3 %)) of bottled water. 62
Figure 4.31 Principal component analysis score plot (PC4 (3 %) vs. PC5 (1 %)) of bottled water. 63
Figure 4.32 PCA loadings line plot for PC 1 (68%). 63
Figure 4.33 PCA loadings line plot for PC 2 (20%). 64

Figure 4.34 PCA loadings line plot for PC 3 (7%). 64
Figure 4.35 PCA loadings line plot for PC 4 (3%). 64
Figure 4.36 PCA loadings line plot for PC 5 (1%). 66
Figure 4.37 3D score plot (LV1 (45.36 %) vs LV2 (22.29 %) vs LV3 (19.63 %)) of mineral (red) and spring (green) water. 67
Figure 4.38 PLS-DA score plot (LV1 (45.36%) vs LV2 (22.29%)) of mineral (red) and spring water (green). 67
Figure 4.39 PLS-DA prediction score plot of mineral water (red) showing the predicted samples. 68
Figure 4.40 PLS-DA prediction score plot of spring water (green) showing the predicted samples. 68
Figure 4.41 PLS-DA regression vector of mineral water in the 1300 – 1600 nm wavelength range.....69
Figure 4.42 PLS-DA regression vector of spring water in the 1300 – 1600 nm wavelength range.....69
Figure 4.43 Aquagrams visualising the spectral patterns of the mineral (blue) and spring (red) water. 70
Figure 4.44 Mean spectra for river water and the different filtrate samples.....72
Figure 4.45 Difference spectra (Savitzky-Golay smoothing with 2nd order polynomial and 21 points) of filtered and unfiltered water. 73
Figure 4.46 Second derivative (calculated with a Savitzky-Golay filter using 2nd order polynomial and 21 points) spectra in the first overtone of water (1300 – 1600 nm) of the river and filtered water samples. 73
Figure 4.47 Principal component analysis score plot (PC1 (74 %) vs. PC2 (14 %)) of the river (RW) and filtered water (activated (AC), black wattle biochar (BW), pine biochar (PN) and sand (SA)).....74
Figure 4.48 Principal component analysis score plot (PC2 (14 %) vs. PC3 (8 %)) of the river (RW) and filtered water (activated (AC), black wattle biochar (BW), pine biochar (PN) and sand (SA)).....75
Figure 4.49 Principal component analysis score plot (PC3 (8 %) vs. PC4 (2 %)) of the river (RW) and filtered water (activated (AC), black wattle biochar (BW), pine biochar (PN) and sand (SA)).....75
Figure 4.50 Principal component analysis score plot (PC4 (2 %) vs. PC5 (1 %)) of the river (RW) and filtered water (activated (AC), black wattle biochar (BW), pine biochar (PN) and sand (SA)).....76

Figure 4.51 PCA loadings line plot for PC 1 (74%). 77
Figure 4.52 PCA loadings line plot for PC 2 (14%). 78
Figure 4.53 PCA loadings line plot for PC 3 (8%). 78
Figure 4.54 PCA loadings line plot for PC 4 (2%). 78
Figure 4.55 PCA loadings line plot for PC 5 (1%). 79
Figure 4.56 PLS regression vector of the filtered river water using activated charcoal (AC) in the 1300 – 1600 nm wavelength range. 80
Figure 4.57 PLS regression vector of the filtered river water using black wattle (BW) biochar in the 1300 – 1600 nm wavelength range. 81
Figure 4.58 PLS regression vector of the filtered river water using pine (PN) biochar in the 1300 – 1600 nm wavelength range. 81
Figure 4.59 PLS regression vector of unfiltered river water (RW) in the 1300 – 1600 nm wavelength range. 82
Figure 4.60 PLS regression vector of the filtered river water using sand (SA) in the 1300 – 1600 nm wavelength range. 82
Figure 4.61 Aquagram visualising the spectral patterns of AC (yellow), BW (blue), PN (orange), RW (black) and SA (purple). 83
Figure 4.62 Difference spectra (Savitzky-Golay smoothing with 2nd order polynomial and 21 points) of the river water collected for the 10 days. 84
Figure 4.63 Second derivative (calculated with a Savitzky-Golay filter using 2nd order polynomial and 21 points) spectra in the first overtone of water (1300 – 1600 nm) of the river water samples. 84
Figure 4.64 Principal component analysis score plot (PC1 (52 %) vs. PC2 (26 %)) of river water collected over a period of ten days. 85
Figure 4.65 Principal component analysis score plot (PC2 (26 %) vs. PC3 (10 %)) of river water collected over a period of ten days. 86
Figure 4.66 Principal component analysis score plot (PC3 (10 %) vs. PC4 (6 %)) of river water collected over a period of ten days. 86
Figure 4.67 Principal component analysis score plot (PC4 (6 %) vs. PC5 (3 %)) of river water collected over a period of ten days. 87
Figure 4.68 Principal component analysis score plot (PC5 (3 %) vs. PC6 (1 %)) of river water collected over a period of ten days. 87
Figure 4.69 PCA loadings line plot for PC 1 (52%). 88
Figure 4.70 PCA loadings line plot for PC 2 (26%). 89
Figure 4.71 PCA loadings line plot for PC 3 (10%). 89

Figure 4.72 PCA loadings line plot for PC 4 (6%). 89
Figure 4.73 PCA loadings line plot for PC 5 (3%). 90
Figure 4.74 Aquagrams visualisng the spectral patterns of river water for the ten days. ..91

List of Tables

Table 2.1 Characteristic water absorbance bands identified 16
Table 2.2 Summary of aquaphotomics applications 19
Table 4.1 The overall performance measures of the calibration, cross-validation and validation PLS-DA models 51
Table 4.2 Performance measures used to evaluate the PLS-DA models of the three spring water sources 54
Table 4.3 Typical mineral content, pH, alkalinity and TDS of the three different sources provided on the labels (mg/L) 58
Table 4.4 The overall performance measures of the calibration, cross-validation and validation PLS-DA models of mineral and spring water 66
Table 4.5 Performance measures used to evaluate the PLS-DA models of mineral and spring water 69
Table 4.6 Typical mineral content, pH, alkalinity and TDS of the different mineral and spring water samples.....71
Table 4.7 PLS statistics of the models.....79
Table 4.8 WAMACS of interest and their assignments for the regression vectors of AC, BW, PN, RW and SA 80
Table 4.9 Physico-chemical data of the river water for the ten days 91

List of abbreviations

CWT	Continuous wavelet transform
EC	Electrical conductivity
EFA	Evolving factor analysis
FT-NIR	Fourier transform near-infrared
HCA	Hierarchical cluster analysis
LDA	Linear discriminant analysis
MCR-ALS	Multivariate curve resolution – alternating least squares
MSC	Multiplicative scatter correction
NIR	Near-infrared
OSC	Orthogonal scatter correction
PC	Principal component
PCA	Principal component analysis
PCA-QCC	Principal component analysis – quality control chart
PCR	Principal component regression
PLS	Partial least squares
PLS-DA	Partial least squares – discriminant analysis
SIMCA	Soft independent modelling of class analogy
SNV	Standard normal variate
TDS	Total dissolved solids
TSS	Total suspended solids
UVT %	Ultraviolet transmission percentage
VSS	Volatile suspended solids
WABS	Water absorbance bands
WAMAC	Water matrix coordinate
WASP	Water spectral pattern
WHO	World Health Organization
WWF	World Wildlife Fund

Chapter 1

Introduction

Water is one of the world's most crucial resources as it is an essential component of all life forms. Although water covers 70% of the world's surface, freshwater is still a rare find (WWF, 2017). Freshwater makes up only 3% of the water found on earth, but two-thirds are either unavailable for use or are in the form of glaciers (WWF, 2017). Globally, it is estimated that 1.1 billion people do not have access to water and another 1.8 billion people only have access to water contaminated with faecal matter (WHO, 2016; WWF, 2017). According to the current water usage and rainfall trends it has been estimated that half of the world's population will be living in water-scarce areas within the next 10 years (WHO, 2016).

The scarcity of freshwater is a growing concern due to the increased demand for water for human consumption and agricultural uses (Hoekstra *et al.*, 2012). Water scarcity and water quality degradation present a major challenge to both developed and developing countries as it is a risk to both humans and the environment (Kirby *et al.*, 2003). Water is not only a major constituent in biological systems, but it also plays a vital role in various processes ranging from agricultural to industrial activities. Approximately 40% of the food demand is supplied by 19% of the irrigated croplands (Molden, 2007). The increasing demand for water as a result of urbanisation, industrialisation and other non-agricultural activities are overshadowing the importance of water required for agricultural purposes, which poses a threat to food security (Hanjra & Qureshi, 2010; Molden *et al.*, 2010).

Food production and energy production industries compete for the demand for freshwater, and each year their demand increases (Kirby *et al.*, 2003). This increased requirement and decreased availability of freshwater certainly affect water quality. The increasing scarcity of water is forcing food manufacturing companies to attempt to optimise the use of water currently available. The waste products of food production and processing industries are a major consideration in the water cycle as it negatively influences the water quality (Kirby *et al.*, 2003).

Water quality is a complex problem since numerous chemical and microbiological analyses are required to provide an indication of the overall water quality (Gowen *et al.*, 2015). Currently, water resources are greatly contaminated due

to natural effects, such as algal blooms, decomposing organic matter and contact with toxic minerals (arsenic, mercury, lead), as well as large-scale human activities (Kovacs *et al.*, 2016). Water is characterised by different quality parameters, such as the concentration of microbiological, organic and inorganic contamination as well as indicators which include pH, total dissolved solids (TDS) and conductivity (Gowen *et al.*, 2012).

The biological and physico-chemical characteristics of water are used to classify water according to its intended use or purpose, such as drinking, food preparation, irrigation and industrial uses (Flörke *et al.*, 2013; Velíšek, 2014). Water quality guidelines have been set by the World Health Organization (WHO), outlining the quality parameters that are required to ensure that the water is safe for human and animal consumption (WHO, 2011). However, water quality standards differ from country to country.

Water quality is the result of the combination of chemical, physical and biological properties of the water (Adu-Manu *et al.*, 2017). Studies by Olivier (2015) and Britz *et al.* (2013) showed that the quality of rivers in the Western Cape has become a rising concern. The increase in river pollution is mainly due to inadequate sanitation amenities and insufficient sewage treatment services throughout South Africa (Olivier, 2015). If polluted water is used for irrigation of fresh produce, it can have a negative effect on food safety (Gulati *et al.*, 2013). With the increasing scarcity of water and increasing demand for safe water, alternative methods for rapid intervention, continuous monitoring and sanitation are required. Conventional methods require intensive sampling systems, multistep sample preparation and manual input, limiting their incorporation into continuous monitoring systems. Therefore, there is a need for a rapid screening method for water quality such as near-infrared spectroscopy.

Near-infrared (NIR) spectroscopy is a non-destructive analytical method, which enables the simultaneous measurement of qualitative and quantitative parameters of a range of different products. NIR spectroscopy is a vibrational spectroscopy technique, which forms part of the electromagnetic range of 800 -2500 nm (Burns & Ciurczak, 2007; Siesler, 2006). The absorption bands in the NIR spectrum corresponds with the overtones and combinations of the carbon-hydrogen (C-H), oxygen-hydrogen (O-H) and nitrogen-hydrogen (N-H) vibrations of molecules (Siesler, 2006). NIR spectroscopy has the potential to become a rapid water screening method

which is cost-effective and improves the speed of water quality testing (Gowen *et al.*, 2012).

Water is considered a hindrance in infrared spectroscopy since water absorbs across the entire frequency range (Tsenkova & Gowen, 2011). However, aquaphotomics is a field of study which uses NIR spectroscopy as a tool to better understand the association between light and water (Gowen *et al.*, 2015; Kovacs *et al.*, 2016). NIR spectroscopy combined with multivariate data analysis forms the basis for aquaphotomics (Tsenkova & Gowen, 2011). Aquaphotomics studies the interaction between water and light, providing information about aqueous and biological systems.

Aquaphotomics studies have shown that it is possible to differentiate between water with different sugar and salt concentrations (Bázár *et al.*, 2015; Gowen *et al.*, 2015), different types of mineral water (Munćan *et al.*, 2014) and pesticides in water (Gowen *et al.*, 2011). Thus, there is a need to investigate the effect water quality differences will have on the spectra. This study aimed at investigating the potential of using aquaphotomics as a rapid screening method for water quality determination. Specific objectives were established to study the effect of different perturbations on the water spectrum. These objectives are as follows:

- Investigate the effect of temperature (20˚C, 25˚C and 30˚C) on the water spectrum
- Differentiate between spring water originating from different sources
- Differentiate between bottled mineral and spring water
- Investigate the effect of filtration media and changes in water quality will have on the spectrum of river water.

1.1 References

Adu-Manu, K. S., Tapparello, C., Heinzelman, W., Katsriku, F. A. & Abdulai, J.-D. (2017). Water quality monitoring using wireless sensor networks: Current trends and future research directions. *ACM Transactions on Sensor Networks (TOSN),* **13,** 4.

Bázár, G., Kovacs, Z., Tanaka, M., Furukawa, A., Nagai, A., Osawa, M., Itakura, Y., Sugiyama, H. & Tsenkova, R. (2015). Water revealed as molecular mirror when measuring low concentrations of sugar with near infrared light. *Anal Chim Acta,* **896,** 52-62.

Britz, T. J., Sigge, G. O., Huisamen, N., Kikine, T., Ackermann, A., Lötter, M., Lamprecht, C. & Kidd, M. (2013). Fluctuations of indicator and index microbes as indication of pollution over three years in the Plankenburg and Eerste Rivers, Western Cape, South Africa. *Water SA,* **39**.

Burns, D. A. & Ciurczak, E. W. (2007). Handbook of near-infrared analysis. CRC press.

Flörke, M., Kynast, E., Bärlund, I., Eisner, S., Wimmer, F. & Alcamo, J. (2013). Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study. *Global Environmental Change,* **23,** 144-156.

Gowen, A. A., Marini, F., Tsuchisaka, Y., De Luca, S., Bevilacqua, M., O'Donnell, C., Downey, G. & Tsenkova, R. (2015). On the feasibility of near infrared spectroscopy to detect contaminants in water using single salt solutions as model systems. *Talanta,* **131,** 609-618.

Gowen, A. A., Tsenkova, R., Bruen, M. & O'donnell, C. (2012). Vibrational Spectroscopy for Analysis of Water for Human Use and in Aquatic Ecosystems. *Critical Reviews in Environmental Science and Technology,* **42,** 2546-2573.

Gowen, A. A., Tsuchisaka, Y., O'Donnell, C. & Tsenkova, R. (2011). Investigation of the potential of near infrared spectroscopy for the detection and quantification of pesticides in aqueous solution. *American Journal of Analytical Chemistry,* **2,** 53.

Gulati, M., Jacobs, I., Jooste, A., Naidoo, D. & Fakir, S. (2013). The water–energy–food security nexus: Challenges and opportunities for food security in South Africa. *Aquatic Procedia,* **1,** 150-164.

Hanjra, M. A. & Qureshi, M. E. (2010). Global water crisis and future food security in an era of climate change. *Food Policy,* **35,** 365-377.

Hoekstra, A. Y., Mekonnen, M. M., Chapagain, A. K., Mathews, R. E. & Richter, B. D. (2012). Global monthly water scarcity: blue water footprints versus blue water availability. *PLoS One,* **7,** e32688.

Kirby, R. M., Bartram, J. & Carr, R. (2003). Water in food production and processing: quantity and quality concerns. *Food control,* **14,** 283-299.

Kovacs, Z., Bazar, G., Oshima, M., Shigeoka, S., Tanaka, M., Furukawa, A., Nagai, A., Osawa, M., Itakura, Y. & Tsenkova, R. (2016). Water spectral pattern as holistic marker for water quality monitoring. *Talanta,* **147,** 598-608.

Molden, D. (2007). Water responses to urbanization. Springer.

Molden, D., Oweis, T., Steduto, P., Bindraban, P., Hanjra, M. A. & Kijne, J. (2010). Improving agricultural water productivity: Between optimism and caution. *Agricultural water management,* **97,** 528-535.

Munćan, J. S., Matija, L., Simić-Krstić, J. B., Nijemčević, S. S. & Koruga, Đ. L. (2014). Discrimination of mineral waters using near infrared spectroscopy and aquaphotomics. *Hemijska industrija,* **68,** 257-264.

Olivier, F. (2015). Evaluating the potential of ultraviolet irradiation for the disinfection of microbiologically polluted irrigation water. Stellenbosch : Stellenbosch University.

Siesler, H. W. (2006). Near-infrared spectroscopy : principles, instruments, applications. Weinheim, Weinheim : Wiley-VCH.

Tsenkova, R. & Gowen, A. (2011). NIR Spectroscopy: a tool for aquaphotomics. In: Pre-conference course, 15th International Conference on Near Infrared Spectroscopy. Pp. 1-22. Cape Town, South Africa.

Velíšek, J. (2014). The chemistry of food. Chichester, West Sussex, UK : Wiley Blackwell.

WHO (2011). Guidelines for drinking-water quality. *WHO chronicle,* **38,** 104-108.

WHO (2016). *Drinking-water.*[Internet document]. http://www.who.int/mediacentre/factsheets/fs391/en/. Accessed 16/04/2017.

WWF (2017). *Water Scarcity.*[Internet document]. https://www.worldwildlife.org/threats/water-scarcity. Accessed 16/04/2017.

Chapter 2

Literature review

2.1 Introduction

Water is a primary constituent of all life forms on earth and is used in a variety of human activities such as agriculture, industrial, household and recreational activities. The overall volume of water on Earth is approximately 1.4 billion km^3, of which only 35 million km^3 is freshwater (Dijkstra & de Roda Husman, 2014). Of this portion of freshwater, only 200 000 km^3 is available for human use and ecosystems (Robertson *et al.*, 2003; UNU-INWEH, 2013).

By the year 2050, it is expected that the population will increase by approximately 1.5 billion people, resulting in a global population of about 9 billion people (Hanjra & Qureshi, 2010). This increase in population will result in an increasing demand for water for agriculture, to meet the food production requirements. Globally, agriculture is one of the activities that require large quantities of water and uses approximately 80% of the available water (Molden, 2007). The demand for water is divided into different sectors, which includes agricultural, municipal and industrial uses. In South Africa, the total amount of water withdrawn per year is approximately 15.5 km^3, of which 62.5% is used for agricultural purposes, 27% is used for municipal use and 10.5% for industrial use (FAO, 2016).

Industrial activities, including food and energy production, compete for the requirement of freshwater, and each year their demand increases (Kirby *et al.*, 2003). The growing demand and decreasing availability of freshwater have a negative impact on its quality. Water scarcity is forcing manufacturing companies to attempt to optimise the use of water currently available. The waste products of food production and processing industries are an important factor in the water cycle since it reduces the water quality (Kirby *et al.*, 2003).

The quality of water is influenced by a variety of factors such as its biological and chemical composition (Gowen *et al.*, 2012). The intended use is determined by the water quality, which is monitored and undergoes various treatment processes to ensure that the quality meets the required standards and regulations. Water quality is a complex problem since numerous chemical and microbiological analyses are

required to provide an indication of the overall quality (Gowen *et al.*, 2015). Currently, water resources are greatly contaminated because of anthropogenic activities such as deforestation, pollution and overexploitation of natural resources (Gowen *et al.,* 2012). Natural effects such as algal blooms, decomposing organic matter and contact with toxic minerals (arsenic, mercury, lead) also have an impact on water quality.

Water quality is characterised by different parameters such as the concentration of bacteria, organic and inorganic pollution as well as indicators which include pH, total dissolved solids (TDS) and conductivity (Gowen *et al.*, 2012). The biological and physico-chemical characteristics of water are used to classify it according to its intended use or purpose, such as drinking water, irrigation and industrial uses (Flörke *et al.*, 2013; Velíšek, 2014). The World Health Organisation has established guidelines, indicating the quality parameters required for water to be considered safe for human consumption and food production (WHO, 2011). However, drinking water quality standards differ from country to country.

Aquaphotomics is a field of study based on the interaction of water and light (Tsenkova, 2009). This interaction provides information about how water interacts with other molecules within biological systems (Gowen *et al.*, 2015; Headrick *et al.*, 2005; Kovacs *et al.*, 2016). The absorbance pattern of water changes depending on the physical or chemical variations in the environment such as temperature changes. These changes also result in changes in the vibrations of water molecules, which can be detected with near-infrared spectroscopy.

Near-infrared (NIR) spectroscopy is a non-destructive analytical technique, which enables the simultaneous measurement of qualitative and quantitative parameters of a range of different products (Pasquini, 2018). The NIR region of the electromagnetic spectrum (780 nm – 2500 nm) is widely used to detect compounds, predict concentrations of solutions and monitor molecular shifts in biological environments (Segtnan *et al.*, 2001; Siesler, 2006; Tsenkova, 2005). NIR spectroscopy has shown that it has the potential to be a cost-effective and faster screening technique for water quality (Gowen *et al.*, 2012).

This literature review addresses the use of near-infrared spectroscopy combined with aquaphotomics as a screening method for water quality. The following topics will be addressed: water resources, water quality, the structure of liquid water, near-infrared spectroscopy, aquaphotomics and the analysis of spectral data.

2.2 Water resources

South Africa has a typical rainfall of around 450 mm per year (DEAT, 2006). This indicates that South Africa is water-stressed since the typical rainfall is well below the global average per year of approximately 860 mm (DWA, 2013; Pitman, 2011). South Africa's water availability faces three main challenges; (1) the spatial and seasonal rainfall patterns, (2) relatively low water levels in rivers most of the time and (3) the location of major urban and industrial developments (DEAT, 2006).

In South Africa, the water requirements of most of the urban areas are met using water from surface resources, such as dams and rivers (DEAT, 2006). Some rural and most arid areas' water requirements are met using groundwater sources (DEAT, 2006). Groundwater is also an important water source in Africa since it makes up the majority of the water supplies (MacDonald *et al.*, 2012). Pollution and exploitation of groundwater sources are major problems in South Africa (DEAT, 2006).

In agriculture, water is important in the production of fresh produce that is nutritious and safe for human consumption. Almost 40 % of the food demands are supplied by irrigated agriculture, which accounts for only 19 % of available cropland (Molden, 2007). The increasing demand for water for non-agricultural uses are threatening food security since the importance of water for agricultural uses are overlooked (Hanjra & Qureshi, 2010; Molden *et al.*, 2010).

2.3 Water quality

Water quality can be described by a variety of factors, which is separated into the subsequent groups: an indicator, microorganisms, organic and inorganic. Indicator factors are colour, conductivity, pH, odour and taste (Gowen *et al.*, 2012). Chemicals such as chlorinated alkanes, benzenes and ethenes are classified as organic parameters, while heavy metals are classified as an inorganic parameter. The major pollutants in water sources, which pose a health risk to humans, are microorganisms and excess dissolved solids (Sampathkumar *et al.*, 2010).

The standard analysis techniques of water quality require intensive sampling plans, sample preparation and require manual inputs which prohibit their integration into continuous screening techniques (Gowen *et al.*, 2015). The turnaround times for standard analysis methods are so slow that most water is consumed or used before the results are available (Gowen *et al.*, 2012).

The physico-chemical characteristics of water affect the functioning of an aquatic environment and its capability to support life forms. When treating water for human consumption and agricultural uses, colour, pH, total dissolved solids and electrical conductivity are parameters used as an indicator of the quality. In terms of aesthetic consideration, the colour of the water is a key aspect of drinking water quality (Sampathkumar *et al.*, 2010). The clarity of the water is an important parameter when treating water for human consumption (Rice *et al.*, 2012). Cloudiness is caused by a variety of particles and is referred to as turbidity (Rahmanian *et al.*, 2015). Water turbidity is associated with the amount of disease-causing microorganisms and the presence of inorganic ions like manganese (DWAF, 1996b; Rahmanian *et al.*, 2015).

The pH of water not only affects the aquatic life of water, but also the physico-chemical properties (Morrison *et al.*, 2001). The solubility and bio-availability of many plant nutrients are dependent on the pH of water (DWAF, 1996a). The toxicity of water is related to the pH, as a decrease in the pH levels can result in an increase in the solubility of elements such as lead, aluminium, boron, copper, cadmium, manganese and iron (Morrison *et al.*, 2001). Water with a low pH can be detrimental to metal pipes and pumping systems as it is corrosive (Rahmanian *et al.*, 2015).

Solids, both dissolved and suspended, can affect the quality of the water unfavourably (APHA, 2012). Dissolved solids refer to the portion of solids which passes through a 2.0 µm filter under specified conditions, while suspended solids are the portion that remains on the filter (APHA, 2012). Total dissolved solids are an indicator of the amount of organic and inorganic matter found in water (Rahmanian *et al.*, 2015). Water with high dissolved solids content is usually of inferior palatability and is unfavourable to the consumer. Domestic sewage can increase the concentration of dissolved salts, which is used as an indicator of the electrical conductivity of the water (Morrison *et al.*, 2001).

The ability of water to carry an electric current is referred to as electrical conductivity, which relies on the concentration, mobility and valence of ions found in the water (DWAF, 1996a; Rice *et al.*, 2012). Dissolved solids such as calcium, magnesium and chloride carry electrical current through water, therefore the mineral content of the water will influence the electrical conductivity (Rahmanian *et al.*, 2015). The concentration of total dissolved solids is directly proportional to the electrical conductivity of the water.

2.3.1 Drinking and irrigation water quality

Potable or drinking water is deemed safe for human consumption and should not pose a risk to human health (Storey *et al.*, 2011). A variety of water types is categorised under potable water. These include bottled water and tap water. In South Africa, bottled water is divided into three categories: natural water, water defined by the origin, and prepared water.

Natural water originates from underground aquifers and is bottled at the source (Dijkstra & de Roda Husman, 2014). Natural water is not allowed to undergo any treatment that will alter the mineral composition. However, natural water can be decanted or filtered to remove unstable constituents (Dijkstra & de Roda Husman, 2014). Spring and mineral water are classified as natural water. Rain, river, stream, snow, mist, glacier and seawater is classified as water defined by origin (Foodstuffs, 2010). These types of water require antimicrobial treatments to ensure that it is fit for human consumption and unsuitable constituents such as iron and manganese can be removed from (Foodstuffs, 2010). Prepared water is from sources such as municipal surface and groundwater. Prepared water requires purification treatments to ensure that the water is safe for human consumption and to meet the standards set for bottled water (Foodstuffs, 2010).

Water used in agricultural practices is sourced from a variety of sources such as large reservoirs, farm dams, rivers, groundwater, municipal supplies and industrial effluent (DWAF, 1996a). The quality of water sourced from ground and surface water sources may vary depending on seasonal conditions such as floods or droughts (DWAF, 1996a). A growing concern in developing countries is the contamination of water sources with human faecal matter (Barnes & Taylor, 2004). Inadequate sewage treatment systems, the release of unprocessed water and informal settlements close to rivers are the main sources of contamination of surface water sources (Barnes & Taylor, 2004).

2.4 The structure of liquid water

Despite the significance and abundance of water in the environment, its structure is still not fully understood (Gowen *et al.*, 2013). At a molecular level, the structure of water is not homogenous but consists of different configurations of water clusters (dimers, trimers, solvation shells) of varying concentrations (Munćan, 2012). Water

molecules cluster together and are held together by hydrogen bonds. These clusters can range in size (**Figure 2.1**) and the simplest water molecular species is a dimer, which consists of two water molecules (Pang, 2014). Solvation shells occur when water molecules cluster around another molecule to which it is attracted to, such as sodium ions. The structure of liquid water is due to the hydrogen bonding network formed between water clusters and the relative orientation and mobility of water molecules are influenced by the neighbouring water molecules (Damodaran & Parkin, 2017).

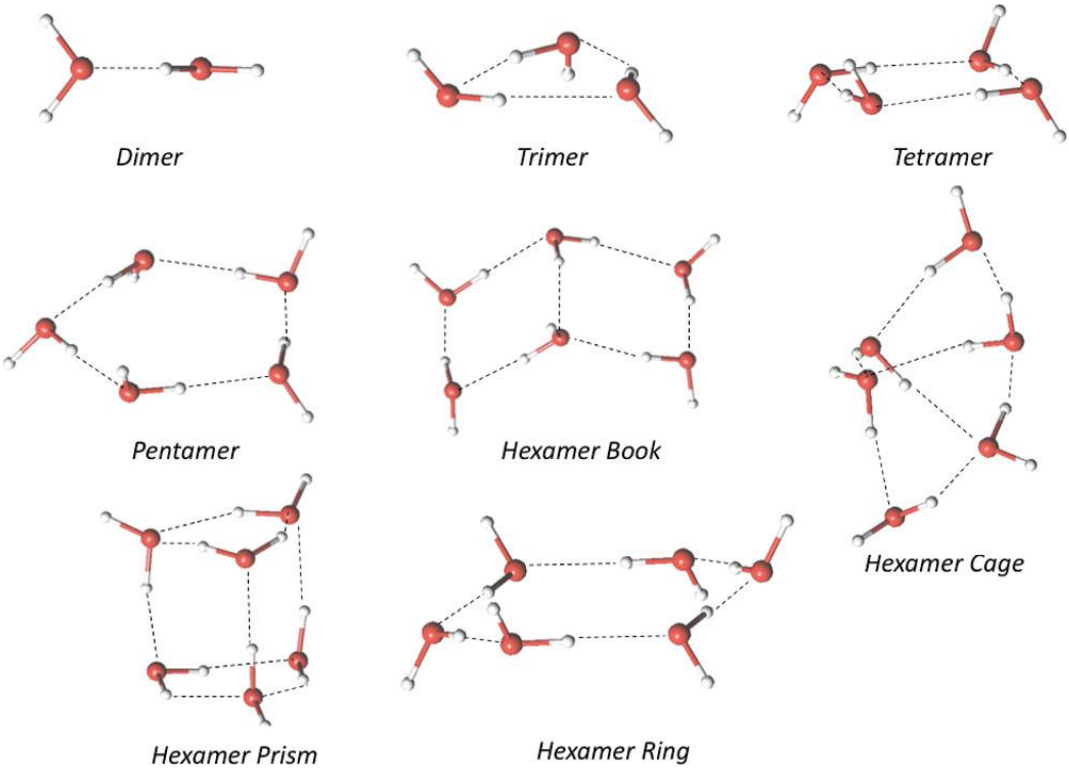

Figure 2.1 Water clusters with different numbers of water molecules and conformational structures (Segarra-Martí *et al.*, 2012).

Over the years, water has been the focus of many studies due to its unusual behaviour and role in many biological and chemical processes (Giangiacomo, 2006; Gowen *et al.*, 2013; Munćan, 2012; Tanaka *et al.*, 1997; Xantheas, 2000). However, these studies have not been able to provide a comprehensive model that can describe the unusual behaviour of water (Xantheas, 2000). The unique behaviour of water, the unusually high boiling and melting point temperatures, is mainly as a result of the hydrogen-bonded (H-bond) network of water molecules (Hasted, 1972; Myneni *et al.*, 2002). Hydrogen bonds occur when a hydrogen atom is bonded to two or more atoms

and acts as a bridge between atoms (Chalmers & Griffiths, 2002). Two types of hydrogen bonds, intermolecular and intramolecular hydrogen bonds, can occur in water (Chalmers & Griffiths, 2002). Hydrogen-bonding occurring between different molecules are referred to as intermolecular, while intramolecular hydrogen-bonding occurs within a molecule between one proton donor and a proton acceptor (Chalmers & Griffiths, 2002). The dipole nature of water molecules results in interactions with other molecules through hydrogen bonds (Velíšek, 2014).

The structure of pure water is a multifaceted and challenging problem, with many theories being suggested (Reid & Fennema, 2007). These theories are incomplete and oversimplified. However, two general structural models, mixture and continuum, have been used to explain the unique behaviour of water (Falk & Ford, 1966; Libnau *et al.*, 1994; Segtnan *et al.*, 2001). The mixture model can be described by an equilibrium mixture of distinct water species with intermolecular hydrogen-bonds being temporarily concentrated in large clusters of water molecules, which occur in a dynamic equilibrium with other more complex species (Reid & Fennema, 2007; Segtnan *et al.*, 2001). The continuum model consists of water molecules which are involved in nearly complete hydrogen bonds (Segtnan *et al.*, 2001). The hydrogen bonds are uniformly distributed throughout the water (Segtnan *et al.*, 2001).

Ions present in a solution alters the hydrogen bonding of water, which can be detected with NIR spectroscopy (Gowen *et al.*, 2013). The hydrogen bond network of the water matrix is influenced by the addition of cations and anions (Vero *et al.*, 2010). This is a result of the charge-dipole interaction between the water molecules and the ions. The effect of the ion depends on the type of ion (Omta *et al.*, 2003). Structure-maker ions increase the strength of the hydrogen bond network, while structure-breaker ions weaken the hydrogen bond network (Omta *et al.*, 2003) The effects of a mixture of inorganic binary salts on the spectra of water can be seen with NIR spectroscopy, due to the specific effect of the cation/anion O-H interactions (Frost & Molt, 1997). It is believed that when a water molecule is positioned next to a cation, it would attract the oxygen atom while repelling the hydrogen atoms within the molecule (Gowen *et al.*, 2013). This results in the weakening of the O-H bond and these changes can be seen in the spectrum of water (Gowen *et al.*, 2013). A similar reaction will occur if an anion is next to a water molecule. However, the anion will attract the hydrogen atoms instead of the oxygen atom.

Changes in spectra due to ions are different from structurally induced changes (Bunzl, 1967). Structurally related variations in the spectra are primarily due to the valance and charge distribution, electron attraction and the ion sizes in the solution (Gowen *et al.*, 2013). These can be observed with a change in the pH of identical ionic salt solutions at different concentrations (Gowen *et al.,* 2013). The changes in the water spectrum can be analysed using aquaphotomics, a method that identifies the unique fingerprint of water bands due to specific perturbations (Tsenkova *et al.,* 2018).

2.5 Near Infrared spectroscopy

Near-infrared (NIR) spectroscopy is a vibrational spectroscopy technique that is based on the interaction of electromagnetic radiation and the different vibrational modes of the covalent bonds in a molecule (Gowen *et al.*, 2012; Osborne *et al.*, 1993) The NIR electromagnetic spectrum covers the range from 780 - 2500 nm. The overtones and combination bands of the fundamental C-H, O-H, and N-H vibrations are found in this wavelength range (Gowen *et al.*, 2012; Lin *et al.*, 2009; Siesler *et al.*, 2008). Overtones are a result of a molecule being excited from the ground state, lowest energy state, to the second or higher vibrational energy level (Burns & Ciurczak, 2007). Combination bands occur when more than one vibrational mode is excited at once (Siesler *et al.*, 2008).

Vibrational energy influences the C-H, O-H and N-H when irradiated by the NIR frequencies (Burns & Ciurczak, 2007; Siesler, 2006). The energy changes result in stretching and bending vibrations of the bonds within molecules. A continuous change in the bond length between two atoms is referred to as stretching vibrations while bending vibrations are due to the change in bond angles between two atoms. The absorption of NIR energy occurs when the vibration of the frequency matches that of the molecular bond within the sample being scanned (Burns & Ciurczak, 2007).

Due to NIR spectral bands consisting of overlapping combination and overtone bands, it can sometimes be difficult to identify chemical species (Burns & Ciurczak, 2007; Siesler, 2006). Since numerous compounds absorb NIR energy, extracting usable information from the spectra for quantitative methods are complicated. Thus, multivariate data analysis methods are required to extract the appropriate information from spectra.

2.5.1 Instrumentation

A NIR spectrophotometer has three main components, a light source, a wavelength isolator and a detector (Ozaki *et al.*, 2006). Fourier transform near-infrared (FT-NIR) spectrophotometers contain an interferometer and are categorised as a multiplex instrument (Pierna *et al.*, 2018). An FT-NIR spectrophotometer produces an interferogram, which is a complex signal that contains all the frequencies that make up the infrared spectrum (Pavia *et al.*, 2008). A graphic illustration of an FT-NIR instrument is shown in **Figure 2.2**. A tungsten halogen lamp (1) produces light that is directed to the first polarizer (2). The polarised light then passes through a double refracting block (3), which splits the light into two orthogonally polarised components with a static phase shift. Two double refracting wedges are placed after the refracting block, the first refracting wedge is stationary (4), while the second refracting wedge (5) constantly moves backwards and forwards. This results in an ongoing phase shift between the light beams. The phase-shifted beams are merged back into one light beam with intensity variation at the second polarizer (6). The light beam then passes through the sample (7) and is detected by the interferogram detector (8), producing an interferogram (9) which is converted to a spectrum by Fourier transforms (Siesler *et al.*, 2008)

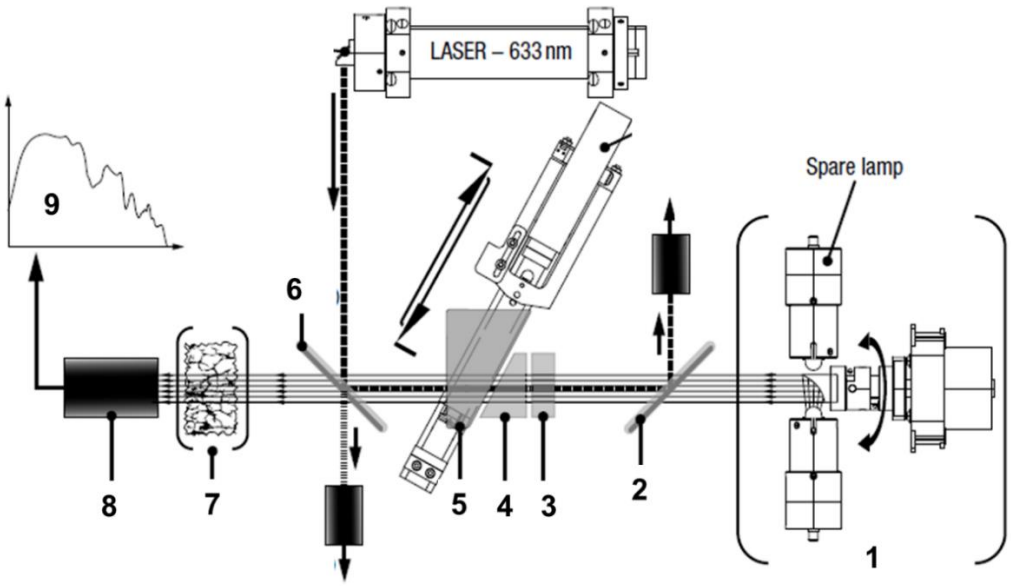

Figure 2.2 Schematic diagram of an FT-NIR instrument with a polarisation interferometer (BÜCHI Labortechnik AG, 2018).

2.6 Aquaphotomics

Aquaphotomics is a scientific field introduced by Professor Roumiana Tsenkova in 2005. This novel method describes the interaction between water and light at each frequency of the NIR electromagnetic spectrum (Kinoshita *et al.*, 2012b; Tsenkova, 2009; Tsenkova & Gowen, 2011). The NIR spectrum is considered an excellent technique for the observation of water, providing information about the molecular structure of water molecules (Tsenkova, 2009). Aquaphotomics provides the possibility for the extraction of information hidden within the NIR spectrum of water, which is not possible with other methods (Bázár *et al.*, 2015).

Over the years, numerous bands of water species have been identified in the infrared range by various studies (Chandler, 2002; Czarnik-Matusewicz *et al.*, 1999; Luck & Schiöberg, 1979; Šašić *et al.*, 2002; Segtnan *et al.*, 2001). In aquaphotomics, these bands are referred to as water absorbance bands (WABS). The spectral database of water absorbance bands of the first (1300-1600 nm), second (610 – 870 nm) and third overtones (870 – 1070 nm) of water consists of more than 500 bands (Tsenkova, 2009; Tsenkova *et al.*, 2015). To date, 12 characteristic water absorbance bands have been identified that occurs in the first overtone of water, regardless of the perturbation observed. These wavelength ranges are referred to as Water matrix coordinates (WAMACS) and are shown in **Table 2.1** with their corresponding assignments for each water matrix coordinate.

Changes at specific water absorbance bands, due to a perturbation, occurring repeatedly throughout the data analysis process, are referred to as activated water absorbance bands (Tsenkova *et al.*, 2018). Activated water absorbance bands are used to describe the water spectral pattern of a system. Water is easily influenced by various factors, resulting in spectral changes. The changes in the water spectral patterns provide information related to the structure and dynamics of the molecular structure of the respective system (Tsenkova, 2009).

Table 2.1 Characteristic water absorbance bands identified

WAMACS	Wavelength range (nm)	Assignment	Reference
C 1	1336 – 1348	v_3	(Siesler *et al.*, 2008)
C 2	1360 – 1366	O-H stretch	(Xantheas, 1995)
C 3	1370 – 1376	$v_1 + v_2$	(Siesler, 2006)
C 4	1380 – 1388	O-H stretch	(Xantheas, 1995)
C 5	1398 – 1418	S_0	(Segtnan et al., 2001)
C 6	1421 – 1430	Water hydration	(Tsenkova, 2009; Williams, 2009)
C 7	1432 – 1444	S_1	(Cattaneo *et al.*, 2009; Siesler, 2006),
C 8	1448 – 1454	$v_2 + v_3$	(Cattaneo *et al.*, 2009; Siesler, 2006)
C 9	1458 – 1468	S_2	(Franks, 1973)
C 10	1472 – 1482	S_3	(Franks, 1973; Siesler, 2006)
C 11	1482 – 1495	S_4	(Tsenkova, 2009)
C 12	1506 - 1516	v_2	(Siesler, 2006)

v_1 - symmetric stretching of the first overtone of water, v_2 - bending of the first overtone of water, v_3 - asymmetric stretching of the first overtone of water, S_0 - free water and free OH or trapped water, S_1 - water molecules with one intermolecular hydrogen bond (dimer), S_2 - water molecules with two intermolecular hydrogen bonds (trimer), S_3 - water molecules with three intermolecular hydrogen bonds (tetramer), S_4 - water molecules with four intermolecular hydrogen bonds (pentamer).

Water spectral patterns are visualised in the form of radar plots, which are referred to as aquagrams. The axes originating from the centre of the aquagram displays the averaged normalised absorbance values at numerous water bands and the absorbance values of the WAMAC's are placed on the relevant radial axes (Bázár *et al.*, 2016). An example of an aquagram displaying water spectral patterns of different types of honey and high fructose corn syrup are shown in **Figure 2.3**.

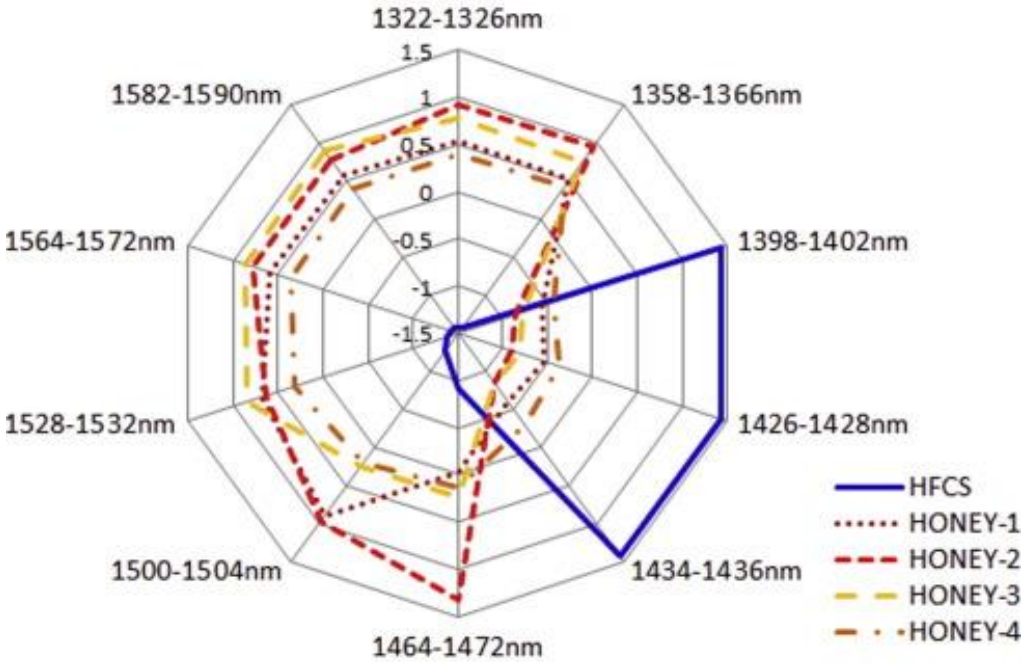

Figure 2.3 Aquagram of pure honey and high fructose corn syrup (Bázár *et al.*, 2016).

The effect of temperature on the water spectrum has been the focus of many studies over the years. The relationship between the hydrogen bonding and spectral changes as a result temperature was investigated by Maeda *et al.* (1997). The study focused on the temperature-dependent spectral variations in the 1100 – 1800 nm wavelength range. Spectra were acquired by scanning water at temperatures ranging from 5 – 85 ˚C. Spectral changes were identified using principal component regression and partial least squares. The study indicated is a relationship between the temperature and hydrogen bonding of water. The differences in the spectra are due to the size of the different water molecule clusters. As the temperature was increased or decreased, changes in the water clusters were observed as spectral changes.

Munćan *et al.* (2014) used aquaphotomics to differentiate between brands of mineral water. The samples were obtained from commercial outlets and the mineral content data was obtained either from the label or the manufacturer. Each sample contained sodium, potassium, calcium, magnesium, chlorine and sulphate ions of varying concentrations. The different mineral water samples each had a unique WASP as a result of the differences in the mineral composition and water molecular species confirmation (Munćan *et al.*, 2014). The differences in the WASPs were used to distinguish between the water samples.

Gowen *et al.* (2015) investigated the use of aquaphotomics as a possible detection method for low concentration contaminants. The authors used a concentration range of 0.002 – 0.1 mol L^{-1} of sodium chloride, potassium chloride, magnesium chloride and aluminium chloride to determine the effect that different salts will have on the spectral pattern of water. They showed that it is possible to predict salt concentration using partial PLS regression. This indicates that aquaphotomics has the potential to be used as a method for the detection of water contamination. In a similar study, Bázár *et al.* (2015) investigated millimolar concentration (0.02 – 100 mM) solutions of mono- and disaccharides to determine the effect of a range of concentrations will have on the water spectrum. The type and concentration of solute will affect the hydrogen bonding of water differently, and these differences can be seen in the NIR spectrum. Glucose, fructose, sucrose and lactose solutions with concentration ranges of 0.02 – 100 mM were analysed to determine the limit of detection of each sugar. The study concluded that it is possible to quantify the mono- and disaccharide solutions at millimolar concentrations.

A study by Kovacs *et al.* (2016) monitored the changes in the water spectral pattern with the addition of sodium chloride, lactose monohydrate and acetic acid solutions at millimolar concentrations (1mM – 100 mM) as possible water contaminants. The structural changes of the water clusters were monitored to determine the lowest level of detection of the three contaminants. The study proved that changes in the water molecular structure as a result of variation in the millimolar concentration of acetic acid, lactose monohydrate and sodium chloride can be detected with near-infrared spectroscopy.

Cattaneo *et al.* (2011) studied the influence of filtration processes on the structure of water. Pure water samples were analysed before and after the filtration process. Spectral variation was detected in the 1333 – 1538 nm wavelength range. These variations are due to changes in the water clusters during the filtration process. The authors concluded that filtration processes influence the hydrogen bond network of water. However further research is required to determine how the different variables in the filtration process influence the hydrogen bond network of water. A summary of aquaphotomics applications is presented in **Table 2.2**.

Table 2.2 Summary of aquaphotomics applications

Application of aquaphotomics approaches	Perturbation	Chemometric techniques	Wavelength range (nm)	References
Sugar detection	Concentration of different sugars	PCA, LDA, PLS	1100 – 1800	(Bázár *et al.*, 2015)
Salts detection	Temperature and inorganic salts	PCA, EFA, MCR-ALS	1300 – 1600	(Gowen *et al.*, 2013)
Heavy metal detection	Metal ions	PLS	680 – 1090, 1110 – 1800	(Putra *et al.*, 2012)
Water filtration system	Filtration process	PCA	1100 – 1800	(Cattaneo *et al.*, 2011)
Pesticide detection	pesticides	PCA, PLS, PLS-DA	400 – 2500	(Gowen *et al.*, 2011)
Mineral water differentiation	Different brands of mineral waters	HCA, SIMCA	1300 – 1600	(Munćan *et al.*, 2014)
Water Monitoring	The concentration of acetic acid, Lactose monohydrate, sodium chloride	PCA-QCC	600 – 1000, 1300 – 1600	(Kovacs *et al.*, 2016)
Honey	Adulteration with high fructose corn syrup	PCA, PCR, PLS	1300 – 1800	(Bázár *et al.*, 2016)
Soybean virus detection	Soybean mosaic virus	PCA, SIMCA	730 – 1025	(Jinendra *et al.*, 2010)

Application of aquaphotomics approaches	Perturbation	Chemometric techniques	Wavelength range (nm)	References
Panda urine	Oestrus concentration	PLS, HCA	1300 – 1600	(Kinoshita *et al.*, 2012a)
Detection of bacteria and extracellular metabolites	*Escherichia coli* and *Staphylococcus aureus*	PLS	680 – 1090, 1100 – 1800	(Nakakimura *et al.*, 2012)
Milk	Somatic cell count	PCA, PLS, SIMCA	1100 – 2500	(Tsenkova, 2006)
Bacteria growth monitoring	The growth of *Lactobacillus spp*	PCA, LDA, PLS, SIMCA	1100 – 1300	(Slavchev *et al.*, 2017)
Salt	Salt concentration	PLS, PCA	1000 – 2500	(Gowen *et al.*, 2015)
Soil quality	Carbon and nitrogen content determination	PCA, PLS	908 – 1676	(Mura *et al.*, 2019)
Effect of glucose on water structure	Addition of glucose	CWT	850 – 2500 nm	(Cui *et al.*, 2016)

PCA – Principal Component Analysis, LDA – Linear Discriminant Analysis, PLS – Partial Least Squares, EFA – Evolving Factor Analysis, MCR-ALS – Multivariate Curve Resolution – Alternating Least Squares, SIMCA – Soft Independent Modelling of Class Analogy, HCA – Hierarchical Cluster Analysis, PCR – Principal Component Regression, PLS-DA - Partial Least Square – Discriminant Analysis, PCA-QCC – Principal Component Analysis – Quality Control Chart, CWT – Continuous wavelet transform

2.7 Near-infrared spectral data analysis

The water spectrum is very complex, and some perturbations cause subtle changes in the spectra, which can easily go unnoticed (Tsenkova *et al.*, 2018). Therefore, aquaphotomics requires multiple data analysis steps to identify activated water absorbance bands, which are used to explain differences between aqueous systems (Tsenkova *et al.*, 2018). Consequently, pre-processing and multivariate data analysis is essential for the extraction of information from spectral data.

2.7.1 Pre-processing

Pre-processing aims to enhance relevant information and reduce or remove the unwanted influences in the spectral data (Rinnan *et al.*, 2009b; Tsenkova *et al.*, 2018). Pre-processing techniques include mean centering, smoothing, derivatives and scatter correction. Mean centering is a technique that involves the subtraction of the mean spectrum from the entire database (Agelet & Hurburgh Jr, 2010; Beebe, 1998). This technique is a pre-processing method most commonly used with principal component analysis (Agelet & Hurburgh Jr, 2010).

Smoothing techniques are used for the removal of noise or background information from spectra, while also increasing the signal-to-noise ratio (Agelet & Hurburgh Jr, 2010; Beebe, 1998). In aquaphotomics, the most common technique to reduce noise is the Savitzky-Golay smoothing filter (Savitzky & Golay, 1964). Smoothing techniques produce an optimum signal-to-noise ratio for spectral data when combined with other pre-processing techniques (Wang *et al.*, 2006). Equation 2.1 presents the Savitzky-Golay smoothing filter algorithm

$$x_j^* = \frac{1}{N}\sum_{h=-k}^{k} c_h x_j + h \qquad \text{(eq. 2.1)}$$

Where:

x_j^*- the new value of the smoothed curve or derivative

N – the normalising constant

k – the number of adjacent values at each side of j

c_j – the coefficient that depends on the polynomial degree used

Spectral derivatives are used to extract relevant information by removing the additive and multiplicative effects from the data (Rinnan *et al.*, 2009a). The first derivative removes only

the baseline, while the second derivative removes the baseline and the linear trend (Rinnan *et al.*, 2009a). In the NIR spectra of aqueous solutions, derivatives can separate overlapping bands and remove baseline variations (Tsenkova *et al.*, 2018). Derivatives also have a downside, it may result in the loss of the original shape of the spectral data and a reduction in the signal-to-noise ratio (Tsenkova *et al.*, 2018).

Scatter correction was designed to reduce any variability between samples due to multiplicative scatter (Rinnan *et al.*, 2009a). Standard normal variant (SNV) and Multiplicative scatter correction (MSC) are the most widely used scatter correction methods (Agelet & Hurburgh Jr, 2010). Multiplicative interferences and particle size scattering are removed with the use of SNV (Barnes *et al.*, 1989). MSC creates a line of best fit for each spectrum to a reference spectrum and then the slope of best fit is used to adjust the spectra (Geladi *et al.*, 1985). This process removes additive and multiplicative differences. The algorithms for SNV and MSC is as follows (Varmuza & Filzmoser, 2011), equations 2.2 – 2.4.

SNV:

$$x_{corr} = \frac{x_{org} - a_0}{a_1} \quad \text{(eq. 2.2)}$$

MSC:

$$x_{org} = b_0 + b_{ref,1}.x_{ref} + e \quad \text{(eq. 2.3)}$$

$$x_{corr} = \frac{x_{org} - b_0}{b_{ref,1}} \quad \text{(eq. 2.4)}$$

Where:

x_{corr} = *corrected spectra*

x_{org} = *original spectra*

x_{ref} = *reference spectra*

a_0 = *the average spectrum*

a_1 = *Standard deviation spectrum*

b_0, $b_{ref,1}$ = *scalar parameters*

e = *un-modelled part of* x_{org}

2.7.2 Multivariate data analysis

Multivariate data analysis is used to extract meaningful information from the spectral data (Wold, 1995). The extraction of relevant information from spectral data for the development of models and prediction of specific properties of unknown samples can be achieved by a number of multivariate data analysis techniques (Manley & Baeten, 2018). Principal component analysis, partial least squares and partial least squares are just three of the many multivariate data analysis techniques existing.

Principal component analysis (PCA) is one of the most popular unsupervised techniques (Cowe & McNicol, 1985). PCA is usually applied before any other multivariate data analysis technique, to identify the main sources of variability and distribution of elements. The variables in a dataset are reduced to principal components (PC), composed of scores and loadings using equation 2.5. Each PC explains a specific amount of the total variance within the dataset, in decreasing order.

$$X = TP' + E \qquad \text{(eq. 2.5)}$$

Where:

X – *represents a matrix*

T - *Score matrix*

P - *loadings matrix*

E - *residual matrix*

Partial least squares (PLS) is a linear regression method, which takes into consideration the explanatory and dependent variables (Varmuza & Filzmoser, 2011). This method aims to find a set of components which provides good linear models for all the Y variables (Garthwaite, 1994). PLS decomposes the X and Y matrices into latent structures in an iterative process (Brereton, 2007). The PLS algorithms are presented in equations 2.6 and 2.7.

$$X = TP^T + E \qquad \text{(eq. 2.6)}$$

$$Y = UQ^T + F \qquad \text{(eq. 2.7)}$$

X – *predictor variables matrix*

Y – *response variable matrix*

T , U – *the score matrices of X and Y, respectively*

P , Q – *the loading matrices of X and Y, respectively*

E , F – *The residual matrices*

Partial least squares discriminant analysis (PLS-DA) is a technique that is a combination between the properties of PLS regression and the discrimination power of a classification method (Ballabio & Consonni, 2013). This technique uses dummy variables, equal to one or zero, instead of measured y-data (Ballabio & Todeschini, 2009; Manley & Baeten, 2018). PLS-DA aims to classify spectra based on its group membership. The PLS-DA algorithm is presented in equation 2.8.

$$y = xb + f \qquad \text{(eq. 2.8)}$$

Where:

$y =$ *dummy matrix*

$x =$ *data matrix*

$b =$ *regression coefficient matrix*

$f =$ *residual matrix*

The loading plots of a PCA and regression vectors of PLS are used to identify any activated water absorbance bands in the observed system (Tsenkova, 2009). The activated water absorbance bands, which occur repeatedly throughout all the data analysis techniques are used to construct the aquagram (Tsenkova *et al.*, 2018).

An aquagram is a way of visualising and comparing the water spectral pattern of a system (Tsenkova *et al.*, 2018). The activated water absorbance bands identified through the data analysis steps are used to construct the aquagram. SNV or MSC is applied to the raw spectral data of the water absorbance bands identified, producing normalised absorbance values that are plotted on the axes originating from the centre. The wavelength of the specific water absorbance bands (or WAMACS) is plotted on the radial axes. The normalisation of an aquagram is computed using equation 2.9.

$$A'_\lambda = \frac{A_\lambda - \mu_\lambda}{\sigma_\lambda} \qquad \text{(eq. 2.9)}$$

Where:

A'_λ - The normalised absorbance values displayed on the aquagram

A_λ - The absorbance values after SNV or MSC

μ_λ- The average spectra of all the samples examined

σ_λ- The standard deviation of all the spectra of the samples

The number of axes originating from the centre of the aquagram depends on the system observed and the number of activated water absorbance bands identified during the data analysis process. The WAMACs of the specific system is placed on the angular axis of the aquagram at each spoke originating from the centre. The radial axis of the aquagram presents the absorbance range of the specific set of WAMACs.

2.8 Conclusion

With the increasing demand for water and water shortages in South Africa, the water treatment industry would benefit from a rapid non-destructive method for water quality determination. As water forms the basis of aquaphotomics, it has the potential to be a technique suited for water quality analysis. However, very little work has been done on water quality using aquaphotomics. Aquaphotomics provides information which usually remains hidden with conventional data analysis techniques.

The water spectrum provides a large amount of information about the water clusters and the effect of different solutes on the water molecules. However, more research is needed to understand how the combination of different physico-chemical parameters such as pH, TDS, TSS and turbidity will have on the water spectrum. With a better understanding of how physico-chemical properties of water influence the water spectrum, aquaphotomics has the potential to be applied in developing a rapid and accurate screening method for water quality.

2.9 References

Agelet, L. E. & Hurburgh Jr, C. R. (2010). A tutorial on near infrared spectroscopy and its calibration. *Critical Reviews in Analytical Chemistry,* **40,** 246-260.

APHA (2012). Standard methods for the examination of water and wastewater. Washington, D.C. : American Public Health Association Press.

Ballabio, D. & Consonni, V. (2013). Classification tools in chemistry. Part 1: linear models. PLS-DA. *Analytical Methods,* **5,** 3790-3798.

Ballabio, D. & Todeschini, R. (2009). Multivariate classification for qualitative analysis. In: Infrared spectroscopy for food quality analysis and control, ed. DW Sun. Pp. 83-104. United Kingdom: Academic Press.

Barnes, J. & Taylor, M. (2004). Health risks assessment in connection with the use of microbiologically contaminated source waters for irrigation. *Water Research Report*, 04.

Barnes, R., Dhanoa, M. S. & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy,* **43,** 772-777.

Bázár, G., Kovacs, Z., Tanaka, M., Furukawa, A., Nagai, A., Osawa, M., Itakura, Y., Sugiyama, H. & Tsenkova, R. (2015). Water revealed as molecular mirror when measuring low concentrations of sugar with near infrared light. *Anal Chim Acta,* **896,** 52-62.

Bázár, G., Romvári, R., Szabó, A., Somogyi, T., Éles, V. & Tsenkova, R. (2016). NIR detection of honey adulteration reveals differences in water spectral pattern. *Food chemistry,* **194,** 873-880.

Beebe, K. R. (1998). Chemometrics : a practical guide. New York, N.Y., New York, N.Y. : Wiley.

Brereton, R. G. (2007). Applied chemometrics for scientists. John Wiley & Sons.

BÜCHI Labortechnik AG (2018). NIRFlex N-500: Technical data sheet. Switzerland: BÜCHI Labortechnik AG.

Bunzl, K. W. (1967). Near-infrared spectra of aqueous solutions of some tetra-n-alkylammonium bromides. *The Journal of Physical Chemistry,* **71,** 1358-1363.

Burns, D. A. & Ciurczak, E. W. (2007). Handbook of near-infrared analysis. CRC press.

Cattaneo, T., Stefania, V., Elena, N. & Vittorio, E. (2011). Influence of filtration processes on aqueous nanostructures by NIR spectroscopy. *Journal of Chemistry and Chemical Engineering,* **5,** 1046-1052.

Cattaneo, T. M., Cabassi, G., Profaizer, M. & Giangiacomo, R. (2009). Contribution of light scattering to near infrared absorption in milk. *Journal of Near Infrared Spectroscopy,* **17,** 337-343.

Chalmers, J. M. & Griffiths, P. R. (2002). Handbook of vibrational spectroscopy. New York, New York : J. Wiley.

Chandler, D. (2002). Hydrophobicity: Two faces of water. *Nature,* **417,** 491.

Cowe, I. A. & McNicol, J. W. (1985). The Use of Principal Components in the Analysis of Near-Infrared Spectra. *Applied Spectroscopy,* **39,** 257-266.

Cui, X., Cai, W. & Shao, X. (2016). Glucose induced variation of water structure from temperature dependent near infrared spectra. *RSC Advances,* **6,** 105729-105736.

Czarnik-Matusewicz, B., Murayama, K., Tsenkova, R. & Ozaki, Y. (1999). Analysis of near-infrared spectra of complicated biological fluids by two-dimensional correlation spectroscopy: protein and fat concentration-dependent spectral changes of milk. *Applied Spectroscopy,* **53,** 1582-1594.

Damodaran, S. & Parkin, K. L. (2017). 2. Water and Ice Relations in Foods. In: Fennema's Food Chemistry (5th Edition). CRC Press.

DEAT (2006). South Africa Environment Outlook: A Report on the State of Environment. Department of Environmental Affairs and Tourism Pretoria.

Dijkstra, A. F. & de Roda Husman, A. M. (2014). Bottled and Drinking Water. In: Food Safety Management. Pp. 347-377. Elsevier.

DWA (2013). National Water Resource Strategy

DWAF (1996a). Agricultural Water Use: Irrigation In: South African Water Quality Guidelines. Pp. 20-76. Pretoria, South Africa: The Government Printer.

DWAF (1996b). South African Water Quality Guidelines: Dosmestic use. (edited by FORESTRY, D. O. W. A. A.). Pp. 29-182. Pretoria, South Africa The Government Printer.

Falk, M. & Ford, T. (1966). Infrared spectrum and structure of liquid water. *Canadian Journal of Chemistry,* **44,** 1699-1707.

FAO (2016). AQUASTAT Main Database, Food and Agriculture Organization of the United Nations (FAO).

Flörke, M., Kynast, E., Bärlund, I., Eisner, S., Wimmer, F. & Alcamo, J. (2013). Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study. *Global Environmental Change,* **23,** 144-156.

Foodstuffs, Cosmetics and Disinfectant Act and Regulations (2010). Act no.54 of 1972, G.N.R. 718/2010. Johannesburg, South Africa: Lex Patria Publishers.

Franks, F. (1973). Water: a comprehensive treatise. New York, Plenum.

Frost, V. J. & Molt, K. (1997). Analysis of aqueous solutions by near-infrared spectrometry (NIRS) III. Binary mixtures of inorganic salts in water. *Journal of molecular structure,* **410,** 573-579.

Garthwaite, P. H. (1994). An Interpretation of Partial Least Squares. *Journal of the American Statistical Association,* **89,** 122-127.

Geladi, P., MacDougall, D. & Martens, H. (1985). Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy,* **39,** 491-500.

Giangiacomo, R. (2006). Study of water–sugar interactions at increasing sugar concentration by NIR spectroscopy. *Food chemistry,* **96,** 371-379.

Gowen, A. A., Amigo, J. M. & Tsenkova, R. (2013). Characterisation of hydrogen bond perturbations in aqueous systems using aquaphotomics and multivariate curve resolution-alternating least squares. *Anal Chim Acta,* **759,** 8-20.

Gowen, A. A., Marini, F., Tsuchisaka, Y., De Luca, S., Bevilacqua, M., O'Donnell, C., Downey, G. & Tsenkova, R. (2015). On the feasibility of near infrared spectroscopy to detect contaminants in water using single salt solutions as model systems. *Talanta,* **131,** 609-618.

Gowen, A. A., Tsenkova, R., Bruen, M. & O'donnell, C. (2012). Vibrational Spectroscopy for Analysis of Water for Human Use and in Aquatic Ecosystems. *Critical Reviews in Environmental Science and Technology,* **42,** 2546-2573.

Gowen, A. A., Tsuchisaka, Y., O'Donnell, C. & Tsenkova, R. (2011). Investigation of the potential of near infrared spectroscopy for the detection and quantification of pesticides in aqueous solution. *American Journal of Analytical Chemistry,* **2,** 53.

Hanjra, M. A. & Qureshi, M. E. (2010). Global water crisis and future food security in an era of climate change. *Food Policy,* **35,** 365-377.

Hasted, J. (1972). Water: A Comprehensive Treatise. *The Physics and Physical Chemistry of Water,* **1,** 255-305.

Headrick, J. M., Diken, E. G., Walters, R. S., Hammer, N. I., Christie, R. A., Cui, J., Myshakin, E. M., Duncan, M. A., Johnson, M. A. & Jordan, K. D. (2005). Spectral Signatures of Hydrated Proton Vibrations in Water Clusters. *Science,* **308,** 1765.

Jinendra, B., Tamaki, K., Kuroki, S., Vassileva, M., Yoshida, S. & Tsenkova, R. (2010). Near infrared spectroscopy and aquaphotomics: Novel approach for rapid in vivo diagnosis of virus infected soybean. *Biochemical and biophysical research communications,* **397,** 685-690.

Kinoshita, K., Miyazaki, M., Morita, H., Vassileva, M., Tang, C., Li, D., Ishikawa, O., Kusunoki, H. & Tsenkova, R. (2012a). Spectral pattern of urinary water as a biomarker of estrus in the giant panda. *Scientific reports,* **2,** 856.

Kinoshita, K., Miyazaki, M., Morita, H., Vassileva, M., Tang, C., Li, D., Ishikawa, O., Kusunoki, H. & Tsenkova, R. (2012b). Spectral pattern of urinary water as a biomarker of estrus in the giant panda. *Sci Rep,* **2,** 856.

Kirby, R. M., Bartram, J. & Carr, R. (2003). Water in food production and processing: quantity and quality concerns. *Food control,* **14,** 283-299.

Kojić, D., Tsenkova, R., Tomobe, K., Yasuoka, K. & Yasui, M. (2014). Water confined in the local field of ions. *ChemPhysChem,* **15,** 4077-4086.

Kovacs, Z., Bazar, G., Oshima, M., Shigeoka, S., Tanaka, M., Furukawa, A., Nagai, A., Osawa, M., Itakura, Y. & Tsenkova, R. (2016). Water spectral pattern as holistic marker for water quality monitoring. *Talanta,* **147,** 598-608.

Libnau, F. O., Kvalheim, O. M., Christy, A. A. & Toft, J. (1994). Spectra of water in the near- and mid-infrared region. *Vibrational Spectroscopy,* **7,** 243-254.

Lin, M., A Rasco, B., G Cavinato, A. & Al-Holy, M. (2009). Infrared (IR) Spectroscopy—Near-Infrared Spectroscopy and Mid-Infrared Spectroscopy. 119-143.

Luck, W. A. P. & Schiöberg, D. (1979). Spectroscopic investigations of the structure of liquid water and aqueous solutions. *Advances in Molecular Relaxation and Interaction Processes,* **14,** 277-296.

Lukacs, M., Bazar, G., Pollner, B., Henn, R., Kirchler, C. G., Huck, C. W. & Kovacs, Z. (2018). Near infrared spectroscopy as an alternative quick method for simultaneous detection of multiple adulterants in whey protein-based sports supplement. *Food control,* **94,** 331-340.

MacDonald, A. M., Bonsor, H. C., Dochartaigh, B. É. Ó. & Taylor, R. G. (2012). Quantitative maps of groundwater resources in africa. *Environmental Research Letters,* **7,** 024009.

Maeda, H., Ozaki, Y., Tanaka, M., Hayashi, N. & Kojima, T. (1997). Near infrared spectroscopy and chemometrics studies of temperature-dependent spectral variations of water: relationship between spectral changes and hydrogen bonds. *Journal of Near Infrared Spectroscopy,* **3,** 191-201.

Manley, M. & Baeten, V. (2018). Spectroscopic technique: near infrared (NIR) spectroscopy. In: Modern techniques for food authentication. Pp. 51-102. Elsevier.

Molden, D. (2007). Water responses to urbanization. Springer.

Molden, D., Oweis, T., Steduto, P., Bindraban, P., Hanjra, M. A. & Kijne, J. (2010). Improving agricultural water productivity: Between optimism and caution. *Agricultural water management,* **97,** 528-535.

Morrison, G., Fatoki, O. S., Persson, L. & Ekberg, A. (2001). Assessment of the impact of point source pollution from the Keiskammahoek Sewage Treatment Plant on the Keiskamma River-pH, electrical conductivity, oxygen-demanding substance (COD) and nutrients. *Water SA,* **27,** 475-480.

Munćan, J. (2012). Comparative Study on Structure and Properties of Water by Infra Red and Opto-Magnetic Spectroscopy. *Contemporary Materials,* **1**.

Munćan, J. S., Matija, L., Simić-Krstić, J. B., Nijemčević, S. S. & Koruga, Đ. L. (2014). Discrimination of mineral waters using near infrared spectroscopy and aquaphotomics. *Hemijska industrija,* **68,** 257-264.

Mura, S., Cappai, C., Greppi, G. F., Barzaghi, S., Stellari, A. & Cattaneo, T. M. P. (2019). Vibrational spectroscopy and Aquaphotomics holistic approach to determine chemical compounds related to sustainability in soil profiles. *Computers and Electronics in Agriculture,* **159,** 92-96.

Myneni, S., Luo, Y., Näslund, L. Å., Cavalleri, M., Ojamäe, L., Ogasawara, H., Pelmenschikov, A., Wernet, P., Väterlein, P. & Heske, C. (2002). Spectroscopic probing of local hydrogen-bonding structures in liquid water. *Journal of Physics: Condensed Matter,* **14,** L213.

Nakakimura, Y., Vassileva, M., Stoyanchev, T., Nakai, K., Osawa, R., Kawano, J. & Tsenkova, R. (2012). Extracellular metabolites play a dominant role in near-infrared spectroscopic quantification of bacteria at food-safety level concentrations. *Analytical Methods,* **4,** 1389-1394.

Omta, A. W., Kropman, M. F., Woutersen, S. & Bakker, H. J. (2003). Negligible effect of ions on the hydrogen-bond structure in liquid water. *Science,* **301,** 347-349.

Osborne, B. G., Fearn, T. & Hindle, P. H. (1993). Practical NIR spectroscopy with applications in food and beverage analysis. Longman scientific and technical.

Ozaki, Y., McClure, W. F. & Christy, A. A. (2006). Near-infrared spectroscopy in food science and technology. John Wiley & Sons.

Pang, X.-F. (2014). Water: molecular structure and properties. World Scientific.

Pavia, D. L., Lampman, G. M., Kriz, G. S. & Vyvyan, J. A. (2008). Introduction to spectroscopy. Cengage Learning.

Pierna, J. A. F., Manley, M., Dardenne, P., Downey, G. & Baeten, V. (2018). Spectroscopic Technique: Fourier Transform (FT) Near-Infrared Spectroscopy (NIR) and Microscopy (NIRM). In: Modern Techniques for Food Authentication. Pp. 103-138. Elsevier.

Pitman, W. (2011). Overview of water resource assessment in South Africa: current state and future challenges. *Water SA,* **37,** 659-664.

Putra, A., Meilina, H. & Tsenkova, R. (2012). Use of near-infrared spectroscopy for determining the characterization metal ion in aqueous solution. In: Proceedings of The Annual International Conference, Syiah Kuala University-Life Sciences & Engineering Chapter.

Rahmanian, N., Ali, S. H. B., Homayoonfard, M., Ali, N., Rehan, M., Sadef, Y. & Nizami, A. (2015). Analysis of physiochemical parameters to evaluate the drinking water quality in the State of Perak, Malaysia. *Journal of Chemistry,* **2015**.

Reid, D. S. & Fennema, O. R. (2007). 2 Water and Ice. *Fennema's food chemistry***,** 17.

Rice, E. W. B., Eaton, R. B., Clesceri, A. D. & Bridgewater, L. S. (2012). Standard methodsfor the examination of water and wastewater.

Rinnan, Å., Berg, F. v. d. & Engelsen, S. B. (2009a). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry,* **28,** 1201-1222.

Rinnan, Å., Nørgaard, L., Berg, F. v. d., Thygesen, J., Bro, R. & Engelsen, S. B. (2009b). Data Pre-processing. 29-50.

Robertson, W. H., Diken, E. G., Price, E. A., Shin, J.-W. & Johnson, M. A. (2003). Spectroscopic determination of the OH− solvation shell in the OH−·(H2O) n clusters. *Science,* **299,** 1367-1372.

Sampathkumar, K., Kavimani, V., Rathnavel, P. & Senthilkumar, P. (2010). Potability studies of drinking water in rural villages of Coimbatore district, Tamilnadu, India. *International Journal of Applied Environmental Sciences,* **5,** 729-740.

Šašić, S., Segtnan, V. & Ozaki, Y. (2002). Self-modeling curve resolution study of temperature-dependent near-infrared spectra of water and the investigation of water structure. *The Journal of Physical Chemistry A,* **106,** 760-766.

Savitzky, A. & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry,* **36,** 1627-1639.

Segarra-Martí, J., Merchán, M. & Roca-Sanjuán, D. (2012). Ab initio determination of the ionization potentials of water clusters (H2O)n (n = 2−6). *The Journal of Chemical Physics,* **136,** 244306.

Segtnan, V. H., Šašić, Š., Isaksson, T. & Ozaki, Y. (2001). Studies on the structure of water using two-dimensional near-infrared correlation spectroscopy and principal component analysis. *Analytical Chemistry,* **73,** 3153-3161.

Siesler, H. W. (2006). Near-infrared spectroscopy : principles, instruments, applications. Weinheim, Weinheim : Wiley-VCH.

Siesler, H. W., Ozaki, Y., Kawata, S. & Heise, H. M. (2008). Near-infrared spectroscopy: principles, instruments, applications. John Wiley & Sons.

Slavchev, A., Kovacs, Z., Koshiba, H., Bazar, G., Pollner, B., Krastanov, A. & Tsenkova, R. (2017). Monitoring of water spectral patterns of lactobacilli development as a tool for

rapid selection of probiotic candidates. *Journal of Near Infrared Spectroscopy,* **25,** 423-431.

Storey, M. V., van der Gaag, B. & Burns, B. P. (2011). Advances in on-line drinking water quality monitoring and early warning systems. *Water Res,* **45,** 741-747.

Tanaka, M., Shibata, A., Hayashi, N., Kojima, T., Maeda, H. & Ozaki, Y. (1997). Discrimination of commercial natural mineral waters using near infrared spectroscopy and principal component analysis. *Journal of Near Infrared Spectroscopy,* **3,** 203-210.

Tsenkova, R. (2005). Visible-near infrared perturbation spectroscopy: Water in action seen as a source of information. . In: 12th International Conference on Near-infrared Spectroscopy. Pp. 607-612. New Zealand.

Tsenkova, R. (2006). Disease Diagnosis Related to Food Safety in Dairy. *Near-Infrared Spectroscopy in Food Science and Technology***,** 379.

Tsenkova, R. (2009). Introduction: Aquaphotomics: dynamic spectroscopy of aqueous and biological systems describes peculiarities of water. *Journal of Near Infrared Spectroscopy,* **17,** 303.

Tsenkova, R. & Gowen, A. (2011). NIR Spectroscopy: a tool for aquaphotomics. In: Pre-conference course, 15th International Conference on Near Infrared Spectroscopy. Pp. 1-22. Cape Town, South Africa.

Tsenkova, R., Kovacs, Z. & Kubota, Y. (2015). Aquaphotomics: near infrared spectroscopy and water states in biological systems. In: Membrane Hydration. Pp. 189-211. Springer.

Tsenkova, R., Muncan, J., Kovacs, Z. & Pollner, B. (2018). Essentials of Aquaphotomics and its Chemometrics Approaches. *Frontiers in Chemistry,* **6,** 363.

UNU-INWEH (2013). Water Security and the Global Water Agenda: A UN-Water Analytical Brief. Ontario, Canada: United Nations University.

Varmuza, K. & Filzmoser, P. (2011). Introduction to multivariate statistical analysis in chemometrics. CRC press.

Velíšek, J. (2014). The chemistry of food. Chichester, West Sussex, UK : Wiley Blackwell.

Vero, S., Tornielli, C. & Cattaneo, T. (2010). Aquaphotomics: wavelengths involved in the study of the speciation of metal ions (Zn2+, Pb2+ and Ag+) in aqueous solutions. *NIR news,* **21,** 11.

Wang, L., Lee, F. S. C., Wang, X. & He, Y. (2006). Feasibility study of quantifying and discriminating soybean oil adulteration in camellia oils by attenuated total reflectance MIR and fiber optic diffuse reflectance NIR. *Food chemistry,* **95,** 529-536.

WHO (2011). Guidelines for drinking-water quality. *WHO chronicle,* **38,** 104-108.

Williams, P. (2009). Influence of water on prediction of composition and quality factors: the Aquaphotomics of low moisture agricultural materials. *Journal of Near Infrared Spectroscopy,* **17,** 315.

Wold, S. (1995). Chemometrics; what do we mean with it, and what do we want from it? *Chemometrics and intelligent laboratory systems,* **30,** 109-115.

Xantheas, S. S. (1995). Ab initio studies of cyclic water clusters (H2O) n, n= 1–6. III. Comparison of density functional with MP2 results. *The Journal of Chemical Physics,* **102,** 4505-4517.

Xantheas, S. S. (2000). Cooperativity and hydrogen bonding network in water clusters. *Chemical Physics,* **258,** 225-231.

Chapter 3

Materials and methods

The experimental work of this project was divided into four separate studies with each focusing on one of the four objectives. These included the effect of different temperatures, water sources, types of water and filter materials on the water spectrum.

3.1 Influence of temperature

3.1.1 Samples

Deionised water was obtained at the start of each experiment using an Elgastat B114 deioniser (Veolia Water, France). Mineral water was sourced from different local retailers the day prior to each experiment and stored at an ambient temperature of 22-23˚C for 24h. Stellenbosch municipal tap water was acquired in the laboratory at the start of each experiment.

3.1.2 Spectral acquisition

From each sample, 850 µL was pipetted into a 1.0 mm quartz cuvette (Hellma Analytics, Germany) and scanned. The water samples (deionised, mineral and tap water) were each scanned in triplicate at the following temperatures: 20˚ C, 25˚ C and 30˚ C.

3.2 Spring water from different sources

3.2.1 Samples

Bottled spring water was sourced from a manufacturer that bottles at three different locations, sources A, B and C. The water samples were stored at ambient temperatures (22 – 23˚C) out of direct sunlight until analysis. An Elgastat B114 deioniser (Veolia Water, France) was used to produce deionised water for each day of scanning. Three samples from each of the three different geographical sources were scanned each day, for a consecutive period of four days. The deionised water was scanned as a control sample between every third sample and used to rinse the cuvettes between samples.

3.2.2 Spectral acquisition

Scanning was done at a temperature of 32 ˚C using the 0.2 mm quartz cuvette (Hellma Analytics, Germany) filled with 80 µL of sample. Each sample was scanned five consecutive times. Deionised water was scanned before the first sample and then again after every fifth sample.

3.3 Different types of bottled water

3.3.1 Samples

A total of 120 bottled water of different brands of mineral (60) and spring (60) water were sourced from a variety of retailers in the Western Cape area. The water bottles were stored at ambient temperature until analysis. Deionised water was obtained at the start of each scanning day using an Elgastat B114 deioniser (Veolia Water, France), which was used as a control sample and to rinse the cuvettes between samples. Sample scanning took place over a period of 19 days.

3.3.2 Spectral acquisition

The water samples were scanned using quartz cuvette (Hellma Analytics, Germany) with a path length of 0.2 mm at a temperature of 32˚C. The cuvette was filled with 80 µL of sample and scanned five consecutive times. Deionised water was scanned before the first sample and then again after every fifth sample.

3.4 River water and different filtration materials

3.4.1 Samples and sample preparation

Water was collected from the Plankenburg River (33˚55’58.6 S, 18˚51’05.4 E) and sampling was done according to the process described by the South African National Standards (SANS) 5667-6 method (SANS, 2006). Sterilised 5 L bottles were used to collect water from the river and were transported back to the department in insulated cooler boxes.

Six filtration columns were constructed containing different filtration material, four columns were constructed with two columns containing 2 L of experimental pine biochar and the other two columns contained 2 L of experimental black wattle biochar. Another two column were constructed containing 2 L of granulated activated charcoal and 300 g of silica

sand, respectively. A schematic of the columns constructed using the different filtration materials is shown in **figure 3.1**.

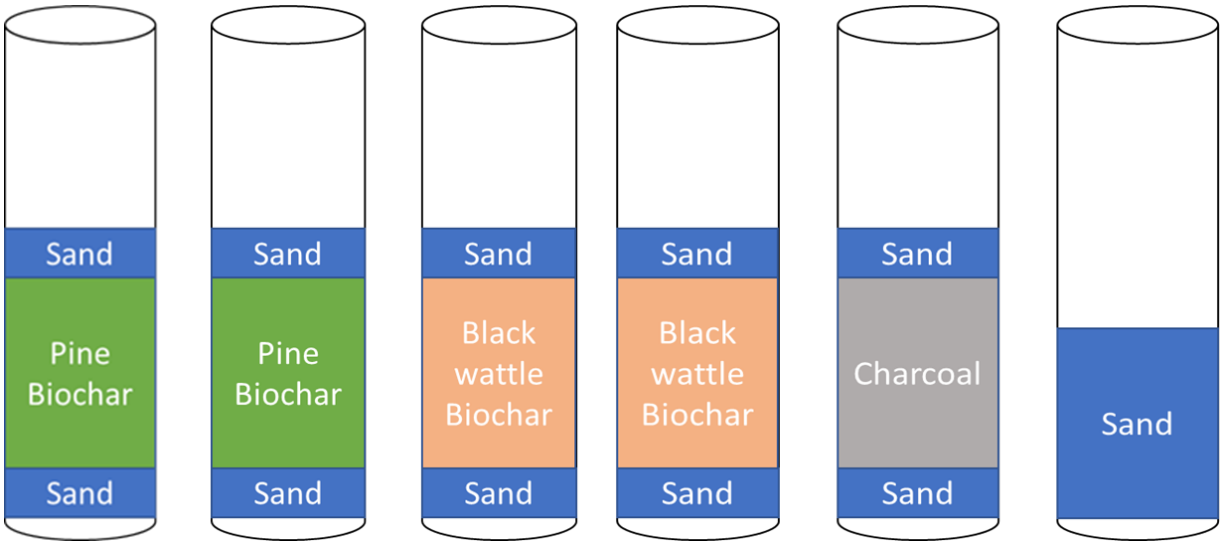

Figure 3.1 Basic schematic of the columns constructed with the different filtration material.

Every 3rd day over a period of 30 days, 2 L of river water was filtered through the six columns and collected. Deionised water was used as a reference sample (used to synchronise the spectra to reduce the effect the day of scanning will have on the water spectrum) and was produced using an Elgastat B114 deioniser (Veolia Water, France). The river water and filtrates were stored at 4˚C and were analysed to determine the physico-chemical properties.

3.4.2 Spectral acquisition

Each sample was scanned at 32˚C using a quartz cuvette (Hellma Analytics, Germany) with a path length of 0.2 mm. From each column 250 ml was collected. The 250 ml river water and filtrate samples were each divided into 10 sub-samples of 25 ml each. From each subsample 80 µL was transferred into the cuvette and scanned five consecutive times. Scanning was done in a randomised order and deionised water was scanned before the first sample, and then scanned again after every 5th sample.

3.5 Physico-chemical analyses

The physico-chemical parameters of the river water samples were determined (in duplicate) using the following methods.

Ultraviolet transmission percentage, turbidity, and electrical conductivity

A hand-held Sense T254 UV-Transmittance Monitor (Berson, The Netherlands) was used to determine the ultraviolet transmission percentage (UVT %) of the river water. A portable Orion AQ3010 Turbidity Meter (Thermo Scientific, USA) was used to determine the turbidity. A portable HI 8733 conductivity meter (Hanna Instruments, USA) was used to determine electrical conductivity (EC) and the total amount of dissolved salts in the water samples. All the instruments were operated according to the manufacturer's instructions and were calibrated using the appropriate standards or deionised water.

Alkalinity and pH

Alkalinity was determined according to Standard Methods (APHA, 2012). Twenty millilitres of sample was titrated with 0.1 N H_2SO_4 until a pH of 4.3 was reached. The alkalinity of each sample was calculated using the standard method. The pH of the river water samples was determined using a pH meter (WTW, Germany). The pH meter was calibrated with the appropriate pH buffers and was operated according to the manufacturer's manual.

Total suspended solids (TSS), volatile suspended solids (VSS) and total dissolved solids (TDS)

The amount of organic and inorganic matter present in water was determined using TSS, TDS and VSS. These methods were done according to the instructions provided by Standard Methods (APHA, 2012). The TDS was determined with the use of a TDS-3 meter (HM Digital), the meter was calibrated according to the manufacturer's manual using a 1000 ppm NaCl solution.

3.6 NIR Instrumentation

All spectra were obtained in the wavelength range of 1000-2500 nm using a Buchi NIRFlex N-500 Fourier transform NIR spectrophotometer (BÜCHI Labortechnik AG, Flawil, Switzerland). The instrument was fitted with a tungsten halogen lamp, a temperature-controlled Indium Gallium Arsenide (InGaAs) detector and a temperature-controlled cuvette holder and was operated using NIRWare software suite. The instrument performed 32 successive scans per sample, at a signal-to-noise ratio of 10 000 and a resolution of 8 cm^{-1} with a data point every 4 cm^{-1} which result in 1501 data points.

3.7 Spectral data analysis

Data analysis was performed using the 1300 – 1600 nm wavelength range, the first overtone of O-H. Spectral data were processed using Unscrambler software v.10.1 (Camo Inc., OSLO, Norway) and MATLAB v 9.2.0.538062 (2017a) (MathWorks, Massachusetts, USA).

The effect of scanning over a number of days was eliminated by synchronising the spectra (Kovacs *et al.*, 2016). This was done by subtracting the average spectrum of the daily deionised water from each spectrum of the same day, and the total average spectrum of all the deionised water scanned was added back to all the sample spectra (Kovacs *et al.*, 2016). Spectra were smoothed and the second derivative calculated using Savitzky-Golay 2nd order polynomial and 21 points (Savitzky & Golay, 1964). Standard normal variate (SNV) (Barnes *et al.*, 1989) was applied to the smoothed data.

Principal component analysis (PCA) was used to identify patterns and visualise the information present in the spectral data set and to identify outliers (Cowe & McNicol, 1985). Partial least squares discriminant analysis was used to classify the spectra into classes based on their class membership. Partial least squares (PLS) regression was used to predict a set of dependent variables from a large set of independent variables (Wold, 1975). Regression was performed using smoothed and SNV transformed data. Activated water absorbance bands were identified using these data analysis techniques. The activated water absorbance bands, which repeatedly occurred throughout the data analysis process were used to describe the water spectral patterns and construct the aquagrams according to the method of Tsenkova *et al.* (2018).

PLS-DA models were developed to differentiate between the different types of water samples. Each data set was spilt into a calibration (70%) and validation (30 %) sets using the Duplex algorithm. The overall performance of the models were validated by calculating the performance measures using the equation 3.1 – 3.8. The classification accuracy was used to prove the effectiveness of the overall model. The probability of a positive response being correctly classified is referred to as sensitivity. While specificity is referred to as the probability of a negative response being correctly classified. Sensitivity and specificity is used together to evaluate the classification algorithm's performance for a single class. The rate of values that measure the accuracy of positive predictions of the model is referred to as precision. The F1 score is used as an overall measure of the accuracy of a model that combines precision and sensitivity.

$$Classification\ accuracy\ (\%) = \frac{TP+TN}{(TP+TN+FP+FN)} \times 100\% \qquad \text{(eq. 3.1)}$$

$$False\ positive\ error\ (\%) = \frac{FP}{(TP+TN+FP+FN)} \times 100\% \quad \text{(eq. 3.2)}$$

$$False\ negative\ error\ (\%) = \frac{FN}{(TP+TN+FP+FN)} \times 100\% \quad \text{(eq. 3.3)}$$

$$Sensitivity/\ recall\ (\%) = \frac{TP}{(TP+FN)} \times 100\% \quad \text{(eq. 3.4)}$$

$$Specificity\ (\%) = \frac{TN}{(TN+FP)} \times 100\% \quad \text{(eq. 3.5)}$$

$$Precision\ (\%) = \frac{TP}{(TP+FP)} \times 100\% \quad \text{(eq. 3.6)}$$

$$F1\ Score\ (\%) = \frac{2 \times Precision \times Sensitivity}{Precision+Sensitivity} \quad \text{(eq. 3.7)}$$

$$Misclassification\ rate\ (\%) = \frac{FP+FN}{(TP+TN+FP+FN)} \times 100\% \quad \text{(eq. 3.8)}$$

Where:

True positives (TP) = positive responses classified as positive responses

True negatives (TN) = negative responses classified as negative responses

False positives (FP) = negative responses classified as positive responses

False negatives (FN) = positive responses classified as negative responses

3.8 References

APHA (2012). Standard methods for the examination of water and wastewater. Washington, D.C. : American Public Health Association Press.

Barnes, R., Dhanoa, M. S. & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy,* **43,** 772-777.

Cowe, I. A. & McNicol, J. W. (1985). The Use of Principal Components in the Analysis of Near-Infrared Spectra. *Applied Spectroscopy,* **39,** 257-266.

Kovacs, Z., Bazar, G., Oshima, M., Shigeoka, S., Tanaka, M., Furukawa, A., Nagai, A., Osawa, M., Itakura, Y. & Tsenkova, R. (2016). Water spectral pattern as holistic marker for water quality monitoring. *Talanta,* **147,** 598-608.

SANS (2006). Method 5667-6. Water quality-Sampling, Part 6: Guidance on sampling of rivers and streams. South African Bureau of Standards Pretoria.

Savitzky, A. & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry,* **36,** 1627-1639.

Tsenkova, R., Muncan, J., Kovacs, Z. & Pollner, B. (2018). Essentials of Aquaphotomics and its Chemometrics Approaches. *Frontiers in Chemistry,* **6,** 363.

Wold, H. (1975). Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach. *Journal of Applied Probability,* **12,** 117-142.

Wold, S., Antti, H., Lindgren, F. & Öhman, J. (1998). Orthogonal signal correction of near-infrared spectra. *Chemometrics and intelligent laboratory systems,* **44,** 175-185.

Chapter 4

Results and Discussion

4.1 The effect of temperature

Water samples were investigated to determine the effect of three temperatures, 20˚C, 25˚C and 30˚C on the NIR spectra.

4.1.1 Spectral analysis

The mean and second derivative spectra of water at the three temperatures are presented in **Figure 4.1 – 4.2**. A broad band is observed in **Figure 4.1** and is due to the overlapping of the absorption related to the different hydrogen-bonded structures at 1450 nm. The overlapping is due to the OH stretching combination of the symmetric stretching and asymmetric stretching modes (v_1+v_3) (Tsenkova, 2009).

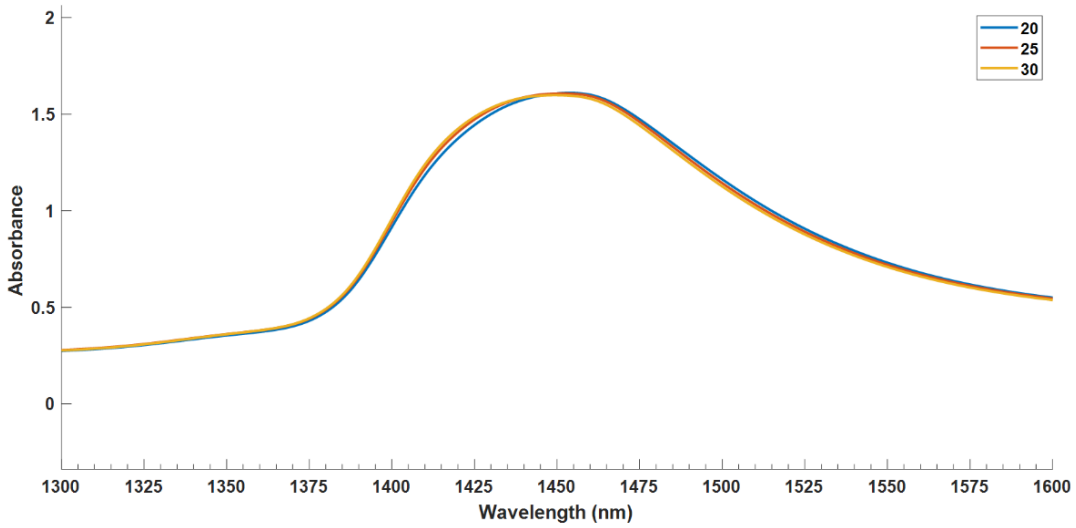

Figure 4.1 Mean spectra obtained at the three different temperatures for the first overtone of water (1300 – 1600 nm).

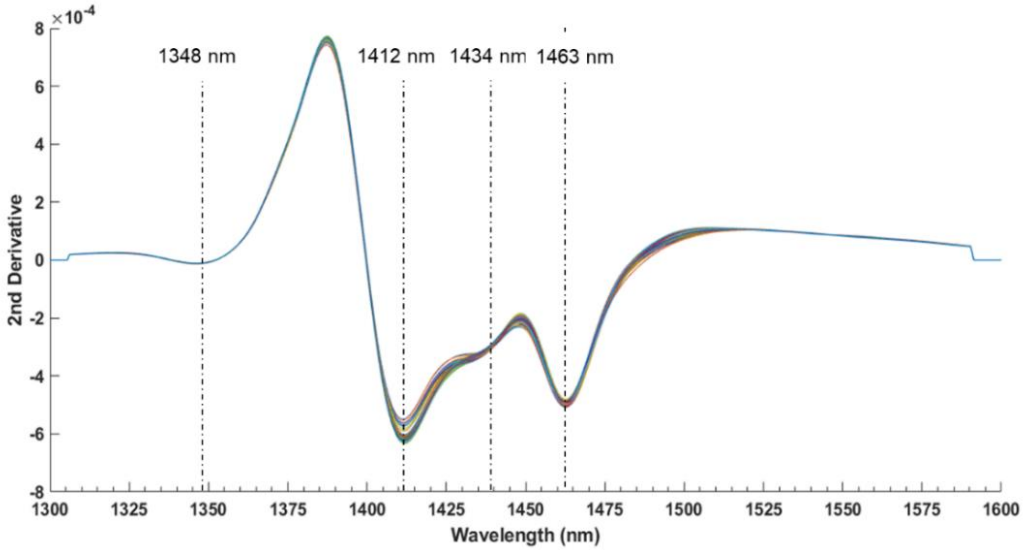

Figure 4.2 Second derivative spectra (calculated with a Savitzky-Golay filter using 2nd order polynomial and 21 points) in the first overtone of water (1300 – 1600 nm) of the water scanned at the three different temperatures.

The 2nd derivative (**Figure 4.2**) of the first overtone indicated four water absorbance bands at 1348 nm, 1412 nm, 1434 nm and 1462 nm. The band at 1348 nm is linked to the asymmetric stretching of the water molecule (Siesler, 2006). Water molecules with different numbers of hydrogen bonds have been assigned to the absorbance bands of 1412 nm, 1434 nm and 1462 nm. The region at 1412 nm is associated with water molecules with little or no intermolecular hydrogen bonds (S_0) (Segtnan et al., 2001), while water molecules with one (S_1) and two (S_2) hydrogen bonds are assigned to the 1434 nm and 1462 nm bands, respectively (Cattaneo *et al.*, 2009; Siesler, 2006). The water absorbance band, 1462 nm, has been assigned to water molecules with two hydrogen bonds (S_2). Changes in intensity are observed at the 1412 nm and 1462 nm band. These changes are due to temperature (Segtnan *et al.*, 2001), with the greatest increase in intensity observed at 1412 nm. This indicates that temperature does have an effect on the spectra, as the increase in temperature weakens the hydrogen bonds resulting in changes in water molecular species sizes (Cui *et al.*, 2016).

4.1.2 Principal component analysis (PCA)

The Savitzky – Golay smoothed and standard normal variant (SNV) corrected data was used to calculate PCA score and loading plots, as seen in **Figure 4.3 – 4.5.** PC 1 and PC 2 describe 98% and 2% of the variation in the data set, respectively. Separation can be seen

in the direction of PC 2 (**Figure 4.3**) and can be attributed to the increase in temperature from 20 to 30˚C.

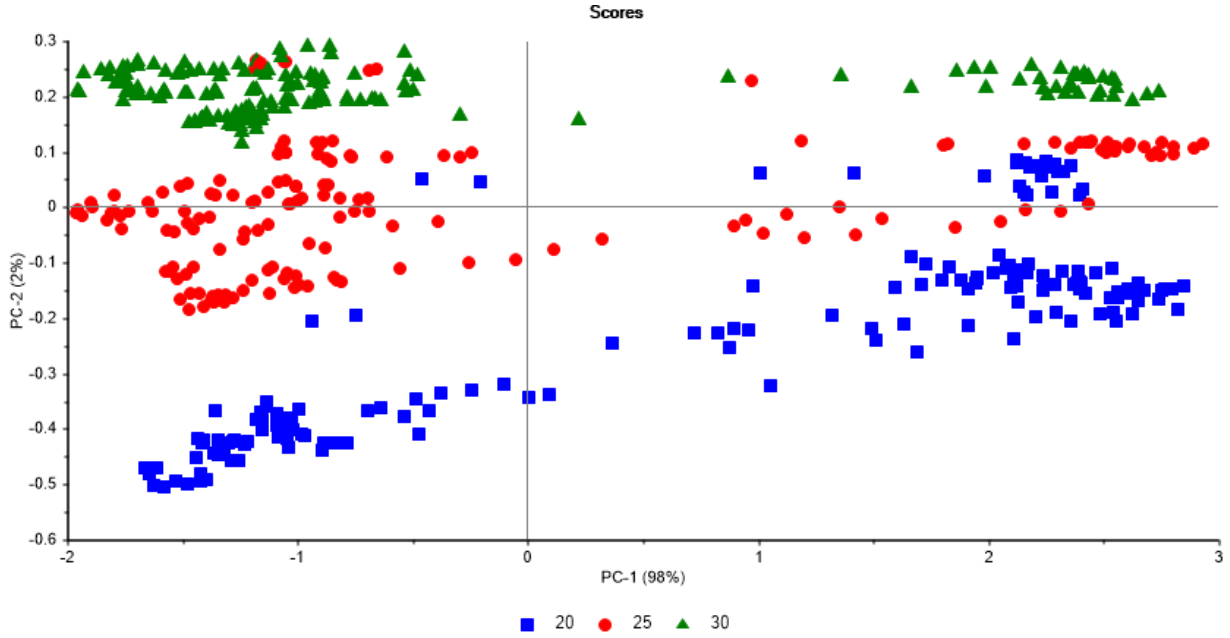

Figure 4.3 Principal component score plot (PC1 (98 %) vs. PC2 (2 %)) of the water scanned at the three different temperatures.

A single band at 1450 nm is observed in loadings plot of PC 1 (**Figure 4.4**) and this band is attributed to the combination of the first overtone of the OH bending and the fundamental OH asymmetric stretching vibration. The 1412 nm band is attributed to water molecules with little or no intermolecular hydrogen bonds (S_0) weakly bonded water molecules and a decrease in strong hydrogen-bonded water molecules.

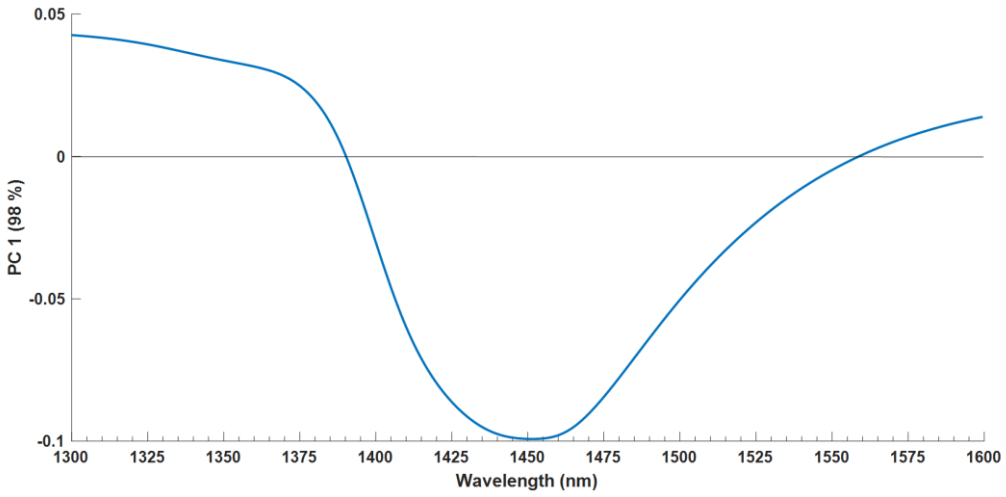

Figure 4.4 First principal component (PC 1) loadings plot for the water scanned at the three different temperatures.

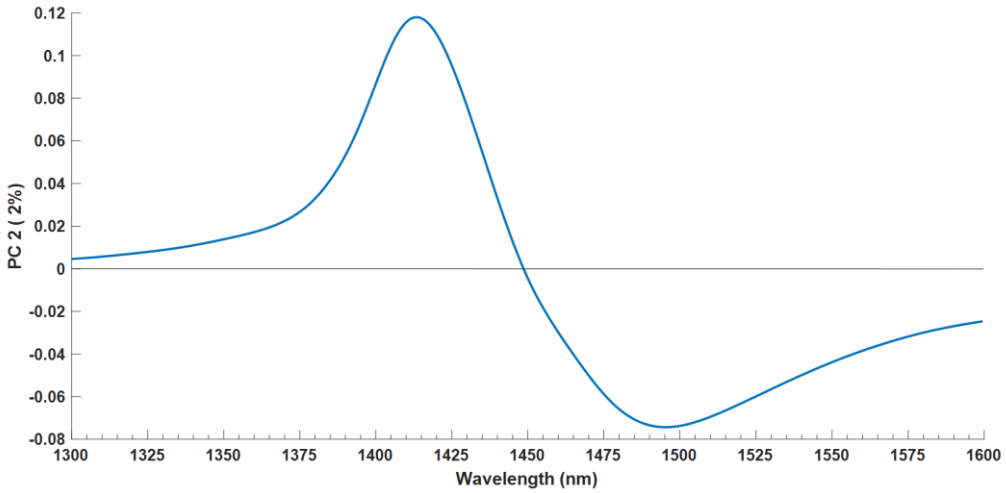

Figure 4.5 Second principal component (PC 2) loadings plot for the water scanned at the three different temperatures.

Two bands are observed in the second principal component loadings plot (**Figure 4.5**), a positive peak at 1412 nm and a negative peak at 1492 nm. (Segtnan et al., 2001). Water molecules with three intermolecular hydrogen bonds or trimers are attributed to the 1492 nm wavelength (Franks, 1973; Siesler, 2006). The positive and negative bands of PC2 indicates that there is a change in hydrogen bonding with an increase in temperature. This is due to the weakening of hydrogen bonds with the increase in temperature, resulting in an increase of weakly bonded water molecules and a decrease in strong hydrogen-bonded water molecules.

These findings are comparable to the results reported in literature (Gowen *et al.*, 2013; Maeda *et al.*, 1997; Segtnan *et al.*, 2001). These studies indicated that in the 1300 nm – 1600 nm region, the variation is predominately described by two bands, 1412 nm and 1492 nm. These two bands represent two types of water clusters which are linked, resulting in a two-state mixture model.

4.1.3 Conclusion

The PCA analysis of the water spectra in the first overtone (1300 – 1600 nm) region at temperatures of 20, 25 and 30˚C showed spectral variation mainly due to two bands at 1412 nm and 1492 nm. These bands arise from weakly and strongly bound hydrogen-bonded water molecules, respectively. Variation in temperature causes changes in the concentration of free and bonded OH groups (Chalmers & Griffiths, 2002). Changes in temperature alter the strength of the hydrogen bonds, which results in spectral changes seen in the water

spectrum (Maeda *et al.*, 1997). An increase in temperature leads to a reduction in water clusters, due to the weakening of hydrogen-bonding (Cui *et al.*, 2016).

4.2 Spring water from different sources

Spring water bottled at three different geographical locations was scanned to determine if it is possible to differentiate between the three sources.

4.2.1 Spectral analysis

The mean spectra of each spring water source, in the 1300 – 1600 nm range, is shown in **Figure 4.6**. The mean spectra of the three water sources are overlapping and appear to be identical, with one dominant band at 1450 nm. This absorbance band, 1450 nm, is attributed to the first overtone of the OH stretching vibration (Luck, 1974). Spectral subtraction was performed by subtracting the average spectrum of deionised water from the mean spectra of each of the spring water sources. This was done to highlight any subtle differences in the spectra. The resulting difference spectra is presented in **Figure 4.7**.

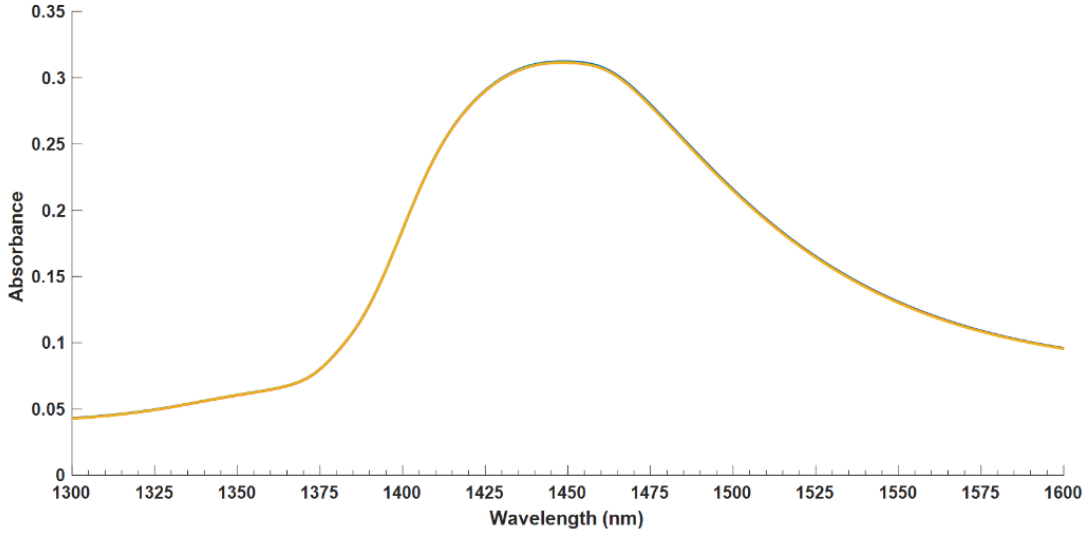

Figure 4.6 Mean (unprocessed) spectra of spring water from three different sources.

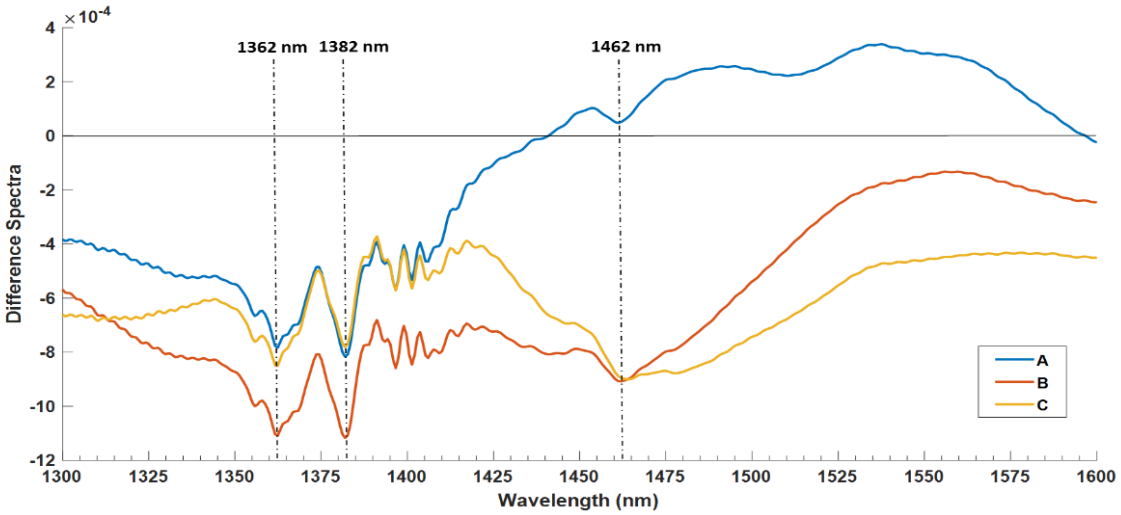

Figure 4.7 Difference spectra (Savitzky-Golay smoothing with 2nd order polynomial and 21 points) of the three different sources.

The difference spectra (**Figure 4.7**) of the three sources revealed water absorbance bands at 1362 nm, 1382 nm and 1462 nm. The bands at 1362 nm and 1382 nm are assigned to the OH stretching in water molecules (Xantheas, 1995). While the band at 1462 nm is assigned to water molecules with two intermolecular hydrogen bonds (S_2) (Franks, 1973).

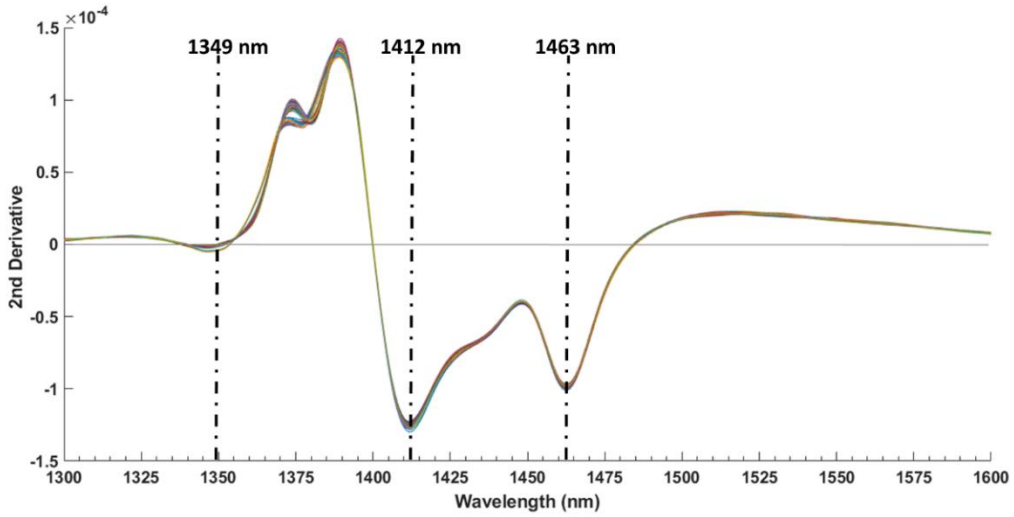

Figure 4.8 Second derivative (calculated with a Savitzky-Golay filter using 2nd order polynomial and 21 points) spectra in the first overtone of water (1300 – 1600 nm) of the three different spring water sources.

The 2nd derivative (**Figure 4.8**) of the first overtone indicated three water absorbance bands at 1349 nm, 1412 nm and 1463 nm. The band at 1349 nm is linked to the asymmetric

stretching of the water molecule (Siesler, 2006). The bands at 1412 nm and 1463 nm have been assigned to water molecules with different numbers of hydrogen bonds. The region of 1412 nm is associated with water molecules with little or no intermolecular hydrogen bonds (S_0) (Segtnan et al., 2001). The water absorbance band, 1463 nm, has been assigned to water molecules with two hydrogen bonds (S_2) (Franks, 1973; Siesler, 2006).

4.2.2 Multivariate data analysis

4.2.2.1 Principal component analysis

Principal component analysis was performed to explore the data by examining the differences between the three sources of spring water. The resulting score and loading plots are presented in **Figure 4.9 – 4.15**.

The score plots (**Figure 4.9 – 4.11**) showed minimal separation between the different spring water sources. PC 1 accounted for 62% of the variance in the data, whereas PC2, PC3 and PC4 accounted for 26%, 7% and 3%, respectively. This indicates that most of the variance in the data set is explained by the first two principal components.

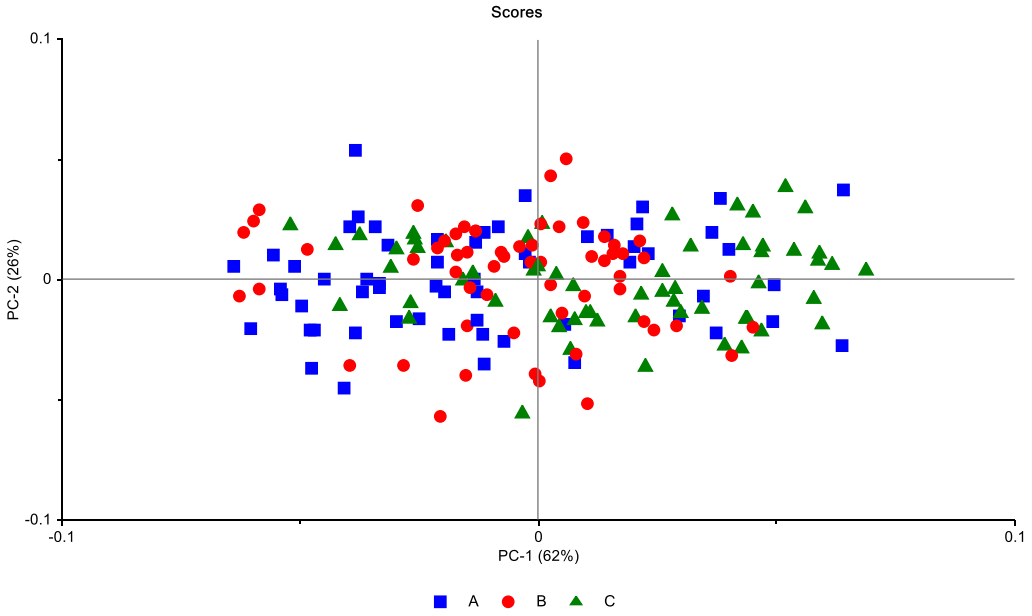

Figure 4.9 Principal component analysis score plot (PC1 (62 %) vs. PC2 (26 %)) of spring water sourced from three different sources.

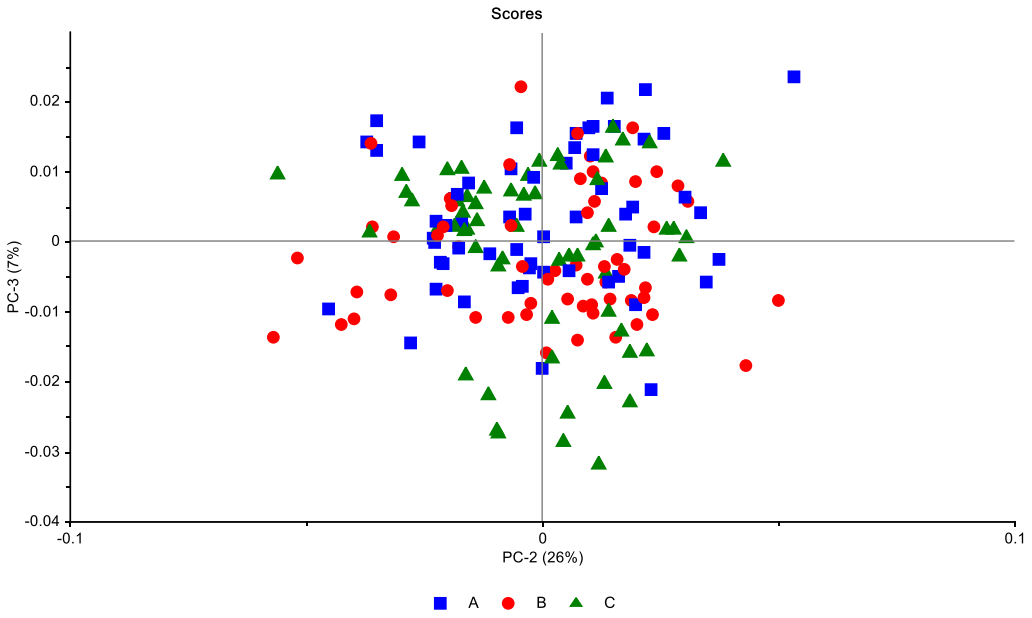

Figure 4.10 Principal component analysis score plot (PC2 (26 %) vs. PC3 (7 %)) of spring water sourced from three different sources.

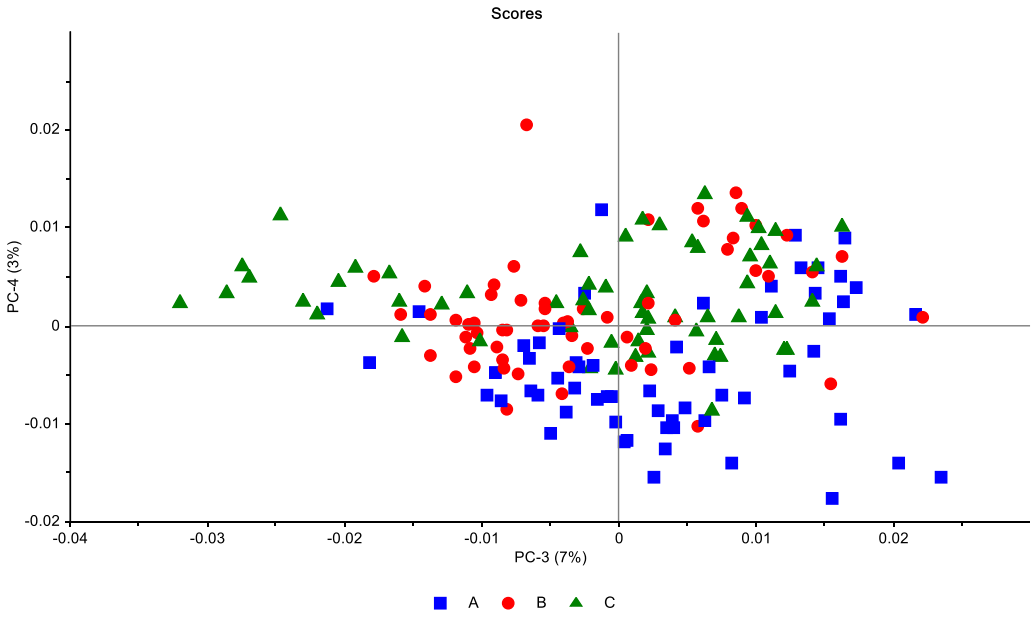

Figure 4.11 Principal component analysis score plot (PC3 (7 %) vs. PC4 (3 %)) of spring water sourced from three different sources.

The variance seen in the direction of PC 1 (**Figure 4.9**) may be attributed to the OH stretching and bending (v_2) of the first overtone of water, while the different water clusters can be the result of the variance in the direction of PC 2 (**Figure 4.10**). The score plots of

PC 2 vs PC 3 (**Figure 4.10**) and PC 3 vs PC 4 (**Figure 4.11**) showed no separation between the three different water sources. The minimal separation between the water sources indicates that there are similarities in their water spectral pattern.

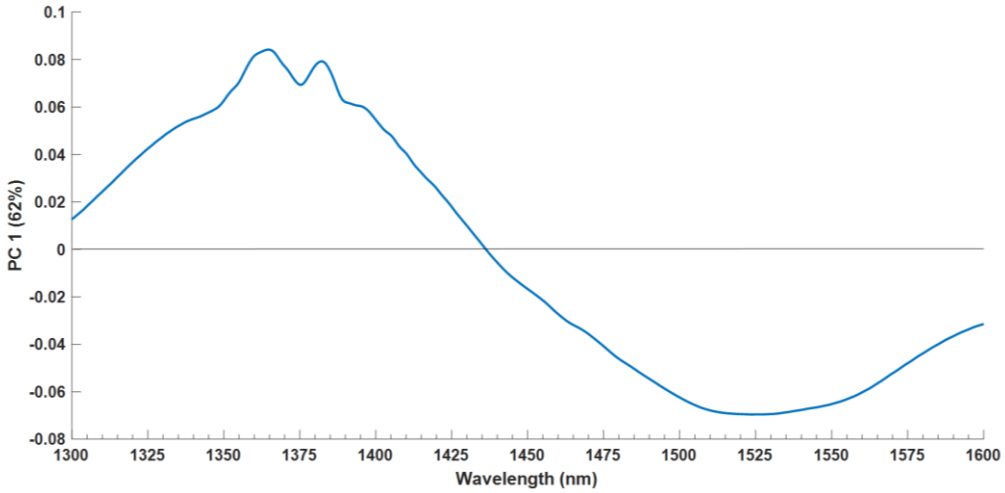

Figure 4.12 PCA loadings line plot for PC 1(62%)

.

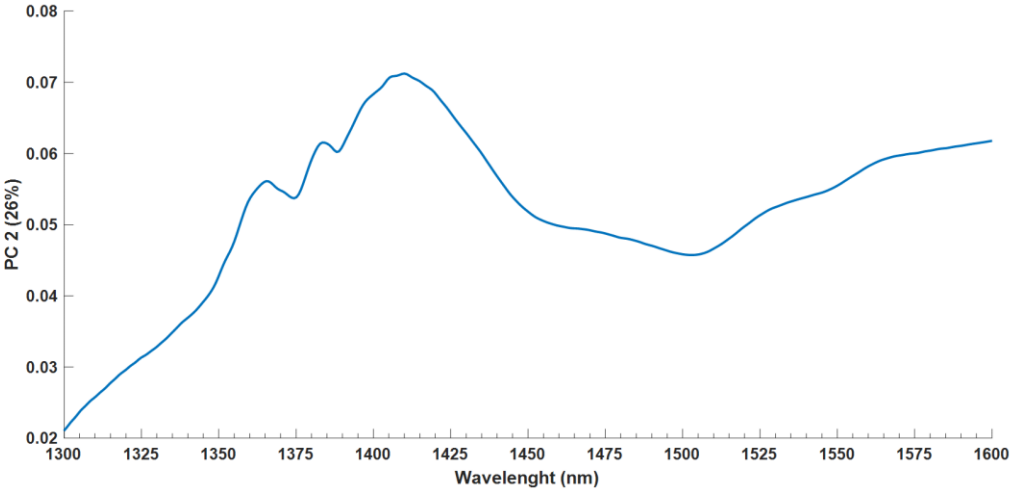

Figure 4.13 PCA loadings line plot for PC 2 (26%)

.

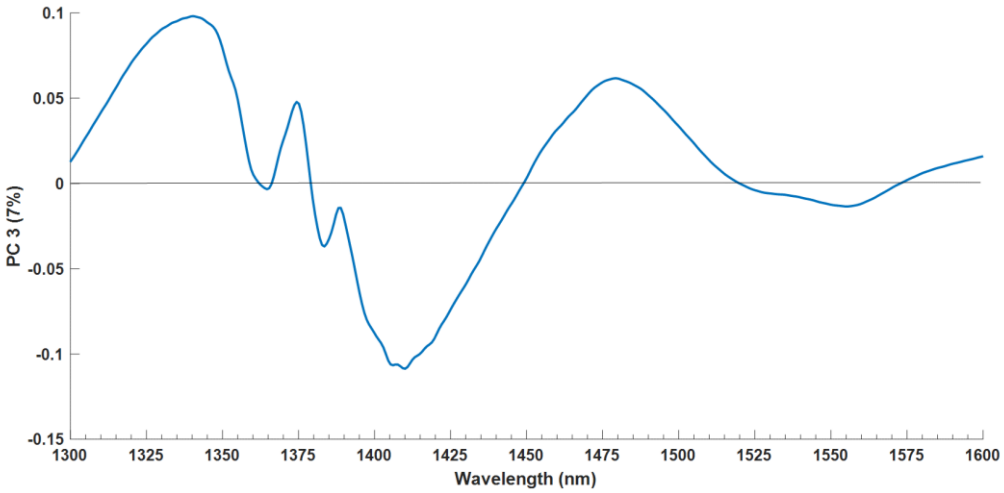

Figure 4.14 PCA loadings line plot for PC 3 (7%)

.

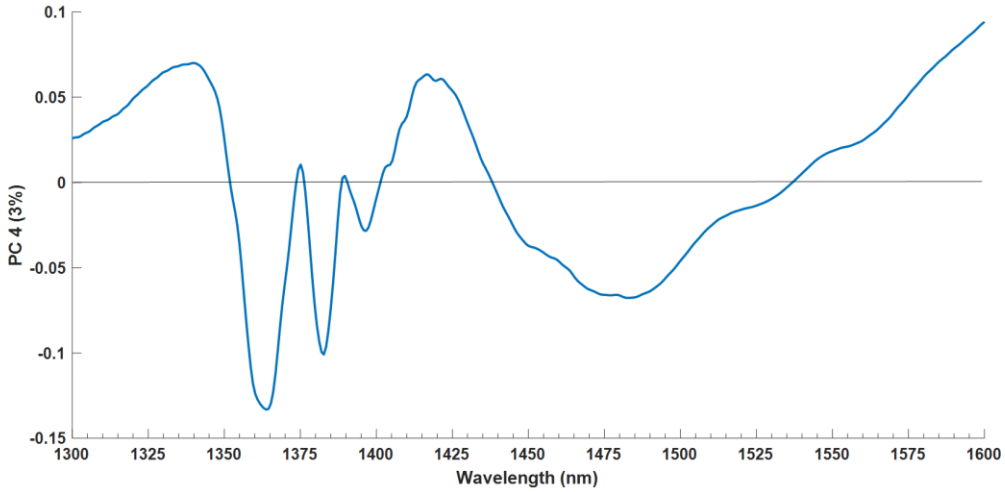

Figure 4.15 PCA loadings line plot for PC 4 (3%).

The loadings plot (**Figure 4.12**) of PC1 indicates three water absorbance bands at 1364 nm, 1382 nm and 1515 nm. The water band 1364 nm has been assigned to stretching of the hydrogen bond in water molecules (Xantheas, 1995). The water band at 1382 nm represents OH stretching (Xantheas, 1995). The 1515 nm water band has been assigned to bending vibration mode of the first overtone of water (Siesler, 2006). The PC2 loadings (**Figure 4.13**) plot revealed water absorbance bands at 1410 nm and 1458 nm. The 1410 nm water band is linked to water molecules with little or no intermolecular hydrogen bonds(S_0) (Segtnan et al., 2001). Water molecules with two hydrogen bonds (S_2) have been assigned to the 1458 nm wavelength (Franks, 1973).

Three water absorbance bands, 1340 nm, 1374 nm and 1478 nm were highlighted in the loadings plot of PC3 (**Figure 4.14**) and PC 4 (**Figure 4.15**). The 1340 nm band is linked to the asymmetric stretching of the water molecule (Siesler, 2006). The 1374 nm band is linked to the symmetrical and asymmetrical stretching of water molecules (Siesler, 2006). Water molecules with three hydrogen bonds (S_3) are assigned to the 1478 nm band (Franks 1973, Siesler, 2006).

4.2.2.2 Partial least squares discriminant analysis (PLS-DA)

The PLS-DA model of the three spring water sources provided satisfactory discrimination results, with an overall classification accuracy and misclassification rate of 89.6 % and 10.32%, respectively. The overall performance measures of the PLS-DA models are given in **Table 4.1**. For the source A class, 38 of the 40 samples were correctly classified, while one sample of source B was misclassified as source A. Forty-one (41) of the 45 source B samples were correctly classified. For source C, 34 of 41 samples were correctly classified and four samples of source B were misclassified as source C.

Table 4.1 The overall performance measures of the calibration, cross-validation and validation PLS-DA models

Model	Number of latent variables	Classification accuracy (%)	Misclassification rate (%)
Calibration	8	89.68	10.32
Cross-validation		86.51	13.49
Validation		81.48	18.52

The score plot of LV1 vs LV2 vs LV3 accounted for 34.23%, 26.89% and 26.10% of the variance, respectively (**Figure 4.16**). The plot shows slight overlapping between the classes. Separation was observed in the direction of LV1 (**Figure 4.17**), where source A is associated with the negative scores and source C with the positive scores.

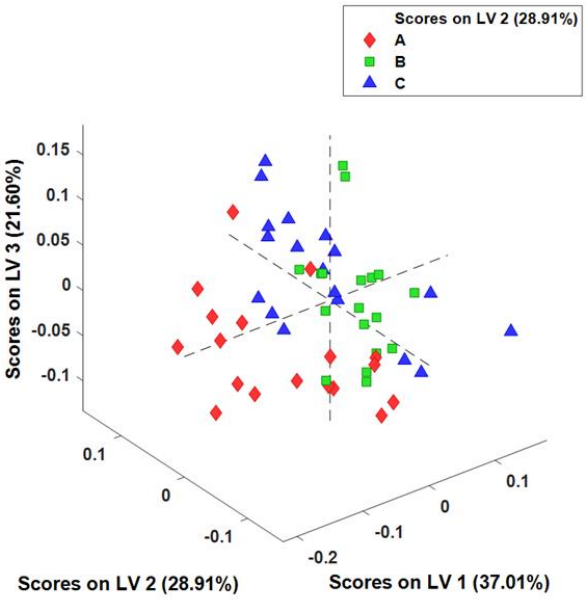

Figure 4.16 3D PLS-DA score plot (LV1 (37.01 %) vs LV2 (28.91 %) vs LV3 (21.60 %)) of source A (red), B (green) and C (blue).

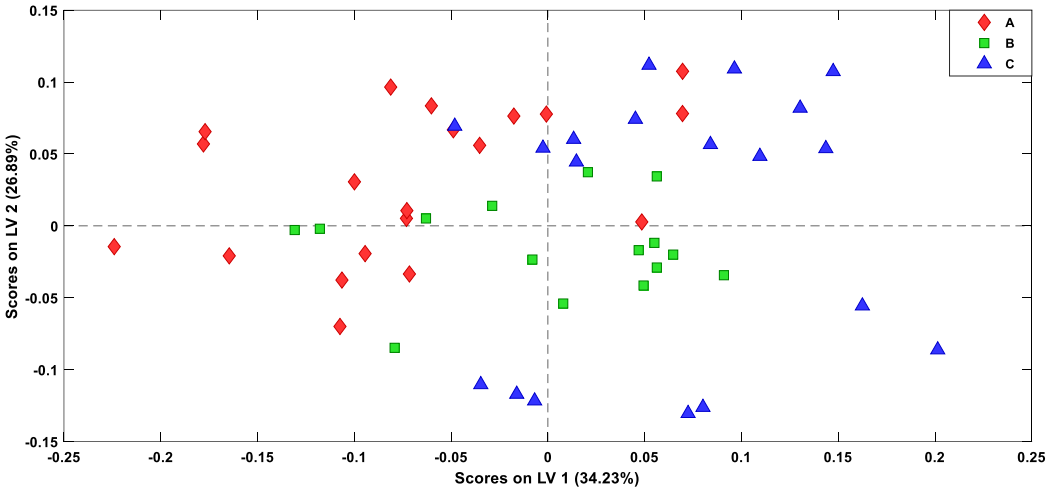

Figure 4.17 PLS-DA score plot (LV1 (34.23 %) vs LV2 (26.89 %)) of source A (red), B (green) and C (blue).

The PLS-DA prediction score plots (**Figure 4.18 – 4.20**) illustrated that nearly all three sources were predicted correctly. The source A (98.26 %) class had the highest classification accuracy, followed by sources C (94.17 %) and B (92.62 %). The model of source A has a sensitivity, specificity and F1 score above 97% (**Table 4.2**). Source B has a sensitivity of 82%, a specificity of 100% and F1 score of 90.11% (**Table 4.2**). Source C

sensitivity of 91.89%, a specificity of 95.18% and F1 score of 90.67%. This confirm that the model is able to classify the water samples correctly.

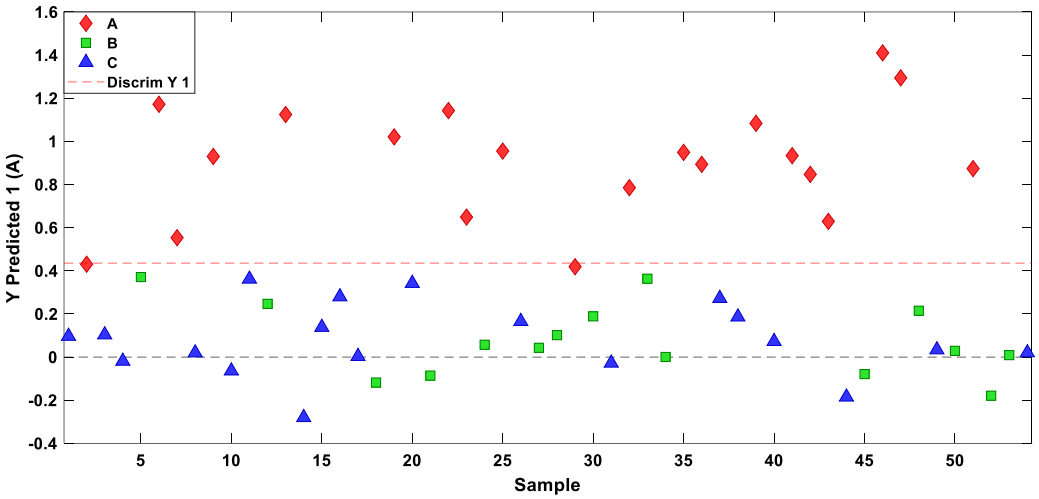

Figure 4.18 PLS-DA prediction score plot of source A (red) showing the predicted objects.

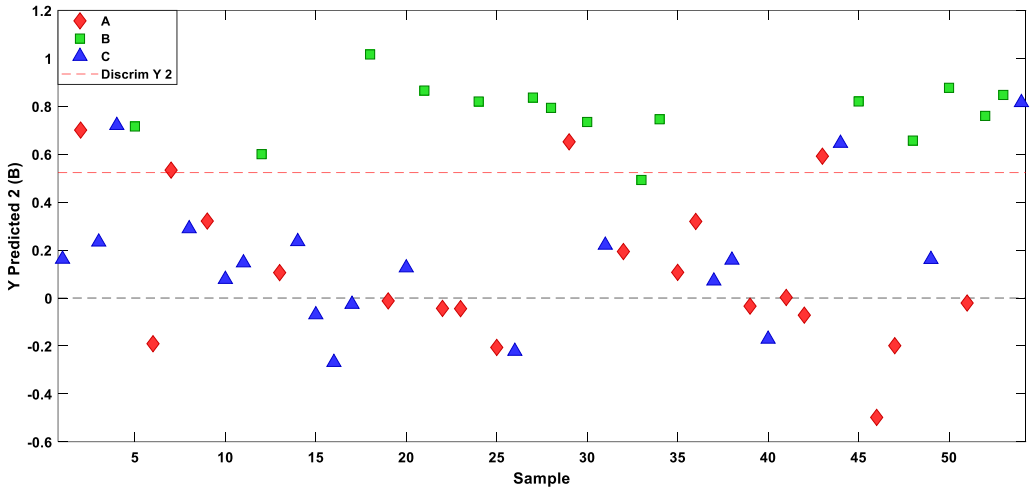

Figure 4.19 PLS-DA prediction score plot of source B (green) showing the predicted objects

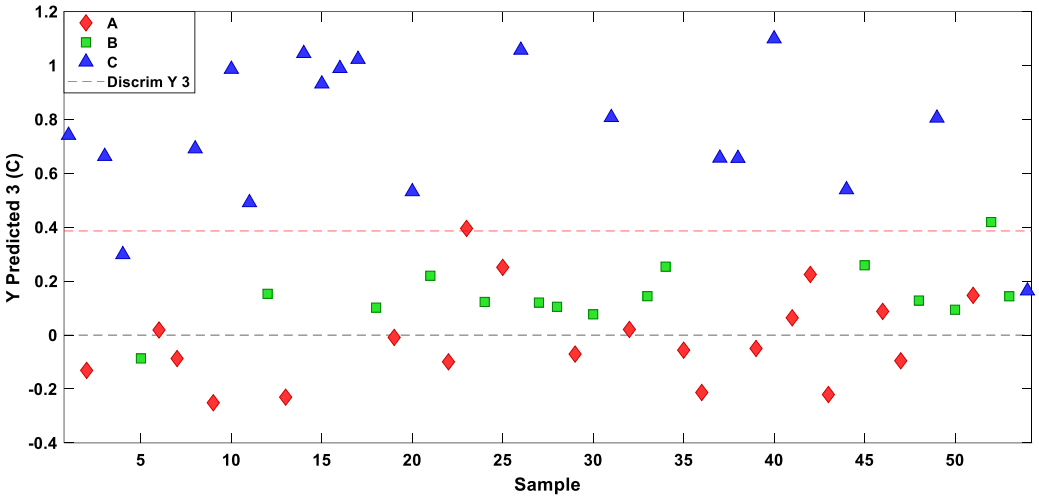

Figure 4.20 PLS-DA prediction score plot of source C (blue) showing the predicted objects

Table 4.2 Performance measures used to evaluate the PLS-DA models of the three spring water sources

Source	Classification accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
A	97.44	97.44	98.68	97.44	97.44	1.74
B	92.62	82	100	100	90.11	7.38
C	94.17	91.89	95.18	89.47	90.67	5.83

The regression vector of source A (**Figure 4.21**) indicates three water absorbance bands at 1362 nm, 1382 nm and 1486 nm. The water band 1362 nm has been assigned to stretching of the hydrogen bond in water molecules, while the band at 1382 nm represents OH stretching (Xantheas, 1995). Water molecules with four intermolecular hydrogen bonds (S_4) are associated with the 1486 nm band (Tsenkova, 2009). Three water absorbance bands, 1382 nm, 1412 nm and 1461 nm, were exhibited by the regression vector of source B (**Figure 4.22**). The 1382 nm band is related to the OH stretch of the first overtone of water (Xantheas, 1995). Free water (S_0) and trimer (S_2) molecules are linked to the 1412 nm and 1461 nm bands, respectively (Franks, 1973; Segtnan et al., 2001). The regression vector of source C (**Figure 4.23**) exhibited two bands at 1380 nm and 1462 nm. These bands have

been assigned to the OH stretch of the first overtone of water and water trimers (S_2), respectively (Franks, 1973; Xantheas, 1995).

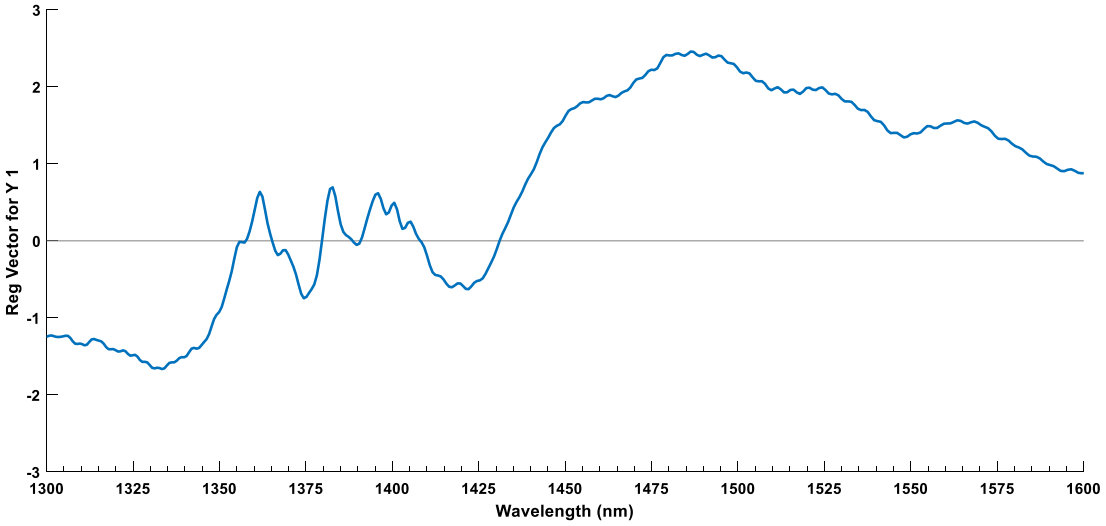

Figure 4.21 PLS-DA regression vector of source A in the 1300 – 1600 nm wavelength range.

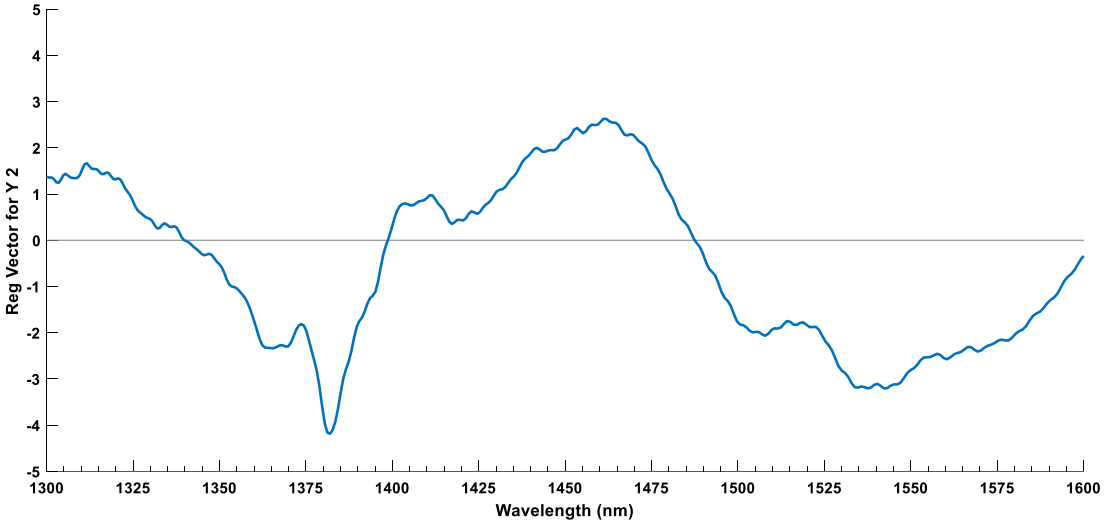

Figure 4.22 PLS-DA regression vector of source B in the 1300 – 1600 nm wavelength range.

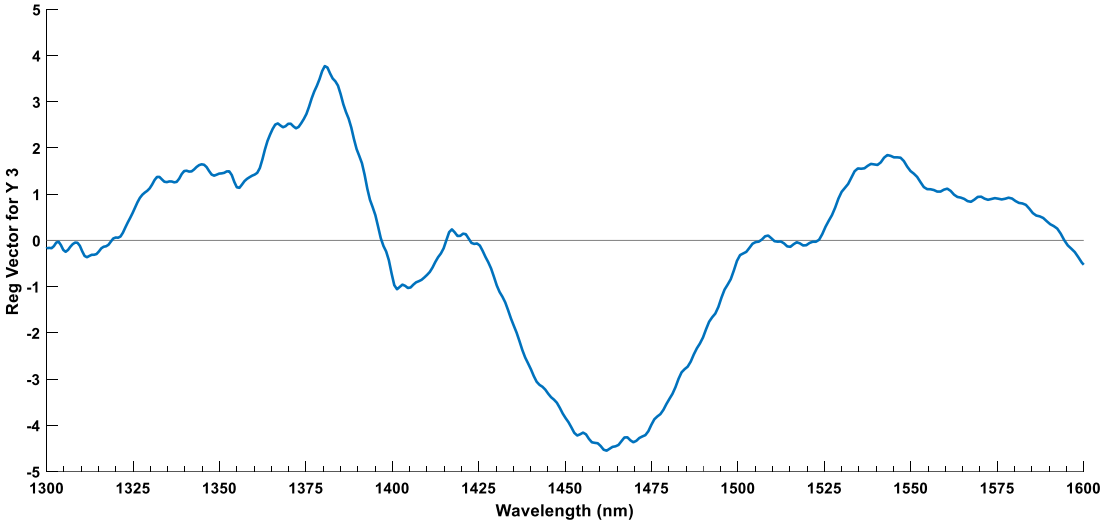

Figure 4.23 PLS-DA regression vector of source C in the 1300 – 1600 nm wavelength range

4.2.3 Aquagrams

The difference in the water spectral patterns of the three sources was visually interpreted with the use of an aquagram (**Figure 4.24**). The aquagram was constructed with the water absorbance bands identified in sections 4.2.1 and 4.2.2.

Typically, the right-hand side of the aquagram (1342 nm - 1426 nm) represents water with weak hydrogen bonds and free water molecules, which can take part in the hydration of solutes (Muncan *et al.*, 2020). The left-hand side of the aquagram (1438 nm – 1515 nm) refers to the different hydrogen-bonded water clusters (S_1, S_2, S_3, S_4) and strongly bound water molecules (Muncan *et al.*, 2020).

Figure 4.24 Aquagrams visualising the spectral patterns of source A (blue), B (red) and C (green).

The spectral pattern of source A is characterised by the high absorbances at 1452 nm, 1462 nm, 1478 nm and 1488 nm. These wavelengths have been assigned to water dimers (S_1), trimers (S_2), tetramer (S_3) and pentamers (S_4), respectively. This indicates that the water molecules present are strongly bonded to each other.

The spectral pattern of source B is similar to that of source A. However, it has lower absorbance values at 1452 nm, 1462 nm, 1478 nm, 1488 nm and a higher absorbance value at 1412 nm. These wavelengths have been assigned to dimers (S_1), trimers (S_2), tetramer (S_3), pentamers (S4) and water molecules with no intermolecular hydrogen bonds (S_0), respectively. Indicating source B contains moderate concentrations of both strongly and weakly bound water.

Source C is characterised by high absorbances at 1342 nm (v_3), 1364 nm (OH stretch), 1374 nm ($v_1 + v_2$), 1382 nm (OH stretch) and 1412 nm (S_0). These wavelengths correspond to free water and molecules with no hydrogen bonds to other molecules. Indicating that source C contains water molecules that are weakly bonded and free to interact with solutes present in the water.

The difference between the three sources can also be explained by the variation in the mineral content of the sources (**Table 4.3**). Source C has a high mineral content, whereas sources A and B have lower mineral contents. Ions present in the water will have an influence on the hydrogen-bonded network of the water (Gowen *et al.*, 2013). Inorganic salts such as NaCl, KCl, $MgCl_2$ and $AlCl_3$ are known to weaken or strengthen the hydrogen-

bond network influence the hydrogen-bond network and are referred to as structure makers or breakers (Gowen *et al.*, 2013). The variations in the concentration of ions present in water will have an influence on the water spectrum (Kovacs *et al.*, 2016).

Table 4.3 Typical mineral content, pH, alkalinity and TDS of the three different sources provided on the labels (mg/L)

Source	A	B	C
Calcium	2	0.4	5.9
Magnesium	1	0.2	2.4
Sodium	8	3.3	46.2
Potassium	1	<0.1	0.5
chloride	13.2	6.5	<5
sulphate	4.4	1.4	10.8
Alkalinity	5.4	3.4	104
Nitrate	0.2	<0.1	<0.1
Fluoride	<1	<0.1	1
TDS	48	19	193
pH	6	5.1	8.5

The results reported were similar to a study by Munćan *et al.* (2014), who applied aquaphotomics to discriminate between mineral waters. The authors indicated that each mineral water sample produced a unique water spectral pattern and that the differences were likely due to the variations in their mineral composition.

4.2.4 Conclusion

It is evident from the water spectral patterns of the three sources, that each water source has a unique arrangement of water species. The differences in the water spectral patterns of the different sources are most likely due to variation in the mineral composition of the three different water sources. Further analysis is however required to determine the effect that the combination of mineral salts and their concentrations will have on the water spectrum, as to date only single salt solutions have been investigated. These studies have proved that it is possible to detect changes in concentrations of the single solutes (Gowen *et al.*, 2013; Gowen *et al.*, 2015; Kovacs *et al.*, 2016).

NIR spectroscopy combined with the aquaphotomics approach could distinguish between the three spring water sources. Each water source had a unique combination of different water clusters. This resulted in each spring water source having a unique water spectral pattern.

4.3 Different types of bottled water

This data set consisted of different mineral and spring water samples, which were bottled at different sources. The aim of this section was to differentiate between mineral and spring water, irrespective of the source.

4.3.1 Spectral analysis

The mean spectrum of each spring water source, in the 1300 – 1600 nm range, is shown in **Figure 4.25**. The mean spectra of the three water sources appear to be identical, with one dominant band at 1450 nm. This absorbance band, 1450 nm, was attributed to the first overtone of the OH stretching vibration (Luck, 1974). Spectral subtraction was performed, by subtracting the average spectrum of deionised water from the average spectra of the spring and mineral water. The resulting difference spectra are shown in **Figure 4.26**.

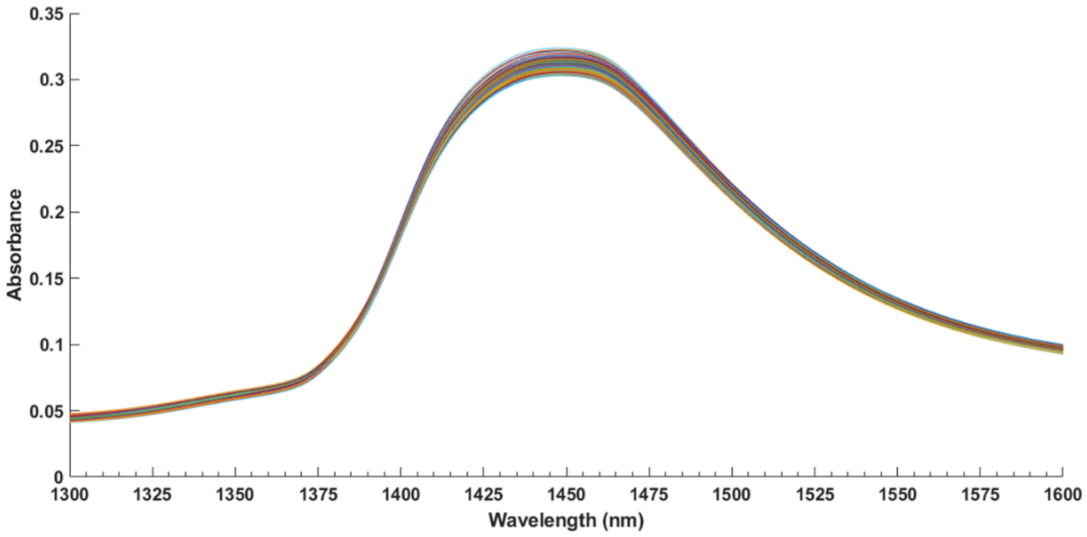

Figure 4.25 Mean spectra of the different brands of bottled water samples.

The difference spectra (**Figure 4.26**) of the spring and mineral water exhibited water absorbance bands at 1373 nm, 1452 nm, 1462 nm and 1472 nm. The water band 1373 nm has been assigned to symmetric stretching and bending vibrational modes of the first overtone of water (v_1+v_2) (Siesler, 2006). While the bending and asymmetric stretching vibrational (v_2+v_3) modes of the first overtone of water has been assigned to 1452 nm (Cattaneo *et al.*, 2009; Siesler, 2006). Water molecules with two (S_2) and three (S_3) hydrogen bonds have been linked to the bands at 1462 nm and 1472 nm, respectively (Franks, 1973; Siesler, 2006).

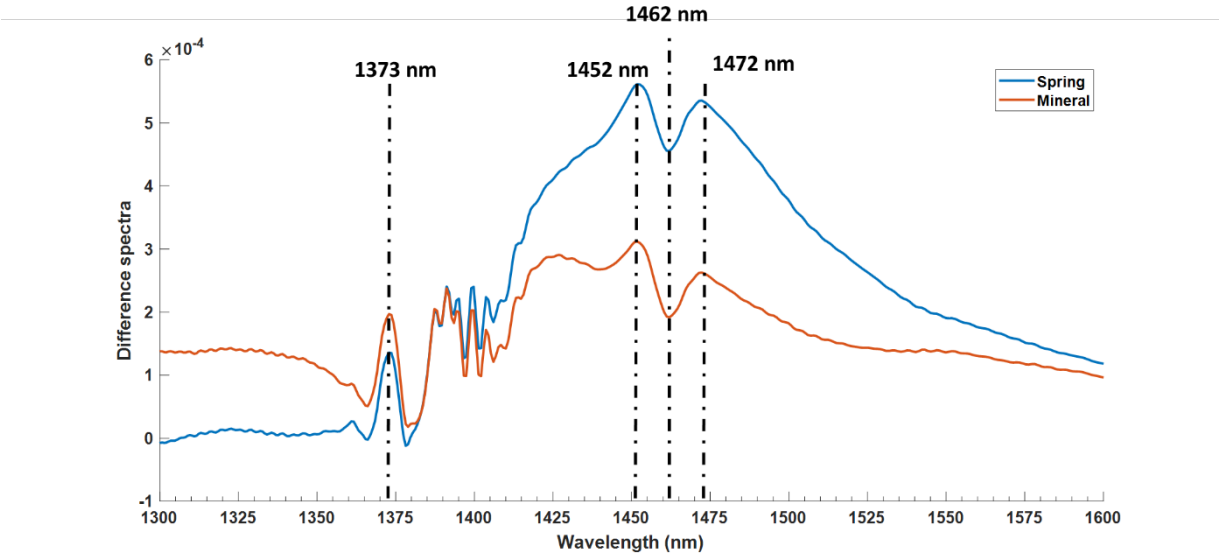

Figure 4.26 Difference spectra (Savitzky-Golay smoothing with 2nd order polynomial and 21 points) of the different brands of bottled water samples.

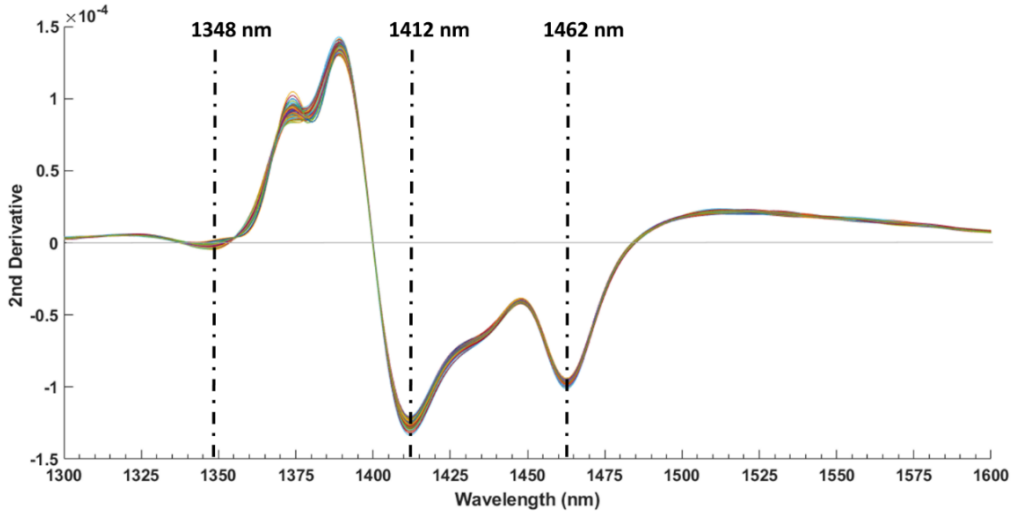

Figure 4.27 Second derivative (calculated with a Savitzky-Golay filter using 2nd order polynomial and 21 points) spectra in the first overtone of water (1300 – 1600 nm) of the different bottled water samples.

The 2nd derivative spectra (**Figure 4.27**) of the spring and mineral water indicated three water absorbance bands at 1348 nm, 1412 nm and 1462 nm. The band at 1348 nm is linked to the asymmetric stretching of the water molecule (Siesler, 2006). The bands at 1412 nm and 1462 nm have been assigned to water molecules with different numbers of hydrogen bonds. The region of 1412 nm is associated with water molecules with little or no

intermolecular hydrogen bonds (S_0) (Segtnan et al., 2001). The water absorbance band, 1462 nm, has been assigned to water molecules with two hydrogen bonds (S_2).

4.3.2 Multivariate data analysis

4.3.2.1 Principal component analysis

Principal component analysis was performed to explore the data by examining the differences between mineral and spring water. The resulting score and loading plots are shown in **Figures 4.28 – 4.35**.

The score plots (**Figures 4.28 – 4.31**) showed no separation between the different spring water sources. PC 1 accounted for 68% of the variance in the data, whereas PC2, PC3, PC4 and PC5 accounted for 20%, 7%, 3% and 1%, respectively. This indicates that most of the variance in the data set is explained by the first two principal components.

The lack of separation between the water sources indicates that there are similarities in their spectral pattern. However, variation in the chemical composition of the water sources can lead to small spectral differences.

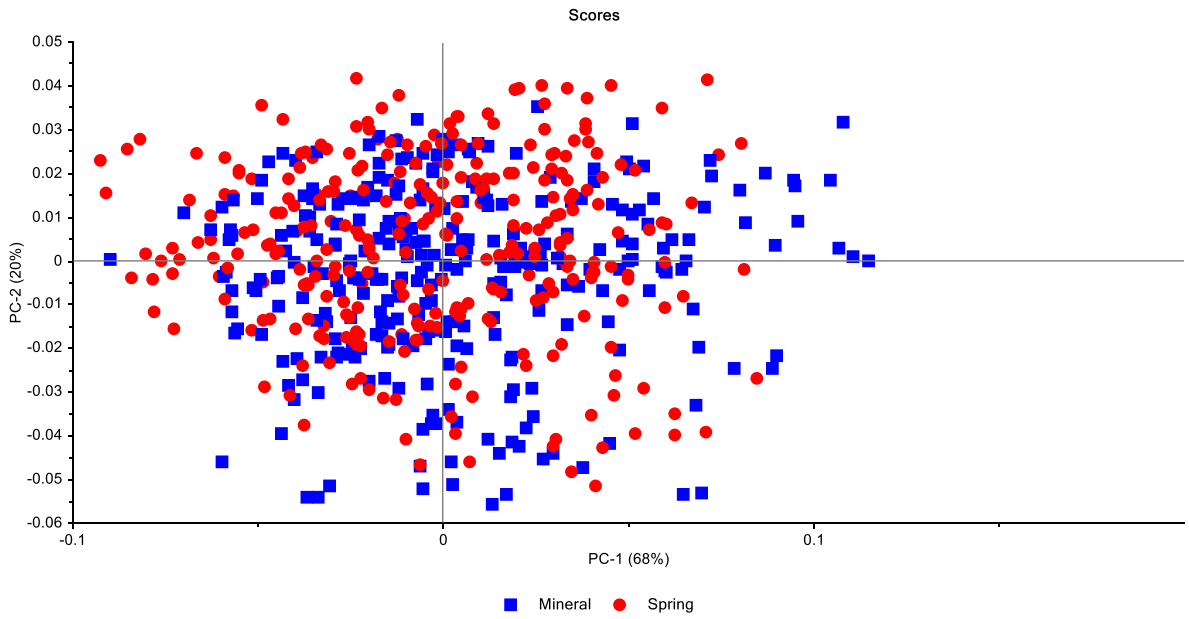

Figure 4.28 Principal component analysis score plot (PC1 (68 %) vs. PC2 (20 %)) of bottled water.

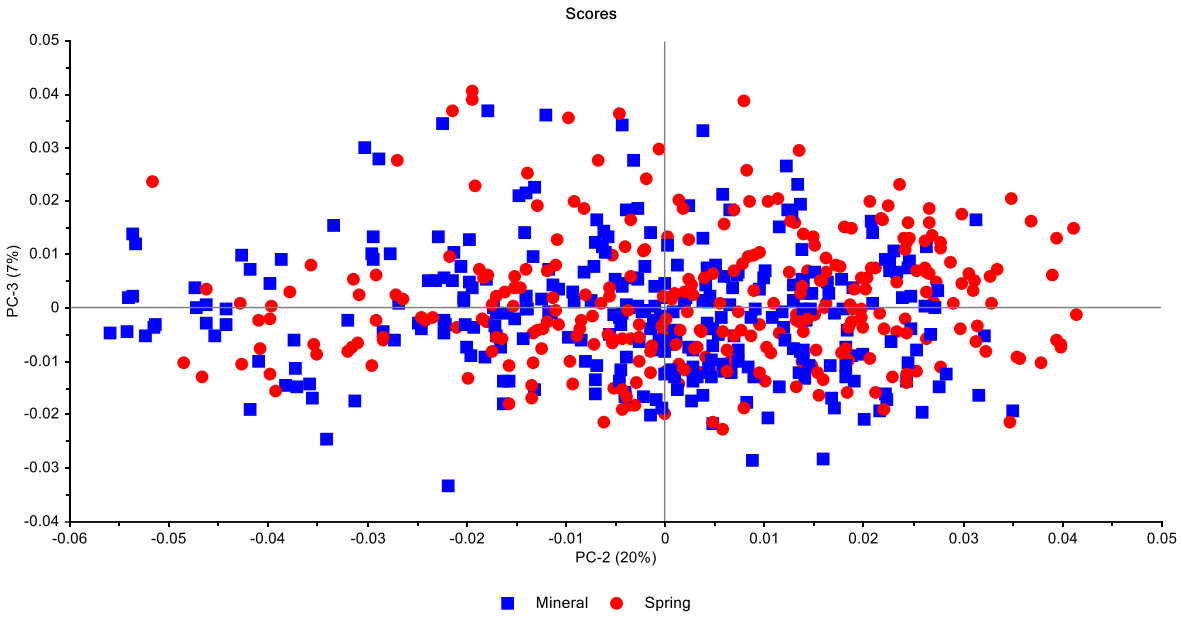

Figure 4.29 Principal component analysis score plot (PC2 (20 %) vs. PC3 (7 %)) of bottled water.

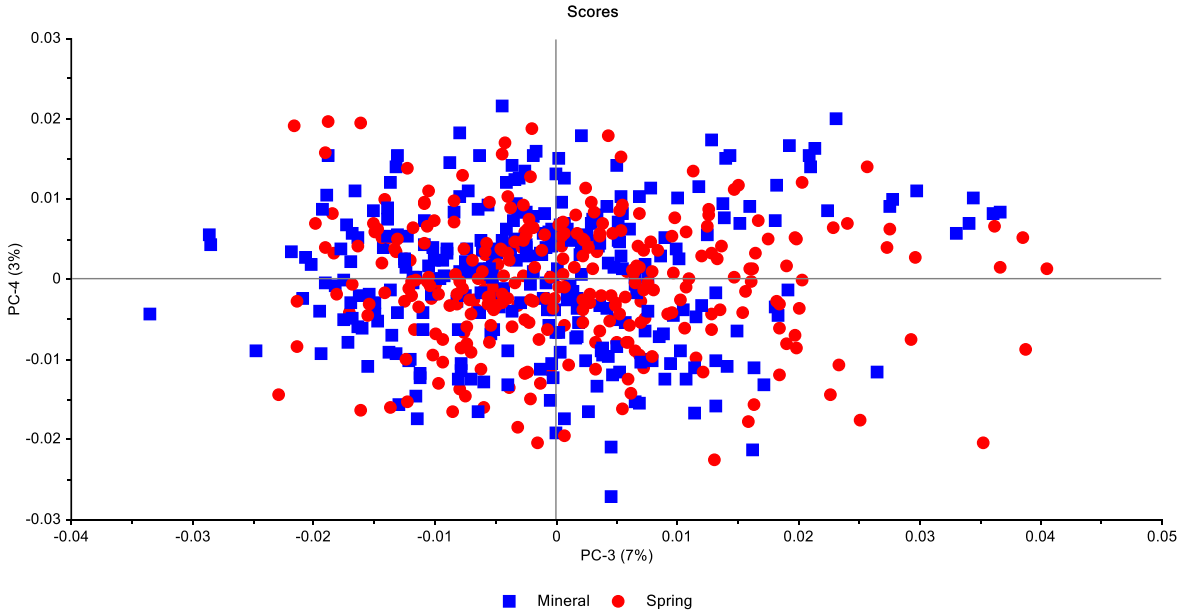

Figure 4.30 Principal component analysis score plot (PC3 (7 %) vs. PC4 (3 %)) of bottled water.

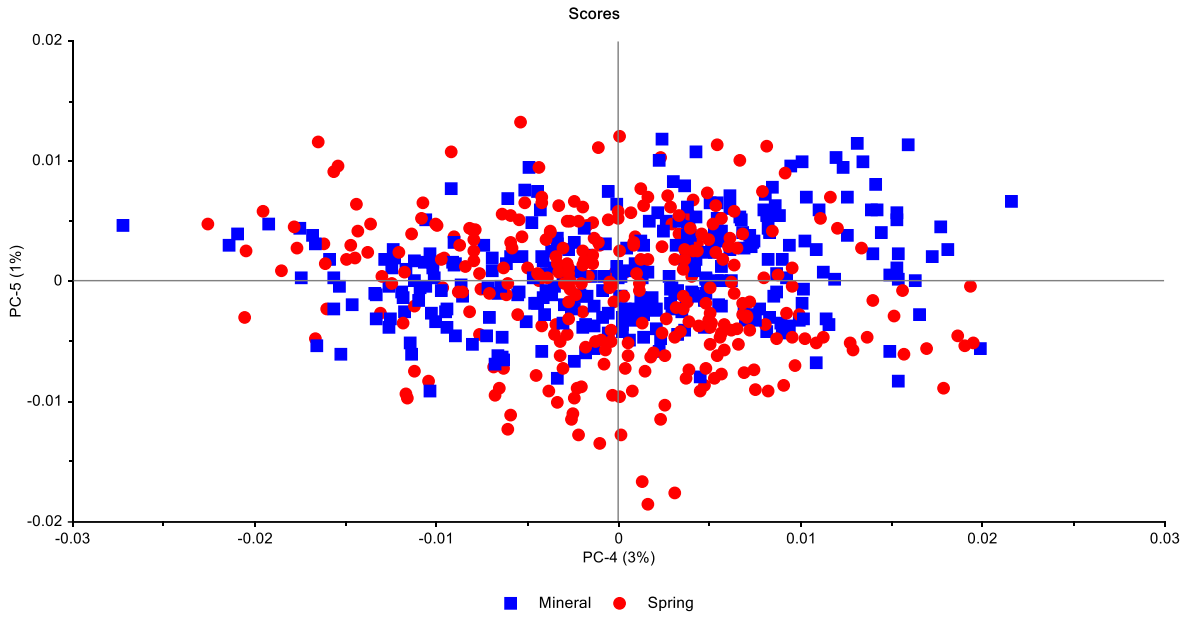

Figure 4.31 Principal component analysis score plot (PC4 (3 %) vs. PC5 (1 %)) of bottled water.

The variance seen in the direction of PC 1 (**Figure 4.28**) may be attributed to the OH stretching and bending (v_2) of the first overtone of water, while the different water clusters can be the result of the variance in the direction of PC 2 (**Figure 4.29**). The score plots of PC 2 vs PC 3 (**Figure 4.30**) and PC 3 vs PC 4 (**Figure 4.31**) showed no separation between the three different water sources. The minimal separation between the water sources indicates that there are similarities between the mineral and spring water samples.

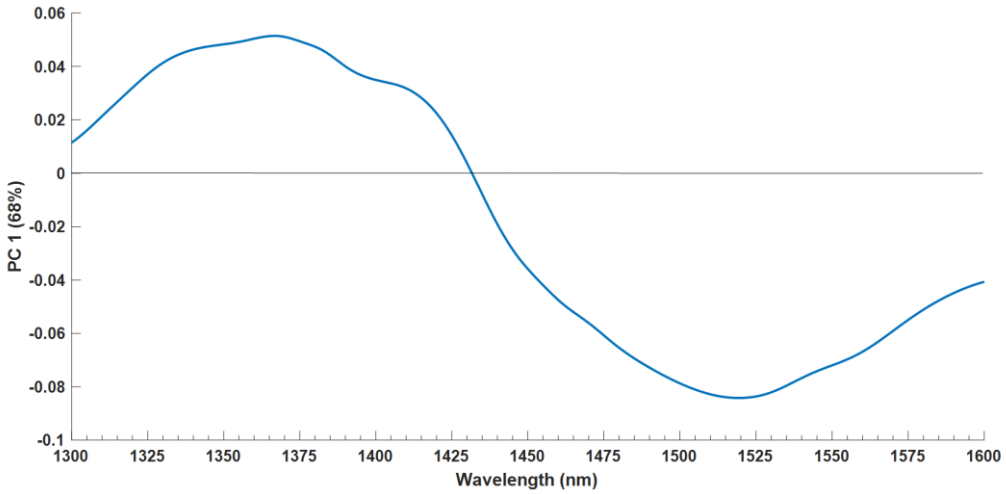

Figure 4.32 PCA loadings line plot for PC 1 (68%).

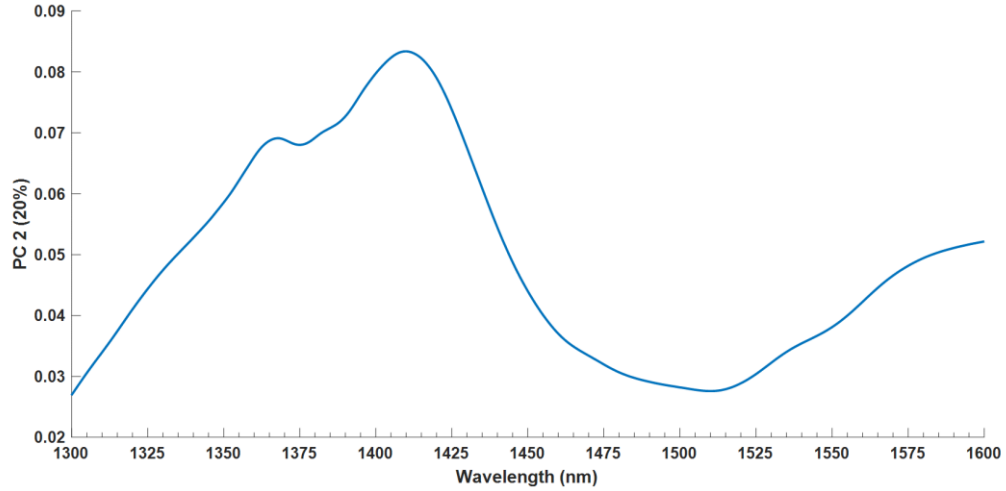

Figure 4.33 PCA loadings line plot for PC 2 (20%).

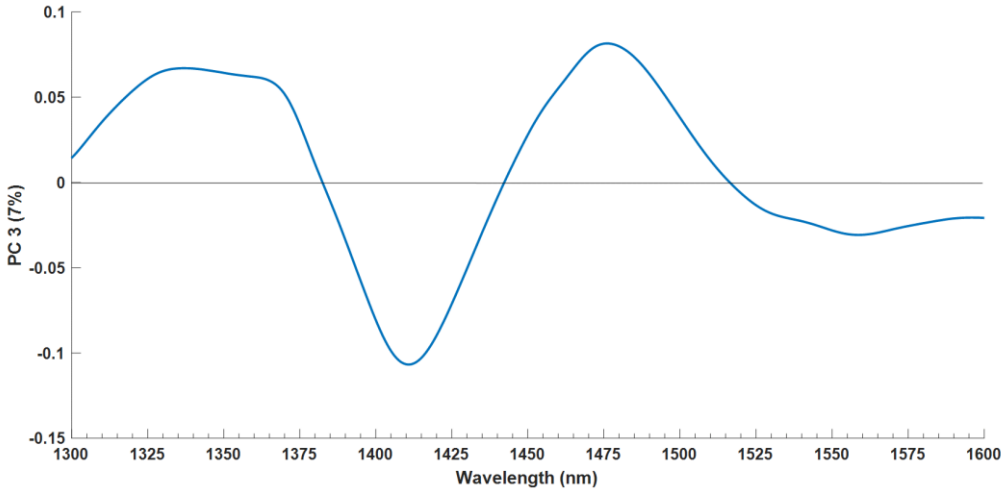

Figure 4.34 PCA loadings line plot for PC 3 (7%).

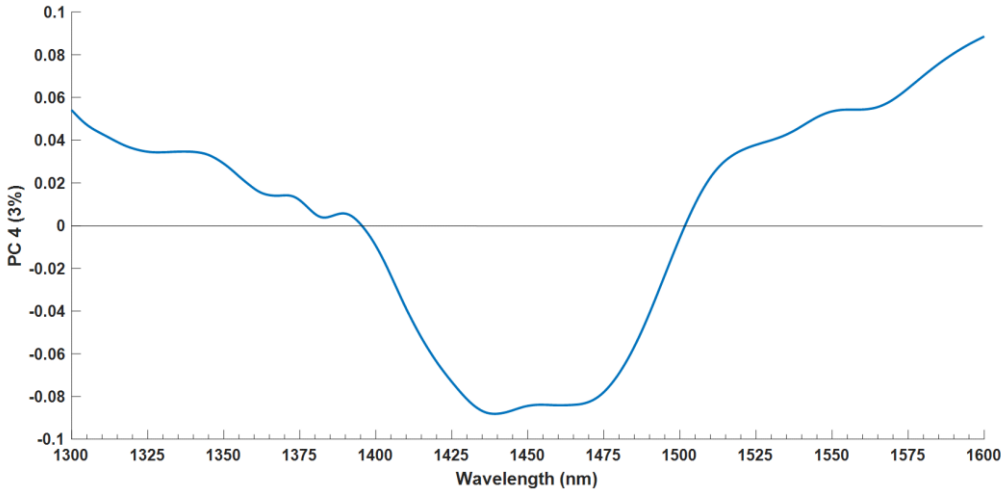

Figure 4.35 PCA loadings line plot for PC 4 (3%).

The PC 1 loadings plot (**Figure 4.32**) exhibited two bands at 1365 nm and 1518 nm. The water band 1365 nm has been assigned to stretching of the hydrogen bond in water molecules (Xantheas, 1995). The 1518 nm water band has been assigned to bending (v_2) vibration mode of the first overtone of water (Siesler, 2006). The loadings plot of PC 2 (**Figure 4.33**) exhibited bands at 1366 nm and 1410 nm. The water band 1366 nm has been assigned to stretching of the hydrogen bond in water molecules (Xantheas, 1995). Water molecules with little or no intermolecular hydrogen bonds (S_0) are linked to the 1410 nm band (Segtnan et al., 2001).

The third PC loadings plot (**Figure 4.34**) exhibited three bands at 1364 nm, 1410 nm and 1476 nm. The band at 1364 nm has been assigned to stretching of the hydrogen bond in water molecules (Xantheas, 1995). The 1412 nm water band is linked to water molecules with little or no intermolecular hydrogen bonds(S_0) (Segtnan et al., 2001). The loadings plot of PC 4 (**Figure 4.35**) exhibited bands at 1341 nm, 1372 nm, 1438 nm, 1467 nm and 1518 nm. The band at 1341 nm is linked to the asymmetric stretching of the water molecule (Siesler, 2006). The band at 1374 nm is linked to the symmetrical and asymmetrical stretching of water molecules (Siesler, 2006). The 1438 nm and 1467 nm bands are assigned to water molecules with one hydrogen bonds (S_1) and two hydrogen bonds (S_2), respectively (Cattaneo *et al.*, 2009; Franks, 1973; Siesler, 2006). The 1518 nm water band has been assigned to bending vibration mode of the first overtone of water (Siesler, 2006).

The loadings plot of PC 5 (**Figure 4.36**) exhibited bands at 1343 nm, 1363nm, 1384 nm, 1422 nm, 1488 nm and 1518 nm. The band at 1343 nm is linked to the asymmetric stretching of the water molecule (Siesler, 2006). The water band 1363 nm has been assigned to stretching of the hydrogen bond in water molecules (Xantheas, 1995). The 1384 nm band corresponds to the OH stretching of molecules in the first overtone of water (Xantheas, 1995). The 1422 nm band is referred to as the water hydration band (Tsenkova, 2009; Williams, 2009). Water molecules with four hydrogen bonds have been assigned to the 1488 nm band (S_2) (Franks, 1973). The 1518 nm water band has been assigned to bending vibration mode of the first overtone of water (Siesler, 2006).

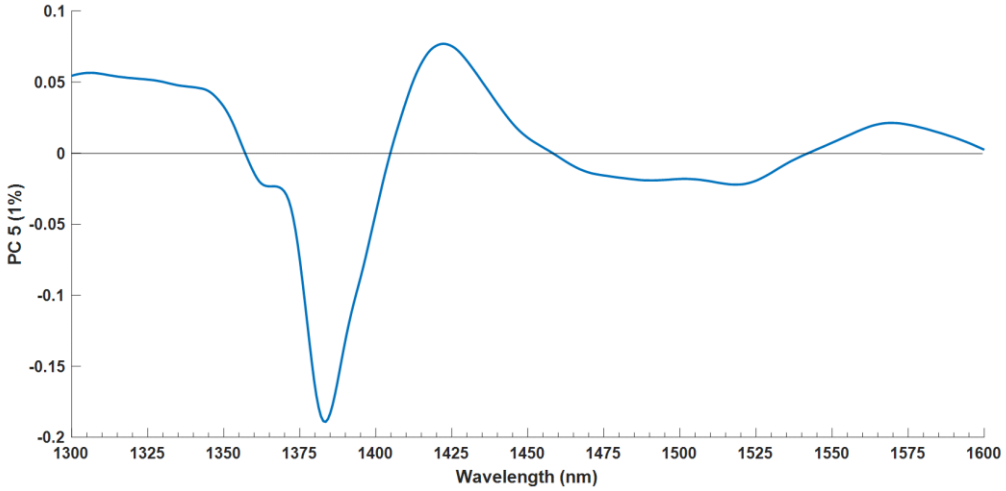

Figure 4.36 PCA loadings line plot for PC 5 (1%).

4.3.2.2 Partial least squares discriminant analysis (PLS – DA)

The PLS-DA model of mineral and spring water provided acceptable discrimination results, with an overall classification accuracy and misclassification rate of 72.62 % and 27.38 %, respectively. The overall performance measures of the PLS-DA models are presented in **Table 4.4**. For the mineral water class, 150 of the 208 samples were correctly classified and 57 of the spring water class was misclassified as mineral water. Out of the 212 spring water samples, 155 samples were correctly classified and 57 samples were misclassified as mineral water.

Table 4.4 The overall performance measures of the calibration, cross-validation and validation PLS-DA models of mineral and spring water

Model	Classification accuracy (%)	Misclassification rate (%)
Calibration	72.62	27.38
Cross – validation	70.48	29.52
Validation	74.44	25.56

The PLS-DA 3D score plot of LV1 vs LV2 vs LV3 accounted for 45.36 %, 22.29 % and 19.63 of the variance, respectively. A large amount of overlap between the classes is observed in **Figures 4.37 – 4.38**, which is accompanied by a large number of samples being incorrectly classified. The lack of separation between the mineral and spring water is likely due to the similarities in the physico-chemical properties of the samples.

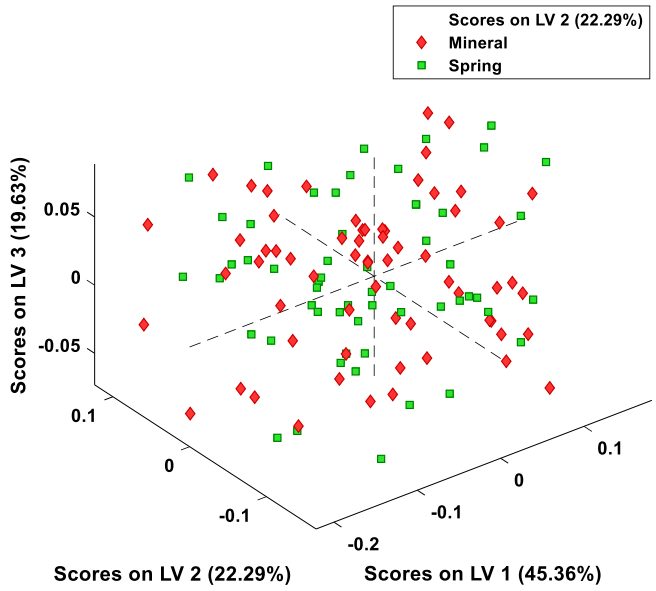

Figure 4.37 3D score plot (LV1 (45.36 %) vs LV2 (22.29 %) vs LV3 (19.63 %)) of mineral (red) and spring (green) water.

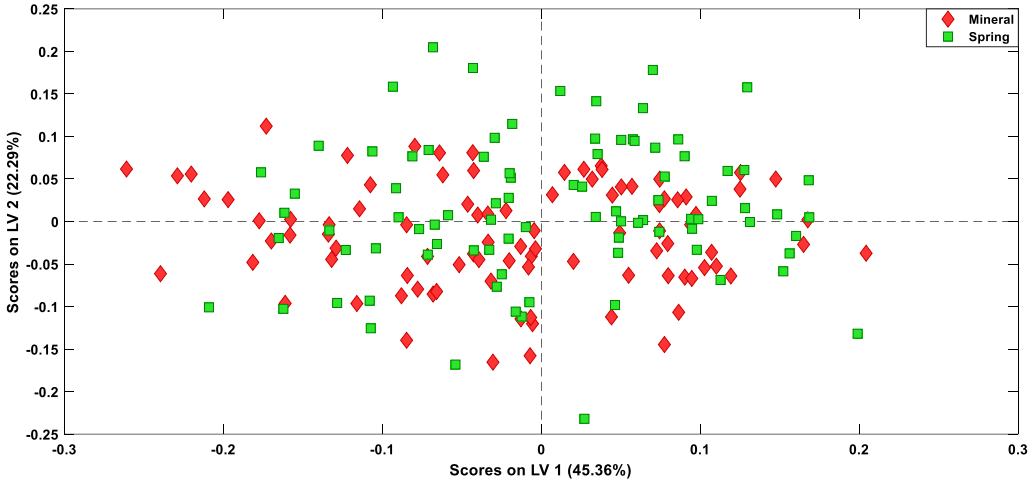

Figure 4.38 PLS-DA score plot (LV1 (45.36%) vs LV2 (22.29%)) of mineral (red) and spring water (green).

The PLS-DA prediction score plots (**Figure 4.39 – 4.40**) illustrated that most of the mineral and spring water samples were correctly predicted. The classification accuracy of the mineral and spring water models was both 72.66 %. The sensitivity, specificity and F1 scores for both mineral and spring water is above 72 % (**Table 4.5**). This indicates that the ability of the model to classify the water types is acceptable. The spectral classification of the water types can be attributed to the variation in mineral content of the water. The

misclassification of the water types is likely due to the mineral content of the water types being similar (**Table 4.6**).

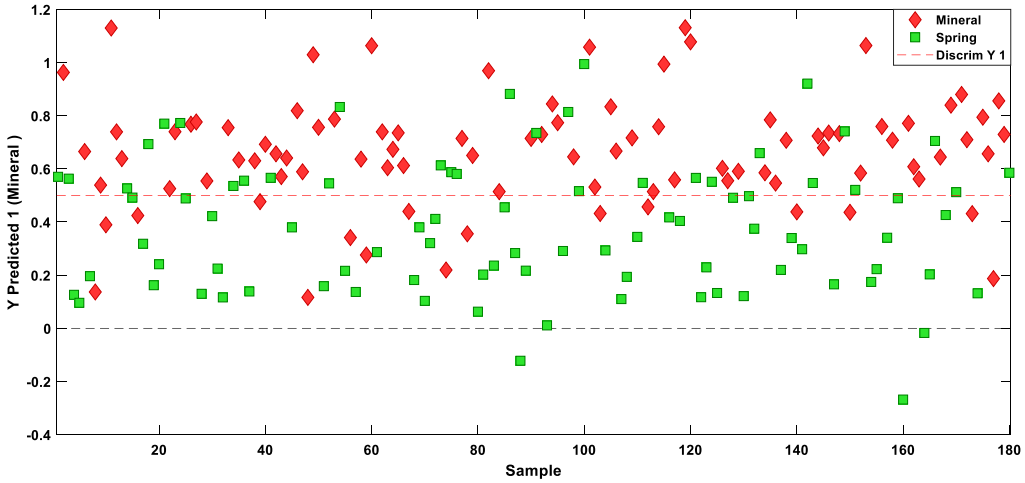

Figure 4.39 PLS-DA prediction score plot of mineral water (red) showing the predicted samples.

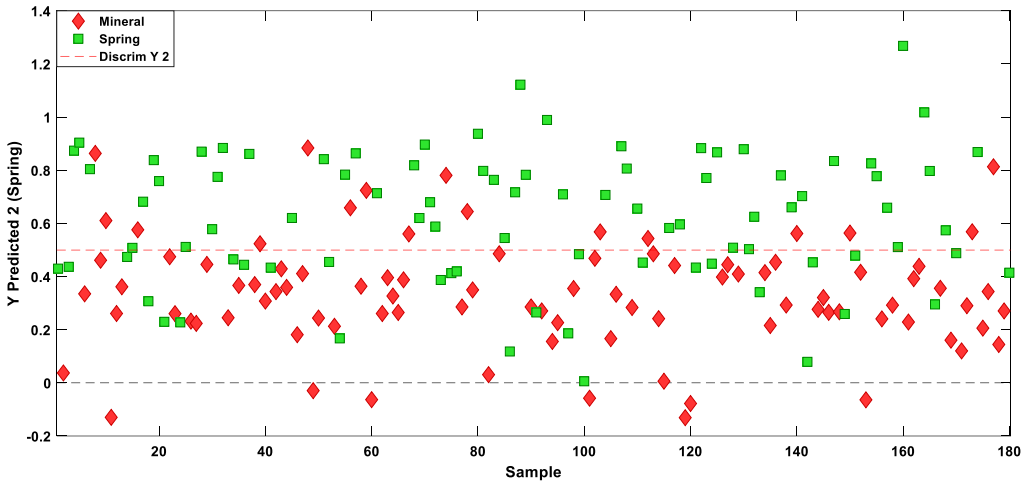

Figure 4.40 PLS-DA prediction score plot of spring water (green) showing the predicted samples.

The PLS-DA regression vectors (**Figure 4.41-4.42**) of mineral and spring water exhibited four bands at 1364 nm, 1384 nm, 1429nm and 1484 nm. The 1364 nm and 1384 nm bands are linked to the OH stretch of water in the first overtone (Xantheas, 1995). The 1429 nm band is referred to as the water hydration band (Tsenkova, 2009; Williams, 2009). Water molecules with four intermolecular hydrogen bonds correspond to the 1484 nm band (Tsenkova, 2009).

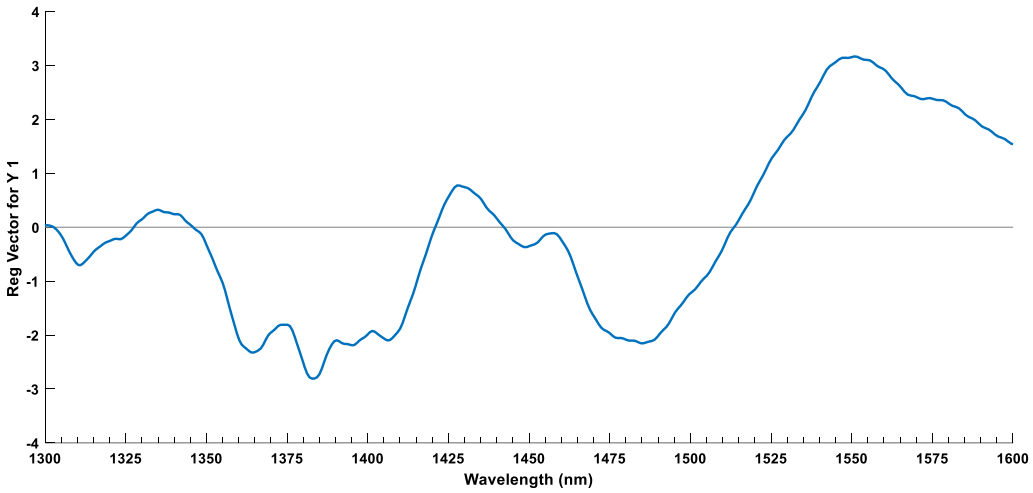

Figure 4.41 PLS-DA regression vector of mineral water in the 1300 – 1600 nm wavelength range.

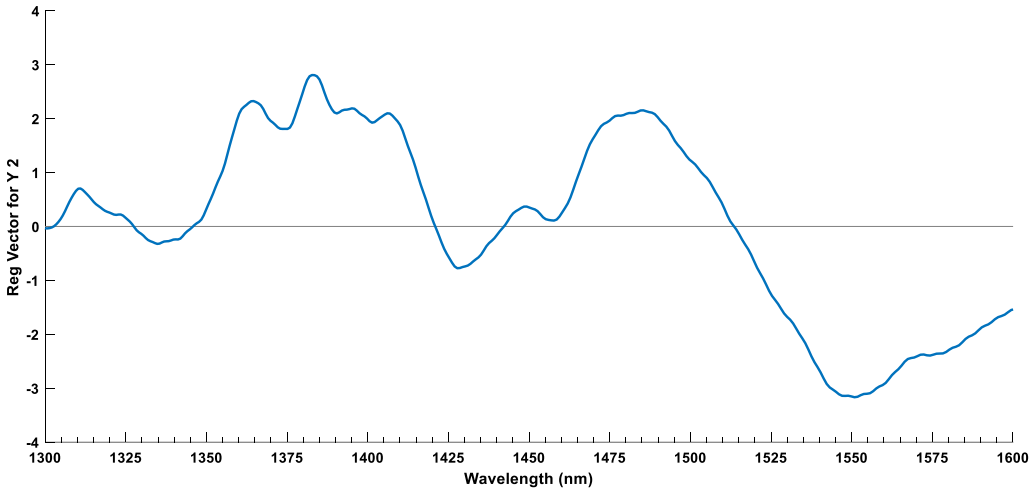

Figure 4.42 PLS-DA regression vector of spring water in the 1300 – 1600 nm wavelength range.

Table 4.5 Performance measures used to evaluate the PLS-DA models of mineral and spring water

Type	Classification accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
Mineral	72.62	72.46	72.77	72.12	72.29	27.38
Spring	72.62	72.77	72.46	73.11	72.94	27.38

4.3.3 Aquagram

The difference between the mineral and spring water was visualised with an aquagram (**Figure 4.43**). The aquagram was constructed with the water absorbance bands which consistently appeared in sections 4.3.1 and 4.3.2. A large difference between the mineral and spring water is observed based on their respective aquagrams. The water spectral pattern of the mineral water has normalised absorbance values which are below zero, while the spring waters' spectral pattern has normalised absorbance values above zero.

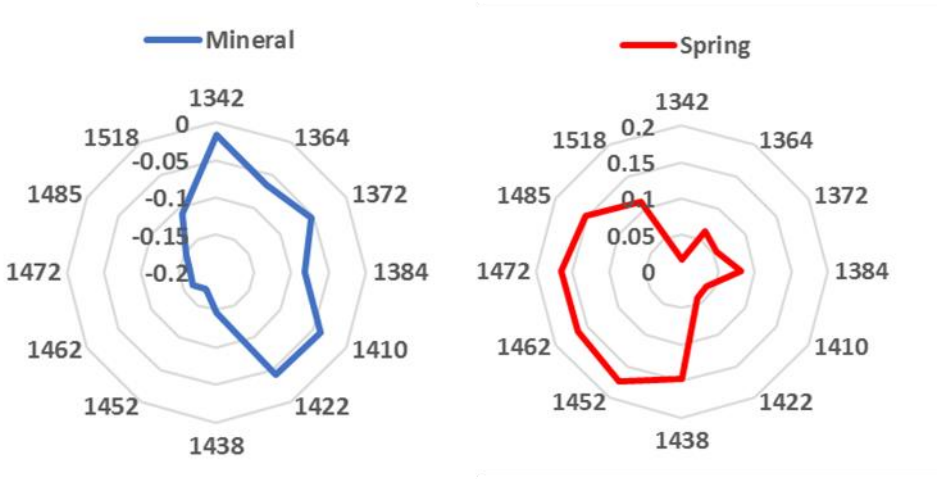

Figure 4.43 Aquagrams visualising the spectral patterns of the mineral (blue) and spring (red) water.

The aquagram of the mineral water indicated high absorbance values at 1410 nm and 1422 nm, which correspond to water molecules with no intermolecular hydrogen bonding (Segtnan et al., 2001) and the water hydration band (Tsenkova, 2009; Williams, 2009), respectively. A noticeable difference is seen between the water spectral patterns of the mineral and spring water. Mineral water has lower absorbance values at all the WAMACS compared to spring water.

Spring water had high absorbance values at 1438nm, 1462 nm, 1472 nm and 1485. These WAMACS correspond to water molecules with one (S_1), two (S_2), three (S_3) and four (S_4) intermolecular hydrogen bonds. Therefore, the spring water can be considered strongly bound water.

The main difference between mineral and spring water is that the mineral content of mineral water is required to remain constant, while variations in the mineral content of spring

water can occur (Ashurst *et al.*, 2017). Therefore, variations in the mineral content of spring water could have resulted in the misclassification rate of spring water as mineral water.

These results are similar to that found in literature, as previous studies indicated that it is possible to classify different water types (Munćan *et al.*, 2014; Tanaka *et al.*, 1997). The water spectrum is very sensitive, therefore, small variances between samples can be detected (Tanaka *et al.*, 1997). These authors also indicated that further research is required to understand the influence the mineral content will have in the water spectral pattern.

Table 4.6 Typical mineral content, pH, alkalinity and TDS of the different mineral and spring water samples

	M1	M2	M3	S1	S2	S3
Calcium	41	41	0.7	0.5	<0.5	10
Magnesium	25.5	25.5	2.7	1.2	<0.5	10
Sodium	6.2	6.2	19.4	10.2	5	3
Potassium	0.7	0.7	0.8	0.8	<1	1
Chloride	7	7	42	19	9	2
Sulphate	8	8	1.8	2	<5	4
Alkalinity (CaCO3)	203	203	1	3	10	65
Nitrate	7	7	0.7	<1	0.2	1
Flouride	<0.2	<0.2	0.2	<1	<0.05	<0.1
TDS	225	225	117	61	<26	83
pH	7.5	7.5	5.6	5	5.2	7.3

4.3.4 Conclusion

With the use of the aquaphotomics approach, it was possible to distinguish between mineral and spring water. The two water types produced two completely different water spectral patterns, indicating that each water type had a different configuration of water species.

4.4 River water and different filtration materials

This data set consisted of river water (RW) that had been collected on ten different days and filtered using different materials (pine biochar (PN), black wattle biochar (BW), activated charcoal (AC) and sand (SA)). The aim of this experiment was to determine whether it is possible to monitor changes in the water spectral patterns due to different filtration media. Furthermore, it was determined whether differences in river water quality over time could be observed through changes in the water spectral patterns.

4.4.1 Filtered river water

4.4.1.1 Spectral analysis

The mean spectrum of the river and filtered water, in the 1300 – 1600 nm range, is shown in **Figure 4.44**. The mean spectra of the river and filtered water appear to be identical, with one dominant band at 1450 nm. This absorbance band, 1450 nm, is attributed to the first overtone of the OH stretching vibration (Luck, 1974).

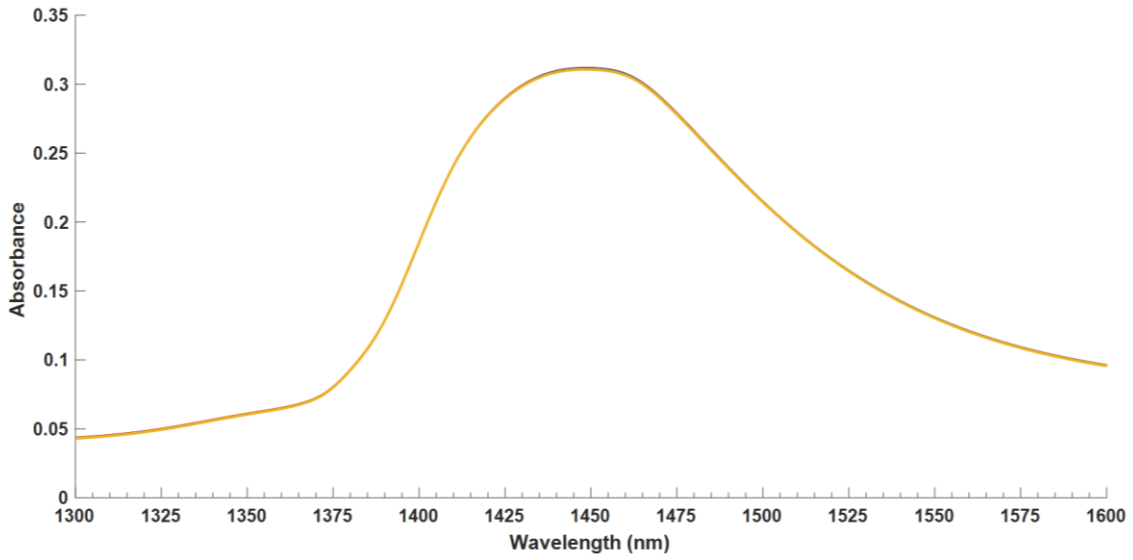

Figure 4.44 Mean spectra for river water and the different filtrate samples.

The difference spectra was calculated by subtracting the average spectrum of deionised water from the average spectra of the river and filtered water. The difference spectra (**Figure 4.45**) of the river and filtered water indicated water absorbance bands at 1380 nm, 1426 nm and 1472 nm. The OH stretch in the first overtone of water is associated with the 1380 nm band (Xantheas, 1995). The 1426 nm band is referred to as the water hydration band (Tsenkova, 2009; Williams, 2009). Water molecules with three intermolecular hydrogen bonds are linked to the 1472 nm band (Franks, 1973; Siesler, 2006).

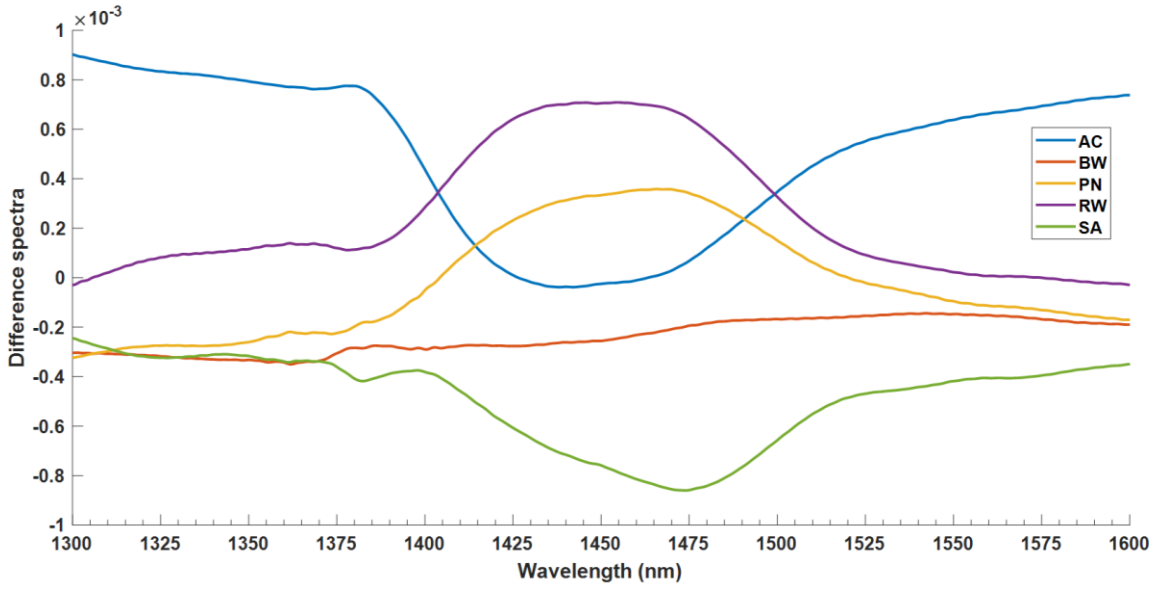

Figure 4.45 Difference spectra (Savitzky-Golay smoothing with 2nd order polynomial and 21 points) of filtered and unfiltered water.

Three water absorbance bands at 1348 nm, 1412 nm and 1462 nm were exhibited by the second derivative spectra (**Figure 4.45**). The asymmetric stretching of the first overtone of water has been linked to the 1348 nm (Siesler, 2006). The 1412 nm and 1426 nm bands have been assigned to water molecules with zero (S_0) and two (S_2) intermolecular hydrogen bonds, respectively (Franks, 1973; Segtnan et al., 2001).

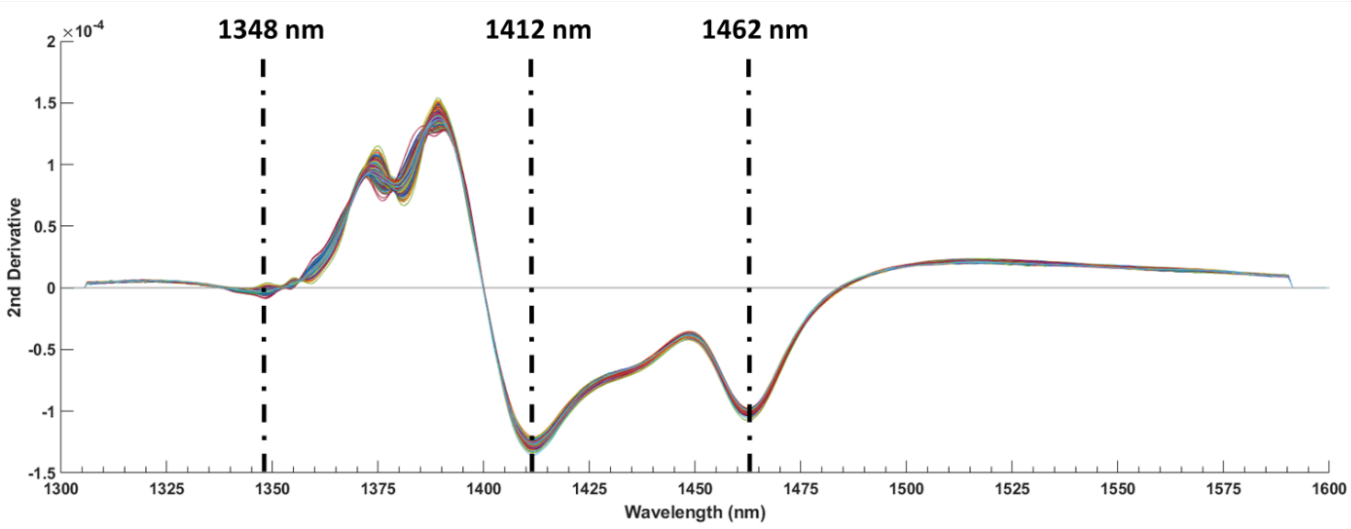

Figure 4.46 Second derivative (calculated with a Savitzky-Golay filter using 2nd order polynomial and 21 points) spectra in the first overtone of water (1300 – 1600 nm) of the river and filtered water samples.

4.4.1.2 Multivariate data analysis

4.4.1.2.1 Principal component analysis

Principal component analysis was performed to explore the data by examining the effect different filter material have on river water. The resulting score and loading plots are shown in **Figure 4.47 -4.55**.

The score plots (**Figure 4.47 – 4.50**) showed no separation between the different river and filtered water samples. PC 1 accounted for 74% of the variance in the data, whereas PC2, PC3, PC4 and PC5 accounted for 14%, 8%, 2% and 1%, respectively. This indicates that most of the variance in the data set is explained by the first two principal components.

The lack of separation between the river and filtered water samples indicates that there are similarities in their spectral patterns. However, variation in the chemical composition of the water sources can lead to small spectral differences.

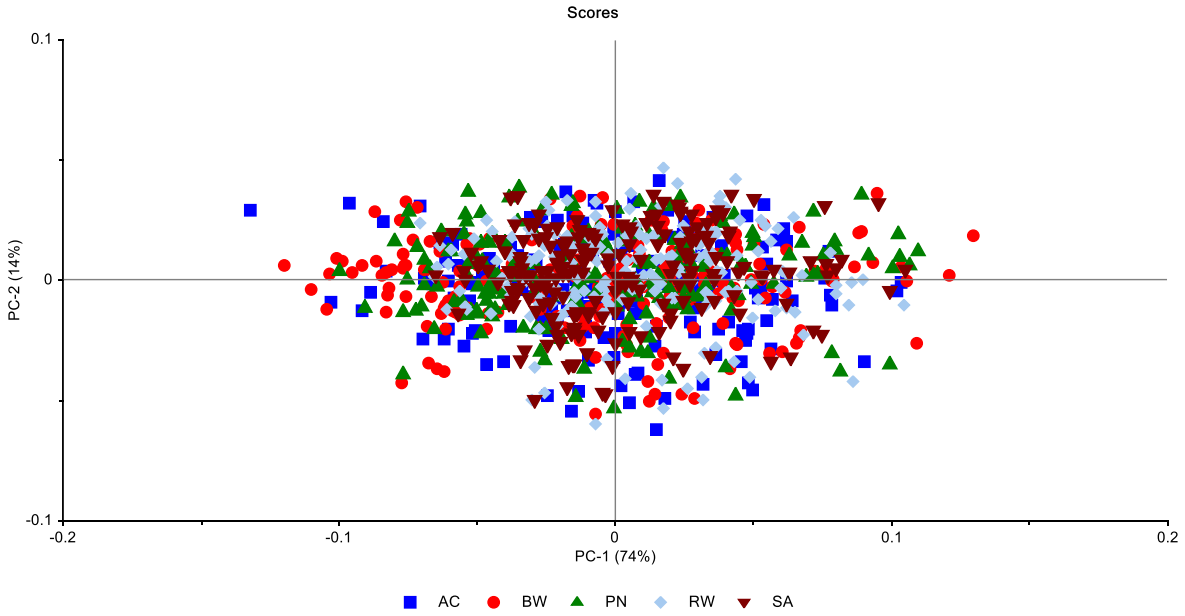

Figure 4.47 Principal component analysis score plot (PC1 (74 %) vs. PC2 (14 %)) of the river (RW) and filtered water (activated (AC), black wattle biochar (BW), pine biochar (PN) and sand (SA)).

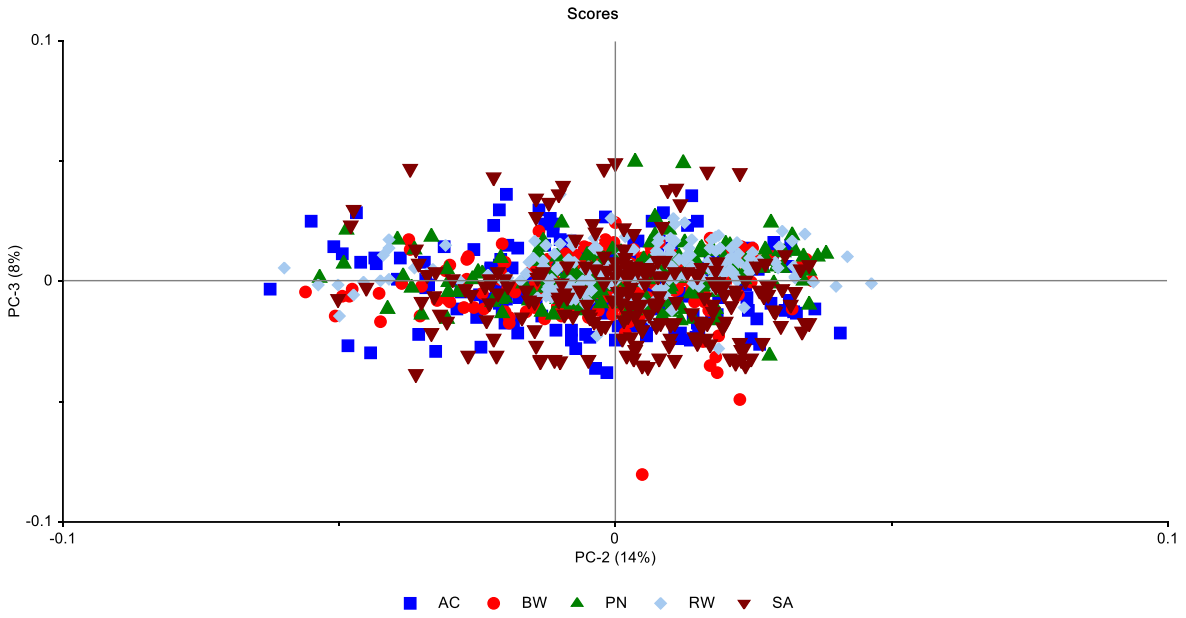

Figure 4.48 Principal component analysis score plot (PC2 (14 %) vs. PC3 (8 %)) of the river (RW) and filtered water (activated (AC), black wattle biochar (BW), pine biochar (PN) and sand (SA)).

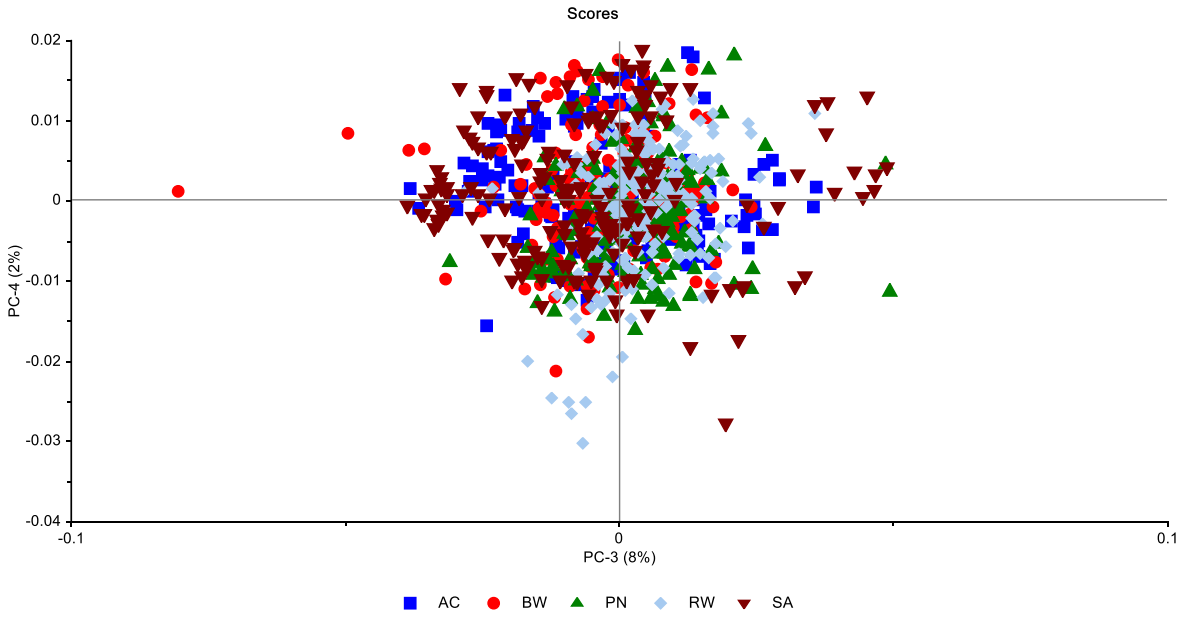

Figure 4.49 Principal component analysis score plot (PC3 (8 %) vs. PC4 (2 %)) of the river (RW) and filtered water (activated (AC), black wattle biochar (BW), pine biochar (PN) and sand (SA)).

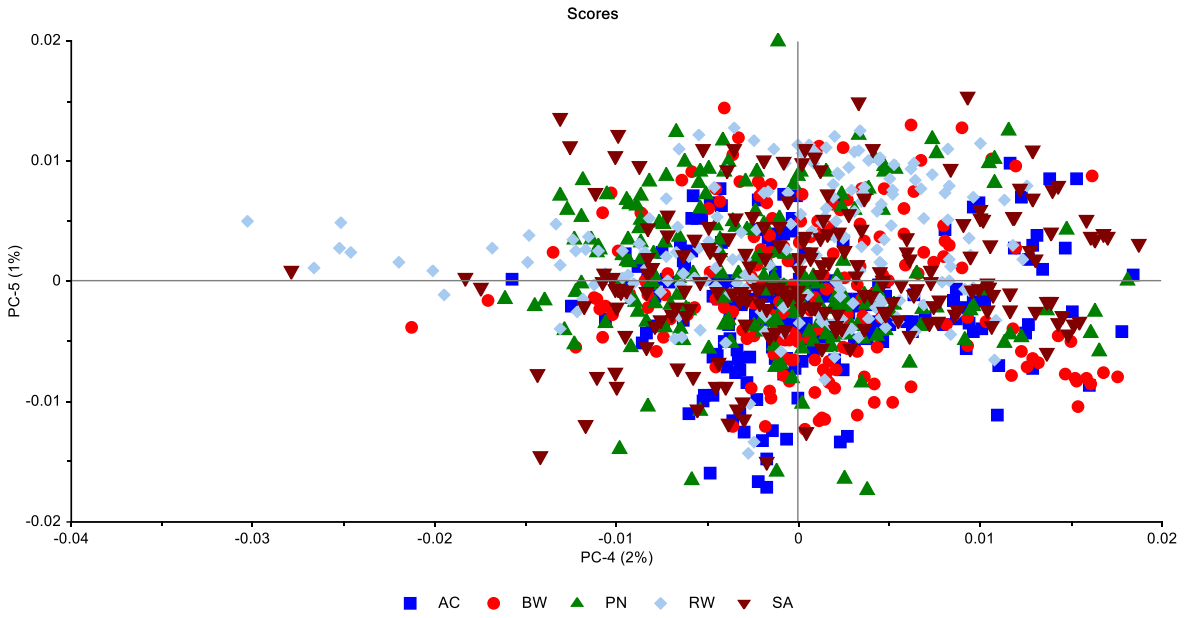

Figure 4.50 Principal component analysis score plot (PC4 (2 %) vs. PC5 (1 %)) of the river (RW) and filtered water (activated (AC), black wattle biochar (BW), pine biochar (PN) and sand (SA)).

The loadings plot of PC 1 (**Figure 4.51**) exhibited a positive band at 1380 nm and a negative band at 1520 nm. The water band at 1380 nm represents OH stretching (Xantheas, 1995). The band, 1380 nm, is associated with the OH stretching in the first overtone of water, while 1520 nm is linked with the bending vibration of the first overtone of water (Siesler, 2006; Xantheas, 1995). The loadings plot of PC 2 (**Figure 4.52**) exhibited four positive bands at 1365 nm, 1383 nm, 1412 nm and 1474 nm. The water band 1365 nm has been assigned to stretching of the hydrogen bond in water molecules (Xantheas, 1995). The water band at 1383 nm represents OH stretching (Xantheas, 1995). The 1412 nm water band is linked to water molecules with little or no intermolecular hydrogen bonds (S_0) (Segtnan et al., 2001). The 1474 nm wavelength band has been assigned to water molecules with three hydrogen bonds (S_3) (Franks, 1973; Siesler, 2006).

The loadings plot of PC 3 (**Figure 4.53**) exhibited three positive bands, 1336 nm, 1362 nm, 1476 nm, and one negative band at 1412 nm. The band at 1336 nm is linked to the asymmetric stretching of the water molecule (Siesler, 2006). The water band 1362 nm has been assigned to stretching of the hydrogen bond in water molecules (Xantheas, 1995). The 1412 nm water band is linked to water molecules with little or no intermolecular hydrogen bonds(S_0) (Segtnan et al., 2001). The 1476 nm wavelength band has been assigned to water molecules with three hydrogen bonds (S_3) (Franks, 1973; Siesler, 2006).

The PC 4 (**Figure 4.54**) loadings plot exhibited four bands, two positive and two negatives, at 1340 nm, 1374 nm, 1443 nm and 1475 nm. The band at 1340 nm is linked to the asymmetric stretching of the water molecule (Siesler, 2006). The band at 1374 nm is linked to the symmetrical and asymmetrical stretching of water molecules (Siesler, 2006) The 1443 nm and 1475 nm bands has been assigned to water molecules with one (S_1) and three (S_3) hydrogen bonds (Cattaneo *et al.*, 2009; Franks, 1973; Siesler, 2006).

Six bands were identified in the PC 5 loadings plot (**Figure 4.55**), 1343 nm,1364 nm,1383 nm, 1423 nm,1478 nm,1516 nm. The band at 1343 nm is linked to the asymmetric stretching of the water molecule (Siesler, 2006). The water band 1364 nm has been assigned to stretching of the hydrogen bond in water molecules (Xantheas, 1995). The water band at 1383 nm represents OH stretching (Xantheas, 1995). The band, 1423 nm, is referred to as the water hydration band (Tsenkova, 2009; Williams, 2009). The 1478 nm wavelength band has been assigned to water molecules with three hydrogen bonds (S_3) (Franks, 1973; Siesler, 2006). The 1516 nm water band has been assigned to bending vibration mode of the first overtone of water (Siesler, 2006).

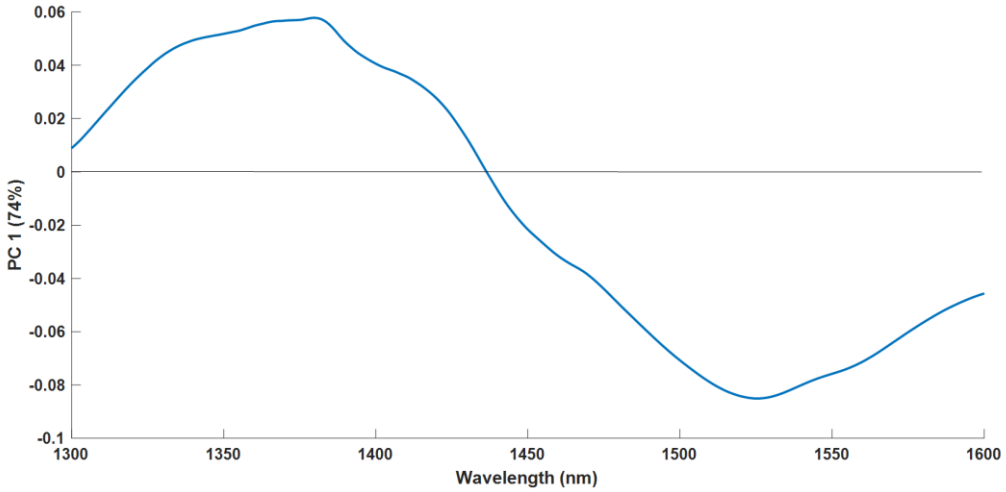

Figure 4.51 PCA loadings line plot for PC 1 (74%).

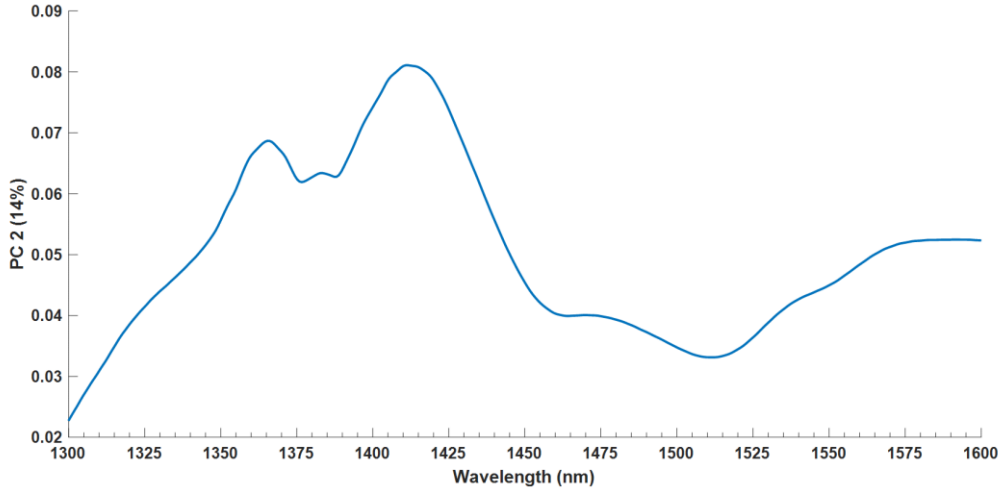

Figure 4.52 PCA loadings line plot for PC 2 (14%).

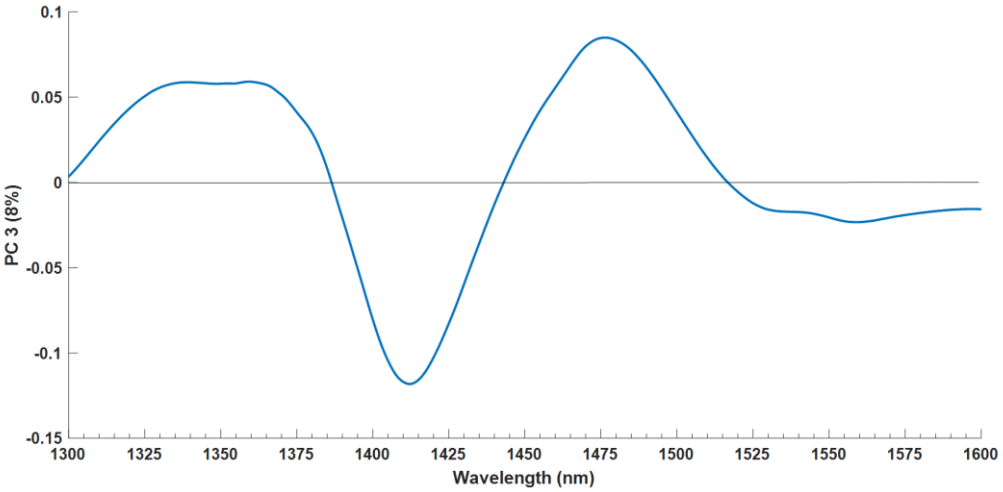

Figure 4.53 PCA loadings line plot for PC 3 (8%).

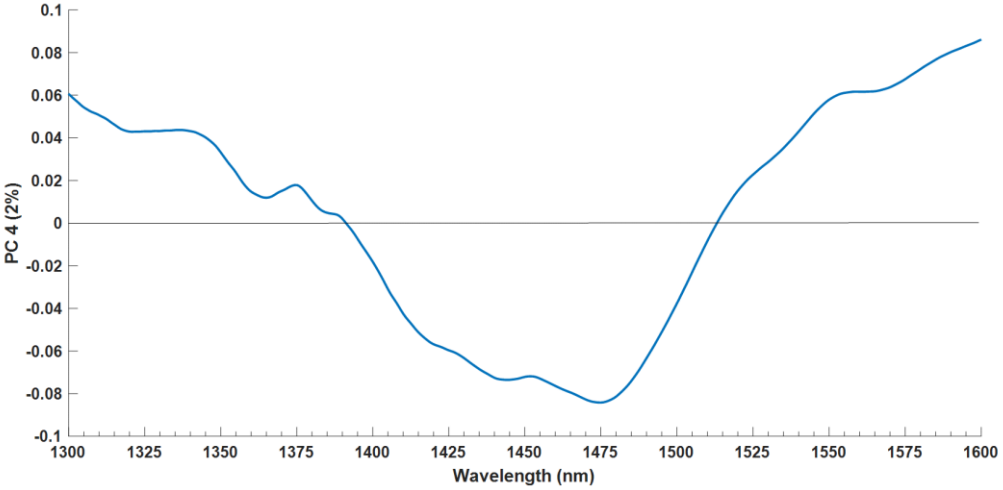

Figure 4.54 PCA loadings line plot for PC 4 (2%).

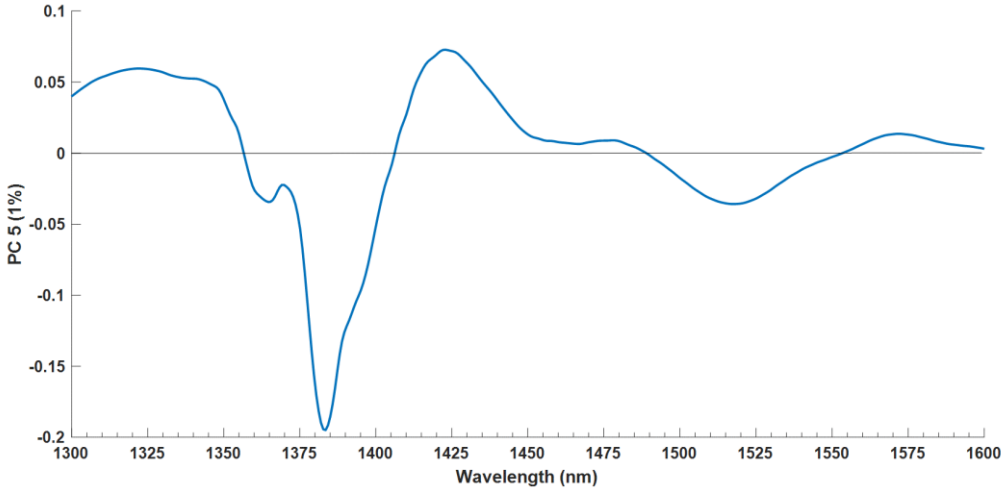

Figure 4.55 PCA loadings line plot for PC 5 (1%).

4.4.1.2.2 Partial least squares regression

Separate PLS regression models were built using smoothed (Savitzky-Golay filter using 2nd order polynomial and 21 points) and SNV transformed spectra in the range of 1300 – 1600 nm for the river (RW) and filtered (AC, BW, PN and SA) water. The relationship between the spectral data and TDS content samples were investigated with the PLS regression models. The accuracy of the models was evaluated with the use of their respective R^2 and RMSECV values (**Table 4.7**).

Table 4.7 PLS statistics of the models

Sample	R^2	RMSECV
AC	0.75	44.06
BW	0.72	91.35
PN	0.57	48.09
RW	0.60	56.04
SA	0.83	94.59

The regression vectors for the river and filtered water is shown in **Figures 4.56 – 4.60**. The important water absorbance bands exhibited in the PLS regression vectors of the river and filtered water correspond with those identified in sections 4.4.1.1 and 4.4.1.2.1. The water absorbance bands exhibited in the regression vectors (**Figures 4.56 – 4.60**) with their respective band assignment is shown in **Table 4.8**.

Table 4.8 WAMACS of interest and their assignments for the regression vectors of AC, BW, PN, RW and SA

Regression vector	Wavelength (nm)	Assignment	Reference
AC	1382	OH stretch	(Xantheas, 1995)
	1415	S_1	(Cattaneo *et al.*, 2009; Siesler, 2006)
	1484	S_4	(Tsenkova, 2009)
BW	1343	V_3	(Siesler, 2006)
	1404	S_0	(Segtnan et al., 2001)
	1428	Water hydration	(Tsenkova, 2009; Williams, 2009)
	1440	S_1	(Cattaneo *et al.*, 2009; Siesler, 2006)
	1485	S_4	(Tsenkova, 2009)
PN	1336	V_3	(Siesler, 2006)
	1365	OH Stretch	(Xantheas, 1995)
	1384	OH Stretch	(Xantheas, 1995)
	1408	S_0	(Segtnan et al., 2001)
	1458	S_2	(Franks, 1973)
RW	1366	OH Stretch	(Xantheas, 1995)
	1383	OH Stretch	(Xantheas, 1995)
SA	1340	V_3	(Siesler, 2006)
	1366	OH Stretch	(Xantheas, 1995)
	1382	OH Stretch	(Xantheas, 1995)
	1413	S_0	(Segtnan et al., 2001)
	1462	S_2	(Franks, 1973)

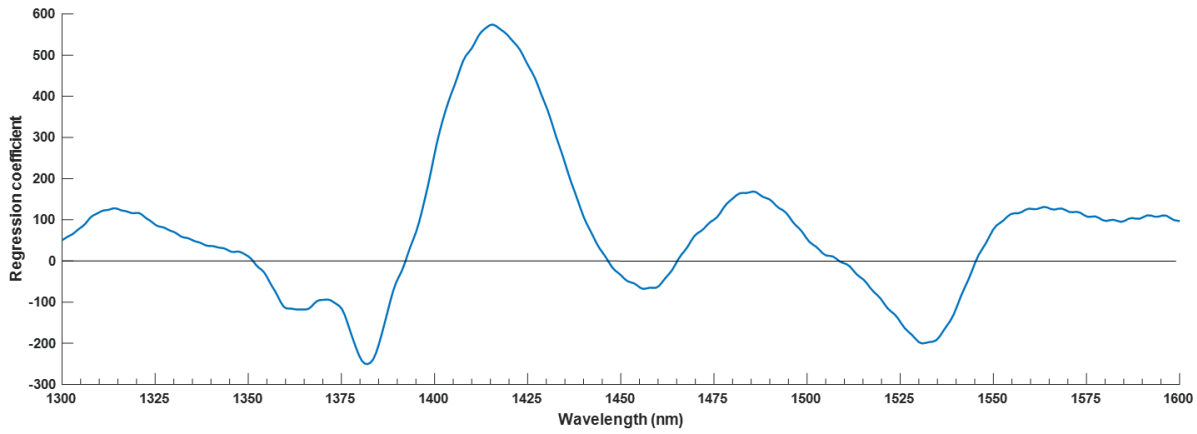

Figure 4.56 PLS regression vector of the filtered river water using activated charcoal (AC) in the 1300 – 1600 nm wavelength range.

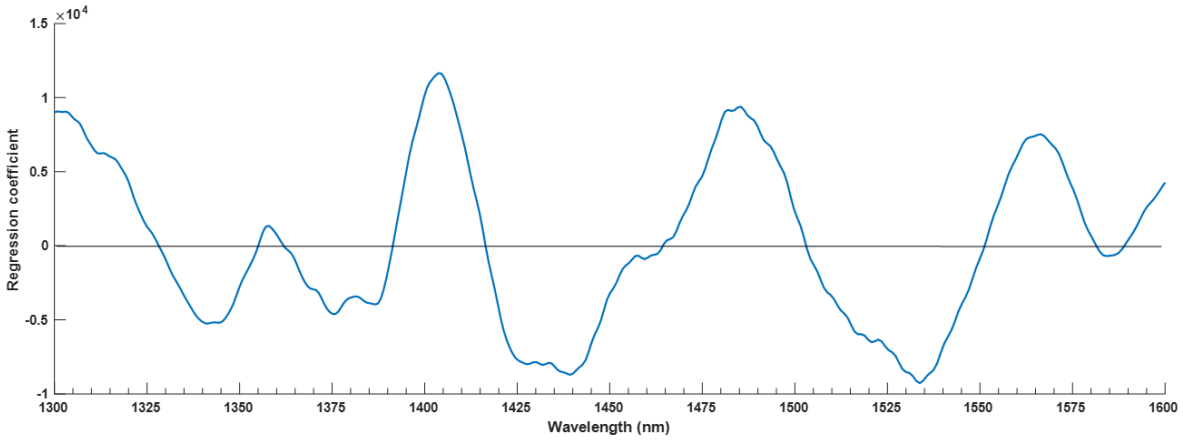

Figure 4.57 PLS regression vector of the filtered river water using black wattle (BW) biochar in the 1300 – 1600 nm wavelength range.

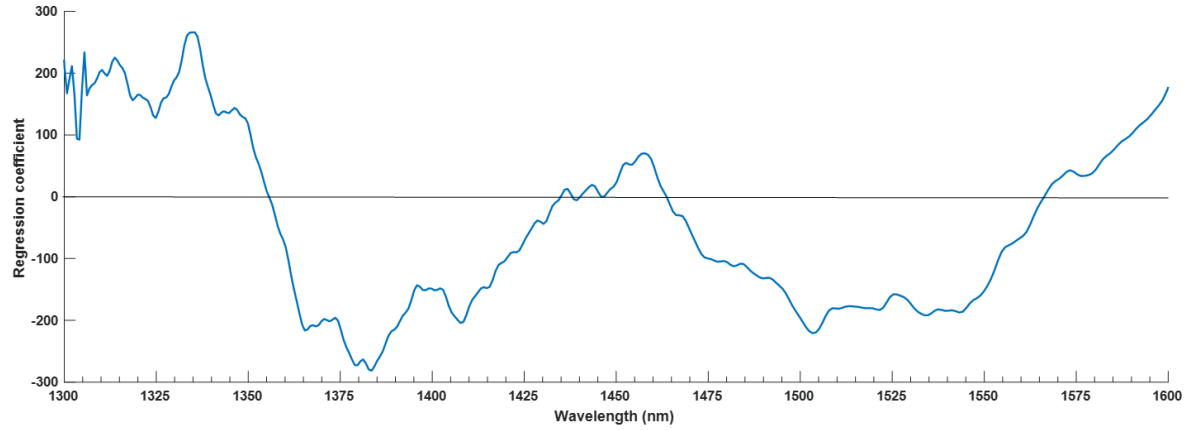

Figure 4.58 PLS regression vector of the filtered river water using pine (PN) biochar in the 1300 – 1600 nm wavelength range.

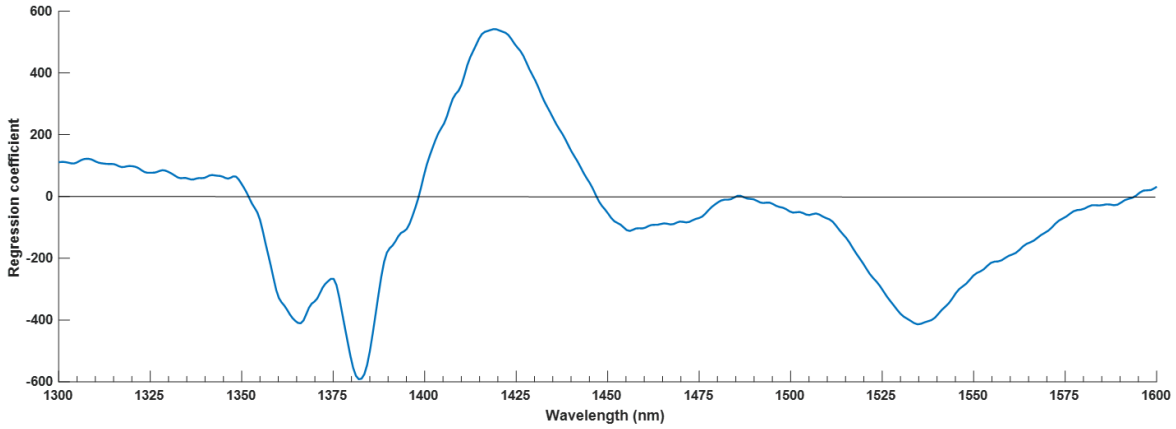

Figure 4.59 PLS regression vector of unfiltered river water (RW) in the 1300 – 1600 nm wavelength range.

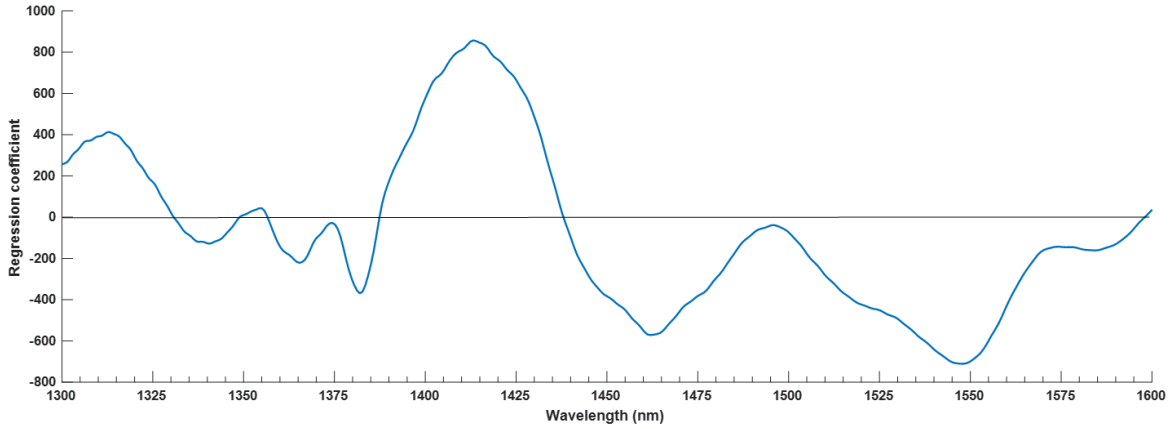

Figure 4.60 PLS regression vector of the filtered river water using sand (SA) in the 1300 – 1600 nm wavelength range.

4.4.1.3 Aquagram

The difference between the river and filtered water was visualised with aquagrams (**Figure 4.61**). The aquagrams were constructed with the water absorbance bands identified in sections 4.4.1.1 and 4.4.1.2.

The spectral pattern of AC and SA is characterised with high absorbance values at 1458 nm, 1462 nm, 1472 and 1485 nm. These wavelengths have been assigned to water molecules with one, two, three and four hydrogen bonds, respectively. This indicates that the molecules have strong hydrogen bonds. The RW spectral pattern indicates that it contains weakly bound molecules. This is due to high absorbance values in the 1348 nm – 1426 nm range and has been linked to free water and water solvation shells. The decrease in absorbance of weakly bonded hydrogen molecules in the spectral pattern of RW to that of AC and SA indicate that some solutes were removed during the filtration process.

The spectral patterns of BW and PN indicate low absorbance values in the 1366 nm – 1426 nm range, which is linked to free and weakly bound water molecules. Compared to the spectral pattern of RW, BW and PN contain less free water molecules than RW. This shows that during the filtration process a rearrangement of water molecular species occurs.

The results reported are similar to the results reported by Muncan *et al.* (2020). The authors monitored the changes in the water spectral patterns due to different filtration steps.

This indicates that with aquaphoptomics it is possible to monitor changes in the water spectral patterns due to filtration steps as a perturbation.

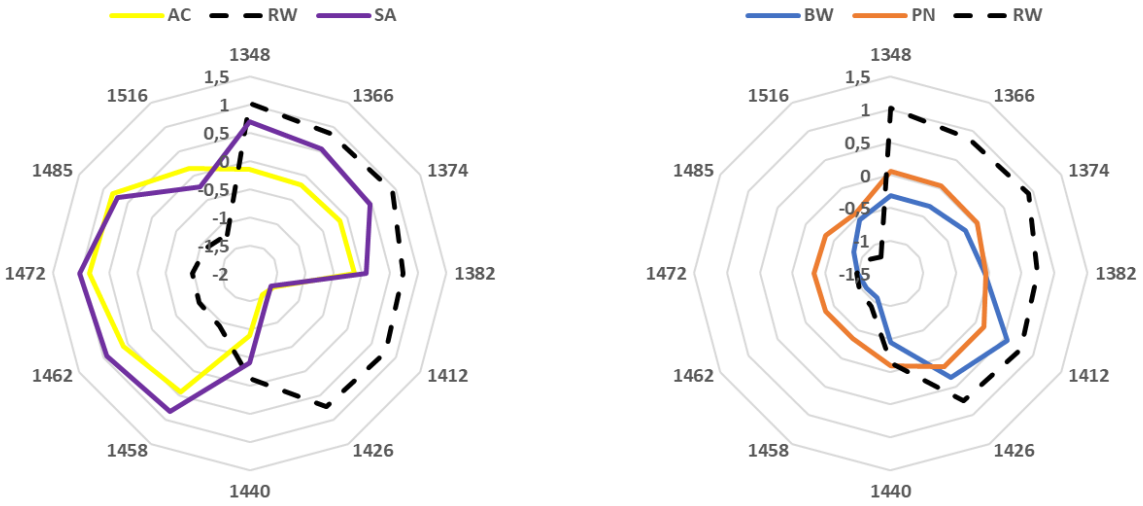

Figure 4.61 Aquagram visualising the spectral patterns of AC (yellow), BW (blue), PN (orange), RW (black) and SA (purple).

4.4.2 Changes in river water over time

4.4.2.1 Spectral analysis

The difference spectra of the river water collected for ten days was calculated by subtracting the average spectra of all the river samples from the average spectra of each day. The resulting difference spectra plot is shown in **Figure 4.62**. The difference spectra exhibited water absorbance bands at 1382 nm, 1416 nm and 1454 nm. The 1382 nm band is linked to the O – H stretch and water solvation shells (Xantheas, 1995). Water molecules with two hydrogen bonds has been linked to the 1416 nm band (Franks, 1973). The bending and asymmetric stretch of the first overtone of water occur at 1454 nm (Siesler, 2006).

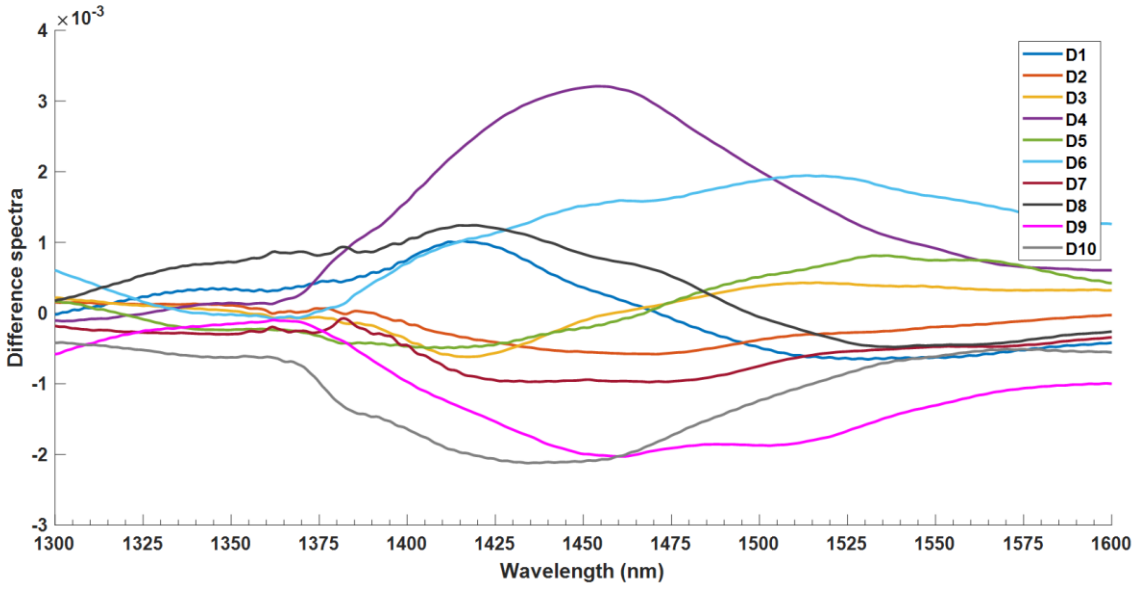

Figure 4.62 Difference spectra (Savitzky-Golay smoothing with 2nd order polynomial and 21 points) of the river water collected for the 10 days.

The 2nd derivative (**Figure 4.63**) of the first overtone indicated four water absorbance bands at 1348 nm, 1412 nm, and 1462 nm. The band at 1348 nm is linked to the asymmetric stretching of the water molecule (Siesler, 2006). Water molecules with different numbers of hydrogen bonds have been assigned to the absorbance bands of 1412 nm, 1434 nm and 1462 nm. The region of 1412 nm is associated with water molecules with little or no intermolecular hydrogen bonds (S_0) (Segtnan et al., 2001). The water absorbance band, 1462 nm, has been assigned to water molecules with two hydrogen bonds (S_2).

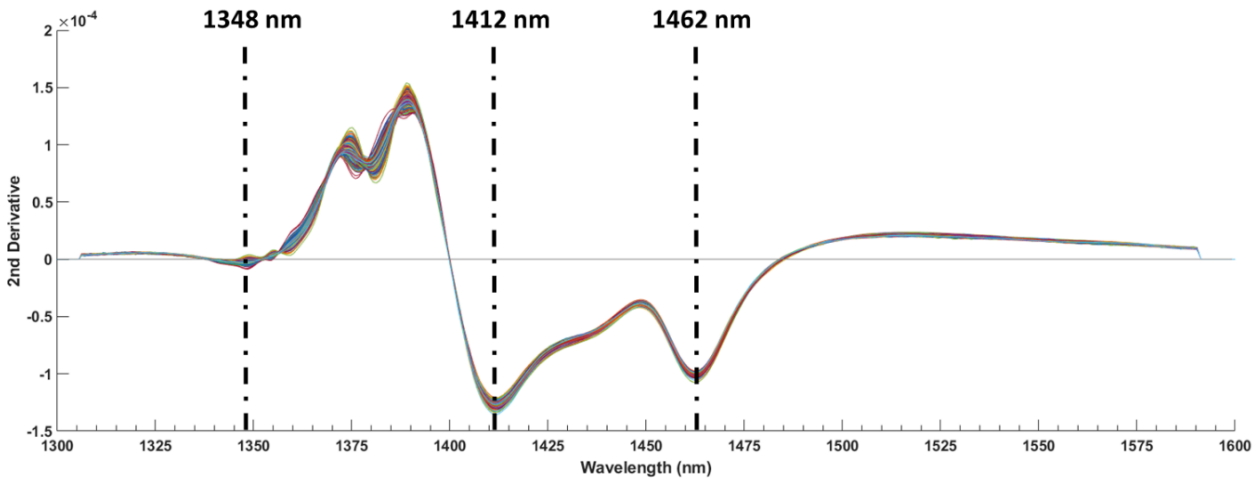

Figure 4.63 Second derivative (calculated with a Savitzky-Golay filter using 2nd order polynomial and 21 points) spectra in the first overtone of water (1300 – 1600 nm) of the river water samples.

4.4.2.2 Multivariate data analysis

4.4.2.1 Principal Component analysis

Principal component analysis was performed to explore the data by examining the changes in the river water over a period of ten days. The resulting score and loading plots are shown in **Figure 4.64 -4.73**.

Separation was observed between the different days in the score plots (**Figure 4.64 – 4.68**). PC 1 accounted for 52% of the variance in the data, whereas PC2, PC3, PC4 and PC5 accounted for 26%, 10%, 6% and 3%, respectively.

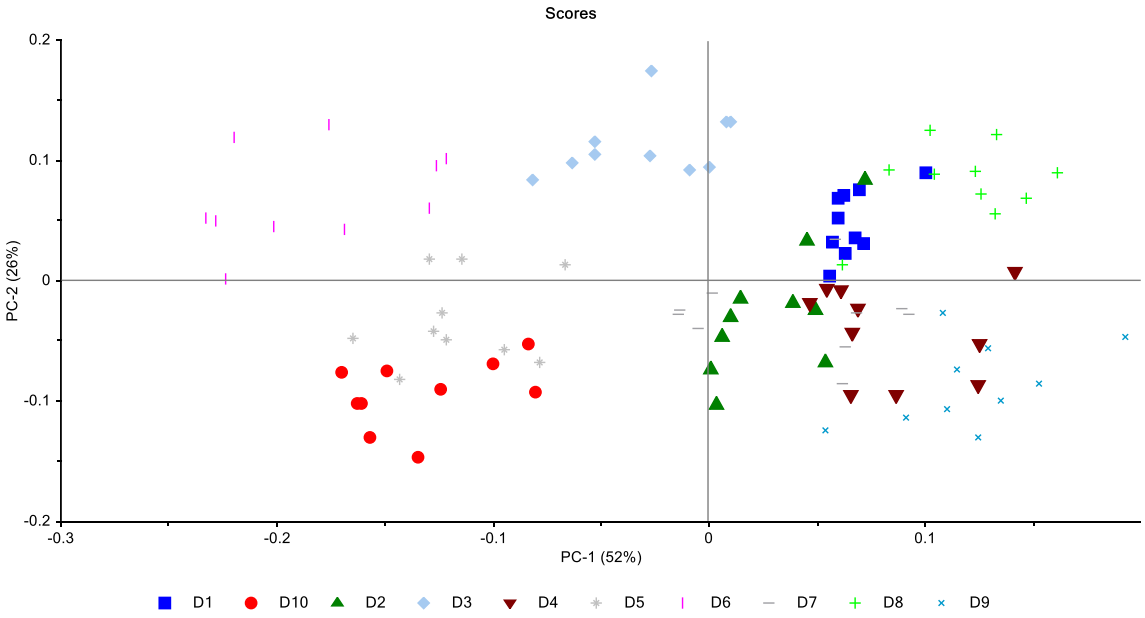

Figure 4.64 Principal component analysis score plot (PC1 (52 %) vs. PC2 (26 %)) of river water collected over a period of ten days.

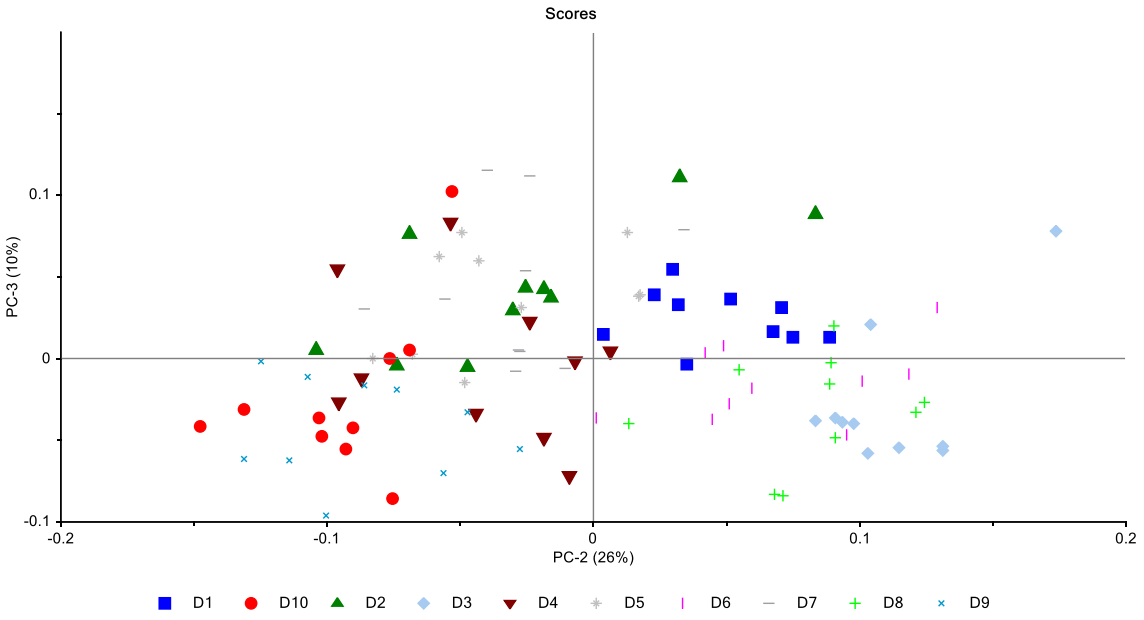

Figure 4.65 Principal component analysis score plot (PC2 (26 %) vs. PC3 (10 %)) of river water collected over a period of ten days.

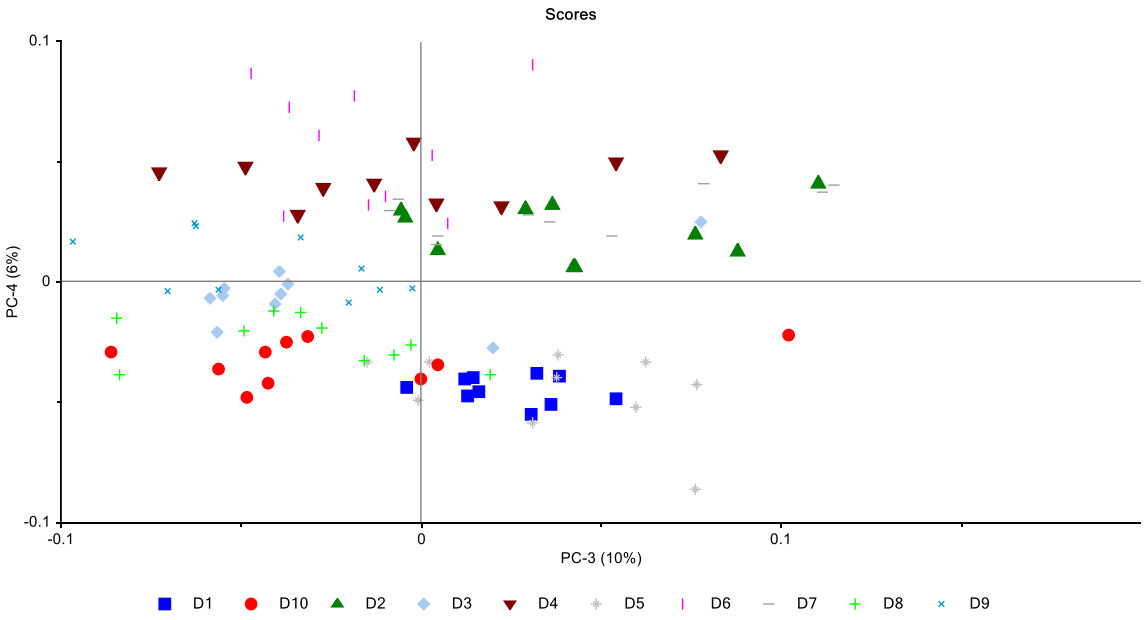

Figure 4.66 Principal component analysis score plot (PC3 (10 %) vs. PC4 (6 %)) of river water collected over a period of ten days.

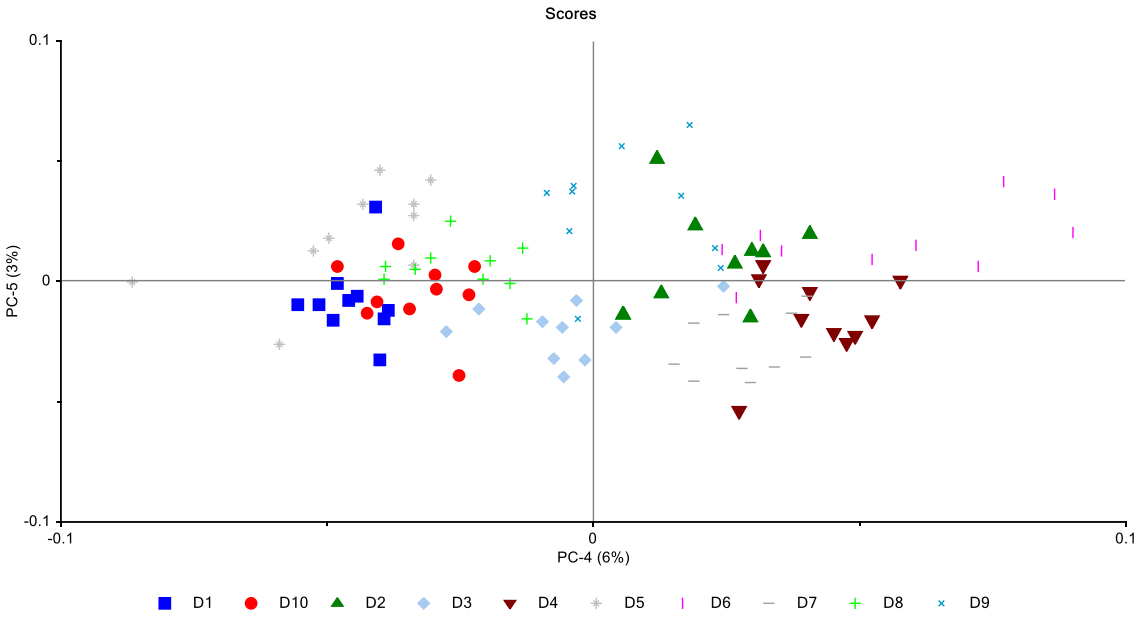

Figure 4.67 Principal component analysis score plot (PC4 (6 %) vs. PC5 (3 %)) of river water collected over a period of ten days.

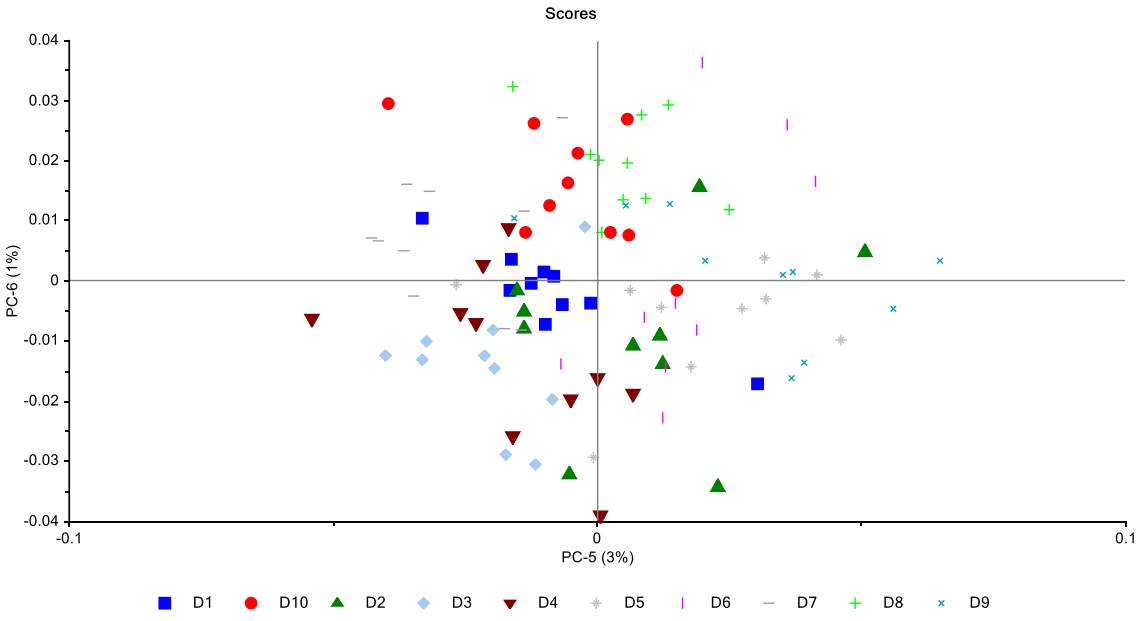

Figure 4.68 Principal component analysis score plot (PC5 (3 %) vs. PC6 (1 %)) of river water collected over a period of ten days.

The loadings plot of PC 1 (**Figure 4.69**) exhibited two water absorbance bands at 1375 nm and 1518 nm. The symmetric stretching and bending vibrational modes (v_1 + v_2) of the first overtone of water have been linked to the 1375 nm band (Siesler, 2006). The

1518 nm wavelength has been assigned to the bending vibrational mode (v_2) of the first overtone of water (Siesler, 2006).

Four absorbance bands, 1365 nm, 1423 nm, 1437 nm and 1484 nm were identified in the loadings plot of PC 2 (**Figure 4.70**). The water band 1365 nm has been assigned to stretching of the hydrogen bond in a water molecule (Xantheas, 1995). The 1423 nm band is linked to water hydration (Tsenkova, 2009; Williams, 2009). Water molecules with one (S_1) and four (S_4) hydrogen bonds have been assigned to the 1437 nm and 1484 nm bands, respectively (Cattaneo *et al.*, 2009; Franks, 1973; Siesler, 2006).

The loadings plot of PC 3 (**Figure 4.71**) and PC 4 (**Figure 4.72**) exhibited bands at 1362 nm, 1366 nm, 1384 nm, 1385 nm, 1454 nm and 1518 nm. The water bands in the region of 1360 - 1366 nm and 1380 - 1384 nm has been assigned to stretching of the hydrogen bond in water molecules (Xantheas, 1995). The band 1454 nm is linked to the bending and asymmetric stretching vibrational modes (v_2+v_3) of the first overtone of water (Cattaneo *et al.*, 2009; Siesler, 2006). The bending vibrational mode (v_2) of the first overtone of water has been assigned to the 1518 nm band (Siesler, 2006).

Four bands, 1338 nm, 1380 nm, 1415 nm and1472 nm, were identified in the loadings plot of PC 5 (**Figure 4.73**). The 1338 nm band is associated with the bending vibrational mode of the first overtone of water (Siesler, 2006). Water molecules with no (S_0) and three (S_3) hydrogen bonds correspond to 1415 nm and 1472 nm bands, respectively (Franks, 1973; Segtnan et al., 2001; Siesler, 2006).

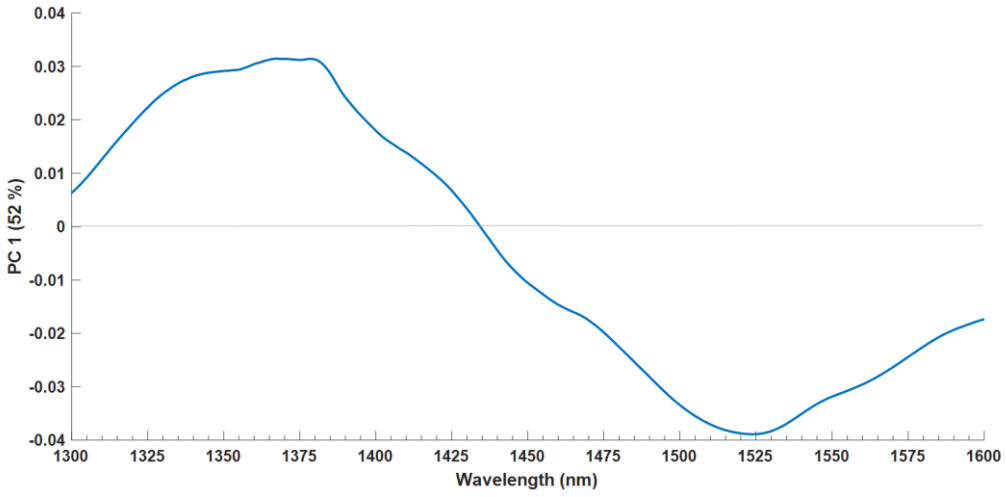

Figure 4.69 PCA loadings line plot for PC 1 (52%).

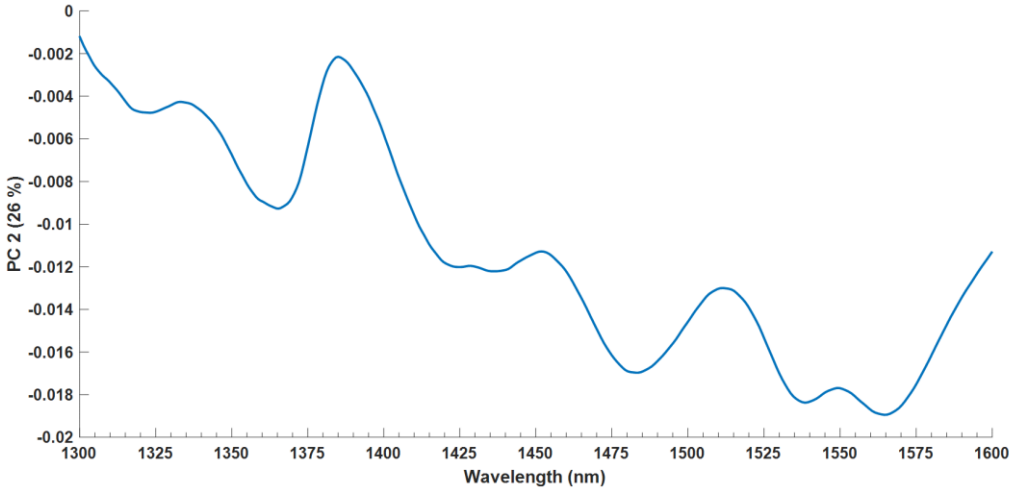

Figure 4.70 PCA loadings line plot for PC 2 (26%).

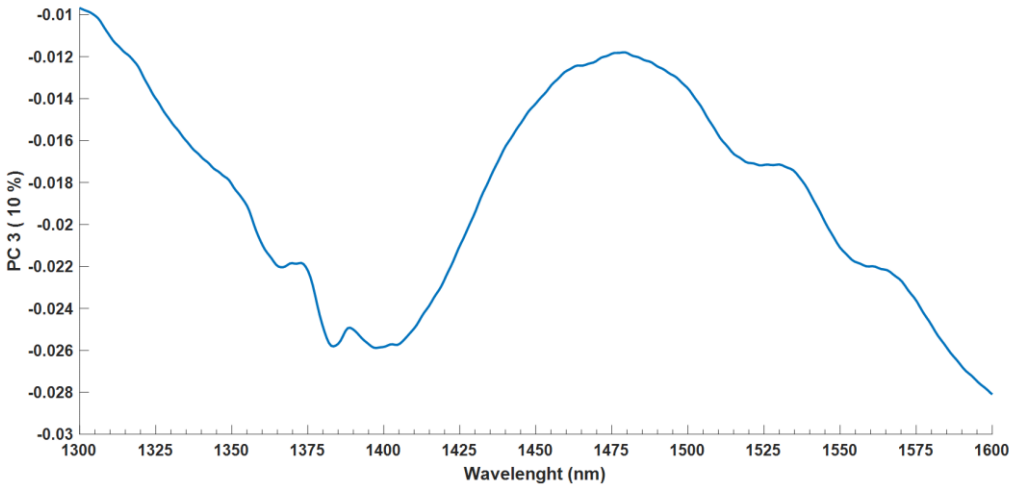

Figure 4.71 PCA loadings line plot for PC 3 (10%).

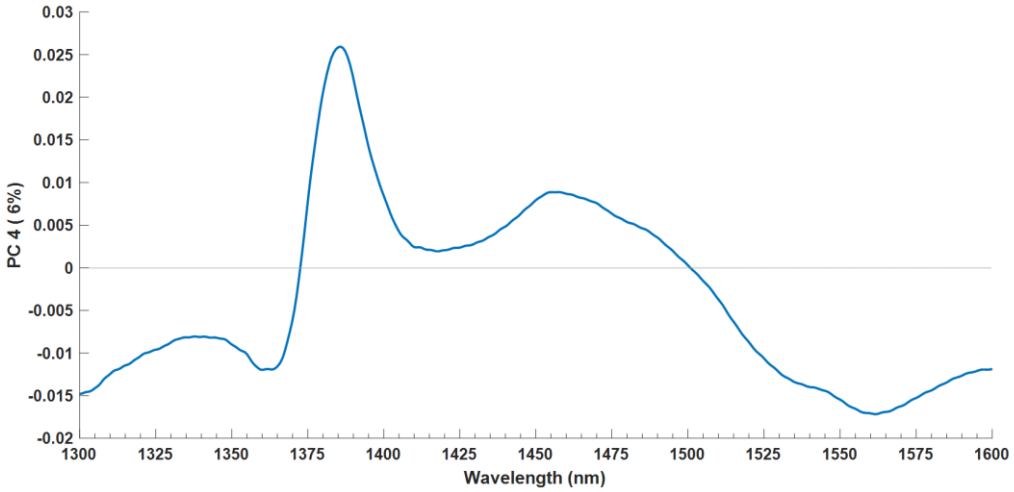

Figure 4.72 PCA loadings line plot for PC 4 (6%).

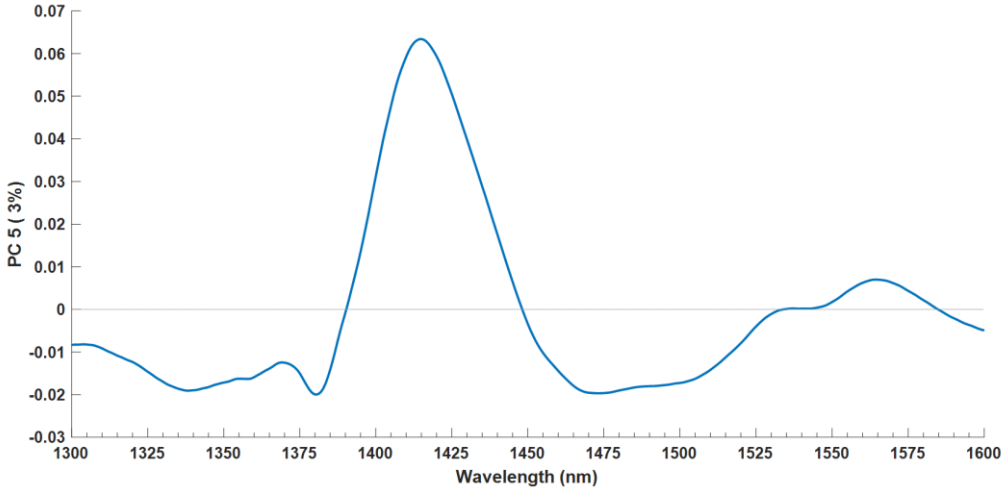

Figure 4.73 PCA loadings line plot for PC 5 (3%).

4.4.2.3 Aquagram

The spectral patterns of the river water samples are shown in **Figure 4.74**. The water samples showed a wide range of variability in the physico-chemical characteristics which are due to the variations in climatic conditions on the different days of sample collection. From the aquagram (**Figure 4.74**), it can be seen that water collected from the same river on different days have different water spectral patterns.

The water spectral patterns of day 1 – 4 is characterised by the presence of free water molecules that can interact with solutes present in the water. This is due to high absorbance values observed from 1348 nm – 1424 nm, indicating that the water molecules are loosely bound. On day 5 and 6, high absorbance values are observed in the 1454 nm – 1518 nm range. This wavelength range is associated with different water molecular species sizes. This indicates that the water is strongly bound to other molecules. A shift back towards loosely bound water molecules is observed from day 7 – 9, with high absorbance values in the 1348 nm – 1424 nm range. An increase in different water clusters is seen on day 10, with high absorbance values in the 1454 nm – 1518 nm range. The differences in the water spectral patterns are likely due to the variation in the physico-chemical characteristics of the ten days.

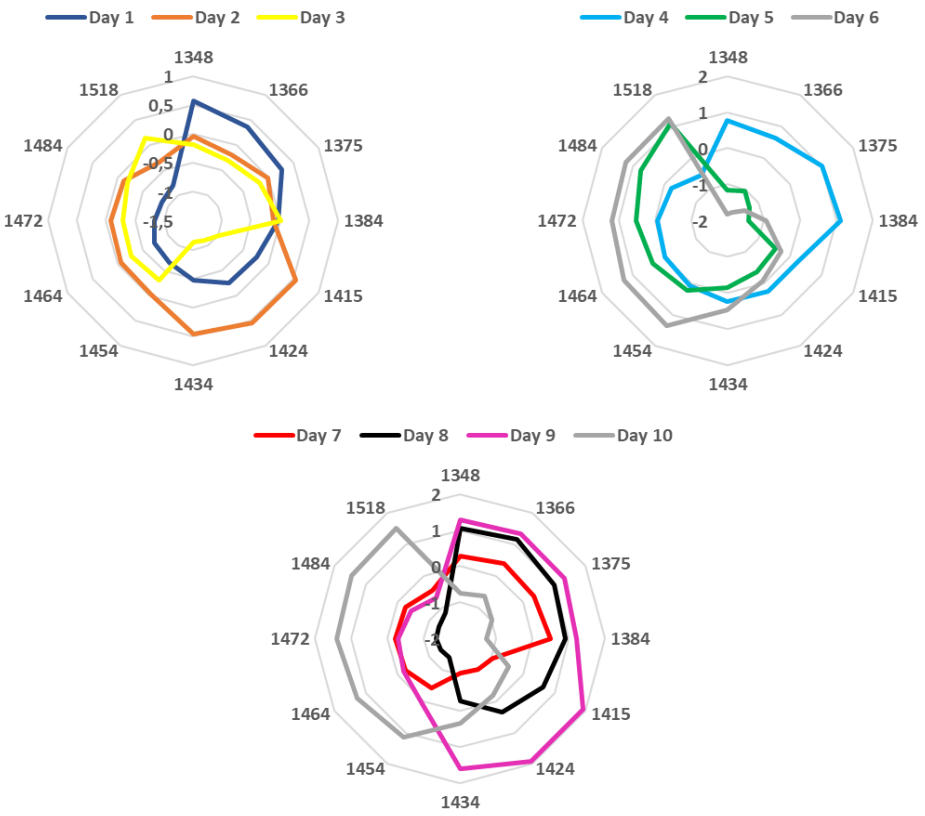

Figure 4.74 Aquagrams visualisng the spectral patterns of river water for the ten days.

The physico-chemical data of the river water for the ten days are shown in **Table 4.9**. River water is complex system, as the physico-chemical characteristics of the water does not remain constant but change continuously due to environmental and human effects. This makes it difficult to determine which parameter influenced the water matrix coordinates.

Table 4.9 Physico-chemical data of the river water for the ten days

	EC (mS/cm)	pH	ALK (mg/l)	TURB (NTU)	TDS (ppm)	TSS (mg)	VSS (mg)	UVT%
D1	0.68	7.14	271.50	16.01	369.50	18.20	11.60	33.85
D2	0.76	5.92	170.00	21.65	398.50	13.00	10.80	21.10
D3	0.52	5.72	122.25	28.85	276.00	32.60	21.40	24.75
D4	0.6	6.19	139.00	16.78	266.50	10.00	10.00	40.70
D5	0.62	7.03	212.75	18.91	278.50	12.20	10.20	37.90
D6	0.73	6.75	250.00	45.50	313.50	21.60	13.40	26.50
D7	0.43	7.04	139.00	8.70	185.00	6.00	4.00	54.90
D8	0.76	6.38	185.75	43.25	453.50	15.20	11.40	31.75
D9	0.72	6.00	125.00	47.00	436.00	9.20	7.00	17.80
D10	0.73	7.13	191.25	19.14	458.50	14.80	9.60	30.10

TDS – Total dissolved solids, EC – Electrical conductivity, ALK – Alkalinity, TURB – Turbidity, TSS – Total suspended solids, VSS – Volatile suspended solids, UVT % – Ultraviolet transmission percentage

4.4.3 Conclusion

With the use of the aquaphotomics approach, it is easy to track changes in water with respect to changes in the water clustering. The effect of different filtration material can be monitored with aquaphotomics since the filtration process alters the hydrogen bonding and water clusters within water. However, further research is required to determine how the different physico-chemical parameters influence the water spectral patterns of the samples.

NIR spectroscopy combined with the aquaphotomics approach can monitor changes in the water spectral patterns of water due to different filtration media. It can also detect changes in the water spectral pattern due to quality changes. However, it is not completely understood how the different physico-chemical parameter influences the spectra of water.

4.5 References

Cattaneo, T. M., Cabassi, G., Profaizer, M. & Giangiacomo, R. (2009). Contribution of light scattering to near infrared absorption in milk. *Journal of Near Infrared Spectroscopy,* **17,** 337-343.

Chalmers, J. M. & Griffiths, P. R. (2002). Handbook of vibrational spectroscopy. New York, New York : J. Wiley.

Cui, X., Cai, W. & Shao, X. (2016). Glucose induced variation of water structure from temperature dependent near infrared spectra. *RSC Advances,* **6,** 105729-105736.

Franks, F. (1973). Water: a comprehensive treatise. New York, Plenum.

Gowen, A. A., Amigo, J. M. & Tsenkova, R. (2013). Characterisation of hydrogen bond perturbations in aqueous systems using aquaphotomics and multivariate curve resolution-alternating least squares. *Anal Chim Acta,* **759,** 8-20.

Gowen, A. A., Marini, F., Tsuchisaka, Y., De Luca, S., Bevilacqua, M., O'Donnell, C., Downey, G. & Tsenkova, R. (2015). On the feasibility of near infrared spectroscopy to detect contaminants in water using single salt solutions as model systems. *Talanta,* **131,** 609-618.

Kojić, D., Tsenkova, R., Tomobe, K., Yasuoka, K. & Yasui, M. (2014). Water confined in the local field of ions. *ChemPhysChem,* **15,** 4077-4086.

Kovacs, Z., Bazar, G., Oshima, M., Shigeoka, S., Tanaka, M., Furukawa, A., Nagai, A., Osawa, M., Itakura, Y. & Tsenkova, R. (2016). Water spectral pattern as holistic marker for water quality monitoring. *Talanta,* **147,** 598-608.

Luck, W. A. (1974). Structure of water and aqueous solutions. Verlag Chemie.

Maeda, H., Ozaki, Y., Tanaka, M., Hayashi, N. & Kojima, T. (1997). Near infrared spectroscopy and chemometrics studies of temperature-dependent spectral variations of water: relationship between spectral changes and hydrogen bonds. *Journal of Near Infrared Spectroscopy,* **3,** 191-201.

Muncan, J., Matovic, V., Nikolic, S., Askovic, J. & Tsenkova, R. (2020). Aquaphotomics approach for monitoring different steps of purification process in water treatment systems. *Talanta,* **206,** 120253.

Munćan, J. S., Matija, L., Simić-Krstić, J. B., Nijemčević, S. S. & Koruga, Đ. L. (2014). Discrimination of mineral waters using near infrared spectroscopy and aquaphotomics. *Hemijska industrija,* **68,** 257-264.

Segtnan, V. H., Šašić, Š., Isaksson, T. & Ozaki, Y. (2001). Studies on the structure of water using two-dimensional near-infrared correlation spectroscopy and principal component analysis. *Analytical Chemistry,* **73,** 3153-3161.

Siesler, H. W. (2006). Near-infrared spectroscopy : principles, instruments, applications. Weinheim, Weinheim : Wiley-VCH.

Tsenkova, R. (2009). Introduction: Aquaphotomics: dynamic spectroscopy of aqueous and biological systems describes peculiarities of water. *Journal of Near Infrared Spectroscopy,* **17,** 303.

Williams, P. (2009). Influence of water on prediction of composition and quality factors: the Aquaphotomics of low moisture agricultural materials. *Journal of Near Infrared Spectroscopy,* **17,** 315.

Xantheas, S. S. (1995). Ab initio studies of cyclic water clusters (H2O) n, n= 1–6. III. Comparison of density functional with MP2 results. *The Journal of Chemical Physics,* **102,** 4505-4517.

Chapter 5

General discussion and conclusion

Water scarcity and quality degradation present a major challenge to both developed and developing countries, as it poses a risk to both the environment and human health (Kirby *et al.*, 2003). With the increasing global population, water resources have been exposed to a variety of pollutants due to anthropogenic activities (Kovacs *et al.*, 2016). Monitoring water quality is a complex problem, as numerous tests are required to provide an overall indication of the quality (Gowen *et al.*, 2015). With the increasing scarcity of water and increasing demand for water, alternative methods for intervention, monitoring and sanitation are required. NIR spectroscopy combined with the aquaphotomics approach has shown that it has the potential to be used as a rapid screening method for water quality (Gowen *et al.*, 2012; Kovacs *et al.*, 2016).

This study aimed at investigating the use of NIR spectroscopy combined with the aquaphotomics approach as a potential screening method for water quality. This was achieved by investigating the effect temperature has on the water spectrum. In addition, bottled water was examined to determine if we could differentiate between various geographical sources and types of water (mineral or spring water). Finally, the effect of a filtration step and day of sample collection was investigated in the water spectral pattern of river water.

The spectral analysis of water in the 1300 – 1600 nm range at the three (20˚C, 25˚C and 30˚C) temperatures indicated the 1412 nm and 1492 nm bands displayed the most variation. This was due to the change in the weakly and strongly hydrogen-bonded water molecules with an increase in temperature. Variations in temperature cause changes in the concentration of free and bonded OH groups (Chalmers & Griffiths, 2002). An increase in temperature leads to a reduction in water clusters, due to the weakening of hydrogen-bonding (Cui *et al.*, 2016).

The PLS-DA model of the spring water could successfully differentiate between the three sources with a classification accuracy of 89.68%. The classification of the three sources was attributed to the variation in their mineral content. The aquagram of the spring water indicated that each source had a unique water spectral pattern. Due to water being a strong absorber in the NIR range, conventional data analysis can easily overlook any subtle changes in the water spectrum. Therefore, aquaphotomics is a good approach as it

highlights the specific wavelengths where changes occurred as a result of a perturbation. Source A was characterised by the large concentration of water clusters, resulting in strongly bonded water. While source C contains a higher concentration of weakly bonded water molecules which are free and able to interact with solutes present in the water. However, source B contains moderate concentrations of both strongly and weakly bound water. This confirms that with use of aquaphotomics it is possible to identify water originating from different sources. The approach can be further developed and implemented as a method to verify the source of origin and prevent fraudulent products from being sold.

The result of the mineral and spring water study proved that it is possible to distinguish between the two types of water. Its PLS-DA model could classify the samples with an accuracy of 72.62 %. Both spring and mineral water is sourced from groundwater sources (Ashurst *et al.*, 2017). However, the main difference between mineral and spring water is that the mineral content of mineral water is required to remain constant (Ashurst *et al.*, 2017). Therefore, variations in the mineral content of spring water could have resulted in the misclassification of spring water as mineral water.

These results are similar to that found in literature, as previous studies indicated that it is possible to classify different water types (Munćan *et al.*, 2014; Tanaka *et al.*, 1997). The water spectrum is very sensitive, therefore, small variances between samples can be detected (Tanaka *et al.*, 1997). These authors also indicated that further research is required to understand the influence the mineral content will have in the water spectral pattern. In their ionic state, the minerals present in the water will either be a structure maker or breaker. Structure-maker ions increase the strength of the hydrogen bond network, while structure-breaker ions weaken the hydrogen bond network (Omta *et al.*, 2003).

The final objective of this study consisted of two parts. The first part investigated the effect filter media have on the water spectral pattern of river water. The water spectral patterns of the filtered and unfiltered water indicated that the different filter material altered the distribution of water clusters. This indicated that it is possible to track variations in the water spectral patterns due to changes in the water clusters in the filtered water. Muncan *et al.* (2020) indicated that aquaphotomics can monitor water as it goes through different filtration steps. The researchers concluded that aquaphotomics can detect variations in the hydrogen-bond network of the water as a result of the different filtration steps. NIR spectroscopy combined with the aquaphotomics approach can be used to monitor the effectiveness of water treatment systems.

In the second part, changes in the physico-chemical changes of river water over a ten-day period were monitored. The water spectral patterns indicated differences, however, with river water being such a complex system, it was impossible to determine how the physico-chemical parameters influenced the water spectral pattern. This was likely due to the physico-chemical parameters not remaining constant and variation observed with each day. It is recommended that the influence of the physico-chemical parameters be studied individually to determine their effect on the water spectral pattern. Currently there is no published work that focuses on the monitoring of river water quality using aquaphotomics.

These results indicate that the aquaphotomics approach has the potential to be used as a rapid screening method for water quality determination. However, this work could be furthered by investigating the effect the individual physico-chemical parameters will have on the water spectral pattern. A study done by Kovacs *et al.* (2016) has shown that the aquaphotomics approach can be used to monitor water quality changes. Aquaphotomics detects structural changes in the hydrogen bonding network of water due to a perturbation.

To conclude, NIR spectroscopy combined with the aquaphotomics approach provides a method to distinguish between water types (mineral or spring) and different sources. With further research this technique has the potential to offer the water industry a rapid method to monitor water quality.

5.1 References

Ashurst, P. R., Hargitt, R. & Palmer, F. (2017). Chapter 9 - Bottled Waters. In: Soft Drink and Fruit Juice Problems Solved (Second Edition) (edited by ASHURST, P. R., HARGITT, R. & PALMER, F.). Pp. 145-160. Woodhead Publishing.

Chalmers, J. M. & Griffiths, P. R. (2002). Handbook of vibrational spectroscopy. New York, New York : J. Wiley.

Cui, X., Cai, W. & Shao, X. (2016). Glucose induced variation of water structure from temperature dependent near infrared spectra. *RSC Advances,* **6,** 105729-105736.

Gowen, A. A., Marini, F., Tsuchisaka, Y., De Luca, S., Bevilacqua, M., O'Donnell, C., Downey, G. & Tsenkova, R. (2015). On the feasibility of near infrared spectroscopy to detect contaminants in water using single salt solutions as model systems. *Talanta,* **131,** 609-618.

Gowen, A. A., Tsenkova, R., Bruen, M. & O'donnell, C. (2012). Vibrational Spectroscopy for Analysis of Water for Human Use and in Aquatic Ecosystems. *Critical Reviews in Environmental Science and Technology,* **42,** 2546-2573.

Kirby, R. M., Bartram, J. & Carr, R. (2003). Water in food production and processing: quantity and quality concerns. *Food control,* **14,** 283-299.

Kovacs, Z., Bazar, G., Oshima, M., Shigeoka, S., Tanaka, M., Furukawa, A., Nagai, A., Osawa, M., Itakura, Y. & Tsenkova, R. (2016). Water spectral pattern as holistic marker for water quality monitoring. *Talanta,* **147,** 598-608.

Muncan, J., Matovic, V., Nikolic, S., Askovic, J. & Tsenkova, R. (2020). Aquaphotomics approach for monitoring different steps of purification process in water treatment systems. *Talanta,* **206,** 120253.

Munćan, J. S., Matija, L., Simić-Krstić, J. B., Nijemčević, S. S. & Koruga, Đ. L. (2014). Discrimination of mineral waters using near infrared spectroscopy and aquaphotomics. *Hemijska industrija,* **68,** 257-264.

Omta, A. W., Kropman, M. F., Woutersen, S. & Bakker, H. J. (2003). Negligible effect of ions on the hydrogen-bond structure in liquid water. *Science,* **301,** 347-349.

Tanaka, M., Shibata, A., Hayashi, N., Kojima, T., Maeda, H. & Ozaki, Y. (1997). Discrimination of commercial natural mineral waters using near infrared spectroscopy and principal component analysis. *Journal of Near Infrared Spectroscopy,* **3,** 203-210.