The evaluation of foodborne pathogenic bacteria using near infrared (NIR) hyperspectral imaging and multivariate image analysis

by

Terri-Lee Kammies

Thesis presented in partial fulfilment of the requirements for the degree of

Master of Science in Food Science

Department of Food Science, Faculty of Agrisciences

UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY

at

Stellenbosch University

1918 · 2018

Department of Food Science

Supervisor: Dr Paul J Williams

Co-supervisors: Prof. Marena Manley

Prof. Pieter Gouws

December 2018

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 and third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: December 2018

ABSTRACT

Near-infrared (NIR) hyperspectral imaging (HSI) and multivariate image analysis (MIA) was investigated for its potential as a rapid analytical method for the identification of foodborne pathogenic bacteria and distinguishing between the various genera and species used.

NIR hyperspectral images of *Bacillus cereus, Escherichia coli, Salmonella enteritidis, Staphylococcus aureus* and a non-pathogenic bacterium, *Staphylococcus epidermidis* were acquired using a sisuChema SWIR (short wave infrared) hyperspectral pushbroom imaging system with a spectral range of 920-2514 nm. Hyperspectral images of streaked out (on Luria Bertani agar) bacteria were acquired after 20, 40 and 60 h growth (37 °C). Principal component analysis (PCA) was applied to mean-centered data, and was used to remove background and bad pixels from images. To investigate the possibility of distinguishing between bacteria, standard normal variate (SNV) correction and the Savitzky-Golay technique (2^{nd} derivative, 3^{rd} order polynomial; 25 point smoothing) was applied to data of growth plates imaged after 20 h. PCA score plots, score images and loading line plots were then evaluated. Bacteria were divided into 3 groups which were merged into mosaics. One group contained bacteria which appeared similar in colour (white) on the growth media (*B. cereus, E. coli* and *S. enteritidis*), another contained 3 Gram positive bacteria (*B. cereus*, *S. aureus* and *S. epidermidis*) and the third contained 2 species of the same genera (*S. aureus* and *S. epidermidis*). On the cleaned images, PCA score plots illustrated distinct chemical differences between colonies which appeared similar in colour on growth media. It was possible to distinguish *B. cereus* from *E. coli* and *S. enteritidis* along PC1 (58.1 % SS) and between *E. coli* and *S. enteritidis* in the direction of PC2 (7.75 % SS). *S. epidermidis* was separated from *B. cereus* and *S. aureus* along PC1 (37.5% SS) and was attributed to variation in amino acid and carbohydrate content. Two clusters were evident in the PC1 vs. PC2 PCA score plot for the images of *S. aureus* and *S. epidermidis*, thus permitting distinction between species. Differentiation between genera, Gram positive and Gram negative bacteria and pathogenic and non-pathogenic species was possible using NIR hyperspectral imaging.

Partial least squares discriminant analysis (PLS-DA) was used as the supervised classification method to differentiate and predict the types of bacteria present. For models built and tested using the 20 h growth plates, the best predictions were made of *B. cereus* and the two *Staphylococcus* species, where results ranged from 82.0-99.96% correctly predicted pixels. However the poorest predictions were made of *E. coli* and *S. enteritidis* where results ranged from 2.34-53.9. To improve on these results, the effect of colony age has on prediction accuracies were investigated, while keeping rapidity in mind. PLS-DA models were built on standard normal variate (SNV) treated data for 20, 40

and 60 h growth plates and tested on a second set of plates. Predictions were improved – the highest overall prediction accuracies, where test plates required the least amount of growth time, was found with models built after 60 h growth and tested using 20 h growth plates. Predictions for bacteria differentiation within these models ranged from 83.1 to 98.8 % correctly predicted pixels.

UITTREKSEL

Naby-infrarooi (NIR) hiperspektrale beelding en meeveranderlike beeld ontleding (MIA) is ondersoek vir die potensiaal daarvan om te dien as a 'n vinnige analitiese metode vir die identifisering van wat voedselbederf patogene bakteriese veroorsaak. Hierdie methods kan ook gebruik word om te onderskei tussen die verskillende genera en species wat voorkom.

NIR hiperspektrale beelde van *Bacillus cereus*, *Escherichia coli*, *Salmonella enteritidis*, *Staphylococcus aureus* en 'n nie-patogeniese bakterie, *Staphylococcus epidermidis* is verkry deur gebruik te maak van 'n sisuChema SWIR (kort golf infrarooi) hiperspectrale beelding stelsel met spectrale reikwydte van 920-2498 nm. HIperspektrale beelde van uitgestreepte (op Luria Bertani agar) bakterieë is verky ná 20, 40 en 60 uur groei tyd (37 ° C). Hoof komponent analise (HKA) is toegepas op middelgesentreerde data, en is gebruik om sodoende agtergrond en dooie piksels te verwyder uit die beeld. Die standaard normale variasie (SNV) restelling en die Savitzky-Golay tegniek (2de afgeleide, 3de orde polinoom, 25 punt gladding) is toegepas op data van groeiplate wat na 20 uur afgeneem word, om die moontlikheid om tussen bakterieë te onderskei. HKA telling grafieke, telling beelde en lading lyne plotte is geëvalueer. Bakterieë is in 3 groepe verdeel wat in mosaïek saamgevoeg is. Een groep bevat bakterieë wat soortgelyk in kleur voorkom (wit) op die groeimedia (*B. cereus, E. coli* en *S. enteritidis*) 3 Gram –positiewe bakterieë (*B. cereus, S. aureus* en *S. epidermidis*), en die derde groep 2 spesies van dieselfde genera (*S. aureus* en *S. epidermidis*). Op diie gekorrigeerde beelde het HK telling grafieke met afsonderlike chemise verskille tussen kolonies geïllustreer wat op die groeimedia soordgelyk gelyk het. Deur die groepering van datapunte te bestudeer, kon daar onderskei word tussen *B. cereus* en *E. coli*, asook *B. cereus* en *S. enteritidis* en tussen *E. coli* en *S. enteritidis* in die rigting van HK 2 (7.75% SS). Om hierde observasie te maak is daar na die HK1 (58.1 % SS) as verwys op die tellings grafiek. Die aminosuur en kolidraatinhoud het did moontlik gemaak om tussen *B. cereus* en *S. aureus* op PC1 (37.5 % SS) te onderskei. Twee groeperings was duidelik in die HK1 vs HK2 HKA-telling grafiek vir die beeld van *S. aureus* en *S. epidermidis*, wat dus onderskied tussen species toelaat. Differensiasie tussen genera, Gram-positiewe en Gram-negatiewe bakterieë asook patogeniese en nie-patogeniese spesies was moontlik met behulp van NIR hiperspektrale beelding.

Parsiële kleinste kwadrate diskriminant analise (PKW-DA) is gebruik as die supervised klassifikasie method om die tipes bakterieë teenwoordig te onderskei en te voorspel. Vir modelle wat gebou en getoets is met behulp van die 20 h groeiplaatjies, is die beste voorspelling gemaak van *B. cereus* en die twee *Staphyloccocus* species,, waar die resultate gewisselhet vanaf wat aanbetref 82.0-99.6 % korrek voorspelde piksels. Die swakste voorspellings is egte gemaak van *E. coli* en *S. enteritidis*, waar

die resultate gewissel het vanaf 2.34-53.9. Om te verbeter op hierdie resultate, is die effek van kolonie ouderdom op voorspellingsakkuraatheid ondersoek, terwyl die tyd wat dit nee mom die toets aft e handel in gedagte gehou word. PKW-DA modelle is gebou op SNV behandelde data vir 20, 40 en 60 h groeiplaatjies en getoets op ‘n tweede stek plate as bevestiging. Voorspellings is op die volgende wyse verbeter: die hoogste algehele voorspellings akkuraatheid, waar toetsplate die minste hoeveelheid groeityd benodig het, is gevind met modelle gebou ná 60 uur groei en getoets met 20 h groeiplaatjies. Voorspellings vir bakteriedifferensiasie binne hierdie modelle het gewissel van 83,1 tot 98,8% korrek voorspelde pixels

ACKNOWLEDGEMENTS

I would like to express my immense gratitude to my supervisors, Dr Paul Williams, Prof Marena Manely and Prof Pieter Gouws for their continuous support, knowledge, patience and motivation throughout my research;

A special thanks to Dr Paul Williams for his persistence in helping me to complete my thesis, even though I was probably the cause of many headaches for him;

All staff and postgraduate students at the Department of Food Science, thank you for a pleasant working environment and assistance when needed. Special thanks to the head of our department, Prof. Sigge;

Prof Alvaro Viljoen, Dr Ilze Vermaak and Carmen Leonard (Tshwane University of Technology, Pretoria) for the use of their SisuChema instrument;

The National Research Fund (NRF) (unique grant nr. 90431, 2015 and Scarce Skills Bursary, 2016) are hereby acknowledged for financial support (any opinion, findings and conclusions or recommendations expressed in this material are those of the author and therefore the NRF does not accept any liability in regard thereto).

To my parents Gregory and Felicia Kammies as well as my siblings, Megan, Neil, Jaime and Stacey Kammies, thank you for your undying support, encouragement and understanding throughout these years, it will always be greatly appreciated.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANN	artificial neural networks
BP-ANN	back-propagation artificial neural network
BW-PLS	weighted partial least squares
CCD	charge coupled device
CMOS	complementary metal oxide semiconductor
Dam	damaged
EB	enrichment broth
ELISA	Enzyme-Linked Immunosorbent Assay
FT-NIR	Fourier transform near infrared
GA	genetic algorithm
HgCdTe	mercury-cadmium-telluride
HSI	hyperspectral imaging
InGaAs	Indium Gallium Arsenide
IR	infrared
kNN	k nearest neighbor
LAMP	loop-mediated isothermal amplification
LB	Luria-Bertani
LDA	linear discriminant analysis
LED	light emitting diode
LPM	lithium chloride phenylethanol-moxalactam
LV	latent variable
MCT	mercury cadmium telluride
MD	mechanically damaged
MIA	multivariate image analysis
MIR	mid-infrared
MPA	multi-purpose analyser
MSC	multiplicative scatter correction
NASBA	nucleic acid sequence-based amplification
NIR	near infrared
NW	Norris-Williams
PC	principal component
PCA	principal component analysis

PCR	polymerase chain reaction
PLS	partial least squares
PLS-DA	partial least squares discriminant analysis
PLS-R	partial least squares regression
PPC	psychotropic plate count
PT	*Pseudomonas tolaasii* infected
R	correlation coefficient
R^2	coefficient of determination
RMSE	root mean square error
ROI	region of interest
SG	Savitzky-Golay
SIMCA	soft independent modelling of class analogy
SNV	standard normal variate
SS	sum of squares
STEC	Shiga toxin-producing *Escherichia coli*
SVM	support vector machines
SWIR	short wave infrared
TH	tungsten-halogen
TVC	total viable count
U	undamaged
vis-NIR	visible-near infrared

CHAPTER 1

INTRODUCTION

The near infrared (NIR) region (700 – 2500 nm) of the electromagnetic spectrum was discovered by Sir William Hershel in the 1800's (Herschel, 1800; Cen & He, 2007). The earliest applications, however, were only reported in the 1950's (Kaye, 1954) and was used to analyse agricultural food samples in the late 1960's (Bengera & Norris, 1968). NIR spectroscopy relies heavily computers, for both, data acquisition and analysis (Hindle, 2007). Thus, the advancement of technology led to increased research and use of the technique in a number of fields (Blanco & Villarroya, 2002). NIR spectroscopy is based on the measurement of bond vibrations of organic molecules. Spectra typically contain information about compounds with carbon, nitrogen, or oxygen bonded to a hydrogen atom (X-H), resulting in a large amount of material which is suitable for NIR analysis. It is non-destructive and unlike alternative instrumental techniques, samples do not require any pre-preparation for measurements to be taken. Additional advantages of this technique include, rapid spectra acquisition and prediction of physical and chemical parameters from a single spectrum. Despite these advantages, NIR spectroscopy relies on average point measurements for quantification and qualitative purposes. This could be limiting for heterogeneous samples (Williams, 2009; Manley, 2014; Burger, 2006).

NIR hyperspectral imaging offers a solution to this. It is a technique which couples NIR spectroscopy with imaging, allowing for both spectral and spatial information acquisition (Koehler *et al.*, 2002; Burger & Gowen, 2011; Geladi *et al.*, 2004). The instrument can be operated with ease and speed, making it ideal in the growing and fast-paced food production industry. When images are acquired, data is arranged into a three dimensional data matrix, also known as a hypercube, which consists of one spectral and two spatial dimensions (Geladi *et al.*, 2004; ElMasry & Sun, 2010). Hyperspectral images contain immense amounts of data, thus multivariate image analysis (MIA) is essential for extraction of relevant information, allowing for sample categorisation or quantitative determinations (Burger & Geladi, 2006; Williams *et al.*, 2009). MIA can be grouped into sub-areas, such as (Prats-Montalbán *et al.*, 2011):

- image preprocessing, which reduces spectral noise and thereby improving the visual quality of the image;
- image compression, which removes redundancies present in the image; and
- image analysis, which is used for classification, usually by returning numeric values and/or different graphs related to image characteristics.

For food related research, among other studies, it has been used to predict lamb composition (Kamruzzaman *et al.*, 2012); assess the quality and marbling in pork (Barbin *et al.*, 2012), surface defects and contaminants on apples (Mehl *et al.*, 2004; ElMasry *et al.*, 2008; Xing *et al.*, 2005); evaluate quality attributes of strawberries (ElMasry *et al.*, 2007); classify maize kernels according to hardness (Williams *et al.*, 2009) and bruising on pickling cucumbers (Ariana *et al.*, 2006).

Apart from quality analysis, rapid food safety tests to ensure microbial safety throughout the production process are becoming increasingly important (ElMasry & Sun, 2010; Gowen *et al.*, 2015). The most commonly used techniques for pathogen detection are often time consuming and require highly skilled workers (Rijpens & Herman, 2002; Mandal *et al.*, 2011; Velusamy *et al.*, 2010). In recent years, hyperspectral imaging has been investigated to detect fungi and bacteria (reviewed by Gowen *et al*, 2015).

Regarding bacterial detection and differentiation, NIR hyperspectral imaging has predominantly been used to measure growth directly on food matrixes, such as chicken carcases (Feng *et al.*, 2013; Feng & Sun, 2013b; Feng & Sun, 2013a), fresh spinach leaves (Siripatrawan *et al.*, 2011), mushrooms (Gaston *et al.*, 2011), pork (Barbin *et al.*, 2013) and beef (Peng *et al.*, 2011). Due to the complex spectral nature of food matrices, it is often challenging to identify bacteria on the surfaces. For this reason, detection of bacteria cultured on solid media in Petri dishes were explored. The first studies were done on *Campylobacter* and non-*Campylobacter* species (Yoon *et al.*, 2009; Yoon *et al.*, 2010), followed by non-Shiga toxin-producing *Escherichia coli* (STEC) species (Windham *et al.*, 2012; Yoon *et al.*, 2013a; Yoon *et al.*, 2013b; Yoon *et al.*, 2015). These studies were all done using wavelengths in the vis-NIR region (400-1000 nm), indicating that classifications were based primarily on colour variations of the colonies. However, bacteria often appear similar in colour on culture media, unless selective media is utilised.

The question that still remains unanswered is whether NIR hyperspectral imaging in the 1100 – 2500 nm spectral range would permit accurate detection of bacteria, using chemical constituents as the key identifying features.

Thus, the aim of this study was to use NIR hyperspectral imaging, while only including wavelengths from 1100 to 2500 nm, to evaluate foodborne pathogenic bacteria cultured on solid growth media. Specific objectives were to:

- use MIA to distinguish between Gram positive and Gram negative bacteria, pathogenic and non-pathogenic species, and, bacteria which appear similar in colour on culture media; and
- assess the impact of age of the bacteria on classification results.

References

Ariana, D.P., Lu, R. & Guyer, D.E. (2006). Near-infrared hyperspectral reflectance imaging for detection of bruises on pickling cucumbers. *Computers and Electronics in Agriculture,* **53,** 60-70.

Barbin, D., Elmasry, G., Sun, D.-W. & Allen, P. (2012). Near-infrared hyperspectral imaging for grading and classification of pork. *Meat Science,* **90,** 259-268.

Barbin, D.F., Elmasry, G., Sun, D.-W., Allen, P. & Morsy, N. (2013). Non-destructive assessment of microbial contamination in porcine meat using nir hyperspectral imaging. *Innovative Food Science & Emerging Technologies,* **17,** 180-191.

Bengera, I. & Norris, K. (1968). Influence of fat concentration on absorption spectrum of milk in near-infrared region. *Israel journal of agricultural research,* **18,** 117.

Blanco, M. & Villarroya, I. (2002). Nir spectroscopy: A rapid-response analytical tool. *TRAC Trends in Analytical Chemistry,* **21,** 240-250.

Burger, J. (2006). Hyperspectral nir image analysis. Swedish University of Agricultural Sciences Umeå.

Burger, J. & Geladi, P. (2006). Hyperspectral nir image regression part ii: Dataset preprocessing diagnostics. *Journal of Chemometrics,* **20,** 106-119.

Burger, J. & Gowen, A. (2011). Data handling in hyperspectral image analysis. *Chemometrics and Intelligent Laboratory Systems,* **108,** 13-22.

Cen, H. & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science & Technology,* **18,** 72-83.

Elmasry, G. & Sun, D.-W. (2010). Principles of hyperspectral imaging technology. *Hyperspectral imaging for food quality analysis and control***,** 3-43.

Elmasry, G., Wang, N., Elsayed, A. & Ngadi, M. (2007). Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *Journal of Food Engineering,* **81,** 98-107.

Elmasry, G., Wang, N., Vigneault, C., Qiao, J. & Elsayed, A. (2008). Early detection of apple bruises on different background colors using hyperspectral imaging. *LWT-Food Science and Technology,* **41,** 337-345.

Feng, Y.-Z., Elmasry, G., Sun, D.-W., Scannell, A.G.M., Walsh, D. & Morcy, N. (2013). Near-infrared hyperspectral imaging and partial least squares regression for rapid and reagentless determination of enterobacteriaceae on chicken fillets. *Food Chemistry,* **138,** 1829-1836.

Feng, Y.-Z. & Sun, D.-W. (2013a). Determination of total viable count (tvc) in chicken breast fillets by near-infrared hyperspectral imaging and spectroscopic transforms. *Talanta,* **105,** 244-249.

Feng, Y.-Z. & Sun, D.-W. (2013b). Near-infrared hyperspectral imaging in tandem with partial least squares regression and genetic algorithm for non-destructive determination and visualization of pseudomonas loads in chicken fillets. *Talanta,* **109,** 74-83.

Gaston, E., Frias, J.M., Cullen, P.J., O'donnell, C. & Gowen, A. (2011). Hyperspectral imaging for the detection of microbial spoilage of mushrooms.

Geladi, P., Burger, J. & Lestander, T. (2004). Hyperspectral imaging: Calibration problems and solutions. *Chemometrics and Intelligent Laboratory Systems,* **72,** 209-217.

Gowen, A.A., Feng, Y., Gaston, E. & Valdramidis, V. (2015). Recent applications of hyperspectral imaging in microbiology. *Talanta,* **137,** 44-53.

Herschel, W. (1800). On the power of penetrating into space by telescopes; with a comparative determination of the extent of that power in natural vision, and in telescopes of various sizes and constructions; illustrated by select observations. In: Abstracts of the Papers Printed in the Philosophical Transactions of the Royal Society of London. Pp. 5-7. The Royal Society.

Hindle, P.H. (2007). Historical development. In: Handbook of near-infrared analysis (edited by BURNS, D. A. & CIURCZAK, E. W.). Pp. 3-6. United States of America: CRC press.

Kamruzzaman, M., Elmasry, G., Sun, D.-W. & Allen, P. (2012). Non-destructive prediction and visualization of chemical composition in lamb meat using nir hyperspectral imaging and multivariate regression. *Innovative Food Science & Emerging Technologies,* **16,** 218-226.

Kaye, W. (1954). Near-infrared spectroscopy: I. Spectral identification and analytical applications. *Spectrochimica Acta,* **6,** 257E1-287E2.

Koehler, F.W., Lee, E., Kidder, L.H. & Lewis, E.N. (2002). Near infrared spectroscopy: The practical chemical imaging solution. *Spectroscopy Europe,* **14,** 12-19.

Mandal, P., Biswas, A., Choi, K. & Pal, U. (2011). Methods for rapid detection of foodborne pathogens: An overview. *American Journal of Food Technology,* **6,** 87-102.

Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chemical Society Reviews,* **43,** 8200-8214.

Mehl, P.M., Chen, Y.-R., Kim, M.S. & Chan, D.E. (2004). Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. *Journal of Food Engineering,* **61,** 67-81.

Peng, Y., Zhang, J., Wang, W., Li, Y., Wu, J., Huang, H., Gao, X. & Jiang, W. (2011). Potential prediction of the microbial spoilage of beef using spatially resolved hyperspectral scattering profiles. *Journal of Food Engineering,* **102,** 163-169.

Prats-Montalbán, J.M., De Juan, A. & Ferrer, A. (2011). Multivariate image analysis: A review with applications. *Chemometrics and Intelligent Laboratory Systems,* **107,** 1-23.

Rijpens, N.P. & Herman, L.M. (2002). Molecular methods for identification and detection of bacterial food pathogens. *Journal of AOAC International,* **85,** 984-995.

Siripatrawan, U., Makino, Y., Kawagoe, Y. & Oshita, S. (2011). Rapid detection of escherichia coli contamination in packaged fresh spinach using hyperspectral imaging. *Talanta,* **85,** 276-281.

Velusamy, V., Arshak, K., Korostynska, O., Oliwa, K. & Adley, C. (2010). An overview of foodborne pathogen detection: In the perspective of biosensors. *Biotechnology Advances,* **28,** 232-254.

Williams, P.J. (2009). Near infrared (nir) hyperspectral imaging for evaluation of whole maize kernels: Chemometrics for exploration and classification. In: Department of Food Science. Stellenbosch: University of Stellenbosch.

Williams, P.J., Geladi, P., Fox, G. & Manley, M. (2009). Maize kernel hardness classification by near infrared (nir) hyperspectral imaging and multivariate data analysis. *Analytica Chimica Acta,* **653,** 121-130.

Windham, W., Yoon, S.-C., Ladely, S., Heitschmidt, J., Lawrence, K., Park, B., Narrang, N. & Cray, W. (2012). The effect of regions of interest and spectral pre-processing on the detection of non-0157 shiga-toxin producing *escherichia coli* serogroups on agar media by hyperspectral imaging. *Journal of Near Infrared Spectroscopy,* **20,** 547-558.

Xing, J., Bravo, C., Jancsók, P.T., Ramon, H. & De Baerdemaeker, J. (2005). Detecting bruises on 'golden delicious' apples using hyperspectral imaging with multiple wavebands. *Biosystems Engineering,* **90,** 27-36.

Yoon, S.-C., Lawrence, K.C. & Park, B. (2015). Automatic counting and classification of bacterial colonies using hyperspectral imaging. *Food and Bioprocess Technology***,** 1-19.

Yoon, S.-C., Windham, W.R., Ladely, S.R., Heitschmidt, J.W., Lawrence, K.C., Park, B., Narang, N. & Cray, W.C. (2013a). Hyperspectral imaging for differentiating colonies of non-o157 shiga-toxin producing escherichia coli (stec) serogroups on spread plates of pure cultures. *Journal of Near Infrared Spectroscopy,* **21,** 81-95.

Yoon, S., Lawrence, K., Siragusa, G., Line, J., Park, B. & Feldner, P. (2009). Hyperspectral reflectance imaging for detecting a foodborne pathogen: Campylobacter. *Transactions of the ASABE,* **52,** 651-662.

Yoon, S.C., Lawrence, K.C., Line, J.E., Siragusa, G.R., Feldner, P.W., Park, B. & Windham, W.R. (2010). Detection of campylobacter colonies using hyperspectral imaging. *Sensing and Instrumentation for Food Quality and Safety,* **4,** 35-49.

Yoon, S.C., Windham, W.R., Ladely, S., Heitschmidt, G.W., Lawrence, K.C., Park, B., Narang, N. & Cray, W.C. (2013b). Differentiation of big-six non-o157 shiga-toxin producing escherichia coli (stec) on spread plates of mixed cultures using hyperspectral imaging. *Journal of Food Measurement and Characterization,* **7,** 47-59.

CHAPTER 2

LITERATURE REVIEW

Introduction

In today's ever-growing population, the food industry has to constantly adjust to increasing demand. This leads to an increase in highly mechanised factories pushing out large volumes of product in the shortest time possible, consequently putting strain on food safety. Food is often globally sourced and distributed, trading of food between countries increases the potential for outbreaks across borders, leading to greater economic and social impact than ever before (Velusamy *et al.*, 2010). It has been reported that the majority of foodborne outbreaks are caused by *Campylobacter, Salmonella, Listeria monocytogenes and Escherichia coli* 0157:H7 (Alocilja & Radke, 2003; Chemburu *et al.*, 2005; Zhao *et al.*, 2014). Conventional methods for enumeration and identification of bacteria can be time consuming and labour intensive (Mandal *et al.*, 2011; Gowen *et al.*, 2015), often taking up to several days to detect (Lazcka *et al.*, 2007). For this reason rapid methods for detection and differentiation are constantly being investigated.

Near-infrared (NIR) spectroscopy is a non-destructive technique with high spectral resolution and has successfully been used to analyse various constituents of food materials (Cen & He, 2007). A drawback of standard NIR spectroscopy for the analysis of bacteria, is that it lacks spatial resolution, making it difficult to record representative measurements from heterogeneous samples (e.g. pathogenic contaminants). NIR hyperspectral imaging could be a solution to this problem, since it is capable of incorporating a spatial dimension to measurements (Gowen *et al.*, 2007; Gowen *et al.*, 2015; Qin *et al.*, 2013).

This review will focus on foodborne pathogenic bacteria and the most widely known conventional methods for pathogenic bacterial identification. Followed by the use of spectroscopic techniques, mainly NIR spectroscopy and NIR hyperspectral imaging together with multivariate image analysis.

Foodborne pathogenic bacteria

The morphology of bacterial cells range from short or long to spherical, curved or spiral and can form long chains or clumps depending on the growth conditions (Cabeen & Jacobs-Wagner, 2005; Bhunia, 2007). The nutrient supply, temperature and gaseous content of its environment affect bacterial growth. Bacterial cells can be defined as either Gram positive or Gram negative depending on their cell wall structure (Kämpfer & Glaeser, 2013). Gram positive bacteria are distinguished by the thick, rigid outer most layer comprised of peptidoglycan (strands of repeating units of *N*-acetylmuramic acid-(β1-4) and *N*-acetylglucosamine (MurNAc-GlcNAc), teichoic acid, lipoteichoic

acid and polysaccharides (Shockman & Barren, 1983; Trüper & Schleifer, 2006). In contrast, Gram negative bacteria are more complex, and have an outer membrane layer, a thin peptidoglycan layer, periplasmic space and an inner cytoplasmic membrane. The outer and inner membranes are mainly made up of a lipid bilayer (lipopolysaccharides and phospholipids) and proteins (Bhunia, 2007; Beveridge, 1999; Trüper & Schleifer, 2006). Division of bacterial cells is composed of two events: division of the nucleoid after replication by segregation and cytokinesis. This strategy involves elongation of the cell, formation of the Z-ring (consisting of FtsZ polymers), followed by development of the septal ring and finally synthesis of the septum which leads to two progeny cells (Kuroiwa *et al.*, 2009; Errington *et al.*, 2003). In a closed system under optimal physical and chemical conditions, the general pattern of bacterial cell development can be divided into four stages, the lag phase, exponential growth phase, stationary phase and death phase (Srivastava & Srivastava, 2013). The lag phase is represented by an initial stationary growth phase followed by slow increase in cell numbers. This occurs when a cell population is introduced into fresh medium, old or damaged cells need time to repair and re-establish their synthesis. During the exponential growth phase, cells multiply logarithmically, there is a maximum rate of growth per organism, and a minimum generation time is maintained. In fast growing bacteria, such as *E. coli* (20 min generation time in optimal conditions (Fehlhaber & Krüger, 1998)), this phase usually only persists for a few hours before the generation time lengthens until bacteria are no longer dividing. This period is known as the stationary growth phase and arises when nutrients decrease. When nutrients become depleted, cells start to die off slowly at first and then with increasing rapidity until the establishment of a logarithmic death phase (Buchanan, 1974; Srivastava & Srivastava, 2013). Various genera of bacteria have different generation times and will therefore not display the growth phases at the same time (Overmann, 2006).

In the food industry, bacteria can be separated into three groups; beneficial bacteria, spoilage bacteria and disease causing bacteria. Beneficial bacteria which mainly consist of lactic acid bacteria, include those involved in food fermentations to produce products such as cultured milk products (yogurt) and fermented meats (pepperoni) and vegetables (sauerkraut) (Bhunia, 2007). Apart from flavour components and acids, during fermentation, these bacteria metabolize complex substrates in the food to produce enzymes and antimicrobial agents which improve the product's shelf-life and prevents pathogen proliferation.

Food spoilage bacteria make food unsuitable for human consumption by producing an undesirable flavour, texture and/or appearance when they grow. The environment created by food spoilage bacteria generally depresses the growth of pathogenic bacteria. Foodborne pathogenic bacteria are responsible for causing illness and sometimes fatalities in humans and animals. They can

be divided into three groups; those responsible for food intoxication (the toxin is formed within the food and ingested), toxicoinfection (the toxin is produced inside the host after ingestion) or foodborne infection (ingestion of infective pathogen) (Bhunia, 2007).

Common foodborne pathogenic bacteria include *E. coli* 0157:H7, *Bacillus cereus, Listeria monocytogenes, Salmonella enterica, Clostridium perfringens, Campylobacter jejuni* and *Staphylococcus aureus* (Law *et al.*, 2013; Zhao *et al.*, 2014). Pathogenic bacteria are of serious concern in the food industry as it may not necessarily alter the aesthetic quality of the product, making contamination challenging to detect by visual inspection or odour detection. Microorganisms are ubiquitous and can grow and survive at a variety of temperatures. Human pathogens are usually mesophilic and thrive at temperatures from 30 to 37 ^{o}C but some such as *L. monocytogenes*, can survive and grow at refrigeration to ambient temperatures (1 to 25 ^{o}C). Therefore detection and identification of the pathogenic bacterial contamination requires multiple microbiological tests (Bhunia, 2007).

Conventional detection methods

Conventional methods for testing foods involve growth in culture media, which begins with pre-enrichment of food samples to increase pathogenic bacterial concentrations. This is followed by isolation, using selective media and then biochemical and serological identification (Fratamico & Bayles, 2005; Zhang, 2013). These methods are reliable, where viable cells multiply into visible colonies on agar and are able to produce qualitative and quantitative results, but lacks specificity (Mandal *et al.*, 2011; Escoriza *et al.*, 2006), and are usually time consuming and labour intensive. Preliminary identification usually requires two to three days, and sometimes more than a week is necessary for confirmation of the pathogen species (Zhao *et al.*, 2014). Farber &Daley, (1994) reviewed different pre-enrichment and plating media, as well as the use of suppressing agents in these media for detection of *Listeria moncytogenes*. They focused on selective media, PALCAM, lithium chloride and phenylethanol-moxalactam (LPM) and Oxford agars as these were most commonly used in Europe and North America. Improvements on the earliest enrichment procedure, cold enrichment, were discussed. This method involves inoculating samples in nutrient broth at 4 ^{o}C in order to suppress the growth of most other microorganisms. *Listeria* spp., however, have a generation time of 1.5 days at this temperature, leading to extended enrichment periods which could take up to three months to complete. Additions of nalidixic acid and acriflavine to enrichment broth (EB) allows for samples to be incubated at 30 ^{o}C for two days. For detection and identification, the suspension can then be plated onto Oxford agar and LPM agars. Thereafter, plates are incubated for 24 to 48 h at 35 and 30 ^{o}C respectively. Additional time required for pre-enrichment is especially

problematic where ready-to-eat products are involved (Zhang, 2013; Ivnitski *et al.*, 1999). Products need to be distributed before these tests can be completed, leading to recalls or consumption of contaminated products. Conventional culture methods are usually combined with other pathogen detection methods such as DNA, antibody or biochemical-based methods to yield more robust results (Velusamy *et al.*, 2010).

It is evident that conventional methods of culturing and colony counting are insufficient in preventing the spread of infections by foodborne bacteria and ensuring the safety of consumers. Therefore rapid methods for detection is in constant demand (Zhao *et al.*, 2014).

Rapid detection methods

Any modification to conventional methods that reduces the analysis time is considered to be a rapid method of detection (Mandal *et al.*, 2011). In the development of rapid methods for identification of microorganisms, emphasis is generally on decreasing the time of analysis and increasing sensitivity (Law *et al.*, 2015; Zhao *et al.*, 2014). Many rapid methods are able to detect pathogens that are present in low concentrations in food, however, each rapid method has its own advantages and limitations. These methods are categorised into immunological-based, nucleic acid-based, biosensor-based and vibrational spectroscopy-based methods (Law *et al.*, 2015).

Immunological-based methods

Immunological methods are based on the specific binding of a bacterial antibody or antigen to antibodies (Fratamico & Bayles, 2005) and are widely used to detect foodborne pathogens (Zhao *et al.*, 2014). The most commonly used method is enzyme-linked immunosorbent assay (ELISA). Six different protocols are described by Hornbeck, (1991). It is based on the detection of specific antibodies, soluble antigens or cell-surface antigens. “Sandwich” ELISA is the most popular due to its sensitivity and robustness. It does however have detection limits ranging from 10^3 – 10^5 cfu mL^{-1}, therefore, for direct pathogen detection, an enrichment of at least 16 – 24 hrs is required (Zhao *et al.*, 2014; Foley & Grant, 2007). ELISA has been widely employed to detect toxins produced by bacteria.

Nucleic acid-based methods

These methods are centred on detecting specific DNA or RNA sequences in the target pathogen. Toxin-related genes can be detected and the presence of an organism is indicated by a positive hybridisation (Law *et al.*, 2015; Rijpens & Herman, 2002). The polymerase chain reaction (PCR), real-time/quantitative PCR, loop-mediated isothermal amplification (LAMP), microarray technology and

nucleic acid sequence-based amplification (NASBA) are the most commonly used nucleic acid-based methods.

PCR is a method where specific DNA sequences in a sample are amplified in a three step process. Firstly, the target DNA is denatured at high temperatures, followed by the annealing step, where two primers (oligonucleotides) are annealed to opposite strands of the DNA. And finally polymerization is performed whereby the primers, complementary to the single-stranded DNA, are extended. When this process is repeated, there is an exponential amplification of the target DNA fragment. The products are then visualised as bands on electrophoresis gel. (Mandal *et al.*, 2011; Law *et al.*, 2015). Real-time PCR on the other hand, combines the DNA amplification and detection steps and therefore does not require electrophoresis gel for detection of PCR results. Instead, it is focused on detection of a fluorescence signal produced during amplification of a DNA target. (Bustin *et al.*, 2005; Huggett *et al.*, 2005). The fluorescent dye used in this technique can be detected and followed in real time as it attaches to the target amplicon (Lazcka *et al.*, 2007). The main limitation of PCR is its inability to distinguish between viable and non-viable cells, since DNA is always present, whether the cell is dead or alive (Velusamy *et al.*, 2010; Rohde *et al.*, 2017)

Biosensor-based methods

Biosensors are analytical devices which convert biological responses to electrical signals (Velusamy *et al.*, 2010). There are two main components in biosensors used for detection of pathogenic bacteria: the bioreceptors (nucleic acids, enzymes or antibodies) which identifies the target analyte and transducer which converts the recognition event into and electrical signal (Velusamy *et al.*, 2010; Ivnitski *et al.*, 1999). Biosensor-transducer systems commonly used for detection of foodborne pathogenic bacteria include, mass-selective, electrochemical and optical-based biosensors (Law *et al.*, 2015; Zhang, 2013). These methods do not require a pre-enrichment step like nucleic-acid and immunological methods. Optical biosensors are favoured for pathogen detection due to it's high sensitivity and specificity, and are divided into several subclasses based on absorption, reflection, refraction, Raman, infrared, chemilumenscence, dispersion, fluorescence and phosphorescence (Velusamy *et al.*, 2010; Zhao *et al.*, 2014). These use changes in light absorption or scattering between reactants and products, or measure light output by a luminescent process (Law *et al.*, 2015). When the pathogen binds to the metal surface (usually gold), the interaction alters its refractive index, resulting in the change of wavelength required for electron resonance (Zhang, 2013; Zhao *et al.*, 2014; Law *et al.*, 2015).

Vibrational spectroscopy

In microorganism detection, vibrational spectra provide molecular fingerprint information of cells (Harz *et al.*, 2009). These spectral fingerprints provide chemical information necessary for characterisation and identification of microbial systems. Raman spectroscopy offers chemical information arising from non-polar bonds (Neugebauer, 2007). It relies on a scattering effect of molecules and record a gain or loss of energy (Lee *et al.*, 2014). Raman scattering is known to have a low scattering efficiency, but can be enhanced with instrumental modifications (Harz *et al.*, 2009; Pelletier & Pelletier, 2011; Hering *et al.*, 2008) or with the use of optical fields, spectral substrates or tuning, to influence the physics of the scattering process (Pelletier & Pelletier, 2011). Infrared (IR) absorption spectra consist of mainly molecular vibrations involving a large change in dipole moment (Harz *et al.*, 2009). Near infrared (NIR) spectra (between 800 and 2500 nm) arises from overtones and combinations of overtones originating from fundamental bond vibrations occurring in the mid-infrared (MIR) (2500 – 25000 nm) region of the electromagnetic spectrum (Manley, 2014; Harz *et al.*, 2009). Overtones and combinations are weaker than those in the MIR region, thus eliminating the need for dilutions – NIR spectra can be collected directly from samples. NIR spectroscopy is a technique which uses certain regions within infrared regions of the electromagnetic spectrum and measures the X-H bond vibrations *i.e.* C-H, O-H and N-H. This allows for a large number of molecules which are suitable for NIR analysis as all organic molecules can be analysed (Manley, 2014; Pelletier & Pelletier, 2011).

Vibrational spectroscopic techniques can offer fast and reliable results in microorganism detection and identification as they are non-invasive, and provide characteristic fingerprints of molecular components (Maity *et al.*, 2013). In recent years, Fourier transform infrared (FT-IR) spectroscopy and Raman microscopy with excitation in the visible or near-infrared regions have demonstrated potential in the application of microbial classification (Harz *et al.*, 2009).

NIR hyperspectral imaging

Principles

NIR hyperspectral imaging is an extension of NIR spectroscopy (Lewis & Kidder, 2010). This analytical technique integrates NIR spectroscopy with conventional digital imaging to obtain an image with both spatial and spectral information of a sample (Geladi *et al.*, 2007; Burger & Geladi, 2006; Williams *et al.*, 2009; Gowen *et al.*, 2007). Thus each pixel in the image can be interpreted as a spectrum (Amigo *et al.*, 2013; Burger & Gowen, 2011), thereby enabling the visualisation of the distribution of different chemical components in a sample, making it ideal for heterogeneous samples (Manley, 2014; Gowen *et al.*, 2007).

Instrumentation

The core components of a hyperspectral imaging system are the light source, wavelength modulator and detector. Tungsten-Halogen (TH) lamps are usually used in vis-NIR hyperspectral systems since it covers a range of wavelengths (400-2500 nm), but may be problematic when measuring biological samples as the heat generated by these bulbs are quite significant (Gowen *et al.*, 2015). Fiber optic line lights can be used to create distance between the TH bulb and sample to overcome this problem (Gaston *et al.*, 2011). Light emitting diodes (LED) are also used but LED bulbs used in the NIR range are more expensive than TH sources. For detectors to record light changes, wavelength modulators split broadband incident light into separate wavelengths. Devices used for this purpose include, tunable filters, spectrographs and bandpass filters. Detectors used for wavelengths 400 – 1000 nm (visible-very near-infrared region) usually utilise cameras with charge coupled device (CCD) or complementary metal oxide semiconductor (CMOS). Longer wavelength systems (1000 – 2500 nm), however, require detectors such as Indium Gallium Arsenide (InGaAs) or mercury cadmium telluride (MCT) detectors (Gowen *et al.*, 2015; Geladi *et al.*, 2007).

Three different camera configurations are used in hyperspectral imaging systems i.e. whiskbroom (point scan), pushbroom (line scan) and staredown (area scan) as illustrated in **Fig 2.1**.

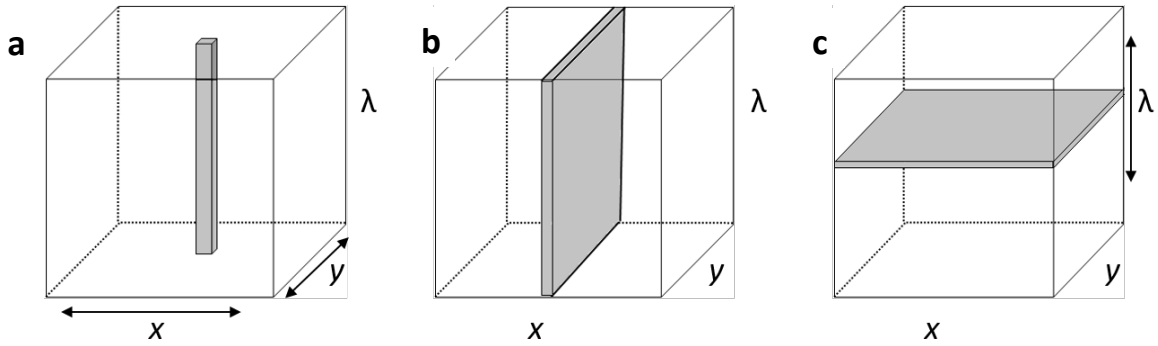

Figure 2.1. Three configurations of hyperspectral imaging systems. **a.** Whiskbroom or point scan configuration – scans sample in point form. **b.** Pushbroom or line scan configuration – scans sample in consecutive lines. **c.** Staredown configuration – scans entire sample (within camera frame), one wavelength at a time.

With the whiskbroom configuration (**Fig 2.1a**), hypercubes are acquired by scanning single points of a sample (pixel by pixel), by either moving the sample or detector. This technique is time consuming, but yields images with high spatial resolution and is most suitable for microscopic imaging (Qin *et al.*, 2013; Gowen *et al.*, 2007; Gowen *et al.*, 2015). The pushbroom method (**Fig 2.1b**) is the fastest, since no filter changes are required (one image can take up to 20 seconds to acquire) and most commonly used with vis-NIR HSI and NIR HSI. Spatial and spectral data is obtained with a two dimensional

detector line-by-line and is well suited for on-line applications (Burger, 2006; Gowen *et al.*, 2007; Geladi *et al.*, 2007; Williams, 2009). Lastly, with the staredown configuration (**Fig 2.1c**), the spectrometer components and the sample remain fixed while the HSI data is obtained by collecting a sequence of images, one wavelength band at a time. Hypercubes consist of a three dimensional stack of images (Williams, 2009; Gowen *et al.*, 2007; Gowen *et al.*, 2015; Burger, 2006).

NIR hyperspectral images

Hyperspectral images are three dimensional multivariate data structures (hypercubes) and consist of hundreds of adjacent wavebands for each spatial position of a sample studied (Gowen *et al.*, 2007; Geladi *et al.*, 2004; Williams, 2013). These hypercubes have two spatial dimensions (pixel coordinates – x and y) and one spectral (wavelength (λ)) dimension (Burger & Gowen, 2011; Geladi *et al.*, 2010). Each pixel represents a spectrum and each spectral dimension acts like a chemical fingerprint, allowing for chemical constituents in the sample to be identified (Koehler *et al.*, 2002; Williams, 2013). Vast amounts of data are contained within each hypercube, with one image from a SisuCHEMA pushbroom imaging system often consisting of more than 100 000 spectra with more than 100 wavelength variables (Burger & Gowen, 2011; Williams, 2013). For this reason, multivariate image analysis (MIA) is an essential tool for approaching and extracting useful information from data of this magnitude.

Chemometrics and multivariate image analysis

The large amount of data obtained from hyperspectral imaging requires multivariate data analysis, which are data analysis techniques employed for exploration and interpretation of hypercubes (Geladi *et al.*, 2007; Esbensen & Geladi, 1989). The fundamental steps involved in hyperspectral image processing usually begins with preprocessing of the original hyperspectral image, followed by image exploration and data reduction/selection, and lastly, quantification or classification of the data (Gowen *et al.*, 2015; Qin *et al.*, 2013).

The general purpose for data analysis is to reduce the dimension of the data while retaining sufficient information to classify and/or quantify chemical and/or physical properties of a sample (Gowen *et al.*, 2015; Qin *et al.*, 2013). Principal component analysis (PCA) is well suited for data reduction and exploration while preserving the original variance within fewer variables (Cowe & McNicol, 1985). The initial step in calculating PCA of hyperspectral images is unfolding the 3 dimensional hypercube into a matrix, on which PCA can be calculated similar to that for regular spectral data (Qin *et al.*, 2013). The residual contains noise and redundant information, therefore

the reduced terms are easier to interpret and have more stability (Geladi, 2003). PCA is calculated using the following equation:

$$\boldsymbol{X} = \boldsymbol{TP}' + \boldsymbol{E} \quad \text{...[eq. 1]}$$

The raw data is held in matrix **X** (unfolded hypercube), where the samples or number of observations are represented in different rows and variables in individual columns of the matrix. The data are mean-centred then decomposed into **T**; the matrix containing the score values, **P'**; the matrix containing the loading values and **E**; the matrix containing residuals. It is possible to reshape the score matrix into a score plot and reorganise this data into a score image. Score images show score values plotted as contour maps of the sample imaged for each principal component (PC). Loading vectors (**p**) may be used for spectral interpretations where loading values are plotted against the wavelengths and show the most prominent spectral variation within a dataset (Geladi *et al.*, 1989; Williams *et al.*, 2012; Williams, 2013; Geladi, 2003; Wold *et al.*, 1987).

Preprocessing

Preprocessing is required to prepare the data for further processing by removing non-chemical biases originating from the instrument or measurement conditions (Qin *et al.*, 2013; Gowen *et al.*, 2015). This is done by transforming data so that the multivariate signals will better adhere to the Beer-Lambert law, which states that absorbance and concentration are proportional (Rinnan *et al.*, 2009a; Swinehart, 1962).

$$A = \varepsilon\, lc \quad \text{...[eq. 2]}$$

Where:

A = absorbance

ε = molar absorptivity

l = path length

c = concentration

Deviation from this linear relationship can originate from various chemical and physical phenomena. Some of these phenomena include molecular interactions, light scatter, stray light, changes in refractive index at high concentrations and change in sample size/path length (Rinnan *et al.*, 2009a). Commonly applied preprocessing methods are divided into two groups, namely: scatter correction methods and spectral derivatives. Background signal noise which affects all the spectra equally is known as constant additive effects. These cause deviations of spectral baseline offsets and slopes and originate from changes in instrumentation, sample orientation and particle size distributions, or packing. Constant additive effects can be compensated for by applying first and second derivative transforms (Burger, 2006; Burger & Geladi, 2006). Spectral derivatives include Savitzky-Golay (SG)

polynomial derivative filters (Savitzky & Golay, 1964) and Norris-Williams (NW) derivatives (Norris, 1959). SG derivations include a smoothing step, where a polynomial is fitted in a systematic window on the raw data, in order to find the derivative at the centre point. (Rinnan *et al.*, 2009b; Burger, 2006). In this moving window averaging method, a window is selected, where the data are fitted by a polynomial of a certain degree (Candolfi *et al.*, 1999). The second derivative is more prone to amplification of noise than the first derivative, therefore a higher number of smoothing points is recommended, as this decreases the overall signal-to-noise ratio (Rinnan *et al.*, 2009a). NW derivation is used to avoid inflation in finite differences and includes two steps. Firstly, smoothing of the spectra by averaging the given number of points (eq. 3).

$$x_{smooth,i} = \frac{\sum_{j=-m}^{m} x_{org,i+j}}{2m+1} \quad \text{...[eq. 3]}$$

Where:

m = number of points in the smoothing window

i = single measurement point in absorbance

The second step differs slightly for first and second derivatives. For the first-order derivation, the difference between two smoothed values with a given gap is calculated (eq. 4), and for second-order derivation, the difference between twice the smoothed value at a point and the smoothed value at a gap distance on either side is calculated (eq. 5). Generally, NW and SG derivations do not give the same estimates (Rinnan *et al.*, 2009b; Norris, 1959).

$$x_i' = x_{smooth,i+gap} - x_{smooth,i-gap} \quad \text{...[eq. 4]}$$

$$x_i'' = x_{smooth,i-gap} - 2 \cdot x_{smooth,i} + x_{smooth,i+gap} \quad \text{...[eq. 5]}$$

Scatter correcting methods used in data analysis is designed to reduce the physical variability between samples due to scatter. The most common scatter correction methods used in data analysis include standard normal variate (SNV) (Barnes *et al.*, 1989) and multiplicative scatter correction (MSC) (Martens *et al.*, 1983; Geladi *et al.*, 1985). Both methods also adjust for baseline shifts between samples. MSC aims to remove non-linearity, caused by scatter, from the data matrix before data modelling. The calculation comprises two steps: estimation and correction of coefficients,

$$x_{org} = b_0 + b_{ref,1} \cdot x_{ref} + e \quad \text{...[eq. 6]}$$

and correction of spectra,

$$x_{corr} = \frac{x_{org} - b_0}{b_{ref,1}} \quad \text{...[eq. 7]}$$

Where:

x_{org} = original sample spectra

x_{ref} = reference spectra

x_{corr} = corrected spectra

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

The average of the calibration spectra is typically used, since the main difficulty encountered with MSC is to select one appropriate reference spectrum, which is most often set as the average of the calibration spectra (Fearn *et al.*, 2009; Rinnan *et al.*, 2009a; Rinnan *et al.*, 2009b). The basic format of the SNV calculation is the same as the second step of the MSC calculation (eq. 7).

$$x_{corr} = \frac{x_{org} - a_0}{a_1} \quad \text{...[eq. 8]}$$

Where:

x_{corr} = corrected spectra

a_0 = average value of spectrum

a_1 = standard deviation of spectrum

Unlike MSC, SNV does not require a common reference spectrum, instead, each observation is processed on its own. Despite this practical advantage, SNV does not involve a least squares fitting the parameter estimation, making this method more prone to noise than MSC (Rinnan *et al.*, 2009a). Although MSC seems to be the more robust method, it should be used with caution, as it may produce additional outliers. Outliers may be accentuated, because it divides the spectrum by a slope that may be small for some spectra (Fearn *et al.*, 2009; Gowen *et al.*, 2015).

The similarities of MSC and SNV is discussed in detail by Dhanoa *et al.*, (1994) and the relation can be presented by the following approximation:

$$\boldsymbol{x}_{MSC} \approx \boldsymbol{x}_{SNV} \cdot \overline{s_x} + \overline{\overline{\boldsymbol{X}}} \quad \text{...[eq. 9]}$$

Where:

$\overline{s_x}$ = average standard deviation of all spectra

$\overline{\overline{\boldsymbol{X}}}$ = grand mean over all spectra

Classification

Once the data has been preprocessed, it is necessary to classify or quantify the data by creating a reliable calibration model. This can be achieved with the supervised classification technique, partial

least squares discriminant analysis (PLS-DA) (Chevallier *et al.*, 2006). PLS-DA is similar to partial least squares (PLS), which is a supervised regression technique that uses the data matrix **X** to explain variation in a reference matrix **Y** while accounting for the variation in **X** (Burger & Geladi, 2006; Davis & Mauer, 2010; Amigo *et al.*, 2015; Wold *et al.*, 2001). It can be used to build models from spectral data to make chemical classifications. When PLS is utilised for classification, it is termed partial least square discriminant analysis (PLS-DA). It consists of PLS regression, where the response variables are replaced with dummy variables describing the categories (Chevallier *et al.*, 2006). This allows for prediction of group membership and thus classification of pixels and classes.

Applications

Infrared spectroscopy

FT-IR is the most common vibrational spectroscopic method applied in bacterial detection and differentiation (Mariey *et al.*, 2001). To discriminate 39 species and subspecies of *Staphylococcus* from *S. aureus*, Lamprell *et al.*, (2006) used Mahalanobis distances in a factorial discriminant analysis to find linear relationships between groups. One model was able to differentiate between *Staphylococcus* species and a second model classified 97 % of 180 *S. aureus* isolates correctly, showing that FT-IR spectroscopy is a method capable of discriminating between species of *Staphylococcus*. It was also possible to differentiate between *Bacillus* species (Beattie *et al.*, 1998). Nine strains – *B. cereus, B. mycoides, B. thuringiensis, B. circulans, B. licheniformis, B. megaterium, B. pumilus, B. subtilis* and and *B. firmus* were discriminated with large Mahalanobis distances between strains and 96 % of spectra correctly assigned to their species groups. More recently, principal component analysis (PCA) was used to explore differences of foodborne pathogens – *B. cereus*, *S. enterica*, *E. coli*, two *L. monocytogenes* strains and two *L. innocua* (Al-Holy *et al.*, 2006). A FT-IR spectrometer was used to collect spectra. An average of 36 scans were collected for each spectrum and 30 spectra were acquired for each bacterial sample. In score plots, Gram negative *E. coli* and *S. enterica* were clustered close to each other and the main variability between all the bacteria was attributed to amide, polysaccharide, lipid and phospholipid differences. Soft independent modelling of class analogy (SIMCA) classification results exhibited approximately 95 % accuracies for all bacteria.

In NIR spectroscopy bacterial detection, spectra was used to correlate fresh salmon fillets with increases in bacterial population after nine days of storage (Tito *et al.*, 2012). The spectra were collected using a fibre optic probe on a Bruker multi-purpose analyser (MPA) over wavelengths 800 – 2500 nm. Each sample spectrum was the average of eight spectra collected from random spots across the sample. It was reported that differences in spectra were most likely due to biochemical

changes in the flesh of the salmon, owing to a combination of autolytic and microbiological proteolysis of muscle proteins. The coefficient of determination (R^2) of calibration and validation partial least squares (PLS) regression models were 0.95 and 0.64 respectively. Using PCA, Alexandrakis *et al.*, (2011) differentiated between inoculated (with *L. innocua, Pseudomonas fluorescens, P. putida, P. mendocina* or *E. coli*) and uninoculated raw chicken. NIR spectroscopy was not able to distinguish between the five bacteria, however, physical changes of the chicken breast muscle due to decomposition caused by bacterial proliferation was again detected. In the raw and Savitzky-Golay (2^{nd} derivative) treated spectra, peaks at 980, 1190 and 1440 nm dominated. These peaks were related to O-H, C-H, NH_2 and $CONH_2$ stretch and bending, first, second and third overtones and combination bands, which could have been due to protein content changes, water absorption and the effect of increased micro flora, their metabolites and storage conditions.

When using NIR spectroscopy, growth substrates of microbial systems present certain challenges in spectral analysis. For example, this technique is extremely sensitive to water, a major component of microbial systems, and exhibits high absorbance in the NIR region of the light spectrum (Gowen *et al.*, 2015). In a study done by Rodriguez-Saona *et al.*, (2004) FT-NIR in conjunction with PCA and SIMCA was investigated to identify *E. coli* spp., *Pseudomonas* spp. and *Bacillus* spp. in liquids, using spiked apple juice as the liquid medium. Bacterial cells were grown in juice, but had to be centrifuged to remove all liquid and create a bacterial film before scanning. Wavelengths 1960-2272 nm enabled differentiation between bacteria genera i.e. *E. coli* species were differentiated from three *Bacillus* species and *Pseudomonas aeruginosa*.

Despite its success in the food microbiology industry, NIR spectroscopy lacks spatial dimension, which may limit heterogeneous sample analysis. NIR hyperspectral imaging provides a solution to this, as it can be used to acquire spectral information of entire samples (Williams, 2013; Williams *et al.*, 2009; Manley, 2014).

NIR hyperspectral imaging

A condensed overview of studies that investigated NIR hyperspectral imaging for detection of pathogenic bacteria on food matrices or growth media is summarized in **Table 2.1**.

Table 2.1. Application of NIR hyperspectral imaging for bacterial detection and classification on food matrixes and growth media

Purpose of study	Wavelengths used (nm)	Microorganisms	Growth medium	Data classification methods	Reference
Identify and differentiate bacteria on food specific cards	1200 – 2350	*Listeria innocua, L. monocytogenes, Bacillus cereus, B. subtilis, Escherichia coli, Salmonella typhimurium, Salmonella entiritidis*	Food cards	PLS1	Dubois *et al.*, (2005)
Determination of TVC on raw chicken fillets	910 – 1700	Total viable count	Chicken fillets	PLSR	Feng &Sun, (2013a)
Differentiate between healthy, mechanically damaged and bacterially damaged mushrooms	445 – 945	*Pseudomonas tolaasii*	Mushrooms	PLS-DA	Gaston *et al.*, (2011)
Determination of Enterobacteriaceae loads on raw chicken fillets	930 – 1450	Enterobacteriaceae	Chicken fillets	BW-PLSR	Feng *et al.*, (2013)
Determination of Pseudomonas loads on raw chicken fillets	897 – 1752	*Pseudomonas*	Chicken fillets	PLS Genetic algorithm	Feng &Sun, (2013b)
Determination of *E. coli* K12 loads on fresh spinach leaves	400 – 1000	*E. coli* K12	Fresh spinach leaves	ANN based on multilayer perceptrons with back propagation algorithm	Siripatrawan *et al.*, (2011)
Determine TVC and PPC at two different temperatures	900 – 1700	Total viable count (TVC) and psychrotrophic plate count(PPC)	Porcine meat	PLS	Barbin *et al.*, (2013)
Measure biochemical changes within beef caused by TVC	400 - 1100	TVC	Raw beef	Multi-linear regression and PLS-R	Peng *et al.*, (2011)
Detecting *Campylobacter* species and non-*Campylobacter* contaminants after 48 h	400 – 900	*Campylobacter*	Campy-Cefex agar, Campy-Line agar and blood agar	Single-band thresholding algorithm	Yoon *et al.*, (2009)
Detecting *Campylobacter*	400 – 900	*Campylobacter*	Campy-Cefex agar and	Spectral feature fitting and	Yoon *et al.*, (2010)

species and non-*Campylobacter* contaminants after 24 h			blood agar	continuum-removed band ratio	
Use Vis-NIR to differentiate between six non-0157 STEC serogroups	400 – 1000	*E. coli* (STEC serogroups)	Rainbow agar	Mahalanobis distance classifiers and PCA-kNN	Windham *et al.*, (2012)
Use VIS-NIR to predict pure cultures	400 – 1000	*E. coli* (STEC serogroups)	Rainbow agar	Mahalanobis distance classifier and *k*NN classifier	Yoon *et al.*, (2013a)
Use classification models built with pure culture plates to predict mixed culture plates	400 – 1000	*E. coli* (STEC serogroups)	Rainbow agar	PCA-kNN	Yoon *et al.*, (2013b)
Automatically count colonies according to their types	400 – 1000	*E. coli* (STEC serogroups)	Rainbow agar	Mahalanobis distance and kNN – based algorithms	Yoon *et al.*, (2015)

PLS1 – partial least squares 1; PLS-R – partial least squares regression; PLS-DA – partial least squares discriminant analysis; BW-PLS -weighted PLS regression; ANN – artificial neural networks; PCA – principal component analysis; kNN – k nearest neighbour

Similarly to NIR spectroscopy, studies using hyperspectral imaging largely correlated biochemical changes in the food structures with spectra, however, with hyperspectral images, prediction maps can be utilised in combination with separately determined bacterial count. An example of this was presented by Peng *et al.*, (2011), where fresh beef samples were imaged (using wavelengths from 400 to 1100 nm), and evaluated for total viable count (TVC) every 12 h over a period of 15 days. After imaging TVC of each sample was measured with the standard plate count method. It was found that the band relating to oxyhaemoglobin absorption (596 nm) correlated to TVC and could be used as a spoilage indicator.

Due to the complex nature of food, it is difficult to identify bacteria, because of interference from the food matrix, one solution to this is to culture and image bacteria on agar media. With relevance to pathogenic bacteria cultured and imaged directly on agar plates, Yoon and co-workers (Yoon *et al.*, 2009; Yoon *et al.*, 2010) conducted two series of studies. The first consisted of two studies focusing on *Campylobacter* identification and the second series focused on differentiation of non-0157 Shiga-toxin producing *E. coli* serogroups (Windham *et al.*, 2012; Yoon *et al.*, 2013b; Yoon *et al.*, 2013a; Yoon *et al.*, 2015). All investigations were conducted using wavelengths in the vis-NIR region (400-1000 nm). In the first investigation, 11 *Campylobacter* were distinguished from 6 non-*Campylobacter* colonies. The bacteria were first cultured in broth for 72 h and then inoculated onto blood agar, Campy-Line agar and Campy-Cefex agar. To avoid confluent growth of colonies and avoid cross-contamination, well-separated point inoculations of 5 µL aliquots were made on allocated spots. The plates were then incubated at 42 ^{o}C for 48 h before image acquisition. Savitzky-Golay smoothing filter (window size: 25; order of moment: 4) was calculated for each spectrum to reduce noise. PCA loadings and theoretical distance measurement, Bhattacharyya distance (Kailath, 1967; Mak & Barnard, 1996) were then used to determine an optimal waveband image for classification of the three classes (*Campylobacter* colonies, non-*Campylobacter* colonies and agar). The best results were obtained with blood agar and cefex agar samples with classification accuracies, for both, above 99 %. In the second study of the first series (Yoon *et al.*, 2010), bacteria were handled in the same way as described above, except that plates were only incubated for 24 h. The results were promising, with up to 99 % classification accuracies for blood agar, however, further in vitro and in situ studies will be required to assess the potential of this technique.

In the second series of studies, conducted on non-0157 Shiga-toxin producing *E. coli* serogroups, six serogroups (026, 045, 0103, 0111, 0121 and 0145) were cultured as spots on Rainbow agar (Windham *et al.*, 2012). Different serogroups appear as different colours on rainbow agar. It was reported that the reason wavelengths from 1000 – 2500 nm were excluded from these studies, was due to interference of water in the agar at certain wavelengths (1450 and 1930 nm) in

this region. Regions of interest representing different classes were selected in the hypercubes to create spectral libraries. Classifications were conducted using a supervised Mahalanobis distance classifier. The models worked well for classification of serogroups 045 and 0121, with an average classification accuracy of 98 %, however, the results for other serogroups were less promising with average classification accuracies varying from 8 – 100 %. This was due to the influence of spectral variability prompted by variations in the growth media and variation in non-Shiga toxin-producing *E.coli* (STEC) populations in the test set images. In order to improve on these results, later work by this research team considered the use of pure and mixed culture spread plates (Yoon *et al.*, 2013a; Yoon *et al.*, 2013b). Classifications were improved, with reported accuracies for cross validated data ranging from 84 – 100 %. This was followed by the last reported study (Yoon *et al.*, 2015) that addressed automatic colony counting and segmentation of touching colonies. The algorithm was developed based on Mahalanobis distance and k nearest neighbours (kNN). The developed colony counting algorithm produced accuracies of above 99 % for the test set. The classification performance for correct colony types ranged from 68.7 – 99.9 % correctly predicted colonies. Despite these positive results, this technique is limiting as it is based on colour variation between bacteria. This method would not be practical where pathogenic bacteria other than the above mentioned *E. coli* species or if specialised media is used. The latter would however be redundant as visual inspection of plates would suffice.

Although bacteria were not scanned on agar plates, a study worth mentioning in this review, was conducted by Dubois *et al.*, (2005). In this study, it was shown that NIR images in the wavelength range of 1200 – 2350 nm could be used to differentiate bacteria deposited on food cards. Suspensions of Gram positive, *Listeria monocytogenes, L. innocua, B. subtilus* and *B. cereus* and Gram negative, *E. coli* and *Samonella* spp. were allowed to dry in the wells of the food cards before being imaged. Gram positive bacteria were successfully separated from Gram negative bacteria based on the intensity of their second derivative spectra. The peak at 1940 nm (O-H stretch + O-H deformation, H_2O) was useful for the discrimination between Gram positive and Gram negative species, possibly due to differences in cell wall composition. It was possible to isolate *B. cereus* using the band at 1685 nm, related to *cis*-double bonds in unsaturated fatty acids, however, it was not possible to distinguish all 5 bacteria using only one wavelength. PLS-DA applied to the full wavelength range showed excellent classification for all species.

Conclusion

Rapid methods for detection and identification of bacteria are in continual demand, especially in the food industry which is under constant pressure and where the consumer's health may be at risk. NIR

hyperspectral imaging has the potential to be used as a tool for pathogen detection and differentiation directly on solid growth media. When used in conjunction with the initial step of conventional detection methods (plating onto solid growth media), it could serve as rapid tool in a microbiology lab to identify pathogenic contaminants and ensure food safety. The spatial feature of hyperspectral imaging allows for characterisation of heterogeneous samples while the spectral feature enables identification of a wide range of multi-constituent surface and sub-surface features. Additional research regarding data analysis and classification is required, especially in terms of which components of bacterial cells are responsible for NIR differentiation. This could assist with building optimal classification models in further investigations. In general, research regarding identification of bacteria on growth media using wavelengths in the NIR region is lacking.

References

Al-Holy, M.A., Lin, M., Al-Qadiri, H., Cavinato, A.G. & Rasco, B.A. (2006). Classification of foodborne pathogens by fourier transform infrared spectroscopy and pattern recognition techniques. *Journal of Rapid Methods & Automation in Microbiology,* **14,** 189-200.

Alexandrakis, D., Downey, G. & Scannell, A.G. (2011). Detection and identification of selected bacteria, inoculated on chicken breast, using near infrared spectroscopy and chemometrics. *Sensing and Instrumentation for Food Quality and Safety,* **5,** 57-62.

Alocilja, E.C. & Radke, S.M. (2003). Market analysis of biosensors for food safety. *Biosensors and Bioelectronics,* **18,** 841-846.

Amigo, J.M., Babamoradi, H. & Elcoroaristizabal, S. (2015). Hyperspectral image analysis. A tutorial. *Analytica Chimica Acta,* **896,** 34-51.

Amigo, J.M., Martí, I. & Gowen, A. (2013). Hyperspectral imaging and chemometrics: A perfect combination for the analysis of food structure, composition and quality. In: Chemometrics in food chemistry (edited by MARINI, F.). Pp. 343-370. Elsevier.

Barbin, D.F., Elmasry, G., Sun, D.-W., Allen, P. & Morsy, N. (2013). Non-destructive assessment of microbial contamination in porcine meat using nir hyperspectral imaging. *Innovative Food Science & Emerging Technologies,* **17,** 180-191.

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

Beattie, S.H., Holt, C., Hirst, D. & Williams, A.G. (1998). Discrimination among *Bacillus cereus*, B. *mycoides* and *B. thuringiensis* and some other species of the genus bacillus by fourier transform infrared spectroscopy. *FEMS Microbiology Letters,* **164,** 201-206.

Beveridge, T.J. (1999). Structures of Gram-negative cell walls and their derived membrane vesicles. *Journal of Bacteriology,* **181,** 4725-4733.

Bhunia, A. (2007). Foodborne microbial pathogens: Mechanisms and pathogenesis. Springer Science & Business Media.

Buchanan, R.E. (1974). Life phases in a bacterial culture. In: Micriobial growth (edited by DAWSON, P. S. S.). Pp. 25-41. Stroudsburg, Pennsylvania: Dowden, Hutchinson & Ross Inc.

Burger, J. (2006). Hyperspectral nir image analysis. Swedish University of Agricultural Sciences Umeå.

Burger, J. & Geladi, P. (2006). Hyperspectral nir image regression part ii: Dataset preprocessing diagnostics. *Journal of Chemometrics,* **20,** 106-119.

Burger, J. & Gowen, A. (2011). Data handling in hyperspectral image analysis. *Chemometrics and Intelligent Laboratory Systems,* **108,** 13-22.

Bustin, S., Benes, V., Nolan, T. & Pfaffl, M. (2005). Quantitative real-time PCR–a perspective. *Journal of Molecular Endocrinology,* **34,** 597-601.

Cabeen, M.T. & Jacobs-Wagner, C. (2005). Bacterial cell shape. *Nature reviews. Microbiology,* **3,** 601.

Candolfi, A., De Maesschalck, R., Jouan-Rimbaud, D., Hailey, P. & Massart, D. (1999). The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra. *Journal of Pharmaceutical and Biomedical Analysis,* **21,** 115-132.

Cen, H. & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science & Technology,* **18,** 72-83.

Chemburu, S., Wilkins, E. & Abdel-Hamid, I. (2005). Detection of pathogenic bacteria in food samples using highly-dispersed carbon particles. *Biosensors and Bioelectronics,* **21,** 491-499.

Chevallier, S., Bertrand, D., Kohler, A. & Courcoux, P. (2006). Application of PLS-DA in multivariate image analysis. *Journal of Chemometrics,* **20,** 221-229.

Cowe, I.A. & Mcnicol, J.W. (1985). The use of principal components in the analysis of near-infrared spectra. *Applied Spectroscopy,* **39,** 257-266.

Davis, R. & Mauer, L. (2010). Fourier transform infrared (FT-IR) spectroscopy: A rapid tool for detection and analysis of foodborne pathogenic bacteria. In: Current research, technology and education topics in applied microbiology and microbial biotechnology (edited by MENDEZ-VILAS, A.). Pp. 1582-1594. Formatex Research Center.

Dhanoa, M., Lister, S., Sanderson, R. & Barnes, R. (1994). The link between multiplicative scatter correction (msc) and standard normal variate (SNV) transformations of NIR spectra. *J. Near Infrared Spectroscopy,* **2,** 43-47.

Dubois, J., Lewis, E.N., Fry, F.S. & Calvey, E.M. (2005). Bacterial identification by near-infrared chemical imaging of food-specific cards. *Food Microbiology,* **22,** 577-583.

Errington, J., Daniel, R.A. & Scheffers, D.-J. (2003). Cytokinesis in bacteria. *Microbiology and Molecular Biology Reviews,* **67,** 52-65.

Esbensen, K. & Geladi, P. (1989). Strategy of multivariate image analysis (mia). *Chemometrics and Intelligent Laboratory Systems,* **7,** 67-86.

Escoriza, M.F., Vanbriesen, J.M., Stewart, S., Maier, J. & Treado, P.J. (2006). Raman spectroscopy and chemical imaging for quantification of filtered waterborne bacteria. *Journal of Microbiological Methods,* **66,** 63-72.

Farber, J. & Daley, E. (1994). Presence and growth of listeria monocytogenes in naturally-contaminated meats. *International Journal of Food Microbiology,* **22,** 33-42.

Fearn, T., Riccioli, C., Garrido-Varo, A. & Guerrero-Ginel, J.E. (2009). On the geometry of snv and msc. *Chemometrics and Intelligent Laboratory Systems,* **96,** 22-26.

Fehlhaber, K. & Krüger, G. (1998). The study of S*almonella enteritidis* growth kinetics using rapid automated bacterial impedance technique. *Journal of Applied Microbiology,* **84,** 945-949.

Feng, Y.-Z., Elmasry, G., Sun, D.-W., Scannell, A.G.M., Walsh, D. & Morcy, N. (2013). Near-infrared hyperspectral imaging and partial least squares regression for rapid and reagentless determination of enterobacteriaceae on chicken fillets. *Food Chemistry,* **138,** 1829-1836.

Feng, Y.-Z. & Sun, D.-W. (2013a). Determination of total viable count (tvc) in chicken breast fillets by near-infrared hyperspectral imaging and spectroscopic transforms. *Talanta,* **105,** 244-249.

Feng, Y.-Z. & Sun, D.-W. (2013b). Near-infrared hyperspectral imaging in tandem with partial least squares regression and genetic algorithm for non-destructive determination and visualization of pseudomonas loads in chicken fillets. *Talanta,* **109,** 74-83.

Foley, S.L. & Grant, K. (2007). Molecular techniques of detection and discrimination of foodborne pathogens and their toxins. In: Foodborne diseases. Pp. 485-510. Springer.

Fratamico, P.M. & Bayles, D.O. (2005). Molecular approaches for detection, identification, and analysis of foodborne pathogens. In: *Foodborne pathogens: Microbiology and molecular biology* (edited by FRATAMICO, P. M., BHUNIA, A. K. & SMITH, J. L.). Pp. 1-13. Norfolk UK: Horizon Scientific Press.

Gaston, E., Frias, J.M., Cullen, P.J., O'donnell, C. & Gowen, A. (2011). Hyperspectral imaging for the detection of microbial spoilage of mushrooms.

Geladi, P. (2003). Chemometrics in spectroscopy. Part 1. Classical chemometrics. *Spectrochimica Acta Part B: Atomic Spectroscopy,* **58,** 767-782.

Geladi, P., Burger, J. & Lestander, T. (2004). Hyperspectral imaging: Calibration problems and solutions. *Chemometrics and Intelligent Laboratory Systems,* **72,** 209-217.

Geladi, P., Grahn, H. & Burger, J. (2007). Multivariate images, hyperspectral imaging: Background and equipment. In: Techniques and applications of hyperspectral image analysis. Pp. 1-15. Southern West Sussex, England: John Wiley & Sons Ltd.

Geladi, P., Grahn, H. & Manley, M. (2010). Data analysis and chemometrics for hyperspectral imaging. In: Raman, infrared, and near-infrared chemical imaging. Pp. 93-107.

Geladi, P., Isaksson, H., Lindqvist, L., Wold, S. & Esbensen, K. (1989). Principal component analysis of multivariate images. *Chemometrics and Intelligent Laboratory Systems,* **5,** 209-220.

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

Gowen, A.A., Feng, Y., Gaston, E. & Valdramidis, V. (2015). Recent applications of hyperspectral imaging in microbiology. *Talanta,* **137,** 44-53.

Gowen, A.A., O'donnell, C.P., Cullen, P.J., Downey, G. & Frias, J.M. (2007). Hyperspectral imaging – an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology,* **18,** 590-598.

Harz, M., Rösch, P. & Popp, J. (2009). Vibrational spectroscopy—a powerful tool for the rapid identification of microbial cells at the single-cell level. *Cytometry Part A,* **75A,** 104-113.

Hering, K., Cialla, D., Ackermann, K., Dörfer, T., Möller, R., Schneidewind, H., Mattheis, R., Fritzsche, W., Rösch, P. & Popp, J. (2008). SERS: A versatile tool in chemical and biochemical diagnostics. *Analytical and Bioanalytical Chemistry,* **390,** 113-124.

Hornbeck, P.V. (1991). Enzyme-linked immunosorbent assays. *Current protocols in immunology***,** 2.1. 1-2.1. 23.

Huggett, J., Dheda, K., Bustin, S. & Zumla, A. (2005). Real-time PCR normalisation; strategies and considerations. *Genes and Immunity,* **6,** 279-284.

Ivnitski, D., Abdel-Hamid, I., Atanasov, P. & Wilkins, E. (1999). Biosensors for detection of pathogenic bacteria. *Biosensors and Bioelectronics,* **14,** 599-624.

Kailath, T. (1967). The divergence and bhattacharyya distance measures in signal selection. *IEEE transactions on communication technology,* **15,** 52-60.

Kämpfer, P. & Glaeser, S.P. (2013). Prokaryote characterization and identification. In: The prokaryotes: Prokaryotic biology and symbiotic associations (edited by ROSENBERG, E., DELONG, E. F., LORY, S., STACKEBRANDT, E. & THOMPSON, F.). Pp. 123-147. Berlin, Heidelberg: Springer Berlin Heidelberg.

Koehler, F.W., Lee, E., Kidder, L.H. & Lewis, E.N. (2002). Near infrared spectroscopy: The practical chemical imaging solution. *Spectroscopy Europe,* **14,** 12-19.

Kuroiwa, T., Misumi, O., Nishida, K., Yagisawa, F., Yoshida, Y., Fujiwara, T. & Kuroiwa, H. (2009). Bacterial, microbody (peroxisome), mitosome, and hydrogenosome division machineries In: International review of cell and molecular biology (edited by JEON, K. W.). Pp. 111-117. California, USA: Elsevier Science.

Lamprell, H., Mazerolles, G., Kodjo, A., Chamba, J., Noel, Y. & Beuvier, E. (2006). Discrimination of *Staphylococcus aureus* strains from different species of Staphylococcus using fourier transform infrared (FTIR) spectroscopy. *International Journal of Food Microbiology,* **108,** 125-129.

Law, J., Ab Mutalib, N., Chan, K. & Lee, L. (2013). Rapid methods for the detection of foodborne bacterial pathogens: Principles, applications, advantages and limitations. *Frontiers in microbiology,* **5**.

Law, J.W.-F., Ab Mutalib, N.-S., Chan, K.-G. & Lee, L.-H. (2015). Rapid methods for the detection of foodborne bacterial pathogens: Principles, applications, advantages and limitations. *Frontiers in microbiology,* **5,** 770.

Lazcka, O., Del Campo, F.J. & Munoz, F.X. (2007). Pathogen detection: A perspective of traditional methods and biosensors. *Biosensors and Bioelectronics,* **22,** 1205-1217.

Lee, K.-M., Herrman, T.J. & Yun, U. (2014). Application of raman spectroscopy for qualitative and quantitative analysis of aflatoxins in ground maize samples. *Journal of Cereal Science,* **59,** 70-78.

Lewis, E.N. & Kidder, L.H. (2010). Technologies and practical considerations for implementing near-infrared chemical imaging. In: Raman, infrared, and near-infrared chemical imaging. Pp. 75-91. John Wiley & Sons, Inc.

Maity, J.P., Kar, S., Lin, C.-M., Chen, C.-Y., Chang, Y.-F., Jean, J.-S. & Kulp, T.R. (2013). Identification and discrimination of bacteria using fourier transform infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* **116,** 478-484.

Mak, B. & Barnard, E. (1996). Phone clustering using the bhattacharyya distance. In: Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on. Pp. 2005-2008. IEEE.

Mandal, P., Biswas, A., Choi, K. & Pal, U. (2011). Methods for rapid detection of foodborne pathogens: An overview. *American Journal of Food Technology,* **6,** 87-102.

Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chemical Society Reviews,* **43,** 8200-8214.

Mariey, L., Signolle, J., Amiel, C. & Travert, J. (2001). Discrimination, classification, identification of microorganisms using FTIR spectroscopy and chemometrics. *Vibrational Spectroscopy,* **26,** 151-159.

Martens, H., Jensen, S. & Geladi, P. (1983). Multivariate linearity transformation for near-infrared reflectance spectrometry. In: Proceedings of the Nordic symposium on applied statistics. Pp. 205-234. Stokkand Forlag Publishers Stavanger, Norway.

Neugebauer, U. (2007). Characterization of bacteria, antibiotics of the fluoroquinolone type and their biological targets DNA and gyrase utilizing the unique potential of vibrational spectroscopy.

Norris, K. (1959). Infra-red spectroscopy and its application to microbiology. *Journal of Hygiene,* **57,** 326-345.

Overmann, J. (2006). Principles of enrichment, isolation, cultivation and preservation of prokaryotes. In: The prokaryotes. Pp. 80-136. Springer.

Pelletier, M.J. & Pelletier, C.C. (2011). Spectrocopic theory for chemical imaging. In: Raman, infrared, and near-infrared chemical imaging (edited by SASIC, S. & OZAKI, Y.). Pp. 16-50. Hoboken, New Jersey: John Wiley & Sons.

Peng, Y., Zhang, J., Wang, W., Li, Y., Wu, J., Huang, H., Gao, X. & Jiang, W. (2011). Potential prediction of the microbial spoilage of beef using spatially resolved hyperspectral scattering profiles. *Journal of Food Engineering,* **102,** 163-169.

Qin, J., Chao, K., Kim, M.S., Lu, R. & Burks, T.F. (2013). Hyperspectral and multispectral imaging for evaluating food safety and quality. *Journal of Food Engineering,* **118,** 157-171.

Rijpens, N.P. & Herman, L.M. (2002). Molecular methods for identification and detection of bacterial food pathogens. *Journal of AOAC International,* **85,** 984-995.

Rinnan, Å., Nørgaard, L., Berg, F.V.D., Thygesen, J., Bro, R. & Engelsen, S.B. (2009a). Chapter 2 - data pre-processing a2 - sun, da-wen. In: Infrared spectroscopy for food quality analysis and control. Pp. 29-50. San Diego: Academic Press.

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

Rodriguez-Saona, L.E., Khambaty, F.M., Fry, F.S., Dubois, J. & Calvey, E.M. (2004). Detection and identification of bacteria in a juice matrix with Fourier transform-near infrared spectroscopy and multivariate analysis. *Journal of Food Protection,* **67,** 2555-2559.

Rohde, A., Hammerl, J.A., Boone, I., Jansen, W., Fohler, S., Klein, G., Dieckmann, R. & Al Dahouk, S. (2017). Overview of validated alternative methods for the detection of foodborne bacterial pathogens. *Trends in Food Science & Technology,* **62,** 113-118.

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

Shockman, G.D. & Barren, J. (1983). Structure, function, and assembly of cell walls of Gram-positive bacteria. *Annual Reviews in Microbiology,* **37,** 501-527.

Siripatrawan, U., Makino, Y., Kawagoe, Y. & Oshita, S. (2011). Rapid detection of *Escherichia coli* contamination in packaged fresh spinach using hyperspectral imaging. *Talanta,* **85,** 276-281.

Srivastava, S. & Srivastava, P.S. (2013). Bacteria and life processes - growth and multiplication. In: Understanding bacteria. Pp. 97-148. Springer Netherlands.

Swinehart, D. (1962). The beer-lambert law. *J. Chem. Educ,* **39,** 333.

Tito, N., Rodemann, T. & Powell, S. (2012). Use of near infrared spectroscopy to predict microbial numbers on atlantic salmon. *Food Microbiology,* **32,** 431-436.

Trüper, H.G. & Schleifer, K.H. (2006). Prokaryote characterization and identification. In: The prokaryotes. Pp. 58-79. Springer.

Velusamy, V., Arshak, K., Korostynska, O., Oliwa, K. & Adley, C. (2010). An overview of foodborne pathogen detection: In the perspective of biosensors. *Biotechnology Advances,* **28,** 232-254.

Williams, P.J. (2009). Near infrared (NIR) hyperspectral imaging for evaluation of whole maize kernels: Chemometrics for exploration and classification. In: Department of Food Science. Stellenbosch: University of Stellenbosch.

Williams, P.J. (2013). Near infrared (NIR) hyperspectral imaging and X-ray computed tomography combined with statistical and multivariate data analysis to study fusarium infection in maize. In: Department of Food Science. Stellenbosch: Stellenbosch University.

Williams, P.J., Geladi, P., Britz, T.J. & Manley, M. (2012). Growth characteristics of three fusarium species evaluated by near-infrared hyperspectral imaging and multivariate image analysis. *Applied Microbiology and Biotechnology,* **96,** 803-813.

Williams, P.J., Geladi, P., Fox, G. & Manley, M. (2009). Maize kernel hardness classification by near infrared (nir) hyperspectral imaging and multivariate data analysis. *Analytica Chimica Acta,* **653,** 121-130.

Windham, W., Yoon, S.-C., Ladely, S., Heitschmidt, J., Lawrence, K., Park, B., Narrang, N. & Cray, W. (2012). The effect of regions of interest and spectral pre-processing on the detection of non-0157 shiga-toxin producing *Escherichia coli* serogroups on agar media by hyperspectral imaging. *Journal of Near Infrared Spectroscopy,* **20,** 547-558.

Wold, S., Esbensen, K. & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems,* **2,** 37-52.

Wold, S., Sjöström, M. & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems,* **58,** 109-130.

Yoon, S.-C., Lawrence, K.C. & Park, B. (2015). Automatic counting and classification of bacterial colonies using hyperspectral imaging. *Food and Bioprocess Technology***,** 1-19.

Yoon, S.-C., Windham, W.R., Ladely, S.R., Heitschmidt, J.W., Lawrence, K.C., Park, B., Narang, N. & Cray, W.C. (2013a). Hyperspectral imaging for differentiating colonies of non-o157 shiga-toxin producing *Escherichia coli* (STEC) serogroups on spread plates of pure cultures. *Journal of Near Infrared Spectroscopy,* **21,** 81-95.

Yoon, S., Lawrence, K., Siragusa, G., Line, J., Park, B. & Feldner, P. (2009). Hyperspectral reflectance imaging for detecting a foodborne pathogen: *Campylobacter*. *Transactions of the ASABE,* **52,** 651-662.

Yoon, S.C., Lawrence, K.C., Line, J.E., Siragusa, G.R., Feldner, P.W., Park, B. & Windham, W.R. (2010). Detection of *Campylobacter* colonies using hyperspectral imaging. *Sensing and Instrumentation for Food Quality and Safety,* **4,** 35-49.

Yoon, S.C., Windham, W.R., Ladely, S., Heitschmidt, G.W., Lawrence, K.C., Park, B., Narang, N. & Cray, W.C. (2013b). Differentiation of big-six non-o157 shiga-toxin producing E*scherichia coli* (STEC) on spread plates of mixed cultures using hyperspectral imaging. *Journal of Food Measurement and Characterization,* **7,** 47-59.

Zhang, G. (2013). Foodborne pathogenic bacteria detection: An evaluation of current and developing methods. *The Meducator,* **1**.

Zhao, X., Lin, C.-W., Wang, J. & Oh, D.H. (2014). Advances in rapid detection methods for foodborne pathogens. *J. Microbiol. Biotechnol,* **24,** 297-312.

UNIVERSITEIT·STELLENBOSCH·UNIVERSITY
jou kennisvennoot • your knowledge partner

DECLARATION BY THE CANDIDATE

With regard to *Chapter 3, pages 28-50*, the nature and scope of my contribution were as follows:

Nature of contribution	Extent of contribution (%)
First author, research, data analysis, initial writing-up	60%

The following co-authors have contributed to *Chapter 3, pages 28-50*:

Name	e-mail address	Nature of contribution	Extent of contribution (%)
Dr Paul J Williams	pauljw@sun.ac.za	Research inputs, proofreading, editorial inputs and corresponding author	15
Prof Marena Manley	mman@sun.ac.za	Research inputs, proofreading, editorial inputs	15
Prof Pieter Gouws	pgouws@sun.ac.za	Research inputs, proofreading, editorial inputs	10

T-L Kammies	21 September 2018 **Date**

Fakulteit AgriWetenskappe
•
Faculty of AgriSciences

Kantoor van die Dekaan • Office of the Dean
Privaatsak/Private Bag X1 • Matieland 7602 • Suid-Afrika/South Africa
Tel: +27 21 808 4792 • Faks/Fax: +27 21 808 2001
ccav@sun.ac.za • www.sun.ac.za/agric

DECLARATION BY CO-AUTHORS

The undersigned hereby confirm that:

1. the declaration above accurately reflects the nature and extent of the contributions of the candidate and the co-authors to *Chapter 3, pages 28-50*,

2. no other authors contributed to *Chapter 3, pages 28-50* besides those specified above, and

3. potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in *Chapter 3, pages 28-50* of this dissertation.

Signature	Institutional affiliation	Date
P. J. Williams	Department of Food Science, Stellenbosch University	21 September 2018
M. M. Manley	Department of Food Science, Stellenbosch University	21 September 2018
P. G. Gouws	Department of Food Science, Stellenbosch University	21 September 2018

CHAPTER 3

Differentiation of foodborne bacteria using NIR hyperspectral imaging and multivariate data analysis

Abstract

The potential for near-infrared (NIR) hyperspectral imaging and multivariate data analysis to be used as a rapid non-destructive tool for detection and differentiation of bacteria was investigated. NIR hyperspectral images were collected of *Bacillus cereus, Escherichia coli, Salmonella enteritidis, Staphylococcus aureus* and *Staphylococcus epidermidis* grown on agar for 20 h at 37 °C. Principal component analysis (PCA) was applied to mean-centered data of three groups of bacteria, each containing different combinations of the bacteria. Standard normal variate (SNV) correction and the Savitzky-Golay technique was applied (2nd derivative, 3rd order polynomial; 25 point smoothing) to wavelengths in the range of 1103 to 2471 nm. Chemical differences between colonies which appeared similar in colour on growth media (*B. cereus, E. coli* and *S. enteritidis*.) were evident in the PCA score plots. It was possible to distinguish *B. cereus* from *E. coli* and *S. enteritidis* along PC1 (58.1 % SS) and between *E. coli* and *S. enteritidis* in the direction of PC2 (7.75 % SS). *S. epidermidis* was separated from *B. cereus* and *S. aureus* along PC1 (37.5% SS) and was attributed to variation in amino acid and carbohydrate content. Two clusters were evident in the PC1 vs. PC2 PCA score plot for the images of *S. aureus* and *S. epidermidis*, thus permitting distinction between species. Differentiation between genera (similarly coloured on growth media), Gram positive and Gram negative bacteria and pathogenic and non-pathogenic species was possible using NIR hyperspectral imaging. Partial least squares discriminant analysis (PLS-DA) models were used to confirm the PCA data. The best predictions were made of *B. cereus* and *Staphylococcus* species, where results ranged from 82.0-99.96% correctly predicted pixels.

Published as: Kammies, T.-L., Manley, M., Gouws, P.A. & Williams, P.J. (2016). Differentiation of foodborne bacteria using NIR hyperspectral imaging and multivariate data analysis. *Applied Microbiology and Biotechnology,* **100,** 9305-9320.

Introduction

Rapid methods for detection and differentiation of pathogenic bacteria is becoming increasingly important in the food industry as conventional microbiological and immunological methods are time-consuming and labour intensive (Mandal *et al.*, 2011; Gowen *et al.*, 2015). Culturing and colony counting could take several days to detect and identify bacteria (Lazcka *et al.*, 2007). This may lead to a delay in product distribution or may result in contaminated fresh produce or ready-to-eat products being distributed before microbiological tests have been completed. For these reasons, there are continuous developments of rapid and reliable detection methods for pathogenic bacteria (Mandal *et al.*, 2011; Gowen *et al.*, 2015). Most current rapid techniques for bacterial detection include the polymerase chain reaction (PCR), real-time-PCR, enzyme-linked immunosorbent assay (ELISA); the more recently developed biosensor technology and spectroscopic techniques (Lazcka *et al.*, 2007).

Fourier transform infrared (FT-IR) spectroscopy is the most common spectroscopic technique used for the discrimination and classification of bacteria (Mariey *et al.*, 2001), and has been used since the late 1950s (Norris, 1959). Maity *et al.*, (2013) studied the identification and discrimination of bacteria (two strains of *Bacillus flexus* and two strains of *Stenotrophomonas maltophilia*) using FT-IR and found that amides (proteins) and lipids were the main source of variation. However, near-infrared (NIR) spectroscopy has become an increasingly popular technique used for various analyses in many industries as it is able to take rapid, accurate and non-destructive measurements of chemical constituents (Manley, 2014). NIR transflectance spectra of *Listeria innocua, Lactococcus lactis* and three *Pseudomonas* species in suspensions were investigated (Alexandrakis *et al.*, 2008). It was reported that partial least square 2 discriminant analysis (PLS2-DA) regression classifications resulted in 100 % accuracy except for *L. innocua* (77%). A spectral database of *Escherichia coli, Pseudomonas aeruginosa, Bacillus subtilis, B. cereus,* and *B. thuringiensis* was developed using FT-NIR and multivariate data analysis (Rodriguez-Saona *et al.*, 2004). Bacterial cells were concentrated on membranes before FT-NIR measurements were taken. Classification models were developed using soft independent modelling of class analogy (SIMCA). *E. coli* spiked apple juice was then analysed, and the information acquired during this data collection was used to identify the bacteria obtained from filtered juice. The models were able to detect *E. coli* on a genera level but not capable of distinguishing between species.

In another study, the total bacterial load on flounder fillets were evaluated using a portable NIR spectrometer (600-1100 nm) (Duan *et al.*, 2014). With the use of wavelet transform and genetic algorithm (GA) in combination with back-propagation artificial neural network (BP-ANN), promising results were achieved. The correlation coefficient (R) and root mean square error (RMSE) was

determined to be 0.985 and 0.095 respectively for the training set, and 0.966 and 0.083 respectively for the validated data. Although accurate, NIR spectroscopy lacks spatial dimension as it only measures a single point in the sample (Williams *et al.*, 2012c). This could pose a problem if a contamination is present or when analyses are done on heterogeneous samples, like bacterial colonies.

Near-infrared hyperspectral imaging (NIR-HSI) is a technique that provides a spatial map of spectral variation of an entire sample resulting in NIR hyperspectral images (Williams *et al.*, 2012e; Gowen *et al.*, 2015). When applied to intact microbial cells, this technique provides characteristic spectral signatures and can therefore be used as a rapid method for detection and differentiation (Windham *et al.*, 2012; Williams *et al.*, 2012e). Hyperspectral data sets are referred to as hypercubes (Gowen *et al.*, 2015) and can be described as a three dimensional cube comprised of one spectral dimension (wavelength - λ) and two spatial dimensions (pixel coordinates – *x* and *y*) (Geladi *et al.*, 2010). Together, these allow identification of various chemical constituents present in samples where each pixel represents a spectrum (Williams *et al.*, 2012c). The size of data obtained often comprises tens of thousands of spectra and could pose problems during analyses. For this reason, multivariate image analysis (MIA) is required (Esbensen & Geladi, 1989).

Principal component analysis (PCA) forms the basis of MIA (Geladi *et al.*, 1989). It is an unsupervised technique commonly used to reduce the dimension of data and classify information in hyperspectral images. The direction of maximum variability in the samples are obtained and used as a new set of axes, known as principal components (PCs) (Cowe & McNicol, 1985). This data is then used when further processing is done. Prediction models may be built using regression techniques that can be used to identify unknown samples. PLS is a regression technique which uses the data matrix **X** to explain variation in a reference matrix **Y** while accounting for the variation matrix in **X** (Williams *et al.*, 2012d; Burger, 2006; Davis & Mauer, 2010). Partial least square discriminant analysis (PLS-DA), similar to PLS regression, is used for classification of hyperspectral images (Chevallier *et al.*, 2006). It consists of PLS regression where the response matrix is composed of dummy variables describing the categories (Williams *et al.*, 2012e; Amigo *et al.*, 2013).

There have been numerous publications reporting the use of the NIR region to detect and quantify bacteria. Rodriguez-Saona *et al.*, (2001) used the NIR region to detect and identify bacterial contamination in liquid samples, while Dubois *et al.*, (2005a) successfully identified bacteria on food cards. He &Sun, (2015) used the 900-1700 nm region to evaluate the potential of NIR hyperspectral imaging to predict *Pseudomonas* spp. distribution in salmon fillets while Feng *et al.*, (2013) quantified Enterobacteriaceae on chicken fillets in the spectral range 930-1450 nm. In a similar

study, *Pseudomonas* spp. loads were determined on chicken fillets in the 900-1700 nm range. This evidence suggests that bacteria are not transparent in the NIR region.

VIS-NIR hyperspectral imaging was investigated as a rapid method for detection of *E. coli* contamination in packaged fresh spinach (Siripatrawan *et al.*, 2011). *E. coli* K12 inoculated spinach, with different initial concentrations, was imaged in the spectral range of 400 to 1000 nm and PCA and artificial neural network (ANN) were used to classify the data. The ANN data was used to construct a prediction map displaying the number of *E. coli* in the sample. The results suggested that this technique can be used to detect *E. coli* contamination in packaged fresh spinach. A number of similar studies have evaluated the use of vis-NIR hyperspectral, NIR hyperspectral and multispectral imaging to detect bacterial contamination on commodities such as pork (Dissing *et al.*, 2012), poultry carcasses (Park *et al.*, 2004), salmon fillets (He & Sun, 2015) and apples (Liu *et al.*, 2007). Each of these studies is promising and suggest that microbial contamination can be detected, using NIR hyperspectral imaging, directly on a product. Although the instrumentation is now more readily available than a decade ago, current food microbiology methods practiced in the food industry still relies heavily on classical techniques *i.e.* culturing, enrichment, enumeration, staining and microscopy. Developing hyperspectral imaging methods that detect and distinguish bacteria on growth media would thus be invaluable.

Williams *et al.*, (2012e) used NIR-HSI to study the growth characteristics and differences between three species each of *Fusarium subglutinans, F. proliferatum* and *F. verticillioides* cultured on solid media. With the use of MIA, it was established that discrimination between species was possible. PLS-DA prediction results were as follows: 16-47 % correctly predicted pixels for *F. verticilloides,* 78-100 % for *F. subglutinans* and 60-80 % for *F. proliferatum*. Visualisation of radial growth over time was also possible.

The use of NIR-HSI for detection and differentiation of pathogenic bacteria on solid growth media has been investigated in two studies using wavelengths in the visible region (400 to 1000 nm). In the first study Yoon *et al.*, (2010) distinguished *Campylobacter* species from non-*Campylobacter* contaminants with 97-99 % detection accuracy. In the second study, six species of non-0157 Shiga toxin-producing *E. coli*, grown on rainbow agar, were differentiated (Windham *et al.*, 2012). Classification accuracies of up to 100 % was obtained for some species. These studies make use of the vis-NIR region that, although information rich, likely distinguished bacteria based on colour, or colour changes in the media caused by the proliferating microorganisms. To date, no investigations using the NIR region (1000-2500 nm) to study bacteria on growth media have been reported.

This study aims to use wavelengths in the NIR region to differentiate between Gram positive and Gram negative bacteria, pathogenic and non-pathogenic bacteria and similar species of the

same genera. The bacteria will include, *Bacillus cereus, Escherichia coli, Salmonella enteritidis*, *Staphylococcus aureus* and *S. epidermidis*.

Materials and methods

Sample preparation

The following bacterial isolates from the culture collection of the Department of Food Science, Stellenbosch University were evaluated: *B. cereus* (ATCC 13061), *E. coli* (ATCC 25922), *S. enteritidis* (ATCC 13076), *S. aureus* (ATCC 25923) and *S. epidermidis* (ATCC 12228). The bacteria were selected as such to include a variety of characteristics, for example:

- Gram positive and Gram negative pathogens;
- Bacteria which appeared similar in colour on the growth media, as well as those which appeared different (to prove that colour does not affect the method);
- *S. epidermidis* was included to determine whether it was possible to distinguish between pathogens and non-pathogens.

Each bacterium was streaked out in duplicate onto solid Luria-Bertani (LB) agar in 100 mm glass Petri dishes and incubated for 20 h at 37 °C (**Fig 3.1**). Although both the spread and streak plate methods can be used for the isolation of pure cultures (Willey *et al.*, 2008), streak plates were preferred as they not only contain single colonies, but there are also large areas of bacterial growth where the initial streak was made. This permits the collection of more spectral data from the bacteria for analyses. LB agar was chosen because it is a common, general growth medium which contains only tryptone, yeast extract, sodium chloride and bacteriological agar powder (Bertani, 1951). To minimise the spectral response from the growth media, only LB agar was used throughout this study. All procedures were performed under aseptic conditions. After incubation the plates were allowed to cool at ambient temperature (approximately 23 °C) for 20 min before imaging.

Figure 3.1. Digital images (Huawei IDEOS S7, 3.15 MP, 2048x1536 pixels) of (from left), *B. cereus E. coli, S. enteritidis*, *S. aureus* and *S. epidermidis* on LB agar in glass Petri dishes (digital image orientation may differ to hyperspectral image).

NIR hyperspectral imaging system and image collection

The SisuChema SWIR (short wave infrared) hyperspectral pushbroom imaging system (Specim, Spectral Imaging Ltd., Oulu, Finland) with Chemadaq version 3.62.183.19 software was used to

acquire images. The system consisted of an imaging spectrograph coupled to a 2D array mercury–cadmium–telluride (HgCdTe) detector with a light source of quartz halogen lamps. A 31 mm lens with a field-of-view of 100 mm and a spatial resolution of 300 µm, at a working distance of 31 mm, was used. The spectral range was 920–2514 nm with a resolution of 10 nm and spectral sampling per pixel of 6.3 nm. The frame rate was 100 Hz at an exposure of 3.0 ms.

Prior to image acquisition, the imaging system was calibrated with white and dark references, where for the white reference a 100 % reflectance standard was used and for the dark reference, the shutter was closed. Thereafter images of the entire Petri dish, without removing the lid, were collected. The lid was not removed avoid to contamination of the agar plates with airborne and environmental microorganisms.

All Petri dishes were allowed to equilibrate to ambient temperature for 20 min before image acquisition. The use of streak plates allowed for many single colonies to be imaged, incorporating more variation within a plate due to differences between colonies. In total (duplicate plates included) 10 images were collected. To compensate for effects of the growth media, images of Petri dishes with only agar were also acquired.

Hyperspectral image analysis

The Evince v.2.7.0 (UmBio AB, Umeå, Sweden) hyperspectral image analysis software package was used to analyse the images. The image calibration and correction to absorbance was done automatically in the Evince software package according to the following equation:

$$I_{\lambda,n} = -log_{10}\left[\left(\frac{S_{\lambda,n} - B_{\lambda,n}}{W_{\lambda,n} - B_{\lambda,n}}\right)\right]$$

Where:

n = Pixel index variable ($n = 1 \dots N$) of the reorganised hypercube

$I_{\lambda,n}$ = Standardised absorbance intensity, pixel n, at wavelength λ

$S_{\lambda,n}$ = Sample image, pixel n, at wavelength λ

$B_{\lambda,n}$ = Dark reference image, pixel n, at wavelength λ

$W_{\lambda,n}$ = White reference image, pixel n, at wavelength λ

Data analysis

PCA was calculated for individual images with three principal components (PC's). To isolate only bacterial growth, images were cleaned, this included removal of bad pixels, background agar and any reflection from the Petri dish, using the brushing technique (Esbensen & Geladi, 1989) on score images and score plots interactively. Evince allowed for interactive image segmentation and does not permit thresholding at wavebands. Clusters were selected in the PCA scores plot and

corresponding components (Petri dish, agar and background) were highlighted in the scores image, removed then PCA was recalculated. This process was repeated iteratively until all irrelevant pixels were removed. Spectral noise regions were removed by excluding wavelengths from 920 to 1097 nm and 2477 to 2514 nm. These cleaned images were then merged to form mosaics. Three mosaics were created, each containing a different combination of bacteria. Group 1 included all bacteria which appeared white/cream in colour on the agar (*B. cereus, E. coli* and *S. enteritidis*). Group 2 included all Gram positive bacteria (*B. cereus, S. aureus* and *S. epidermidis*) and group 3 included the two *Staphylococcus* species (*S. aureus* and *S. epidermidis*). PCA was recalculated and standard normal variant (SNV) (Barnes *et al.*, 1989) and Savitzky Golay smoothing filter (Savitzky & Golay, 1964) (2^{nd} derivative, 3^{rd} order polynomial; 25 point smoothing) were applied. Score images, score plots and loading line plots of the mosaics as well as mean spectra of each bacterium were used to investigate differences, if any, between the bacteria. Mean spectra were obtained by computing the mean of each bacterium, then correcting it with Savitzky Golay smoothing filter (2^{nd} derivative, 3^{rd} order polynomial; 25 point smoothing) and SNV. A schematic of this process is shown in **Fig 3.2** where *S. aureus* and *S. epidermidis* were used as an example. After background removal and image correction, each image comprised at least 15000 pixels per bacteria. All calculations were done on the duplicate plates individually, yielding similar results, thus duplicate plates were deemed sufficient.

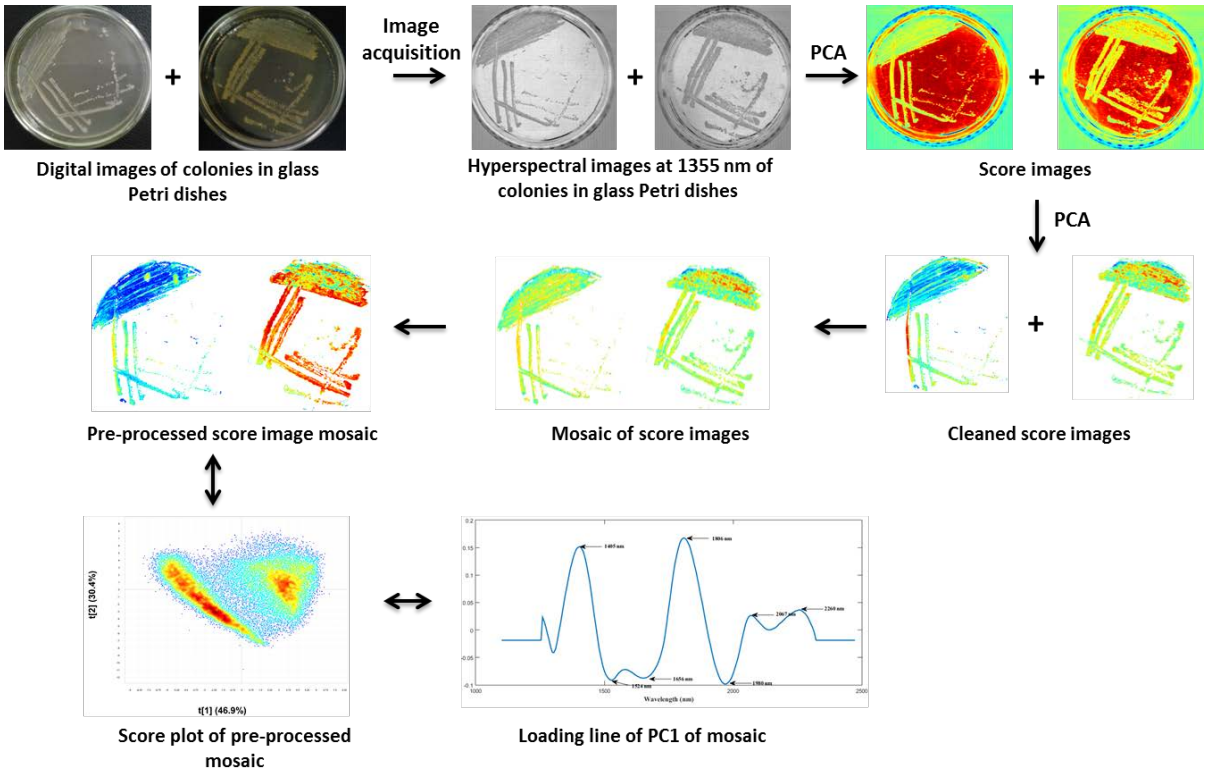

Figure 3.2. Schematic of image processing and data analysis sequence used in the study.

PLS-DA models were calculated for each group. The calibration models included the first set of 5 Petri dishes and the models were independently validated on the duplicate plates (test sets). SNV, Savitzky Golay smoothing filter (2nd derivatives, 3rd order polynomial; 25 point smoothing) were applied to all models and each organism was individually predicted. PLS-DA models use the coefficient of determination (R^2) as a diagnostic and was used to select the optimum number of components, i.e. the number of components which resulted in the highest R^2 value and prediction images with the best classification of samples. In this work pixels were considered as samples, thus a larger number of samples (>15000 pixels) per bacteria. Because of the large number of pixels in each image, it was decided to use the R^2, prediction image(s) and number of correctly classified pixels as an indication of the optimal number of factors and thus model complexity.

Results

Group 1 – bacteria which appear similar in colour on agar media

The score image of PC1 (58.1 % sum of squares (SS)) shows the bacteria (*B. cereus* (i), *E. coli* (ii) and *S. enteritidis* (iii)) according to the amplitude of pixels, where warm colours (yellow/red) are indicative of high score values whereas cold colours (blue/cyan) show lower score values (**Fig 3.3a**). Although there was variation in score values within the bacterial growth, *E. coli* and *S. enteritidis* (ii & iii) appeared to be similar while *B. cereus* was different. To better understand this, the score plot of PC1 vs. PC2 (7.75 % SS) **Fig 3.3b** was studied. The score plot displays the density of pixels of all three bacteria where the highest density of pixels are coloured red/yellow and lower density of pixels is coloured blue/cyan. There are two distinct clusters in the direction of PC1, however when each bacteria was selected in the score image and observed in the score plot, three clusters were evident. For improved visualisation of the differences, pixels were grouped and coloured (*B. cereus* is blue, *S. enteritidis* is yellow and *E. coli* is red) (**Fig 3.3c & d**). A clear separation of *B. cereus* from *S. enteritidis* and *E. coli* is now evident along PC1 (**Fig 3.3d**). This confirms that the chemical composition of *B. cereus* (blue) differs considerably from *E. coli* (red) and *Salmonella* (yellow). The negative region in the score plot along PC1 mainly contained pixels belonging to *B. cereus.* Thus, *B. cereus* was associated with the two negatively loaded peaks in the loading of PC1 (**Fig 3.4 a**). The first peak at 1480 nm (O-H stretch, first overtone (intramolecular H-bond), sugar) (Osborne *et al.*, 1993) was likely associated with a sugar and the other with protein, 2130 nm (N-H stretch + C=O stretch, amino acid). *S. enteritidis* and *E. coli* were associated with positively loaded peaks at 1693 nm (CH_3) and 1974 nm (N-H asymmetric stretch + amide II, protein) which may represent lipids (Dubois *et al.*, 2005b) and protein (Osborne *et al.*, 1993) respectively (**Fig 3.4 a**). A summary of the band assignments and chemical composition can be found in **Table 3.1**. In the direction of PC2 it was

possible to distinguish between *S. enteritidis* and *E. coli*. The majority of pixels representative of *E. coli* were found in the positive region along PC2, whereas pixels belonging to *S. enteritidis* were found in the negative region (**Fig 3.3d**). *E. coli* was associated with the positively loaded peak at 2310 nm (C-H stretch + C-H deformation) while *S. enteritidis* was associated with a negatively loaded peak at 2167 nm (2 x amide I + amide III, protein) (Osborne *et al.*, 1993) (**Fig 3.4 b**). To obtain a comprehensive description of the spectral variation within and between the bacteria, studying the mean spectra was essential.

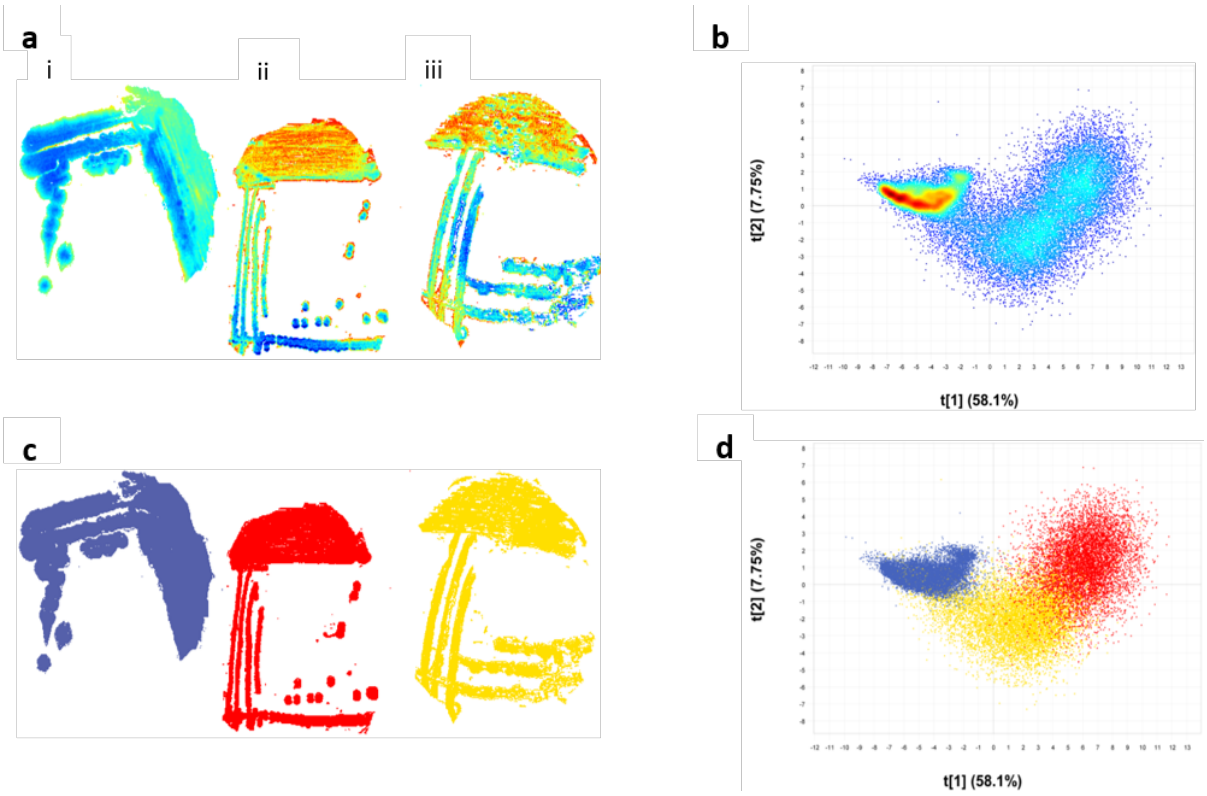

Fig. 3.3 a PC1 score image of (i) *B. cereus,* (ii) *E. coli* and (iii) *S. enteritidis*; **b** score plot of PC1 vs. PC2 for the mosaic of *B. cereus, E. coli* and *S. enteritidis*; **c** score image showing bacteria coloured according to class membership, *i.e. B. cereus* (blue)*, E. coli* (red) and *S. enteritidis* (yellow). **d** Projection of class membership onto score plot permitting easier visualisation of clusters.

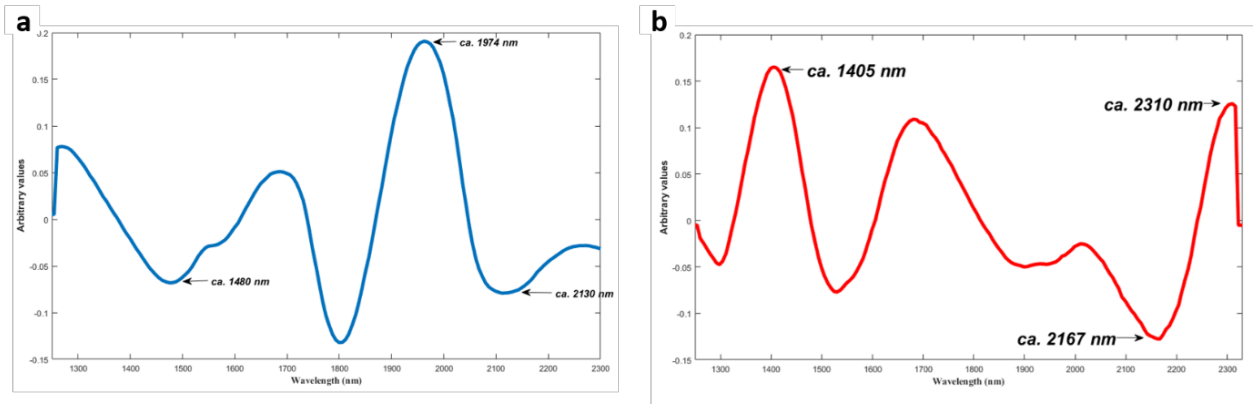

Fig 3.4a Loading line of PC1 indicating the variables (sugar, carbohydrate and protein) responsible for variation between *B. cereus* (negative loadings) and *E. coli* and *S. enteritidis* (positive loadings. **b** Loading line plot of PC2 showing the variables (carbohydrate and protein) responsible for variation between *E. coli* (positive loadings) and *S. enteritidis* (negative loadings).

In the mean spectra (**Fig 3.5**) *E. coli* (red) had more prominent peaks (most negative second derivative value) at 1424 nm (N-H stretch, first overtone, CONH) and 1781 nm (C-H stretch, first overtone, carbohydrate) (Osborne *et al.*, 1993) than *B. cereus* (blue) and *S. enteritidis* (yellow). At 1937 nm (C=O stretch, second overtone – CONH) *B. cereus* had the most prominent peak followed by *S. enteritidis* and *E. coli* with no peak.

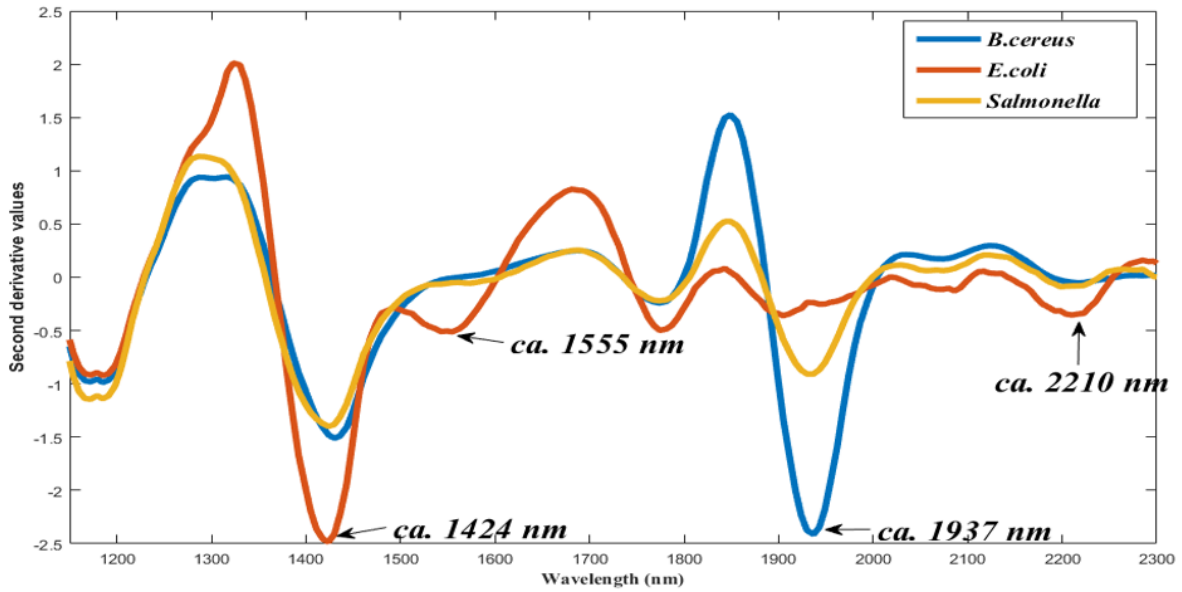

Fig 3.5 SNV and Savitzky-Golay (2nd derivative, 3rd polynomial order; 25 points) corrected mean spectra of group 1 (*B. cereus*, *E.coli* and *S. enteritidis*), with major peaks representing protein and carbohydrate

Table 3.1 Wavelengths of interest and assignments of bands for group 1 – group 3 for loadings and mean spectra, respectively

Wavelength (nm)	Band assignment	Structure	Bacteria association
Group 1 – Loading line of PC1			
1480	O-H stretch first overtone		*B. cereus*
2130	N-H stretch + C=O	Protein	*B. cereus*

	stretch		
1693	CH_3	Lipids	*E. coli & S. enteritidis*
1974	N-H asymmetric stretch + amide II	Protein	*E. coli & S. enteritidis*
Loading line of PC2			
2310	C-H stretch + C-H deformation	Lipids	*E. coli*
2167	2 x amide I + amide III	Protein	*S. enteritidis*
Mean spectra			
1424	N-H stretch, first overtone, CONH	Protein	
1781	C-H stretch, first overtone	Carbohydrate	
1937	C=O stretch, second overtone	Protein	
Group 2 – Loading line of PC1			
1405	O-H stretch, ROH	Teichoic acid	*S. epidermidis*
1980	N-H asymmetrical stretch + amide II	Protein	*B. cereus & S. aureus*
1824	O-H stretch + 2 x C-O stretch	Carbohydrate	
Group 3 – Loading line of PC 1			
1405	O-H stretch, ROH	Teichoic acid	*S. epidermidis & S. aureus*
1524	N-H stretch – intramolecular bond	Protein	*S. epidermidis & S. aureus*
1980	N-H asymmetric stretch + amide II	Protein	*S. epidermidis & S. aureus*
2067	N-H symmetrical stretch + amide II	Protein	*S. epidermidis & S. aureus*
Mean spectra			
1374	C-H stretch + C-H deformation		*S. aureus*
1436	O-H stretch, first overtone/C-H stretch	Carbohydrate/CH	*S. epidermidis*

1937	+ C-H deformation C=O stretch, second overtone	Protein

Group 2 – all Gram positive bacteria

The score image of PC1 (**Fig 3.6a**) shows distinct differences in score values in all three bacteria. The highest score values can be found within *S. epidermidis* (iii) whereas the lower score values are found in *B. cereus* (i) and *S. aureus* (ii). This suggests that *B. cereus* and *S. aureus* are more similar to each other than *S. epidermidis*. In the score plot of PC1 (37.5 % SS) vs. PC2 (33.8 % SS) two distinct clusters were observed (**Fig 3.6b**). Pixels of *S. epidermidis* (non-pathogenic) were predominantly located in the positive region in the direction of PC1, whereas those belonging to *B. cereus* and *S. aureus* (pathogenic) were overlapping in the negative part. This is visualised in **Fig 3.6d** and corresponds to the score image (**Fig 3.6c**). Three prominent peaks in the loadings of PC1 (**Fig 3.7**) (two positively loaded and one negatively loaded) accounted for the variation. Since pixels of *B. cereus* and *S. aureus* were located in the negative region of PC1 in the score plot, they were represented by negatively loaded peaks in the loading line plot. The peak at 1980 nm (N-H asymmetrical stretch + amide II) is characteristic of a protein (Osborne *et al.*, 1993). *S. epidermidis* was located in the positive region of the score plot and was therefore associated with positively loaded peaks at 1405 nm (O-H stretch, ROH) and 1824 nm (O-H stretch + 2 x C-O stretch, carbohydrate).

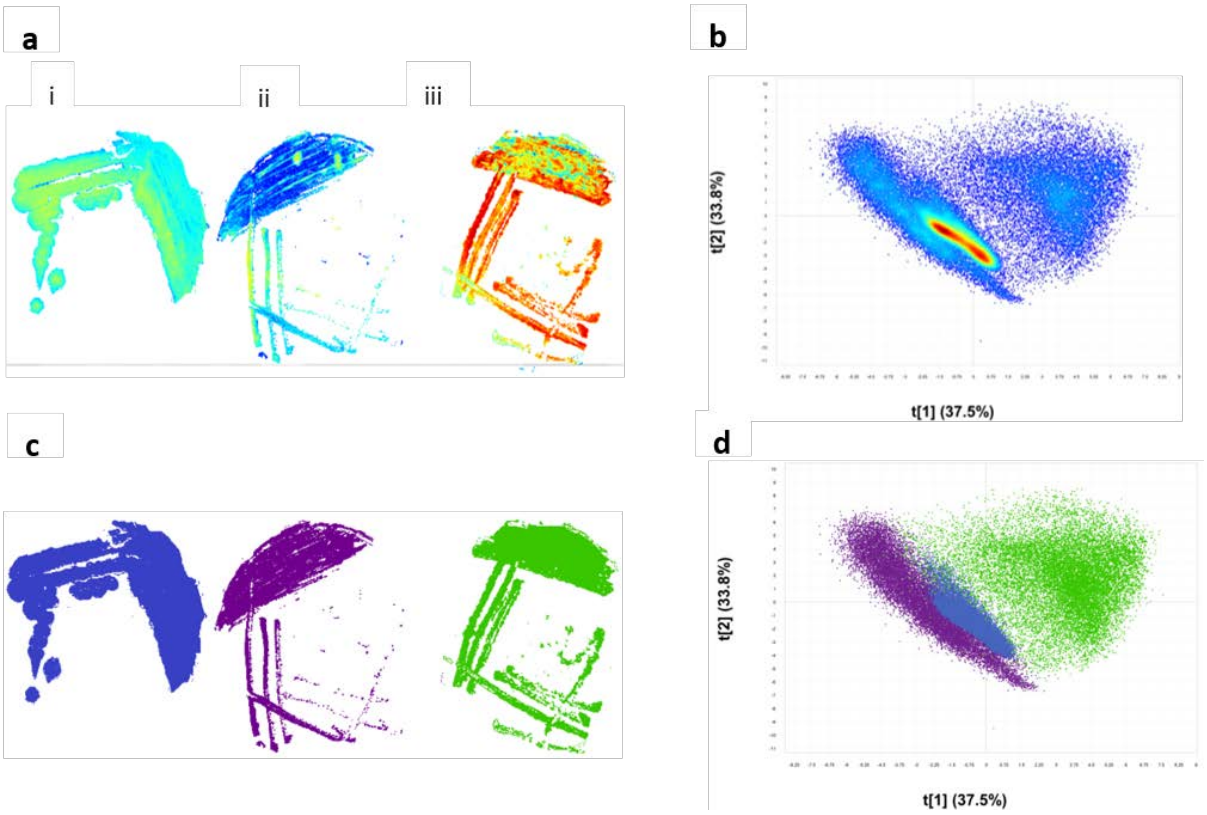

Fig 3.6a. PCA score image of *B. cereus* (i), *S. aureus* (ii) and *S. epidermidis* (iii). **b** score plot of PC1 vs PC2 for the mosaic of *B. cereus*, *S. aureus* and *S. epidermidis* **c** Score image showing bacteria coloured according to class membership **d** projection of class membership onto the score plot permitting easier visualisation of clusters.

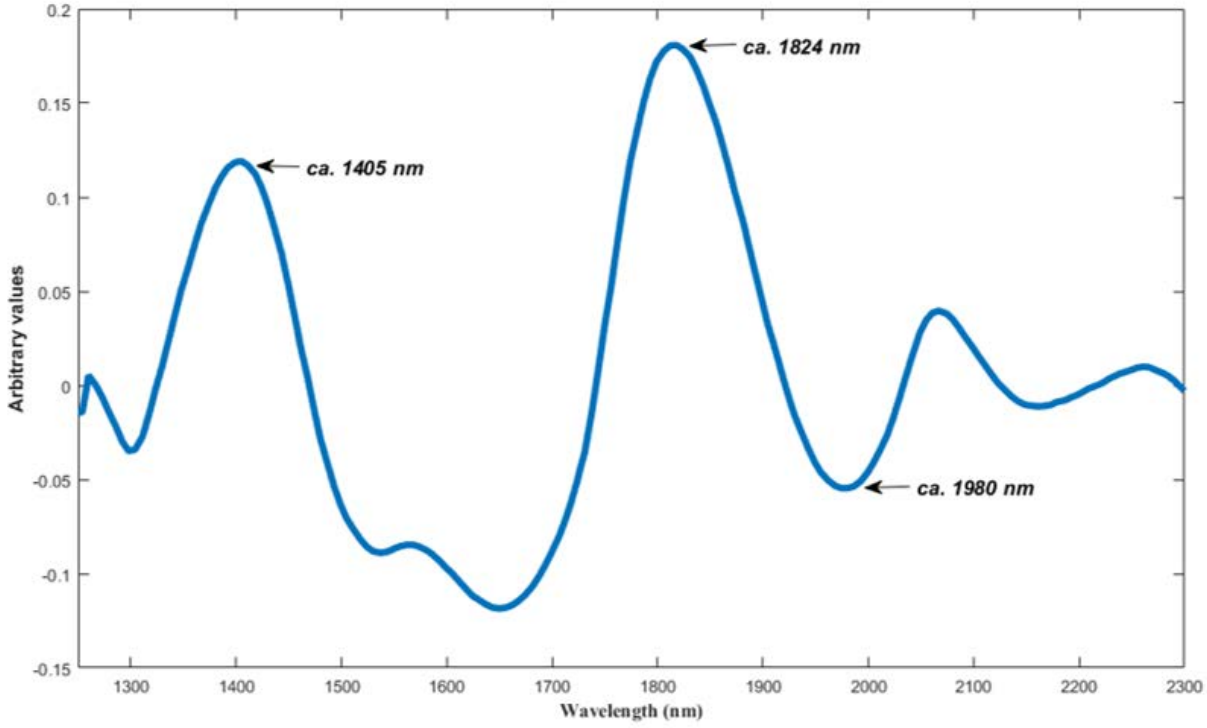

Fig 3.7 Loading line plot of PC1 showing the variables (teichoic acid, carbohydrates and protein) responsible for variation between *B. cereus* and *S. aureus* (negative loadings) and *S. epidermidis* (positive loadings).

The mean spectra (**Fig 3.8**) shows that *B. cereus* and *S. epidermidis* were fairly similar and only differ slightly at peaks 1436 nm (N-H stretch, first overtone – $CONH_2$) and 1937 nm (C=O, second overtone – CONH/O-H stretch + O-H deformation – H_2O). *S. aureus* differed considerably in absorption at these peaks.

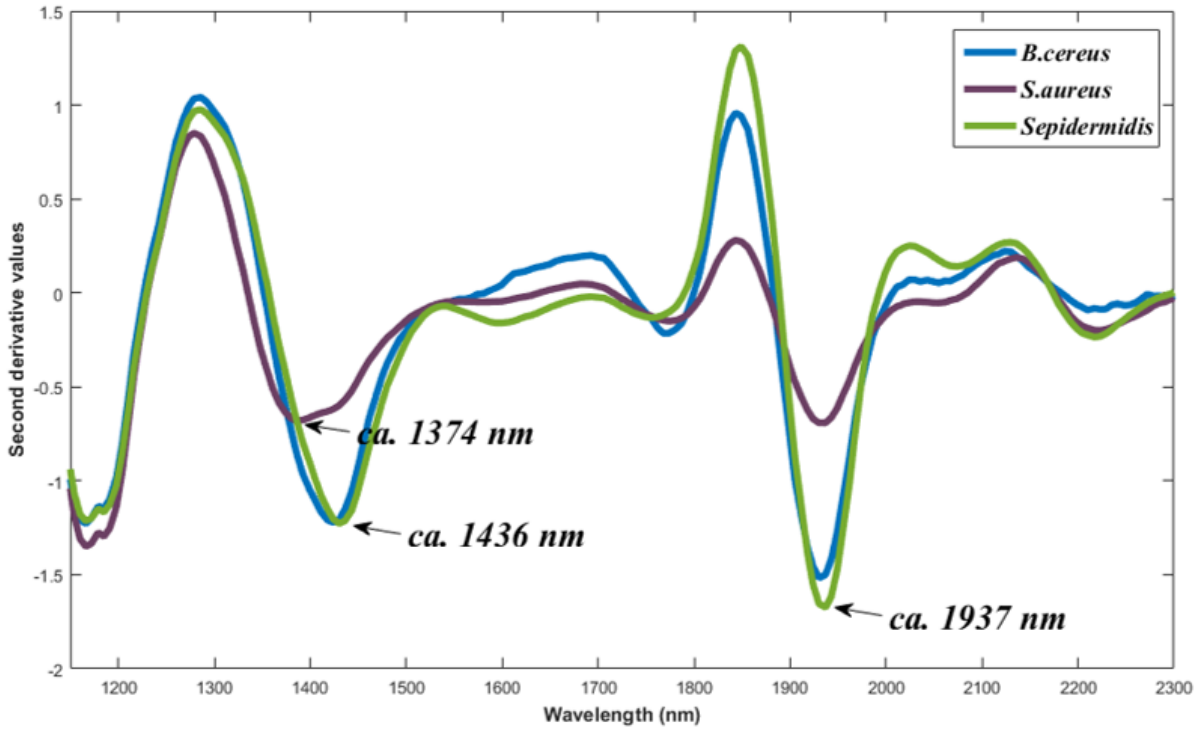

Fig 3.8 SNV and Savitzky Golay (2nd derivative, third polynomial; 25 points) corrected mean spectra of group 2 (*B. cereus, S. aureus* and *S. epidermidis)*, with major peaks representing proteins.

Group 3 – pathogenic and non-pathogenic species of the same genera

Species differentiation was performed by studying the spectra of *S. aureus* and *S. epidermidis*. The score image of PC1 (*S. aureus* (i) and *S. epidermidis* (ii)) shows a marked difference, in score values, between the two organisms (**Fig 3.9a.**) In the score plot (**Fig 3.9b**), two clusters were evident with a cumulative % SS for PC1 vs. PC2 of 77.8. PC1 and PC2 were required to distinguish between *S. aureus* and *S. epidermidis*. **Fig 3.9 c & d** displays the pixels belonging to *S. aureus* in purple and those of *S. epidermidis* in green. Separation of clusters can be explained by the loading line of PC1 (**Fig 3.10 a)** showing positively loaded peaks at 1405 nm (O-H stretch, first overtone from ROH) and 2067 nm (N-

H symmetrical stretch + amide II) (Osborne *et al.*, 1993). Negatively loaded peaks were found at 1524 nm (N-H stretch – intramolecular bond) and 1980 nm (N-H asymmetric stretch + amide II).

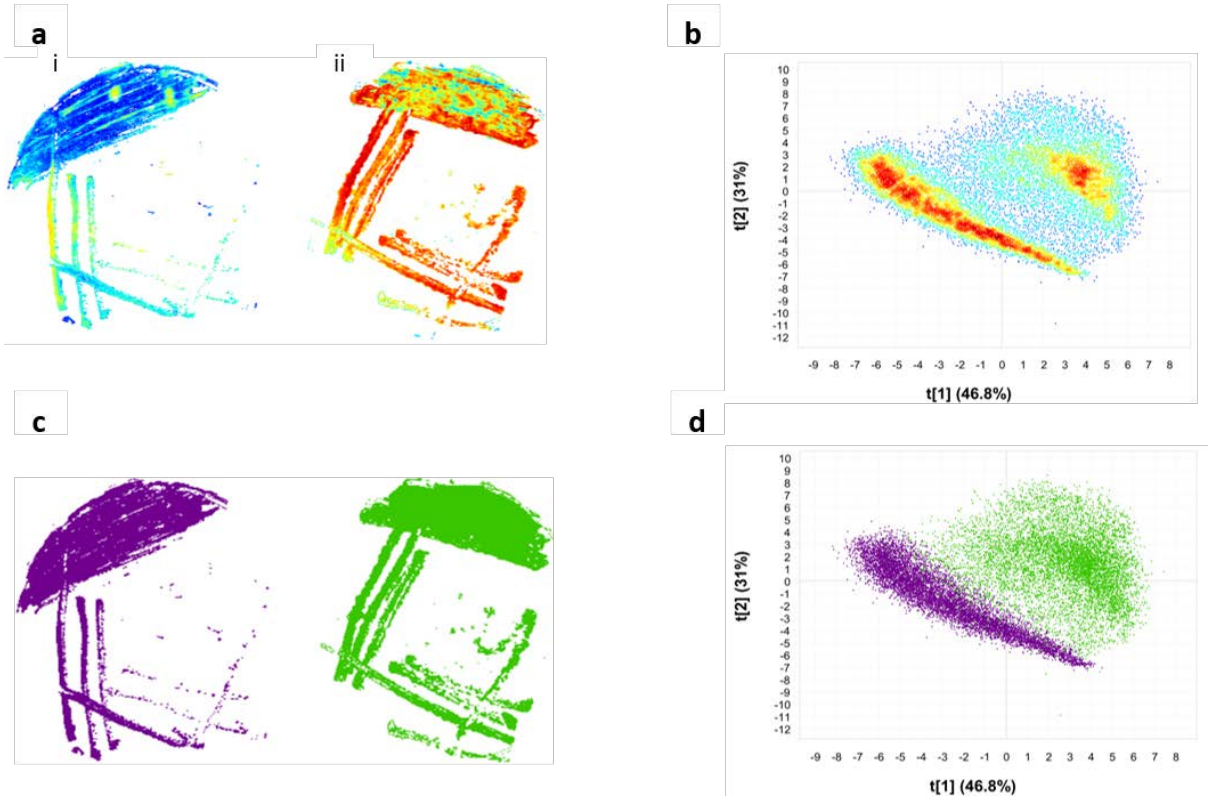

Fig 3.9a. PCA score image of *S. aureus* (i) and *S. epidermidis* (ii). **b** score plot of PC1 vs. PC2 for the mosaic of *S. aureus* and *S. epidermidis* **c** Score image showing bacteria coloured according to class membership **d** projection of class membership onto the score plot permitting easier visualisation of clusters

As previously mentioned, the mean spectra of these two species of *Staphylococcus* (**Fig 10 b**) differ at two spectral bands, 1436 and 1937 nm. These peaks belong to an O-H stretch, first overtone of ArOH and N-H stretch, first overtone of $CONH_2$ respectively (Osborne *et al.*, 1993).

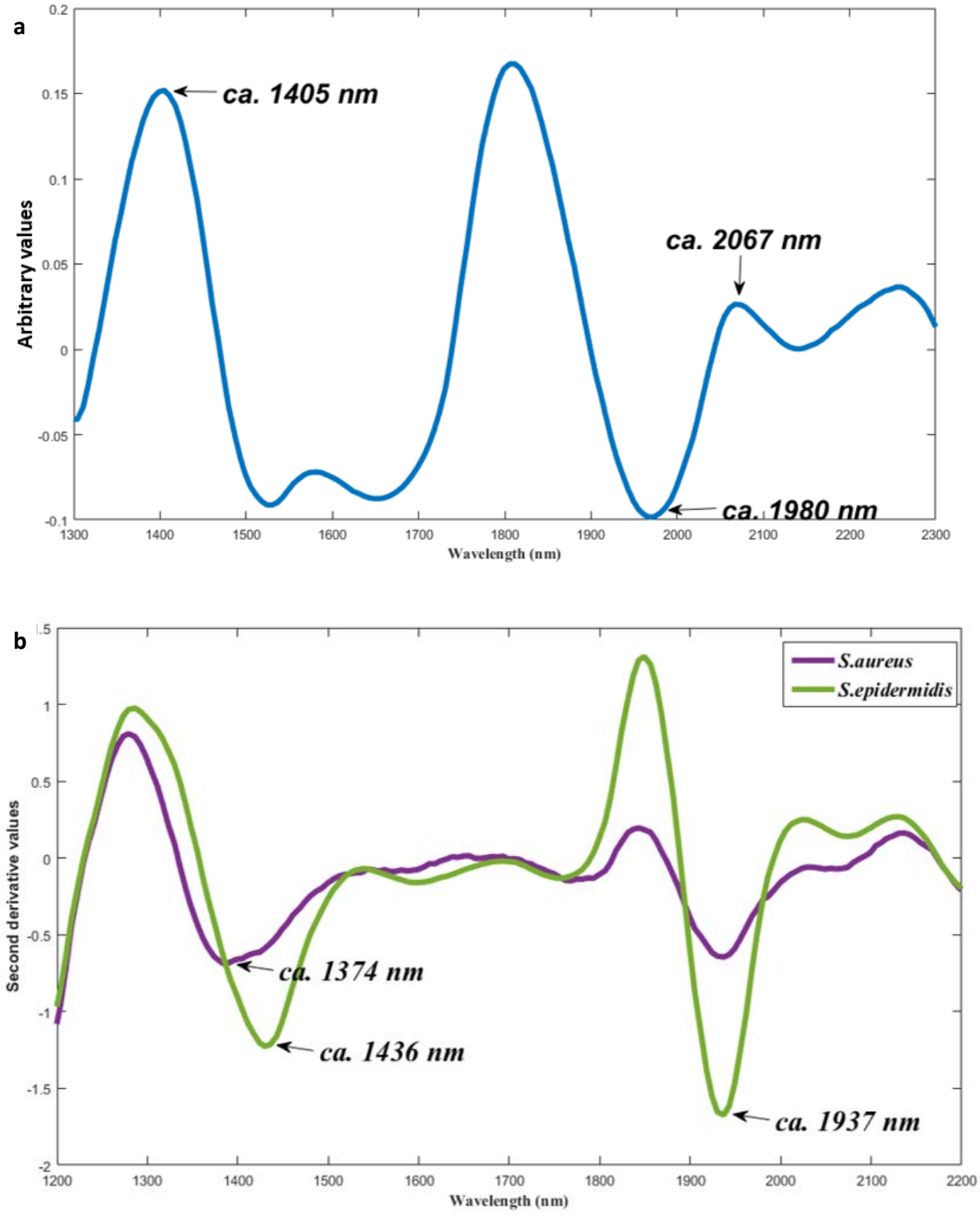

Fig 3.10 a Loading line plot of PC1 displaying the variables (teichoic acid, carbohydrates and protein) responsible for variation between *S. aureus* and *S. epidermidis.* **b** SNV and Savitzky Golay (2nd derivative, 3rd polynomial; 25 points) corrected mean spectra of group 3 (*S. aureus* and *S. epidermidis),* with major peaks representing proteins.

Table 3.2 shows the PLS-DA classification results where the model for groups 1 and 2 was calculated with four components, while group 3 was calculated with three. Coefficients of determination (R^2) for groups 1, 2 and 3 were 0.41, 0.74 and 0.89, respectively. For group 1, 78.2 % of *B. cereus* pixels and 53.9 % of *S. enteritidis* were correctly predicted, while only 2.34 % of pixels belonging to *E. coli* were predicted correctly. Group 2 had a higher percentage of correctly predicted pixels than group 1, where 93.0 % of *B. cereus,* 86.7 % of *S. aureus* and 82.0 % of *S. epidermidis* were correctly predicted. When predicting the *Staphylococcus* species, 99.96 % and 91.0 % of *S. aureus* and *S. epidermidis* pixels were predicted correctly.

Table 3.2 PLS-DA results of groups 1-3 showing classification results as percentage pixels correctly predicted

			Group 1		
Predicted as	*% B. cereus*	*% E. coli*	*% S. enteritidis*	*% S. aureus*	*% S. epidermidis*
B. cereus	**78.2**	26.3	16.6	-	-
E. coli	2.62	**2.34**	8.11	-	-
S. enteritidis	5.53	33.6	**53.9**	-	-
Not classified	13.6	37.7	21.3	-	-
			Group 2		
B. cereus	**93.0**	-	-	5.87	1.20
S. aureus	1.37	-	-	**86.7**	2.25
S. epidermidis	0.268	-	-	0.084	**82.0**
Not classified	5.78	-	-	7.35	14.6
			Group 3		
S. aureus	-	-	-	**99.96**	2.14
S. epidermidis	-	-	-	0.04	**91.0**
Not classified	-	-	-	0.00	6.87

Values in bold represent correctly predicted pixels of each bacteria whereas values which are not in bold represent misclassified or unclassified pixel percentages

DISCUSSION

The loading line plot of group 1 (**Fig 3.4a**) explains the variation between the two prominent clusters in the direction of PC1. Pixel distribution from negative to positive score values were as a result of differences in Gram positive and Gram negative cell walls. *B. cereus*, a Gram positive bacteria, has an exceptionally thick peptidoglycan layer (40-80 % by weight of the cell wall) – a polymer comprising of alternating residues of sugars; N-acetylglucoamine and N-acetylmuramic acid (Schleifer & Kandler, 1972). Furthermore, the cell wall also has proteins and teichoic acid embedded within it. Gram negative bacteria however (*E. coli* and *S. enteritidis*), have an outer membrane which consists mainly of lipopolysaccharides and proteins with an extremely thin peptidoglycan layer (10 % by weight of the cell wall) (Schleifer & Kandler, 1972; Davis & Mauer, 2010). The negatively loaded peak at 1480 nm most likely indicates a difference in the sugar (content or structure) of *B. cereus* compared to *E. coli* and *S. enteritidis*, and is possibly due to the increased level of peptidoglycan in *B. cereus*. The two peaks, characteristic of protein (2130 nm and 1974 nm) (Osborne *et al.*, 1993), indicate a distinct difference in the protein (content or structure) of the bacterial cell walls. Gram positive bacteria are known to have completely different protein structures to that of Gram negative bacteria (Navarre & Schneewind, 1999). They also contain teichoic and lipoteichoic acids, which consists of glycerol phosphate or ribitol phosphate joined together via phosphodiester bonds, not present in the cell walls of Gram negative bacteria (Navarre & Schneewind, 1999). In the loading plot of PC2, the peak which occurred at 1405 nm was likely due to the variation in the teichoic acid and/or lipoteichoic acid. Additional structures and compounds, such as sugars or D-alanine attached to teichoic acid, differ from one bacterial species to another (Navarre & Schneewind, 1999). This might account for the variation in teichoic acid, potentially represented by the 1405 nm peak.

Indeed, water was shown to be a significant influence in previous studies (Windham *et al.*, 2012; Yoon *et al.*, 2013). However, a number of other researchers utilised this region to successfully distinguish between foodborne bacteria (Dubois *et al.*, 2005b; Rodriguez-Saona *et al.*, 2001), differentiate fungal pathogens and study their growth on agar media (Williams *et al.*, 2012a; Williams *et al.*, 2012b). Furthermore, Dubois *et al.*, (2005b) indicted that the peak at 1940 nm (O-H stretch + O-H deformation) was useful to distinguish between Gram negative and Gram positive bacteria. It is well-known that the cell walls of Gram negative and Gram positive bacteria differ vastly (Mingeot-Leclercq & Decout, 2016) and that water is the dominant constituent of bacterial cells (Nakakimura *et al.*, 2012), hence the most important absorbance feature. It is thus possible that Gram negative and Gram positive bacteria have contrasting water contents, may interact with water differently and bind water in a different way. These all impact the absorption of water in the NIR

region and can be used to distinguish between bacteria, and quantify them (Nakakimura *et al.*, 2012; Slavchev *et al.*, 2015).

Differentiation between Gram negative bacteria is more complex, as they are quite similar in cell wall composition (Beveridge, 1999). The Gram negative cell wall consists of a lipid-rich outer and inner membrane, a thin layer of peptidoglycan, a gel-like matrix in the periplasmic space as well as inter-membrane proteins (Beveridge, 1999; Mingeot-Leclercq & Decout, 2016). Despite this, separation between *E. coli* and *S. enteritidis* was evident along PC2 (**Fig 3.3b**) and was explained by the corresponding loading line plot (**Fig 3.4b**) and the mean spectra (**Fig 3.5)**. The loading peak (2310 nm) indicates that *E. coli* is comprised of more lipids than *S. enteritidis*. This was confirmed with the mean spectra (**Fig 3.5**) where a larger absorption peak was observed at 2310 nm for *E. coli*. *S. enteritidis* contained more protein (2167 nm) compared to *E. coli.* This was established by studying the mean spectra, where the peak at 1937 nm (protein/water) was non-existent in the spectrum of *E. coli*.

The overlapping of *B. cereus* and *S. aureus* pixels (**Fig 3.6d**) indicated that the major chemical components in their cell walls were similar and differed considerably from that of *S. epidermidis*. In the loading line plot of PC1 (**Fig 3.7**), the peak at 1405 nm was once again present. This peak was possibly associated with differences in teichoic and/or lipoteichoic acid as well as species specific decorations in the cell walls of Gram positive bacteria (Schleifer & Kandler, 1972). The two peaks at 1649 nm and 1824 nm suggested that the carbohydrates in the cell walls of *S. epidermidis* differ from that of *S. aureus* and *B. cereus*. These differences could lie within the length and structure of the glycan units of peptidoglycan (Schleifer & Kandler, 1972) or the presence of other mono- or disaccharides embedded in the cell walls. Furthermore, pathogenic bacteria (*S. aureus* and *B. cereus*) require structures such as protein A and other cell wall-anchored surface proteins to adhere to components of the host tissues which are present, to a lesser extent, in non-pathogenic bacteria (Navarre & Schneewind, 1999). These proteins are likely represented by the negatively loaded peak at 1980 nm (N-H asymmetric stretch + amide II). Since the bulk of pixels representing *S. aureus* and *B. cereus* were located in the negative region of PC1, it is possible that the peak was indicative of their cell wall proteins.

The separation of pixels in the score plot of PC1 vs. PC2 (**Fig. 3.9b)** was likely due to differences in peptidoglycan and teichoic acid composition. Schleifer &Kandler, (1972), reported that the peptidoglycan in *S. aureus* contains large amounts of glycine, whereas in *S. epidermidis* it is comprised of serine. The two peaks related to protein, content or composition (1980 nm and 2067 nm) (Osborne *et al.*, 1993) (**Fig 3.10a**), partly explains the differences between the two staphylococci. Differences in teichoic acid (1405 nm) also contributed to the variation, and may be

attributed to structural dissimilarities, *i.e*. teichoic acid in *S. aureus* is comprised of ribitol units whereas that of *S. epidermidi*s consists of glycerol units. This confirms that the most distinct properties which separate the two *Staphylococcus* species are their protein and teichoic acid structures.

It was apparent that most pixels were predicted correctly, except *E. coli* and *S. enteritidis* where only 2.34 % of 18006 pixels and 53.9 % of 13351 pixels, respectively, were predicted correctly (**Table 3.2**). The low R^2 of 0.41 for group 1 was due to the overlapping pixels of *E.coli* and *S. enteritidis*. Cell wall components of Gram negative bacteria do not differ immensely from each other, resulting in similar spectral features, thus the model could not clearly distinguish between the two. This also resulted in high (33.6 %) misclassification of *E. coli* pixels as *S. enteritidis*. The good prediction results (82.0 – 99.96 % correctly predicted pixels) for groups 2 and 3 were attributed to the distinct differences in cell wall composition. Pixels which could not be identified (Not classified in **Table 3.2**) were observed mainly on the edges of colonies and areas where growth was not very pronounced i.e. a thin layer covering the agar. This could be due to interference from the agar (mixture of spectra). The lack of consistency in correctly predicted pixels, highlights the weakness in using pixel-wise classification for the purpose of bacterial colony differentiation. Object-wise classification is recommended where the entire bacterial growth is identified as one object (Kucheryavskiy, 2013).

In conclusion, NIR-HSI and MIA techniques was successfully used to separate Gram positive and Gram negative bacteria, differentiate between bacteria which appear similar in colour and distinguish between pathogenic and non-pathogenic bacteria. It was also possible to recognise variation among Gram positive bacteria and distinguish between similar species of the same genera on solid growth media.

Differences in teichoic acid, protein structures and lipid content were found to be the possible sources of variation in the analysed images. To test the reliability of the PLS-DA models for discrimination between bacteria, a model containing more replicates of the same bacteria should be used in a training set. Colonies of the same species can vary significantly, therefore more replicates would contribute variation, possibly making improving classification accuracy. Although chemical variances can provide a more comprehensive representation of bacterial content, this study did not include verification (with NIR spectroscopy) of the pure chemical components (teichoic acid) mentioned.

This research aimed to demonstrate that NIR hyperspectral imaging, in the spectral range of 900-2500 nm, can be used to study foodborne bacteria on growth media. The performance of some of the classification models were not encouraging, however other results indicated that this method

is promising for use as a tool in the food microbiology laboratory. The proposed method could considerably reduce the time spent on pathogen detection, leading to a reduction in contaminated product reaching the market and being consumed or recalled. Future work should include inoculation of bacteria on different growth media as well a time series study, where bacterial growth is followed using NIR hyperspectral imaging.

REFERENCES

Alexandrakis D, Downey G, Scannell AG (2008) Detection and identification of bacteria in an isolated system with near-infrared spectroscopy and multivariate analysis. J Agric Food Chem 56(10):3431-3437

Amigo JM, Martí I, Gowen A (2013) Hyperspectral imaging and chemometrics: A perfect combination for the analysis of food structure, composition and quality. In: Marini F (ed) Chemometrics in Food Chemistry. 1st edn. Elsevier, pp 343-370

Barnes R, Dhanoa M, Lister SJ (1989) Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl Spectrosc 43(5):772-777

Bertani G (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. J Bacteriol 62(3):293-300

Beveridge TJ (1999) Structures of gram-negative cell walls and their derived membrane vesicles. J Bacteriol 181(16):4725-4733

Burger J (2006) Hyperspectral NIR image analysis. Swedish University of Agricultural Sciences Umeå

Chevallier S, Bertrand D, Kohler A, Courcoux P (2006) Application of PLS-DA in multivariate image analysis. J Chemometrics 20(5):221-229

Cowe IA, McNicol JW (1985) The use of principal components in the analysis of near-infrared spectra. Appl Spectrosc 39(2):257-266 doi:Doi 10.1366/0003702854248944

Davis R, Mauer L (2010) Fourier transform infrared (FT-IR) spectroscopy: a rapid tool for detection and analysis of foodborne pathogenic bacteria. In: Mendez-Vilas A (ed) Current research, technology and education topics in applied microbiology and microbial biotechnology. 2nd edn. Formatex Research Center, pp 1582-1594

Dissing BS, Papadopoulou OS, Tassou C, Ersbøll BK, Carstensen JM, Panagou EZ, Nychas G-J (2012) Using multispectral imaging for spoilage detection of pork meat. Food Bioprocess Technol:1-12 doi:10.1007/s11947-012-0886-6

Duan C, Chen C, Khan MN, Liu Y, Zhang R, Lin H, Cao L (2014) Non-destructive determination of the total bacteria in flounder fillet by portable near infrared spectrometer. Food Control 42:18-22 doi:10.1016/j.foodcont.2014.01.023

Dubois J, Neil Lewis E, Fry Jr FS, Calvey EM (2005) Bacterial identification by near-infrared chemical imaging of food-specific cards. Food Microbiol 22(6):577-583

Esbensen K, Geladi P (1989) Strategy of multivariate image analysis (MIA). Chemometrics Intellig Lab Syst 7(1):67-86

Feng Y-Z, ElMasry G, Sun D-W, Scannell AGM, Walsh D, Morcy N (2013) Near-infrared hyperspectral imaging and partial least squares regression for rapid and reagentless determination of Enterobacteriaceae on chicken fillets. Food Chem 138(2–3):1829-1836 doi:http://dx.doi.org/10.1016/j.foodchem.2012.11.040

Geladi P, Grahn H, Manley M (2010) Data Analysis and Chemometrics for Hyperspectral Imaging Raman, Infrared, and Near-Infrared Chemical Imaging. John Wiley & Sons, Inc., pp 93-107

Geladi P, Isaksson H, Lindqvist L, Wold S, Esbensen K (1989) Principal component analysis of multivariate images. Chemometrics Intellig Lab Syst 5(3):209-220 doi:http://dx.doi.org/10.1016/0169-7439(89)80049-8

Gowen AA, Feng Y, Gaston E, Valdramidis V (2015) Recent applications of hyperspectral imaging in microbiology. Talanta 137:44-53 doi:10. 1016/j.talanta.2015.01.012

He H-J, Sun D-W (2015) Toward enhancement in prediction of Pseudomonas counts distribution in salmon fillets using NIR hyperspectral imaging. LWT - Food Science and Technology 62(1, Part 1):11-18 doi:http://dx.doi.org/10.1016/j.lwt.2015.01.036

Kucheryavskiy S (2013) A new approach for discrimination of objects on hyperspectral images. Chemometrics Intellig Lab Syst 120(0):126-135 doi:http://dx.doi.org/10.1016/j.chemolab.2012.11.009

Lazcka O, Del Campo FJ, Munoz FX (2007) Pathogen detection: A perspective of traditional methods and biosensors. Biosensors Bioelectron 22(7):1205-1217

Liu Y, Chen Y-R, Kim MS, Chan DE, Lefcourt AM (2007) Development of simple algorithms for the detection of fecal contaminants on apples from visible/near infrared hyperspectral reflectance imaging. J Food Eng 81(2):412-418

Maity JP, Kar S, Lin C-M, Chen C-Y, Chang Y-F, Jean J-S, Kulp TR (2013) Identification and discrimination of bacteria using Fourier transform infrared spectroscopy. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 116:478-484 doi:http://dx.doi.org/10.1016/j.saa.2013.07.062

Mandal P, Biswas A, Choi K, Pal U (2011) Methods for rapid detection of foodborne pathogens: an overview. Am J Food Technol 6(2):87-102

Manley M (2014) Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. Chem Soc Rev 43(24):8200-8214

Mariey L, Signolle J, Amiel C, Travert J (2001) Discrimination, classification, identification of microorganisms using FTIR spectroscopy and chemometrics. Vib Spectrosc 26(2):151-159

Mingeot-Leclercq MP, Decout JL (2016) Bacterial lipid membranes as promising targets to fight antimicrobial resistance, molecular foundations and illustration through the renewal of aminoglycoside antibiotics and emergence of amphiphilic aminoglycosides. Medchemcomm 7(4):586-611 doi:10.1039/c5md00503e

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(5):1389-1394

Navarre WW, Schneewind O (1999) Surface Proteins of Gram-Positive Bacteria and Mechanisms of Their Targeting to the Cell Wall Envelope. Microbiol Mol Biol Rev 63(1):174-229

Norris KP (1959) Infra-red spectroscopy and its application to microbiology. The Journal of Hygiene 57(3):326-345

Osborne BG, Fearn T, Hindle PH (1993) Practical NIR Spectroscopy with Applications in Food and Beverage Analysis, 2nd edn. Longman Scientific & Technical, Essex, England

Park B, Lawrence K, Windham W, Smith D (2004) Multispectral imaging system for fecal and ingesta detection on poultry carcasses. J Food Process Eng 27(5):311-327 doi:10.1111/j.1745-4530.2004.00464.x

Rodriguez-Saona LE, Khambaty FM, Fry FS, Calvey EM (2001) Rapid Detection and Identification of Bacterial Strains By Fourier Transform Near-Infrared Spectroscopy. J Agric Food Chem 49(2):574-579 doi:10.1021/jf000776j

Rodriguez-Saona LE, Khambaty FM, Fry FS, Dubois J, Calvey EM (2004) Detection and identification of bacteria in a juice matrix with Fourier transform-near infrared spectroscopy and multivariate analysis. J Food Prot 67(11):2555-2559

Savitzky A, Golay MJ (1964) Smoothing and differentiation of data by simplified least squares procedures. Anal Chem 36(8):1627-1639

Schleifer KH, Kandler O (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriol rev 36(2):407-477

Siripatrawan U, Makino Y, Kawagoe Y, Oshita S (2011) Rapid detection of *Escherichia coli* contamination in packaged fresh spinach using hyperspectral imaging. Talanta 85(1):276-281

Slavchev A, Kovacs Z, Koshiba H, Nagai A, Bázár G, Krastanov A, Kubota Y, Tsenkova R (2015) Monitoring of Water Spectral Pattern Reveals Differences in Probiotics Growth When Used for Rapid Bacteria Selection. PLoS ONE 10(7):e0130698 doi:10.1371/journal.pone.0130698

Willey JM, Sherwood LM, Prescott LM (2008) Prescott, Harley, and Klein's Microbiology-7th international edition New York: McGraw-Hill Higher Education

Williams P, Geladi P, Britz T, Manley M (2012a) Growth characteristics of three *Fusarium* species evaluated by near-infrared hyperspectral imaging and multivariate image analysis. Appl Microbiol Biotechnol 96(3):803-13 doi:10.1007/s00253-012-4380-x

Williams P, Geladi P, Britz T, Manley M (2012b) Near-infrared (NIR) hyperspectral imaging and multivariate image analysis to study growth characteristics and differences between species and strains of members of the genus *Fusarium*. Anal Bioanal Chem 404(6):1759-1769 doi:10.1007/s00216-012-6313-z

Williams PJ, Geladi P, Britz TJ, Manley M (2012c) Growth characteristics of three *Fusarium* species evaluated by near-infrared hyperspectral imaging and multivariate image analysis. Appl Microbiol Biotechnol 96(3):803-813

Williams PJ, Geladi P, Britz TJ, Manley M (2012d) Investigation of fungal development in maize kernels using NIR hyperspectral imaging and multivariate data analysis. J Cereal Sci 55(3):272-278 doi:http://dx.doi.org/10.1016/j.jcs.2011.12.003

Williams PJ, Geladi P, Britz TJ, Manley M (2012e) Near-infrared (NIR) hyperspectral imaging and multivariate image analysis to study growth characteristics and differences between species and strains of members of the genus *Fusarium*. Anal Bioanal Chem 404(6-7):1759-1769

Windham W, Yoon S-C, Ladely S, Heitschmidt J, Lawrence K, Park B, Narrang N, Cray W (2012) The effect of regions of interest and spectral pre-processing on the detection of non-0157 Shiga-toxin producing *Escherichia coli* serogroups on agar media by hyperspectral imaging. J Near Infrared Spectrosc 20(5):547-558

Yoon S-C, Windham W, Ladely S, Heitschmidt J, Lawrence K, Park B, Nareng N, Cray W (2013) Hyperspectral imaging for differentiating colonies of non-O157 Shiga-toxin producing *Escherichia coli* (STEC) serogroups on spread plates of pure cultures. J Near Infrared Spectrosc 21(2):81-95

Yoon SC, Lawrence KC, Line JE, Siragusa GR, Feldner PW, Park B, Windham WR (2010) Detection of *Campylobacter* colonies using hyperspectral imaging. Sens Instrum Food Qual Saf 4(1):35-49 doi:10.1007/s11694-010-9094-0

CHAPTER 4

Effect of foodborne bacteria colony age on near-infrared (NIR) hyperspectral image analysis

Abstract

Near-infrared hyperspectral imaging (NIR-HSI) and multivariate image analysis (MIA) was used to distinguish between foodborne pathogenic bacteria, *Bacillus cereus, Escherichia coli, Salmonella enteritidis, Staphylococcus aureus* and a non-pathogenic bacterium, *Staphylococcus epidermidis*. Hyperspectral images of streaked out (on Luria Bertani agar) bacteria were acquired after 20, 40 and 60 h growth (37 °C) using a SisuCHEMA hyperspectral pushbroom imaging system with a spectral range of 920-2514 nm. Principal component analysis (PCA) was calculated before three different preprocessing methods; standard normal variate (SNV), Savitzky-Golay (1st derivative, 2nd order polynomial, 15 point smoothing) and Savitzky-Golay (2nd derivative, 3rd order polynomial, 15 point smoothing) were applied separately to mean centred data. SNV provided more distinct clustering in the PCA score plots and was therefore used as the only preprocessing method. Partial least squares discriminant analysis (PLS-DA) models were built for each growth period and tested on a second set of plates to evaluate the effect the age of the colony has on classification accuracies. The highest overall prediction accuracies, where test plates required the least amount of growth time, was found with models built after 60 h growth and tested after 20 h growth. Predictions for bacteria differentiation within these models ranged from 83.1 to 98.8 % correctly predicted pixels.

Introduction

Recently near-infrared (NIR) hyperspectral imaging (HSI) has been explored as a tool to differentiate pathogenic bacteria from each other as well as from non-pathogenic bacteria (Dubois *et al.*, 2005; Yoon *et al.*, 2010; Kammies *et al.*, 2016; Yoon *et al.*, 2013). NIR-HSI is a non-destructive technique which combines imaging and spectroscopy. It provides both spectral and spatial information within a three dimensional data structure known as a hypercube (two spatial and one spectral dimension) (Burger & Geladi, 2006) and is advantageous when analysing heterogeneous samples (Manley, 2014). The NIR region spans the wavelength range 800-2500 nm and spectra contain information about major X-H bonds in chemical compounds. Peaks observed are due to overtone vibrations of hydrogen bonding to carbon, oxygen and nitrogen. This results in a large number of organic molecules which are measurable with NIR analysis (Blanco & Villarroya, 2002).

NIR imaging instruments provide vast amounts of data rapidly, for example a hypercube from a SisuCHEMA (Specim, Spectral Imaging Ltd, Oulu, Finland) imaging system can contain close to

100 000 spectra with 242 wavelength channels (Williams *et al.*, 2012; Burger & Geladi, 2006). For this amount of data, chemometrics, a well-known discipline that allows extraction of data in a multivariate manner is required (Amigo *et al.*, 2015). The fundamental technique in the application of chemometric tools in NIR-HSI is principal component analysis (PCA) (Geladi *et al.*, 1989; Wold *et al.*, 1987). This is the most common approach used to reduce complexity in a hypercube.

Physical and chemical phenomena such as molecular interactions, scattering of particles, changes in refractive index at high concentrations, changes in sample path length and stray light can cause deviations from linear relationships (Rinnan *et al.*, 2009a). With preprocessing, these effects can be largely eliminated to improve exploratory information, multivariate regression or classification models (Burger & Geladi, 2006; Rinnan *et al.*, 2009a). A large, diverse, number of preprocessing methods is available for NIR data as the spectra can be significantly affected by nonlinearities introduced by light scatter. The most widely used preprocessing techniques in NIR spectral data analysis are divided into two categories, namely: spectral derivatives and scatter correction methods. The two main methods used for scatter correction are standard normal variate (SNV) (Barnes *et al.*, 1989) and multiplicative scatter correction (MSC) (Martens *et al.*, 1983; Geladi *et al.*, 1985). These are often used interchangeably with NIR spectra, as the equations for the transformations take the same form and therefore produce the same results in most practical applications (Rinnan *et al.*, 2009b; Fearn *et al.*, 2009). MSC calculation consists of two steps: estimation and correction of coefficients,

$$x_{org} = b_0 + b_{ref,1} \cdot x_{ref} + e \qquad \text{...[eq. 1]}$$

and correction of spectra,

$$x_{corr} = \frac{x_{org} - b_0}{b_{ref,1}} \qquad \text{...[eq. 2]}$$

Where:

x_{org} = original sample spectrum

x_{ref} = reference spectrum

x_{corr} = corrected spectrum

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

The basic equation for SNV has the same form as the second step of the MSC calculation (eq. 2):

$$x_{corr} = \frac{x_{org} - a_0}{a_1} \qquad \text{...[eq. 3]}$$

Where:

x_{corr} = corrected spectrum

a_0 = average value of spectrum

a_1 = standard deviation of spectrum

Spectral derivatives are represented by Norris-Williams derivatives (Norris, 1959) and Savitzky-Golay polynomial derivative filters (Savitzky & Golay, 1964). Smoothing of spectra, which reduces narrow spikes by attempting to remove random noise in data, is used in both methods before application of derivatives (Varmuza & Filzmoser, 2016; Rinnan *et al.*, 2009b).

Following preprocessing and exploration of data (e.g. with PCA), regression or classification techniques are used to develop models for qualitative analysis. Partial least squares discriminant analysis (PLS-DA) is a supervised discriminative technique used to classify data using dummy variables for class membership (Chevallier *et al.*, 2006; Manley *et al.*, 2008). Gowen *et al.*, (2010) used this technique to discriminate between undamaged (U), and damaged (Dam, mechanically damaged (MD) and *Pseudomonas tolaasii* infected (PT)) mushrooms as well as between MD and PT mushrooms. For test sets, PLS-DA analysis showed the sensitivity of the U/Dam model as 0.832 but none of the U mushrooms were misclassified as Dam (specificity = 1.000). For the MD/PT model almost none of the MD mushrooms were classified as PT (specificity = 0.984), but this model had the lowest sensitivity (0.661). In another study, PLS and linear discriminant analysis (LDA) was used to access total viable count (TVC) and psychotropic plate count (PPC) on porcine meat over 21 days (Barbin *et al.*, 2013). Chemical changes in the meat, due to bacterial growth were used to predict whether samples were commercially sound (those which had a bacterial number lower than log 6 CFU/g). LDA models showed over 95 % accuracy for spoilage detection and PLS produced coefficients of determination of 0.82 and 0.85 for TVC and PPC respectively.

Bacteria multiply by division of single cells. When an old or damaged culture is introduced into fresh media, the culture will undergo the following growth phases in a closed system (Srivastava & Srivastava, 2013; Buchanan, 1974):

1. Lag phase – during this phase, initially, stationary growth is observed where damaged cells are repaired and synthesis of certain enzymes required for growth is resumed. Followed by a slow increase in cell population.
2. Exponential growth phase – once bacteria have acclimatised to the environment, cells multiply logarithmically. During this phase cell division is at a maximum and generation time is at a minimum for fast growing bacteria.

3. Stationary growth phase – the rate of growth of the bacteria decreases and generation time increases until there is practically no increase in numbers.
4. Death phase – numbers of bacterial cells start to decline slowly as nutrients decrease and then with increasing rapidity, until a logarithmic decrease is seen when nutrients become depleted.

Various genera of bacteria have different generation times and will therefore not display the growth phases at the same time (Buchanan, 1974; Srivastava & Srivastava, 2013). The approximate generation time for bacteria used in the current study was as follows: *E. coli* and *S. enteritidis* – 20 min (Fehlhaber & Krüger, 1998), *S. aureus* and *B. cereus* – 30 min (Keren *et al.*, 2004; Collins & Richmond, 1962) and *S. epidermidis* – 45 min (Gottenbos *et al.*, 2001).

A study was conducted by Yoon *et al.*, (2009), where NIR hyperspectral imaging was used to facilitate the detection and identification of *Campylobacter* and non-*Campylobacter* species, using wavelengths 400 – 900 nm. The colonies were incubated for 48 h and a two-step pattern classification algorithm was developed to differentiate *Campylobacter* colonies from non-*Campylobacter* colonies (*Sphingomonas paucimobilis*, *Acinetobacter baumannii*, *Brevundimonas diminuta*, *Ochrobacterium* sp. and *Flavobacterium odoratum*). Three different agars (Campy-Cefex, Campy-Line and blood agar) were used as growth media in this study. Cefex cultures had the highest classification accuracies (up to 99%). In an attempt to increase the rapidity of this detection and identification method, a follow-up study was done (Yoon *et al.*, 2010). The same organisms were incubated for a shorter period (24 h) before being imaged, and a comparison was done between the results of the two growth times. Three different classification methods (band ratio algorithm, spectral feature fitting and single band thresholding algorithm) were applied. With the single band thresholding algorithm, it was found that the overall classification accuracy of the 48 h Cefex cultures (96.81 %) was higher than the overall 24 h Cefex cultures (85.95 %). Blood agar cultures, incubated for 24 h, showed up to 99 % accuracy with the band ratio algorithm. These results demonstrated that it was possible to detect Campylobacter using NIR HSI after 24 h of growth.

The present study aimed to use different preprocessing methods to find the best clustering and score value differences between bacteria in score plots and score images, respectively. Thereafter, bacteria were classified over a period of three days to evaluate the effect colony age has on classification results of foodborne bacteria.

Materials and methods

Sample preparation

Bacterial isolates of *B. cereus* (ATCC 13061), *E. coli* (ATCC 25922), *S. enteritidis* (ATCC 13076), *S. aureus* (ATCC 25923) and *S. epidermidis* (ATCC 12228) were acquired from the culture collection of the Department of Food Science, Stellenbosch University. Two sets of streak plates were prepared for each bacterium using Luria-Bertani (LB) agar in 100 mm glass Petri dishes and incubated for 60 h at 37 °C. Plates were removed from the incubator every 20 h, digitally imaged, and then imaged with the hyperspectral system. All inoculations were performed under aseptic conditions. To reduce the temperature to approximately ambient, plates were left at 21 °C for 20 min before imaging.

NIR hyperspectral imaging system and image acquisition

A SisuCHEMA short wave infrared camera (Specim, Spectral Imaging Ltd, Oulu, Finland) was used to acquire NIR hyperspectral images of each bacterium after each 20 h growth period. The sensor consists of a 2D array mercury-cadmium-telluride (HgCdTe) detector, with a light source of quartz halogen lamps as described on pg 32. Once the imaging system was calibrated (as described on pg 33), images of the entire Petri dish, with the lid on, were collected every 20 h over a period of 60 h. In total, 30 images were collected (5 different bacteria, imaged thrice, including test plates).

Data analysis

Individual images were imported into Evince v.2.7.0 (UmBio AB, Umeå, Sweden) hyperspectral image analysis software package. PCA was calculated with three PC's before each image was cleaned of background, agar, bad pixels and reflection from the Petri dish using the brushing technique (Esbensen & Geladi, 1989) on score images and score plots interactively. Wavelengths from 920 to 1097 nm and 2477 to 2514 nm were removed as these areas were noisy. Mosaics were then constructed with these cleaned individual images before PCA was recalculated. Bacteria were divided into groups showed in **Table 4.1**. Group 1 consisted of Gram positive and Gram negative bacteria which appeared similar in colour (white) on the agar. Group 2 consisted of only Gram positive bacteria, and group 3 contained two species of the same genera of bacteria (a pathogenic and non-pathogenic species). Separate mosaics were then constructed for each group. One was constructed for each of the 20, 40 and 60 h growth periods. A total of nine mosaics were constructed.

Table 4.1 Groupings of bacteria used in mosaics.

Group 1	Group 2	Group 3
B. cereus	*B. cereus*	*S. aureus*
E. coli	*S. aureus*	*S. epidermidis*
S. enteritidis	*S. epidermidis*	

To acquire more defined clusters in PCA score plots, and thereby improve subsequent classification results, three preprocessing treatments were tested. Score plots and score images were inspected for clustering and colour differences. In score images, warm colours (yellow/red) are indicative of high score values whereas cold colours (blue/cyan) show lower score values. A good result was achieved with distinct clustering of pixels of each bacterium, with little overlap and varying colour in score images. As previously mentioned, SNV and MSC are used interchangeably, therefore preprocessing methods which were tested included SNV and Savitzky-Golay first derivative (2^{nd}-order polynomial; 15 point smoothing) and second derivative (3^{rd}-order polynomial; 15 point smoothing). Baseline correction was excluded as this was done in combination with smoothing when Savitzky-Golay polynomial derivatives are calculated. The data was then transferred to MATLAB v.9.1 for further processing and to calculate PLS-DA models for each group, at each time interval. These models were validated with test sets of mosaics made from the duplicate plates from each time interval. The data analysis process is shown in **Fig. 4.1**. For PLS-DA models, cross-validation (venetian blinds with 10 splits and 20 samples per split) was used to select the optimum number of latent variables (LV's) for each model.

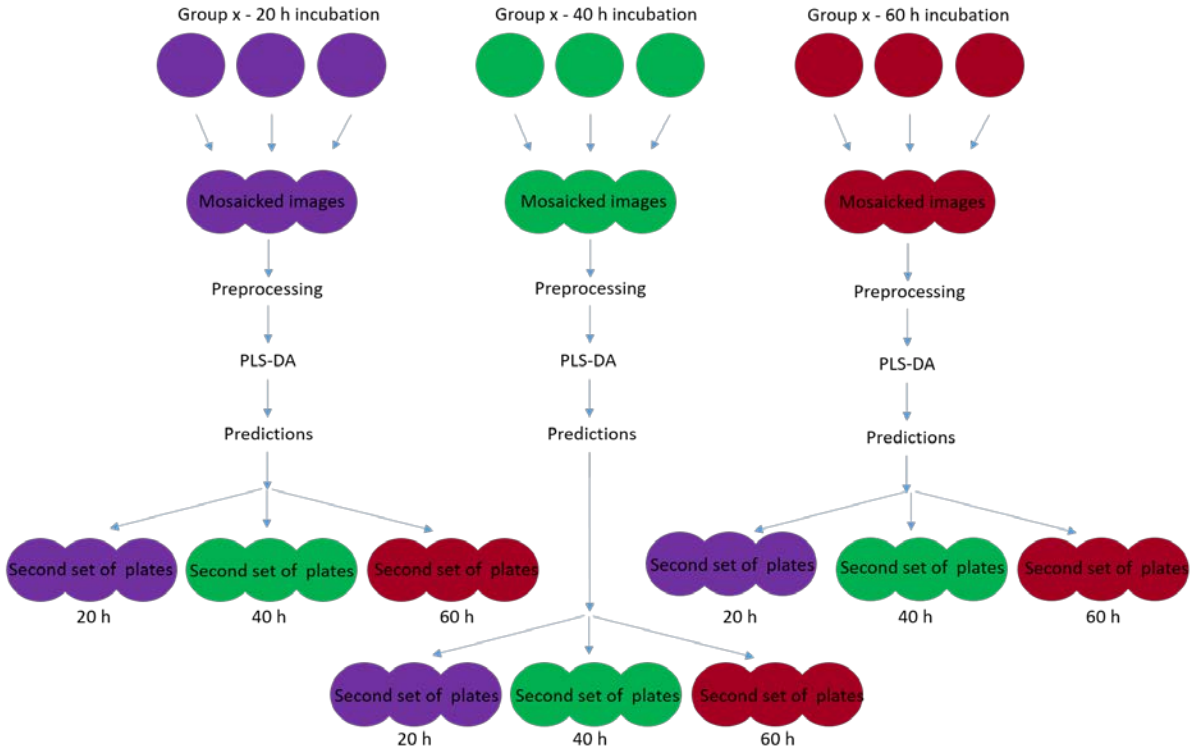

Figure 4.1. Schematic of data analysis process used in this study. PLS-DA models built for each time increment (20, 40 and 60 h growth), tested on a second set of plates of bacteria imaged after 20, 40 and 60 h growth.

Results and Discussion

Digital images of each Petri dish were collected after each growth period, before hyperspectral image acquisition using a Huawei IDEOS S7, 3.15MP, 2048 × 1536 pixels. An example of the digital images is shown in **Fig 4.2**

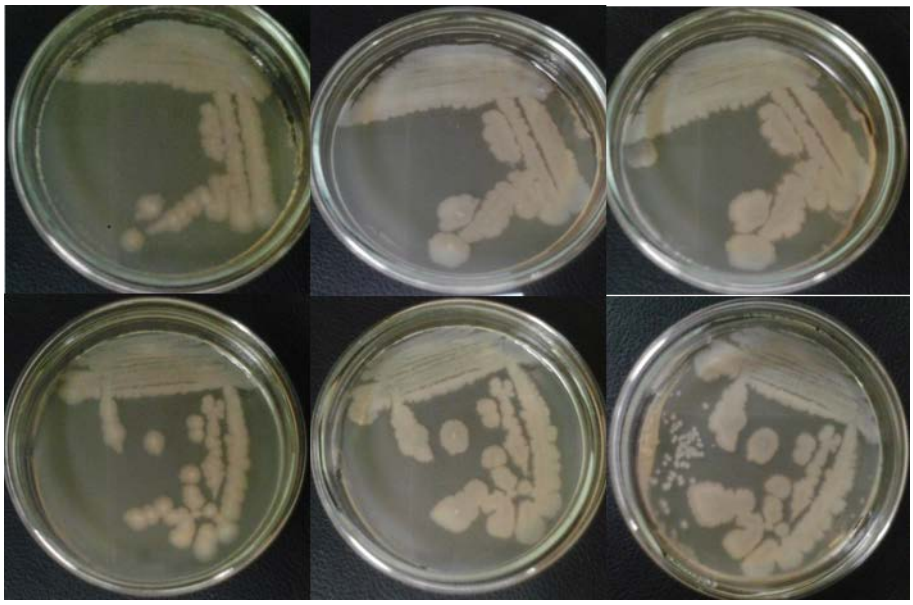

Figure 4.2. Digital images of *B. cereus* after 20, 40 and 60 h (left to right) growth at 37 °C with the second set of plates in the bottom row.

Preprocessing

From the exploratory PCA plots, it was found that using SNV alone produced the best clustering in score plots for all groups over the three days. Bacteria were imaged in glass Petri dishes with the lid on, when images where acquired, irregularities in the glass and reflection of the agar could have caused some scattering of light. Moreover, the elevation of all colony morphologies was raised, this slight curvature could have contributed to light scatter as well. As there were a large number of sources for possible scattering, the removal of it using SNV was an essential step in reducing these effects, in order to achieve higher classification accuracies.

Score plots and score images of groups 1 and 2 after 40 h growth were used as examples (**Figs. 4.3, 4.4, 4.5** and **4.6**). For group one (**Fig. 4.3**), Gram positive *B. cereus* was represented by the cluster with a red centre (highest density of pixels) in each score plot, whereas Gram negative *E. coli* and *S. enteritidis* were represented by clusters with a yellow/cyan centre. Score plots of raw data, as well as first and second derivative preprocessed data (**Fig. 4.3a, c & d**) showed only one or two clusters. The score plot of SNV transformed data was the only one which showed three clusters (**Fig.**

4.3b.), *B. cereus* was separated from *E. coli* and *S. enteritidis* along PC 1 (30 % sum of squares (SS)) and *E. coli* was separated from *S. enteritidis* along PC2 (11.8 % SS).

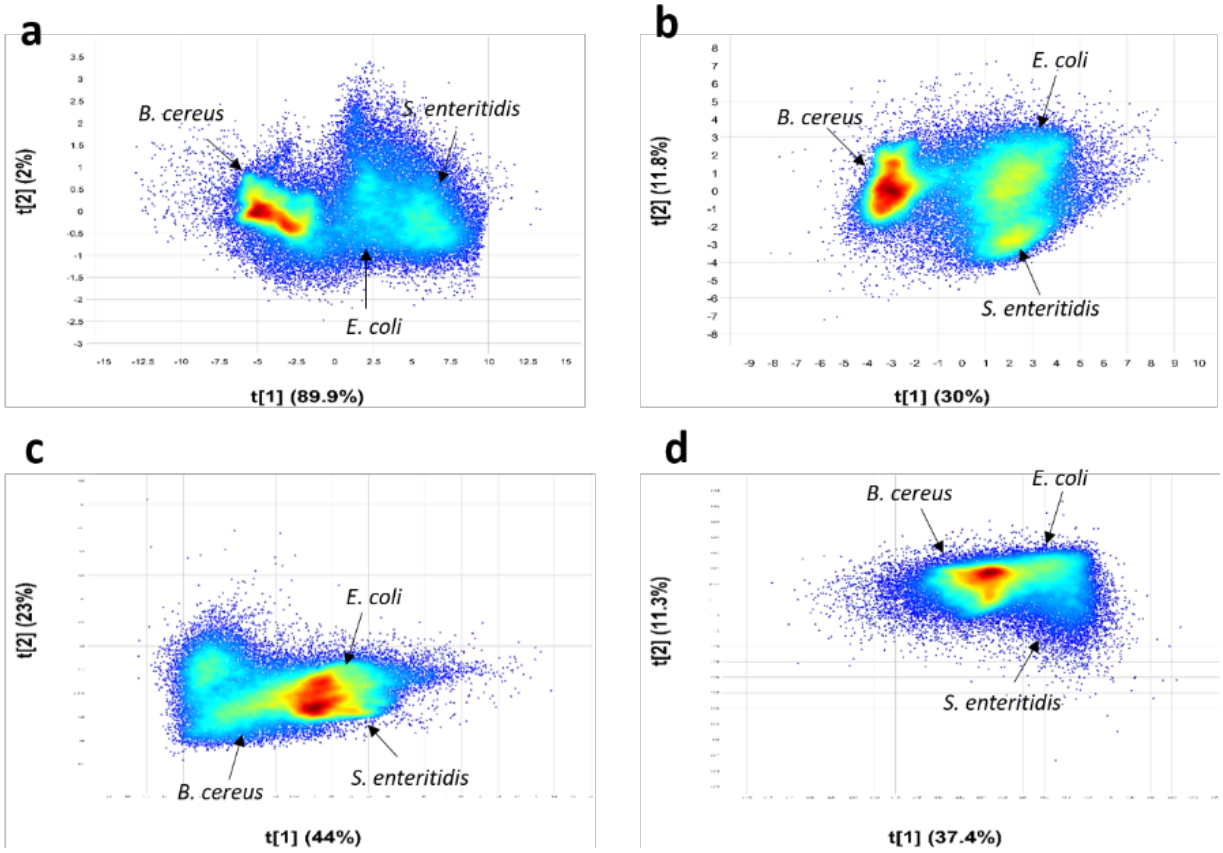

Figure 4.3. Score plots of PC1 vs. PC2 for the mosaic of group 1 (*B. cereus*, *E. coli* and *S. enteritidis*) with different preprocessing methods applied. **a.** Raw data. **b.** SNV. **c.** Savitzky-Golay (1st derivative, 2nd order polynomial, 15 point smoothing). **d.** Savitzky-Golay (2nd derivative, 3rd order polynomial, 15 point smoothing).

In the score image of the raw data (**Fig. 4.4a**), all three bacteria appear different when compared to each other. This could possibly be due to scattering effects caused by varying morphology of bacterial colonies on the growth media. *E. coli* and *S. enteritidis* appear to be similar to each other in the preprocessed score images (**Figs. 4.4b, c & d**), whereas *B. cereus* was different, indicating the presence of notable chemical variances between Gram positive *B. cereus* and the two Gram negative bacteria - *E. coli* and *S. enteritidis*. Differences between Gram-positive and Gram-negative bacteria are due to distinct variation in cell wall composition (Salton, 1953; Navarre & Schneewind, 1999; Beveridge, 1999). In a previous study, it was found that differences in protein and carbohydrate structures in Gram-positive and Gram-negative bacteria as well as absence of teichoic acid in the cell walls of Gram-negative bacteria were the main factors contributing to variation in the score plots (Kammies *et al.*, 2016).

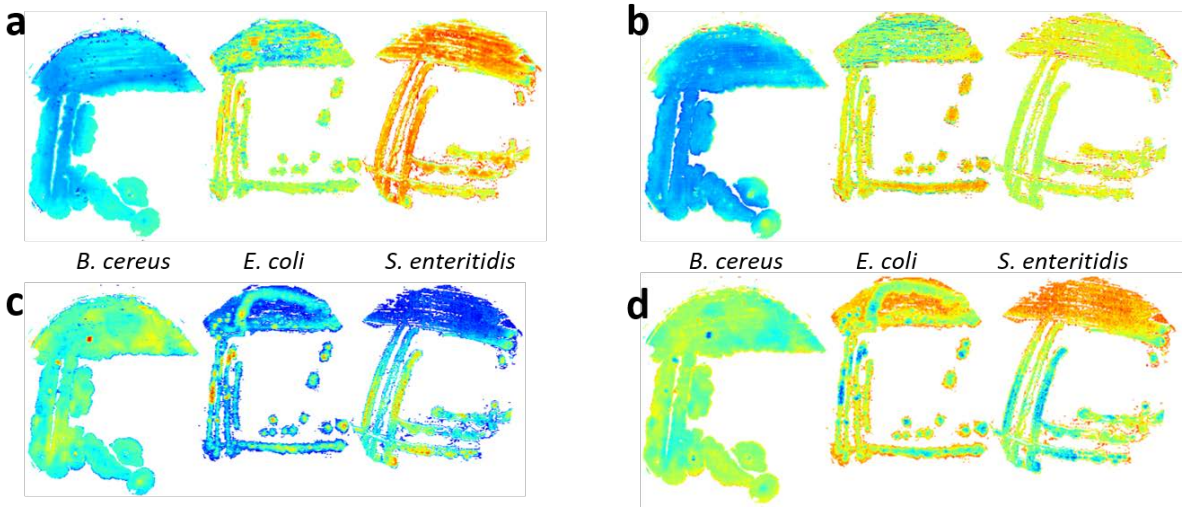

Figure 4.4. Score image of PC1 for the mosaic of group 1 (*B. cereus*, *E. coli* and *S. enteritidis*) with different preprocessing methods applied. **a.** Raw data. **b.** SNV. **c.** Savitzky-Golay (1st derivative, 2nd order polynomial, 15 point smoothing). **d.** Savitzky-Golay (2nd derivative, 3rd order polynomial, 15 point smoothing). Warm colours (yellow/red) are indicative of high score values and cold colours (blue/cyan) show lower score values.

In group 2, the score plot of the SNV transformed data (**Fig**. **4.5b**) again showed the clearest clusters of the three bacteria. Where the two pathogens (*B. cereus* and *S. aureus*) were separated from non-pathogenic *S. epidermidis*. The score image of this pre-processing method (**Fig. 4.6b**) also showed the most variation in score values between the three Gram positive bacteria. Both pathogens appeared to have mostly low score values, however *B. cereus* appeared dark blue while *S. aureus* was cyan. Contrastingly, *S. epidermidis* had mainly higher score values. In the Savitzky-Golay preprocessed score images (**Figs. 4.6c & d**), the two *Staphylococcus* species appeared very similar. These results suggest that scattering effects caused by bacteria morphology and the other sources mentioned previously, play a major role in bacteria differentiation.

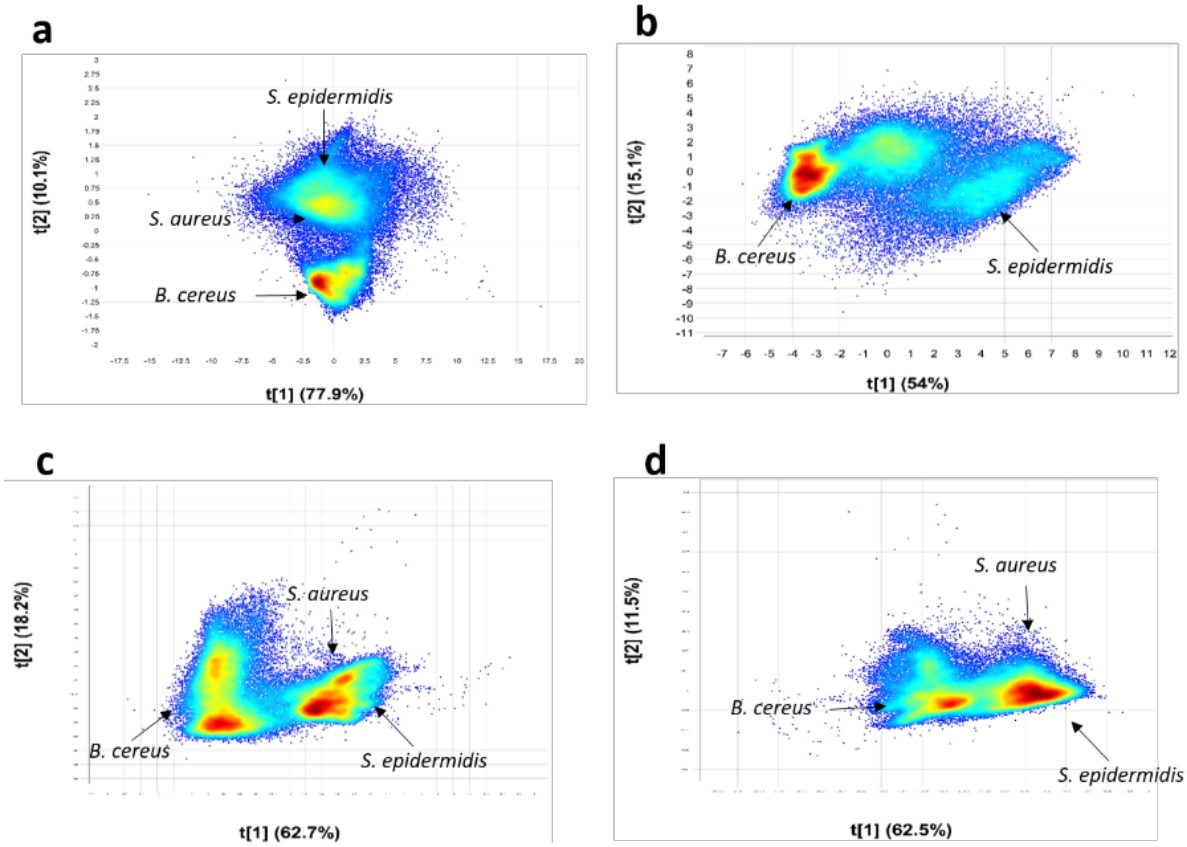

Figure 4.5. Score plots of PC1 vs. PC2 for the mosaic of group 2 (*B. cereus*, *S. aureus* and *S. epidermidis*) with different pre-processing methods applied. **a.** Raw data. **b.** SNV. **c.** Savitzky-Golay (1st derivative, 2nd order polynomial, 15 point smoothing). **d.** Savitzky-Golay (2nd derivative, 3rd order polynomial, 15 point smoothing).

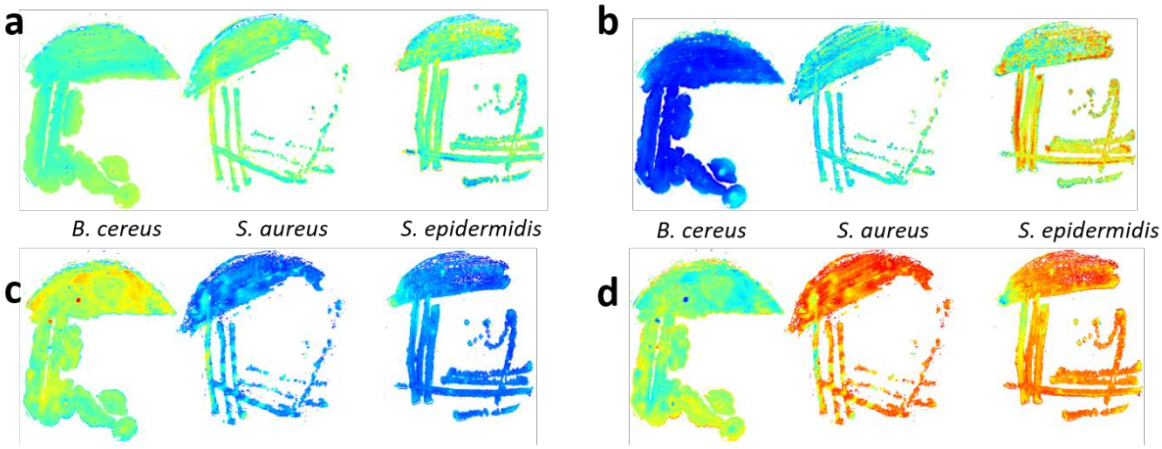

Figure 4.6. Score image of PC1 for the mosaic of group 2 (*B. cereus*, S. aureus and S. epidermidis) with different pre-processing methods applied. **a.** Raw data. **b.** SNV. **c.** Savitzky-Golay (1st derivative, 2nd order polynomial, 15 point smoothing). **d.** Savitzky-Golay (2nd derivative, 3rd order polynomial, 15

point smoothing). Warm colours (yellow/red) are indicative of high score values and cold colours (blue/cyan) show lower score values.

Partial least squares discriminant analysis

The results from the previous chapter indicated that it was possible to distinguish between different bacteria in score plots and score images after 20 h growth, however some classification results were not satisfactory. This prompted an investigation into using older cultures to build PLS-DA models and classify younger colonies, in order to improve classifications while keeping rapidity in mind. For example, models were built after 60 h growth and used to predict 20 h growth plates

The performance of the 20 h model for group 1 (**Table 4.2**) was good, except for classification of *B. cereus* and *E. coli* tested on the second set of 20 h growth plates, where only 69.1 and 70.6 % of pixels were correctly classified. As streak plates were used for this study, and pixel-wise classifications were done, each pixel was regarded as a colony, for this reason these results were deemed unsatisfactory. In practice, if 30 % of pathogenic colonies on a plate was not identified as such, it would mean that the detection method being applied is inadequate since a high percentage of possibly harmful contamination goes undetected. Therefore, in this study any classification accuracy under 90 % correctly predicted pixels was considered low. The prediction accuracy of the 20 h model for group 1 was nevertheless 16 % higher for *B. cereus* than in the 40 h model. However, for *E. coli*, this increased by 12.4 %. Classification results from the 60 h model showed an overall decrease in accuracy, except for *B. cereus* and *E. coli* after 20 h growth, where it increased to 96.9 and 83.1 %, respectively. R^2 for 20, 40 and 60 h models were 0.45, 0.57 and 0.54 respectively. The lowest prediction accuracies for group 1 was for *B. cereus* with the 20 and 40 h models, tested using the 20 h growth plates (69.1 and 53.1 % correctly predicted pixels respectively). The mean spectra of the SNV corrected data, of each bacterium on the respective days, were studied to obtain information about chemical compounds within each bacterium to help explain the classification results. At 1937 nm (C=O, second overtone, CONH/O-H stretch + O-H deformation, H_2O)(Osborne *et al.*, 1993), the spectrum of *B. cereus* after 20 h growth (**Fig 4.7**) was closer to that of *E. coli* after 40 h growth. Whereas after 40 and 60 h growth, the peak intensities were nearly the same throughout. This could explain the low prediction accuracy of *B. cereus* using the 20 h growth plate for the 40 h model where it might have been misclassified as *E. coli*.

There was an increase in prediction results for *E. coli* with the 40 and 60 h models, tested using the 20 h growth plates. However, for the 60 h model, predictions for 40 and 60 h growth plates decreased. With the 20 h model as a base, the only increase in correctly predicted pixels for *S. enteritidis* was in the 40 h model tested using the 60 h growth plate - an increase of 1.9 % (87.9 to

89.8 %) was observed. The best result for *B. cereus* and *S. enteritidis* was seen with the 20 h model, tested using the 40 h growth plate, with 99.8 and 96.0 % correctly predicted pixels respectively, and for *E. coli* it was in the 60 h model, tested using the 60 h of growth plate, with 93.1 % correctly predicted pixels.

Table 4.2. PLS-DA results of group 1 showing classification results of 20, 40 and 60 h models as percentage pixels correctly predicted

	B. cereus	***E. coli***	***S. enteritidis***
% Correctly predicted pixels of validation data – model built from training data generated after 20 h growth (R^2 = 0.45)			
20 h growth	69.1	70.6	90.3
40 h growth	99.8	85.9	96.0
60 h growth	96.0	82.4	87.9
% Correctly predicted pixels of validation data – model built from training data generated after 40 h growth (R^2 = 0.57)			
20 h growth	53.1	83.0	81.6
40 h growth	95.8	86.5	95.2
60 h growth	92.3	93.1	89.8
% Correctly predicted pixels of validation data – model built from training data generated after 60 h growth (R^2 = 0.54)			
20 h growth	96.9	83.1	86.4
40 h growth	98.8	75.3	95.5
60 h growth	93.9	78.2	84.4

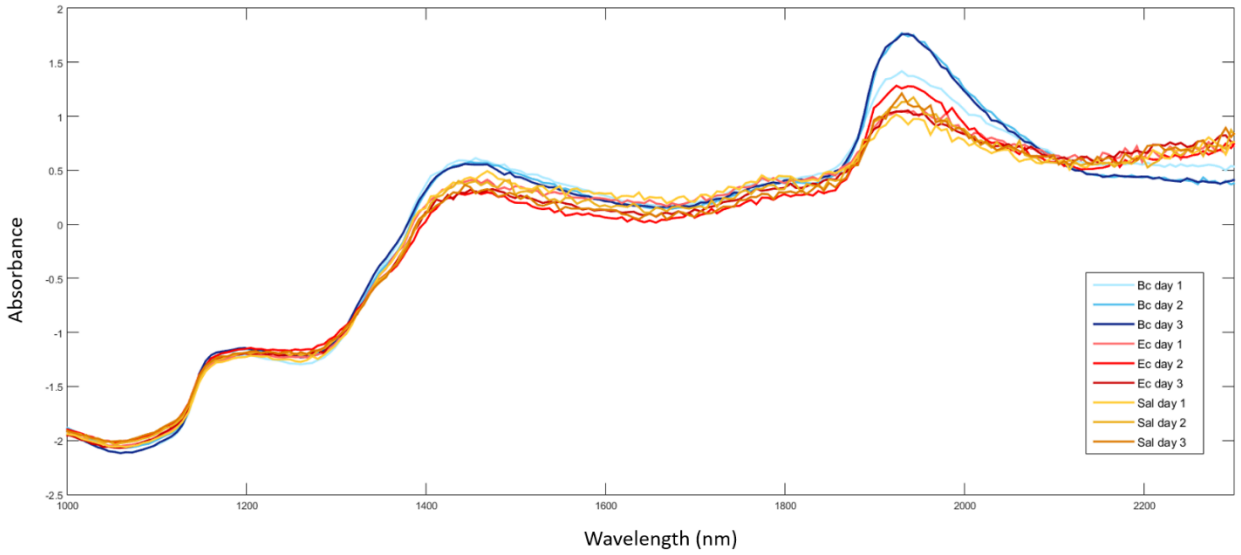

Figure 4.7. SNV corrected mean spectra of group 1 (*B. cereus* (Bc), *E. coli* (Ec) and *S. enteritidis* (Sal)) after 20, 40 and 60 h growth, with the major peak at 1937 nm (C=O, second overtone, CONH/O-H stretch + O-H deformation, H_2O)

For group 2 (**Table 4.3**), the R^2 for 20, 40 and 60 h models were 0.80, 0.75 and 0.69 respectively. An overall decrease in prediction accuracy was observed, except for *S. aureus*, where classification of pixels increased in consecutive models. The highest prediction accuracy for *B. cereus* was 99.9 % with the 20 h model, tested using the 40 h growth plate. For both *S. aureus* and *S. epidermidis*, the 40 h model, showed the highest prediction accuracies tested using the 40 h growth plates, with 97.3 and 99.3 % correctly predicted pixels respectively. In the 20 and 40 h models, *S. aureus* had the lowest prediction accuracies, with only 24.6 % for the 20 h model tested using the 60 h growth plate and 61.4 % for the 40 h model, tested using 60 h growth plate as well. These poor results could possibly be attributed to the slow initial growth of *S. aureus* (Sutherland *et al.*, 1994). This is seen spectrally (**Fig 4.8**) where the intensity of the peak at 1937 nm (C=O, second overtone, CONH/O-H stretch + O-H deformation, H_2O) (Osborne *et al.*, 1993) for *S. aureus* is much lower after 20 h growth than 40 h and 60 h growth. This peak could be related to protein or moisture constituents, which are key components in all bacterial cell walls, as cells increase in population over time, protein and moisture content in colonies increase as well, leading to a higher peak intensity. After 60 h growth the spectral profile of *S. aureus* is more similar to that of *B. cereus* after 40 and 60 h growth which could have led to misclassifications for this growth period.

Table 4.3. PLS-DA results of group 2 showing classification results of 20, 40 and 60 h models as percentage pixels correctly predicted

	B. cereus	***S. aureus***	***S. epidermidis***
% Correctly predicted pixels of validation data – Models built from training data generated after 20 h growth (R^2 = 0.80)			
20 h growth	77.6	96.7	98.2
40 h growth	99.9	71.8	99.2
60 h growth	97.0	24.6	98.4
% Correctly predicted pixels of validation data – Models built from training data generated after 40 h growth (R^2 = 0.75)			
20 h growth	47.4	97.3	99.3
40 h growth	98.4	95.0	95.2
60 h growth	89.6	61.4	95.1
% Correctly predicted pixels of validation data – Models built from training data generated after 60 h growth (R^2 = 0.69)			
20 h growth	96.9	89.1	97.6
40 h growth	97.5	89.1	97.4
60 h growth	93.9	89.1	97.3

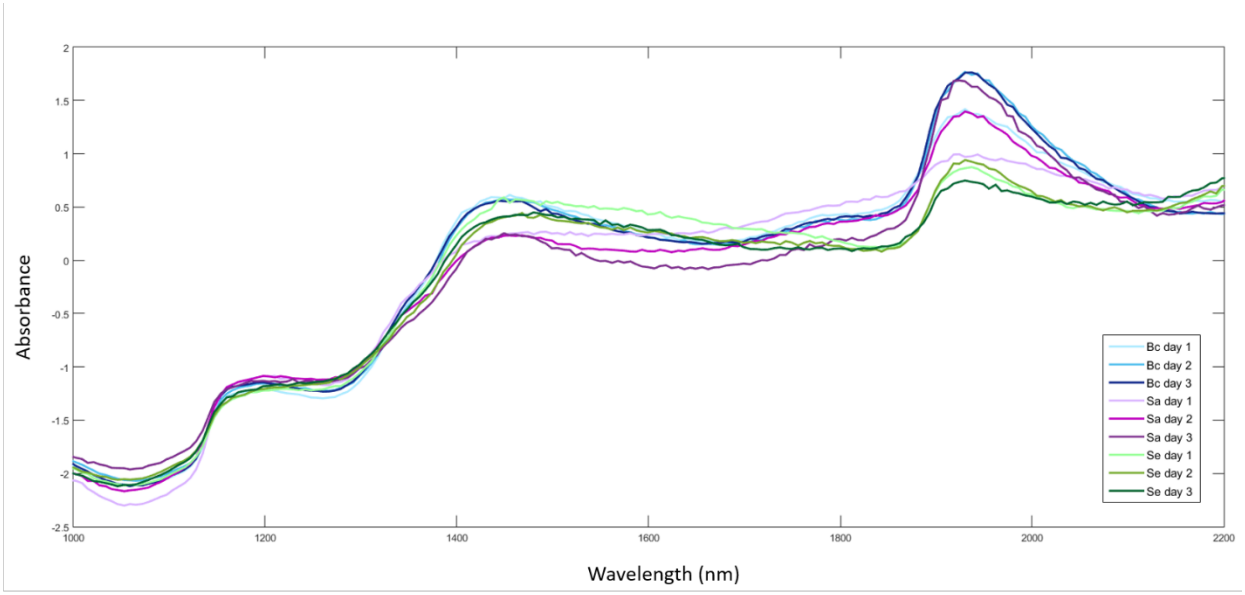

Figure 4.8. SNV corrected mean spectra of group 2 (*B. cereus* (Bc), *S. aureus* (Sa) and *S. epidermidis* (Se)) after 20, 40 and 60 h growth, with the major peak at 1937 nm (C=O, second overtone, CONH/O-H stretch + O-H deformation, H_2O).

For group 3 (**Table 4.4**), R^2 for the 20, 40 and 60 h models were 0.80, 0.83 and 0.83 respectively. Model performances for this group were the best out of the three groups with no prediction accuracy under 90 %. The highest prediction accuracies for *S. aureus* was seen in the 20 h model, tested using the 20 h growth plate, with 99.8 % correctly predicted pixels and with the 60 h model, tested using the 20 h growth plate for *S. epidermidis*, where 98.8 % of pixels were correctly predicted. These two species are closely related (Gillaspy & Landolo, 2009), thus these results are quite encouraging for applications of this technique for differentiation of pathogenic and non-pathogenic species.

Table 4.4. PLS-DA results of group 3 showing classification results of 20, 40 and 60 h models as percentage pixels correctly predicted

	S. aureus	***S. epidermidis***
% Correctly predicted pixels of validation data – Models built from training data generated after 20 h growth (R^2 = 0.80)		
20 h growth	99.8	97.4
40 h growth	99.6	98.2
60 h growth	98.3	90.6
% Correctly predicted pixels of validation data – Models built training from data generated after 40 h growth (R^2 = 0.83)		
20 h growth	99.3	97.0
40 h growth	99.2	97.4
60 h growth	96.0	97.3
% Correctly predicted pixels of validation data – Models built from training data generated after 60 h growth (R^2 = 0.83)		
20 h growth	98.4	98.8
40 h growth	98.4	94.7
60 h growth	97.8	92.8

Conclusions

The age of the colonies used to build models for pathogen differentiation proved to play a role in prediction accuracies. Of the three preprocessing methods applied, SNV produced the best clustering and was therefore used as the only preprocessing method throughout data classification. Classification accuracies for most models were encouraging, regardless of the day of prediction. Deviations in the spectra due to different generation times of the various genera used in this study could have caused misclassification of bacteria.

Granted that the most rapid classifications were desired, it was found that 60 h models with predictions made using 20 h growth plates provided the overall highest prediction accuracies in each group. For this reason, it was the proposed period of growth for model calibration if this method is used for differentiation of bacteria. Future research should include using different types of classification methods, such as support vector machines, to improve models even more, however this classification method requires a large amount of samples (hundreds) so was therefore not included in this study. Additionally, object-wise classification of bacterial colonies using different plating methods (such as spot plates) for models could also improve results. In practise, spread plates would be the selected plating method used in this detection technique, since this would provide single colonies on the surface of the solid agar. Object-wise classification models would allow for entire colonies to be recognised as one bacterium. For example, if more than 50 % of pixels in a colony is recognised as *B. cereus*, the entire colony would be classed as such, allowing for a more definitive result. This could greatly increase the accuracy of this method.

References

Amigo, J.M., Babamoradi, H. & Elcoroaristizabal, S. (2015). Hyperspectral image analysis. A tutorial. *Analytica Chimica Acta,* **896,** 34-51.

Barbin, D.F., Elmasry, G., Sun, D.-W., Allen, P. & Morsy, N. (2013). Non-destructive assessment of microbial contamination in porcine meat using nir hyperspectral imaging. *Innovative Food Science & Emerging Technologies,* **17,** 180-191.

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

Beveridge, T.J. (1999). Structures of Gram-negative cell walls and their derived membrane vesicles. *Journal of Bacteriology,* **181,** 4725-4733.

Blanco, M. & Villarroya, I. (2002). Nir spectroscopy: A rapid-response analytical tool. *TRAC Trends in Analytical Chemistry,* **21,** 240-250.

Buchanan, R.E. (1974). Life phases in a bacterial culture. In: Micriobial growth (edited by DAWSON, P. S. S.). Pp. 25-41. Stroudsburg, Pennsylvania: Dowden, Hutchinson & Ross Inc.

Burger, J. & Geladi, P. (2006). Hyperspectral nir image regression part ii: Dataset preprocessing diagnostics. *Journal of Chemometrics,* **20,** 106-119.

Chevallier, S., Bertrand, D., Kohler, A. & Courcoux, P. (2006). Application of PLS-DA in multivariate image analysis. *Journal of Chemometrics,* **20,** 221-229.

Collins, J. & Richmond, M. (1962). Rate of growth of *Bacillus cereus* between divisions. *Microbiology,* **28,** 15-33.

Dubois, J., Lewis, E.N., Fry, F.S. & Calvey, E.M. (2005). Bacterial identification by near-infrared chemical imaging of food-specific cards. *Food Microbiology,* **22,** 577-583.

Esbensen, K. & Geladi, P. (1989). Strategy of multivariate image analysis (MIA). *Chemometrics and Intelligent Laboratory Systems,* **7,** 67-86.

Fearn, T., Riccioli, C., Garrido-Varo, A. & Guerrero-Ginel, J.E. (2009). On the geometry of SNV and MSC. *Chemometrics and Intelligent Laboratory Systems,* **96,** 22-26.

Fehlhaber, K. & Krüger, G. (1998). The study of *Salmonella enteritidis* growth kinetics using rapid automated bacterial impedance technique. *Journal of Applied Microbiology,* **84,** 945-949.

Geladi, P., Isaksson, H., Lindqvist, L., Wold, S. & Esbensen, K. (1989). Principal component analysis of multivariate images. *Chemometrics and Intelligent Laboratory Systems,* **5,** 209-220.

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

Gillaspy, A.F. & Landolo, J.J. (2009). Staphylococcus. In: Encyclopedia of microbiology (edited by SCHAECHTER, M.). Pp. 293-303. San Diego, USA: Elsevier Science.

Gottenbos, B., Grijpma, D.W., Van Der Mei, H.C., Feijen, J. & Busscher, H.J. (2001). Antimicrobial effects of positively charged surfaces on adhering Gram-positive and Gram-negative bacteria. *Journal of Antimicrobial Chemotherapy,* **48,** 7-13.

Gowen, A.A., Taghizadeh, M. & O'donnell, C.P. (2010). Using hyperspectral imaging for quality evaluation of mushrooms. In: Hyperspectral imaging for food quality analysis and control (edited by SUN, D.-W.). Pp. 403-430. San Diego: Academic Press.

Kammies, T.-L., Manley, M., Gouws, P.A. & Williams, P.J. (2016). Differentiation of foodborne bacteria using NIR hyperspectral imaging and multivariate data analysis. *Applied Microbiology and Biotechnology,* **100,** 9305-9320.

Keren, I., Kaldalu, N., Spoering, A., Wang, Y. & Lewis, K. (2004). Persister cells and tolerance to antimicrobials. *FEMS Microbiology Letters,* **230,** 13-18.

Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chemical Society Reviews,* **43,** 8200-8214.

Manley, M., Downey, G. & Baeten, V. (2008). Spectroscopic technique: Near-infrared (NIR) spectroscopy. In: Modern techniques for food authentication (edited by SUN, D.-W.). Pp. 65-99. Elsevier Inc.

Martens, H., Jensen, S. & Geladi, P. (1983). Multivariate linearity transformation for near-infrared reflectance spectrometry. In: Proceedings of the Nordic symposium on applied statistics. Pp. 205-234. Stokkand Forlag Publishers Stavanger, Norway.

Navarre, W.W. & Schneewind, O. (1999). Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiology and Molecular Biology Reviews,* **63,** 174-229.

Norris, K. (1959). Infra-red spectroscopy and its application to microbiology. *Journal of Hygiene,* **57,** 326-345.

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

Rinnan, Å., Nørgaard, L., Berg, F.V.D., Thygesen, J., Bro, R. & Engelsen, S.B. (2009a). Chapter 2 - data pre-processing a2 - sun, da-wen. In: Infrared spectroscopy for food quality analysis and control. Pp. 29-50. San Diego: Academic Press.

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

Salton, M. (1953). Studies of the bacterial cell wall: Iv. The composition of the cell walls of some Gram-positive and Gram-negative bacteria. *Biochimica et Biophysica Acta,* **10,** 512-523.

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

Srivastava, S. & Srivastava, P.S. (2013). Bacteria and life processes - growth and multiplication. In: Understanding bacteria. Pp. 97-148. Springer Netherlands.

Sutherland, J., Bayliss, A. & Roberts, T. (1994). Predictive modelling of growth of *Staphylococcus aureus*: The effects of temperature, ph and sodium chloride. *International Journal of Food Microbiology,* **21,** 217-236.

Varmuza, K. & Filzmoser, P. (2016). Preprocessing: Smoothing and differentiation. In: Introduction to multivariate statistical analysis in chemometrics. Pp. 283-284. New York: CRC press.

Williams, P.J., Geladi, P., Britz, T.J. & Manley, M. (2012). Growth characteristics of three fusarium species evaluated by near-infrared hyperspectral imaging and multivariate image analysis. *Applied Microbiology and Biotechnology,* **96,** 803-813.

Wold, S., Esbensen, K. & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems,* **2,** 37-52.

Yoon, S.-C., Windham, W.R., Ladely, S.R., Heitschmidt, J.W., Lawrence, K.C., Park, B., Narang, N. & Cray, W.C. (2013). Hyperspectral imaging for differentiating colonies of non-o157 shiga-toxin producing *Escherichia coli* (STEC) serogroups on spread plates of pure cultures. *Journal of Near Infrared Spectroscopy,* **21,** 81-95.

Yoon, S., Lawrence, K., Siragusa, G., Line, J., Park, B. & Feldner, P. (2009). Hyperspectral reflectance imaging for detecting a foodborne pathogen: *Campylobacter*. *Transactions of the ASABE,* **52,** 651-662.

Yoon, S.C., Lawrence, K.C., Line, J.E., Siragusa, G.R., Feldner, P.W., Park, B. & Windham, W.R. (2010). Detection of *Campylobacte*r colonies using hyperspectral imaging. *Sensing and Instrumentation for Food Quality and Safety,* **4,** 35-49.

CHAPTER 5

General discussion and conclusions

Foodborne pathogens comprise of bacteria, viruses, fungi and parasites (Zhao *et al.*, 2014). There are 31 pathogens which have been identified as causing foodborne illnesses. The majority of illnesses, hospitalisations and deaths are caused by Norovirus, *Salmonella*, *Campylobacter*, *Staphylococcus aureus*, *Listeria monocytogenes*, *Clostridium perfringens*, *Toxoplasma gondii* and *Escherichia coli* O157:H7 (Velusamy *et al.*, 2010; Zhao *et al.*, 2014; Chamberlain *et al.*, 1988). The widespread occurrence of foodborne diseases in developed and developing countries indicate underlying food safety flaws in adequate pathogen detection (Zhao *et al.*, 2014). Conventional methods for pathogen detection and identification are labour intensive and time consuming – often taking two to three days for initial results (enrichment and growth on specialised media) and up to a week for species identification. Pure culture identifications are typically required for regulatory and legal purposes (Qin *et al.*, 2013). For this reason, rapid methods for pathogen detection and improvements on existing methods are constantly being investigated.

In this study near infrared (NIR) hyperspectral imaging, together with multivariate image analysis (MIA) was investigated as a tool to be used in food microbiology laboratories to decrease detection and differentiation/identification time of foodborne pathogenic bacteria.

Bacteria are defined as Gram positive or Gram negative depending on their cell wall structure. Gram-positive cell walls are quite simple, and are defined by their thick peptidoglycan layer, teichoic and lipoteichoic acids and polysaccharides (Schleifer & Kandler, 1972). Whereas Gram negative bacterial cell walls are more complex, with an outer membrane, a thin peptidoglycan layer, periplasmic space and an inner cytoplasmic membrane (Shockman & Barren, 1983; Trüper & Schleifer, 2006; Bhunia, 2007; Beveridge, 1999). Additionally, pathogenic bacteria require structures such as cell wall-anchored surface proteins to adhere to host cell tissues (Navarre & Schneewind, 1999). In this study, exploratory principal component analysis (PCA) was essential for successfully differentiating pathogenic bacteria. In order to expand on which chemical differences were responsible for variance seen in the organisms investigated, bacteria were divided into three groups. The first group contained three bacteria which appeared similar in colour on the agar media (*Bacillus cereus*, *Escherichia coli* and *Salmonella enteritidis*), the second had three Gram-positive bacteria (*B. cereus*, *Staphylococcus aureus* and *Staphylococcus epidermidis*), and the third comprised of a pathogenic and non-pathogenic species of the same genus (*S. aureus* and *S. epidermidis*). Clustering of pixels according to bacterial genera was evident along principal components (PC's) 1 and 2 in score plots, despite overlapping of pixels of all bacteria. Differentiation was found to be based on cell wall variances across genera, species and pathogenicity. It was apparent that this unsupervised

technique could be used to distinguish between pathogenic bacteria. Group 1 showed that it was possible to differentiate between Gram-positive and Gram negative pathogenic bacteria as well as between pathogenic bacteria which appear similar in colour on solid growth media. Group 2 illustrated that Gram-positive pathogenic bacteria appeared to be more similar to each other in score plots than to the non-pathogenic bacteria used in this study. The final group (3) was used to distinguish between a pathogenic and non-pathogenic species of the same genera.

Spectrally, protein, sugars, carbohydrates, lipids and teichoic and/or lipoteichoic acids were found to be the main contributing variables responsible for variation between the bacteria analysed. These components were attributed to cell wall composition. In loading lines 1 and 2, protein structures were represented by peaks at 2130, 1974, 2167, 1980, 1524 and 2067 nm. Carbohydrates were represented by peaks at 1781 and 1824 nm. Lipids were represented by peaks at 1693 and 2310 nm and teichoic and/or lipoteichoic acid by 1405 nm. In the mean spectra, proteins were represented by peaks at 1424, 1937 and 1436 nm and carbohydrate by 1781 nm. In order to use this information in the food industry, it was necessary for the data to be evaluated using multivariate data analysis to build classification models. For that reason, partial least squares-discriminant analysis (PLS-DA) was applied to determine whether good discrimination between foodborne pathogenic bacteria could be estimated. Initially, PLS-DA models were evaluated using the shortest growth time (20 h) and the same pre-processing used for the PCA models (standard normal variate (SNV) correction and Savitzky-Golay (2^{nd} derivative, 3^{rd} order polynomial; 25 point smoothing), however, certain prediction results were unsatisfactory. The poorest results were of *E. coli* and *S. enteritidis*, where only 2.34 and 53.9 % of pixels were correctly predicted respectively. Prediction results were deemed inadequate if it was below 90 %, as a large number of misclassified pixels could result in potentially harmful bacteria not being identified as such. In an attempt to improve these predictions, it was decided to test models built separately after three growth time increments (20, 40 and 60 h growth). Each model was validated using a test set of plates, where predictions were made on each of the three time increment models for each of the three groups. Results showed major improvements in prediction accuracies of, especially when SNV was used as the only pre-processing method. The prediction accuracies for *B. cereus* was up to 99.9 %, *E. coli*, 93.1 %, *S. enteritidis*, 96, 0 %, *S. aureus*, 99.8 % and *S. enteritidis*, up to 99.8 %, depending on the model used. It was observed that differentiation between *E. coli* and *S. enteritidis* was more complex than any of the other bacteria. This could have been due to the Gram-negative nature of their cell walls – Gram-negative cell walls are known to be quite similar in composition, among genera and species (Beveridge, 1999). This study however, did not include verification (with NIR spectroscopy) of pure chemical components (teichoic or lipoteichoic acid) mentioned. This could be included in further

investigations, by comparing spectra of the chemical components and the bacteria to observe whether the peak at 1405 nm and perhaps any others were common in both. This could provide a more comprehensive result of chemical components involved in differentiation of bacteria.

Although NIR hyperspectral imaging has previously been used for differentiation of foodborne bacteria on agar (Yoon *et al.*, 2009; Windham *et al.*, 2012), these studies utilised wavelength ranges in the vis-NIR region. Discrimination between bacterial genera and species in these studies were mostly found to be based on colour differences, limiting detection to pathogens which are already identifiable on the selective media

In this study, it was clear that differences of each bacterium, such as cell structure, pathogenicity and age of colonies have an influence on how models should be built to obtain optimal prediction accuracies. It was conducted to prove that chemical differences between foodborne pathogenic bacteria can be detected on solid growth media, using NIR hyperspectral imaging with wavelengths above 1100 nm. A small amount of the 31 identified foodborne pathogens were investigated, therefore, future work should include more of these, particularly those with longer identification times, like *Listeria monocytogenes*. A method with decreased detection/identification time could be especially beneficial to the ready-to-eat food sector and those products with a short shelf life.

To improve prediction results even more, object–wise classifications (Kucheryavskiy, 2013) should be explored, where entire colonies can be identified as one bacterium, according to what the majority of the colony is predicted as. Spot plates could be used to build calibrations, however for practicality it should be tested on spread plates, as this would be the most rapid method of producing single colonies in practice. Furthermore, improvements might also be possible when using other classification methods, such as Soft Independent Modelling of Class Analogy (SIMCA) (Wold & Sjöström, 1977), support vector machines (SVM) (Cortes & Vapnik, 1995; Burges, 1998) or artificial neural networks (ANN).(Rosenblatt, 1962)

NIR hyperspectral imaging offers the ability to rapidly distinguish between foodborne pathogenic bacteria on agar, to efficiently prevent contaminated products from entering the food chain. While it was proven that the NIR region can be utilised to distinguish and classify foodborne pathogenic bacteria, additional work on classification of colonies is still required.

References

Beveridge, T.J. (1999). Structures of gram-negative cell walls and their derived membrane vesicles. *Journal of Bacteriology,* **181,** 4725-4733.

Bhunia, A. (2007). Foodborne microbial pathogens: Mechanisms and pathogenesis. Springer Science & Business Media.

Burges, C.J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery,* **2,** 121-167.

Chamberlain, J.S., Gibbs, R.A., Rainer, J.E., Nguyen, P.N. & Thomas, C. (1988). Deletion screening of the duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucleic Acids Research,* **16,** 11141-11156.

Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine learning,* **20,** 273-297.

Kucheryavskiy, S. (2013). A new approach for discrimination of objects on hyperspectral images. *Chemometrics and Intelligent Laboratory Systems,* **120,** 126-135.

Navarre, W.W. & Schneewind, O. (1999). Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiology and Molecular Biology Reviews,* **63,** 174-229.

Qin, J., Chao, K., Kim, M.S., Lu, R. & Burks, T.F. (2013). Hyperspectral and multispectral imaging for evaluating food safety and quality. *Journal of Food Engineering,* **118,** 157-171.

Rosenblatt, F. (1962). Principles of neurodynamics.

Schleifer, K.H. & Kandler, O. (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriological reviews,* **36,** 407-477.

Shockman, G.D. & Barren, J. (1983). Structure, function, and assembly of cell walls of gram-positive bacteria. *Annual Reviews in Microbiology,* **37,** 501-527.

Trüper, H.G. & Schleifer, K.H. (2006). Prokaryote characterization and identification. In: The prokaryotes. Pp. 58-79. Springer.

Velusamy, V., Arshak, K., Korostynska, O., Oliwa, K. & Adley, C. (2010). An overview of foodborne pathogen detection: In the perspective of biosensors. *Biotechnology Advances,* **28,** 232-254.

Windham, W., Yoon, S.-C., Ladely, S., Heitschmidt, J., Lawrence, K., Park, B., Narrang, N. & Cray, W. (2012). The effect of regions of interest and spectral pre-processing on the detection of non-0157 shiga-toxin producing *escherichia coli* serogroups on agar media by hyperspectral imaging. *Journal of Near Infrared Spectroscopy,* **20,** 547-558.

Wold, S. & Sjöström, M. (1977). Simca: A method for analyzing chemical data in terms of similarity and analogy. ACS Publications.

Yoon, S., Lawrence, K., Siragusa, G., Line, J., Park, B. & Feldner, P. (2009). Hyperspectral reflectance imaging for detecting a foodborne pathogen: Campylobacter. *Transactions of the ASABE,* **52,** 651-662.

Zhao, X., Lin, C.-W., Wang, J. & Oh, D.H. (2014). Advances in rapid detection methods for foodborne pathogens. *J. Microbiol. Biotechnol,* **24,** 297-312.