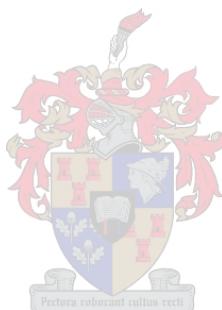


Rapid differentiation of South African game meat using portable near-infrared (NIR) spectroscopy

by

Kiah Payne



*Thesis presented in partial fulfilment of the requirements for the degree
of Master of Science (Food Science)
in the Faculty of AgriSciences
at
Stellenbosch University*

Supervisor: Dr Paul J Williams

Co-supervisor: Prof Marena Manley

Co-supervisor: Prof Louwrens C Hoffman

April 2019

Declaration

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

Date: April 2019

Abstract

The South African game meat industry operates as a free-market enterprise; however, this can create certain problems for producers and consumers. For example, in South Africa no standardised meat cuts exist and there are no quality standards in place for game. Therefore, allowing the legal sale of inferior quality game meat. Due to a general lack of regulations as well as varying carcass dressings, the chance that a species may be mislabelled or substituted is greatly increased.

In recent years, meat authenticity awareness has increased, as there have been incidences where meat has been fraudulently mislabelled. Typical cases involve the intentional substitution of high value raw ingredients with inferior species or materials, the addition of non-declared proteins from several origins, or the marketing of frozen-thawed meat as fresh. This type of food fraud concerns consumers in terms of economic loss, food allergies, religious compliance, and food safety. This study aimed to investigate a feasible alternative to the manual, tedious and time-consuming conventional analytical methods used for meat differentiation and authentication that could provide the meat industry with a rapid, non-destructive, accurate and reliable solution in the near future.

Near-infrared (NIR) spectroscopy combined with multivariate data analysis (MDA) techniques were used to rapidly differentiate between South African game species, irrespective of the treatment (fresh or previously frozen) or the muscle type as well as determine these individual classes (fresh; previously frozen; frozen period; muscle type) per species. Meat samples of four game species [black wildebeest (*Connochaetes gnou*), zebra (*Equus quagga burchelli*), springbok (*Antidorcas marsupialis*), ostrich (*Struthio camelus*)] were scanned at *ca.* 23° C. Spectra were collected with a portable MicroNIR OnSite spectrophotometer (Viavi Solutions Inc., Milpitas, USA), in the range of 908 – 1676 nm, after which the data was analysed using MDA.

It was possible to differentiate between game species, irrespective of the treatment (fresh or previously frozen), the frozen period or the muscle type. The partial least squares discriminant analysis (PLS-DA) model was successful and achieved accuracies ranging 89.8 – 93.2%. It was also possible to distinguish between fresh and previously frozen meat, and with low accuracies determine the frozen period. The principal component analysis (PCA) score plots illustrated good separation between the fresh and frozen-thawed samples, however, the frozen-thawed samples exhibited an overlap between the individual frozen periods. Therefore, lacking separation and distinct clustering for the different frozen periods. Thus, the MDA models were not effective when trying to classify the different frozen periods. The PLS-DA models could however discriminate between the fresh and previously frozen meat, irrespective of the frozen period or muscle type. The black wildebeest (99.2%), zebra (94.4 and 99.3%), springbok (100%) and ostrich (90 and 98.3%) models achieved good overall accuracies.

Lastly, this study found that the ostrich muscles could be distinguished from each other with a 100% accuracy. Furthermore, the results suggested that discrimination of the different muscle types for both zebra and springbok was less sufficient due to the lack of separation between the different muscles.

Misclassification mostly occurred between muscles that are anatomically located near to one another. Therefore, the samples with spectral similarities were grouped to form a two-group class discrimination model. The PLS-DA results showed that it was possible to differentiate between the forequarters and hindquarters of the zebra (90.3%) and springbok (97.9%) muscles.

The results showed NIR spectroscopy's potential as a rapid and non-destructive method for species identification, fresh and previously frozen meat differentiation as well as muscle type determination. Furthermore, this technique has the potential of providing the South African game meat industry with an alternative technique to the current manual, destructive and time-consuming authentication methods.

Uittreksel

Die Suid-Afrikaanse wildsvleisbedryf word as 'n vryemarkonderneming bestuur; dit kan egter sekere probleme vir produsente en verbruikers voortbring. Byvoorbeeld, in Suid-Afrika is daar geen gestandardiseerde vleissnitte nie en daar is geen kwaliteitstandarde in plek vir wild nie. Daarom, word die wettige verkoop van miderwaardige kwaliteit wildsvlies toegelaat. As gevolg van 'n algemene gebrek aan regulasies sowel as verskeie karkasse, word die kans dat 'n spesie dalk verkeerdelik geetiketteer of vervang word, aansienlik verhoog.

In onlangse jare het die bewustheid van vleis-egtheid toegeneem, aangeien daar voorvalle was waar vleis bedrieglig geetiketteer was. Tipiese gevalle behels die opsetlike vervanging van hoë-waarde rou bestanddele met minderwaardige spesies of materiale, die toevoeging van nie-verklaarde proteïene vanaf verskeie oorspronge, of the bemarking van bevroe-ontdooide vleis as vars. Hierdie soort vodeselbedrog het betrekking op verbruikers in terme van ekonomiese verlies, voedselallergieë, godsdienstige nakoming en voedselveiligheid. Hierdie studie beoog om 'n haalbare alternatief vir die huidige met-die-hand, vervelige en tydrowende konvensionele analitiese metodes vir vleisdifferensiasie en verifikasié te ondersoek, wat die vleisbedryf in die nabye toekoms 'n vinnige, nie-vernieterende, akkurate en betroubare oplossing sal bied.

Naby-infrarooi (NIR) spektroskopie gekombineer met meerveranderlike data analise (MDA) tegnieke is gebruik om vinnig tussen Suid-Afrikaanse wildspesies te onderskei, ongeag die behandelling (vars of voorheen bevroe) of die spier tipe sowel as om hierdie individuele klasse te bepaal (vars; voorheen bevroe; bevroe tydperk; spier tipe) per spesie. Vleismonsters van vier wildspesies [swartwildebees (*Connochaetes gnou*), sebra (*Equus quagga burchelli*), springbok (*Antidorcas marsupialis*), volstruis (*Struthio camelus*)] is geskandeer teen ca. 23° C. Spekta is versamel met 'n draagbare MicroNIR OnSite spektrofotometer (Viavi Solutions Inc., Milpitas, USA), in die golflengte reeks van 908 – 1676 nm, waarna die data met MDA geanalyseer is.

Dit was moontlik om tussen wildspesies te onderkei, ongeag die behandelling (vars of voorheen bevroe), die bevroe tydperk of die spier tipe. Die parsiële kleinste kwadrate diskriminantanalise (PLS-DA) model was suksesvol en het akkuraatheid bereik was 89.8 – 93.2% dek. Dit was ook moontlik om tuseen vars en voorheen bevroe vleis te onderskei, en met lae akkuraatheid die bevroe periode te bepaal. Die hoofkomponentanalise (PCA) telling-beeldetelling het goeie skeiding tussen die vars en bevroe ontdooide monsters geïllustreer, maar die bevroe-ontdooide monsters het 'n oorvleueling tussen die individuele bevroe tydperke getoon. Daarom ontbreek skeiding en afsonderlike groepering vir die verskillende bevroe tydperke. Die MDA modelle was dus nie effektief om die verskillende bevroe tydperke te probeer klassifiseer nie. Die PLS-DA modelle kon egter onderskei tussen die vars en voorheen bevroe vleis, ongeag die bevroe tydperk of die spier tipe. Die swartwildebees (99.2%), sebra (94.4 en 99.3%), springbok (100%) en volstruis (90 en 98.3%) modelle het goeie algehele akkuraatheid behaal.

Laastens het hierdie studie bevind dat die volstruisspiere van mekaar onderskei kan word met 'n 100% akkuraatheid. Verder het die resultate voorgestel dat die diskriminasie van die verskillende spier tipes vir beide sebra en springbok minder voldoende was weens die gebrek aan skeiding tussen die verskillende spiere. Misklassifikasie het meestal plaasgevind tussen spiere wat anatomies naby mekaar geleë is. Daarom is die monsters met spektrale ooreenkomsste gegroepeer om 'n tweegroep-klas diskriminasie model te vorm. Die PLS-DA resultate het getoon dat dit moontlik was om tussen die voor- en agterkwarte van die sebra (90.3%) en springbok (97.9%) spiere te onderskei.

Die resultate het NIR spektroskopie se potensiaal getoon as 'n vinnige en nie-verniegende metode vir die identifikasie van spesies, vars en vorheen bevrore vleis differensiasie sowel as spier-tipe bepaling. Verder het hierdie tegniek die potensiaal om die Suid-Afrikaanse wildsvleisbedryf te voorsien met 'n alternatiewe tegniek vir die huidige met-die-hand, destruktiewe en tydrowende verifikasiemetodes.

Acknowledgements

I would like to thank my supervisors, Dr Williams, Prof Manley (Department of Food Science, Stellenbosch University) and Prof Hoffman (The University of Queensland, Australia), for the continues support of my research. Their patience, guidance, motivation and vast knowledge were greatly appreciated throughout the completion of my research and writing of my thesis.

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

Prof Huck and PhD students Anel Beganovic and Christian Kirchler (Institute of Analytical Chemistry and Radiochemistry, Innsbruck University, Austria) for welcoming Dr Williams and myself so nicely in Austria, as well as their interest in this collaborative project and expert advice while introducing me to the Unscrambler X software. It was extremely insightful and beneficial to work alongside experts in the field of Chemometrics and spectral data analysis;

The Department of Animal Sciences, Stellenbosch University for supplying all game meat samples and assisting with sample preparation;

All the staff and post-graduate students from the Departments of Animal Sciences and Food Science (Stellenbosch University) for their help, assistance, providing a friendly work environment and support when needed. A special thanks to Anchen Lombard for assisting with travel arrangements; and

The National Research Foundation (NRF) [Thutuka Grant Holder Bursary (94031), 2017 and 2018] is 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).

This research is supported by the South African Research Chairs Initiative (SARChI) and funded by the South African Department of Science and Technology (UID: 84633), as administered by the National Research Foundation (NRF) of South Africa. The financial assistance of the NRF towards this research is hereby acknowledged as are the financial contributions of the Technology for Human Resource and Industrial Program: (THRIP/64/19/04/2017) and Wildlife Ranching South Africa. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the NRF.

I would also like to thank Robyn Vice for helping me compile my thesis chapters – Dude ek sou dit nie sonder jou hulp kon doen nie;

Thank you to my family and friends, without whom I would not have been able to complete this study. My parents, Steffan and Beneal and my brother Troy have been my primary support system throughout this journey and I would like to thank them for their continual love, support, motivation and encouragement.

Lastly, to my fiancé Richard Frederick Edwards, thank you for your willingness to help me, whether it be lab work or proof reading my thesis, all your time and effort is highly appreciated my love. Thank you for the continues love, support and encouragement throughout this journey and completion of my research.

Table of Contents

Declaration	i
Abstract	ii
Uittreksel	iv
Acknowledgements	vi
List of Figures.....	xiii
List of Tables.....	xx
List of Abbreviations.....	xxvi
Chapter 1 Introduction	1
1.1 References	5
Chapter 2 Near-infrared (NIR) spectroscopy for meat authentication: a review	12
2.1 Introduction.....	12
2.2 Food fraud	13
2.3 Meat fraud.....	15
2.3.1 Meat origin and substitution	16
2.3.2 Meat treatment	18
2.4 Conventional analytical detection methods.....	18
2.4.1 Immunological methods	19
2.4.2 Electrophoretic and chromatographic methods	19
2.4.3 DNA-based procedures.....	20
2.4.4 Enzyme activity determination.....	21
2.4.5 Nuclear magnetic resonance	22
2.4.6 Electron microscopy	22
2.5 Near-infrared (NIR) spectroscopy.....	23
2.5.1 Principles of NIR spectroscopy	24
2.5.1.1 What is NIR spectroscopy?	24
2.5.1.2 NIR instrumentation and spectral acquisition.....	25
2.5.1.3 Reflectance and transmittance mode	25
2.5.1.4 Interactance mode	26
2.5.1.5 Filter-based instruments	27
2.5.1.6 LED based instruments	27
2.5.1.7 Dispersive instruments	27
2.5.1.8 Interferometric instruments.....	27
2.5.1.9 Benchtop and portable handheld devices.....	28
2.6 Chemometrics and multivariate data analysis (MDA).....	29
2.6.1 Pre-processing	30
2.6.2 Spectral processing.....	32

2.6.3 Principal component analysis (PCA)	33
2.6.4 Soft independent modelling of class analogy (SIMCA).....	34
2.6.5 K-nearest neighbour (KNN).....	35
2.6.6 Linear discriminant analysis (LDA) and partial least squares discriminant analysis (PLS-DA)	37
2.6.7 Support vector machines (SVM).....	38
2.7 Applications of NIR spectroscopy in meat authentication	41
2.8 Future perspective.....	46
2.9 Conclusion	47
2.10 References	48
Chapter 3 Materials and Methods	62
3.1 Samples, sampling and sample preparation.....	62
3.2 NIR instrumentation	63
3.3 Spectral acquisition	63
3.4 Spectral analysis	65
3.4.1 Pre-processing	66
3.5 Exploratory data analysis (EDA).....	67
3.5.1 Principal component analysis (PCA) and difference spectra.....	67
3.6 Multivariate data analysis (MDA)	67
3.6.1 Model development	67
3.6.1.1 Soft independent modelling of class analogy (SIMCA).....	68
3.6.1.2 K-nearest neighbour (KNN).....	68
3.6.1.3 Discriminant analysis (DA)	69
3.6.1.4 Partial least squares discriminant analysis (PLS-DA)	69
3.6.2 Performance measures.....	70
3.7 Hierarchical model development	71
3.8 References	71
Chapter 4 Results and Discussion.....	73
4.1 Species determination	73
4.1.1 Spectral analysis	73
4.1.2 Species determination irrespective of muscle type	74
4.1.2.1 Exploratory data analysis (EDA).....	75
4.1.2.1.1 Principal component analysis (PCA)	75
4.1.2.1.2 Difference spectra	79
4.1.2.2 Multivariate data analysis (MDA): Model development	82
4.1.2.2.1 Soft independent modelling of class analogy (SIMCA).....	82
4.1.2.2.2 K-nearest neighbour (KNN)	85
4.1.2.2.3 Discriminant analysis (DA)	89

4.1.2.2.4 Partial least squares discriminant analysis (PLS-DA)	91
4.1.2.3 Optimal model selection.....	93
4.1.2.4 Exploratory data analysis (EDA).....	94
4.1.2.4.1 Principal component analysis (PCA)	94
4.1.2.5 Multivariate data analysis (MDA): Optimal model development	96
4.1.2.5.1 Linear discriminant analysis (LDA) [mahalanobis distance)	96
4.1.2.5.2 Partial least squares discriminant analysis (PLS-DA)	98
4.1.2.6 Optimal model selection.....	100
4.1.2.7 Conclusion	101
4.1.3 Species determination irrespective of frozen-period.....	101
4.1.3.1 Exploratory data analysis (EDA).....	101
4.1.3.1.1 Principal component analysis (PCA)	101
4.1.3.1.2 Difference spectra	104
4.1.3.2 Multivariate data analysis (MDA): Model development	107
4.1.3.2.1 Soft independent modelling of class analogy (SIMCA).....	107
4.1.3.2.2 K-nearest neighbour (KNN)	110
4.1.3.2.3 Discriminant analysis (DA)	114
4.1.3.2.4 Partial least squares discriminant analysis (PLS-DA)	115
4.1.3.3 Optimal model selection.....	118
4.1.3.4 Exploratory data analysis (EDA).....	119
4.1.3.4.1 Principal component analysis (PCA)	119
4.1.3.5 Multivariate data analysis (MDA): Optimal model development	119
4.1.3.5.1 Quadratic discriminant analysis (QDA).....	119
4.1.3.5.2 Partial least squares discriminant analysis (PLS-DA)	122
4.1.3.6 Optimal model selection.....	124
4.1.3.7 Conclusion	124
4.2 Fresh vs. previously frozen meat determination.....	125
4.2.1 Spectral analysis	125
4.2.2 Exploratory data analysis (EDA).....	126
4.2.2.1 Principal component analysis (PCA)	126
4.2.2.2 Difference spectra	129
4.2.3 Fresh vs. previously frozen meat determination irrespective of muscle type	130
4.2.3.1 Multivariate data analysis (MDA): Model development	130
4.2.3.1.1 Soft independent modelling of class analogy (SIMCA).....	130
4.2.3.1.2 K-nearest neighbour (KNN)	133
4.2.3.1.3 Discriminant analysis (DA)	137
4.2.3.1.4 Partial least squares discriminant analysis (PLS-DA)	140

4.2.3.2 Optimal model selection.....	144
4.2.3.3 Exploratory data analysis (EDA).....	145
4.2.3.3.1 Principal component analysis (PCA)	145
4.2.3.4 Multivariate data analysis (MDA): Optimal model development	147
4.2.3.4.1 Discriminant analysis (DA)	147
4.2.3.4.2 Partial least squares discriminant analysis (PLS-DA)	151
4.2.3.5 Optimal model selection.....	155
4.2.3.6 Conclusion	157
4.2.4 Fresh vs. previously frozen meat determination as well as frozen period.....	157
4.2.4.1 Multivariate data analysis (MDA): Model development	157
4.2.4.1.1 Soft independent modelling of class analogy (SIMCA).....	157
4.2.4.1.2 K-nearest neighbour (KNN)	161
4.2.4.1.3 Discriminant analysis (DA)	166
4.2.4.1.4 Partial least squares discriminant analysis (PLS-DA)	170
4.2.4.2. Optimal model selection	175
4.2.4.3 Exploratory data analysis (EDA).....	176
4.2.4.3.1 Principal component analysis (PCA)	176
4.2.4.4 Multivariate data analysis (MDA): Optimal model development	176
4.2.4.4.1 Quadratic discriminant analysis (QDA).....	176
4.2.4.4.1.1 Discrimination of fresh vs. frozen period	176
4.2.4.4.1.2 Discrimination of fresh vs. previously frozen, irrespective of frozen period	177
4.2.4.4.2 Partial least squares discriminant analysis (PLS-DA)	181
4.2.4.4.2.1 Discrimination of fresh vs. frozen period	181
4.2.4.4.2.2 Discrimination of fresh vs. previously frozen, irrespective of frozen period	182
4.2.4.5 Optimal model selection.....	186
4.2.4.6 Conclusion	189
4.3 Muscle type determination	189
4.3.1 Spectral analysis	189
4.3.2 Exploratory data analysis (EDA).....	191
4.3.2.1 Principal component analysis (PCA)	191
4.3.2.2 Difference spectra	194
4.3.3 Multivariate data analysis (MDA): Model development	197
4.3.3.1 Soft independent modelling of class analogy (SIMCA).....	197
4.3.3.2 K-nearest neighbour (KNN).....	200
4.3.3.3 Discriminant analysis (DA)	204
4.3.3.4 Partial least squares discriminant analysis (PLS-DA)	207
4.3.4 Optimal model selection.....	212

4.3.5 Exploratory data analysis (EDA).....	212
4.3.5.1 Principal component analysis (PCA)	212
4.3.6 Multivariate data analysis (MDA): Optimal model development	214
4.3.6.1 Discriminant analysis (DA)	214
4.3.6.2 Partial least squares discriminant analysis (PLS-DA)	217
4.3.7 Optimal model selection.....	220
4.3.8 Two-group discrimination model evaluation	221
4.3.9 Conclusion	222
4.4 Hierarchical model development	223
4.5 References	228
Chapter 5 General Discussion and Conclusion	232
5.1 References	238
Addendum A Supplementary information pertaining to Chapter 4:	241
4.1 Species determination	241
Addendum B Supplementary information pertaining to Chapter 4:.....	255
4.2 Fresh vs. previously frozen meat determination.....	255
Addendum C Supplementary information pertaining to Chapter 4:.....	288
4.3 Muscle type determination	288
Addendum D Supplementary information pertaining to Chapter 4:	300
4.4 Hierarchical model development	300

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

List of Figures

Figure 2.1 Diagram representing potential meat fraud problems [Adapted from (Ballin, 2010)].....	17
Figure 2.2 Schematic representation of the interaction of radiation and matter (Anonymous, 2017).....	25
Figure 2.3 (a) Near infrared reflectance, and (b) near infrared transmittance basic instrument configurations (Siesler, 2008a).	26
Figure 2.4 Typical setup of the VIS-NIR Flame spectrophotometer (Ocean Optics, Inc., Florida, USA).....	26
Figure 2.5 Block diagram of an interferometer and related electronics, generally used in a FT-NIR instrument (Nielsen, 2010).....	28
Figure 2.6 (a) Buchi NIRFlex N-500 (benchtop FT-NIR spectrometer) (NIRSolutionsTM, Flawil, Switzerland), and (b) MicroNIR OnSite spectrophotometer (portable handheld NIR spectrometer) (Viavi Solutions Inc., Milpitas, USA).	29
Figure 2.7 A data matrix of size I^*K	30
Figure 2.8 A data matrix sized (I^*K) is condensed to smaller sized matrices (I^*A) and (A^*K) that are easier to interpret and comprehend in addition to containing all the relevant information. Noise and other disturbances are left in the residual matrix of size (I^*K) . [Adapted from (Esbensen et al., 2002; Geladi, 2003)].	33
Figure 2.9 Principle of the SIMCA method. Each symbol represents a spectrum. Three classes are independently modelled by principal component analysis (Sirven et al., 2007).	35
Figure 2.10 Schematic representation of (a) Euclidean distance and (b) Manhattan distance (Brereton, 2003b).....	36
Figure 2.11 Example of a confusion matrix where TP is the true positive, TN the true negative, FP the false positive and FN the false negative.	37
Figure 2.12 The simulation of the k-nearest neighbour method, $k = 5$. Each class is represented by a symbol and independently modelled by principal component analysis (Zhang, 2017).	37
Figure 2.13 Diagrammatic depiction of the PLS-DA method. Y , a dummy matrix consisting of ones and zeros (representing class membership), is paired with X to obtain b , the regression coefficients and the residual, F	38
Figure 2.14 An example of SVM and linear separating hyperplanes. The support vectors, marked with grey squares, define the margin of largest separation between the two classes (Cortes & Vapnik, 1995).	39
Figure 2.15 An RBF kernel using diverse hyper-parameters in order to create unique SVM solutions (Mueller & Massaron, 2016).	40
Figure 2.16 (a) A sigmoid kernel, and (b) a polynomial kernel applied to the same data (Mueller & Massaron, 2016).....	40
Figure 3.1 Schematic illustrating the randomisation process of the LTL and FF muscle samples. Each muscle was cut into nine steaks, with each steak representing an anatomical position (P1 – P9). The anatomical position of the steaks was numbered from the dorsal (top) side of the muscle. Each piece of steak was randomly assigned to a month (m1 – m9), therefore ensuring that the same anatomical position is not scanned on the same month, for each muscle sample. 63	
Figure 3.2 Experimental set-up, blooming of samples and spectral collection of game meat steaks.....	64
Figure 3.3 Schematic of the spectral acquisition process.	65

Figure 3.4 Schematic of the spectral data analysis process. Firstly, the smoothed data (pre-processed with SNV + detrend) was subjected to various unsupervised techniques in order to identify the optimal supervised method. After the optimal method was identified, the smoothed data was pre-processed using various techniques and models were developed using the previously identified optimal methods. This was done to see whether or not the method could be optimised. Thereafter the models were compared and an optimal model was selected.....	66
Figure 4.1 Unprocessed mean spectra for ostrich (blue), springbok (red) and zebra (green), irrespective of the muscle type.	73
Figure 4.2 Unprocessed mean spectra for black wildebeest (blue), springbok (red), zebra (green) and ostrich (light blue), irrespective of the frozen period.....	74
Figure 4.3 PCA analysis (SNV + detrend pre-processed) of species for ostrich (blue), springbok (red) and zebra (green) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC1 (86%) vs. PC2 (6%); and (b) PCA score plot of PC2 (6%) vs. PC3 (3%).	75
Figure 4.4 PCA analysis (SNV + detrend pre-processed) of species for ostrich (blue), springbok (red) and zebra (green) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC3 (3%) vs. PC4 (2%); and (b) PCA score plot of PC4 (2%) vs. PC5 (1%).	75
Figure 4.5 (a) PCA loadings line plot and (b) correlation loadings for PC1 (86%) with interpretable bands at 982, 1093, 1180, 1329, 1422 and 1589 nm.....	76
Figure 4.6 (a) PCA loadings line plot and (b) correlation loadings for PC2 (6%) with bands at 957, 1106, 1193, 1298, 1440 and 1595 nm.....	77
Figure 4.7 (a) PCA loadings line plot and (b) correlation loadings for PC3 (3%) with interpretable bands at 1143, 1378 and 1515 nm.....	77
Figure 4.8 (a) PCA loadings line plot and (b) correlation loadings for PC4 (2%) with an interpretable band at 1645 nm.....	77
Figure 4.9 (a) PCA loadings line plot and (b) correlation loadings for PC5 (1%) with an interpretable band at 1031 nm.....	78
Figure 4.10 Difference spectra of zebra and the remaining species, irrespective of the muscle type, with absorption bands at 1075, 1124, 1285, 1298, 1366, 1409, 1465 nm.....	81
Figure 4.11 Difference spectra of springbok and the remaining species, irrespective of the muscle type, with absorption bands at 939, 1075, 1118, 1162, 1205, 1279, 1341, 1403, 1465, 1527, 1645 nm.....	81
Figure 4.12 Difference spectra of ostrich and the remaining species, irrespective of the muscle type, with absorption bands at 957, 994, 1155, 1366, 1496, 1645nm.	82
Figure 4.13 SIMCA classification (SNV + detrend pre-processed) resulting in a 71.7% and 68.5% classification accuracy for zebra and springbok, respectively. SIMCA classification score plot of ostrich (red), springbok (green) and zebra (blue), illustrating the predicted objects. (a) Score plot for predicted zebra (top) vs. remaining species (bottom) and (b) score plot of predicted springbok (top) vs. remaining species (bottom).	84

Figure 4.14 SIMCA classification (SNV + detrend pre-processed) resulting in a 94.1% classification accuracy for ostrich. SIMCA classification score plot of ostrich (red), springbok (green) and zebra (blue), illustrating the predicted objects, where the top row indicates the objects predicted as ostrich and the bottom row the objects predicted as the remaining species.....	84
Figure 4.15 KNN ($k = 5$) classification (SNV + detrend pre-processed) resulting in an overall classification accuracy of 78.5%. KNN classification score plot of ostrich (red), springbok (green) and zebra (blue), illustrating the predicted objects. (a) Score plot of predicted ostrich (top) vs. remaining species (bottom), (b) score plot of predicted springbok (top) vs. remaining species (bottom) and (c) score plot of predicted zebra (top) vs. remaining species (bottom).	86
Figure 4.16 PCA analysis (SNV + detrend pre-processed) of species for ostrich (blue), springbok (red) and zebra (green) classes. Minimal class separation was observed. Scores illustrated as KS-calibration PCA score plot of PC2 (7%) vs. PC3 (4%).	86
Figure 4.17 PLS-DA scores plot (SNV + detrend pre-processed) of ostrich (red), springbok (green) and zebra (blue). (a) 3D scores plot of LV1 (84.94%) vs. LV2 (6.06%) vs. LV3 (1.32%) and (b) score plot of LV2 (6.06%) vs. LV3 (1.32%), colour coded per class species.	91
Figure 4.18 PLS-DA (8 LVs) (SNV + detrend pre-processed) resulting in an overall classification accuracy of 95.7%. PLS-DA prediction score plot of ostrich (red), springbok (green) and zebra (blue), illustrating the predicted objects. (a) Score plot of objects predicted as ostrich [above red line (Y1)] vs. remaining species [below red line (Y1)], (b) score plot of objects predicted as springbok [above red line (Y2)] vs. remaining species [below red line (Y2)] and (c) score plot of objects predicted as zebra [above red line (Y3)] vs. remaining species [below red line (Y3)].	92
Figure 4.19 PCA analysis (SNV + detrend pre-processed) of species for black wildebeest (blue), zebra (green), springbok (light blue) and ostrich (red) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC1 (86%) vs. PC2 (6%). (b) PCA loadings line plot and (c) correlation loadings for PC1 with interpretable bands at 982, 1093, 1174, 1323, 1422 and 1601 nm.....	102
Figure 4.20 PCA analysis (SNV + detrend pre-processed) of species black wildebeest (blue), zebra (green), springbok (light blue) and ostrich (red) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (6%) vs. PC3 (4%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 1515 and 1645 nm.....	102
Figure 4.21 PCA analysis (SNV + detrend pre-processed) of species black wildebeest (blue), zebra (green), springbok (light blue) and ostrich (red) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC3 (4%) vs. PC4 (3%); and (b) PCA score plot of PC4 (3%) vs. PC5 (1%).	103
Figure 4.22 (a) PCA loadings line plot and (b) correlation loadings for PC3 (4%) with interpretable bands at 1273 and 1372 nm.....	104
Figure 4.23 Difference spectra of black wildebeest and the remaining species, irrespective of frozen period, with absorption bands at 939, 1062, 1118, 1298, 1366 nm.	106
Figure 4.24 Difference spectra of zebra and the remaining species, irrespective of frozen period, with absorption bands at 945, 976, 1149, 1162, 1316, 1341, 1366, 1403, 1645 nm.....	106
Figure 4.25 Difference spectra of springbok and the remaining species, irrespective of frozen period, with absorption bands at 932, 982, 1062, 1118, 1168, 1186, 1341, 1403, 1645 nm.....	107

Figure 4.26 Difference spectra of ostrich and the remaining species, irrespective of frozen period, with absorption bands at 970, 976, 1118, 1149, 1162, 1310, 1341, 1366, 1397, 1403 nm.....	107
Figure 4.27 SIMCA classification (SNV + detrend pre-processed) resulting in a 65.5% and 60.4% classification accuracy for black wildebeest and zebra, respectively. SIMCA classification score plot of black wildebeest (red), ostrich (green), springbok (blue) and zebra (turquoise), illustrating the predicted objects. (a) Score plot for predicted black wildebeest (top) vs. remaining species (bottom) and (b) score plot of predicted zebra (top) vs. remaining species (bottom).	109
Figure 4.28 SIMCA classification (SNV + detrend pre-processed) resulting in a 62.1% and 61.9% classification accuracy for springbok and ostrich, respectively. SIMCA classification score plot of black wildebeest (red), ostrich (green), springbok (blue) and zebra (turquoise), illustrating the predicted objects. (a) Score plot for predicted springbok (top) vs. remaining species (bottom) and (b) score plot of predicted ostrich (top) vs. remaining species (bottom).	109
Figure 4.29 KNN ($k = 5$) classification (SNV + detrend pre-processed) resulting in an overall classification accuracy of 73.6%. KNN classification score plot of black wildebeest (red), ostrich (green), springbok (blue) and zebra (turquoise), illustrating the predicted objects. (a) Score plot for predicted black wildebeest (top) vs. remaining species (bottom), (b) score plot of predicted zebra (top) vs. remaining species (bottom), (c) score plot of predicted springbok (top) vs. remaining species (bottom) and (d) score plot of predicted ostrich (top) vs. remaining species (bottom).	111
Figure 4.30 PCA analysis (SNV + detrend pre-processed) of species for black wildebeest (blue), ostrich (red), zebra (green) and springbok (light blue) classes. Minimal class separation was observed. Scores illustrated as KS-calibration PCA score plot of PC2 (6%) vs. PC3 (4%).	111
Figure 4.31 PLS-DA scores plot (SNV + detrend pre-processed) of black wildebeest (red), ostrich (green), springbok (blue) and zebra (turquoise). (a) 3D score plot of LV1 (84.72%) vs. LV2 (5.74%) vs. LV3 (3.36%), and (b) score plot of LV1 (84.72%) vs. LV3 (3.36%), colour coded per class species.	116
Figure 4.32 PLS-DA (6 LVs) (SNV + detrend pre-processed) resulting in an overall classification accuracy of 78.9%. PLS-DA prediction score plot of black wildebeest (red), ostrich (green), springbok (blue) and zebra (turquoise), illustrating the predicted objects. (a) Score plot of objects predicted as black wildebeest [above red line (Y1)] vs. remaining species [below red line (Y1)], (b) score plot of objects predicted as ostrich [above red line (Y2)] vs. remaining species [below red line (Y2)], (c) score plot of objects predicted as zebra [above red line (Y4)] vs. remaining species [below red line (Y4)] and (d) score plot of objects predicted as springbok [above red line (Y3)] vs. remaining species [below red line (Y3)].	117
Figure 4.33 Standard normal variate (SNV) pre-processed mean spectra for fresh (blue) and frozen-thawed (red) black wildebeest, zebra, springbok and ostrich.	125
Figure 4.34 PCA analysis (SNV + detrend pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (88%) vs. PC2 (6%). (b) PCA loadings line plot and (c) correlation loadings for PC1 with interpretable bands at 982, 1093, 1180, 1329, 1422 and 1589 nm.....	126
Figure 4.35 PCA analysis (SNV + detrend pre-processed) of zebra (LTL) [frozen up to 9 months] illustrating a slight overlap between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (78%) vs. PC2 (10%). (b) PCA loadings line plot and (c) correlation loadings for PC1 with interpretable bands at 982, 1093, 1180, 1329, 1422 and 1589 nm.....	127

- Figure 4.36** PCA analysis (SNV + detrend pre-processed) of fresh vs frozen period [frozen up to 9 months] for all four species illustrating good separation of the fresh samples with an overlap between frozen period classes. Scores illustrated as (a) PCA score plot of PC1 (88%) vs. PC2 (7%) for black wildebeest (LTL), (b) PCA score plot of PC1 (78%) vs. PC2 (10%) for zebra (LTL), (c) PCA score plot of PC1 (83%) vs. PC2 (7%) for springbok (LTL) and (d) PCA score plot of PC1 (81%) vs. PC2 (10%) for ostrich (FF)..... 128
- Figure 4.37** (a) Difference spectra of zebra (LTL) fresh and frozen-thawed samples, irrespective of the frozen period (1 – 9 months). (b) Difference spectra of zebra fresh and frozen-thawed samples, irrespective of the muscle type (LTL, BF, SM, ST, IS, SS, fillet)..... 130
- Figure 4.38** PCA analysis (SNV + detrend pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating good separation between fresh (**blue**) and frozen-thawed (**red**) classes. Scores illustrated as KS-calibration PCA score plot of PC1 (89%) vs. PC2 (5%). 134
- Figure 4.39** PCA analysis (SNV + detrend pre-processed) of (a) zebra (LTL, BF, SM, ST, IS, SS, fillet) and (b) ostrich (BD, FF) [frozen up to 1 month] illustrating minimal separation between fresh (**blue**) and frozen-thawed (**red**) classes. Scores illustrated as (a) KS-calibration PCA score plot of PC1 (87%) vs. PC2 (6%) for zebra, and (b) KS-calibration PCA score plot of PC1 (72%) vs. PC2 (18%) for ostrich. 135
- Figure 4.40** PLS-DA scores plot (SNV + detrend pre-processed) of fresh (**red**) and frozen-thawed (**green**) meat samples. (a) score plot of LV1 (87.58%) vs. LV2 (5.47%) for zebra and (b) score plot of LV1 (89.09%) vs. LV2 (2.85%) for springbok, colour coded per treatment class. 141
- Figure 4.41** PLS-DA models (SNV + detrend pre-processed) indicating satisfactory overall classification accuracies. PLS-DA prediction score plot of (a) zebra, (b) springbok and (c) ostrich, illustrating the predicted fresh (**red**) and frozen-thawed (**green**) objects. (a) Score plot (3 LVs) for zebra (95.1%) of objects predicted as fresh [above red line (Y1)] vs. frozen-thawed [below red line (Y1)], (b) score plot (2 LVs) for springbok (100%) of objects predicted as fresh [above red line (Y1)] vs. frozen-thawed [below red line (Y1)] and (c) score plot (1 LV) for ostrich (90%) of objects predicted as fresh [above red line (Y1)] vs. frozen-thawed [below red line (Y1)]. 141
- Figure 4.42** PCA analysis (SNV + detrend pre-processed) of black wildebeest (LTL) [frozen up to 9 months] illustrating good separation between fresh (**blue**) and frozen-thawed (**red**) classes. Scores illustrated as KS-calibration PCA score plot of PC1 (86%) vs. PC2 (8%). 162
- Figure 4.43** PCA analysis (SNV + detrend pre-processed) of springbok (LTL) [frozen up to 9 months] illustrating good separation with a slight overlap between fresh (**blue**) and frozen-thawed (**red**) classes. Scores illustrated as KS-calibration PCA score plot of PC1 (82%) vs. PC2 (7%). 163
- Figure 4.44** PCA analysis (SNV + detrend pre-processed) of (a) zebra (LTL) and (b) ostrich (FF) [frozen up to 9 months] illustrating minimal separation between fresh (**blue**) and frozen-thawed (**red**) classes. Scores illustrated as (a) KS-calibration PCA score plot of PC1 (79%) vs. PC2 (10%) for zebra, and (b) KS-calibration PCA score plot of PC1 (79%) vs. PC2 (11%) for ostrich. 164
- Figure 4.45** PLS-DA score plots (SNV + detrend pre-processed) of fresh (**red/purple**) and frozen-thawed (**green/peach**) meat samples. (a) score plot of LV1 (86.26%) vs. LV2 (7.89%) for black wildebeest, (b) score plot of LV1 (78.63%) vs. LV2 (8.61%) for zebra, (c) score plot of LV1 (82.28%) vs. LV2 (6.07%) for springbok and (d) score plot of LV1 (78.49%) vs. LV2 (9.44%) for ostrich, colour coded per treatment class. 172
- Figure 4.46** PLS-DA models (SNV + detrend pre-processed) resulting in satisfactory overall classification accuracies. PLS-DA prediction score plot of (a) black wildebeest (b) zebra, (c) springbok and (d) ostrich, illustrating the predicted fresh (**red/purple**) and frozen-thawed (**green/peach**) objects. (a) Score plot (4 LVs)

for black wildebeest (96%) of objects predicted as fresh [above red line (Y1)] vs. frozen-thawed [below red line (Y1)], (b) score plot (2 LVs) for zebra (88.9%) of objects predicted as fresh [above red line (Y1)] vs. frozen-thawed [below red line (Y1)], (c) score plot (4 LVs) for springbok (97.5%) of objects predicted as fresh [above red line (Y1)] vs. frozen-thawed [below red line (Y1)] and (d) score plot (3 LVs) for ostrich (91.7%) of objects predicted as fresh [above red line (Y1)] vs. frozen-thawed [below red line (Y1)]. 173

Figure 4.47 Standard normal variate (SNV) pre-processed mean spectra for various zebra muscle types [(BF) biceps femoris (**blue**), (Fillet) psoas major (**red**), (IS) infraspinatus (**green**), (LTL) longissimus thoracis et lumborum (**light blue**), (SM) semimembranosus (**maroon**), (SS) supraspinatus (**grey**), (ST) semitendinosus (**purple**)], irrespective of the meat being fresh or previously frozen..... 190

Figure 4.48 Standard normal variate (SNV) pre-processed mean spectra for various springbok muscle types [(BF) *biceps femoris* (**blue**), (*Fillet*) *psoas major* (**red**), (IS) *infraspinatus* (**green**), (LTL) *longissimus thoracis et lumborum* (**light blue**), (SM) *semimembranosus* (**maroon**), (SS) *supraspinatus* (**grey**), (ST) *semitendinosus* (**purple**)], irrespective of the meat being fresh or previously frozen..... 190

Figure 4.49 Standard normal variate (SNV) pre-processed mean spectra for various ostrich muscle types [(BD) *gastrocnemius* (**blue**) and (FF) *iliofibularis* (**red**)], irrespective of the meat being fresh or previously frozen. 191

Figure 4.50 PCA analysis (SNV + detrend pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating minimal separation between classes. Scores illustrated as (a) PCA score plot of PC1 (88%) vs. PC2 (6%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with absorption bands at 963, 1075, 1162, 1366, 1453 and 1589 nm. 192

Figure 4.51 PCA analysis (SNV + detrend pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [fresh or previously frozen samples] illustrating minimal separation between classes. Scores illustrated as (a) PCA score plot of PC1 (90%) vs. PC2 (4%). (b) PCA loadings line plot and (c) correlation loadings for PC1 with interpretable bands at 1484 and 1639 nm..... 193

Figure 4.52 PCA analysis (SNV + detrend pre-processed) of ostrich (BD, FF) [fresh and previously frozen samples] illustrating a slight overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (76%) vs. PC2 (16%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 1279 and 1372 nm..... 193

Figure 4.53 Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 smoothing points) treated mean spectra of zebra illustrating the spectral differences between the six muscle types [(BF) biceps femoris (**blue**), (Fillet) psoas major (**red**), (IS) infraspinatus (**green**), (LTL) longissimus thoracis et lumborum (**light blue**), (SM) semimembranosus (**maroon**), (SS) supraspinatus (**grey**), (ST) semitendinosus (**purple**)], with absorption bands at 976, 1162, 1335, 1403 and 1626 nm. 196

Figure 4.54 Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 smoothing points) treated mean spectra of springbok illustrating the spectral differences between the six muscle types [(BF) *biceps femoris* (**blue**), (*Fillet*) *psoas major* (**red**), (IS) *infraspinatus* (**green**), (LTL) *longissimus thoracis et lumborum* (**light blue**), (SM) *semimembranosus* (**maroon**), (SS) *supraspinatus* (**grey**), (ST) *semitendinosus* (**purple**)], with absorption bands at 976, 1162, 1335, 1403 and 1626 nm. 196

Figure 4.55 Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 smoothing points) treated mean spectra of ostrich illustrating the spectral differences between the two muscle types [(BD) *gastrocnemius* (**blue**) and (FF) *iliofibularis* (**red**)], with absorption bands at 976, 1162, 1341, 1403 and 1639 nm..... 197

Figure 4.56 PCA analysis (SNV + detrend pre-processed) of ostrich muscles [fresh and previously frozen samples] illustrating a slight overlap between BD (**blue**) and FF (**red**) classes. Scores illustrated as KS-calibration PCA score plot of PC1 (75%) vs. PC2 (16%). 201

Figure 4.57 PCA analysis (SNV + detrend pre-processed) of (a) zebra (LTL, BF, SM, ST, IS, SS, fillet) and (b) springbok (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating minimal separation between the different muscle classes. Scores illustrated as (a) KS-calibration PCA score plot of PC1 (88%) vs. PC2 (6%) for zebra, and (b) KS-calibration PCA score plot of PC1 (90%) vs. PC2 (4%) for ostrich. 202

Figure 4.58 PLS-DA (SNV + detrend pre-processed) (a) score plot of LV1 (86.8%) vs. LV2 (6.84%) vs. LV3 (2.87%) for zebra, (b) score plot of LV1 (89.09%) vs. LV2 (4.45%) vs. LV3 (3.06%) for springbok, (c) score plot of LV1 (73.49%) vs. LV2 (18.01%) vs. LV3 (4.6%) for ostrich, colour coded per muscle type class. 208

Figure 4.59 PLS-DA models (SNV + detrend pre-processed) for muscle discrimination, resulting in satisfactory individual classification accuracies. PLS-DA prediction score plot of (a) zebra (b) springbok and (c) ostrich, illustrating the predicted objects. (a) Score plot (6 LVs) for zebra (91.5%) of objects predicted as BF [above red line (Y1)] vs. remaining muscles [below red line (Y1)], (b) score plot (4 LVs) for springbok (79.7%) of objects predicted as Fillet [above red line (Y2)] vs. remaining muscles [below red line (Y2)], (c) score plot (5 LVs) for ostrich (100%) of objects predicted as BD [above red line (Y1)] vs. FF [below red line (Y1)]. 210

List of Tables

Table 2.1 Generally cited, high-profile cases of fraud that have affected global food safety	14
Table 2.2 Selected applications of conventional analytical detection methods used for meat authentication.....	18
Table 2.3 A summary of selected applications of NIR spectroscopy and chemometric techniques for the authentication and classification of meat and meat products.....	44
Table 3. 1 General information regarding the species, origin and muscle type used for this study.	62
Table 4.1 SIMCA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for species classification.	82
Table 4.2 The performance measures used to assess the SIMCA classification models of the three species, pre-processed with SNV + detrend.	84
Table 4.3 KNN model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for species classification.	85
Table 4.4 The performance measures used to assess the KNN classification models of the three species, pre-processed with SNV + detrend.	88
Table 4.5 DA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for species discrimination.	89
Table 4.6 The performance measures used to assess the LDA [mahalanobis distance] models of the three species, pre-processed with SNV + detrend.	90
Table 4.7 PLS-DA model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for species discrimination.	91
Table 4.8 The performance measures used to assess the PLS-DA models of the three species, pre-processed with SNV + detrend.	93
Table 4.9 An overview of the accuracies for the various classification and discrimination models, pre-processed with SNV + detrend to distinguish between species.....	93
Table 4.10 An overview of the accuracies of the LDA [mahalanobis distance] models with various pre-processing techniques applied for species discrimination.....	96
Table 4.11 LDA [mahalanobis distance] model calibration and validation results to assess the overall performance of the SNV+detrend+SGd ₂ (9) corrected data for species discrimination.....	97
Table 4.12 LDA [mahalanobis distance] model calibration and validation results to assess the overall performance of the SGd ₁ (7) corrected data for species discrimination.	97
Table 4.13 The performance measures used to assess the LDA [mahalanobis distance] models of the three species, pre-processed with SGd ₁ (7).	98
Table 4.14 An overview of the accuracies of the PLS-DA models with various pre-processing techniques applied for species discrimination.....	98

Table 4.15 The performance measures used to assess the PLS-DA models of the three species, pre-processed with SGd ₁ (7)	100
Table 4.16 An overview of the accuracies of the LDA [mahalanobis distance] and PLS-DA models with various pre-processing techniques applied to distinguish between species	100
Table 4.17 SIMCA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for species classification	108
Table 4.18 The performance measures used to assess the SIMCA classification models of the four species, pre-processed with SNV + detrend	109
Table 4.19 KNN model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for species classification	110
Table 4.20 The performance measures used to assess the KNN classification models of the four species, pre-processed with SNV + detrend	113
Table 4.21 DA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for species discrimination	114
Table 4.22 The performance measures used to assess the QDA models of the four species, pre-processed with SNV + detrend	115
Table 4.23 PLS-DA model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for species discrimination	116
Table 4.24 The performance measures used to assess the PLS-DA models of the four species, pre-processed with SNV + detrend	118
Table 4.25 An overview of the accuracies for the various classification and discrimination models, pre-processed with SNV + detrend to distinguish between species.....	118
Table 4.26 An overview of the classification rates of the QDA models with various pre-processing techniques applied for species discrimination.....	120
Table 4.27 QDA model calibration and validation results to assess the overall performance of the SNV+SGd ₂ (7) corrected for species discrimination	121
Table 4.28 QDA model calibration and validation results to assess the overall performance of the SGd ₁ (7) corrected data for species discrimination	121
Table 4.29 The performance measures used to assess the QDA models of the four species, pre-processed with SGd ₁ (7).....	121
Table 4.30 An overview of the accuracies of the PLS-DA models with various pre-processing techniques applied for species discrimination.....	122
Table 4.31 The performance measures used to assess the PLS-DA models of the four species, pre-processed with SGd ₂ (9).	123
Table 4.32 An overview of the accuracies of the QDA and PLS-DA models with various pre-processing techniques applied to distinguish between species.....	124

Table 4.33 SIMCA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for fresh or previously frozen meat classification.	131
Table 4.34 The performance measures used to assess the SIMCA classification models for fresh vs. frozen-thawed meat of the three species, pre-processed with SNV + detrend.	132
Table 4.35 KNN model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for fresh or previously frozen meat classification.	133
Table 4.36 The performance measures used to assess the KNN classification models for fresh vs. frozen-thawed meat of the three species, pre-processed with SNV + detrend.	136
Table 4.37 Optimal DA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for fresh or previously frozen meat discrimination.	137
Table 4.38 The performance measures used to assess the DA models for fresh vs. frozen-thawed meat of the three species, pre-processed with SNV + detrend.	139
Table 4.39 PLS-DA model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for fresh or previously frozen meat discrimination.	140
Table 4.40 The performance measures used to assess the PLS-DA models for fresh vs. frozen-thawed meat of the three species, pre-processed with SNV + detrend.	143
Table 4.41 An overview of the accuracies for the various classification and discrimination models, pre-processed with SNV + detrend to distinguish between fresh or previously frozen meat, irrespective of the muscle type.	144
Table 4.42 An overview of the classification accuracies of the DA models with various pre-processing techniques applied for fresh or previously frozen meat discrimination.	149
Table 4.43 Springbok LDA model calibration and validation results to assess the overall performance of the SNV + SGd ₂ (7/9) and SNV + detrend + SGd ₂ (7/9) corrected data for fresh or previously frozen meat discrimination.	150
Table 4.44 Zebra QDA model calibration and validation results to assess the overall performance of the SNV + SGd ₂ (7/9) and SGd ₂ (7/9) corrected data for fresh or previously frozen meat discrimination.	150
Table 4.45 The performance measures used to assess the optimal QDA models (6 PCs) for fresh vs. frozen-thawed zebra meat.	150
Table 4.46 Ostrich QDA model calibration and validation results to assess the overall performance of the SNV + detrend + SGd ₂ (7) and SGd ₁ (5) corrected data for fresh or previously frozen meat discrimination.	151
Table 4.47 The performance measures used to assess the optimal QDA models (6 PCs) for fresh vs. frozen-thawed ostrich meat.	151
Table 4.48 An overview of the classification accuracies of the PLS-DA models with various pre-processing techniques applied for fresh vs. frozen-thawed meat.	154
Table 4.49 The performance measures used to assess the optimal PLS-DA (5LVs) model for fresh vs. frozen-thawed zebra meat.	155

Table 4.50 The performance measures used to assess the optimal PLS-DA (3LVs) model for fresh vs. frozen-thawed ostrich meat	155
Table 4.51 An overview of the classification accuracies of the DA and PLS-DA models with various pre-processing techniques applied for fresh or previously frozen meat discrimination.....	156
Table 4.52 SIMCA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen period classification.....	157
Table 4.53 SIMCA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen-thawed meat classification, irrespective of frozen period.....	158
Table 4.54 The performance measures used to assess the SIMCA classification models for fresh vs. frozen-thawed meat of the four species, pre-processed with SNV + detrend.	160
Table 4.55 Optimal KNN model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen period classification.	161
Table 4.56 Optimal KNN model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen-thawed meat classification, irrespective of frozen period.	162
Table 4.57 The performance measures used to assess the KNN classification models for fresh vs. frozen-thawed meat of the four species, pre-processed with SNV + detrend.	165
Table 4.58 Optimal DA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen period classification.....	166
Table 4.59 Optimal DA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen-thawed meat classification, irrespective of frozen period.	167
Table 4.60 The performance measures used to assess the DA models for fresh vs. frozen-thawed meat of the four species, pre-processed with SNV + detrend.....	169
Table 4.61 PLS-DA model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen period classification.	170
Table 4.62 PLS-DA model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen-thawed meat classification, irrespective of frozen period.	171
Table 4.63 The performance measures used to assess the PLS-DA models for fresh vs. frozen-thawed meat of the four species, pre-processed with SNV + detrend.....	174
Table 4.64 An overview of the accuracies for the various classification and discrimination models, pre-processed with SNV + detrend to distinguish between fresh or previously frozen meat, irrespective of the frozen period, as well as determine the frozen period.	175
Table 4.65 An overview of the classification accuracies of the QDA models (6 PCs) for fresh vs. frozen period with various pre-processing techniques applied.	179

Table 4.66 An overview of the classification rates of the QDA models (6 PCs) for fresh vs. previously frozen meat, irrespective of the frozen period, with various pre-processing techniques applied.....	180
Table 4.67 An overview of the classification accuracies of the PLS-DA models for fresh vs. frozen period with various pre-processing techniques applied.....	184
Table 4.68 An overview of the classification accuracies of the PLS-DA models for fresh vs. previously frozen meat, irrespective of the frozen period, with various pre-processing techniques applied.....	185
Table 4.69 An overview of the classification rates of the DA and PLS-DA models for fresh vs. frozen period with various pre-processing techniques applied.....	187
Table 4.70 An overview of the classification rates of the DA and PLS-DA models for fresh vs. previously frozen meat, irrespective of the frozen period, with various pre-processing techniques applied.....	188
Table 4.71 SIMCA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for muscle type classification.....	197
Table 4.72 The performance measures used to assess the SIMCA classification models for different muscle types of the three species, pre-processed with SNV + detrend.....	199
Table 4.73 KNN model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for muscle type classification.....	200
Table 4.74 The performance measures used to assess the KNN classification models for different muscle types of the three species, pre-processed with SNV + detrend.....	203
Table 4.75 Optimal DA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for muscle type discrimination.....	204
Table 4.76 The performance measures used to assess the DA models for different muscle types of the three species, pre-processed with SNV + detrend	206
Table 4.77 PLS-DA model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for muscle type discrimination	207
Table 4.78 The performance measures used to assess the PLS-DA models for different muscle types of the three species, pre-processed with SNV + detrend	211
Table 4.79 An overview of the accuracies for the various classification and discrimination models, pre-processed with SNV + detrend to distinguish between the different muscle types.....	212
Table 4.80 An overview of the accuracies of the DA models with various pre-processing techniques applied for muscle type discrimination.....	216
Table 4.81 An overview of the classification accuracies of the PLS-DA models with various pre-processing techniques applied for muscle type discrimination	219
Table 4.82 An overview of the classification accuracies of the DA and PLS-DA models with various pre-processing techniques applied to distinguish between the muscle types.....	221
Table 4.83 An overview of the classification rates of the two-group distribution QDA and PLS-DA models for zebra and springbok with various pre-processing techniques applied.....	221

Table 4.84 An overview of the multilevel hierarchical model classification accuracies of the PLS-DA models with various pre-processing techniques applied for determining the species, fresh vs. previously frozen meat as well as the muscle type..... 225

Table 4.85 An overview of the multilevel hierarchical model classification accuracies of the PLS-DA models with various pre-processing techniques applied for determining the species, fresh vs. previously frozen meat as well as the frozen period. 226

List of Abbreviations

3D	three dimensional
AOTF	Acousto-Optic Tunable Filters
BD	<i>gastrocnemius</i> muscle
BF	<i>biceps femoris</i> muscle
BSE	bovine spongiform encephalopathy
CDA	canonical discriminant analysis
CE	capillary electrophoresis
CGE	capillary gel electrophoresis
CITES	Convention on International Trade in Endangered Species of Wild Fauna and Flora
CV	cross-validation
D	diffusion coefficients
DA	discriminant analysis
DESIR	dry extract spectroscopy by infrared reflectance
DT	detrend
EC	electrochemical
ELISA	enzyme-linked immunosorbent assay
FDA	factorial discriminant analysis
FF	<i>iliofibularis</i> muscle
FN	false negative
FP	false positive
FT	Fourier Transform
FT	frozen-thawed
FT-NIR	Fourier Transform Near-Infrared
InGaAs	Indium Gallium Arsenide
IS	<i>infraspinatus</i> muscle
IUCN	The International Union for Conservation of Nature
KNN	K-nearest neighbour
KS	Kennard-Stone
LC	liquid chromatography
LDA	linear discriminant analysis
LED	light emitting diodes
LTL	<i>longissimus thoracis et lumborum</i> muscle
LVs	latent variables
m	month

MDA	multivariate data analysis
MIA	multivariate image analysis
Mid-FTIR	Mid-Fourier Transform Infrared
Mid-IR	Mid-Infrared
MLR	multiple linear regression
MSC	multiplicative scattering correction
MSI	multi-spectral imaging
MT	magnetization transfer
NIR	near-infrared
NIR	near-infrared reflectance
NIT	near-infrared transmittance
NMR	nuclear magnetic resonance
P	position (anatomical)
PbS	lead sulphide
PC	principal component
PCs	principal components
PCA	principal component analysis
PCBs	polychlorinated biphenyl's
PCR	polymerase chain reaction
PCR	principal component regression
PLS	partial least squares
PLSR	partial least squares regression
PLS-DA	partial least squares discriminant analysis
Q^2	cross-validation coefficient of determination
QDA	quadratic discriminant analysis
R^2	coefficient of determination
RBF	radial basis function
RMSECV	root mean square error of cross-validation
RMSEP	root mean square error of prediction
SEM	scanning electron microscopy
SG	Savitzky-Golay
SGd ₁ (5)	Savitzky-Golay (1 st derivative, 2 nd order polynomial, 5 smoothing points)
SGd ₁ (7)	Savitzky-Golay (1 st derivative, 2 nd order polynomial, 7 smoothing points)
SGd ₂ (7)	Savitzky-Golay (2 nd derivative, 2 nd order polynomial, 7 smoothing points)
SGd ₂ (9)	Savitzky-Golay (2 nd derivative, 2 nd order polynomial, 9 smoothing points)
SIMCA	soft independent modelling of class analogy

SM	<i>semimembranosus</i> muscle
SNV	standard normal variate
SS	<i>supraspinatus</i> muscle
ST	<i>semitendinosus</i> muscle
SVM	support vector machines
TN	true negative
TP	true positive
USDA	United States Department of Agriculture
UV	ultra-violet
Vis-NIR	visible near-infrared

Chapter 1

Introduction

Meat and meat products represent a vital component of the human diet (reviewed by Ballin & Lametsch, 2008). Meat is deemed a high-value product, leading to an increased demand and consumption in most developed countries (reviewed by Prieto *et al.*, 2009). Presently, many consumers are concerned about the meat they eat (Ballin, 2010), as there have been incidences where meat has been fraudulently mislabelled. Thus, the quality and safety of this commodity is of growing concern to consumers, governmental control authorities and retailers (reviewed by Ballin & Lametsch, 2008) as well as all involved in the food industry from breeders and processors to manufacturers.

Consumers' increased health awareness has led to the demand for lean meat (Hoffman & Wiklund, 2006). Game meat is reported to have a fat content between 2 and 3%, which is lower than that of beef, pork and mutton/lamb (Schönfeldt, 1993). Viljoen (1999) also reported that it is lower in saturated- and higher in polyunsaturated fatty acids, compared to beef. Therefore, the South African game meat industry is continuing to grow locally and internationally (Bekker *et al.*, 2011), as consumers are in search of low kilojoule and low cholesterol meat products (Hoffman & Wiklund, 2006). Environmental concern amongst consumers has led to an increase in demand for free range and organic products, along with an increase in demand for products derived from natural production methods (Steenkamp, 1997). South African game meat can be classified as organic, as the game meat ranching is in accordance with the requirements for organic agricultural enterprises (Hoffman & Bigalke, 1999). Game meat has also gained a lot of attention in an increasingly health-aware market for its natural origin and lack of antibiotics, anabolic steroids, hormones and other additives (D'Amato *et al.*, 2013).

The South African game meat industry operates as a free-market enterprise, with the advantage of creating opportunities for individual game farmers and game meat processors (Hoffman *et al.*, 2004). The labelling of game meat also relies mainly or exclusively on wholesalers and manufacturers (D'Amato *et al.*, 2013). Free-market enterprise can however create certain problems for producers and consumers. For example, in South Africa no standardised meat cuts exist and there are no quality standards in place for game (Hoffman & Bigalke, 1999; Hoffman, 2001). This permits the legal sale of inferior quality game meat (Hoffman *et al.*, 2004). Due to a general lack of regulations as well as varying carcass dressings, such as head- and skin-off, the chance that a species may be mislabelled or substituted is greatly increased (D'Amato *et al.*, 2013).

Another concerning matter is the possibility of intentional distribution of endangered species in the market. In South Africa, the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) (CITES, 2017) listed the following ungulate species as endangered: bontebok (*Damaliscus pygargus pygargus/D. dorcas dorcas/D. p. Dorcas*) (CITES Appendix II); cape mountain zebra (*Equus zebra zebra*) (CITES Appendix I); southern white rhinoceros (*Ceratotherium simum simum*) (CITES Appendix I); black

rhinoceros (*Diceros bicornis*) (CITES Appendix I); and the African elephant (*Loxodonta africana*) (CITES Appendix II). The International Union for Conservation of Nature (IUCN) Red List of Threatened Species 2018 (IUCN, 2018) categorized these species as either ‘vulnerable’ (*E. zebra*, *L. africana*, *D. p. pygargus*) or ‘critically endangered’ (*Diceros bicornis*, *Ceratotherium simum simum*). Therefore, the identity of game meat is of mutual interest for both the meat industry and protection of biodiversity (D’Amato *et al.*, 2013). Despite the potential for a sustainable game meat market, very limited research has been undertaken on game meat fraud.

In recent years, there has been increased awareness in meat authenticity (Ballin, 2010; Cawthorn *et al.*, 2013; D’Amato *et al.*, 2013; Premanandh, 2013; reviewed by Sentandreu & Sentandreu, 2014). Accurate labelling of meat products is important to inform consumer preference, as meat forms a major part of the human food chain. Product preference is not just defined by consumer liking, but is also a function of lifestyle choices (e.g. vegetarianism, veganism and organic foods), religion (e.g. Halaal or Kosher), diet and concerns regarding optimal health (Ballin, 2010). Furthermore, regulations regarding fair-trade demands accurate labelling of products. A concerning matter is that meat and meat products can be attractive targets for adulteration in numerous ways (reviewed by Ballin & Lametsch, 2008). Typical cases involve the intentional substitution of high value raw ingredients with inferior species or materials, the addition of non-declared proteins from several origins, or the marketing of frozen-thawed meat as fresh (reviewed by Ballin & Lametsch, 2008; Alamprese *et al.*, 2016).

Frequently, meat from different origins present no obvious visual differences and could thus encourage adulteration and food fraud. The accurate classification of muscles is also critical for pricing, authentication and categorisation of meat (Kamruzzaman *et al.*, 2011), as some meats (muscles, grades) are more valuable to the consumer than others. The morphological characteristics of the muscles are lost when processed into minced meat or emulsified/manufactured meat products (Meza-Márquez *et al.*, 2010; Alamprese *et al.*, 2016). This makes it difficult to differentiate between muscle types and different species, creating the opportunity to fraudulently replace or substitute premium quality materials or species with grades that are inferior (Downey & Beauchêne, 1997b). This type of food fraud concerns consumers in terms of economic loss, food allergies, religious compliance, and food safety (Dean *et al.*, 2006). Meat fraud was suspected to be occurring on the South African market, and a study by Cawthorn *et al.* (2013) confirmed this suspicion with special focus on processed meats. A case study by D’Amato *et al.* (2013), exposed the high levels of unreliability of commercial labelling of game meat in South Africa. The extensive substitution and/or mislabelling of meat and wild game has important implications and is therefore an important authenticity issue.

Another well-recognised form of fraud is selling frozen-thawed meat as fresh, since fresh meat is considered to be superior and of higher value compared to frozen meat. Although frozen storage is an effective protective measure against microbiological deterioration of meat, its quality attributes and organoleptic properties suffer. Freezing negatively affects the meat, resulting in freezer burn, increased drip

loss and decreased juiciness (reviewed by Leygonie *et al.*, 2012). Therefore, thawed meat becomes drier and less tasty due to the loss of micro nutrients and water in the drip (Barbin *et al.*, 2013). Consumers thus associate frozen meat with an inferior quality compared to fresh meat. It is difficult to distinguish between fresh meat and meat that has been previously frozen, since they are similar in appearance (Barbin *et al.*, 2013). This type of problem is extensive throughout the food industry, making it difficult to identify retailers involved in fraudulent activities (Barbin *et al.*, 2013). The identification of fresh and frozen-thawed meat is therefore an important authenticity issue.

Various conventional analytical methods were proposed, used and evaluated to prevent retailers from offering fraudulent meat products (Ballin, 2010; Premanandh, 2013). Some reported methods for species determination are based on immunological detection and DNA-based procedures such as the enzyme linked immunosorbent assay (ELISA) (Martín *et al.*, 1988; Patterson & Jones, 1990; Smith, 1992; Jha *et al.*, 2003; Cawthorn *et al.*, 2013) and the polymerase chain reaction (PCR and real-time PCR) techniques (Calvo *et al.*, 2001; Calvo *et al.*, 2002; Rodriguez *et al.*, 2003; Vasconcellos *et al.*, 2003; Fajardo *et al.*, 2007; Fajardo *et al.*, 2008; Kesmen *et al.*, 2009; Soares *et al.*, 2010). Electrophoretic and chromatographic methods (Siebert *et al.*, 1994; Lerma-García *et al.*, 2009; Vallejo-Cordoba *et al.*, 2010; Hung *et al.*, 2011; Montowska & Pospiech, 2011; Mazorra-Manzano *et al.*, 2012) have also shown potential for this purpose. Enzyme activity determination (HADH method) (Gottesmann & Hamm, 1983; Chen *et al.*, 1988; Toldrá *et al.*, 1991; Sharma *et al.*, 1994; Ellerbroek *et al.*, 1995; Duflos *et al.*, 2002), nuclear magnetic resonance (NMR) (Guilheneuf *et al.*, 1997; Evans *et al.*, 1998; Mortensen *et al.*, 2006) and scanning electron microscopy (SEM) (Carroll *et al.*, 1981; Sen & Sharma, 2004) have been evaluated and used to discriminate between fresh and frozen-thawed meat.

Although these methods have shown potential to discriminate between species as well as fresh and frozen-thawed meat, no single satisfactory method has yet been developed. Instead a combination of techniques are utilised in practice to obtain reliable results (reviewed by Ballin & Lametsch, 2008). While most of the conventional techniques are able to detect low levels of adulteration with a high reliability, they are destructive, time-consuming, labour intensive and expensive. To date, no analytical authentication methods have yet been published for the identification of meat cuts or muscle types. In the meat industry, educated staff manually differentiate between primary meat cuts through visual inspection (Ballin, 2010). However, the visual authentication process is more problematic when meat is cut into steaks resulting in secondary meat cuts. Muscle authentication is further complicated since the names of primary and secondary meat cuts vary between countries (Ballin, 2010). These factors make the establishment of objective visual criteria for the purpose of authenticating specific meat cuts challenging. Thus, there is a need for rapid, reliable, robust and simple alternatives for the authentication of meat. Ideally the techniques should have the potential to be implemented on- or at-line in an abattoir or factory, be non-destructive and have a high level of accuracy as well as reproducibility, to evaluate meat samples for the purpose of authentication. Near-infrared (NIR) spectroscopy is a technique that meets all of these requirements (Manley & Baeten, 2018).

NIR spectroscopy is a suitable, rapid, non-destructive and accurate analytical technique which is popular and well established in meat science research as well as throughout the food industry (Downey & Beauchêne, 1997b; Downey & Beauchêne, 1997a; Rannou & Downey, 1997; Thyholt & Isaksson, 1997; Ding & Xu, 1999; Blanco & Villarroya, 2002; Pasquini, 2003; Cozzolino & Murray, 2004; reviewed by Ballin & Lametsch, 2008; reviewed by Prieto *et al.*, 2009; Alamprese *et al.*, 2013; Schmutzler *et al.*, 2015; Alamprese *et al.*, 2016). NIR spectroscopy is based on the interaction of NIR light with O-H, C-H, C=O and N-H vibrations in molecules, where molecules absorb energy from light with wavelengths of 780 to 2500 nm (Siesler, 2008; Manley, 2014). Therefore, the acquired spectra reveals information about the sample and its constituents, depending on how the light was either absorbed by the molecules, or scattered. Spectral bands in the NIR region tend to be broad, extensive and generally overlap, making it difficult to determine discrete chemical species (reviewed by Workman, 1993) and the analytical information obtained is influenced by a number of chemical, physical and structural variables (Blanco & Villarroya, 2002). Compounding this problem is the fact that many compounds absorb energy throughout the entire NIR region and slight spectral variances might be caused by differences between samples. Therefore, making it difficult and impractical to make a distinction between spectra with the naked eye. Thus, various multivariate analysis techniques can be implemented to extract the analytical data confined in the NIR spectra. Hence, the use of chemometrics is necessary for the decomposition and interpretation of NIR spectroscopic data (Blanco & Villarroya, 2002). The basic techniques required for effective data decomposition, interpretation, cluster analysis and quantification include: principal component analysis (PCA); principal component regression (PCR) (Hotelling, 1957; Kendall, 1957; Jeffers, 1967); multiple linear regression (MLR) (Bottenberg & Ward, 1963), linear discriminant analysis (LDA) (Fisher, 1936); partial least squares (PLS) (Wold, 1975); and partial least squares-discriminant analysis (PLS-DA) (Wold *et al.*, 1987).

NIR spectroscopy calibrations have been developed, within the meat sector, for the quantitative prediction of the chemical (Viljoen, 2003; Prevolnik *et al.*, 2010), physical (ElMasry *et al.*, 2012) and sensory (Liu & Chen, 2001; Barbin *et al.*, 2012; Kamruzzaman *et al.*, 2013) characteristics of meat. NIR spectroscopy has also been successfully used in discriminant analysis to recognise a specimen without the need of any chemical analysis, e.g., the discrimination between different types of ground beef samples (Prieto *et al.*, 2009); the differentiation between beef breeds (Alomar *et al.*, 2003); and the discrimination between fresh and frozen-thawed beef and lamb meat (Downey & Beauchêne, 1997a; Downey & Beauchêne, 1997b; Thyholt & Isaksson, 1997; Ropodi *et al.*, 2018). The technique has also successfully been used to discriminate between beef and kangaroo meat (Ding & Xu, 1999); beef, pork, chicken, turkey and lamb meat (Rannou & Downey, 1997; Cozzolino & Murray, 2004); and to detect and quantify adulterants in meat and minced beef (Ortiz-Somovilla *et al.*, 2005; Ortiz-Somovilla *et al.*, 2007; Meza-Márquez *et al.*, 2010; Alamprese *et al.*, 2013; Schmutzler *et al.*, 2015; Alamprese *et al.*, 2016) as well as to differentiate between beef cuts (Mitsumoto *et al.*, 1991) and discriminate between three muscle types of bovine meat (Alomar *et al.*, 2003).

NIR spectroscopy, combined with multivariate data analysis techniques, has proven to be an appropriate alternative as a rapid and non-destructive method for species and muscle type identification as well as the detection of fresh or frozen-thawed meat. However, previous studies did not combine and compare the different classes (species, muscles) and treatments (fresh, frozen-thawed) or investigate different frozen periods and the effect thereof on the differentiation. There is thus a need to combine the classes and treatments investigated previously, as well as to investigate the effect of different frozen periods on the differentiation accuracy. Despite the broad availability of literature on NIR spectroscopy applications to determine meat composition and quality (reviewed by Prieto *et al.*, 2009), no studies were found on ostrich (*Struthio camelus*) and South African game meat species, such as; black wildebeest (*Connochaetes gnou*), zebra (*Equus quagga burchelli*) and springbok (*Antidorcas marsupialis*), consequently requiring further investigation.

The aim of this dissertation was to rapidly differentiate between South African game species, irrespective of the treatment (fresh or previously frozen) or the muscle type as well as determine these individual classes (fresh; previously frozen; frozen period; muscle type) per species, using near-infrared (NIR) spectroscopy combined with multivariate data analysis. Specific objectives were established to develop methods and models that:

- permits the rapid differentiation of South African game species, irrespective of the meat being fresh, previously frozen and; the frozen period or the muscle type;
- enables the differentiation of fresh and frozen-thawed meat, irrespective of the frozen period or muscle type;
- determines the different frozen periods for the frozen-thawed meat (frozen 1 – 9 months); and
- enables the differentiation between the different muscle types for each species, irrespective of the treatment (fresh or previously frozen).

1.1 References

- Alamprese, C., Amigo, J.M., Casiraghi, E. & Engelsen, S.B. (2016). Identification and quantification of turkey meat adulteration in fresh, frozen-thawed and cooked minced beef by FT-NIR spectroscopy and chemometrics. *Meat Science*, **121**, 175-181.
- Alamprese, C., Casale, M., Sinelli, N., Lanteri, S. & Casiraghi, E. (2013). Detection of minced beef adulteration with turkey meat by UV-vis, NIR and MIR spectroscopy. *LWT-Food Science and Technology*, **53**, 225-232.
- Alomar, D., Gallo, C., Castaneda, M. & Fuchslocher, R. (2003). Chemical and discriminant analysis of bovine meat by near infrared reflectance spectroscopy (NIRS). *Meat Science*, **63**, 441-450.
- Ballin, N.Z. (2010). Authentication of meat and meat products. *Meat Science*, **86**, 577-587.
- Ballin, N.Z. & Lametsch, R. (2008). Analytical methods for authentication of fresh vs. thawed meat - A review. *Meat Science*, **80**, 151-158.

- Barbin, D.F., ElMasry, G., Sun, D.-W. & Allen, P. (2012). Predicting quality and sensory attributes of pork using near-infrared hyperspectral imaging. *Analytica Chimica Acta*, **719**, 30-42.
- Barbin, D.F., Sun, D.-W. & Su, C. (2013). NIR hyperspectral imaging as non-destructive evaluation tool for the recognition of fresh and frozen-thawed porcine *longissimus dorsi* muscles. *Innovative Food Science & Emerging Technologies*, **18**, 226-236.
- Bekker, J.L., Hoffman, L.C. & Jooste, P.J. (2011). Knowledge of stakeholders in the game meat industry and its effect on compliance with food safety standards. *International Journal of Environmental Health Research*, **21**, 341-363.
- Blanco, M. & Villarroya, I. (2002). NIR spectroscopy: a rapid-response analytical tool. *Trends in Analytical Chemistry*, **21**, 240-250.
- Bottenberg, R.A. & Ward, J.H. (1963). Applied multiple linear regression. (PRL-TDR-63-6): Lackland AF Base, Texas.
- Calvo, J., Zaragoza, P. & Osta, R. (2001). Random amplified polymorphic DNA fingerprints for identification of species in poultry pate. *Poultry Science*, **80**, 522-524.
- Calvo, J.H., Osta, R. & Zaragoza, P. (2002). Quantitative PCR detection of pork in raw and heated ground beef and pate. *Journal of Agricultural and Food Chemistry*, **50**, 5265-5267.
- Carroll, R., Cavanaugh, J. & Rorer, F. (1981). Effects of frozen storage on the ultrastructure of bovine muscle. *Journal of Food Science*, **46**, 1091-1094.
- Cawthorn, D.-M., Steinman, H.A. & Hoffman, L.C. (2013). A high incidence of species substitution and mislabelling detected in meat products sold in South Africa. *Food Control*, **32**, 440-449.
- Chen, M.-T., Yang, W.-D. & Guo, S.-L. (1988). Differentiation between fresh beef and thawed frozen beef. *Meat Science*, **24**, 223-226.
- CITES (2017). Convention of International Trade in Endangered Species of Wild Fauna and Flora: The CITES Appendices [WWW document]. <http://www.cites.org/eng/app/applications.php>. Accessed 19/07/2018
- Cozzolino, D. & Murray, I. (2004). Identification of animal meat muscles by visible and near infrared reflectance spectroscopy. *LWT-Food Science and Technology*, **37**, 447-452.
- D'Amato, M.E., Alechine, E., Cloete, K.W., Davison, S. & Corach, D. (2013). Where is the game? Wild meat products authentication in South Africa: a case study. *Investigative genetics*, **4**, 6.
- Dean, N., Murphy, T.B. & Downey, G. (2006). Using unlabelled data to update classification rules with applications in food authenticity studies. *Journal of the Royal Statistical Society: Applied Statistics*, **55**, 1-14.
- Ding, H. & Xu, R. (1999). Differentiation of Beef and Kangaroo Meat by Visible/Near-Infrared Reflectance Spectroscopy. *Journal of Food Science*, **64**, 814-817.
- Downey, G. & Beauchêne, D. (1997a). Authentication of fresh vs. frozen-then-thawed beef by near-infrared reflectance spectroscopy of dried drip juice. *LWT-Food Science and Technology*, **30**, 721-726.

- Downey, G. & Beauchêne, D. (1997b). Discrimination between fresh and frozen-then-thawed beef *m. longissimus dorsi* by combined visible-near infrared reflectance spectroscopy: A feasibility study. *Meat Science*, **45**, 353-363.
- Duflos, G., Le Fur, B., Mulak, V., Becel, P. & Malle, P. (2002). Comparison of methods of differentiating between fresh and frozen-thawed fish or fillets. *Journal of the Science of Food and Agriculture*, **82**, 1341-1345.
- Ellerbroek, L.I., Lichtenberg, G. & Weise, E. (1995). Differentiation between fresh and thawed meat by an enzyme profile test. *Meat Science*, **40**, 203-209.
- ElMasry, G., Sun, D.-W. & Allen, P. (2012). Near-infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef. *Journal of Food Engineering*, **110**, 127-140.
- Evans, S.D., Nott, K., Kshirsagar, A.A. & Hall, L.D. (1998). The effect of freezing and thawing on the magnetic resonance imaging parameters of water in beef, lamb and pork meat. *International journal of food science & technology*, **33**, 317-328.
- Fajardo, V., González, I., López-Calleja, I., Martín, I., Rojas, M., García, T., Hernandez, P. & Martín, R. (2007). PCR identification of meats from chamois (*Rupicapra rupicapra*), pyrenean ibex (*Capra pyrenaica*), and mouflon (*Ovis ammon*) targeting specific sequences from the mitochondrial D-loop region. *Meat Science*, **76**, 644-652.
- Fajardo, V., González, I., Martín, I., Rojas, M., Hernández, P.E., García, T. & Martín, R. (2008). Real-time PCR for detection and quantification of red deer (*Cervus elaphus*), fallow deer (*Dama dama*), and roe deer (*Capreolus capreolus*) in meat mixtures. *Meat Science*, **79**, 289-298.
- Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, 179-188.
- Gottesmann, P. & Hamm, R. (1983). New biochemical methods of differentiating between fresh meat and thawed, frozen meat. *Fleischwirtschaft*, **63**, 219-221.
- Guilheneuf, T.M., Parker, A.D., Tessier, J.J. & Hall, L.D. (1997). Authentication of the effect of freezing/thawing of pork by quantitative magnetic resonance imaging. *Magnetic Resonance in Chemistry*, **35**, 112-118.
- Hoffman, L.C. (2001). The effect of different culling methodologies on the physical meat quality attributes of various game species. In: Proceedings of the 5th International Wildlife Ranching Symposium Pp. 212-221. Pretoria, South Africa.
- Hoffman, L.C. & Bigalke, R.C. (1999). Utilising wild ungulates from southern Africa for meat production: potential research requirements for the new millennium. In: Proceedings of the 37th Congress of the Wildlife Management Association of South Africa. Pp. 20-21. Pretoria, South Africa.
- Hoffman, L.C., Muller, M., Schutte, D.W. & Crafford, K. (2004). The retail of South African game meat: current trade and marketing trends. *South African Journal of Wildlife Research*, **34**, 123-134.
- Hoffman, L.C. & Wiklund, E. (2006). Game and venison-meat for the modern consumer. *Meat Science*, **74**, 197-208.

- Hotelling, H. (1957). The relations of the newer multivariate statistical methods to factor analysis. *British Journal of Statistical Psychology*, **10**, 69-79.
- Hung, C.-J., Ho, H.-P., Chang, C.-C., Lee, M.-R., Franje, C.A., Kuo, S.-I., Lee, R.-J. & Chou, C.-C. (2011). Electrochemical profiling using copper nanoparticle-plated electrode for identification of ostrich meat and evaluation of meat grades. *Food Chemistry*, **126**, 1417-1423.
- IUCN (2018). The International Union for Conservation of Nature and Natural Resources (IUCN): Red List of Threatened Species 2018 [WWW document]. <http://www.iucnredlist.org/species>. Accessed 19/07/2018
- Jeffers, J. (1967). Two case studies in the application of principal component analysis. *Applied Statistics*, **16**, 225-236.
- Jha, V.K., Kumar, A. & Mandokhot, U.V. (2003). Indirect enzyme-linked immunosorbent assay in detection and differentiation of cooked and raw pork from meats of other species. *Journal of Food Science & Technology*, **40**, 254-256.
- Kamruzzaman, M., ElMasry, G., Sun, D.-W. & Allen, P. (2011). Application of NIR hyperspectral imaging for discrimination of lamb muscles. *Journal of Food Engineering*, **104**, 332-340.
- Kamruzzaman, M., ElMasry, G., Sun, D.-W. & Allen, P. (2013). Non-destructive assessment of instrumental and sensory tenderness of lamb meat using NIR hyperspectral imaging. *Food Chemistry*, **141**, 389-396.
- Kendall, M.G. (1957). *A Course in Multivariate Analysis*. London: Charles Griffin & Co. Ltd.
- Kesmen, Z., Gulluce, A., Sahin, F. & Yetim, H. (2009). Identification of meat species by TaqMan-based real-time PCR assay. *Meat Science*, **82**, 444-449.
- Lerma-García, M., Herrero-Martínez, J., Ramis-Ramos, G., Mongay-Fernández, C. & Simó-Alfonso, E. (2009). Prediction of the curing time of Spanish hams using peptide profiles established by capillary zone electrophoresis. *Food Chemistry*, **113**, 635-639.
- Leygonie, C., Britz, T.J. & Hoffman, L.C. (2012). Impact of freezing and thawing on the quality of meat: Review. *Meat Science*, **91**, 93-98.
- Liu, Y. & Chen, Y.-R. (2001). Two-dimensional visible/near-infrared correlation spectroscopy study of thawing behavior of frozen chicken meats without exposure to air. *Meat Science*, **57**, 299-310.
- Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chemical Society Reviews*, **43**, 8200-8214.
- Manley, M. & Baeten, V. (2018). Spectroscopic Technique: Near Infrared (NIR) Spectroscopy. In: *Modern Techniques for Food Authentication* (edited by D.-W. Sun). Pp. 51-102. Oxford UK: Elsevier.
- Martín, R., Azcona, J.I., García, T., Hernández, P.E. & Sanz, B. (1988). Sandwich ELISA for detection of horse meat in raw meat mixtures using antisera to muscle soluble proteins. *Meat Science*, **22**, 143-153.

- Mazorra-Manzano, M.A., Torres-Llanez, M.J., González-Córdova, A.F. & Vallejo-Cordoba, B. (2012). A capillary electrophoresis method for the determination of hydroxyproline as a collagen content index in meat products. *Food Analytical Methods*, **5**, 464-470.
- Meza-Márquez, O.G., Gallardo-Velázquez, T. & Osorio-Revilla, G. (2010). Application of mid-infrared spectroscopy with multivariate analysis and soft independent modeling of class analogies (SIMCA) for the detection of adulterants in minced beef. *Meat Science*, **86**, 511-519.
- Mitsumoto, M., Maeda, S., Mitsuhashi, T. & Ozawa, S. (1991). Near-infrared spectroscopy determination of physical and chemical characteristics in beef cuts. *Journal of Food Science*, **56**, 1493-1496.
- Montowska, M. & Pospiech, E. (2011). Differences in two-dimensional gel electrophoresis patterns of skeletal muscle myosin light chain isoforms between *Bos taurus*, *Sus scrofa* and selected poultry species. *Journal of the Science of Food and Agriculture*, **91**, 2449-2456.
- Mortensen, M., Andersen, H.J., Engelsen, S.B. & Bertram, H.C. (2006). Effect of freezing temperature, thawing and cooking rate on water distribution in two pork qualities. *Meat Science*, **72**, 34-42.
- Ortiz-Somovilla, V., España-España, F., De Pedro-Sanz, E. & Gaitán-Jurado, A. (2005). Meat mixture detection in Iberian pork sausages. *Meat Science*, **71**, 490-497.
- Ortiz-Somovilla, V., España-España, F., Gaitán-Jurado, A., Pérez-Aparicio, J. & De Pedro-Sanz, E. (2007). Proximate analysis of homogenized and minced mass of pork sausages by NIRS. *Food Chemistry*, **101**, 1031-1040.
- Pasquini, C. (2003). Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, **14**, 198-219.
- Patterson, R.L.S. & Jones, S.J. (1990). Review of current techniques for the verification of the species origin of meat. *Analyst*, **115**, 501-506.
- Premanandh, J. (2013). Horse meat scandal – A wake-up call for regulatory authorities. *Food Control*, **34**, 568-569.
- Prevolnik, M., Čandek-Potokar, M. & Škorjanc, D. (2010). Predicting pork water-holding capacity with NIR spectroscopy in relation to different reference methods. *Journal of Food Engineering*, **98**, 347-352.
- Prieto, N., Roehe, R., Lavin, P., Batten, G. & Andres, S. (2009). Application of near infrared reflectance spectroscopy to predict meat and meat products quality: A review. *Meat Sci*, **83**, 175-186.
- Rannou, H. & Downey, G. (1997). Discrimination of raw pork, chicken and turkey meat by spectroscopy in the visible, near-and mid-infrared ranges. *Analytical Communications*, **34**, 401-404.
- Rodriguez, M.A., García, T., González, I., Asensio, L., Hernández, P.E. & Martin, R. (2003). Qualitative PCR for the detection of chicken and pork adulteration in goose and mule duck foie gras. *Journal of the Science of Food and Agriculture*, **83**, 1176-1181.
- Ropodi, A.I., Panagou, E.Z. & Nychas, G.J.E. (2018). Rapid detection of frozen-then-thawed minced beef using multispectral imaging and Fourier transform infrared spectroscopy. *Meat Science*, **135**, 142-147.

- Schmutzler, M., Beganovic, A., Böhler, G. & Huck, C.W. (2015). Methods for detection of pork adulteration in veal product based on FT-NIR spectroscopy for laboratory, industrial and on-site analysis. *Food Control*, **57**, 258-267.
- Schönfeldt, H. (1993). Nutritional content of venison. In: The venison industry - Research requirements and possibilities Pp. 51-60. Meat industry centre, Irene Agricultural research Council Pretoria, South Africa.
- Sen, A.R. & Sharma, N. (2004). Effect of freezing and thawing on the histology and ultrastructure of buffalo muscle. *Asian Australasian Journal of Animal Sciences*, **17**, 1291-1295.
- Sentandreu, M.Á. & Sentandreu, E. (2014). Authenticity of meat products: Tools against fraud. *Food Research International*, **60**, 19-29.
- Sharma, N.K., Srivastava, A.K., Gill, J.P.S. & Joshi, D.V. (1994). Differentiation of meat from food animals by enzyme assay. *Food Control*, **5**, 219-221.
- Siebert, S., Beneke, B. & Bentler, W. (1994). Bef, pork and sheep meat-detecting previous frozen treatment by isoelectric-focusing in polyamide acryl gel (PAGIF) during routine diagnosis. *Fleischwirtschaft*, **74**, 417-420.
- Siesler, H.W. (2008). Basic principles of near-infrared spectroscopy. In: *Handbook of Near-Infrared Analysis* (edited by D.A. Burns & E.W. Ciurczak). Pp. 7-19. Boca Raton: CRC Press, Taylor & Francis Group.
- Smith, C. (1992). Application of immunoassay to the detection of food adulteration. In: *Food safety and quality assurance - Applications of immunoassay system*. . Pp. 13-32. London: Elsevier Applied Science.
- Soares, S., Amaral, J.S., Mafra, I. & Oliveira, M.B.P. (2010). Quantitative detection of poultry meat adulteration with pork by a duplex PCR assay. *Meat Science*, **85**, 531-536.
- Steenkamp, J.-B.E.M. (1997). Dynamics in consumer behavior with respect to agricultural and food products. In: *Agricultural marketing and consumer behavior in a changing world*. Pp. 143-188. Boston: Kluwer Academic Publishers.
- Thyholt, K. & Isaksson, T. (1997). Differentiation of frozen and unfrozen beef using near-infrared spectroscopy. *Journal of the Science of Food and Agriculture*, **73**, 525-532.
- Toldrá, F., Torrero, Y. & Flores, J. (1991). Simple test for differentiation between fresh pork and frozen/thawed pork. *Meat Science*, **29**, 177-181.
- Vallejo-Cordoba, B., Rodríguez-Ramírez, R. & González-Córdova, A.F. (2010). Capillary electrophoresis for bovine and ostrich meat characterisation. *Food Chemistry*, **120**, 304-307.
- Vasconcellos, L.P.d.M.K., Tambasco-Talhari, D., Pereira, A.P., Coutinho, L.L. & Regitano, L.C.d.A. (2003). Genetic characterization of Aberdeen Angus cattle using molecular markers. *Genetics and Molecular Biology*, **26**, 133-137.
- Viljoen, J.J. (1999). A comparison of the lipid components of springbok meat with those of beef and the related importance on aspects of health. *South African Journal of Science and Technology*, **18**, 51-53.

- Viljoen, M. (2003). *The use of near infrared reflectance spectroscopy (NIRS) for the chemical analysis of meat and feedstuffs.* (MSc Thesis). South Africa: University of Stellenbosch.
- Wold, H. (1975). Path models with latent variables: The NIPALS approach. In: *Quantitative sociology: International perspectives on mathematical and statistical modeling.* Pp. 307-357. New York: Academic Press.
- Wold, S., Geladi, P., Esbensen, K. & Öhman, J. (1987). Multi-way principal components-and PLS-analysis. *Journal of chemometrics*, **1**, 41-56.
- Workman, J. (1993). A brief review of the near infrared measurement technique. *NIR news*, **4**, 8-16.

Chapter 2

Near-infrared (NIR) spectroscopy for meat authentication: a review

2.1 Introduction

Meat and meat products represent a vital component of the human diet (reviewed by Ballin & Lametsch, 2008). Meat is deemed a high-value product, leading to an increased demand and consumption in most developed countries (reviewed by Prieto *et al.*, 2009). Presently, many consumers are concerned about the meat they eat (Ballin, 2010). Thus, the quality and safety of this commodity is of growing concern to consumers, governmental control authorities and retailers (reviewed by Ballin & Lametsch, 2008) as well as all involved in the food industry from breeders and processors to manufacturers.

Meat authenticity and traceability are important issues in modern society (Premanandh, 2013), as incidences regarding meat adulteration and fraud have become more sophisticated and mainstream (Cawthorn *et al.*, 2013). Food fraud is a collective term used for the deliberate and intentional substitution, addition, tampering or misrepresentation of food for economic gain (Spink & Moyer, 2011a). Therefore, the main driving force of meat adulteration can be attributed to the ever-increasing prices of commercial meat products, the globalisation of food trade and the increased processing of meat into value-added products (Cawthorn *et al.*, 2013). Consumers have the right to expect that the information provided with meat products is correct. Accurate labelling is thus important to inform consumer choice regarding aspects of lifestyle, religion, diet and health concerns, hence affecting the choice of one product over another. In addition, accurate labelling is vital to support fair-trade (Ballin, 2010). A concerning matter is that meat and meat products can be attractive targets for adulteration in numerous ways (reviewed by Ballin & Lametsch, 2008). Typical cases involve the intentional substitution of high value raw ingredients with inferior species or materials, the addition of non-declared proteins from several origins, or the marketing of frozen-thawed meat as fresh (reviewed by Ballin & Lametsch, 2008; Alamprese *et al.*, 2016).

Formerly, meat has not often been associated with adulteration because it was traditionally marketed as fresh and domestically prepared (Downey & Beauchêne, 1997b; Nakayinsige *et al.*, 2012). Currently, meat is more frequently consumed as processed and ready-to-eat convenience products (Nakayinsige *et al.*, 2012; Cawthorn *et al.*, 2013). In these products, the morphological characteristics of the muscles are lost when processed into mince, hamburgers, patties, sausages, meatballs and pâtés (Meza-Márquez *et al.*, 2010; Alamprese *et al.*, 2016). This makes it difficult to differentiate between muscle types and different species, creating the opportunity to fraudulently replace or substitute premium quality materials or species by grades that are inferior (Downey & Beauchêne, 1997b). This type of food fraud concerns consumers in terms of economic loss, food allergies, religious compliance, and food safety (Dean *et al.*, 2006).

Another well-recognised form of fraud is selling frozen-thawed meat as fresh, since fresh meat is considered to be superior and of higher value compared to frozen meat. It is very difficult to distinguish between fresh meat and meat that has been processed by freezing, since they are very similar in appearance (Barbin *et al.*, 2013). This type of problem is extensive throughout the food industry, making it difficult to identify fraudulent retailers (Barbin *et al.*, 2013).

In the foreseeable future, the meat industry will gradually move away from tedious and time-consuming analytical methods toward rapid, non-destructive, non-invasive, reproducible and reliable analytical techniques (Cozzolino & Murray, 2004; reviewed by Prieto *et al.*, 2009; Barbin *et al.*, 2013; Premanandh, 2013; Alamprese *et al.*, 2016). Particularly applicable are techniques to differentiate between meat samples for the purpose of classification assessment. Ideally the techniques should have the potential to be implemented on- or at-line in an abattoir or factory, be non-destructive and have a high level of accuracy as well as reproducibility, to evaluate meat samples for the purpose of authentication. Near-infrared (NIR) spectroscopy is a technique that meets all of these requirements. NIR spectroscopy is a suitable, rapid, non-destructive and accurate analytical technique which is popular and well established in meat science research as well as throughout the food industry (Downey & Beauchêne, 1997b; Downey & Beauchêne, 1997a; Rannou & Downey, 1997; Thyholt & Isaksson, 1997; Ding & Xu, 1999; Blanco & Villarroya, 2002; Pasquini, 2003; Cozzolino & Murray, 2004; reviewed by Ballin & Lametsch, 2008; reviewed by Prieto *et al.*, 2009; Alamprese *et al.*, 2013; Schmutzler *et al.*, 2015; Alamprese *et al.*, 2016). Over time, this technology has made considerable advances and is capable of acquiring an abundance of data, regarding a sample, in a single, rapid, non-destructive measurement (Bokobza, 1998; Blanco & Villarroya, 2002; Schmutzler *et al.*, 2015).

This review addresses and describes the different types of food fraud, and the possible utilisation of near-infrared (NIR) spectroscopy for meat differentiation, classification and authentication. Near-infrared spectroscopy will be discussed with regards to the history, working principle and theoretical background. Various spectroscopic acquisition configurations and principles of the commonly used multivariate techniques are discussed. The numerous applications of this technique, to differentiate and authenticate meat and meat products, are reviewed to illustrate the future potential of NIR spectroscopy in the meat industry. For the purpose of this review the term food fraud will be used as a collective term when referring to food tampering and misrepresentation. The term adulteration will be used when referring to substitution and addition, whereas the term authentication will be used when referring to the process of testing the meat's authenticity.

2.2 Food fraud

Food fraud, or the act of defrauding buyers of food and food ingredients for economic gain, has been problematic for the food industry throughout history (Johnson, 2014). Food fraud has been reported with increasing regularity over the past few years with the earliest cases involving wine (Accum, 1820; Wilson,

2008), spices, tea (Accum, 1820; Foster, 2011) and olive oil (Accum, 1820; Mueller, 2011). Worldwide, it is not conclusively known how widespread food fraud is. This is due to the fact that those committing fraud do not intend to cause physical harm, and attempt to avoid detection. Since most incidences generally do not result in a food safety risk, they go undetected and often consumers do not notice a quality deviation.

Despite there being very few health risks associated with the majority of food fraud, there have been cases in which it has led to potential and actual public health risks. **Table 2.1** contains examples of the most generally cited, high-profile cases of fraud that have affected global food safety. These highlight the major incidents that have recently occurred, yet food fraud perseveres and the detection thereof remains important (Ellis *et al.*, 2015). Furthermore, food adulterations are becoming progressively more sophisticated.

Table 2.1 Generally cited, high-profile cases of fraud that have affected global food safety.

Food product	Adulterant	Country	Year	Reference
Wine	Methanol	Italy	1980	Kuteifan <i>et al.</i> (1998)
	Diethylene glycol	Austria	1985	Hesser (1986)
Olive oil	Colza oil, intended for industrial use, was sold as olive oil and lead to toxic oil syndrome	Spain	1981	Pestana and Munoz (1982)
Beef	Scrapie-infected cattle feed leading to bovine spongiform encephalopathy (BSE)	UK	1980s/1990s	Wilesmith <i>et al.</i> (1992); Anderson <i>et al.</i> (1996)
Chilli powder and tomato-based products	Carcinogenic Sudan I-IV dyes	Various countries in the EU	2003	Calbiani <i>et al.</i> (2004)
Pork	Polychlorinated biphenyl's (PCBs) and dioxins via industrial oil contaminated animal feed	Ireland	2008	Kennedy <i>et al.</i> (2009); Marnane (2012)
Milk	Melamine	China	2008	Ingelfinger (2008); Xin and Stone (2008)

Nowadays, food and food ingredients associated with food fraud include olive oil, honey, milk and dairy products, meat products, fish, grain-based foods, fruit juices, wine and alcoholic beverages, organic foods, spices, coffee, tea and some extensively processed foods (Johnson, 2014).

Reports indicate that, in some markets, fish and seafood fraud may be extensive (Buck, 2010). The main issue in fish fraud is the mislabelling or substitution of species of a high value with inferior value and quality species (Cawthorn *et al.*, 2012; Cawthorn *et al.*, 2015). The species used for substitution could expose consumers to unexpected allergens and may also be associated with certain types of food poisoning (Buck, 2010). Correspondingly, the substitution of olive oil with additional types of legume, seed or nut oils could have unintentional consequences, if consumed by individuals with certain food allergies. The presence of unapproved chemicals in honey from China have been documented, indicating that the honey may contain certain unapproved antibiotics or other agricultural chemicals (Schneider, 2011). Reports also indicate the substitution and dilution of some fruit juices with juice from rotten fruit, containing toxic mould (FDA, 2004).

The risks posed by other types of food fraud are not well-documented and the consequences of the adulteration may be less harmful, or the risks may never be known (Johnson, 2014). Generally, those committing the fraud are the only individuals who have knowledge thereof. Thus, exposition of food fraud cases may only happen once an investigation into a serious public health event has occurred (Spink & Moyer, 2011b). Some fraud cases, on the other hand, might never be exposed despite having the potential to contribute to enduring health concerns (Johnson, 2014).

Not all food fraud cases result in a food safety crisis, with the sole purpose of some being to deceive the buyer with regards to his/her expectation of the product. The vast majority of food fraud cases involve the replacement of a high-value product with an inexpensive or lower quality substitute (Dennis, 1998). A concerning matter is that meat and meat products can be attractive targets for numerous adulterations and are often associated with such food fraud cases (Cozzolino & Murray, 2004).

2.3 Meat fraud

In recent years, there has been increased awareness in meat authenticity (Ballin, 2010; Cawthorn *et al.*, 2013; D'Amato *et al.*, 2013; Premanandh, 2013; reviewed by Sentandreu & Sentandreu, 2014). Consumers are concerned and becoming more aware of the meat they eat. Accurate labelling of meat products is important to inform consumer preference as it forms a major part of the human food chain. Product preference is not just defined by consumer liking, but is also a function of lifestyle choices (e.g. vegetarianism, veganism and organic foods), religion (e.g. Halaal or Kosher), diet and concerns regarding optimal health (Ballin, 2010). Furthermore, regulations regarding fair-trade, demands accurate labelling of products.

Consumers' increased health awareness has led to the demand for lean meat (Hoffman & Wiklund, 2006). Game meat is reported to have a fat content between 2 and 3%, which is lower than that of beef, pork and mutton/lamb (Schönfeldt, 1993). Viljoen (1999) also reported that it is lower in saturated- and higher in polyunsaturated fatty acids, compared to that of beef. Therefore, the South African game meat

industry is continuing to grow locally and internationally (Bekker *et al.*, 2011), as consumers are in search of low kilojoule and low cholesterol meat products (Hoffman & Wiklund, 2006). The South African game meat industry operates as a free-market enterprise, with the advantage of creating opportunities for individual game farmers and game meat processors (Hoffman *et al.*, 2004). Free-market enterprise can however create certain problems for producers and consumers. For example, in South Africa no standardised meat cuts exist and there are no quality standards in place for game (Hoffman & Bigalke, 1999; Hoffman, 2001). Therefore, allowing the legal sale of inferior quality game meat (Hoffman *et al.*, 2004). Despite the potential for a sustainable game meat market, very limited research has been undertaken on game meat fraud.

Meat fraud is grouped into four main categories (**Fig. 2.1**): meat origin, meat substitution, meat treatment and non-ingredient addition. These are then further subcategorised as follows: meat origin – species, sex, breed, feed intake, slaughter age, meat cuts, farmed versus wild meat, conventional versus organic meat and geographic origin; meat substitution – meat species, protein and fat; meat treatment – meat preparation and fresh versus thawed meat; non-meat ingredient addition – water and additives.

The key meat fraud issues, e.g., origin (Ballin, 2010; Primrose *et al.*, 2010), species (Ballin, 2010; Primrose *et al.*, 2010; Alamprese *et al.*, 2013; Premanandh, 2013) and treatment (previously frozen meat marketed as fresh) (reviewed by Ballin & Lametsch, 2008; Alamprese *et al.*, 2016) will be discussed and reviewed.

2.3.1 Meat origin and substitution

Frequently, meat from different origins present no obvious visual differences and could thus create problems of adulteration and food fraud. The morphological characteristics of muscles are removed by the production of minced meat or emulsified/manufactured meat products, making it difficult to differentiate between muscle types. Incidences in the UK and Europe in early 2013 reported the mis-labelling of beef products (Premanandh, 2013). These products were essentially found to contain 80 – 100% horse meat instead of beef, which the supplier withheld from authorities (Premanandh, 2013). This case is considered as economic adulteration in which partial or full substitution of high value species with low value species occurs, in the meat industry, for profitable gains. Although the incident eventually did not result in public health concerns, recent debates surrounding the horse meat scandal forces authorities to implement strict regulations on food adulteration (Premanandh, 2013). Similar incidences are not simply concerning for importation and the meat packers, but also at restaurant and retail level, where substitution is more easily concealed. The horsemeat scandal was consequently a wake-up call for food management and safety in South Africa (reviewed by Clark, 2013). Meat fraud was suspected to be occurring on the South African market and a study by Cawthorn *et al.* (2013) confirmed this suspicion with special focus on processed meats. A case study by D'Amato *et al.* (2013) exposed the high levels of unreliability of commercial labelling of game meat in South Africa. The extensive substitution of meat and wild game has important implications and is therefore an important authenticity issue.

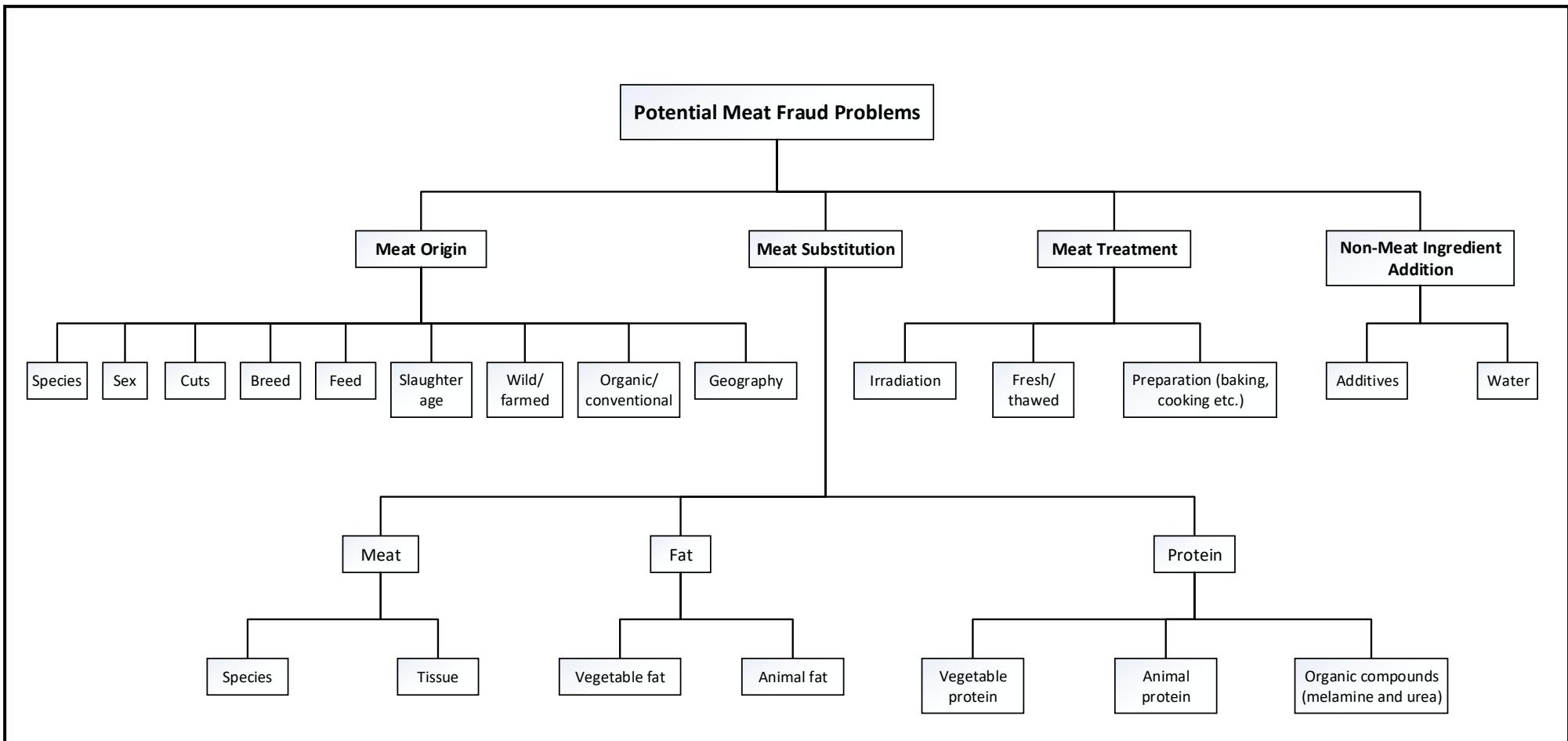


Figure 2.1 Diagram representing potential meat fraud problems [Adapted from (Ballin, 2010)].

2.3.2 Meat treatment

Extending the shelf-life of meat is imperative due to its perishable nature (Downey & Beauchêne, 1997b). It is well established that freezing, over long periods, is a safe and effective form of preservation for meat (Barbin *et al.*, 2013). Although frozen storage is an effective protective measure against microbiological deterioration of meat, its quality attributes and organoleptic properties suffer. Ice crystal growth is a crucial phenomenon, causing mechanical damage, denaturation of proteins and osmotic removal of water (Thyholt & Isaksson, 1997; reviewed by Leygonie *et al.*, 2012). Cells, mitochondria and organelles are disrupted in varying degrees, causing a release of enzymes and other components into the meat drip juice. Freezing negatively affects the meat, resulting in freezer burn, increased drip loss and decreased juiciness. Therefore, thawed meat becomes drier and less tasty due to the loss of micro nutrients and water in the drip (Barbin *et al.*, 2013). Consumers thus associate frozen meat with an inferior quality compared to fresh meat (chilled post-slaughter and stored at refrigeration temperature) and this association could explain the difference in price. Because of the similarities in appearance of fresh and frozen-thawed meat, it is very difficult for consumers to distinguish between them. This makes it attractive for retailers to fraudulently sell previously frozen meat as chilled fresh (Barbin *et al.*, 2013). The identification of fresh and frozen-thawed meat is therefore an important authenticity issue.

2.4 Conventional analytical detection methods

Various methods were proposed, used and evaluated to prevent retailers from offering fraudulent meat products (Ballin, 2010; Premanandh, 2013). The general information of all the methods reviewed is given in **Table 2.2**. An overview is given on different selected analytical methods that have been traditionally used to assess meat authenticity.

Table 2.2 Selected applications of conventional analytical detection methods used for meat authentication.

Authenticity problem	Analytical technique	Reference
Species determination	Immunological methods (ELISA)	Martín <i>et al.</i> (1988); Patterson and Jones (1990); Smith (1992); Jha <i>et al.</i> (2003); Cawthorn <i>et al.</i> (2013)
	Electrophoretic and chromatographic methods	Siebert <i>et al.</i> (1994); Lerma-García <i>et al.</i> (2009); Vallejo-Cordoba <i>et al.</i> (2010); Hung <i>et al.</i> (2011); Montowska and Pospiech (2011); Mazorra-Manzano <i>et al.</i> (2012)
	DNA-based procedures (PCR and real-time PCR techniques)	Calvo <i>et al.</i> (2001); Calvo <i>et al.</i> (2002); Rodriguez <i>et al.</i> (2003); Vasconcellos <i>et al.</i> (2003); Fajardo <i>et al.</i> (2007); Fajardo <i>et al.</i> (2008); Kesmen <i>et al.</i> (2009); Soares <i>et al.</i> (2010)

Table 2.2 (Continued)

Authenticity problem	Analytical technique	Reference
Differentiating between fresh and frozen-thawed meat	Enzyme activity determination (HADA method)	Gottesmann and Hamm (1983); Chen <i>et al.</i> (1988); Toldrá <i>et al.</i> (1991); Sharma <i>et al.</i> (1994); Ellerbroek <i>et al.</i> (1995); Duflos <i>et al.</i> (2002)
	Nuclear magnetic resonance (NMR)	Guilheneuf <i>et al.</i> (1997); Evans <i>et al.</i> (1998); Mortensen <i>et al.</i> (2006)
	Electron microscopy (SEM)	Carroll <i>et al.</i> (1981); Sen and Sharma (2004)

2.4.1 Immunological methods

Enzyme linked immunosorbent assay (ELISA), is possibly the most generally used method amongst the different immunological assays for meat species determination (reviewed by Sentandreu & Sentandreu, 2014). ELISA-based techniques have been used in research for the detection of meat from different species within a food product (Martín *et al.*, 1988; Jha *et al.*, 2003; Cawthorn *et al.*, 2013). It is based on the cultivation of antibodies or antisera, with the capability of binding to a protein of interest (Reid *et al.*, 2006), therefore permitting both the qualitative and quantitative detection of that protein. This approach has one main advantage; proteins can be recognised and quantified exclusively as the antibodies or antisera are produced to respond to the specific protein of interest. Despite its advantage, the need for specific antibodies raises some limitations. One of the main challenges of the ELISA approach is the production of a protein-specific antibody. This requires the antibodies of the species or tissue in question to be highly specific. If this is not the case, cross-reactions may occur, resulting in false positive cases (reviewed by Sentandreu & Sentandreu, 2014). This can be particularly problematic when differentiating between closely related species. This is evidenced by the inability of the commercial ELISA kits to differentiate between chicken and turkey and instead identifying both as poultry meat (Giovannacci *et al.*, 2004). Another limitation concerns the analysis of highly processed meat products. Extreme processing conditions like very high temperatures may alter the 3° structure of the proteins, thereby decreasing the ability to recognise the specific antibody (Nakyinsige *et al.*, 2012). This can result in false negatives or underestimated results of the quantified ingredient/constituent. Some innovations have reported strategies to overcome this problem by producing antibodies raised against thermostable proteins (reviewed by Kreuz *et al.*, 2012; Sentandreu & Sentandreu, 2014). Although this method holds ample potential for the authentication of foodstuffs, its authentication capabilities are restricted due to limited advances of the technique (Reid *et al.*, 2006).

2.4.2 Electrophoretic and chromatographic methods

Protein electrophoresis (mono- and bi-dimensional) is a technique used to reveal mobility differences in the target proteins of different species of meat (Montowska & Pospiech, 2011). However, results obtained from this technique can be ambiguous, unclear and inconclusive due to restrictions in terms of repeatability as

well as low discriminating power (reviewed by Sentandreu & Sentandreu, 2014). Sample preparation and analysis is also time-consuming. Capillary electrophoresis (CE) has thus been assayed as an alternative authentication technique. CE allows for the rapid and efficient separation of charged components present in small sample volumes (Li, 1992). The principle of CE, is that separations occur due to the different electrophoretic mobilities of ions in the specific electrophoretic media inside small capillaries. A standard CE instrument can be used to perform a variety of different modes of capillary electrophoresis (Li, 1992), with capillary gel electrophoresis (CGE) being the main mode used for the separation of proteins, polynucleotides and DNA fragments. In CGE the mechanism of separation is based on “molecular sieving” (Li, 1992). The analytes move through the pores of the gel-filled column and is then separated due to solute size differences. Vallejo-Cordoba *et al.* (2010) demonstrated the capability of this approach to distinguish between ostrich and bovine meat by creating electrophoretic protein profiles. However, this technique has limited abilities when authenticating products containing mixed meats, as discrimination is predominantly based on quantitative differences of proteins found in both species. CE has also been used as an approach to compare peptide profiles (Lerma-García *et al.*, 2009), as well as determine the hydroxyproline content in meat products as a guide of the collagen content (Mazorra-Manzano *et al.*, 2012). Often liquid chromatography (LC) is coupled to electrochemical (EC) detection for accurate hydroxyproline content determination. Electrochemical detection has one main advantage, namely, the possibility of direct analysis of amino acids and peptides without undergoing prior derivatisation. The approach of coupling LC to EC has also been used for the specific detection of ostrich meat and its differentiation from beef, pork and chicken meat (Hung *et al.*, 2011).

2.4.3 DNA-based procedures

DNA-based analysis is used as an alternative to overcome the reported limitations of methods based on protein analysis and other target compounds (reviewed by Sentandreu & Sentandreu, 2014). The main advantages of DNA-based procedures are their high discriminating power and level of sensitivity. The majority of work associated with the development of DNA analysis has focussed on the amplification of DNA in targeted regions using a technique called polymerase chain reaction (PCR) (Reid *et al.*, 2006). This technique allows for identification of species based on a DNA fragment with a unique 1° structure (sequence) (reviewed by Sentandreu & Sentandreu, 2014). This feature allows for the acquisition of unambiguous results, making the assays efficient and reliable. The principle of PCR is that it copies a specific DNA fragment multiple times, providing a large amount of genetic material of a specific region to be analysed. A variety of methods is used for end-point analysis, with electrophoretic techniques (i.e. gel electrophoresis) being the most common. PCR was evaluated for the authentication of meat and meat-based products (Calvo *et al.*, 2002). The study proved that PCR could be used to detect adulterated minced beef with both pork as well as chicken meat. In addition, a real-time PCR assay was developed for differentiating between chicken and turkey meat (Kesmen *et al.*, 2012), and between horse and donkey meat (Kesmen *et al.*, 2009). The main

difference between conventional PCR and real-time PCR is the analysis of the results. Real-time PCR is more rapid as the detection of amplifications occurs earlier in the reaction (Heid *et al.*, 1996), where a fluorescent dye system is used to detect the products during amplicon formation (reviewed by Rebrikov & Trofimov, 2006). These DNA methodologies were effective at identifying chicken as an adulterant in goose and duck foie gras (Rodriguez *et al.*, 2003) and the adulteration of poultry pâté with pork (Calvo *et al.*, 2001). In other work it was successfully used to identify the exact breed of cattle and could therefore successfully differentiate between the different breeds (Vasconcellos *et al.*, 2003). It was also used to differentiate between meat mixtures from different breeds of deer (Fajardo *et al.*, 2008). Therefore, the PCR approach could be used to address the issue of discrimination between species that are closely related or between different breeds belonging to the same species (Fajardo *et al.*, 2008).

Despite the success, genetic methods also show a few limitations, especially when analysing processed meat products (reviewed by Sentandreu & Sentandreu, 2014). The complexity of food matrices is a hurdle for the development of standardised extraction protocols. Therefore, in order to guarantee reproducible amounts of DNA extraction from samples, the protocols need to be optimised for each specific case (López-Andreo *et al.*, 2012). Degradation of DNA can occur due to the processing of meat products, where it is exposed to changes in pH and high temperatures which could also compromise cellular integrity (reviewed by Sentandreu & Sentandreu, 2014). Consequently, this impairs the quantitative measurements, as the amount of DNA present in the product is not a true reflection of the real amount of the source material (Primrose *et al.*, 2010). Another drawback is the shortening of the DNA fragments. The implication of shortened fragments is that they are used to perform PCR assays and will result in an increase in the cross-reactions with different species (Hird *et al.*, 2006). In that respect, different strategies have been devised to overcome some of these drawbacks (Soares *et al.*, 2013). A real-time PCR approach was developed for the purpose of quantifying pork at varying levels in processed meat products. Binary mixtures of known concentrations of pork in poultry meat was used to construct a standard curve to obtain reliable quantitative results. The curve was used to control variations in DNA extractions and efficiency of amplification (reviewed by Sentandreu & Sentandreu, 2014).

2.4.4 Enzyme activity determination

The β -hydroxyacyl-CoA-dehydrogenase (HADH) method (Gottesmann & Hamm, 1983) is generally the most used enzymatic method to differentiate between fresh and previously frozen meat. This technique is based on the principle that the freezing and thawing process results in a release of the enzyme HADH. Freezing disrupts the mitochondria, releasing this enzyme into the sarcoplasm and resulting in an elevated HADH activity in the muscle press-juice (Chen *et al.*, 1988). Unfortunately, this method has a few drawbacks. When meat is subjected to long-term storage or inadequate freezing conditions the protease activity is prolonged or increased (reviewed by Ballin & Lametsch, 2008). This phenomenon can, in theory, result in HADH digestion and therefore erroneously indicate that the sample is fresh. Another disadvantage is that

alternative freezing techniques (slow freezing, air blast freezing, cryogenic freezing, among others) could prevent the disruption of the mitochondria and as a result prevent HADH release. In order to overcome this problem an additional analytical method, e.g., scanning electron microscopy, could be used to confirm the results (reviewed by Ballin & Lametsch, 2008). The HADH technique is only applicable to whole pieces of meat, as grinding of the meat disrupts the mitochondria, resulting in HADH release.

2.4.5 Nuclear magnetic resonance

Nuclear magnetic resonance (NMR) has shown potential in discriminating fresh from thawed meat (reviewed by Ballin & Lametsch, 2008). NMR spectroscopy involves the analysis of the energy absorption by atomic nuclei with non-zero spins in the presence of a magnetic field (Reid *et al.*, 2006). The presence of nuclei in the surrounding molecules affect the energy absorption of the atomic nuclei, causing small modifications to the external magnetic field. This technique can therefore obtain comprehensive information regarding the molecular structure of foodstuff. During freezing the water binding capacity and water distribution is altered and therefore NMR can be used to detect these changes (reviewed by Ballin & Lametsch, 2008). Magnetic resonance analysis measures the longitudinal relaxation times (T_1 values), transverse relaxation times (T_2 values), magnetization transfer (MT) rates and apparent water diffusion coefficients (D), in order to authenticate the effect of freezing/thawing in meat. Studies reported a significant decrease in the T_2 values of whole pork that had been thawed compared to fresh pork (Guilheneuf *et al.*, 1997; Mortensen *et al.*, 2006), and an increased temperature in the freezer led to a slower T_2 (Mortensen *et al.*, 2006). The T_1 was affected similarly, with it decreasing in thawed meat in comparison to the fresh pork (Guilheneuf *et al.*, 1997), lamb and beef (Evans *et al.*, 1998). The MT rate was investigated and when compared to fresh meat, an increased value was observed in thawed pork (Guilheneuf *et al.*, 1997), lamb and beef (Evans *et al.*, 1998). Increasing MT rate can be attributed to two factors, namely; a decreased moisture content and the myofibrillar proteins denaturing in meat that has been frozen/thawed. The main advantage of NMR spectroscopy is its high level of specificity and accuracy in food characterisation (Reid *et al.*, 2006). Despite its benefits, from an industrial point of view, instrumentation, running costs and routine analysis can be expensive and complicated, requiring skilled workers to carry out the tests.

2.4.6 Electron microscopy

Scanning electron microscopy (SEM) focusses on the histological changes in frozen meats by examining it on a microscopic or ultra-structural level (Sen & Sharma, 2004). The freezing conditions determine the size of ice crystals. The formation of large extracellular ice crystals can be attributed to slow freezing, whereas rapid freezing produces numerous, well distributed, small ice crystals throughout the meat tissue (Fennema *et al.*, 1973; Martino *et al.*, 1998; Sanz *et al.*, 1999; reviewed by Leygonie *et al.*, 2012). The size as well as location of the ice crystals are responsible for the degree to which the meat is damaged by freezing (reviewed by Ballin & Lametsch, 2008). Studies performed on whole bovine (Carroll *et al.*, 1981) and buffalo meat (Sen &

Sharma, 2004) reported that the degree of microstructure deterioration was attributed to the freeze and thaw conditions applied to the meat. This phenomenon made it possible to detect differences between fresh and previously frozen meat, however, the authors suggested that more research is required to establish the exact relationship. The main drawback of this method is that microstructural damage may be difficult to detect when the meat is subjected to certain freeze-thaw conditions, making it difficult and sometimes impossible to discriminate fresh from thawed meat (reviewed by Ballin & Lametsch, 2008).

Although these methods have shown potential to discriminate between species as well as fresh and frozen-thawed meat, no single satisfactory method has yet been developed. Instead a combination of techniques are utilised in practice to obtain reliable results (reviewed by Ballin & Lametsch, 2008). While most of the traditional techniques are able to detect low levels of adulteration with a high reliability, they are destructive, time-consuming, labour intensive and expensive. Thus, there is a need for rapid, reliable, robust and simple alternatives for the authentication of meat. NIR spectroscopy, combined with multivariate data analysis techniques, has proven to be an appropriate alternative for species identification and detection of fresh or frozen-thawed meat.

2.5 Near-infrared (NIR) spectroscopy

In 1800, William Herschel discovered NIR energy while measuring heat energy of solar emission (Herschel, 1800). He was the first to document the first NIR spectrum while observing radiation beyond the red portion of the visible spectrum (Herschel, 1800). The NIR region (780 – 2500 nm), within the electromagnetic spectrum, consists of absorption bands corresponding to overtones and combinations of fundamental O-H, C-H, C=O and N-H vibrations (Bokobza, 1998; Barton, 2002; Bokobza, 2002; Williams, 2006; Cen & He, 2007). These molecular bonds experience vibrational energy changes when irradiated with NIR frequencies (Barton, 2002; Bokobza, 2002; Cen & He, 2007; Siesler, 2008b). Two vibration patterns occur, namely stretching and bending. When the inter-atomic distances along the axis between adjacent atoms are subjected to a continuous change, it is known as a stretching vibration. A change in the bond angle between adjacent atoms is known as a bending vibration (reviewed by Workman, 1993). Absorption of energy occurs when the light energy is equal to the frequency of the molecular bond.

Spectral bands in the NIR region tend to be broad, extensive and generally overlap, making it difficult to determine discrete chemical species (reviewed by Workman, 1993) and the analytical information obtained is influenced by a number of chemical, physical and structural variables (Blanco & Villarroya, 2002). Compounding this problem is the fact that many compounds absorb energy throughout the entire NIR region and slight spectral variances might be caused by differences between samples, making it difficult and impractical to make a distinction with the naked eye. Therefore, various multivariate analysis techniques can be implemented to extract the analytical data, confined in the NIR spectra. Hence, the use of chemometrics is necessary for the decomposition and interpretation of NIR spectroscopic data (Blanco & Villarroya, 2002). The basic techniques required for effective data decomposition, interpretation, cluster analysis and

quantification include: principal component analysis (PCA), principal component regression (PCR) (Hotelling, 1957; Kendall, 1957; Jeffers, 1967), multiple linear regression (MLR) (Bottenberg & Ward, 1963), linear discriminant analysis (LDA) (Fisher, 1936), partial least squares (PLS) (Wold, 1975) and partial least squares-discriminant analysis (PLS-DA) (Wold *et al.*, 1987b). Advances and refinement in the instrumentation and the technology have required that newer techniques be developed to analyse the larger data sets.

Early in the 1960s, Karl Norris, of the United States Department of Agriculture (USDA), introduced NIR spectroscopy as an accepted and recognised analytical tool (Norris, 1962; Norris, 1964). Since then the application of NIR spectroscopy is more widespread and is found in a great number of diverse fields such as food and agriculture (Szabo *et al.*, 1991; Lammertyn *et al.*, 2000; McGlone *et al.*, 2002; Lovett *et al.*, 2004; Lovett *et al.*, 2005; Han *et al.*, 2006; Siuda *et al.*, 2006; Juhász *et al.*, 2007; Nicolai *et al.*, 2007; Ortiz-Somovilla *et al.*, 2007); medicine (Sakudo *et al.*, 2006; Tisdall *et al.*, 2007; Wolf *et al.*, 2007; Ciurczak & Igne, 2014); the pharmaceutical industry (Reich, 2005; Rodionova *et al.*, 2005; Luypaert *et al.*, 2007; Roggo *et al.*, 2007; Jamrógiewicz, 2012) and animal science (Cozzolino & Murray, 2004; Cozzolino *et al.*, 2006; Landau *et al.*, 2006).

The growing interest in this technique is due to its advantages over alternative instrumental techniques. With technological advances and the increasing quality and safety concern, it has become vital to trace molecules, chemicals or constituents within samples. NIR spectroscopy can, without sample pre-treatment, record spectra of liquid and solid samples, implement continuous methodologies (on-line or in-line application), provide instant spectra and predict chemical and physical parameters from a single measurement (Blanco & Villarroya, 2002). These features make it particularly attractive for straightforward, rapid, non-destructive sample characterisation and could therefore lead to superior quality control and effective process monitoring within a factory, particularly for on-line, in-line and at-line applications.

2.5.1 Principles of NIR spectroscopy

2.5.1.1 What is NIR spectroscopy?

NIR spectroscopy is a non-destructive, rapid and accurate analytical technique that incorporates spectroscopy and applied statistics to analyse samples (Williams, 2006). NIR spectroscopy is based on the interaction of NIR light with O-H, C-H, C=O and N-H vibrations in molecules, where molecules absorb energy from light of different wavelengths (Siesler, 2008b). When the electromagnetic radiation, e.g. infrared light, interacts with the sample, the radiation can be absorbed, transmitted, reflected, scattered or subjected to photoluminescence (**Fig. 2.2**). Therefore, the ability to acquire a spectrum depends completely on the detection of light energy, due to light of a particular wavelength being reflected from the object or transmitted through the object. Additional light of particular wavelengths is absorbed and in spectrochemical methods the absorbed radiation is measured. Therefore, the acquired spectra reveals information about the sample and its constituents depending on how the light was either absorbed by the molecules, or scattered.

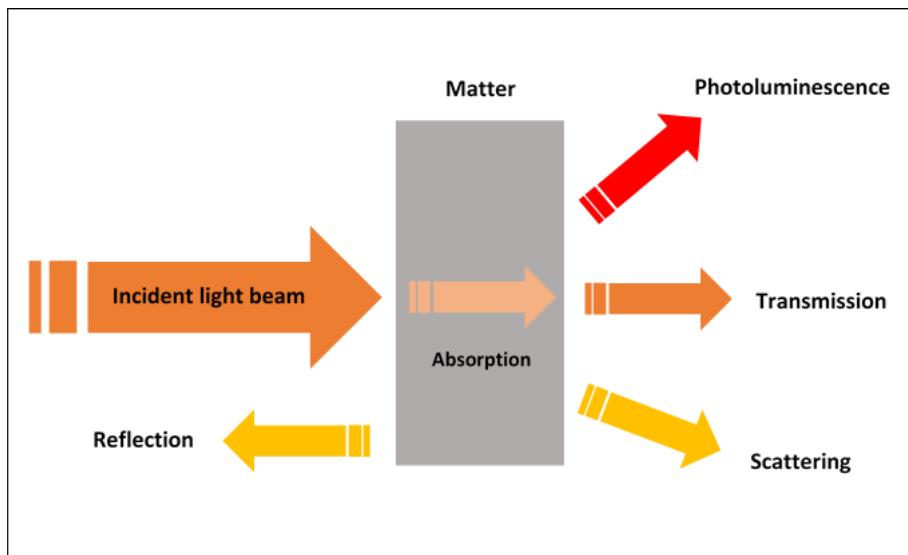


Figure 2.2 Schematic representation of the interaction of radiation and matter (Anonymous, 2017).

2.5.1.2 NIR instrumentation and spectral acquisition

A typical NIR instrument mainly consists of a light source, a wavelength selection system, a detector and a signal processor, all coupled to a computer (Blanco & Villarroya, 2002; Williams, 2006; Cen & He, 2007). The detectors often utilized for the NIR spectral region are based on silicon, InGaAs (Indium Gallium Arsenide) and PbS (lead sulphide) photoconductive materials. Systems responsible for wavelength selection, separation and measurement include discrete narrow band-pass filters, the monochromator and the acousto-optic tunable filter (AOTF). The sample can be irradiated from either above or below and the instrument is composed of a light source, generally tungsten halogen lamps or light-emitting diodes (LEDs) (Blanco & Villarroya, 2002; Cen & He, 2007). There are various existing instrument configurations that allow for different spectral acquisition. The two basic measurement modes most frequently used in modern NIR analysis are reflectance and transmittance (Fig. 2.3). Various measurement modes and NIR instruments exist and are available as either benchtop or portable handheld devices. These instruments can be of two types and are usually differentiated and classified based on the way the wavelengths are scanned or selected (Pasquini, 2003), namely discrete wavelength and whole spectrum. These instruments include: filter-based instruments, LED based instruments, dispersive instruments and interferometric instruments.

2.5.1.3 Reflectance and transmittance mode

Often reflectance and transmittance modes are termed NIR- and NIT measurements, respectively (Williams, 2006). For reflectance mode, primarily used for solid or granular samples, the light source and the detector are situated on the same side of the sample. Two types of reflectance occur, e.g. specular (surface) reflectance and diffuse reflectance (Williams, 2006). The specular reflectance does not contain any information about the sample, whereas diffuse reflectance contains information concerning the sample's composition and characteristics. It is therefore essential to only measure the diffusely reflected radiation.

However, in transmittance mode, often used for liquid samples, the light source and the detector are on opposite sides of the sample. The light from the illuminating source thus passes through the sample and is received by the detector.

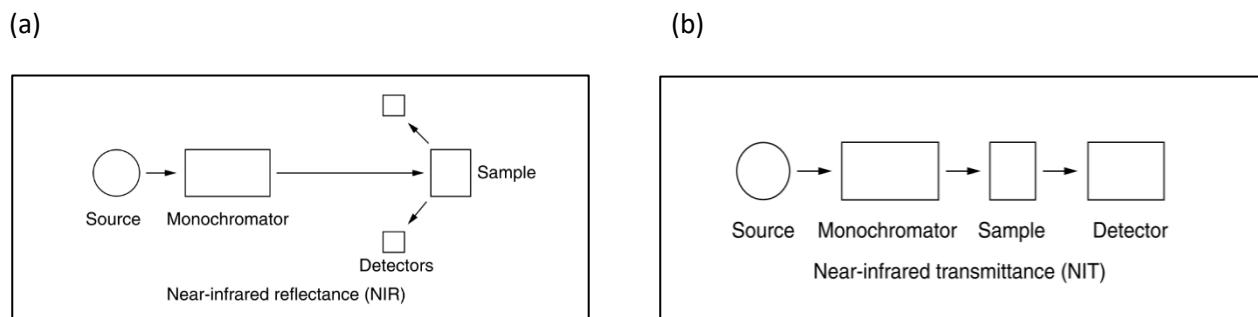


Figure 2.3 (a) Near infrared reflectance, and (b) near infrared transmittance basic instrument configurations (Siesler, 2008a).

2.5.1.4 Interactance mode

Samples can also be measured in interactance mode where a higher probability is given to the incident beam to interact with the sample (Pasquini, 2003). In order to collect spectra in interactance mode, a special instrument needs to be used. The interactance instrument is a type of device where the instrument and the sample comes into contact via fibre optics, with the signal being transmitted back to the detector along the fibre optic cable (Williams, 2006). The application of fibre optics with reflectance instruments can be utilised in order to scan the sample from a distance. The transmission efficiency is dependent on the distance. A longer wavelength range shortens the effective distance for practical purposes due to the increased risk of absorbers interfering with the transmission of spectral data (Williams, 2006). When the instrument is implemented with a small wavelength range, the data can be transmitted over numerous meters and even further when using a few selected wavelengths. Due to the sensing head spanning a small area, several scans should be taken of the sample and then averaged. Alternatively, the sample may be placed on a turntable or belt, ensuring that several scans are taken.



Figure 2.4 Typical setup of the VIS-NIR Flame spectrophotometer (Ocean Optics, Inc., Florida, USA).

2.5.1.5 Filter-based instruments

The instruments based on Acousto-Optical Tunable Filters (AOTF) are modern scan spectrophotometers (Pasquini, 2003) that uses radio-frequency signals for wavelength selection. The refractive index of a birefringent crystal (generally TeO₂, suitable for NIR regions) is altered by the signals to transmit light of a given wavelength (Blanco & Villarroya, 2002). AOTFs are capable of reaching extremely high scan speeds over a broad spectral region due to the absence of moving parts. Therefore, AOTFs are characterised by their reliability and reproducibility of wavelength scans, making these instruments suitable for difficult measurement conditions, as experienced in production plants (Blanco & Villarroya, 2002).

2.5.1.6 LED based instruments

In discrete-wavelength spectrophotometers, light source filters or light-emitting diodes (LEDs) are used to select a few wavelengths for specific molecules for distinct applications (Blanco & Villarroya, 2002). LED-based instruments are known to be straightforward and robust, due to the absence of moving parts, making them ideal for use in portable devices for in-field analysis (Blanco & Villarroya, 2002; Pasquini, 2003).

2.5.1.7 Dispersive instruments

The first infrared instruments were dispersive spectrophotometers, comprising of three fundamental parts: a radiation source, a monochromator and a detector (Abbas *et al.*, 2012). A broad radiation spectrum is dispersed by the monochromator, resulting in a series of electromagnetic energy bands covering a specific wavelength range. Dispersive components (prisms or gratings) are used jointly with variable-slit mechanisms, mirrors and filters. Two types of detectors are implemented in this instrumentation: thermal detectors and photon detectors. The thermal and photon detector measures the heating effect produced by infrared radiation as well as the resultant interaction that occurs between the infrared radiation and the semiconductor, respectively (Abbas *et al.*, 2012). Thermal detectors provide a linear response over a wide range of frequencies, whereas photon detectors have shorter response times and higher sensitivity. Therefore, in dispersive instruments the radiation reflects or passes through the sample and is dispersed by a monochromator into component frequencies. By directing the beam onto the detector, it is possible for the instrument to produce an electrical signal which is then converted into a recorded spectral response.

2.5.1.8 Interferometric instruments

Another type of instrument is spectrophotometers based on the use of interferometers and Fourier transform. The interferometer's spectral data is pre-processed using Fourier transform (FT) (Fourier, 1822), and the instruments are referred to as FT-NIR instruments. In FT-NIR instruments, all wavelengths reach the detector simultaneously and are processed with FT to obtain a typical NIR spectrum (Nielsen, 2010; Abbas *et al.*, 2012). In the interferometer, the NIR beam is split and then recombined, by reflecting the split beams

back, with the use of mirrors (**Fig. 2.5**). Since all the wavelengths are measured at once, the spectra are acquired rapidly, with an improved signal-to-noise ratio.

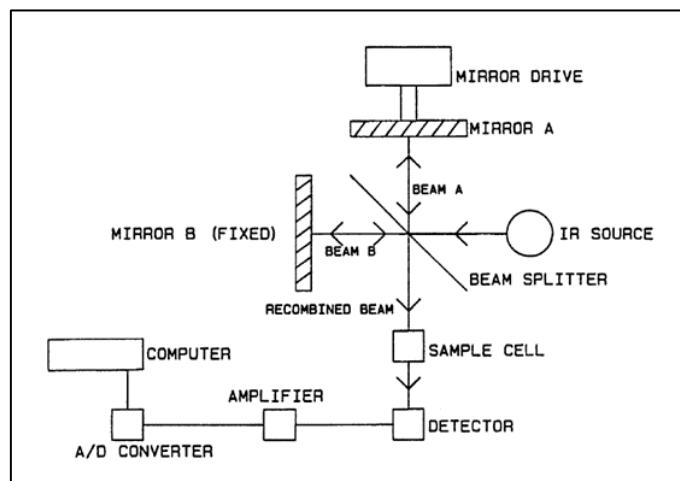


Figure 2.5 Block diagram of an interferometer and related electronics, generally used in a FT-NIR instrument (Nielsen, 2010).

2.5.1.9 Benchtop and portable handheld devices

Advances in NIR spectroscopy instrumentation have led to the development of miniaturized spectrometers, offering a wide range of applications (Zamora-Rojas *et al.*, 2012). These portable handheld devices make it possible to obtain instantaneous results at any location. Benchtop and portable handheld devices (**Fig. 2.6**) differ in various ways such as size, resolution and the wavelength range of the instrument (Henn *et al.*, 2016). Although the use of portable spectrometers have some advantages, such as ease of use and a wide range of applications, it lacks in other areas such as high-end performance, resolution and individualisation. Many studies have been done to compare the performance of these two devices (Herberholz *et al.*, 2010; Zamora-Rojas *et al.*, 2012; Plans *et al.*, 2013; Wilkerson *et al.*, 2013; Henn *et al.*, 2016). These studies determined that the results obtained, using portable devices, are comparable to those achieved by benchtop spectrometers, under laboratory conditions. Therefore much work and research is needed to evaluate the performance of these portable devices under production conditions (reviewed by Dos Santos *et al.*, 2013).

(a)



(b)



Figure 2.6 (a) Buchi NIRFlex N-500 (benchtop FT-NIR spectrometer) (NIRsolutionsTM, Flawil, Switzerland), and (b) MicroNIR OnSite spectrophotometer (portable handheld NIR spectrometer) (Viavi Solutions Inc., Milpitas, USA).

2.6 Chemometrics and multivariate data analysis (MDA)

The characteristically broad and extensively overlapped bands of NIR spectra contains analytical information which is non-selective and influenced by various physical, chemical and structural variables (Blanco & Villarroya, 2002). Additionally, sample variation may result in slight spectral differences that are indistinguishable with the naked eye. Therefore, chemometrics is required to extract all the relevant information obtained using NIR spectroscopy (Blanco & Villarroya, 2002).

Chemometrics is a chemical discipline which utilises mathematics, statistics and computational devices. This technique is used to select the optimal procedure for measurement of certain characteristics and allows for effective design of experiments to provide the most relevant chemical information (Massart *et al.*, 1988; Brown, 1995; Bokobza, 1998). Multivariate data analysis (MDA) is another frequently used term and is often confused with chemometrics. Chemometrics is the science, whereas the application of specific chemometric techniques is called MDA. In order to relate properties, being either physical or chemical, of the sample to energy absorption in the NIR wavelength range, the chemometric approach is commonly utilised (Bokobza, 1998).

Chemometrics involves the extraction of meaningful information about the *objects* (samples) and *variables* (measurement results) from a *data matrix* (data table comprising of K variables determined for I objects), as illustrated in **Figure 2.7** (Geladi, 2003). This implies that there is a significant amount of redundant information within the matrices, which can be reduced to obtain only the relevant data. The purpose of chemometrics is therefore to decompose these frequently oversized matrices, resulting in a reduced number of components, that are more manageable for the purpose of data analysis. The reduced terms are easier to comprehend, more stable and the residuals consist of noise and/or less valuable information (Geladi, 2003). The multivariate techniques that are used the most often allows for grouping of samples that have similar characteristics for two purposes, namely; for the establishment of classification

methods for an unknown sample (qualitative analysis), or to establish methods that have the ability to define a specific characteristic or property of a sample (quantitative analysis) (Blanco & Villarroya, 2002). The chemometric approach can therefore be broken down into two phases, spectral pre-processing and spectral processing (Brereton, 2003a).

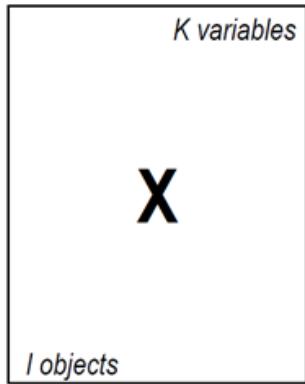


Figure 2.7 A data matrix of size $I \times K$

2.6.1 Pre-processing

Spectroscopic data often contain disturbances caused by scattering due to surface inhomogeneities (Williams, 2009). Pre-processing is generally performed so as to eliminate specific non-chemical biases from the acquired spectra and to prepare the data before the next processing steps, in order to develop simple and robust models (Blanco & Villarroya, 2002). A vast majority of techniques exist and the most widely used are multiplicative scattering correction (MSC) (Geladi *et al.*, 1985), standard normal variate (SNV) (Barnes *et al.*, 1989), de-trending (DT) and derivatives including Savitzky-Golay (Savitzky & Golay, 1964).

The foundation of MSC is that it allows for the separation of chemically absorbed light from scattering caused by physical disturbances. It accomplishes this because chemically absorbed light has different scattering characteristics, based on the wavelength dependencies, when compared to the characteristics of scattering from physical disturbances (Geladi *et al.*, 1985; Isaksson & Næs, 1988). Subsequently, discriminating between absorption and scatter is possible, by means of data from various wavelengths. The scatter for an ideal sample is calculated and compared to the individual scattering for each sample and each spectrum is corrected to produce a uniform scattering for all the spectra compared to the ideal.

The correction is executed based on the assumption that the scatter coefficients are the same among all the samples at every wavelength in the NIR region. In MSC, the mean spectrum is calculated from all the spectra in a defined data set (Pizarro *et al.*, 2004). Subsequently, the absorbance values of the sample are subjected to a least square linear regression and these values are compared to the sample spectrum at the specific wavelength in the mean spectrum. This function provides a linear equation (equation 2.1), with a fixed intercept and slope (Maleki *et al.*, 2007):

$$x_{ik} = a_i + b_i \bar{x} + e_{ik} \quad \dots\text{equation 2.1}$$

Where x_i represents an individual spectrum, \bar{x} the mean spectrum and e_{ik} the residual spectrum, which is representative of the chemical information in the data set.

Subsequently, the intercept value is subtracted from each data point in the spectrum. MSC therefore rotates each spectrum in order to fit it as closely as possible to the mean spectrum (Osborne *et al.*, 1993). Eventually, each absorbance value in the resultant spectra is divided by the slope value, as shown in equation 2.2 (Maleki *et al.*, 2007):

$$x_{ik}(\text{new}) = \frac{[x_{ik}(\text{old}) - a_i]}{b_i} \quad \dots\text{equation 2.2}$$

The constants a_i (intercept) and b_i are used for the correction of each optical value (e.g. reflectance) of the spectrum x_{ik} . MSC endeavours to remove additive and multiplicative effects caused by scattering of light, thereby decreasing the variation in the spectra that cannot be attributed to the concentration of the analyte in question.

Standard normal variate (SNV) (Barnes *et al.*, 1989) centres and scales discrete spectra, having a similar effect to that of MSC (Osborne *et al.*, 1993). SNV, however, has one main practical difference, each spectrum is standardised using the data exclusively from that spectrum. In other words, not using the mean spectrum of any set. The spectra are converted arithmetically by calculating the standard normal variate (equation 2.3) at every wavelength, removing the deviation from the slope discreetly for each sample

$$SNV = \left(\frac{x_i - \bar{x}_i}{s_i} \right) \quad \dots\text{equation 2.3}$$

Where x_i represents the spectral measurement from the entire spectrum, \bar{x}_i the mean of each spectrum and s_i the standard deviation of each spectrum.

De-trending (DT) accounts for the variation in baseline shift and curvilinearity (Barnes *et al.*, 1989), where the effects of these offsets can be removed, with the use of a polynomial regression. The de-trend function subtracts the mean or best-fit line from the data set. Removing a trend from the data set allows the analysis to be focussed on the variations in the data around the trend. Even though trends can be meaningful, some types of analyses provide better understanding once these trends are removed. The removal of trend effects from the data often depends on the objectives of the analysis. De-trending often includes SNV transformation, but if the interest is mainly focussed on shape differences in the spectra then de-trending may be carried out in the absence of SNV (Barnes *et al.*, 1989).

Derivatives are common transformations applied to spectral data (Esbensen *et al.*, 2002). The first or second derivatives are functional pre-processing techniques to eliminate noise. The first derivative is used for correcting baseline shifts and the second derivative is an alternative to handling scatter effects (Esbensen *et al.*, 2002). Derivatives can also be calculated based on the Savitzky-Golay smoothing method (Savitzky &

Golay, 1964). This method fits a simple polynomial to a running local region of the sample vector (Kenneth *et al.*, 1998). A window width is selected and the data is used to calculate the smoothed value in the middle of the window. The centre point is then replaced with the polynomial estimate of that point. In the case of derivatives, this centre point is replaced with the derivative of the polynomial at that point. When using derivatives, it is critical to consider the window width for the polynomial fit (Kenneth *et al.*, 1998). Using too small a window size might cause too little smoothing to take place, resulting in derivatives with poor signal-to-noise. Then again, using too large a window size, might smooth out features. Therefore, the optimal window size is dependent on the data, as smoothing away features may or may not be disadvantageous to the primary analysis (Kenneth *et al.*, 1998). When selecting a window width, one should always consider the noise level, the number of data points and the sharpness of the features. A reasonable window width and polynomial order is selected by applying several combinations to optimise the pre-processing conditions and evaluate the final results.

2.6.2 Spectral processing

Prior to exploring and understanding the numerous techniques employed in chemometrics, a few definitions as well as notations must be stated. A *matrix* is a rectangular array of numbers and functions (Gorsuch, 1974). The use of a capital letter signifies an entire array.

Listed is the applicable nomenclature often used:

- x Single scalar vector
- \mathbf{x} Vector of scalar values
- \mathbf{X} Two-dimensional matrix of scalar values
- $\mathbf{\underline{X}}$ Three-dimensional array of scalar values
- \mathbf{X}^T The transpose of the matrix \mathbf{X}

Multivariate data analysis (MDA) involves the extraction of information from a matrix or three-way array (Geladi, 2003). The most applicable chemometric technique is determined by different goals, data generation methods and the nature of the sample. Data exploration is the first step in MDA. This fundamentally involves familiarising oneself with the data by means of examining it for capturing phenomena without any preconceived notions or conclusions. This allows for the detection of certain clusters or clustering of objects and the identification of outliers and gradients between clusters. The aforementioned is the fundamental principle of a powerful and most frequently used chemometric technique known as principal component analysis (PCA) (Cowe & McNicol, 1985). PCA is an effective tool for exploration of data as it places the large $I \times K$ matrix into a more confined space, allowing for easier analysis (Jolliffe, 1986). PCA is accepted to fall into the unsupervised classification category.

2.6.3 Principal component analysis (PCA)

PCA is the most basic “workhorse” of all MDA (Esbensen *et al.*, 2002). The primary focus of PCA is to analyse a data set containing many variables that are interrelated and in turn reduce the dimensionality of this data set, whilst preserving the inherent variation (Cowe & McNicol, 1985; Jolliffe, 1986; Massart *et al.*, 1988; Bokobza, 1998; Esbensen *et al.*, 2002; Geladi, 2003; Reich, 2005; Roggo *et al.*, 2007). This is attained by transformation to new variables that are called principal components (PC’s). The variables are uncorrelated and ordered in such a manner so that the utmost variation, present in the original variables, is retained by the first few PC’s (Jolliffe, 1986; Massart *et al.*, 1988; Bokobza, 1998; Esbensen *et al.*, 2002; Geladi, 2003; Reich, 2005; Roggo *et al.*, 2007) as depicted in equation 2.4:

$$\text{PCA: } \mathbf{X} = \mathbf{t}_1 \mathbf{p}'_1 + \mathbf{t}_2 \mathbf{p}'_2 + \dots + \mathbf{t}_A \mathbf{p}'_A + \mathbf{E} \quad \dots\text{equation 2.4}$$

where \mathbf{X} represents a matrix ($I*K$), generally mean-centered, \mathbf{t}_A signifies the *score* values for the a^{th} component, \mathbf{p}'_A the *loading* values for the a^{th} component and \mathbf{E} the *residual* matrix ($I*K$). The components are generally referred to as *latent variables* or, more commonly, *principal components*. A schematic representation of equation four can be observed geometrically, as shown in **Figure 2.8**.

Once data decomposition has occurred, line and scatter plots can be compiled to visualise the scores and loadings which allow the data space to be effectively interpreted. It is essential to review the scores and loading plots when exploring the data. The scores plots illustrate the correlation between the objects, whilst the loadings give insight into the variation that occurs within the variables. The scores and loadings are concurrently used in order to understand the essence of the data, for instance the cause of the variation and the driving component. This phenomenon is particularly useful in spectroscopy, as the loading line plot indicates the variable(s) (wavelength) accountable for the variation. Therefore, the presence of variation in the samples is indicated by the score plot while the loading line plot indicates the variable responsible.

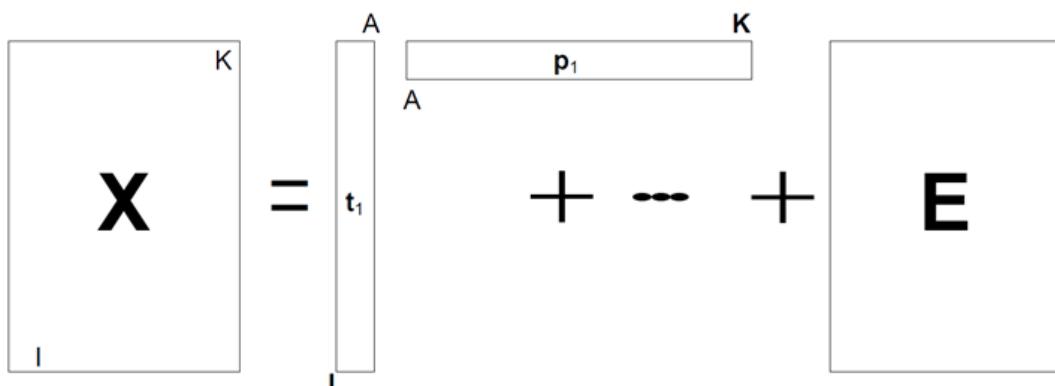


Figure 2.8 A data matrix sized ($I*K$) is condensed to smaller sized matrices ($I*A$) and ($A*K$) that are easier to interpret and comprehend in addition to containing all the relevant information. Noise and other disturbances are left in the residual matrix of size ($I*K$). [Adapted from (Esbensen *et al.*, 2002; Geladi, 2003)].

PCA is therefore the first step in MDA (Wold *et al.*, 1987a), followed by further calculations on the new variable space. Due to its success in dimensionality reduction, it is an outstanding tool for NIR spectral data analysis and has been evaluated in various applications (Cowe & McNicol, 1985; Sato, 1994; Brigger *et al.*, 2000; Chang *et al.*, 2001; Cozzolino & Murray, 2004; He *et al.*, 2006; He *et al.*, 2007; Mishra *et al.*, 2015).

Although PCA is a very successful technique, the one disadvantage of the loadings line plot is that the loadings are assigned arbitrary values which are either positive or negative (Williams, 2009). This requires complex interpretation of these plots as it may not be clear when to analyse the peaks or troughs when deciding on the essential absorption areas. On the other hand, when classifying samples, it is often necessary to differentiate between classes using models built from unsupervised classification techniques. This is followed by supervised classification, a technique that classifies data according to known information. The most frequently used techniques, in discrimination studies, are soft independent modelling of class analogy (SIMCA) (Wold & Sjöström, 1977), K-nearest neighbour (KNN) (Sebestyen, 1962), linear discriminant analysis (LDA) or partial least squares discriminant analysis (PLS-DA) (Wold *et al.*, 1987b) and support vector machines (SVM) (Cortes & Vapnik, 1995).

2.6.4 Soft independent modelling of class analogy (SIMCA)

In the 1970s, Svante Wold presented a classification method known as soft independent modelling of class analogies (SIMCA) (Wold & Sjöström, 1977). SIMCA is a classification method designed based on spectral similarities. This method is effective when classifying high-dimensional observations since it incorporates PCA for dimension reduction (Branden & Hubert, 2005). Classification models are built using PC scores to model the shape and position of the objects formed by the samples in row space, for class definition. Construction of a multidimensional box is a requirement for each class, where after the classification of a future sample is performed by calculating in which box (class) the sample lies (**Fig. 2.9**).

The goal of this classification method is to assign new objects to the class to which they show the largest similarity rather than identical behaviour. In this context similarity has a geometrical meaning and is measured by the distance between objects (Massart *et al.*, 1988). Therefore, a small distance means a high similarity. This approach specifically allows the objects to display intrinsic individualities as well as their common patterns, but only common properties of the classes are modelled.

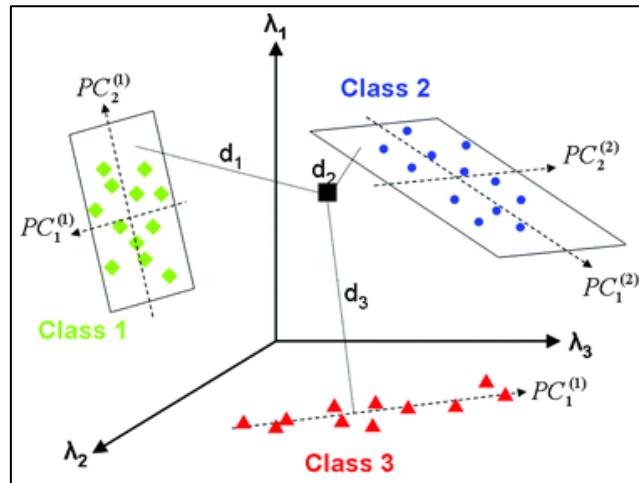


Figure 2.9 Principle of the SIMCA method. Each symbol represents a spectrum. Three classes are independently modelled by principal component analysis (Sirven *et al.*, 2007).

2.6.5 K-nearest neighbour (KNN)

Fix and Hodges (1951) presented a non-parametric method for pattern classification, generally known as the k-nearest neighbour rule (Fix & Hodges, 1951). K-nearest neighbour (KNN) is a fairly simple, yet fundamental method of classification. This discriminant analysis technique was developed for when determination of dependable parametric estimates of densities are difficult or not known (Peterson, 2009). K-nearest neighbour is performed on the PC scores. The classification of an unknown is performed by calculating the distance between the unknown and a set of samples with known class membership (training set). The most commonly used distance measurements are given by Euclidean or alternatively Manhattan and Mahalanobis distances (Massart *et al.*, 1988; Brereton, 2003b). The Euclidean distance is calculated between samples k and l in multiple measurements as:

$$d_{kl} = \sqrt{\sum_{j=1}^J (x_{kj} - x_{lj})^2} \quad \dots\text{equation 2.5}$$

where x_k and x_l are the sample coordinates of k and l in the j th measurement of the row space (Massart *et al.*, 1988; Brereton, 2003b). The Manhattan distance is somewhat differently defined to the Euclidean distance and is given by:

$$d_{kl} = \sum_{j=1}^J |x_{kj} - x_{lj}| \quad \dots\text{equation 2.6}$$

The difference between the Euclidean and Manhattan distances are illustrated in **Figure 2.10**. As depicted the Manhattan distance will always be greater than the Euclidean distance.

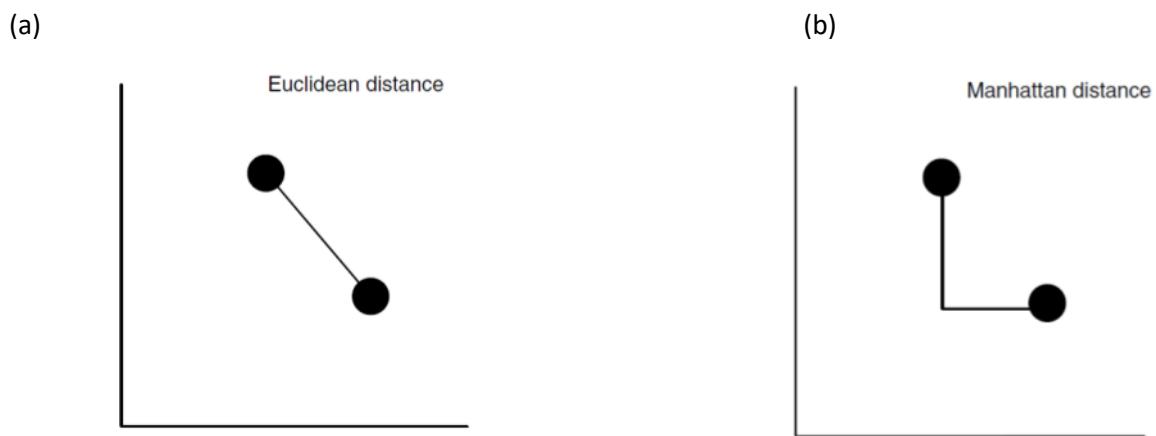


Figure 2.10 Schematic representation of (a) Euclidean distance and (b) Manhattan distance (Brereton, 2003b).

The Mahalanobis distance method is very popular in chemometrics (Brereton, 2003b). Although Mahalanobis is superficially similar to the Euclidean distance, it emphasizes some measurements over others (Massart *et al.*, 1988) and takes into account that some variables may be correlated (Brereton, 2003b). The distance between objects k and l is best defined in matrix terms by:

$$d_{kl} = \sqrt{(x_k - x_l) \cdot C^{-1} \cdot (x_k - x_l)'} \quad \dots\text{equation 2.7}$$

where C represents the variance-covariance matrix of the variables. This measure differs from the Euclidean distance where the inverse of the variance-covariance matrix is inserted as a scaling factor (Brereton, 2003b). However, the application of this method is not as straight forward and can only be applied where the number of objects exceeds the number of measurements (variables). Otherwise the variance-covariance matrix would not have an inverse due to insufficient degrees of freedom for measurements.

Literature shows several other related distance measures that are used to determine the proximity between samples in a row space. Euclidean distance is often used in conjunction with KNN and a classification is made using the closest K samples (**Fig. 2.12**).

To select the optimal K, a cross-validation procedure is applied to a set of data with known class identities. Each sample in the training set is treated as an unknown and is classified using the remaining training set samples. This is repeated using different numbers of nearest neighbours (K) for the classification. A confusion matrix is used to describe the optimal K and performance of a classification model (**Fig. 2.11**).

$n = 125$	Predicted: NO	Predicted: YES	
Actual: NO	$TN = 10$	$FP = 10$	20
Actual: YES	$FN = 5$	$TP = 100$	105
	15	110	

Figure 2.11 Example of a confusion matrix where **TP** is the true positive, **TN** the true negative, **FP** the false positive and **FN** the false negative.

In **Figure 2.12**, there are three classes with the ultimate goal being to classify the unknown sample, X_u . A value of $k = 5$ neighbours was used along with the Euclidean distance. Four of the five closest neighbours belonged to ω_1 and the other belonged to ω_3 . The unknown sample, X_u was therefore assigned to the predominant class, ω_1 .

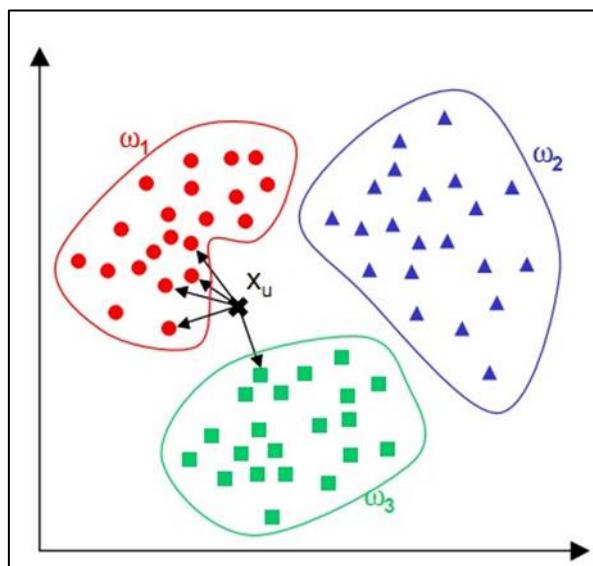


Figure 2.12 The simulation of the k -nearest neighbour method, $k = 5$. Each class is represented by a symbol and independently modelled by principal component analysis (Zhang, 2017).

2.6.6 Linear discriminant analysis (LDA) and partial least squares discriminant analysis (PLS-DA)

NIR spectroscopy is perfectly suited for discriminant studies. Products are compared with each other, to identify irregularities or impurities by means of classification (Brereton, 2003c; Williams, 2013). Linear discriminant analysis (LDA) works by calculating an optimal linear projection which maximises the inter-class variance while at the same time minimising the intra-class variance (Fisher, 1936). Objects (samples) are classified by calculating the distance to the centre of each class. Objects are then assigned to the class with

the smallest distance. Partial least squares discriminant analysis (PLS-DA) is similar to partial least squares (PLS) regression, as it uses the latent variable approach to find fundamental relations between two matrices (\mathbf{X} and \mathbf{Y}) (Wold *et al.*, 2001a; Wold *et al.*, 2001b; Esbensen *et al.*, 2002). PLS uses the y-data structure to decompose \mathbf{X} so that the outcome constitutes an optimal regression vector (Williams, 2013). PLS-DA works similarly, but instead of measuring y-data, dummy variables are used which are indicators of groups (Westerhuis *et al.*, 2008). The objective of this method is to successfully predict group membership, hence the classification of spectra to classes. During the developmental procedure of the discriminant models, a calibration matrix is constructed using every sample by assigning random values to dummy variables. Therefore, the calibration matrix (\mathbf{Y}) represents the class memberships using numbers such as ones and zeros, and is then paired with the training set (\mathbf{X}). Once completed, PLS is applied in the typical way (Fig. 2.13) (Barker & Rayens, 2003).

For example, if the sample spectrum belongs to the pre-defined correct group then a value of one would be assigned to that spectrum. A zero value would indicate that the spectrum does not belong to the correct group. Often a threshold of 0.5 is used, and predicted values above this threshold will be classified as a one and values below 0.5 will be classified as a zero. Misclassification occurs when a sample is classified as an incorrect group, therefore receiving an incorrect binary value (1 or 0). PLS-DA generally outperforms LDA, when the classes are closely related, because it overcomes the collinearity problems often associated with LDA.

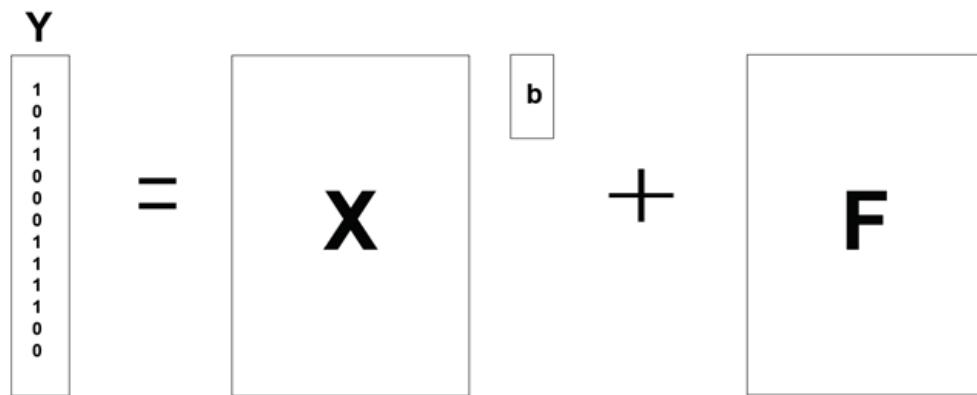


Figure 2.13 Diagrammatic depiction of the PLS-DA method. \mathbf{Y} , a dummy matrix consisting of ones and zeros (representing class membership), is paired with \mathbf{X} to obtain \mathbf{b} , the regression coefficients and the residual, \mathbf{F} .

2.6.7 Support vector machines (SVM)

Support vector machines (SVM) was invented in 1963 by Vladimir Vapnik and was initially designed for separating hyperplanes in pattern recognition problems, used for linear threshold classifiers (Vapnik, 1963). Later the method was generalized for constructing non-linear separating functions (Cortes & Vapnik, 1995) and has subsequently been introduced for non-linear NIR calibrations (Agelet & Hurlburgh, 2010). Currently,

SVM is extensively used among data scientists and applied to an incredible range of problems, from medical diagnosis to image recognition and textual classification (Mueller & Massaron, 2016).

Although SVM consists of complex mathematics, the method was initiated by a simple idea. Known as a learning machine for two-group classification problems (Cortes & Vapnik, 1995), SVM is performed by separating two groups with one line (**Fig. 2.14**) (Mueller & Massaron, 2016).

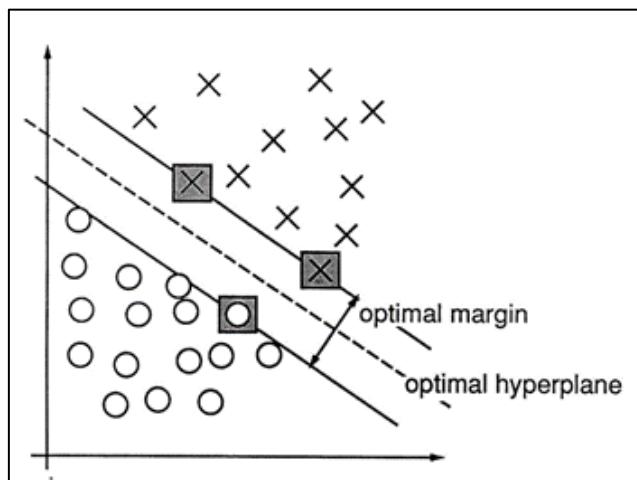


Figure 2.14 An example of SVM and linear separating hyperplanes. The support vectors, marked with grey squares, define the margin of largest separation between the two classes (Cortes & Vapnik, 1995).

SVM divides the data set into two groups by placing a *separating line* (the line with the largest margin) in the middle of the *margin* (the empty space between the boundaries of the classes). This phenomenon is described as *maximum margin* or *optimal margin hyperplane* and therefore indicates the maximum separation between the two groups. The data points near the line and thus forming part of the boundaries are known as the *support vectors* (Mueller & Massaron, 2016). The support vectors are used as training patterns and convey all relevant information about the classification problem (Hearst *et al.*, 1998). The SVM approach is thus quite simple and ultimately based on an optimal distance calculation between data points and the hyperplane (Mueller & Massaron, 2016).

SVM is an optimization technique, where *kernel functions* are used for nonlinear separating hyperplanes to evaluate the robustness of the classifier model. A kernel is a similarity function that requires two inputs and calculates how similar they are. These kernel functions include linear, radial basis function (RBF), polynomial and sigmoid. In practice, RBF is commonly used since it is faster than other kernels. The RBF works in a simple, effective way and can map and approximate almost any nonlinear function (Mueller & Massaron, 2016). RBF uses normal curves around the data points, creating a margin around every support vector (**Fig. 2.15**). Additionally, this kernel is adaptable to different learning strategies, resulting in a bended or smoother curve hyperplane.

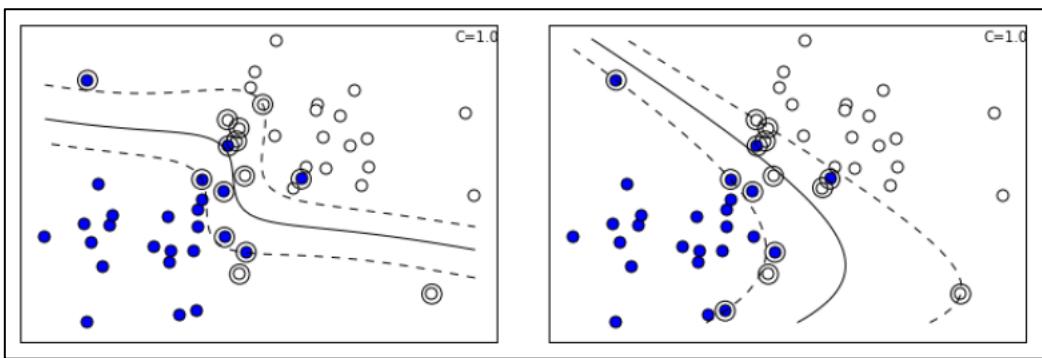


Figure 2.15 An RBF kernel using diverse hyper-parameters in order to create unique SVM solutions (Mueller & Massaron, 2016).

Compared to RBF, the polynomial and sigmoid kernels are less adaptable and display more favouritism (Mueller & Massaron, 2016). If the sigmoid features a single bend, the polynomial function can have as many bends or separated hyperplanes as its set degree. Keeping in mind that a higher degree takes longer to compute. A graphic representation is provided (Fig. 2.16), illustrating what can be done with the polynomial and sigmoid kernels.

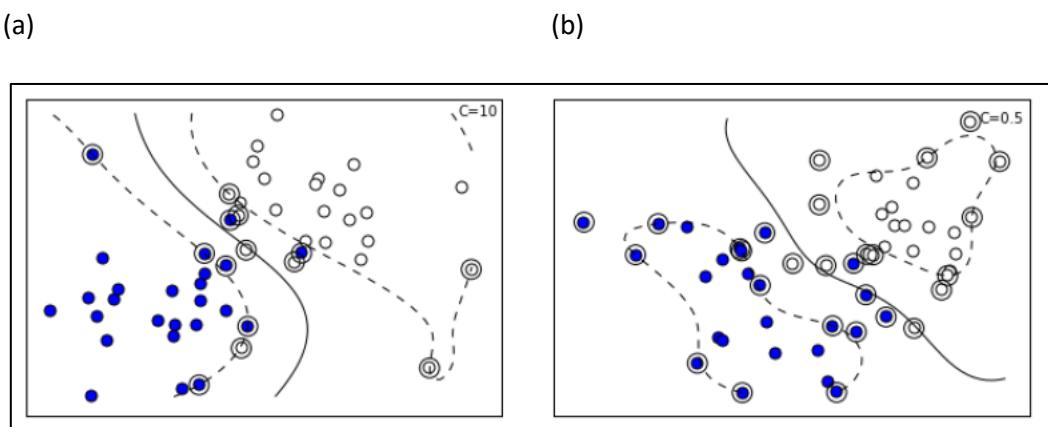


Figure 2.16 (a) A sigmoid kernel, and (b) a polynomial kernel applied to the same data (Mueller & Massaron, 2016).

Even though SVMs are intricate learning algorithms, they are easily executed by a classification software function. Therefore, SVM has been recognized as a powerful method for learning certain structures from data for prediction.

Chemometrics is therefore a powerful data reduction scientific tool, used for qualitative and quantitative data analysis. Qualitatively it is used for the classification or grouping of unknown samples, with similar characteristics and quantitatively for the determination of analytes in samples.

2.7 Applications of NIR spectroscopy in meat authentication

In the food industry, NIR spectroscopy has been recognised for its ability as a rapid and non-destructive method that requires small amounts of sample, with minimal sample preparation and handling. In food and agricultural applications, it has gained much attention as an instrument for on/in-line monitoring of foods and beverages (reviewed by Huang *et al.*, 2008). NIR spectroscopy has shown potential to predict the quality of meat and meat products (reviewed by Prieto *et al.*, 2009), and in meat science it was successfully applied for authentication purposes, without the need of any chemical analysis (Rannou & Downey, 1997; Ding & Xu, 1999; Alomar *et al.*, 2003; Cozzolino & Murray, 2004; Prieto *et al.*, 2009; Meza-Márquez *et al.*, 2010; Morsy & Sun, 2013; Alamprese *et al.*, 2016). The technique is well-established as a means of quality assessment and prediction in the food industry (Williams, 2013). The applications of NIR spectroscopy in meat science are numerous. The general information of selected studies reviewed is given in **Table 2.3**.

Much work was done regarding the identification and discrimination of different species as well as the detection of fresh and frozen-thawed meat. Visible-, near- and mid-infrared spectroscopy was successfully used to rapidly discriminate between three types of uncooked meat, namely; pork, chicken and turkey (Rannou & Downey, 1997). Separate discriminant models, using principal component analysis (PCA) and factorial discriminant analysis (FDA), were initially developed in the mid-IR and visible-near-IR regions. The best predictive model correctly classified 86.5% of the samples using mid-IR data (Rannou & Downey, 1997), while for the visible-near-IR data an optimum classification of 91.9% was achieved. Alternatively, the researchers have shown that using combined visible-, near- and mid-IR data may produce the most accurate discriminant models, with a classification accuracy of 94.6%. However, this is not feasible in practice, as these combination instruments are very expensive.

In similar studies a combination of visible-near-infrared spectroscopy was used to discriminate between beef and kangaroo meat (Ding & Xu, 1999), and beef, pork, chicken and lamb meat (Cozzolino & Murray, 2004). The classification models developed in both these studies performed well. Ding and Xu (1999) reported that NIR spectroscopy, in conjunction with canonical discriminant analysis (CDA) and stepwise multiple linear regression (MLR), could differentiate between beef and kangaroo meat with an overall classification accuracy of up to 100%. Using the visible (Vis) and NIR region from 400 to 2500 nm, Cozzolino and Murray (2004) reported that they were able to identify and authenticate meat from different species. Different meat species were compared using both PLS regression and PCR. However, these two methods are generally used for quantification purposes and not classification. Alternative classification techniques such as LDA, SIMCA and PLS-DA are routinely used and would have been the correct techniques. However, quantification techniques can be used to do classification if the classification is based on quantities such as moisture, protein and lipid content. The PLS regression method was superior to the PCR method, giving the best classification results. Both classification methods were performed in the three wavelength segments (400 – 700; 1100 – 2500; 400 – 2500 nm) used. The PCR method achieved the best classification, using the NIR region, with an accuracy of 94%. For the PLS method, the best classification was achieved in

the combined Vis + NIR region, correctly classifying 96% of the samples (Cozzolino & Murray, 2004). These results indicate the potential of Vis- and NIR spectra as means of rapid authentication and identification of different meat species.

The technique has also been successfully used to detect and quantify adulterations in meat and minced beef (Ortiz-Somovilla *et al.*, 2005; Gaitán-Jurado *et al.*, 2008; Meza-Márquez *et al.*, 2010). In a study by Meza-Márquez *et al.* (2010), Mid-Fourier transform infrared (MID-FTIR) spectroscopy and chemometric methods were used to rapidly detect adulterants in minced beef. Mixtures of minced meat and adulterants (horse meat, fat beef trimmings and textured soy protein) were prepared and calibration models were built for each adulterant. Chemometric methods, PCA and PLS regression, were developed to detect and quantify the adulteration of minced beef with the adulterant (Meza-Márquez *et al.*, 2010). The results showed that the chemical composition of the samples influenced the spectra, therefore, presenting some significant internal chemical characteristics of the minced meat mixtures. Good correlations between the absorbance spectra and the percentage of adulteration were obtained (Meza-Márquez *et al.*, 2010). Additionally, a method using SIMCA was developed to discriminate between samples that were adulterated and samples that were unadulterated. SIMCA models of the MID-FTIR spectra were effective for the discrimination of adulterated from unadulterated samples and the resulting classification accuracy was 100%.

Schmutzler *et al.* (2015) developed and compared three different methods based upon NIR to detect the adulteration of veal sausage with pork meat. These methods include: a FT-NIR benchtop device fit for laboratory use, a fibre optic probe suitable for industrial in- and on-line application and a handheld spectrophotometer ready for on-site analysis. Different percentages of pork and pork fat were used as adulterant in the veal sausages. Calibration models were built for each adulteration percentage and analysed, respectively. Subsequent measurements were performed through quartz cuvettes as well as directly through the polymer packaging. PCA models were developed for each setup and the scores were used as input data for support vector machines (SVM) classification and validation (Schmutzler *et al.*, 2015). The results showed that meat and fat adulterations were detected at the lowest levels of contamination in this experiment (10%) when applying the laboratory and industrial setup. For both setups a classification accuracy of 100% was obtained when using quartz cuvettes. The classification accuracy for the industrial fibre optics setup decreased when measurements were acquired by scanning the sample directly through the packaging. This decrease was attributed to the influence of the polymer packaging, thus resulting in a classification accuracy of 91.7%. The handheld device was able to discriminate between unadulterated and adulterated samples, with an adulteration level of 10% for meat contamination. The use of quartz cuvettes resulted in a 100% classification accuracy. However, the same was not observed for measurements performed directly through the polymer packaging. The SVM classification accuracy decreased to 83.3% due to losses in signal intensity and resulting overlap of the PCA clusters (Schmutzler *et al.*, 2015). Regarding the fat adulteration, the handheld device was less successful for detecting low levels of contamination (10%). The classification accuracy for measurements performed through the quartz cuvette and polymer packaging

was 83.3 and 75%, respectively. However, the classification accuracy improved and was 100% successful with adulteration levels between 20 and 40%. This represents a complete and well-designed study as it incorporates all aspects of the meat industry, from the development of the method at lab-scale, to the use of NIR for on-site authentication as well as at point of sale, through the packaging.

The rapid identification and quantification of minced beef that had been adulterated with turkey meat was evaluated using Fourier transform near-infrared (FT-NIR) spectroscopy and multivariate analysis (Alamprese *et al.*, 2016). In this study the samples were analysed as raw, frozen-thawed and cooked, using different multivariate regression and class-modelling strategies. PLS regression models were developed that predicted the level to which turkey meat was added to minced beef. These models performed well, with a prediction R^2 higher than 0.884 and RMSEP lower than 10.8% (Alamprese *et al.*, 2016). PLS regression models were thus able to accurately predict adulteration levels higher than 20%. PLS-DA models were developed and applied for the classification of each sample type (fresh, frozen-thawed, cooked) into two classes. The classes were separated by the pre-defined adulteration threshold of 20%. PLS-DA showed prediction values of sensitivity and specificity higher than 0.84 and 0.76, respectively (Alamprese *et al.*, 2016). PLS-DA models were thus able to accurately distinguish between low (<20%) and high ($\geq 20\%$) levels of adulteration. Alamprese *et al.* (2016) therefore demonstrated the potential of FT-NIR spectroscopy, for the rapid identification and quantification of adulterated minced beef in samples irrespective of them being fresh, frozen-thawed or cooked.

The suitability of NIR, as a rapid and objective technique, was successfully evaluated to discriminate fresh from frozen-thawed meat (Downey & Beauchêne, 1997b; Downey & Beauchêne, 1997a; Thyholt & Isaksson, 1997; Ropodi *et al.*, 2018). The use of combined visible-near infrared reflectance spectroscopy and two chemometric techniques were evaluated for the rapid classification of meat samples as either fresh or frozen-thawed (Downey & Beauchêne, 1997b). Factorial discriminant analysis (FDA) and SIMCA was used for the discrimination of beef muscle samples. The results showed that FDA of combined visible-near infrared reflectance spectra has the potential to discriminate between fresh and frozen-thawed samples of meat, with a 95% accuracy (Downey & Beauchêne, 1997b). The alternative SIMCA approach was less successful, due to the similarity of the spectral data, hence resulting in lower discrimination accuracies. Downey and Beauchêne (1997b) advised that the research requires extension and should include a range of muscle types, different animal species and a number of freeze-thaw treatments, for further validation.

Table 2.3 A summary of selected applications of NIR spectroscopy and chemometric techniques for the authentication and classification of meat and meat products.

Application	Spectroscopy	Important wavelengths (nm)	Chemometric techniques	R ²	Reference
Meat species					
Discrimination of raw pork, chicken & turkey meat	visible-, near- and mid-infrared	400-750; 400-1100; 1100-2498; 400-2498	PCA*; FDA	86.5; 91.9; 94.6%**	Rannou and Downey (1997)
Discrimination of beef and kangaroo meat	visible-near-infrared	500-750; 900-1100; 400-2500	PCA*; CDA; MLR	83 - 100%**	Ding and Xu (1999)
Discrimination of beef, pork chicken & lamb meat	visible-near-infrared	400-700; 1100-2500; 400-2500	PCA*; PLSR; PCR	>80%**	Cozzolino and Murray (2004)
Detection of adulterants in minced beef	MID-FTIR	-	PCA*; PLSR; SIMCA	0.99; 100%**	Meza-Márquez <i>et al.</i> (2010)
Detection of minced beef adulteration with turkey meat	UV-vis, NIR and MIR	418; 546; 578; 980; 1041; 1200; 1450; 1740-1800; 1920; 2440; 2985; 5556-11111; 6079; 6452; 6878; 7133; 8333-10526	PCA*; LDA; PLSR	>55%; 81 - 99%**	Alamprese <i>et al.</i> (2013)
Detection of pork adulteration in veal product	FT-NIR	1683; 1693; 1697; 1704; 1731; 1737; 1764; 1770; 1821	PCA*; SVM		Schmutzler <i>et al.</i> (2015)
Detection of minced beef adulterated with turkey meat	FT-NIR	970; 1188; 1200; 1327; 1450; 1457; 1661; 1800; 1950; 1108-1408; 1800-1920	PCA*; PLSR; PLS-DA	>0.88; 100%**	Alamprese <i>et al.</i> (2016)
Fresh vs frozen-thawed					
Discrimination between fresh and frozen-thawed beef	visible-near-infrared	650-700; 758; 762; 964; 650-1100; 1070	PCA*; FDA; SIMCA	95%**	Downey and Beauchêne (1997b)
Authentication of fresh vs frozen-then-thawed beef	NIR	1188; 1450; 1470; 1488; 1696; 1700; 1738; 1742; 1748; 1930; 1946; 1980; 2056; 2060; 2176; 2192; 2290; 2298; 2326; 2350	PCA*; PLSR; FDA; SIMCA	93 - 95%; 39 - 66%**	Downey and Beauchêne (1997a)
Differentiation of frozen and unfrozen beef	NIR; DESIR	590; 614; 646; 704; 2058; 2248	PCA*; PLSR; KNN	90 - 100%**	Thyholt and Isaksson (1997)
Detection of frozen-then-thawed minced beef	MSI; FTIR	510-650; 750; 910; 970; 6061-6173; 6410-6536	PCA*; PLSDA; SVM	83.3 - 100%**	Ropodi <i>et al.</i> (2018)

*PCA used for exploratory statistical analysis only

**Classification accuracy (R² not available)

(PCA) principal component analysis; (LDA) linear discriminant analysis; (FDA) factorial discriminant analysis; (CDA) canonical discriminant analysis; (PLSR) partial least squares regression; (PCR) principal component regression; (MLR) multiple linear regression; (PLS-DA) partial least squares discriminant analysis, (SIMCA) soft independent modelling of class analogy; (KNN) k-nearest neighbour; (SVM) support vector machines.

In another study, beef was investigated to determine whether it had been frozen and thawed or if it was fresh, using NIR diffuse reflectance spectroscopy (Thyholt & Isaksson, 1997). Intact beef and drip or centrifuged meat juice of muscles from cattle were used as samples. The samples were analysed as fresh, frozen-thawed and refrozen-thawed to create classification models. The meat juices were analysed using dry extract spectroscopy by infrared reflectance (DESIR) (Thyholt & Isaksson, 1997). The KNN method correctly classified 100% of the centrifuged meat juice samples into fresh and frozen-thawed classes. KNN and PLS regression was performed to classify the samples as either fresh, frozen-thawed or refrozen-thawed. These methods gave results between 90 and 100% (Thyholt & Isaksson, 1997), with the centrifuged meat juice samples resulting in the highest accuracy. This was concluded to be due to the chemical components in the meat juice that leached out due to the damage caused by the ice crystals that were not present in the fresh meat's juice. Therefore, the exudate from the meat had to be evaluated to determine its composition and the intact beef was evaluated to determine whether the meat had undergone specific structural or chemical changes.

Ropodi *et al.* (2018) investigated the use of multi-spectral imaging (MSI) and Fourier transform infrared (FTIR) spectroscopy to rapidly detect whether minced beef had been frozen and thawed. Multi-spectral imaging and FTIR spectra were immediately acquired for the fresh minced samples. The samples were then frozen (-20 °C) and kept for seven and thirty-two days. Data was acquired similarly from both groups after the samples were thawed (Ropodi *et al.*, 2018). A total of 105 multi-spectral images and FTIR spectra were collected and further analysed using PLS-DA and SVM. The samples were divided into training, test and independent validation sets. The PLS-DA and SVM model results showed that the MSI data had a 100% overall classification accuracy. While for the FTIR data, the PLS-DA model yielded 93.3 and 96.7% overall classification accuracies for the test and external validation sets, respectively. The SVM model results were substandard compared to PLS-DA but still acceptable, yielding classification accuracies of 86.7 and 83.3% for testing and external validation, respectively. Ropodi *et al.* (2018) concluded that the data analysis used in conjunction with rapid methods could be a useful tool to monitor compliance to label regulations and for detection of frozen-then-thawed minced beef specifically.

These are only a few applications highlighting the use of NIR spectroscopy and multivariate data analysis (MDA) on meat and meat products. The simplicity of NIR spectroscopy and MDA render them useful techniques for food quality, safety and authentication inspection. Results from these few studies showed NIR spectroscopy's potential as a rapid and non-destructive method for species identification, to differentiate fresh and frozen-thawed meat as well as to detect numerous adulterations in meat. However, to date, no studies were found on ostrich and South African game meat species, such as black wildebeest, zebra and springbok, consequently requiring further investigation.

2.8 Future perspective

Conventional NIR spectroscopy has one shortcoming: it only provides spectral information about an entire sample in one spectrum. No spatial information is gleaned from conventional NIR spectroscopy; however, hyperspectral imaging incorporates both spatial and spectral information to overcome this problem (Feng *et al.*, 2013). Goetz *et al.* (1985) was the first to use NIR hyperspectral imaging, for remote sensing, as a new technique that combines digital imaging and spectroscopy. This technology was later adapted and evolved for laboratory purposes and has since been applied in countless applications; frequently used in the pharmaceutical, food and agricultural industries. NIR hyperspectral imaging is superior to that of conventional NIR spectroscopy as it provides spatial information, thus allowing the visualisation of constituents within a measured sample (Geladi *et al.*, 2007; Burger & Gowen, 2011). NIR hyperspectral imaging has shown potential to predict the quality of meat and meat products (reviewed by ElMasry *et al.*, 2012a; ElMasry *et al.*, 2012b; Kamruzzaman *et al.*, 2012b; Kamruzzaman *et al.*, 2013a; Kamruzzaman *et al.*, 2016; Ma *et al.*, 2016), and in meat science it was successfully applied for authentication purposes, without the need of any chemical analysis (Kamruzzaman *et al.*, 2011; Kamruzzaman *et al.*, 2012a; Kamruzzaman *et al.*, 2013b; Kamruzzaman *et al.*, 2015; Pu *et al.*, 2015).

A study by Kamruzzaman *et al.* (2012a) classified pork, beef and lamb using PCA and PLS-DA. For the spectral based model, six optimal wavelengths (957, 1071, 1121, 1144, 1368 and 1394 nm) were selected for the basis of discrimination and the resulting classification accuracy was 98.67%. In 2013, Kamruzzaman *et al.* (2013b), took their previous work a step further and developed a non-destructive method that could detect and quantify the level of adulteration in minced lamb meat. Minced lamb samples were adulterated with pork (2-40%) and the spectral data was used to develop a PLSR model to predict the level of adulteration (Kamruzzaman *et al.*, 2013b). The PLSR model resulted in a good prediction when using the 910-1700 nm spectral range, with R^2_{cv} of 0.99 and a RMSECV of 1.37%. Four important wavelengths were then selected and a multiple linear regression (MLR) model was constructed to predict adulteration. The MLR model resulted in a R^2_{cv} of 0.98 and a RMSECV of 1.45%. Each pixel in the image was subjected to the MLR model and subsequent prediction maps were obtained to illustrate the adulteration distribution. In a similar study by Kamruzzaman *et al.* (2015) a visible near-infrared (Vis-NIR) hyperspectral imaging system was investigated for the first time. Horsemeat was used as the adulterant in minced beef, with the levels of adulteration ranging from 2-50% (w/w). Four important wavelengths (515, 595, 650 and 880 nm) were used to develop and optimise a calibration model, resulting in the best PLSR model. The level of adulteration by horsemeat was predicted with a R^2 of 0.98 and a SEP of 2.20%.

A study by Ma *et al.* (2015) made use of Vis-NIR hyperspectral imaging to classify pork meat that was fresh or frozen-thawed. Classification algorithms were developed with PLS-DA models using (A) reflected spectra at full wavelengths, (B) at eight optimal wavelengths (448, 460, 552, 583, 588, 609, 624 and 673 nm), (C) extracted textures and (D) fused variables combining spectra at optimal wavelengths and textures. The results showed that the best classification of 97.73% was achieved by applying (D). This confirmed that

textural analysis could be used as a classification tool to discriminate between fresh and frozen-thawed meat (Ma *et al.*, 2015). Pu *et al.* (2015) followed up with a similar study, using the same wavelength range (400 – 1000 nm), but made use of six feature wavelengths (400, 446, 477, 516, 592 and 686 nm) instead of eight. As found in the previous study, the combination of textural and spectral features yielded the highest classification accuracy (Pu *et al.*, 2015) when classifying fresh and frozen-thawed pork meat.

These studies demonstrated that hyperspectral imaging, paired with multivariate analysis and image processing shows a high potential as a rapid and objective method for the classification of different animal species (Kamruzzaman *et al.*, 2012a; Kamruzzaman *et al.*, 2013b) as well as fresh and frozen-thawed meat (Barbin *et al.*, 2013; Ma *et al.*, 2015; Pu *et al.*, 2015). By optimising wavelength selection, it is possible to produce a cheap multispectral imaging instrument that can be used as rapid and reliable alternative to traditional analytical methods (Kamruzzaman *et al.*, 2015).

2.9 Conclusion

Consumers are more educated and knowledgeable about the health and safety aspects of meat products. This leads to an increasing demand for transparency with regards to the products and processes in the food industry. Therefore, rapid and non-invasive analytical methods are frequently required in order to perform routine analyses, ensuring accurate and reliable data.

NIR spectroscopy, a frequently used technique in the food industry, provides rapid, accurate, reliable, precise and reproducible results. Regarding meat authenticity, NIR spectroscopy would be an excellent screening method to be used in production plants with the capability to rapidly provide information about the food composition and potentially fraudulent activity. Various methods to prevent food fraud have been proposed in the past, but these methods are destructive, time-consuming and expensive. NIR spectroscopy could prove to be beneficial since it was shown to be a sensitive and rapid technique in the authentication of a wide variety of food products. In addition, it has the advantage of being easy to use in conjunction with chemometric analysis for more conclusive classification. Therefore, NIR spectroscopy together with chemometrics has the potential of becoming a reliable tool for the rapid identification and authentication of species and muscles in, fresh as well as frozen-thawed, meat products.

However, further research is needed with regards to instrumentation, software and chemometrics to refine this technique. The majority of the research performed, with the exception of one study, make use of benchtop NIR devices which are mostly suitable for laboratory setups. Therefore, research is required to investigate the use of portable handheld devices suitable for industrial as well as on-site setups. Research into these areas is required, specifically in relation to the application of this technique to more exotic and alternative meat species, such as ostrich and South African game meat, as these are increasingly being consumed locally and internationally.

2.10 References

- Abbas, O., Dardenne, P. & Baeten, V. (2012). Near-infrared, mid-infrared, and Raman spectroscopy. In: *Chemical analysis of food: Techniques and applications* (edited by Y. Pico). Pp. 59-91. San Diego, USA: Elsevier Inc.
- Accum, F.C. (1820). *A Treatise on Adulterations of Food, and Culinary Poisons: Exhibiting the Fraudulent Sophistications of Bread, Beer, Wine, Spirituous Liquors, Tea, Coffee, Cream, Confectionery, Vinegar, Mustard, Pepper, Cheese, Olive Oil, Pickles, and Other Articles Employed in Domestic Economy. And Methods of Detecting Them.* London: True Legacy Books.
- Agelet, L.E. & Hurlburgh, C.R. (2010). A tutorial on near infrared spectroscopy and its calibration. *Critical Reviews in Analytical Chemistry*, **40**, 246-260.
- Alamprese, C., Amigo, J.M., Casiraghi, E. & Engelsen, S.B. (2016). Identification and quantification of turkey meat adulteration in fresh, frozen-thawed and cooked minced beef by FT-NIR spectroscopy and chemometrics. *Meat Science*, **121**, 175-181.
- Alamprese, C., Casale, M., Sinelli, N., Lanteri, S. & Casiraghi, E. (2013). Detection of minced beef adulteration with turkey meat by UV-vis, NIR and MIR spectroscopy. *LWT-Food Science and Technology*, **53**, 225-232.
- Alomar, D., Gallo, C., Castaneda, M. & Fuchslocher, R. (2003). Chemical and discriminant analysis of bovine meat by near infrared reflectance spectroscopy (NIRS). *Meat Science*, **63**, 441-450.
- Anderson, R.M., Donnelly, C.A., Ferguson, N.M., Woolhouse, M.E.J., Watt, C.J., Udy, H.J., MaWhinney, S., Dunstan, S.P., Southwood, T.R.E. & Wilesmith, J.W. (1996). Transmission dynamics and epidemiology of BSE in British cattle. *Nature*, **382**, 779.
- Anonymous (2017). Principles of spectroscopy: Interaction of radiation and matter. [WWW document]. <https://www.docsity.com/en/spectroscopy-and-chemistry-bf/889779/>. Accessed 25/04/2017
- Ballin, N.Z. (2010). Authentication of meat and meat products. *Meat Science*, **86**, 577-587.
- Ballin, N.Z. & Lametsch, R. (2008). Analytical methods for authentication of fresh vs. thawed meat - A review. *Meat Science*, **80**, 151-158.
- Barbin, D.F., Sun, D.-W. & Su, C. (2013). NIR hyperspectral imaging as non-destructive evaluation tool for the recognition of fresh and frozen-thawed porcine *longissimus dorsi* muscles. *Innovative Food Science & Emerging Technologies*, **18**, 226-236.
- Barker, M. & Rayens, W. (2003). Partial least squares for discrimination. *Journal of chemometrics*, **17**, 166-173.
- 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.
- Barton, F.E. (2002). Theory and principles of near infrared spectroscopy. *Spectroscopy Europe*, **14**, 12-19.

- Bekker, J.L., Hoffman, L.C. & Jooste, P.J. (2011). Knowledge of stakeholders in the game meat industry and its effect on compliance with food safety standards. *International Journal of Environmental Health Research*, **21**, 341-363.
- Blanco, M. & Villarroya, I. (2002). NIR spectroscopy: a rapid-response analytical tool. *Trends in Analytical Chemistry*, **21**, 240-250.
- Bokobza, L. (1998). Near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **6**, 3-18.
- Bokobza, L. (2002). Origin of near-infrared absorption bands. In: *Near Infrared Spectroscopy: Principles, Instruments, Applications* (edited by S.Y. Ozaki & H.M. Heise). Pp. 11-41. Weinheim: Wiley-VCH Verlag.
- Bottenberg, R.A. & Ward, J.H. (1963). Applied multiple linear regression. (PRL-TDR-63-6): Lackland AF Base, Texas.
- Branden, K.V. & Hubert, M. (2005). Robust classification in high dimensions based on the SIMCA method. *Chemometrics and intelligent laboratory systems*, **79**, 10-21.
- Brereton, R.G. (2003a). *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. West Sussex, England: John Wiley & Sons, Ltd.
- Brereton, R.G. (2003b). Pattern recognition. In: *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. Pp. 183 - 225. West Sussex, England: John Wiley & Sons, Ltd.
- Brereton, R.G. (2003c). Pattern recognition. In: *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. Pp. 230. West Sussex, England: John Wiley & Sons, Ltd.
- Brigger, I., Chaminade, P., Desmaële, D., Peracchia, M.T., d'Angelo, J., Gurny, R., Renoir, M. & Couvreur, P. (2000). Near infrared with principal component analysis as a novel analytical approach for nanoparticle technology. *Pharmaceutical Research*, **17**, 1124-1132.
- Brown, S.D. (1995). Has the chemometrics revolution ended? Some views on the past, present and future of chemometrics. *Chemometrics and intelligent laboratory systems*, **30**, 49-58.
- Buck, E.H. (2010). Seafood marketing: combating fraud and deception. Congressional Research Service, Library of Congress.
- Burger, J. & Gowen, A. (2011). Data handling in hyperspectral image analysis. *Chemometrics and intelligent laboratory systems*, **108**, 13-22.
- Calbiani, F., Careri, M., Elviri, L., Mangia, A., Pistara, L. & Zagnoni, I. (2004). Development and in-house validation of a liquid chromatography-electrospray-tandem mass spectrometry method for the simultaneous determination of Sudan I, Sudan II, Sudan III and Sudan IV in hot chilli products. *Journal of Chromatography A*, **1042**, 123-130.
- Calvo, J., Zaragoza, P. & Osta, R. (2001). Random amplified polymorphic DNA fingerprints for identification of species in poultry pate. *Poultry Science*, **80**, 522-524.
- Calvo, J.H., Osta, R. & Zaragoza, P. (2002). Quantitative PCR detection of pork in raw and heated ground beef and pate. *Journal of Agricultural and Food Chemistry*, **50**, 5265-5267.

- Carroll, R., Cavanaugh, J. & Rorer, F. (1981). Effects of frozen storage on the ultrastructure of bovine muscle. *Journal of Food Science*, **46**, 1091-1094.
- Cawthorn, D.-M., Duncan, J., Kastern, C., Francis, J. & Hoffman, L.C. (2015). Fish species substitution and misnaming in South Africa: an economic, safety and sustainability conundrum revisited. *Food Chemistry*, **185**, 165-181.
- Cawthorn, D.-M., Steinman, H.A. & Hoffman, L.C. (2013). A high incidence of species substitution and mislabelling detected in meat products sold in South Africa. *Food Control*, **32**, 440-449.
- Cawthorn, D.-M., Steinman, H.A. & Witthuhn, R.C. (2012). DNA barcoding reveals a high incidence of fish species misrepresentation and substitution on the South African market. *Food Research International*, **46**, 30-40.
- 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.
- Chang, C.-W., Laird, D.A., Mausbach, M.J. & Hurlburgh, C.R. (2001). Near-infrared reflectance spectroscopy—principal components regression analyses of soil properties. *Soil Science Society of America Journal*, **65**, 480-490.
- Chen, M.-T., Yang, W.-D. & Guo, S.-L. (1988). Differentiation between fresh beef and thawed frozen beef. *Meat Science*, **24**, 223-226.
- Ciurczak, E.W. & Igne, B. (2014). *Pharmaceutical and medical applications of near-infrared spectroscopy*. Boca Raton, Florida: CRC Press.
- Clark, V. (2013). South Africa's dirty secret: news. *South African Food Review*, **40**, 18-19.
- Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine learning*, **20**, 273-297.
- Cowe, I.A. & McNicol, J.W. (1985). The use of principal components in the analysis of near-infrared spectra. *Applied spectroscopy*, **39**, 257-266.
- Cozzolino, D. & Murray, I. (2004). Identification of animal meat muscles by visible and near infrared reflectance spectroscopy. *LWT-Food Science and Technology*, **37**, 447-452.
- Cozzolino, D., Vadell, A., Ballesteros, F., Galletta, G. & Barlocco, N. (2006). Combining visible and near-infrared spectroscopy with chemometrics to trace muscles from an autochthonous breed of pig produced in Uruguay: a feasibility study. *Analytical and Bioanalytical Chemistry*, **385**, 931-936.
- D'Amato, M.E., Alechine, E., Cloete, K.W., Davison, S. & Corach, D. (2013). Where is the game? Wild meat products authentication in South Africa: a case study. *Investigative genetics*, **4**, 6.
- Dean, N., Murphy, T.B. & Downey, G. (2006). Using unlabelled data to update classification rules with applications in food authenticity studies. *Journal of the Royal Statistical Society: Applied Statistics*, **55**, 1-14.
- Dennis, J.M. (1998). Recent developments in food authentication. *Analyst*, **123**, 151-156.
- Ding, H. & Xu, R. (1999). Differentiation of Beef and Kangaroo Meat by Visible/Near-Infrared Reflectance Spectroscopy. *Journal of Food Science*, **64**, 814-817.

- Dos Santos, C.A.T., Lopo, M., Páscoa, R.N. & Lopes, J.A. (2013). A review on the applications of portable near-infrared spectrometers in the agro-food industry. *Applied spectroscopy*, **67**, 1215-1233.
- Downey, G. & Beauchêne, D. (1997a). Authentication of fresh vs. frozen-then-thawed beef by near-infrared reflectance spectroscopy of dried drip juice. *LWT-Food Science and Technology*, **30**, 721-726.
- Downey, G. & Beauchêne, D. (1997b). Discrimination between fresh and frozen-then-thawed beef *m. longissimus dorsi* by combined visible-near infrared reflectance spectroscopy: A feasibility study. *Meat Science*, **45**, 353-363.
- Duflos, G., Le Fur, B., Mulak, V., Becel, P. & Malle, P. (2002). Comparison of methods of differentiating between fresh and frozen-thawed fish or fillets. *Journal of the Science of Food and Agriculture*, **82**, 1341-1345.
- Ellerbroek, L.I., Lichtenberg, G. & Weise, E. (1995). Differentiation between fresh and thawed meat by an enzyme profile test. *Meat Science*, **40**, 203-209.
- Ellis, D.I., Muhamadali, H., Haughey, S.A., Elliott, C.T. & Goodacre, R. (2015). Point-and-shoot: rapid quantitative detection methods for on-site food fraud analysis – moving out of the laboratory and into the food supply chain. *Analytical Methods*, **7**, 9401-9414.
- ElMasry, G., Kamruzzaman, M., Sun, D.-W. & Allen, P. (2012a). Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: a review. *Critical Reviews in Food Science and Nutrition*, **52**, 999-1023.
- ElMasry, G., Sun, D.-W. & Allen, P. (2012b). Near-infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef. *Journal of Food Engineering*, **110**, 127-140.
- Esbensen, K.H., Guyot, D., Westad, F. & Houmoller, L.P. (2002). *Multivariate data analysis in practice: An introduction to multivariate data analysis and experimental design*. Pp. 598. Oslo: CAMO Process AS.
- Evans, S.D., Nott, K., Kshirsagar, A.A. & Hall, L.D. (1998). The effect of freezing and thawing on the magnetic resonance imaging parameters of water in beef, lamb and pork meat. *International journal of food science & technology*, **33**, 317-328.
- Fajardo, V., González, I., López-Calleja, I., Martín, I., Rojas, M., García, T., Hernandez, P. & Martín, R. (2007). PCR identification of meats from chamois (*Rupicapra rupicapra*), pyrenean ibex (*Capra pyrenaica*), and mouflon (*Ovis ammon*) targeting specific sequences from the mitochondrial D-loop region. *Meat Science*, **76**, 644-652.
- Fajardo, V., González, I., Martín, I., Rojas, M., Hernández, P.E., García, T. & Martín, R. (2008). Real-time PCR for detection and quantification of red deer (*Cervus elaphus*), fallow deer (*Dama dama*), and roe deer (*Capreolus capreolus*) in meat mixtures. *Meat Science*, **79**, 289-298.
- FDA (2004). Patulin in apple juice, apple juice concentrates and apple juice products.
- Feng, Y., ElMasry, G., Sun, D., 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.

- Fennema, O.R., Powrie, W.D. & Marth, E.H. (1973). Low-temperature preservation of foods and living matter. Pp. 151-227. New York: Marcel Dekker.
- Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, 179-188.
- Fix, E. & Hodges, J.L. (1951). Discriminatory analysis, nonparametric discrimination: Consistency properties. Randolph Field, Texas: USAF School of Aviation Medicine.
- Foster, S. (2011). A Brief History of Adulteration of Herbs, Spices, and Botanical Drugs. *HerbalGram*, **92**, 42-57.
- Fourier, J. (1822). *Théorie analytique de la chaleur, par M. Fourier*. Pp. 639. Paris: Chez Firmin Didot, père et fils.
- Gaitán-Jurado, A., Ortiz-Somovilla, V., España-España, F., Pérez-Aparicio, J. & De Pedro-Sanz, E. (2008). Quantitative analysis of pork dry-cured sausages to quality control by NIR spectroscopy. *Meat Science*, **78**, 391-399.
- Geladi, P. (2003). Chemometrics in spectroscopy. Part 1. Classical chemometrics. *Spectrochimica Acta Part B: Atomic Spectroscopy*, **58**, 767-782.
- Geladi, P., Grahn, H.F. & Burger, J. (2007). Multivariate images, hyperspectral imaging: background and equipment. In: *Techniques and applications of hyperspectral image analysis* (edited by H.F. Grahn & P. Geladi). Pp. 1-14. Chichester, West Sussex: John Wiley & Sons Ltd.
- Geladi, P., MacDougall, D. & Martens, H. (1985). Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied spectroscopy*, **39**, 491-500.
- Giovannacci, I., Guizard, C., Carlier, M., Duval, V., Martin, J.L. & Demeulemester, C. (2004). Species identification of meat products by ELISA. *International journal of food science & technology*, **39**, 863-867.
- Goetz, A.F.H., Vane, G., Solomon, J.E. & Rock, B.N. (1985). Imaging spectrometry for earth remote sensing. *Science*, **228**, 1147-1153.
- Gorsuch, R.L. (1974). *Factor Analysis*. Pp. 370. Philadelphia: W.B. Saunders Company.
- Gottesmann, P. & Hamm, R. (1983). New biochemical methods of differentiating between fresh meat and thawed, frozen meat. *Fleischwirtschaft*, **63**, 219-221.
- Guilheneuf, T.M., Parker, A.D., Tessier, J.J. & Hall, L.D. (1997). Authentication of the effect of freezing/thawing of pork by quantitative magnetic resonance imaging. *Magnetic Resonance in Chemistry*, **35**, 112-118.
- Han, D., Tu, R., Lu, C., Liu, X. & Wen, Z. (2006). Nondestructive detection of brown core in the Chinese pear 'Yali' by transmission visible-NIR spectroscopy. *Food Control*, **17**, 604-608.
- He, Y., Li, X. & Deng, X. (2007). Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model. *Journal of Food Engineering*, **79**, 1238-1242.

- He, Y., Li, X. & Shao, Y. (2006). Discrimination of varieties of apple using near infrared spectra based on principal component analysis and artificial neural network model. *Guang Pu Xue Yu Guang Pu Fen Xi*, **26**, 850-853.
- Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J. & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems*, **13**, 18-28.
- Heid, C.A., Stevens, J., Livak, K.J. & Williams, P.M. (1996). Real time quantitative PCR. *Genome Research*, **6**, 986-994.
- Henn, R., Schwab, A. & Huck, C.W. (2016). Evaluation of benchtop versus portable near-infrared spectroscopic method combined with multivariate approaches for the fast and simultaneous quantitative analysis of main sugars in syrup formulations. *Food Control*, **68**, 97-104.
- Herberholz, L., Kolomiets, O. & Siesler, H.W. (2010). Quantitative analysis by a portable near infrared spectrometer: Can it replace laboratory instrumentation for *in situ* analysis? *NIR news*, **21**, 6-8.
- Herschel, W. (1800). Investigation of the powers of the prismatic colours to heat and illuminate objects; with remarks, that prove the different refrangibility of radiant heat. To which is added, an inquiry into the method of viewing the sun advantageously, with telescopes of large apertures and high magnifying powers. . *Philosophical Transactions of the Royal Society of London*, **90**, 255-283.
- Hesser, L. (1986). Diethylene glycol toxicity. *Food and Chemical Toxicology*, **24**, 261-263.
- Hird, H., Chisholm, J., Sánchez, A., Hernandez, M., Goodier, R., Schneede, K., Boltz, C. & Popping, B. (2006). Effect of heat and pressure processing on DNA fragmentation and implications for the detection of meat using a real-time polymerase chain reaction. *Food Additives and Contaminants*, **23**, 645-650.
- Hoffman, L.C. (2001). The effect of different culling methodologies on the physical meat quality attributes of various game species. In: Proceedings of the 5th International Wildlife Ranching Symposium Pp. 212-221. Pretoria, South Africa.
- Hoffman, L.C. & Bigalke, R.C. (1999). Utilising wild ungulates from southern Africa for meat production: potential research requirements for the new millennium. In: Proceedings of the 37th Congress of the Wildlife Management Association of South Africa. Pp. 20-21. Pretoria, South Africa.
- Hoffman, L.C., Muller, M., Schutte, D.W. & Crafford, K. (2004). The retail of South African game meat: current trade and marketing trends. *South African Journal of Wildlife Research*, **34**, 123-134.
- Hoffman, L.C. & Wiklund, E. (2006). Game and venison-meat for the modern consumer. *Meat Science*, **74**, 197-208.
- Hotelling, H. (1957). The relations of the newer multivariate statistical methods to factor analysis. *British Journal of Statistical Psychology*, **10**, 69-79.
- Huang, H., Yu, H., Xu, H. & Ying, Y. (2008). Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. *Journal of Food Engineering*, **87**, 303-313.

- Hung, C.-J., Ho, H.-P., Chang, C.-C., Lee, M.-R., Franje, C.A., Kuo, S.-I., Lee, R.-J. & Chou, C.-C. (2011). Electrochemical profiling using copper nanoparticle-plated electrode for identification of ostrich meat and evaluation of meat grades. *Food Chemistry*, **126**, 1417-1423.
- Ingelfinger, J.R. (2008). Melamine and the global implications of food contamination. *New England Journal of Medicine*, **359**, 2745-2748.
- Isaksson, T. & Næs, T. (1988). The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. *Applied spectroscopy*, **42**, 1273-1284.
- Jamrógiewicz, M. (2012). Application of the near-infrared spectroscopy in the pharmaceutical technology. *Journal of Pharmaceutical and Biomedical Analysis*, **66**, 1-10.
- Jeffers, J. (1967). Two case studies in the application of principal component analysis. *Applied Statistics*, **16**, 225-236.
- Jha, V.K., Kumar, A. & Mandokhot, U.V. (2003). Indirect enzyme-linked immunosorbent assay in detection and differentiation of cooked and raw pork from meats of other species. *Journal of Food Science & Technology*, **40**, 254-256.
- Johnson, R. (2014). Food Fraud and "Economically Motivated Adulteration" of Food and Food Ingredients. In: Congressional Research Service. Pp. 1-45. Washington, D.C.
- Jolliffe, I.T. (1986). *Principal Component Analysis*. Pp. 478. London: Springer.
- Juhász, R., Gergely, S., Szabóki, Á. & Salgó, A. (2007). Correlation between NIR spectra and RVA parameters during germination of maize. *Cereal Chemistry*, **84**, 97-101.
- Kamruzzaman, M., Barbin, D., ElMasry, G., Sun, D.-W. & Allen, P. (2012a). Potential of hyperspectral imaging and pattern recognition for categorization and authentication of red meat. *Innovative Food Science & Emerging Technologies*, **16**, 316-325.
- Kamruzzaman, M., ElMasry, G., Sun, D.-W. & Allen, P. (2011). Application of NIR hyperspectral imaging for discrimination of lamb muscles. *Journal of Food Engineering*, **104**, 332-340.
- Kamruzzaman, M., ElMasry, G., Sun, D.-W. & Allen, P. (2012b). 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.
- Kamruzzaman, M., ElMasry, G., Sun, D.-W. & Allen, P. (2013a). Non-destructive assessment of instrumental and sensory tenderness of lamb meat using NIR hyperspectral imaging. *Food Chemistry*, **141**, 389-396.
- Kamruzzaman, M., Makino, Y. & Oshita, S. (2016). Hyperspectral imaging for real-time monitoring of water holding capacity in red meat. *LWT-Food Science and Technology*, **66**, 685-691.
- Kamruzzaman, M., Makino, Y., Oshita, S. & Liu, S. (2015). Assessment of visible near-infrared hyperspectral imaging as a tool for detection of horsemeat adulteration in minced beef. *Food and Bioprocess Technology*, **8**, 1054-1062.

- Kamruzzaman, M., Sun, D.-W., ElMasry, G. & Allen, P. (2013b). Fast detection and visualization of minced lamb meat adulteration using NIR hyperspectral imaging and multivariate image analysis. *Talanta*, **103**, 130-136.
- Kendall, M.G. (1957). *A Course in Multivariate Analysis*. London: Charles Griffin & Co. Ltd.
- Kennedy, J., Delaney, L., McGloin, A. & Wall, P.G. (2009). Public perceptions of the dioxin crisis in Irish pork. In: UCD Geary Institute Discussion Paper Series. University College Dublin: Geary Institute.
- Kenneth, R.B., Randy, J.P. & Seasholtz, M.B. (1998). Preprocessing. In: *Chemometrics: A Practical Guide*. Pp. 41-45. New York: John Wiley and Sons Inc.
- Kesmen, Z., Gulluce, A., Sahin, F. & Yetim, H. (2009). Identification of meat species by TaqMan-based real-time PCR assay. *Meat Science*, **82**, 444-449.
- Kesmen, Z., Yetiman, A.E., Şahin, F. & Yetim, H. (2012). Detection of chicken and turkey meat in meat mixtures by using Real-Time PCR Assays. *Journal of Food Science*, **77**, 167-173.
- Kreuz, G., Zagon, J., Broll, H., Bernhardt, C., Linke, B. & Lampen, A. (2012). Immunological detection of osteocalcin in meat and bone meal: a novel heat stable marker for the investigation of illegal feed adulteration. *Food Additives & Contaminants: Part A*, **29**, 716-726.
- Kuteifan, K., Oesterle, H., Tajahmady, T., Gutbub, A. & Laplatte, G. (1998). Necrosis and haemorrhage of the putamen in methanol poisoning shown on MRI. *Neuroradiology*, **40**, 158-160.
- Lammertyn, J., Peirs, A., De Baerdemaeker, J. & Nicolai, B. (2000). Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. *Postharvest Biology and Technology*, **18**, 121-132.
- Landau, S., Glasser, T. & Dvash, L. (2006). Monitoring nutrition in small ruminants with the aid of near infrared reflectance spectroscopy (NIRS) technology: A review. *Small Ruminant Research*, **61**, 1-11.
- Lerma-García, M., Herrero-Martínez, J., Ramis-Ramos, G., Mongay-Fernández, C. & Simó-Alfonso, E. (2009). Prediction of the curing time of Spanish hams using peptide profiles established by capillary zone electrophoresis. *Food Chemistry*, **113**, 635-639.
- Leygonie, C., Britz, T.J. & Hoffman, L.C. (2012). Impact of freezing and thawing on the quality of meat: Review. *Meat Science*, **91**, 93-98.
- Li, S.F.Y. (1992). *Capillary electrophoresis: principles, practice and applications*. Pp. 1-555. Amsterdam: Elsevier.
- López-Andreo, M., Aldeguer, M., Guillén, I., Gabaldón, J.A. & Puyet, A. (2012). Detection and quantification of meat species by qPCR in heat-processed food containing highly fragmented DNA. *Food Chemistry*, **134**, 518-523.
- Lovett, D.K., Deaville, E.R., Givens, D.I., Finlay, M. & Owen, E. (2005). Near infrared reflectance spectroscopy (NIRS) to predict biological parameters of maize silage: effects of particle comminution, oven drying temperature and the presence of residual moisture. *Animal Feed Science and Technology*, **120**, 323-332.

- Lovett, D.K., Deaville, E.R., Mould, F., Givens, D.I. & Owen, E. (2004). Using near infrared reflectance spectroscopy (NIRS) to predict the biological parameters of maize silage. *Animal Feed Science and Technology*, **115**, 179-187.
- Luypaert, J., Massart, D. & Vander Heyden, Y. (2007). Near-infrared spectroscopy applications in pharmaceutical analysis. *Talanta*, **72**, 865-883.
- Ma, J., Pu, H., Sun, D.-W., Gao, W., Qu, J.-H. & Ma, K.-Y. (2015). Application of Vis–NIR hyperspectral imaging in classification between fresh and frozen-thawed pork *Longissimus Dorsi* muscles. *International Journal of Refrigeration*, **50**, 10-18.
- Ma, J., Sun, D.-W. & Pu, H. (2016). Spectral absorption index in hyperspectral image analysis for predicting moisture contents in pork longissimus dorsi muscles. *Food Chemistry*, **197**, 848-854.
- Maleki, M., Mouazen, A., Ramon, H. & De Baerdemaeker, J. (2007). Multiplicative scatter correction during on-line measurement with near infrared spectroscopy. *Biosystems Engineering*, **96**, 427-433.
- Marnane, I. (2012). Comprehensive environmental review following the pork PCB/dioxin contamination incident in Ireland. *Journal of Environmental Monitoring*, **14**, 2551-2556.
- Martín, R., Azcona, J.I., García, T., Hernández, P.E. & Sanz, B. (1988). Sandwich ELISA for detection of horse meat in raw meat mixtures using antisera to muscle soluble proteins. *Meat Science*, **22**, 143-153.
- Martino, M.N., Otero, L., Sanz, P. & Zaritzky, N. (1998). Size and location of ice crystals in pork frozen by high-pressure-assisted freezing as compared to classical methods. *Meat Science*, **50**, 303-313.
- Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y. & Kaufman, L. (1988). *Chemometrics: A Textbook*. Pp. 477. Amsterdam: Elsevier Science Publishers.
- Mazorra-Manzano, M.A., Torres-Llanez, M.J., González-Córdova, A.F. & Vallejo-Cordoba, B. (2012). A capillary electrophoresis method for the determination of hydroxyproline as a collagen content index in meat products. *Food Analytical Methods*, **5**, 464-470.
- McGlone, V.A., Jordan, R.B. & Martinsen, P.J. (2002). Vis/NIR estimation at harvest of pre-and post-storage quality indices for 'Royal Gala' apple. *Postharvest Biology and Technology*, **25**, 135-144.
- Meza-Márquez, O.G., Gallardo-Velázquez, T. & Osorio-Revilla, G. (2010). Application of mid-infrared spectroscopy with multivariate analysis and soft independent modeling of class analogies (SIMCA) for the detection of adulterants in minced beef. *Meat Science*, **86**, 511-519.
- Mishra, P., Herrero-Langreo, A., Barreiro, P., Roger, J.M., Diezma, B., Gorretta, N. & Lleó, L. (2015). Detection and quantification of peanut traces in wheat flour by near infrared hyperspectral imaging spectroscopy using principal-component analysis. *Journal of Near Infrared Spectroscopy*, **23**, 15-22.
- Montowska, M. & Pospiech, E. (2011). Differences in two-dimensional gel electrophoresis patterns of skeletal muscle myosin light chain isoforms between *Bos taurus*, *Sus scrofa* and selected poultry species. *Journal of the Science of Food and Agriculture*, **91**, 2449-2456.
- Morsy, N. & Sun, D.W. (2013). Robust linear and non-linear models of NIR spectroscopy for detection and quantification of adulterants in fresh and frozen-thawed minced beef. *Meat Science*, **93**, 292-302.

- Mortensen, M., Andersen, H.J., Engelsen, S.B. & Bertram, H.C. (2006). Effect of freezing temperature, thawing and cooking rate on water distribution in two pork qualities. *Meat Science*, **72**, 34-42.
- Mueller, J.P. & Massaron, L. (2016). Going a step beyond using Support Vector Machines. In: *Machine Learning for Dummies*. Pp. 297-313. Canada: John Wiley & Sons, Inc.
- Mueller, T. (2011). *Extra virginity: The sublime and scandalous world of olive oil*. W.W. Norton & Company.
- Nakyinsige, K., Che Man, Y.B. & Sazili, A.Q. (2012). Halal authenticity issues in meat and meat products. *Meat Science*, **91**, 207-214.
- Nicolai, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I. & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*, **46**, 99-118.
- Nielsen, S.S. (2010). Infrared Spectroscopy. In: *Food Analysis*. Pp. 407-420. New York, USA: Springer.
- Norris, K.H. (1962). Instrumentation of infrared radiation. *Transactions of the ASAE*, **5**, 17-20.
- Norris, K.H. (1964). Simple spectroradiometer for 0.4 to 1.2 micron region. *Transactions of the ASAE*, **7**, 240-242.
- Ortiz-Somovilla, V., España-España, F., De Pedro-Sanz, E. & Gaitán-Jurado, A. (2005). Meat mixture detection in Iberian pork sausages. *Meat Science*, **71**, 490-497.
- Ortiz-Somovilla, V., España-España, F., Gaitán-Jurado, A., Pérez-Aparicio, J. & De Pedro-Sanz, E. (2007). Proximate analysis of homogenized and minced mass of pork sausages by NIRS. *Food Chemistry*, **101**, 1031-1040.
- Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*. Essex, England: Longman Scientific and Technical.
- Pasquini, C. (2003). Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, **14**, 198-219.
- Patterson, R.L.S. & Jones, S.J. (1990). Review of current techniques for the verification of the species origin of meat. *Analyst*, **115**, 501-506.
- Pestana, A. & Munoz, E. (1982). Anilides and the Spanish toxic oil syndrome. *Nature*, **298**, 608.
- Peterson, L.E. (2009). K-nearest neighbor. *Scholarpedia*, **4**, 1883.
- Pizarro, C., Esteban-Díez, I., Nistal, A.J. & González-Sáiz, J.M. (2004). Influence of data pre-processing on the quantitative determination of the ash content and lipids in roasted coffee by near infrared spectroscopy. *Analytica Chimica Acta*, **509**, 217-227.
- Plans, M., Simó, J., Casañas, F., Sabaté, J. & Rodriguez-Saona, L. (2013). Characterization of common beans (*Phaseolus vulgaris* L.) by infrared spectroscopy: comparison of MIR, FT-NIR and dispersive NIR using portable and benchtop instruments. *Food Research International*, **54**, 1643-1651.
- Premanandh, J. (2013). Horse meat scandal – A wake-up call for regulatory authorities. *Food Control*, **34**, 568-569.

- Prieto, N., Roehe, R., Lavin, P., Batten, G. & Andres, S. (2009). Application of near infrared reflectance spectroscopy to predict meat and meat products quality: A review. *Meat Sci*, **83**, 175-186.
- Primrose, S., Woolfe, M. & Rollinson, S. (2010). Food forensics: methods for determining the authenticity of foodstuffs. *Trends in Food Science & Technology*, **21**, 582-590.
- Pu, H., Sun, D.-W., Ma, J. & Cheng, J.-H. (2015). Classification of fresh and frozen-thawed pork muscles using visible and near infrared hyperspectral imaging and textural analysis. *Meat Science*, **99**, 81-88.
- Rannou, H. & Downey, G. (1997). Discrimination of raw pork, chicken and turkey meat by spectroscopy in the visible, near-and mid-infrared ranges. *Analytical Communications*, **34**, 401-404.
- Rebrikov, D. & Trofimov, D.Y. (2006). Real-time PCR: a review of approaches to data analysis. *Applied Biochemistry and Microbiology*, **42**, 455-463.
- Reich, G. (2005). Near-infrared spectroscopy and imaging: basic principles and pharmaceutical applications. *Advanced Drug Delivery Reviews*, **57**, 1109-1143.
- Reid, L.M., O'Donnell, C.P. & Downey, G. (2006). Recent technological advances for the determination of food authenticity. *Trends in Food Science & Technology*, **17**, 344-353.
- Rodionova, O.Y., Houmøller, L.P., Pomerantsev, A.L., Geladi, P., Burger, J., Dorofeyev, V.L. & Arzamastsev, A.P. (2005). NIR spectrometry for counterfeit drug detection: a feasibility study. *Analytica Chimica Acta*, **549**, 151-158.
- Rodriguez, M.A., García, T., González, I., Asensio, L., Hernández, P.E. & Martin, R. (2003). Qualitative PCR for the detection of chicken and pork adulteration in goose and mule duck foie gras. *Journal of the Science of Food and Agriculture*, **83**, 1176-1181.
- Roggo, Y., Chalus, P., Maurer, L., Lema-Martinez, C., Edmond, A. & Jent, N. (2007). A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of Pharmaceutical and Biomedical Analysis*, **44**, 683-700.
- Ropodi, A.I., Panagou, E.Z. & Nychas, G.J.E. (2018). Rapid detection of frozen-then-thawed minced beef using multispectral imaging and Fourier transform infrared spectroscopy. *Meat Science*, **135**, 142-147.
- Sakudo, A., Suganuma, Y., Kobayashi, T., Onodera, T. & Ikuta, K. (2006). Near-infrared spectroscopy: promising diagnostic tool for viral infections. *Biochemical and Biophysical Research Communications*, **341**, 279-284.
- Sanz, P., De Elvira, C., Martino, M., Zaritzky, N., Otero, L. & Carrasco, J. (1999). Freezing rate simulation as an aid to reducing crystallization damage in foods. *Meat Science*, **52**, 275-278.
- Sato, T. (1994). Application of principal-component analysis on near-infrared spectroscopic data of vegetable oils for their classification. *Journal of the American Oil Chemists' Society*, **71**, 293-298.
- Savitzky, A. & Golay, M.J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, **36**, 1627-1639.

- Schmutzler, M., Beganovic, A., Böhler, G. & Huck, C.W. (2015). Methods for detection of pork adulteration in veal product based on FT-NIR spectroscopy for laboratory, industrial and on-site analysis. *Food Control*, **57**, 258-267.
- Schneider, A. (2011). Tests Show Most Store Honey Isn't Honey. In: Food Safety News. Pp. 1-10.
- Schönfeldt, H. (1993). Nutritional content of venison. In: The venison industry - Research requirements and possibilities Pp. 51-60. Meat industry centre, Irene Agricultural research Council Pretoria, South Africa.
- Sebestyen, G. (1962). Pattern recognition by an adaptive process of sample set construction. *IRE Transactions on Information Theory*, **8**, 82-91.
- Sen, A.R. & Sharma, N. (2004). Effect of freezing and thawing on the histology and ultrastructure of buffalo muscle. *Asian Australasian Journal of Animal Sciences*, **17**, 1291-1295.
- Sentandreu, M.Á. & Sentandreu, E. (2014). Authenticity of meat products: Tools against fraud. *Food Research International*, **60**, 19-29.
- Sharma, N.K., Srivastava, A.K., Gill, J.P.S. & Joshi, D.V. (1994). Differentiation of meat from food animals by enzyme assay. *Food Control*, **5**, 219-221.
- Siebert, S., Beneke, B. & Bentler, W. (1994). Bef, pork and sheep meat-detecting previous frozen treatment by isoelectric-focusing in polyamide acryl gel (PAGIF) during routine diagnosis. *Fleischwirtschaft*, **74**, 417-420.
- Siesler, H.W. (2008a). Basic principles of near-infrared spectroscopy. In: *Handbook of Near-Infrared Analysis* (edited by D.A. Burns & E.W. Ciurczak). Pp. 71. Boca Raton: CRC Press, Taylor & Francis Group.
- Siesler, H.W. (2008b). Basic principles of near-infrared spectroscopy. In: *Handbook of Near-Infrared Analysis* (edited by D.A. Burns & E.W. Ciurczak). Pp. 7-19. Boca Raton: CRC Press, Taylor & Francis Group.
- Sirven, J.-B., Salle, B., Mauchien, P., Lacour, J.-L., Maurice, S. & Manhes, G. (2007). Feasibility study of rock identification at the surface of Mars by remote laser-induced breakdown spectroscopy and three chemometric methods. *Journal of Analytical Atomic Spectrometry*, **22**, 1471-1480.
- Siuda, R., Balcerowska, G. & Sadowski, C. (2006). Comparison of the usability of different spectral ranges within the near ultraviolet, visible and near infrared ranges (UV-VIS-NIR) region for the determination of the content of scab-damaged component in blended samples of ground wheat. *Food Additives and Contaminants*, **23**, 1201-1207.
- Smith, C. (1992). Application of immunoassay to the detection of food adulteration. In: *Food safety and quality assurance - Applications of immunoassay system*. . Pp. 13-32. London: Elsevier Applied Science.
- Soares, S., Amaral, J.S., Mafra, I. & Oliveira, M.B.P. (2010). Quantitative detection of poultry meat adulteration with pork by a duplex PCR assay. *Meat Science*, **85**, 531-536.
- Soares, S., Amaral, J.S., Oliveira, M.B.P. & Mafra, I. (2013). A SYBR Green real-time PCR assay to detect and quantify pork meat in processed poultry meat products. *Meat Science*, **94**, 115-120.

- Spink, J. & Moyer, D.C. (2011a). Defining the public health threat of food fraud. *Journal of Food Science*, **76**, 157-163.
- Spink, J. & Moyer, D.C. (2011b). Defining the public health threat of food fraud. *Journal of Food Science*, **76**.
- Szabo, L., Pinter, L. & Alfoldi, Z. (1991). The NIR technique as a fast and cheap tool for improving silage maize breeding and evaluation. *NOVENYTERMELES*, **40**, 135-140.
- Thyholt, K. & Isaksson, T. (1997). Differentiation of frozen and unfrozen beef using near-infrared spectroscopy. *Journal of the Science of Food and Agriculture*, **73**, 525-532.
- Tisdall, M.M., Tachtsidis, I., Leung, T.S., Elwell, C.E. & Smith, M. (2007). Near-infrared spectroscopic quantification of changes in the concentration of oxidized cytochrome c oxidase in the healthy human brain during hypoxemia. *Journal of biomedical optics*, **12**, 024002-024007.
- Toldrá, F., Torrero, Y. & Flores, J. (1991). Simple test for differentiation between fresh pork and frozen/thawed pork. *Meat Science*, **29**, 177-181.
- Vallejo-Cordoba, B., Rodríguez-Ramírez, R. & González-Córdova, A.F. (2010). Capillary electrophoresis for bovine and ostrich meat characterisation. *Food Chemistry*, **120**, 304-307.
- Vapnik, V. (1963). Pattern recognition using generalized portrait method. *Automation and remote control*, **24**, 774-780.
- Vasconcellos, L.P.d.M.K., Tambasco-Talhari, D., Pereira, A.P., Coutinho, L.L. & Regitano, L.C.d.A. (2003). Genetic characterization of Aberdeen Angus cattle using molecular markers. *Genetics and Molecular Biology*, **26**, 133-137.
- Viljoen, J.J. (1999). A comparison of the lipid components of springbok meat with those of beef and the related importance on aspects of health. *South African Journal of Science and Technology*, **18**, 51-53.
- Westerhuis, J.A., Hoefsloot, H.C., Smit, S., Vis, D.J., Smilde, A.K., van Velzen, E.J., van Duijnhoven, J.P. & van Dorsten, F.A. (2008). Assessment of PLSDA cross validation. *Metabolomics*, **4**, 81-89.
- Wilesmith, J., Ryan, J., Hueston, W. & Hoinville, L. (1992). Bovine spongiform encephalopathy: epidemiological features 1985 to 1990. *The Veterinary Record*, **130**, 90-94.
- Wilkerson, E.D., Anthon, G.E., Barrett, D.M., Sayajon, G.F.G., Santos, A.M. & Rodriguez-Saona, L.E. (2013). Rapid assessment of quality parameters in processing tomatoes using hand-held and benchtop infrared spectrometers and multivariate analysis. *Journal of Agricultural and Food Chemistry*, **61**, 2088-2095.
- Williams, P. (2006). Near-infrared technology - Getting the best out of light. In: *A short course in the practical implementation of near-infrared spectroscopy for the user*. Nanaimo, Canada: PDK Projects, Inc.
- Williams, P.J. (2009). *Near infrared (NIR) hyperspectral imaging for evaluation of whole maize kernels: chemometrics for exploration and classification*. (MSc Thesis). South Africa: 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* (PhD Thesis). South Africa: University of Stellenbosch.
- Wilson, B. (2008). *Swindled: The Dark History of Food Fraud - From Poisoned Candy to Counterfeit Coffee*. Princeton University Press.
- Wold, H. (1975). Path models with latent variables: The NIPALS approach. In: *Quantitative sociology: International perspectives on mathematical and statistical modeling*. Pp. 307-357. New York: Academic Press.
- Wold, S., Esbensen, K. & Geladi, P. (1987a). Principal component analysis. *Chemometrics and intelligent laboratory systems*, **2**, 37-52.
- Wold, S., Geladi, P., Esbensen, K. & Öhman, J. (1987b). Multi-way principal components-and PLS-analysis. *Journal of chemometrics*, **1**, 41-56.
- Wold, S. & Sjöström, M. (1977). SIMCA: a method for analyzing chemical data in terms of similarity and analogy. In: *Chemometrics: Theory and Application*. Pp. 243 - 282. ACS Publications.
- Wold, S., Sjöström, M. & Eriksson, L. (2001a). PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, **58**, 109-130.
- Wold, S., Trygg, J., Berglund, A. & Antti, H. (2001b). Some recent developments in PLS modeling. *Chemometrics and intelligent laboratory systems*, **58**, 131-150.
- Wolf, M., Ferrari, M. & Quaresima, V. (2007). Progress of near-infrared spectroscopy and topography for brain and muscle clinical applications. *Journal of biomedical optics*, **12**, 062104-062114.
- Workman, J. (1993). A brief review of the near infrared measurement technique. *NIR news*, **4**, 8-16.
- Xin, H. & Stone, R. (2008). Chinese probe unmasks high-tech adulteration with melamine. *Science*, **322**, 1310-1311.
- Zamora-Rojas, E., Pérez-Marín, D., De Pedro-Sanz, E., Guerrero-Ginel, J. & Garrido-Varo, A. (2012). Handheld NIRS analysis for routine meat quality control: Database transfer from at-line instruments. *Chemometrics and intelligent laboratory systems*, **114**, 30-35.
- Zhang, W. (2017). Machine Learning Approaches to Predicting Company Bankruptcy. *Journal of Financial Risk Management*, **6**, 364.

Chapter 3

Materials and Methods

3.1 Samples, sampling and sample preparation

The trial was approved by the Animal ethics committee at Stellenbosch University (Ethical clearance number: SU-ACUM14-001SOP). Meat from four different South African game species [black wildebeest (*Connochaetes gnou*), zebra (*Equus quagga burchelli*), springbok (*Antidorcas marsupialis*) and ostrich (*Struthio camelus*)] were obtained from several game and ostrich farms across the Western Cape, South Africa (**Table 3.1**). The animals were randomly selected (for age, sex) and slaughtered according to standard South African procedures and regulations (DAFF, 2000) either on-farm or in a registered abattoir and processed further at the Department of Animal Sciences, Stellenbosch University (SU). From each ungulate game carcass (black wildebeest, zebra and springbok), seven muscles [*longissimus thoracis et lumborum* (LTL), *biceps femoris* (BF), *semimembranosus* (SM), *semitendinosus* (ST), *infraspinatus* (IS), *supraspinatus* (SS) and *psoas major* (fillet)] were excised from the left side and two muscles [*gastrocnemius* (BD) and *iliofibularis* (FF)] from the left side of each ostrich carcass. The subcutaneous fat and visible sinews were removed from the excised muscles. The muscles were excised post rigor, placed in labelled plastic bags and transported to the Department of Food Science, Stellenbosch University (SU) for storage at 3° C until analysed.

Table 3. 1 General information regarding the species, origin and muscle type used for this study.

Species	Code	Origin	Muscle type	No. of animals
Black wildebeest	BWB	Bredasdorp	LTL, BF, SM, ST, IS, SS, fillet	10
Zebra	ZEB	Bredasdorp Elandsberg Nature Reserve	LTL, BF, SM, ST, IS, SS, fillet	22
Springbok	SB	Witsand, Heidelberg	LTL, BF, SM, ST, IS, SS, fillet	19
Ostrich	OST	KKI abattoir, Oudtshoorn	BD, FF	10
Total				61

(LTL) *longissimus thoracis et lumborum*; (BF) *biceps femoris*; (SM) *semimembranosus*; (ST) *semitendinosus*; (IS) *infraspinatus*; (SS) *supraspinatus*; (fillet) *psoas major*; (BD) *gastrocnemius*; (FF) *iliofibularis*.

The muscles [BF, SM, ST, IS, SS, fillet (ungulates) and BD (ostrich)] were cut into 1.5 – 2.0 cm thick steaks 24h post mortem and storage at 3° C. The *longissimus thoracis et lumborum* (LTL) muscle from each ungulate carcass and the *iliofibularis* (FF) muscle from each ostrich carcass was cut into nine 1.5 – 2.0 cm thick steaks and numbered (P1 – P9) from the dorsal side of the muscle. Each piece of steak was then assigned to a month (m1 – m9) in randomised order as shown in **Figure 3.1**. The samples were randomised to ensure that as much variation as possible was incorporated into the data, which is required for multivariate data analysis. All the steak samples were laid down and allowed to bloom at ambient temperature (ca. 23°

C) for *ca.* 30 – 60 min after which the near-infrared (NIR) spectra were acquired. The samples were then placed in labelled bags, vacuum sealed and stored at -20° C until analysed at the next designated time period (monthly; **Figure 3.1**).

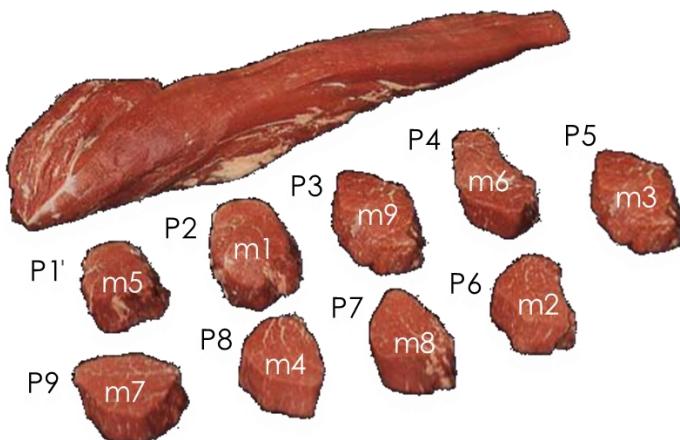


Figure 3.1 Schematic illustrating the randomisation process of the LTL and FF muscle samples. Each muscle was cut into nine steaks, with each steak representing an anatomical position (P1 – P9). The anatomical position of the steaks was numbered from the dorsal (top) side of the muscle. Each piece of steak was randomly assigned to a month (m1 – m9), therefore ensuring that the same anatomical position is not scanned on the same month, for each muscle sample.

3.2 NIR instrumentation

Near-infrared reflectance spectra were acquired with the MicroNIR OnSite spectrophotometer (Viavi Solutions Inc., Milpitas, USA). The illumination source comprised of two integrated vacuum tungsten lamps coupled to a linear variable filter and a 128-pixel Indium Gallium Arsenide (InGaAs) photodiode array detector. Individual spectra were acquired within the spectral range of 908 – 1676 nm at <12.5 nm resolution with a 6.2 nm pixel-to-pixel interval and a pixel size/pitch of 30 µm x 250 µm/50 µm. The samples were scanned at an optimal pathlength of 3 mm (through a glass petri dish) and the measuring time per sample spectrum was 0.25 – 0.5 sec.

3.3 Spectral acquisition

In **Figure 3.2** the experimental set-up, blooming of samples and spectral collection procedure are illustrated. After the samples were left to bloom [development of metmyoglobin (browning)] for *ca.* 30 – 60 min at room temperature (*ca.* 23° C), the surface of the meat was blotted dry, with an absorbent tissue paper, to remove the excess moisture before collecting the NIR spectra. Firstly, all the meat samples were scanned as fresh and the spectra was used to establish a baseline for each piece of meat per species. The LTL and FF muscle samples were then vacuum packed and stored at -20° C for up to nine months, thereafter the samples were thawed and the same procedure was repeated to acquire the new set of scans for the frozen-thawed meat. The meat samples (m1 – m9) were thawed in one-month intervals according to the following procedure: the samples were removed from the -20° C storage and placed at 3° C to thaw for *ca.* 18 h prior to analysis. The

samples were then removed from the packaging, blotted dry and laid down to bloom at ambient temperature (*ca.* 23° C) for *ca.* 30 – 60 min after which the NIR scans were acquired. Therefore, ten spectral datasets (F0 – F9) were obtained at monthly intervals for each of the LTL (ungulate) and FF (ostrich) muscle samples per species, respectively. The monthly intervals indicate periods of time after frozen storage. The remaining muscle samples [BF, SM, ST, IS, SS, fillet (ungulates) and BD (ostrich)] were also scanned as described before (F0) being vacuum packed and stored at -20° C for one month (F1). Thereafter, the samples were thawed (removed from -20° C, placed at 3° C for *ca.* 18 h) and the same procedure was repeated to acquire the new set of scans (F1) for the frozen-thawed meat. Two spectral datasets (F0 – F1) were collected at two time points (0 and 1 month), for each of the muscle samples per species. The time points indicate whether the samples are fresh (F0) or frozen-thawed (F1) after one month of frozen storage. **Figure 3.3** provides a schematic of the above-mentioned spectral acquisition process.

Preliminary experiments indicated that performing triplicate measurements was sufficient for variation coverage within a sample. Therefore, each sample was scanned in triplicate, through a glass petri dish (3 mm pathlength), while moving the NIR spectrophotometer across the sample. Moving the spectrophotometer across the sample ensured that the whole piece of meat was scanned and that most of the variation within one sample was covered. Throughout this process, the samples were prepared and scanned under the same conditions.

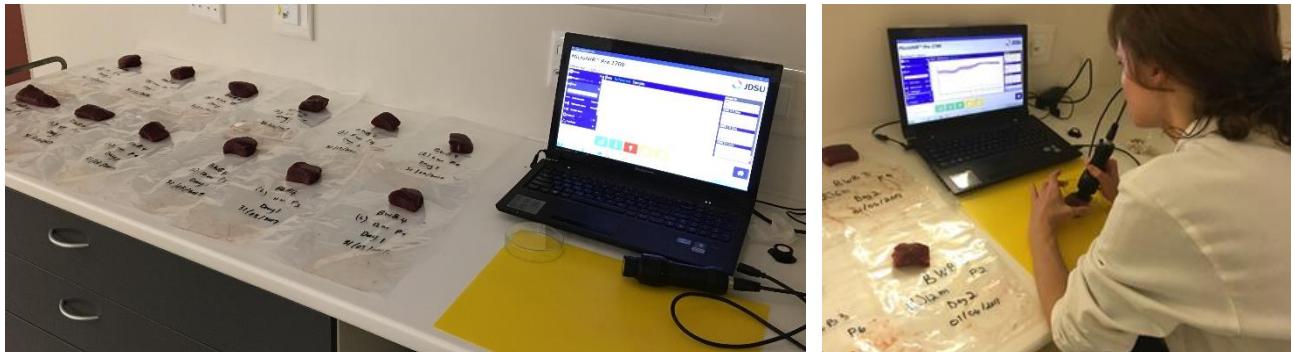


Figure 3.2 Experimental set-up, blooming of samples and spectral collection of game meat steaks.

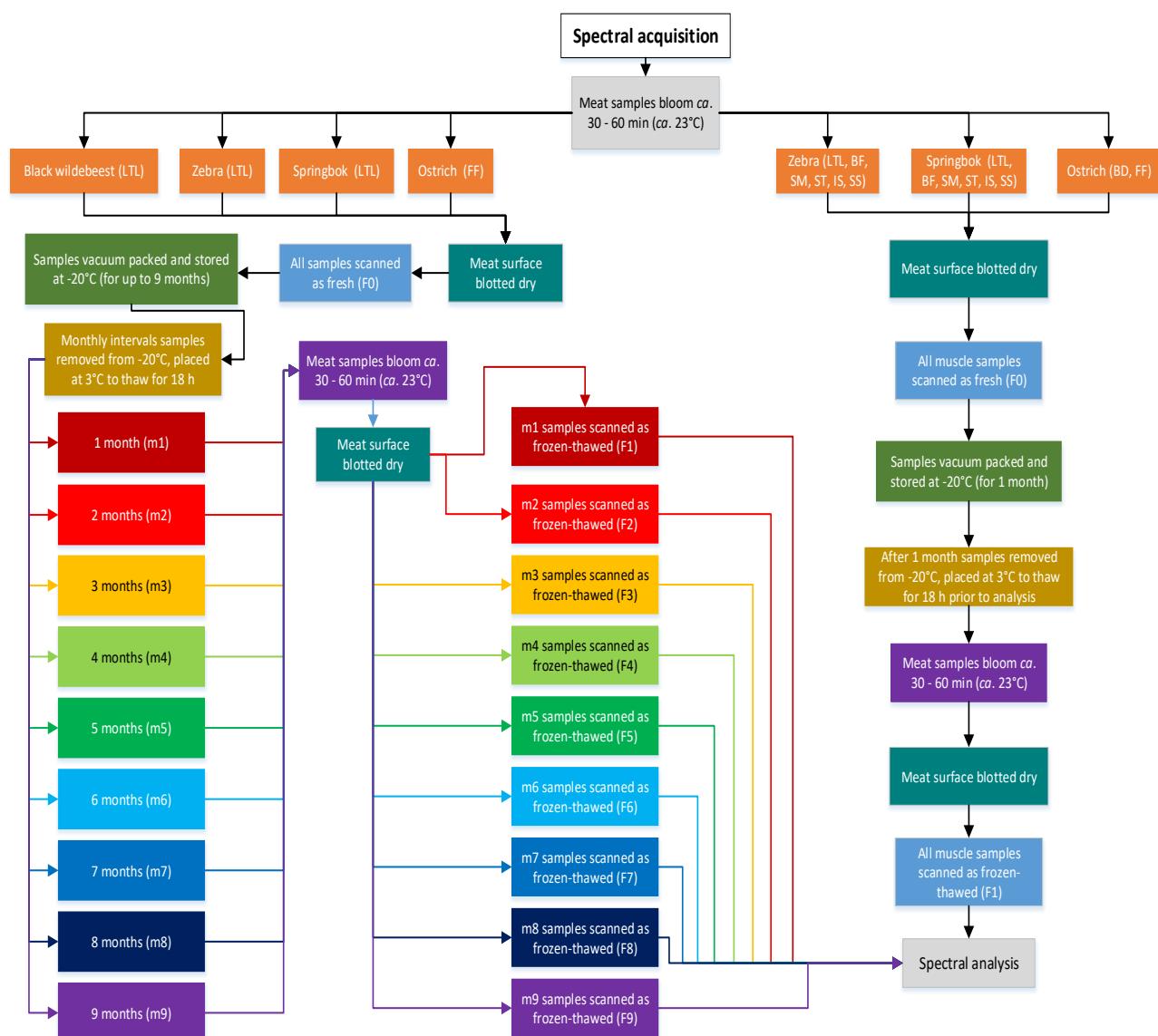


Figure 3.3 Schematic of the spectral acquisition process.

3.4 Spectral analysis

Spectral data were analysed using The Unscrambler X10.5 (Camo Software AS., Oslo, Norway) and the PLS_Toolbox [Solo] (Eigenvector Research Inc., Wenatchee, WA) multivariate data analysis software packages. The triplicate spectra obtained for each sample were averaged prior to pre-processing, data exploration and model development. In addition, spectra were reduced to a 920 – 1651 nm wavelength range, as noise was observed at both ends of the spectral range. The mean spectrum of each class set (species, muscle type, fresh vs. frozen-thawed) was computed between 920 and 1651 nm and plotted on one graph to investigate, determine and compare the chemical properties. **Figure 3.4** presents a schematic of the spectral data analysis process.

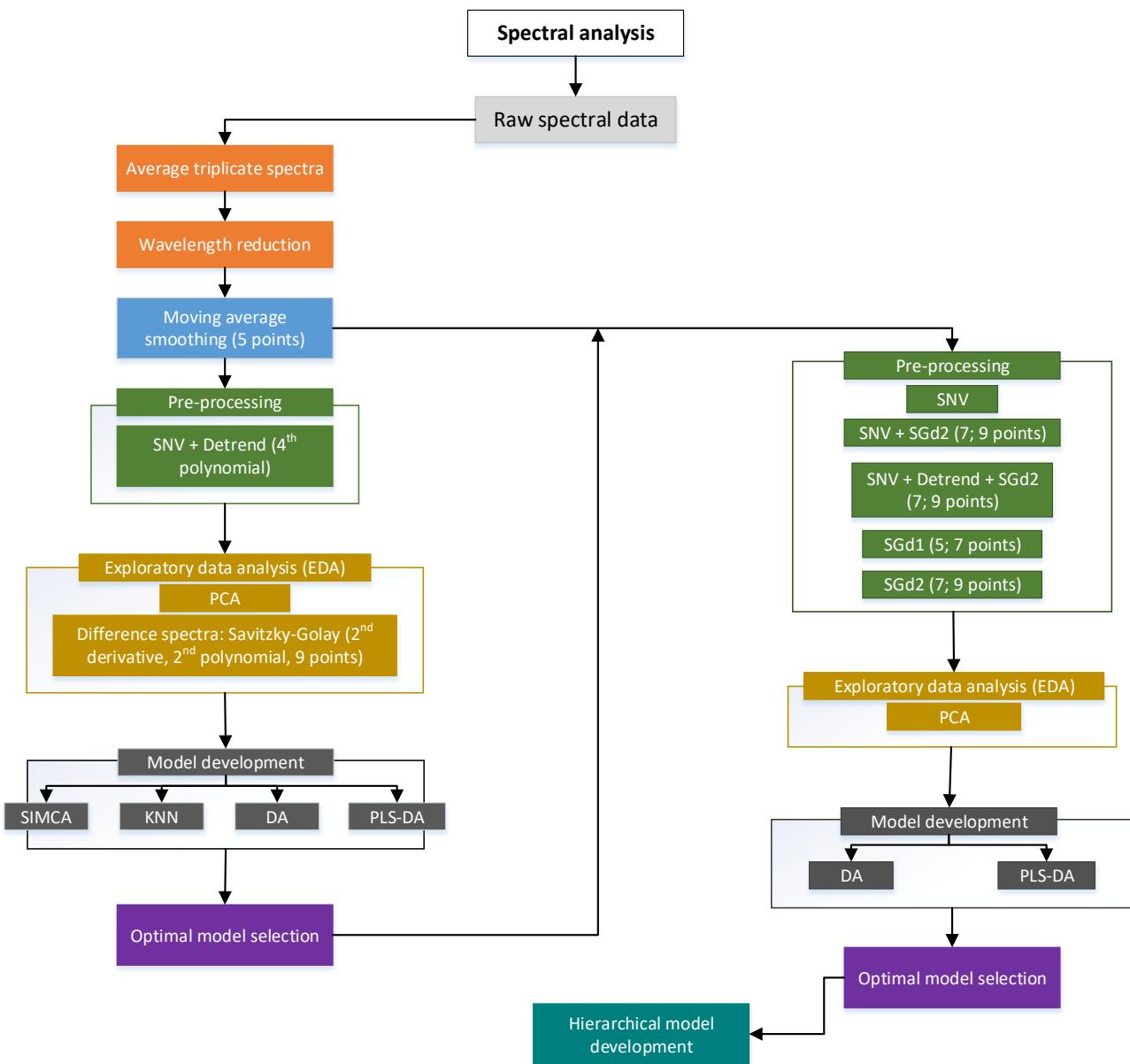


Figure 3.4 Schematic of the spectral data analysis process. Firstly, the smoothed data (pre-processed with SNV + detrend) was subjected to various unsupervised techniques in order to identify the optimal supervised method. After the optimal method was identified, the smoothed data was pre-processed using various techniques and models were developed using the previously identified optimal methods. This was done to see whether or not the method could be optimised. Thereafter the models were compared and an optimal model was selected.

3.4.1 Pre-processing

Pre-processing was performed to eliminate specific non-chemical biases from the acquired spectra and therefore prepare the data before the next processing steps, in order to develop simple and robust models (Blanco & Villarroya, 2002). Prior to pre-processing the spectroscopic data was subjected to moving average smoothing (5 points) which helps to reduce the noise in the data without reducing the number of variables. Several pre-processing techniques [standard normal variate (SNV) (Barnes *et al.*, 1989), de-trending (DT) and first and second derivatives (Savitzky-Golay) (Savitzky & Golay, 1964)] were evaluated in combination with different cross-validation methods (leave-one-out and venetian blinds) to determine which combination

would yield the best results. Based on literature and previous studies concerning meat, the use of SNV + detrend pre-processing was first evaluated. Thereafter, the data were subjected to various pre-processing techniques in different combinations to see whether or not the classification accuracy would improve.

3.5 Exploratory data analysis (EDA)

3.5.1 Principal component analysis (PCA) and difference spectra

Principal component analysis (PCA) was applied to the mean-centred absorbance spectra. For consistency, all the PCA models were calculated with seven principal components (PCs). Subsequently, the PCA scores plots were used to explore the data and detect potential- and/or true outliers. In order to identify outliers, the scores plot was examined in combination with the influence plot. In an influence plot, the F residuals of the samples were plotted against their leverages. The leverages were used to determine the influence of a sample in the PCA model. Leverages measure the distance from the projected sample (i.e. its model approximation) to the centre (mean points). Samples with high leverages have a stronger influence on the model than other samples and they may or may not be outliers (potential outliers), but they are influential. Before removing these potential outliers from the datasets, the origin of the sample deviation had to be investigated. The potential outliers' spectral information was investigated to detect the primary reason for the deviation. When an outlier was identified, it was removed, except for occasions when it could not be attributed to spectral inconsistencies as it may then contain valuable information. In contrast, an influential outlier (high residual + high leverage) is the worst case as it is indicative of a true outlier. Consequently, the outliers were removed and the PCA (Wold *et al.*, 1987a) models were recalculated to further explore the data, where scores plots, loading line plots and correlation loadings were used to locate and identify clustering and wavelengths of interest.

Another form of data exploration was calculating the difference spectra. The mean spectra of each class set (species, muscle type, fresh vs. frozen-thawed) was subjected to Savitzky-Golay transformation (2nd derivative, 2nd polynomial order, 9 smoothing points). The difference spectra were then calculated and used to identify wavelengths, indicative of spectral differences. Exploratory data analysis was then followed by supervised classification, a technique that classifies data according to known information.

3.6 Multivariate data analysis (MDA)

3.6.1 Model development

Supervised classification and discrimination models were developed to characterise the game meat samples, differentiate between species and muscle types, determine whether there were differences between the fresh and previously frozen meat samples and at what stage these changes could be detected. The supervised classification and discrimination techniques evaluated were; soft independent modelling of class analogy (SIMCA) (Wold & Sjöström, 1977), K-nearest neighbour (KNN) (Sebestyen, 1962), discriminant analysis [linear (LDA), quadratic (QDA)] and partial least squares discriminant analysis (PLS-DA) (Wold *et al.*,

1987b). After the data was subjected to various pre-processing techniques and outliers were removed, the following step was to split the data into a calibration and validation set by means of the Kennard-Stone (KS) algorithm (Kennard & Stone, 1969). In summary, the KS algorithm selects sample pairs with the largest Euclidean distance of x-vectors (predictors). Thereafter, samples are sequentially selected to maximise the Euclidean distance between x-vectors of the already selected- and remaining samples. This is then done repeatedly until the defined number of samples are attained. For each sample pair; i and j , the Euclidean distance in x space is defined as (Kennard & Stone, 1969):

$$d_x(i,j) = \|x_i - x_j\| = \sqrt{\sum_{k=1}^M (x_{ik} - x_{jk})^2} \quad i, j \in [1, N] \quad \dots \text{equation 3.1}$$

where M is the number of variables in x space and N is the number of samples. x_{ik} and x_{jk} are the k th variable for samples i and j , respectively. The number of samples selected for the calibration set comprised of approximately two-thirds (70 %) of the total dataset whilst the remaining one-third (30 %) was selected as the independent validation set. Upon completion of the development of calibration models, pre-processing techniques were optimised and subsequently performed to evaluate their performance. SIMCA, KNN, DA and PLS-DA models were constructed to differentiate between the multiple classes, independently. Firstly, models were calculated to distinguish between the four different game species (black wildebeest, zebra, springbok and ostrich), regardless of the meat being fresh, previously frozen or muscle type. For each of the four species, separate SIMCA, KNN, DA and PLS-DA models were then calculated to distinguish between; fresh vs. frozen-thawed meat, frozen periods and different muscle types.

3.6.1.1 Soft independent modelling of class analogy (SIMCA)

The aim of SIMCA is to classify samples based on spectral similarities (Wold & Sjöström, 1977). Classification models were built using PC scores to model the shape and position of the objects formed by the samples in a row space, for class definition. A multidimensional box was constructed for each class, where after the classification of a future sample was performed by calculating in which box (class) the sample lies. The SIMCA models were calculated on the calibration data using venetian blinds cross-validation. The models were then applied to the independent validation data and a classification output was generated. The goal of this classification method was to assign new objects to the class to which they show the largest similarity rather than identical behaviour. In this context similarity has a geometrical meaning and is measured by the distance between objects (Massart *et al.*, 1988). Therefore, a small distance means a high similarity.

3.6.1.2 K-nearest neighbour (KNN)

K-nearest neighbour (KNN) is a distance-based non-parametric discriminant classification method (Fix & Hodges, 1951). KNN models were performed on the PC scores. The classification of an unknown was performed by calculating the distance between the unknown and a set of samples with a known class

membership (training set). Euclidean distance was used in conjunction with KNN, to measure the proximity between samples in a row space. A classification was then made using the closest K samples. To select the optimal K , a venetian blinds cross-validation procedure was applied to a set of data with known class identities. Each sample in the training set was treated as an unknown and was classified using the remaining training set samples. This was repeated using different numbers of nearest neighbours ($k = 2$; $k = 3$; $k = 5$) for the classification. A confusion matrix was then used to describe the optimal K and performance of the calibration model. The models were then applied to the independent validation data and a classification output was generated.

3.6.1.3 Discriminant analysis (DA)

Discriminant analysis (DA) models were constructed using PC scores. The number of PCs to be used was determined by means of the explained variance plot (x-variance vs. PCs), therefore taking the lowest value of variance (%) before the curve started to plateau. For every dataset, different combinations of pre-processing techniques, number of PCs and distance calculations (linear, mahalanobis, quadratic) were evaluated by means of leave-one-out full cross-validation. Linear discriminant analysis (LDA) (euclidean- and mahalanobis distance) works by calculating an optimal linear projection which maximises the inter-class variance while at the same time minimising the intra-class variance (Fisher, 1936). While the quadratic discriminant analysis (QDA) calculates a non-linear decision boundary using a quadratic function. The best model combination, resulting in the lowest misclassification rate, was then selected and calibration models were built. Objects (samples) were classified by calculating the distance to the centre of each class. Objects were then assigned to the class with the smallest distance. The models were then applied to the independent validation data and a classification output was generated.

3.6.1.4 Partial least squares discriminant analysis (PLS-DA)

Partial least squares discriminant analysis (PLS-DA) is similar to partial least squares (PLS) regression, as it uses the latent variable approach to find fundamental relations between two matrices (\mathbf{X} and \mathbf{Y}) (Wold *et al.*, 2001a; Wold *et al.*, 2001b; Esbensen *et al.*, 2002). PLS uses the y-data structure to decompose \mathbf{X} so that the outcome constitutes an optimal regression vector (Williams, 2013). PLS-DA works similarly, but instead of measuring y-data, dummy variables are used which are indicators of groups (Westerhuis *et al.*, 2008). The objective of this method was to successfully predict group membership, hence the classification of spectra to classes. The PLS-DA models were calculated on the calibration data using venetian blinds cross-validation to optimise the number of latent variables. During the developmental procedure of the discriminant models, a calibration matrix was constructed using every sample by assigning random values to dummy variables. Therefore, the calibration matrix (\mathbf{Y}) represented the class memberships using numbers such as ones and zeros, and was then paired with the training set (\mathbf{X}) (Barker & Rayens, 2003). Once completed, PLS was applied in the typical manner.

PLS-DA calibration models were then applied to the independent validation data and a classification output was generated. For example, if the sample spectrum belonged to the pre-defined correct group then a value of one was assigned to that spectrum. A zero-value indicated that the spectrum did not belong to the correct group. A threshold of 0.5 was used, and predicted values above this threshold were classified as a one and values below 0.5 were classified as a zero. Misclassification occurred when a sample was classified as an incorrect group, therefore receiving an incorrect binary value (1 or 0).

3.6.2 Performance measures

The overall performance of the individual models, in combination with the various pre-processing techniques, were validated by calculating the classification accuracy (equation 3.2), false positive error (equation 3.3) and false negative error (equation 3.4). The efficacy of the overall model was illustrated by the classification accuracy. A false positive occurred when a negative response (incorrect class) was incorrectly classified as a positive response (correct class) and a false negative was when a positive response (correct class) was incorrectly classified as a negative response (incorrect class). Consequently, the false positive error describes the misclassification of an incorrect class in the model as a correct class (e.g. springbok classified as zebra), while the false negative error describes the misclassification of a correct class to another incorrect class in the model (zebra classified as springbok/black wildebeest/ostrich).

The sensitivity (equation 3.5), specificity (equation 3.6), precision (equation 3.7), F1 score (equation 3.8) and misclassification rate (equation 3.9) were also calculated as part of the performance measures. Sensitivity (recall or true positive rate), describes the probability that a positive response would be correctly classified. It describes how often the positive response prediction was correct, therefore describing the efficacy of the model according to classes. Specificity (true negative rate), describes the probability that a negative response would be correctly classified and how often the prediction was correct. Sensitivity, along with specificity were used to evaluate the performance of the classification algorithm for a single class. The predictive power of the model is defined by its precision, hence calculating the predicted value for each class. F1 score describes the weighted mean between precision and recall (sensitivity). The misclassification rate (classification error) describes how often the classifier was incorrect.

$$\text{Classification accuracy (\%)} = \frac{\text{TP} + \text{TN}}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \times 100\% \quad \dots\text{equation 3.2}$$

$$\text{False positive error (\%)} = \frac{\text{FP}}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \times 100\% \quad \dots\text{equation 3.3}$$

$$\text{False negative error (\%)} = \frac{\text{FN}}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \times 100\% \quad \dots\text{equation 3.4}$$

$$\text{Sensitivity or Recall (\%)} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \times 100\% \quad \dots\text{equation 3.5}$$

$$\text{Specificity (\%)} = \frac{\text{TN}}{(\text{TN} + \text{FP})} \times 100\% \quad \dots\text{equation 3.6}$$

$$\text{Precision (\%)} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \times 100\% \quad \dots\text{equation 3.7}$$

$$\text{F1 Score (\%)} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \quad \dots\text{equation 3.8}$$

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

Where:

True Positives (TP) = Positive response correctly classified as a positive response

True Negatives (TN) = Negative response correctly classified as a negative response

False Positives (FP) = Negative response incorrectly classified as a positive response

False Negative (FN) = Positive response incorrectly classified as a negative response

3.7 Hierarchical model development

Due to the complexity of the data, it was not possible to classify multiple classes (species, fresh vs. frozen-thawed, frozen period, muscle type), using a single classification and/or discrimination model (SIMCA, KNN, DA, PLS-DA). Therefore, to solve the classification/discrimination problems and handle the increased detail of the data, it had to be divided into sub-groups with individual models. A multilevel hierarchical model was constructed by selecting between the general categories (species), then sub-models were used to sub-divide those general categories using increasingly specific classification and differentiation models. The specific classification and differentiation models were used to sub-divide the species categories into multiple class categories (e.g. fresh vs. frozen-thawed, frozen period and muscle type). The multilevel hierarchical model was then applied to the independent validation data and a classification output was generated.

3.8 References

- Barker, M. & Rayens, W. (2003). Partial least squares for discrimination. *Journal of chemometrics*, **17**, 166-173.
- 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.
- Blanco, M. & Villarroya, I. (2002). NIR spectroscopy: a rapid-response analytical tool. *Trends in Analytical Chemistry*, **21**, 240-250.
- DAFF (2000). Meat Safety Act (Act No. 119 of 1990, No. R. 342). In: *Department of Agriculture, Forestry and Fisheries (DAFF)*. Pretoria, South Africa: Government Printer.
- Esbensen, K.H., Guyot, D., Westad, F. & Houmoller, L.P. (2002). *Multivariate data analysis in practice: An introduction to multivariate data analysis and experimental design*. Pp. 598. Oslo: CAMO Process AS.
- Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, 179-188.

- Fix, E. & Hodges, J.L. (1951). Discriminatory analysis, nonparametric discrimination: Consistency properties. Randolph Field, Texas: USAF School of Aviation Medicine.
- Kennard, R.W. & Stone, L.A. (1969). Computer aided design of experiments. *Technometrics*, **11**, 137-148.
- Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y. & Kaufman, L. (1988). *Chemometrics: A Textbook*. Pp. 477. Amsterdam: Elsevier Science Publishers.
- Savitzky, A. & Golay, M.J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, **36**, 1627-1639.
- Sebestyen, G. (1962). Pattern recognition by an adaptive process of sample set construction. *IRE Transactions on Information Theory*, **8**, 82-91.
- Westerhuis, J.A., Hoefsloot, H.C., Smit, S., Vis, D.J., Smilde, A.K., van Velzen, E.J., van Duijnhoven, J.P. & van Dorsten, F.A. (2008). Assessment of PLSDA cross validation. *Metabolomics*, **4**, 81-89.
- 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* (PhD Thesis). South Africa: University of Stellenbosch.
- Wold, S., Esbensen, K. & Geladi, P. (1987a). Principal component analysis. *Chemometrics and intelligent laboratory systems*, **2**, 37-52.
- Wold, S., Geladi, P., Esbensen, K. & Öhman, J. (1987b). Multi-way principal components-and PLS-analysis. *Journal of chemometrics*, **1**, 41-56.
- Wold, S. & Sjöström, M. (1977). SIMCA: a method for analyzing chemical data in terms of similarity and analogy. In: *Chemometrics: Theory and Application*. Pp. 243 - 282. ACS Publications.
- Wold, S., Sjöström, M. & Eriksson, L. (2001a). PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, **58**, 109-130.
- Wold, S., Trygg, J., Berglund, A. & Antti, H. (2001b). Some recent developments in PLS modeling. *Chemometrics and intelligent laboratory systems*, **58**, 131-150.

Chapter 4

Results and Discussion

4.1 Species determination

4.1.1 Spectral analysis

The mean spectrum of each species was computed between 920 and 1651 nm and plotted to investigate, determine and compare the chemical properties (**Figure 4.1 – 4.2**). The mean spectra of the four species followed a similar trend with comparable absorption bands, however the intensity of the bands varied. The intensity differences cannot be attributed to the internal chemical composition, as the unprocessed spectra may contain physical effects such as light scattering. Three prominent absorption bands were exhibited at 970, 1193 and 1428 nm.

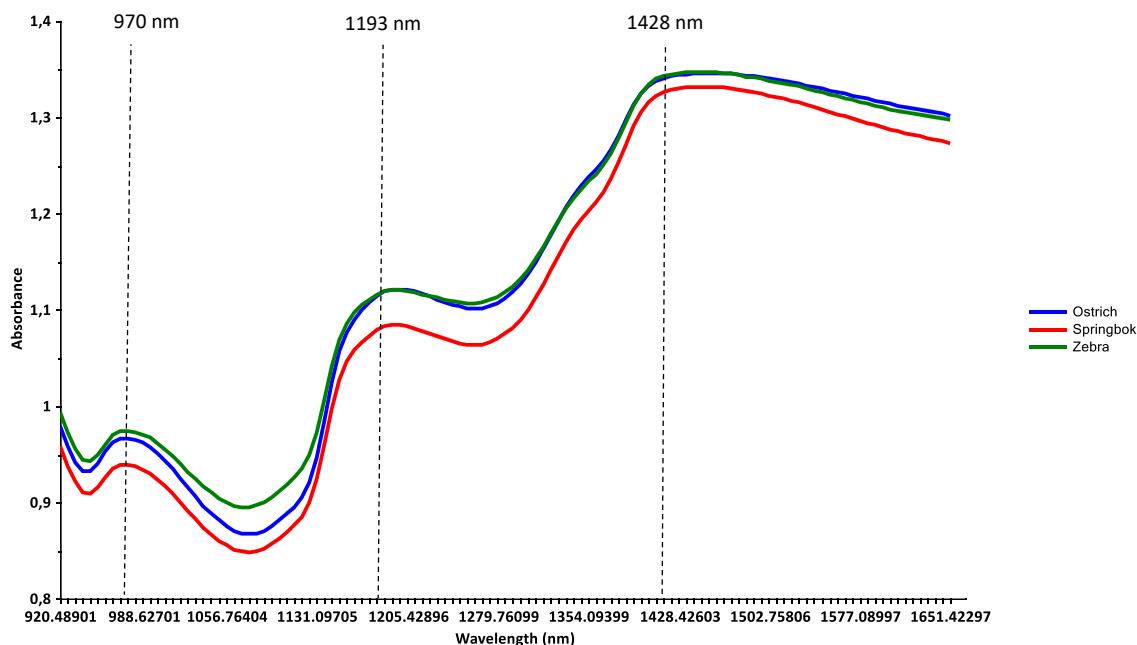


Figure 4.1 Unprocessed mean spectra for ostrich (blue), springbok (red) and zebra (green), irrespective of the muscle type.

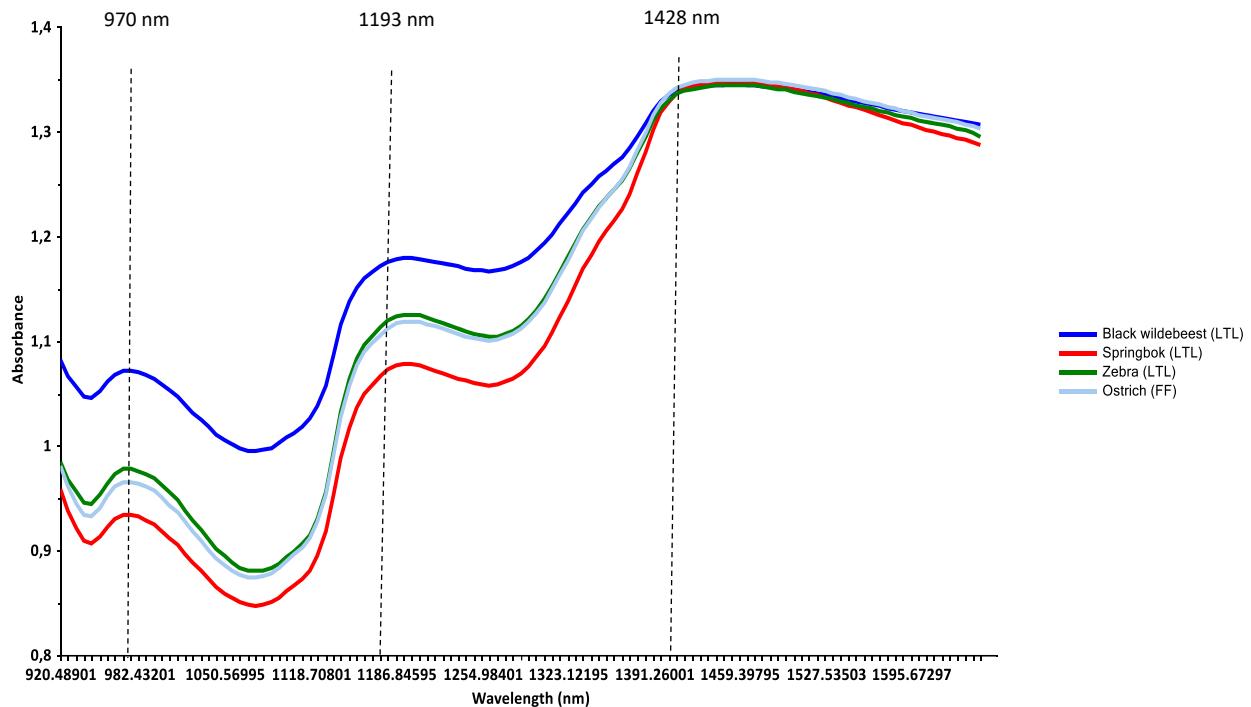


Figure 4.2 Unprocessed mean spectra for black wildebeest (blue), springbok (red), zebra (green) and ostrich (light blue), irrespective of the frozen period.

The band at 970 nm is related to the O-H second stretch overtone associated with water (Ding & Xu, 2000; Cozzolino & Murray, 2004; Barbin *et al.*, 2013b). The 1193 nm band indicates the presence of fat (C-H stretch second overtone) as specified by Osborne *et al.* (1993). Lastly, the 1428 nm band represents the N-H stretch first overtone related to the CONH₂ group associated with the peptide bonds in proteins (Osborne *et al.*, 1993). The broad band at 1428 nm also has a contribution from moisture around 1420 and 1440 nm (O-H stretch first overtone). Therefore, the broad band at 1428 nm may be related to protein and moisture content (Osborne *et al.*, 1993).

4.1.2 Species determination irrespective of muscle type

This data set consisted of three species [zebra (*Equus quagga burchelli*), springbok (*Antidorcas marsupialis*), ostrich (*Struthio camelus*)], multiple muscle types for each species and the meat samples were frozen for one month. The aim of **Section 4.1.2** was to differentiate between the three game species, irrespective of the treatment (fresh or previously frozen) or the muscle type as well as to determine whether the treatments and muscle type had an effect on the species classification accuracies.

4.1.2.1 Exploratory data analysis (EDA)

4.1.2.1.1 Principal component analysis (PCA)

Minimal class separation between the species was observed in the PCA score plots (**Figure 4.3 – 4.4**) of the SNV + detrend (4th order polynomial) corrected data. PC1 accounted for 86% of the variance (**Figure 4.3a**), and PC2 6% (**Figure 4.3b**), whereas PC3 (**Figure 4.4a**), PC4 and PC5 (**Figure 4.4b**) only accounted for approximately 3%, 2% and 1% of the variance, respectively. This illustrates that the variation, seen as the separation, was explained in the first two components.

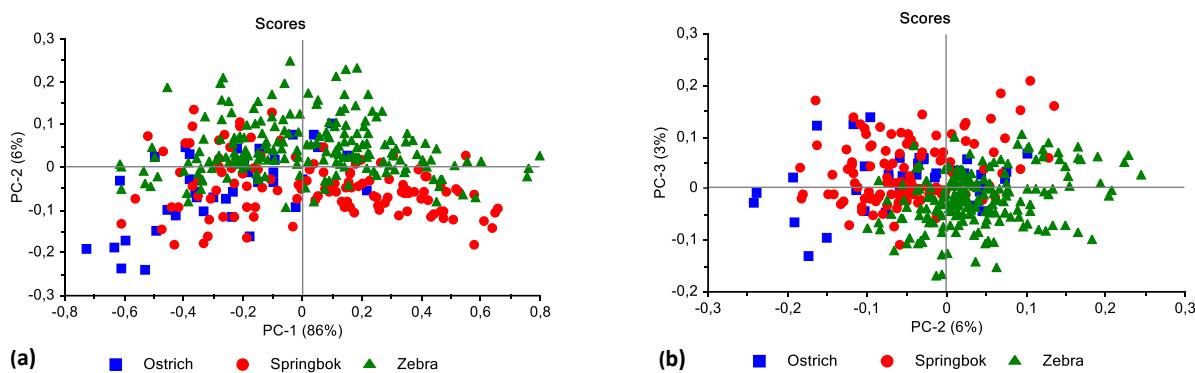


Figure 4.3 PCA analysis (SNV + detrend pre-processed) of species for ostrich (blue), springbok (red) and zebra (green) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC1 (86%) vs. PC2 (6%); and (b) PCA score plot of PC2 (6%) vs. PC3 (3%).

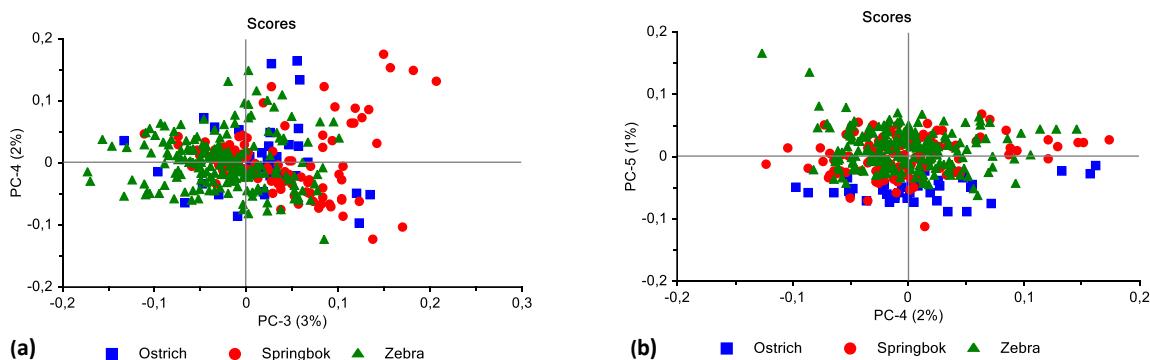


Figure 4.4 PCA analysis (SNV + detrend pre-processed) of species for ostrich (blue), springbok (red) and zebra (green) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC3 (3%) vs. PC4 (2%); and (b) PCA score plot of PC4 (2%) vs. PC5 (1%).

The variance observed in the direction of PC1 (**Figure 4.3a**) may be attributed to the intra-class separation, while PC2 (**Figure 4.3b**) can be attributed to the variances between the three classes due to differences in macronutrient composition. This separation was predominantly due to differences in moisture, fat, protein and pH as illustrated in the loadings plot (**Figure 4.6a**). The score plots shown in **Figure 4.4** (PC3 vs. PC4 and PC4 vs. PC5) illustrated no distinct clusters for the separation between species. However, the third-, fourth- and fifth PC illustrates the intra-class variance of each species and the separation could be

attributed to moisture, fat, protein, pH and consequently meat tenderness. This internal variance was likely due to animal-to-animal variance and the anatomical location of different muscle types used in this study. However, the overall separation between the species was not affected by this internal variation.

The lack of separation between the different species indicates the similarity in their spectral signatures; however, there are numerous components (e.g. physical and chemical) that could differ and lead to slight spectral differences. The reasons for the differences are most likely attributed to the fluctuation of macronutrient composition. To study the relationship between the samples' spectra and explain the origin of variance between the species, one must use the loadings line plot in combination with the score plot. The scores depict the correlation between the objects whereas the loadings express the variation within the variables. Therefore, the loadings can be interpreted as the relative weightings of the spectral variables in each PC and the plots show how each variable (wavelength band) contributed to the separation (Esbensen *et al.*, 2002). A variable with positive loadings correlates to the positive side of the corresponding PCs score plot and exhibits the intensity of the spectral response (Esbensen *et al.*, 2002). The same is revealed by a variable with negative loadings correlating with the negative side of the score plot.

The loadings line plot was also examined in conjunction with the correlation loadings plot. With the standard loadings plot it is not always clear which variables are significant for explaining the variance. Therefore, the correlation loadings (**Figure 4.5b – 4.9b**) were used to identify variables with high loadings (i.e. close to +1 or -1) that consequently indicates whether or not the loading is interpretable. In the correlation loadings plot the red dashed lines indicate how much variance is considered by the model. The upper/outer lines indicate 100% of the explained variance for the given variable, whereas the inner lines indicate 50%. Values that lie within the upper and outer bounds of the plot are therefore interpretable and modelled by the latent variable. Those occurring between the two inner bounds are not. Therefore, the correlation loading plots are especially useful when identifying and interpreting important wavelengths.

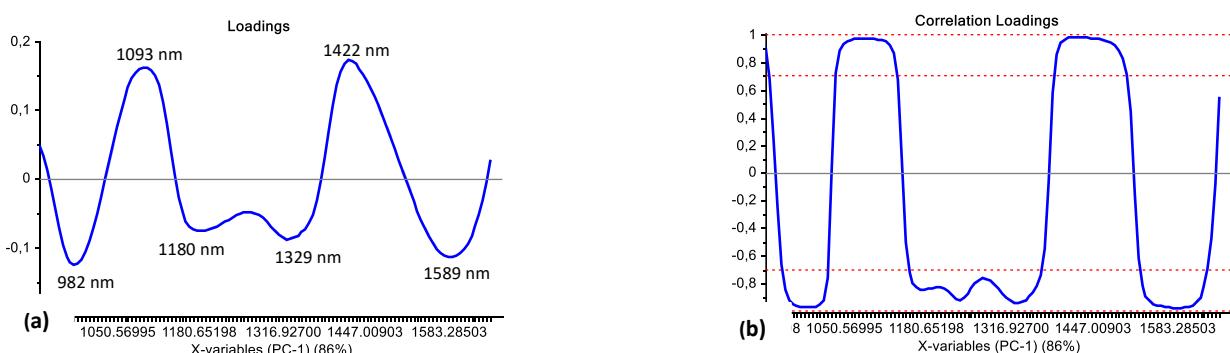


Figure 4.5 (a) PCA loadings line plot and (b) correlation loadings for PC1 (86%) with interpretable bands at 982, 1093, 1180, 1329, 1422 and 1589 nm.

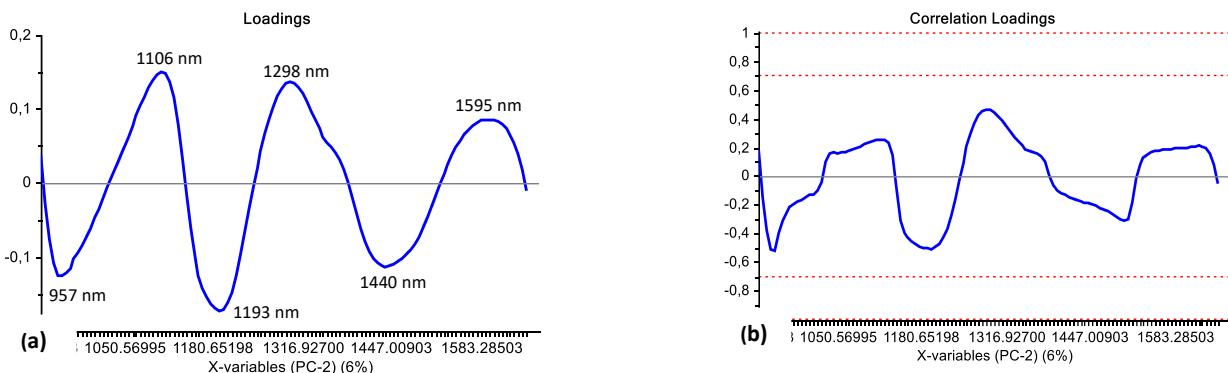


Figure 4.6 (a) PCA loadings line plot and (b) correlation loadings for PC2 (6%) with bands at 957, 1106, 1193, 1298, 1440 and 1595 nm.

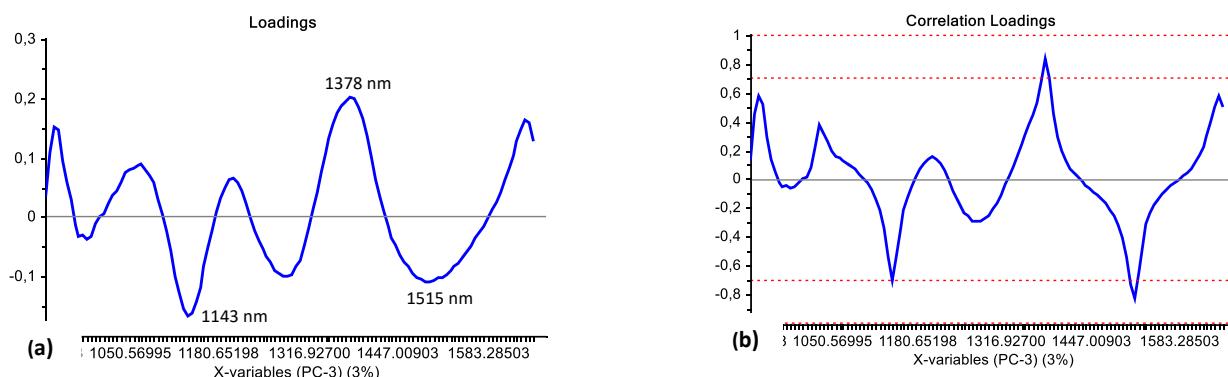


Figure 4.7 (a) PCA loadings line plot and (b) correlation loadings for PC3 (3%) with interpretable bands at 1143, 1378 and 1515 nm.

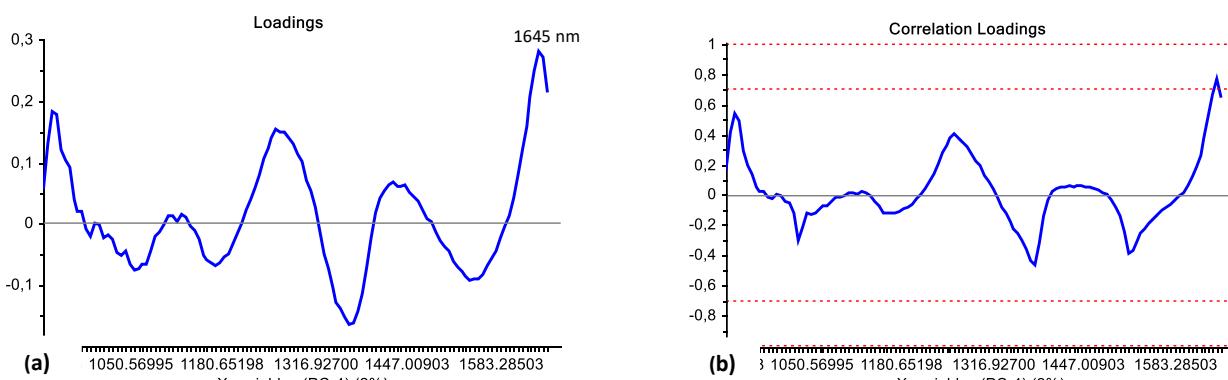


Figure 4.8 (a) PCA loadings line plot and (b) correlation loadings for PC4 (2%) with an interpretable band at 1645 nm.

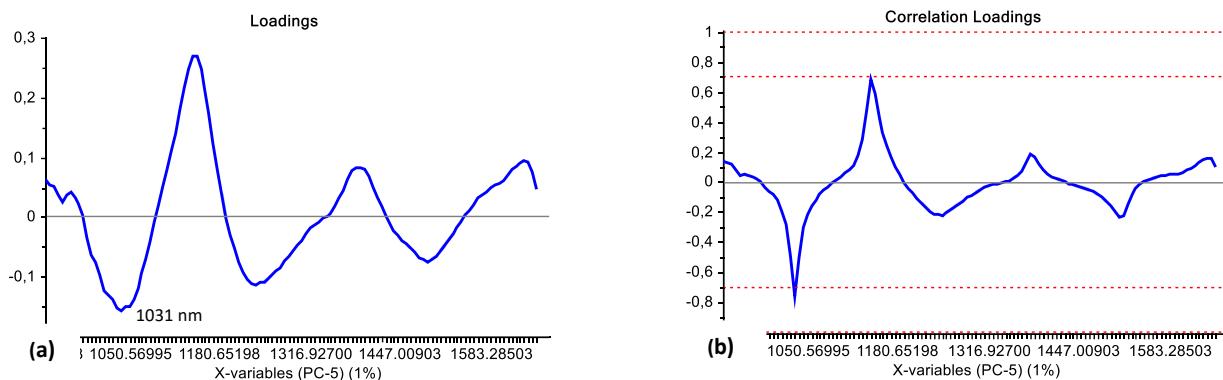


Figure 4.9 (a) PCA loadings line plot and (b) correlation loadings for PC5 (1%) with an interpretable band at 1031 nm.

The loadings line plot of PC1 (**Figure 4.5a**) exhibited interpretable positive bands at 1093 and 1422 nm and negative bands at 982, 1180, 1329 and 1589 nm. When evaluating the positive and negative loading bands of PC1, in combination with the score plot of PC1 (86%) vs. PC2 (6%) (**Figure 4.3a**), the intra-class species separation is mainly based on the positive spectral bands at 1422 nm, which is related to moisture (Osborne *et al.*, 1993), and 1093 nm, associated with pH (ElMasry *et al.*, 2012). The positive bands correspond with the positive side of the score plot (**Figure 4.3a**), thus illustrating that the intra-class separation is due to the difference between fresh and frozen-thawed meat samples. These results will be discussed thoroughly in **Section 4.2**.

Although the loadings plot of PC2 (**Figure 4.6a**) had bands at 957, 1106, 1193, 1298, 1440 and 1595 nm, these were not interpretable as the variables only contributed to 50% of the explained variance, thus explaining the lack of separation between the species illustrated in **Figure 4.3b**. The correlation loading plot illustrated and confirmed this (**Figure 4.6b**). These bands are associated with pH (957 nm) (Barbin *et al.*, 2012b), moisture and tenderness (1106 nm) (ElMasry *et al.*, 2012), fat (1193 nm) (Barbin *et al.*, 2013a) and protein (1298 and 1440 nm) (Kamruzzaman *et al.*, 2012b).

The score plots shown in **Figure 4.4a – b** illustrated a major overlap of the three classes. Although the PCA results exhibited the lack of separation between species, the PCs, nevertheless, displayed intra-class variance for each species. Although the PCs accounted for minimal data variation, the loading line plots of PC3 (**Figure 4.7a**), PC4 (**Figure 4.8a**) and PC5 (**Figure 4.9a**) displayed interpretable absorption bands which explained the separation within each species. PC3 (3%) had a positive band at 1378 nm (C-H stretch + C-H deformation) and negative bands at 1143 (C-H stretch second overtone) and 1515 nm (N-H stretch first overtone and O-H stretch first overtone). These bands are associated with meat tenderness (1143 nm) (Kamruzzaman *et al.*, 2013), pH (1378 nm) (ElMasry *et al.*, 2012), protein and moisture (1515 nm) (Osborne *et al.*, 1993; Barbin *et al.*, 2013a). PC4 (2%) had a positive band at 1645 nm (C-H stretch first overtone) and is related to moisture, fat and protein as stated by Barbin *et al.* (2013a). Lastly, PC5 (1%) displayed a negative band at 1031 nm (N-H stretch second overtone) which is associated with protein (Osborne *et al.*, 1993) and meat tenderness (ElMasry *et al.*, 2012). The internal variation was likely due to the different muscle types

used as well as their diverse anatomical locations. PCA gave a global representation of the data, thus the need for further exploratory data analysis to show the specific differences in the spectral data.

4.1.2.1.2 Difference spectra

The difference spectra of each species illustrated that the separation between the species was a result of numerous absorption bands (**Figure 4.10 – 4.12**). From this, it was deduced that ten prominent and regular occurring absorption bands were exhibited at 939, 957, 1075, 1118, 1162, 1366, 1403, 1409, 1465 and 1645 nm. The absorption band at 939 nm is related to the third overtones of C-H stretching modes from different structures and indicates the presence of fat and protein (Osborne *et al.*, 1993; Barbin *et al.*, 2012a). Therefore, the difference spectra (**Figure 4.11**) illustrates that springbok has a lower protein content compared to that of zebra and ostrich, which corresponds to literature. Hoffman and Wiklund (2006) reported a protein content of *ca.* 18.8 g/100g for springbok which is considerably lower than that of zebra (*ca.* 22.3 g/100 g) (Hoffman *et al.*, 2016) and ostrich (*ca.* 20.9 g/100 g) (Sales, 1996).

Kamruzzaman *et al.* (2012b) found that the band at 957 nm is related to the water content of lamb as well as the presence of fat. The spectral differences between ostrich, zebra and springbok can therefore be ascribed to the higher moisture content of ostrich (*ca.* 76.6 g/100 g) (Sales, 1996) in comparison with zebra (*ca.* 76.4 g/100 g) (Hoffman *et al.*, 2016) and springbok (*ca.* 74.7 g/100 g) (Hoffman & Wiklund, 2006). From these results and the difference spectra (**Figure 4.12**), it is also evident that zebra has a higher moisture content than springbok. This phenomenon is supported by the current findings illustrated in the PCA scores plot (**Figure 4.3b**) as the zebra samples were separated by a higher absorption band at 1106 nm (**Figure 4.6a**), indicative of moisture (ElMasry *et al.*, 2012). This wavelength (957 nm) was also identified when categorising red meat samples based on physiochemical characteristics (Kamruzzaman *et al.*, 2012a).

The 1075 nm band represents the C-H stretch combination vibrations (Osborne *et al.*, 1993) and can be associated with fat of beef (ElMasry *et al.*, 2013) and pH of turkey (Iqbal *et al.*, 2013). Because beef, zebra and springbok are all ungulates the absorption band at 1075 nm was ascribed to the presence of fat. The difference spectra of zebra (**Figure 4.10**) and springbok (**Figure 4.11**) illustrated that there was little difference between these two species due to the semi-straight line. This indicates that both these species have a similar fat content and literature could support this. Studies reported a fat content of *ca.* 1.7 g/100 g for springbok (Hoffman & Wiklund, 2006; Hoffman *et al.*, 2007b), and *ca.* 1.5 g/100 g for zebra (Hoffman *et al.*, 2016). When evaluating the difference spectra illustrated in **Figure 4.10 – 4.11**, it is observed that there is a substantial difference between the ostrich, zebra and springbok species. Because ostrich is part of the class Aves, similar to turkey, the absorption band was ascribed to the difference in pH. The spectral differences between ostrich and the remaining species can therefore be ascribed to the higher pH value of ostrich (*ca.* 6.1) (Sales, 1996; Leygonie *et al.*, 2012b) in comparison with zebra (*ca.* 5.7) (Onyango *et al.*, 1998) and springbok (*ca.* 5.8) (Hoffman *et al.*, 2007a).

Barbin *et al.* (2012b) and ElMasry *et al.* (2012) found that the bands at 1118, 1162, 1366 and 1645 nm (second and first stretch overtones and C-H combination/deformation bands) can be related to pH and is therefore, responsible for the spectral differences at these absorption bands. The band at 1645 nm can also be associated with water content, fat and protein of pork (Barbin *et al.*, 2013a) as well as water content and protein of ham (Talens *et al.*, 2013). When comparing the spectra to the chemical characteristics reported in literature, it was possible to conclude that the absorption band at 1645 nm can be ascribed to protein. When examining the difference spectra of springbok (**Figure 4.11**), it is observed that the protein difference for zebra is greater than that of ostrich, thus suggesting that the protein content of springbok differs less from ostrich and more from zebra. This observation is supported by literature as Hoffman and Wiklund (2006) reported springbok to have a protein content of *ca.* 18.8 g/100g which is considerably lower and differs substantially from that of zebra (*ca.* 22.3 g/100 g) (Hoffman *et al.*, 2016) and to a lesser extent from ostrich (*ca.* 20.9 g/100 g) (Sales, 1996). A clear moisture band is observed, due to O-H stretching and O-H bending combinations, at 1403 and 1409 nm (Osborne *et al.*, 1993; Liu & Chen, 2001). These bands were also found to be related to the pH of beef (ElMasry *et al.*, 2012). Lastly, the 1465 nm band represents the N-H stretch first overtone related to the CONH₂ group associated with the peptide bonds in proteins (Osborne *et al.*, 1993).

The difference spectra indicated the specific differences in the spectral data, responsible for the separation between the three different game species. These results indicate that the physiochemical characteristics were similar to that obtained with conventional analysis methods (Sales, 1996; Onyango *et al.*, 1998; Van Zyl & Ferreira, 2004; Hoffman & Wiklund, 2006; Hoffman *et al.*, 2007a; Hoffman *et al.*, 2007b; Leygonie *et al.*, 2012b; Hoffman *et al.*, 2016). Therefore, the differences and separation between species can be explained by examining the physiochemical characteristics e.g. moisture, fat, protein and pH. Although no NIR studies have been done on the differentiation of South African game species, previous studies were conducted to differentiate between conventional species e.g. chicken, turkey, pork, beef, lamb (McElhinney *et al.*, 1999); beef, pork, lamb, chicken (Cozzolino & Murray, 2004); beef, pork, lamb (Kamruzzaman *et al.*, 2012a) as well as unconventional species e.g. beef and kangaroo (Ding & Xu, 1999); cattle, llama, horse (Mamani-Linares *et al.*, 2012). These studies suggest that it is possible to determine the different species based on their physiochemical characteristics, therefore indicating that zebra, springbok and ostrich, irrespective of the treatment (fresh or previously frozen) or the muscle type, has the potential to be successfully differentiated.

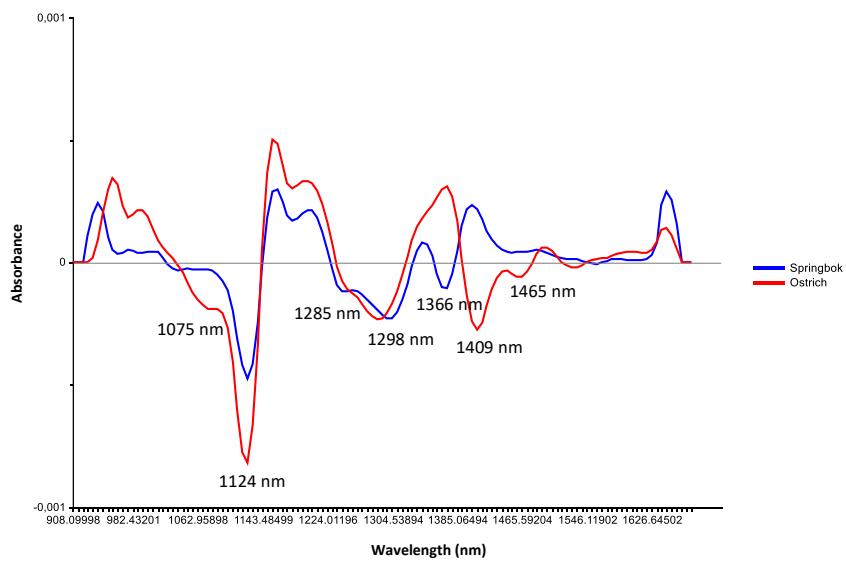


Figure 4.10 Difference spectra of zebra and the remaining species, irrespective of the muscle type, with absorption bands at 1075, 1124, 1285, 1298, 1366, 1409, 1465 nm.

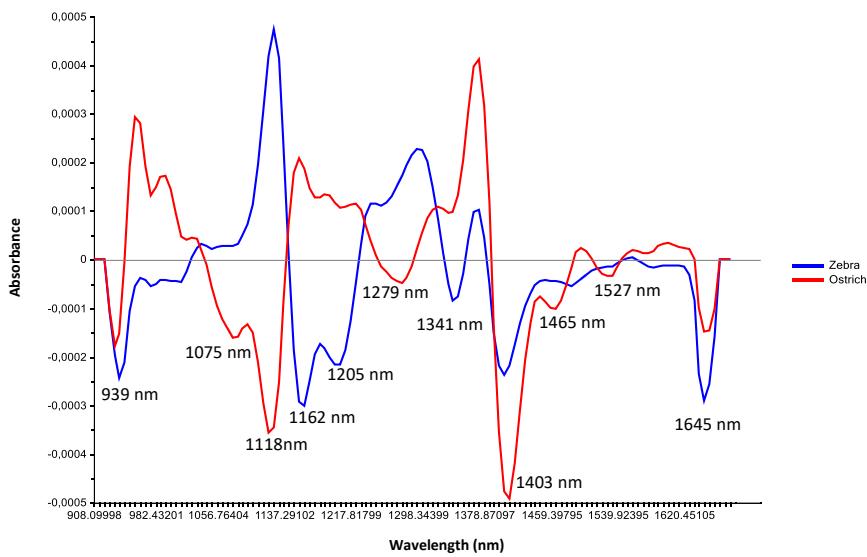


Figure 4.11 Difference spectra of springbok and the remaining species, irrespective of the muscle type, with absorption bands at 939, 1075, 1118, 1162, 1205, 1279, 1341, 1403, 1465, 1527, 1645 nm.

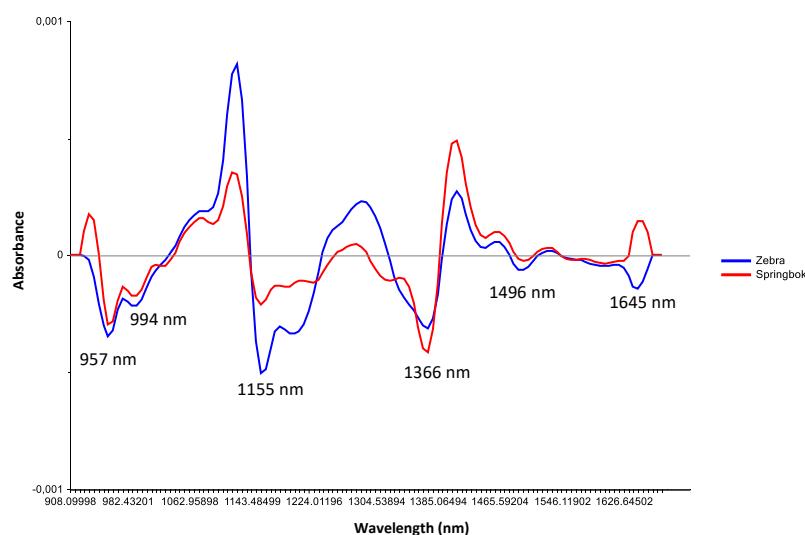


Figure 4.12 Difference spectra of ostrich and the remaining species, irrespective of the muscle type, with absorption bands at 957, 994, 1155, 1366, 1496, 1645nm.

4.1.2.2 Multivariate data analysis (MDA): Model development

4.1.2.2.1 Soft independent modelling of class analogy (SIMCA)

The SIMCA model, consisting of zebra, springbok and ostrich, gave unsatisfactory classification results. The SIMCA calibration model achieved an overall classification accuracy of 68.5% and a misclassification rate of 31.5% (**Table 4.1**). The classification results for the validation model decreased, resulting in an overall classification accuracy of 59.7% and a misclassification rate of 40.3% (**Table 4.1**). These unsatisfactory classification results confirmed what was seen in the PCA score plot (**Figure 4.3a**). The lack of separation can be ascribed to the samples' spectral similarities, and as the SIMCA algorithm aims to classify samples based on spectral similarities, the classification accuracy was low. In this context similarity has a geometrical meaning and is measured by the distance between objects (Massart *et al.*, 1988). Therefore, a small distance means a high similarity.

Table 4.1 SIMCA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for species classification.

Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
Calibration	68.5	31.5
Validation	59.7	40.3

(SIMCA) Soft independent modelling of class analogy; (SNV) Standard normal variate.

The performance measures for each individual SIMCA model is given in **Table 4.2** whilst the score plots were used to illustrate the limited classification abilities of the models (**Figure 4.13 – 4.14**). The SIMCA score plot (**Figure 4.13 – 4.14**) illustrates how well the model classified the individual samples. The ostrich

class achieved the highest classification accuracy (94.1%), followed by zebra (71.7%) and springbok (68.5%) as shown in **Table 4.2**. The PCA score plot (**Figure 4.3b**) and loadings line plot (**Figure 4.6**) support these results, as a slight separation was observed between ostrich and the other two species. The separation and correct classification of the ostrich was mainly attributed to the lower fat content and higher pH, which accounted for the spectral differences. These properties for ostrich was considerably different, compared to the other species, and a correlation was observed in the difference spectra (**Figure 4.10 – 4.12**) as well as the loadings (**Figure 4.6**), which could be compared to literature (Sales, 1996; Onyango *et al.*, 1998; Van Zyl & Ferreira, 2004; Hoffman & Wiklund, 2006; Hoffman *et al.*, 2007a; Hoffman *et al.*, 2007b; Leygonie *et al.*, 2012b; Hoffman *et al.*, 2016). The ostrich also achieved a sensitivity and specificity of 94.1%, suggesting an effective model.

The classification accuracy of zebra and springbok was lower due to the increased false positive and false negative objects classified in these species (**Figure 4.13 – 4.14**). The lack of separation can be ascribed to the samples' spectral similarities which corresponds to their similar physiochemical characteristics as reported in literature (Hoffman & Wiklund, 2006; Hoffman *et al.*, 2007a; Hoffman *et al.*, 2007b; Hoffman *et al.*, 2016). The sensitivity and specificity for zebra was 83.1% and 62.7%, respectively. This suggests that the model is very sensitive for predicting a true positive (zebra) as correct, but not very specific when predicting a true negative (ostrich and springbok) as correct. The sensitivity and specificity for the springbok was 50.3% and 85.3%. Thus, this model is less sensitive and more specific when classifying objects. These results suggest that spectra of zebra and springbok are similar due to their similar physiochemical characteristics and for this reason, the SIMCA model achieved lower classification accuracies. The ostrich can, with a high level of accuracy, be distinguished from the other two species with a small misclassification rate (**Table 4.2**). McElhinney *et al.* (1999) reported similar results when discriminating between chicken, turkey, pork, beef and lamb meat. The authors found that beef and lamb resemble each other closely, as do chicken and turkey. Thus, their SIMCA models were less accurate and did not perform well due to the spectral similarities of the samples. Therefore, the SIMCA results in the current study indicates that these game species cannot be differentiated because of their spectral similarities.

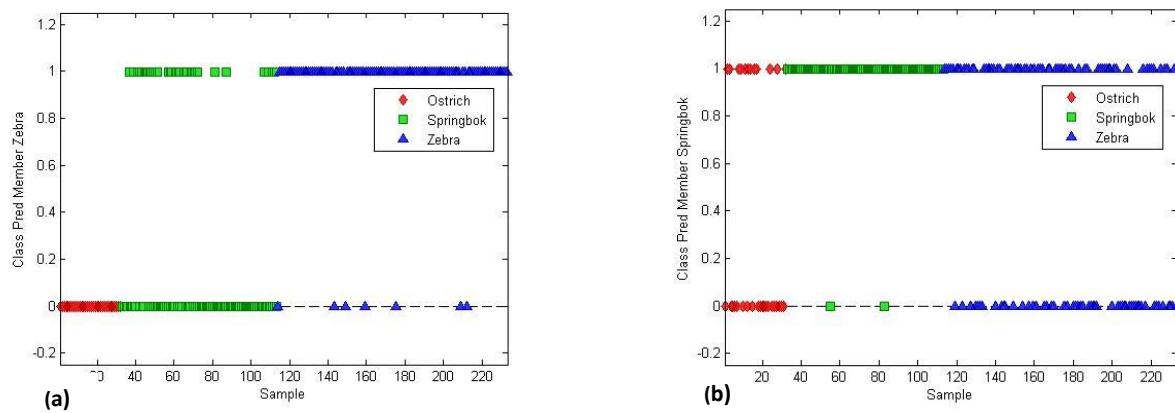


Figure 4.13 SIMCA classification (SNV + detrend pre-processed) resulting in a 71.7% and 68.5% classification accuracy for zebra and springbok, respectively. SIMCA classification score plot of ostrich (red), springbok (green) and zebra (blue), illustrating the predicted objects. (a) Score plot for predicted zebra (top) vs. remaining species (bottom) and (b) score plot of predicted springbok (top) vs. remaining species (bottom).

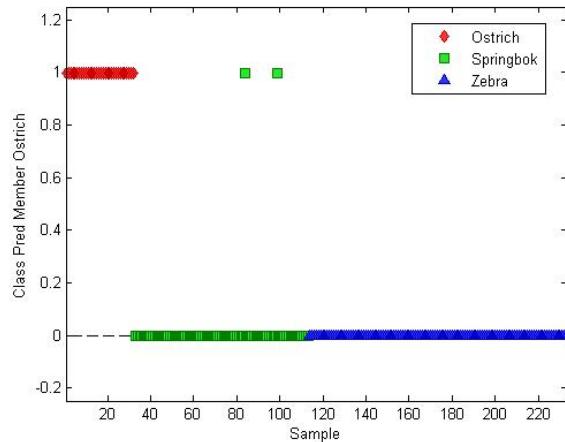


Figure 4.14 SIMCA classification (SNV + detrend pre-processed) resulting in a 94.1% classification accuracy for ostrich. SIMCA classification score plot of ostrich (red), springbok (green) and zebra (blue), illustrating the predicted objects, where the top row indicates the objects predicted as ostrich and the bottom row the objects predicted as the remaining species.

Table 4.2 The performance measures used to assess the SIMCA classification models of the three species, pre-processed with SNV + detrend.

Species	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
Zebra	71.7	21.1	7.4	83.1	62.7	63.1	71.8	28.4
Springbok	68.5	7.7	23.9	50.3	85.3	76.0	60.5	31.5
Ostrich	94.1	5.0	0.8	94.1	94.1	72.7	82.1	5.9

(SIMCA) Soft independent modelling of class analogy; (SNV) Standard normal variate.

4.1.2.2.2 K-nearest neighbour (KNN)

Table 4.3 KNN model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for species classification.

Number of neighbours (k)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
2	Calibration	74.7	25.3
	Cross-validation	82.4	17.6
	Validation	82.5	17.5
3	Calibration	78.1	21.9
	Cross-validation	78.0	22.0
	Validation	85.0	15.0
5	Calibration	78.5	21.5
	Cross-validation	79.8	20.2
	Validation	85.8	14.2

(KNN) K -nearest neighbours; (SNV) Standard normal variate.

The overall model accuracy of KNN(2) (74.7%), KNN(3) (78.1%) and KNN(5) (78.5%) suggested that five nearest neighbours would provide the best classification (Table 4.3). The KNN(5) calibration model achieved an overall classification accuracy of 78.5%, with improved classification accuracies for the cross-validation- (79.8%) and validation model (85.8%) (Table 4.3). These improved accuracies are indicative of a classification model that is not over-fitted (Miller, 2005). Although this model had fairly high accuracies, a few misclassified objects were observed amongst the three species (Figure 4.15a – c). This accounted for the misclassification rate of 21.5% for the overall model (Table 4.3). The sensitivity and specificity for the zebra was 87.1% and 86.1%, respectively (Table 4.4). This indicates that the model has a high probability of correctly classifying the zebra (Figure 4.15c). The lower sensitivity for both the springbok (71.4%) and ostrich (60%) reveals that the model was less suited for predicting these two species. This phenomenon can be explained by referring to the PCA score plot (Figure 4.16), as k -nearest neighbour was performed on the PC scores. The KS-calibration PCA score plot illustrates an overlap between the three classes, with a larger number of ostrich samples displaying a close distance to the springbok samples. Therefore, the ostrich samples are assigned to the predominant class, springbok. Hence, explaining why the model is less suited for predicting the ostrich samples.

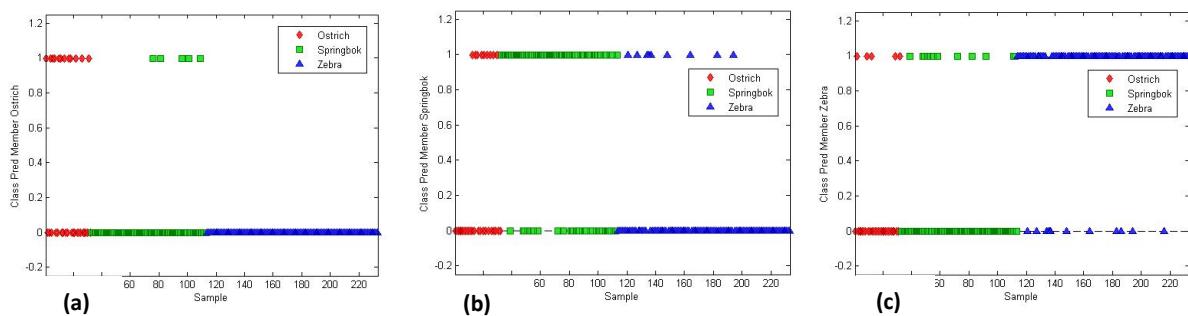


Figure 4.15 KNN ($k = 5$) classification (SNV + detrend pre-processed) resulting in an overall classification accuracy of 78.5%. KNN classification score plot of ostrich (red), springbok (green) and zebra (blue), illustrating the predicted objects. (a) Score plot of predicted ostrich (top) vs. remaining species (bottom), (b) score plot of predicted springbok (top) vs. remaining species (bottom) and (c) score plot of predicted zebra (top) vs. remaining species (bottom).

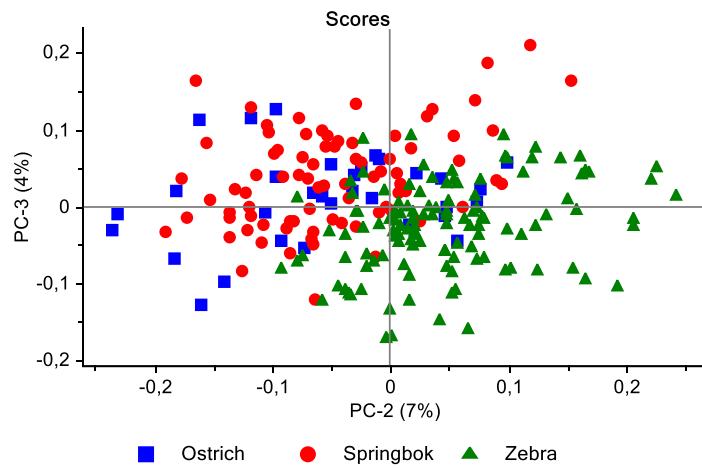


Figure 4.16 PCA analysis (SNV + detrend pre-processed) of species for ostrich (blue), springbok (red) and zebra (green) classes. Minimal class separation was observed. Scores illustrated as KS-calibration PCA score plot of PC2 (7%) vs. PC3 (4%).

The decreased model accuracy of KNN(2) and KNN(3) is mainly attributed to the misclassification of objects assigned to the springbok and ostrich classes. These classes were misclassified due to the extensive overlap observed in the PCA score plots (Figure 4.16) which can be ascribed to the closely related protein content of springbok (ca. 18.8 g/100 g) (Hoffman & Wiklund, 2006; Hoffman *et al.*, 2007b) and ostrich (ca. 20.9 g/100 g) (Sales, 1996). The sensitivity of both springbok and ostrich in KNN(2) is 64.4% and 47.9%, respectively. While the specificity is 82.5% and 94.4%, therefore confirming the model's ability to accurately predict the zebra even though it struggles with the springbok and ostrich. The reason for this is that the protein content for zebra is higher (ca. 22.3 g/100 g), as stated in literature, and differs from the other species, thus causing the zebra samples to separate in the positive direction of PC2 (Figure 4.16). The same phenomenon is observed in KNN(3) (Addendum A, Figure A2), where the sensitivity is 74.3% (springbok) and 45% (ostrich), and the specificity is 83% (springbok) and 92.1% (ostrich) (Table 4.4). From these results, it

can be said that the zebra can, with a high level of accuracy, be distinguished from the other two species with a misclassification rate below 14% (**Table 4.4**).

Therefore, the improved classification of zebra can be attributed to the increased and differing protein content that was observed in the difference spectra, loadings and confirmed by literature. The KNN results reported in this study were difficult to compare to those of McElhinney *et al.* (1999) due to differences in sample presentation (homogenised vs. whole steaks) and the use of statistical procedures, especially when referring to pre-processing [Savitzky-Golay (2nd derivative, 15 smoothing points)]. However, McElhinney *et al.* (1999) similarly reported that meat species can be distinguished based on the spectral response to compositional differences. These researchers also found that samples that resemble each other closely were difficult to differentiate, thus accounting for the greatest source of misclassification. This demonstrated that the KNN models were less accurate and did not perform well due to the close relationship between the species. Therefore, the KNN results in the current study indicates that game species, irrespective of the treatment (fresh or previously frozen) or the muscle type, can successfully be discriminated if the species are substantially different.

Table 4.4 The performance measures used to assess the KNN classification models of the three species, pre-processed with SNV + detrend.

Number of neighbours (k)	Species	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
2	Zebra	86.1	12.4	1.5	96.9	76.0	79.2	87.2	13.9
	Springbok	75.7	10.9	13.5	64.4	82.5	69.1	66.7	24.4
	Ostrich	83.7	4.3	12.0	47.9	94.4	71.9	57.5	16.4
3	Zebra	89.7	5.4	4.9	91.6	86.9	90.8	91.2	10.3
	Springbok	80.2	11.5	8.4	74.3	83.0	67.9	71.0	19.8
	Ostrich	83.5	6.4	10.1	45.0	92.1	56.3	50.0	16.5
5	Zebra	86.7	5.7	7.6	87.1	86.2	90.0	88.5	13.3
	Springbok	80.3	9.2	10.5	71.4	85.4	74.1	72.7	19.8
	Ostrich	87.1	8.1	4.8	60.0	90.8	46.9	52.6	12.9

(KNN) K-nearest neighbours; (SNV) Standard normal variate.

4.1.2.2.3 Discriminant analysis (DA)

Table 4.5 DA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for species discrimination.

Number of principal components (PCs)	Method (distance calculations)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
5	Linear	Calibration	86.7	13.3
		Validation	80.8	19.2
	Quadratic	Calibration	88.8	11.2
		Validation	86.7	13.3
	Mahalanobis	Calibration	89.7	10.3
		Validation	86.7	13.3
6	Linear	Calibration	87.6	12.5
		Validation	80.0	20.0
	Quadratic	Calibration	88.8	11.2
		Validation	85.8	14.2
	Mahalanobis	Calibration	90.1	9.9
		Validation	87.5	12.5

(DA) Discriminant analysis; (SNV) Standard normal variate.

The overall accuracy of the linear [5 PCs (86.7%); 6 PCs (87.6%)], quadratic [5 PCs and 6 PCs (88.8%) and mahalanobis [5 PCs (89.7%) and 6 PCs (90.1%)] models, suggests that a model with 6 PCs would provide better discrimination. These results can be explained by investigating the performance measures (**Table 4.6** and **Addendum A, Table A1 – A2**) of the different discrimination models.

The mahalanobis discriminant analysis models presented the best discrimination results. The LDA [mahalanobis distance] (5 PCs) calibration model achieved an overall classification accuracy of 89.7% and a misclassification rate of 10.3% (**Table 4.5**). This suggests that the model can accurately distinguish between the three species. The model also revealed that the discrimination of ostrich was nearly all correctly predicted. The classification accuracy (98.1%), sensitivity (93.8%) and specificity (98.9%) confirmed this (**Table 4.6**).

The LDA [mahalanobis distance] model with 6 PCs achieved better results than that of the model with 5 PCs. The classification accuracy was improved to 90.1%, consequently lowering the misclassification rate to 9.9%. These improved model results were also observed in the improvement of the sensitivity (>83%) and the specificity (>90%) (**Table 4.6**), therefore confirming the model's ability to accurately differentiate between the three species. From both the LDA [mahalanobis distance] (5 PCs and 6 PCs) results it can be said that the ostrich can, with a high level of accuracy, be distinguished from the other two species with a misclassification rate below 3% (**Table 4.6**). Although the model exhibited an excellent overall calibration

accuracy as well as good performance measures for each treatment (**Table 4.6**), the decreased validation accuracy suggests that the classification model was over-fitted (Miller, 2005). The disadvantage of a model that has been over-fitted is that it is very specific to the exact data used to build the model and as a result more sensitive to the deviating conditions (Miller, 2005). Consequently, the over-fitted model has a low probability of correctly predicting the species classes when using an independent validation set.

In general, it is observed that the LDA [mahalanobis distance] model achieved the best classification results. The LDA [mahalanobis distance] model achieved better results because the algorithm was able to calculate an optimal linear projection between the classes. The cluster overlap observed between zebra and springbok in the PCA score plot (**Figure 4.16**) explains why the mahalanobis distance algorithm was less successful to calculate an optimal linear projection between these species. The lack of separation between the clusters can be attributed to their similar physiochemical characteristics (e.g. fat content and pH), as previously discussed in **Section 4.1.2.1**.

These overall model results (90.1%) are comparable to that reported by McElhinney *et al.* (1999). These researchers were able to classify beef, lamb, pork, chicken and turkey with a 91.3% accuracy with no pre-processing and a 97.4% accuracy with Savitzky-Golay 2nd derivative pre-processing, concluding that NIR spectroscopy in conjunction with DA is a suitable method for this application. In another study, Ding and Xu (1999) reported that with SNV + detrend pre-processing they could differentiate between beef and kangaroo meat with a 100% accuracy. These authors also reported that the spectral separation between species may be due to their chemical and physical differences (Ding & Xu, 1999). The results from previous studies also suggest that the species classification can be improved by using different types of pre-processing techniques. The results in the current study therefore show that DA can be used to discriminate between species, irrespective of the treatment (fresh or previously frozen) or the muscle type.

Table 4.6 The performance measures used to assess the LDA [mahalanobis distance] models of the three species, pre-processed with SNV + detrend.

Number of principal components (PCs)	Species	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
5	Zebra	91.3	6.6	2.2	95.5	87.4	87.5	91.3	8.7
	Springbok	89.7	3.0	7.3	81.3	95.1	91.4	86.0	10.3
	Ostrich	98.1	0.9	0.9	93.8	98.9	93.8	93.8	1.9
6	Zebra	92.1	4.8	3.1	94.0	90.2	90.8	92.4	7.9
	Springbok	90.1	3.7	6.0	83.7	93.9	88.9	86.2	9.9
	Ostrich	97.7	1.4	0.9	93.6	98.4	90.6	92.1	2.3

(LDA) Linear discriminant analysis; (SNV) Standard normal variate.

4.1.2.2.4 Partial least squares discriminant analysis (PLS-DA)

The PLS-DA model, consisting of zebra, springbok and ostrich, gave satisfactory discrimination results. The PLS-DA calibration model achieved an overall classification accuracy of 95.7% and a misclassification rate of 4.3% (**Table 4.7**). The PLS-DA score plots for the three species are given in **Figure 4.17a – b**. The 3D scores plot [LV1 (84.94%) vs. LV2 (6.06%) vs. LV3 (1.32%)] demonstrates a minimal overlap between the classes that is indicative of satisfactory model calibration. The score plot of LV2 (6.06%) vs. LV3 (1.32%) (**Figure 4.17b**) exhibited the best species separation, with three clusters in the score space. The springbok and zebra exhibited the best separation in the direction of LV2, while the separation between ostrich and the remaining two species was best described in the direction of LV3. The zebra was predominantly associated with the positive scores in both latent variables and the springbok with the negative scores in LV2 and positive scores in LV3. The ostrich was predominantly associated with the negative scores in both latent variables. Therefore, the score plot (**Figure 4.17b**) exhibited that species separation is best described in the direction of both LV2 and LV3.

Table 4.7 PLS-DA model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for species discrimination.

Number of latent variables (LVs)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
8	Calibration	95.7	4.3
	Cross-validation	94.4	5.6
	Validation	88.3	11.7

(PLS-DA) Partial least squares discriminant analysis; (SNV) Standard normal variate.

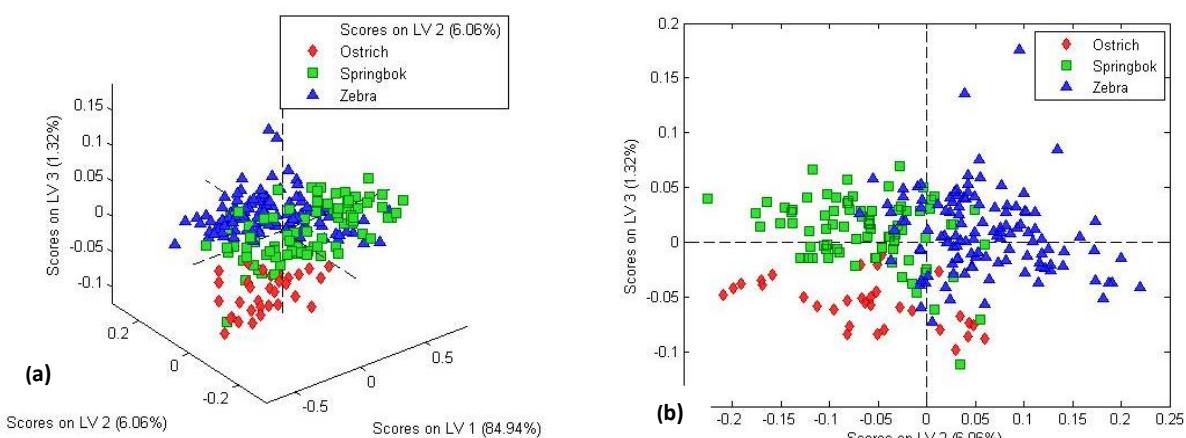


Figure 4.17 PLS-DA scores plot (SNV + detrend pre-processed) of ostrich (red), springbok (green) and zebra (blue). (a) 3D scores plot of LV1 (84.94%) vs. LV2 (6.06%) vs. LV3 (1.32%) and (b) score plot of LV2 (6.06%) vs. LV3 (1.32%), colour coded per class species.

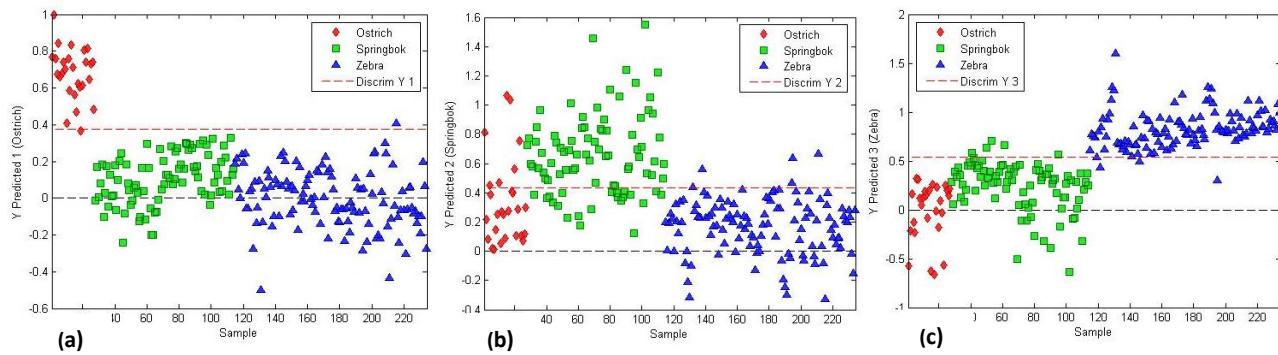


Figure 4.18 PLS-DA (8 LVs) (SNV + detrend pre-processed) resulting in an overall classification accuracy of 95.7%. PLS-DA prediction score plot of ostrich (red), springbok (green) and zebra (blue), illustrating the predicted objects. (a) Score plot of objects predicted as ostrich [above red line (Y1)] vs. remaining species [below red line (Y1)], (b) score plot of objects predicted as springbok [above red line (Y2)] vs. remaining species [below red line (Y2)] and (c) score plot of objects predicted as zebra [above red line (Y3)] vs. remaining species [below red line (Y3)].

The PLS-DA prediction score plots (**Figure 4.18a, c**) illustrated that the ostrich and zebra were nearly all correctly predicted. This was confirmed by the performance measures (**Table 4.8**). The zebra class resulted in the highest classification accuracy (98.2%), followed by ostrich (97.4%) and springbok (95.7%). The sensitivity, specificity and F1 score for zebra was above 98% (**Table 4.8**). The specificity and F1 score for ostrich were above 91%, and the sensitivity was 86%. These results confirms the model's ability to classify the zebra and ostrich species with little fault. Although springbok exhibited a classification accuracy, sensitivity, specificity and F1 score above 93%, indicative of an effective model, the classification accuracy decreased due to the increased false positive and false negative errors (**Figure 4.18b**). The spectral classification of zebra, ostrich and springbok can be attributed to their physiochemical differences as discussed in **Section 4.1.2.1**. Even though the PLS-DA model exhibited classification accuracies of 95.7% (calibration) and 94.4% (cross-validation), indicative of an effective model, the validation accuracy (88.3%) decreased. This indicated that the model was over-fitted (Miller, 2005) and therefore not as effective as suggested by the calibration- and cross-validation accuracies. The results in the current study suggest that PLS-DA can be used to discriminate between species. It was also evident that the model has a high probability of correctly predicting zebra and ostrich, with a high level of accuracy and small misclassification rate (**Table 4.8**).

The results reported herein were comparable to the PLS-DA results reported by McElhinney *et al.* (1999). The authors suggested that species discrimination was based on compositional chemical differences and primarily influenced by colour variations which are linked to the myoglobin concentration and state. Cozzolino and Murray (2004) reported similar results (96%) for the identification of beef, lamb, pork and chicken meat and a discrimination was made between the species based on intra-muscular fat, fatty acids and moisture. Lastly, in a study by Mamani-Linares *et al.* (2012), the researchers were able to differentiate between beef, llama and horse meat with accuracies between 89 and 100%. All the results from the

mentioned studies were comparable to that of the current study, thereby reinforcing that PLS-DA can be used to discriminate between zebra, springbok and ostrich species.

Table 4.8 The performance measures used to assess the PLS-DA models of the three species, pre-processed with SNV + detrend.

Number of latent variables (LVs)	Species	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
	Zebra	98.2	0.9	0.9	98.3	98.1	98.3	98.3	1.8
8	Springbok	95.7	3.0	1.3	96.1	95.5	91.4	93.7	4.3
	Ostrich	97.4	0.4	2.2	86.1	99.5	96.9	91.2	2.6

(PLS-DA) Partial least squares discriminant analysis; (SNV) Standard normal variate.

4.1.2.3 Optimal model selection

After the models [pre-processed with SNV + detrend (4th order polynomial)] were calculated to distinguish between the three different game species (zebra, springbok and ostrich), regardless of the treatment (fresh or previously frozen) or muscle type, the models were compared to identify the one with the best results.

Table 4.9 reveals that the LDA [mahalanobis distance] and PLS-DA models produced the best results.

Table 4.9 An overview of the accuracies for the various classification and discrimination models, pre-processed with SNV + detrend to distinguish between species.

Model	Number of nearest neighbour (k), PCs or LVs	Classification accuracy (%)	Misclassification rate (%)
SIMCA	-	68.5	31.5
KNN	2	74.7	25.3
KNN	3	78.1	21.9
KNN	5	78.5	21.5
LDA	5	86.7	13.3
QDA	5	88.8	11.2
MDA	5	89.7	10.3
LDA	6	87.6	12.5
QDA	6	88.8	11.2
MDA	6	90.1	9.9
PLS-DA	8	95.7	4.3

(SNV) Standard normal variate; (PCs) Principal components; (LVs) Latent variables; (SIMCA) Soft independent modelling of class analogy; (KNN) K-nearest neighbour; (LDA) Linear discriminant analysis; (QDA) Quadratic discriminant analysis; (MDA) Mahalanobis discriminant analysis; (PLS-DA) Partial least squares discriminant analysis.

These results raised the important question of whether or not the classification accuracy would improve if data were subjected to different pre-processing techniques. Hence, the data was subjected to various pre-processing techniques [SNV; SNV + Savitzky-Golay (2nd derivative, 2nd order polynomial, 7/9 points) [SGd₂(7/9)] , SNV + detrend + Savitzky-Golay (2nd derivative, 2nd order polynomial, 7/9 points) [SGd₂(7/9)], Savitzky-Golay (1st derivative, 2nd order polynomial, 5/7 points) [SGd₁(5/7)] and Savitzky-Golay (2nd derivative, 2nd order polynomial, 7/9 points) [SGd₂(7/9)]] in different combinations. After PCA models were calculated, the data explored and outliers removed the MDA- and PLS-DA models were recalculated and evaluated.

4.1.2.4 Exploratory data analysis (EDA)

4.1.2.4.1 Principal component analysis (PCA)

The score plots of PC2 vs. PC3 were used to investigate the difference between the species, as the data shows that the variation, seen as the separation, was explained in the second component. Although the data variance explained by PC2 improved when using the various pre-processing techniques, minimal class separation between the species was still observed in the results of the PCA score plots (**Addendum A, Figure A3 – A11**). The variance observed in PC2 can be attributed to the variances between the three classes due to differences in macronutrient composition of the biological samples. This separation was predominantly due to differences in moisture, fat, protein and pH as illustrated in the loading plots (**Addendum A, Figure A3b – A11b**).

The PC2 score plot of the SNV corrected data accounted for 10% of the data variance (**Addendum A, Figure A3a**) and the loadings plot (**Addendum A, Figure A3b**) exhibited an interpretable positive band at 957 nm and a negative band at 1106 nm. These are associated with pH (957 nm) (Barbin *et al.*, 2012b), moisture and meat tenderness (1106 nm) (ElMasry *et al.*, 2012). Although SNV removed the scatter effects, by centering and scaling each spectrum, the separation was insufficient as an overlap between species was observed (**Addendum A, Figure A3a**).

The score plots (PC2) of the SNV+SGd₂(7) and SNV+SGd₂(9) corrected data accounted for 11% and 9% of the data variance, respectively (**Addendum A, Figure A4a – A5a**). The loading plots (**Addendum A, Figure A4b – A5b**) had interpretable bands at 1131 nm (negative), 1205 nm (positive), 1242 nm (negative) and 1298 nm (negative), thus suggesting that the separation between the three species is a result of variance in the fat and moisture contents (Kamruzzaman *et al.*, 2012b), pH (ElMasry *et al.*, 2012) as well as protein content (Kamruzzaman *et al.*, 2012b). Once again, the pre-processing improved the explained variance between the species, and though the separation in PC2 was more prominent, an overlap between the species was still evident (**Addendum A, Figure A4a – A5a**). The improved species separation can be ascribed to the combined pre-processing of SNV + Savitzky-Golay second derivative. The second derivative is a functional pre-processing technique and is used to amplify wavebands, in order to identify bands of importance that best describe the separation in the data (Esbensen *et al.*, 2002).

The PC2 loadings plot of the SNV+detrend+SGd₂(7) and SNV+detrend+SGd₂(9) corrected data illustrated a minimal class separation between the species (**Addendum A, Figure A6b – A7b**). These plots had an interpretable positive band at 1205 nm and interpretable negative bands at 1131, 1242 and 1298 nm. These results suggest that the separation between the three species is a result of variance in the fat and moisture content (Kamruzzaman *et al.*, 2012b), pH (ElMasry *et al.*, 2012) as well as protein content (Kamruzzaman *et al.*, 2012b). Although the explained variance of PC 2 improved, and the separation between species was better, it was still insufficient as a species overlap was observed in the score plots (**Addendum A, Figure A6a – A7a**). The combined pre-processing of SNV + detrend + Savitzky-Golay second derivative exhibited similar results as SNV + Savitzky-Golay second derivative, thus indicating that detrend pre-processing did not improve the data for species separation.

The PCA results of the SGd₁(5) and SGd₁(7) corrected data are shown in **Addendum A, Figure A8 – A9**. These results suggested that Savitzky-Golay first derivative would be an effective pre-processing technique to apply for species separation. The Savitzky-Golay first derivative is a functional pre-processing technique as it eliminates noise and corrects for the baseline shift (Esbensen *et al.*, 2002). The explained variance for PC2 in both score plots improved and accounted for 24% of the data variance. The loading plot for both pre-processing techniques showed interpretable bands at 1031 nm (negative) and 1143 nm (positive) (**Addendum A, Figure A8b – A9b**). The band at 1031 nm (N-H stretch second overtone) is associated with protein (Osborne *et al.*, 1993) as well as meat tenderness (ElMasry *et al.*, 2012). The band at 1143 nm is associated with the C-H stretch second overtone found in aromatic groups (Osborne *et al.*, 1993) and corresponds to the tenderness of meat (Kamruzzaman *et al.*, 2013).

The PC2 score plots of the SGd₂(7) and SGd₂(9) corrected data accounted for 35% and 36% of the data variance, respectively (**Addendum A, Figure A10a – A11a**), hence exhibiting the best PCA results for species separation. The loading line plots (**Addendum A, Figure A10b – A11b**) had interpretable positive bands at 1069, 1118 and 1372 nm and negative interpretable bands at 970, 1155 and 1329 nm. As with the previous PCA results, the separation between the three species is due to differences in moisture (Osborne *et al.*, 1993; ElMasry *et al.*, 2013), protein (Osborne *et al.*, 1993; Barbin *et al.*, 2013a) and pH (ElMasry *et al.*, 2012).

The results from the various pre-processing techniques suggests that the Savitzky-Golay first- and second derivative transformations would be sufficient and best suited for species separation. The Savitzky-Golay first- and second derivative transformations enhanced the differences in macronutrient composition of the biological samples. This separation was predominantly due to differences in moisture, protein and pH as illustrated in the loading plots (**Addendum A, Figure A3b – A11b**). Even though Savitzky-Golay derivatives showed the best PCA results for separation, all the pre-processing techniques were used for further data analysis and model evaluation.

4.1.2.5 Multivariate data analysis (MDA): Optimal model development

4.1.2.5.1 Linear discriminant analysis (LDA) [mahalanobis distance]

The overall accuracy of the mahalanobis discriminant models suggested that a model with 6 PCs would provide better discrimination. The LDA [mahalanobis distance] (6 PCs) models for the various types of pre-processing gave similar results. There are no substantial differences between the pre-processing techniques (**Table 4.10**). The LDA [mahalanobis distance] model, pre-processed with SNV+detrend+SGd₂(9) and SGd₁(7), presented the best discrimination results. These results can be explained by investigating the performance measures (**Table 4.11 – 4.13**).

Table 4.10 An overview of the accuracies of the LDA [mahalanobis distance] models with various pre-processing techniques applied for species discrimination.

Pre-processing	Number of PCs	Classification accuracy (%)	Misclassification rate (%)	Number of PCs	Classification accuracy (%)	Misclassification rate (%)
SNV		77.4	22.7		90.6	9.4
SNV+SGd ₂ (7)		79.9	20.1		85.5	14.5
SNV+SGd ₂ (9)		79.5	20.5		84.2	15.8
SNV+detrend+SGd ₂ (7)		79.0	21.0		90.1	9.9
SNV+detrend+SGd ₂ (9)	5	80.3	19.7	6	91.9	8.2
SGd ₁ (5)		81.7	18.3		83.8	16.2
SGd ₁ (7)		84.3	15.7		91.5	8.5
SGd ₂ (7)		83.0	17.0		88.9	11.1
SGd ₂ (9)		82.1	17.9		84.3	15.7

(LDA) Linear discriminant analysis [mahalanobis distance]; (PCs) Principal components; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

The LDA [mahalanobis distance] (6 PCs) calibration model achieved an overall classification accuracy of 91.9% for SNV+detrend+SGd₂(9) and 91.5% for SGd₁(7) with misclassification rates <9% (**Table 4.10**). Although both models had a calibration accuracy of 91%, further investigation shows that the model with an 89.8% validation accuracy is optimal (**Table 4.12**). Therefore, the SGd₁(7) pre-processed data resulted in a more robust model and is less affected by variation of intrinsic parameters in meat. This suggests that the model can accurately distinguish between the three species, and is most likely ascribed to the fact that Savitzky-Golay first derivative eliminates noise and corrects for the baseline shift (Esbensen *et al.*, 2002). The derivatives were also calculated based on a smoothing method by removing unwanted features from the data, and consequently enhancing features that are advantageous to the primary analysis. The model also revealed that the discrimination of ostrich was nearly perfect (97.7%). The classification accuracy (97.7%), sensitivity (96.6%) and specificity (99.5%) confirmed this (**Table 4.13**). The improved classification of ostrich

can be further explained by referring to the PCA scores- and loadings plot (**Addendum A, Figure A9a - b**). The scores illustrate a separation of ostrich from the other species in the positive direction of PC2. This positive separation can be attributed to the positive absorption band at 1143 nm (**Addendum A, Figure A9b**). A study by Kamruzzaman *et al.* (2013) found that this absorption band relates to meat tenderness. Meat tenderness has previously been quantified for these specific species and the authors have concluded that springbok (Hoffman *et al.*, 2007a) is the most tender of the species in question followed by ostrich (Leygonie *et al.*, 2012b). Therefore, these studies can explain the variation in the data (**Addendum A, Figure A9a**). The scores plot corroborate the literature as the clustering of the classes mirrored the tenderness reported.

The LDA [mahalanobis distance] model, pre-processed with SGd₁(7), achieved slightly better results than that of the original model pre-processed with SNV + detrend. To illustrate this, the performance measures of SGd₁(7) (**Table 4.13**) and SNV + detrend (**Table 4.6**) were shown. The classification accuracy for differentiating between the species was 90.1% for SNV + detrend and 91.5% for SGd₁(7). The reason for the slight differences between these two pre-processing techniques is most likely due to the fact that both algorithms essentially do the same thing. SNV removes the scatter effects (noise) by centering and scaling each spectrum (Barnes *et al.*, 1989) and detrend reduces the baseline shift and curvature in the spectroscopic data (Barnes *et al.*, 1989). Savitzky-Golay first derivative does exactly the same as SNV + detrend, however the derivatives are calculated with an additional smoothing effect which removes the unwanted features and as a result enhances the features that are advantageous for the analysis (Savitzky & Golay, 1964). Therefore, the SGd₁(7) pre-processing achieved better results as the separation due to differences in protein content and especially meat tenderness was enhanced.

Table 4.11 LDA [mahalanobis distance] model calibration and validation results to assess the overall performance of the SNV+detrend+SGd₂(9) corrected data for species discrimination.

Number of principal components (PCs)	Model [SNV+detrend+SGd ₂ (9)]	Classification accuracy (%)	Misclassification rate (%)
6	Calibration	91.9	8.2
	Validation	85.8	14.2

(LDA) Linear discriminant analysis [mahalanobis distance]; (SNV) Standard normal variate; [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points).

Table 4.12 LDA [mahalanobis distance] model calibration and validation results to assess the overall performance of the SGd₁(7) corrected data for species discrimination.

Number of principal components (PCs)	Model [SGd ₁ (7)]	Classification accuracy (%)	Misclassification rate (%)
6	Calibration	91.5	8.5
	Validation	89.8	10.2

(LDA) Linear discriminant analysis [mahalanobis distance]; [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

Table 4.13 The performance measures used to assess the LDA [mahalanobis distance] models of the three species, pre-processed with SGd₁(7).

Number of principal components (PCs)	Species	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
6	Zebra	93.1	2.6	4.3	94.6	94.8	91.3	92.9	6.9
	Springbok	91.9	5.6	2.6	86.3	91.1	93.2	89.6	8.1
	Ostrich	97.7	0.5	1.8	96.6	99.5	87.5	91.8	2.3

(LDA) Linear discriminant analysis [mahalanobis distance]; [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

4.1.2.5.2 Partial least squares discriminant analysis (PLS-DA)

The PLS-DA models for the various types of pre-processing gave similar results. The overall accuracy of the PLS-DA models, suggested that the models pre-processed with SNV+SGd₂(7), SNV+SGd₂(9), SGd₁(7), SGd₂(7) and SGd₂(9) would provide the best discrimination. There are no substantial differences between these mentioned pre-processing techniques (**Table 4.14**).

Table 4.14 An overview of the accuracies of the PLS-DA models with various pre-processing techniques applied for species discrimination.

Pre-processing	Calibration			CV	Validation
	Number of LVs	Classification accuracy (%)	Misclassification rate (%)		
SNV	4	84.2	15.8	80.8	86.4
SNV+SGd ₂ (7)	5	94.9	5.1	93.6	89.0
SNV+SGd ₂ (9)	5	93.2	6.8	89.7	88.1
SNV+detrend+SGd ₂ (7)	4	91.4	8.6	88.8	88.3
SNV+detrend+SGd ₂ (9)	3	87.6	12.5	86.3	84.2
SGd ₁ (5)	5	88.5	11.5	86.4	89.0
SGd ₁ (7)	7	93.2	6.8	88.9	94.1
SGd ₂ (7)	6	94.0	6.0	90.2	89.0
SGd ₂ (9)	6	94.5	5.5	93.2	60.2

(PLS-DA) Partial least squares discriminant analysis; (LVs) Latent variables; (CV) Cross-validation; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

The PLS-DA calibration models achieved an overall classification accuracy of 94.9% for SNV+SGd₂(7), 93.2% for SNV+SGd₂(9), 93.2% for SGd₁(7), 94.0% for SGd₂(7) and 94.5% for SGd₂(9) with misclassification

rates <7% (**Table 4.14**). Although all five models exhibited calibration accuracies between 93.2% and 94.9%, further investigation showed that the model with a 94.1% validation accuracy was optimal (**Table 4.14**). Therefore, the SGd₁(7) (93.2%) pre-processed data resulted in a more robust model and is less affected by variation of intrinsic parameters in meat. This suggests that the model can accurately distinguish between the three species. These results can be explained by investigating the performance measures (**Table 4.15**).

The reason why SGd₁(7) achieved the best results was discussed in **Section 4.1.2.5**. The model also revealed that the ostrich class resulted in the highest classification accuracy (98.2%), followed by zebra (94.4%) and springbok (93.6%) (**Table 4.15**). Therefore, the probability of accurately predicting ostrich was higher as confirmed by the sensitivity (96.6%), specificity (99.5%) and F1 score (93.3%) in **Table 4.15**. The improved classification of ostrich was discussed in **Section 4.1.2.5**.

The PLS-DA model, pre-processed with SGd₁(7), resulted in a slightly lower classification accuracy compared to the original model pre-processed with SNV + detrend (95.7%). Although the SNV + detrend model had a higher accuracy and better individual species discrimination performance measures (**Table 4.15**), the CV (94.4%) and validation (88.3%) results suggest that the model has a good predictive power for the training set, but the independent validation set was under-fitted (Miller, 2005). This can also be indicative of a calibration model that has been over-fitted (Miller, 2005). Therefore, the PLS-DA model, pre-processed with SGd₁(7), is considered to be better than the original SNV + detrend model. The SGd₁(7) model achieved lower classification (93.2%) and CV (88.9%) results, but the independent validation set achieved a higher accuracy (94.1%). This indicates that the model is more robust and not over-fitted to the training set (Miller, 2005). The reason for the slight differences between these two pre-processing techniques was discussed in **Section 4.1.2.5**. Therefore, the SGd₁(7) pre-processing achieved better results as the separation due to differences in protein content and meat tenderness were enhanced.

Table 4.15 The performance measures used to assess the PLS-DA models of the three species, pre-processed with SGd₁(7).

Number of latent variables (LVs)	Species	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
7	Zebra	94.4	2.2	3.5	92.8	93.6	95.4	94.1	5.6
	Springbok	93.6	3.4	3.0	92.6	94.9	91.7	92.2	6.4
	Ostrich	98.2	1.4	0.5	96.6	99.5	90.3	93.3	1.8

(PLS-DA) Partial least squares discriminant analysis; [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

4.1.2.6 Optimal model selection

After the models for the various types of pre-processing were calculated, the models with the best results were compared to identify the optimal model (**Table 4.16**). It was concluded that the PLS-DA model, pre-processed with Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points) was the best for differentiating between zebra, springbok and ostrich, irrespective of the treatment (fresh or previously frozen) or the muscle type. The PLS-DA models achieved better results as it generally outperforms LDA, when the classes are closely related, because it overcomes the collinearity problems often associated with LDA. SGd₁(7) outperformed the combined pre-processing of SNV + detrend, as the Savitzky-Golay transformation was found to have enhanced the differences in protein content and tenderness, which was predominantly the contributors for species separation. Ding and Xu (1999) also reported that the use of derivatives improved the classification accuracy of their models, therefore making these results comparable. Lastly, these results illustrate that samples that resemble each other closely were difficult to differentiate, thus accounting for the greatest source of misclassification. This suggests that greater and enhanced spectral differences between species, would result in improved classification accuracies and better prediction models.

Table 4.16 An overview of the accuracies of the LDA [mahalanobis distance] and PLS-DA models with various pre-processing techniques applied to distinguish between species.

Model	Pre-processing	Number of PCs or LVs	Calibration		Validation
			Classification accuracy (%)	Misclassification rate (%)	Classification accuracy (%)
LDA	SNV + detrend	6	90.1	9.9	87.5
PLS-DA	SNV + detrend	8	95.7	4.3	88.3
LDA	SGd ₁ (7)	6	91.5	8.5	89.8
PLS-DA	SGd ₁ (7)	8	93.2	6.8	94.1

(LDA) Linear discriminant analysis [mahalanobis distance]; (PLS-DA) Partial least squares discriminant analysis; (PCs) Principal components; (LVs) Latent variables; (SNV) Standard normal variate; [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

4.1.2.7 Conclusion

NIR spectroscopy combined with MDA could accurately distinguish between the three species, irrespective of the treatment (fresh or previously frozen) or the muscle type. The PLS-DA discrimination model, data pre-processed with Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points), yielded the best results and could effectively distinguish the zebra, springbok and ostrich from one another with a 93.2% accuracy. Throughout the study, the samples that resembled each other closely were difficult to differentiate. This is attributed to the samples' spectral similarities, thus accounting for the greatest source of misclassification. It can thus be deduced that greater and enhanced spectral differences between species, would improve the classification accuracies and better the prediction models. In addition, it was found that the treatment (fresh or previously frozen) and muscle type does not influence the accuracy of the model for species discrimination.

4.1.3 Species determination irrespective of frozen-period

This data set consisted of four species [black wildebeest (*Connochaetes gnou*), zebra (*Equus quagga burchelli*), springbok (*Antidorcas marsupialis*), ostrich (*Struthio camelus*)], one muscle type for each species and the meat samples were frozen for a period of nine months. The aim of **Section 4.1.3** was to differentiate between the four game species, irrespective of the treatment (fresh or previously frozen) or the frozen period (1 – 9 months), as well as to determine whether the treatments (fresh or previously frozen) and different frozen periods had an effect on the species classification accuracies.

4.1.3.1 Exploratory data analysis (EDA)

4.1.3.1.1 Principal component analysis (PCA)

Minimal class separation between the species was observed in the results of the PCA score plots (**Figure 4.19 – 4.21**) of the SNV + detrend corrected data. PC1 accounted for 86% of the variance (**Figure 4.19a**), and PC2 6% (**Figure 4.20a**), whereas PC3 (**Figure 4.21a**), PC4 and PC5 (**Figure 4.21b**) only accounted for approximately 4%, 3% and 1% of the variance, respectively. This illustrates that the variation, seen as the separation, was explained in the first two components.

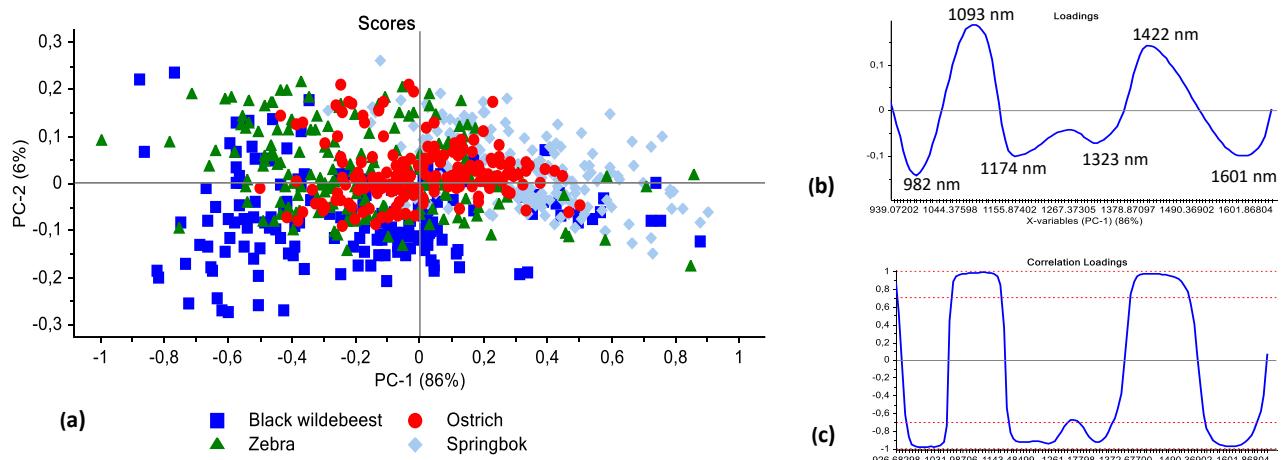


Figure 4.19 PCA analysis (SNV + detrend pre-processed) of species for black wildebeest (**blue**), zebra (**green**), springbok (**light blue**) and ostrich (**red**) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC1 (86%) vs. PC2 (6%). (b) PCA loadings line plot and (c) correlation loadings for PC1 with interpretable bands at 982, 1093, 1174, 1323, 1422 and 1601 nm.

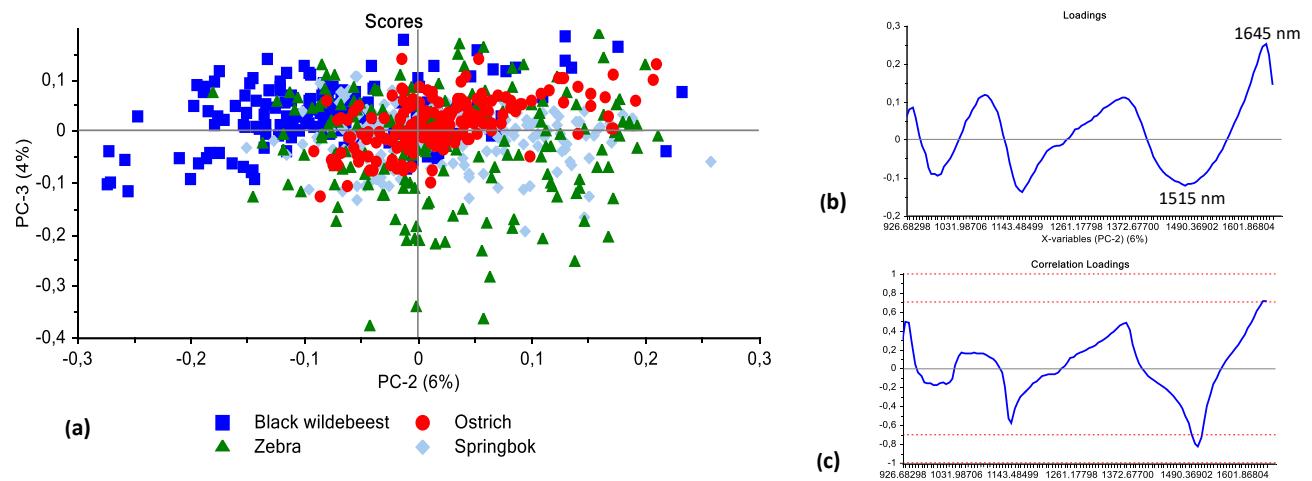


Figure 4.20 PCA analysis (SNV + detrend pre-processed) of species black wildebeest (**blue**), zebra (**green**), springbok (**light blue**) and ostrich (**red**) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (6%) vs. PC3 (4%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 1515 and 1645 nm.

The variance observed in the direction of PC1 may be attributed to the intra-class separation, while PC2 can be attributed to the variances between the four classes due to differences in macronutrient composition. This separation was predominantly due to differences in moisture, fat, protein and pH as illustrated in the loadings plot (**Figure 4.19b**, **4.20b**, **4.22a – b**). The loadings line plot of PC1 (**Figure 4.19b**) exhibited interpretable positive bands at 1093 and 1422 nm and negative bands at 982, 1174, 1323 and 1601 nm. When investigating the positive and negative loading bands of PC1, in combination with the score plot of PC1 (86%) vs. PC2 (6%) (**Figure 4.19a**), it is clear that the intra-class species separation is mainly based on

the positive spectral bands at 1093 nm, associated with pH (ElMasry *et al.*, 2012), and 1422 nm, which is related to moisture (Osborne *et al.*, 1993). The positive bands correspond with the positive side of the score plot (**Figure 4.19a**), thus illustrating that the intra-class separation is due to the difference between fresh and frozen-thawed meat samples. The PC2 loadings line plot (**Figure 4.20b**) exhibited interpretable bands at 1515 nm (negative) and 1645 nm (positive), thus suggesting that separation between the four species is a result of variance in the moisture-, fat- and protein content as well as the pH. The 1515 nm band (N-H stretch first overtone and O-H stretch first overtone) is associated with protein and moisture (Osborne *et al.*, 1993; Barbin *et al.*, 2013a), and 1645 nm (C-H stretch first overtone) is related to moisture, fat and protein as stated by Barbin *et al.* (2013a).

When investigating the loadings plot and comparing the suggested chemical results to literature, it was possible to conclude that protein- and fat content were the main factors contributing to the species separation. Black wildebeest had the highest reported protein content (*ca.* 24.3 g/100 g) (Hoffman *et al.*, 2009) followed by zebra (*ca.* 22.3 g/100 g) (Hoffman *et al.*, 2016), ostrich (*ca.* 20.9 g/100 g) (Sales, 1996) and springbok (*ca.* 18.8 g/100 g) (Hoffman & Wiklund, 2006; Hoffman *et al.*, 2007b). These results from literature were comparable to that illustrated in **Figure 4.20a**. The species clustered from highest- to lowest protein content in the direction of PC2. Furthermore, zebra and ostrich clustered in the centre of the PC score plot as they had similar protein contents. A similar phenomenon was observed for the fat contents, except that species with a lower fat content clustered to the left of the scores plot and those with a higher fat content clustered to the right. It was also observed that species with similar fat contents clustered together, resulting in an overlap. According to literature, the fat contents are as follows: *ca.* 0.5 g/100 g for ostrich (Sales, 1996), *ca.* 0.9 g/100 g for black wildebeest (Hoffman *et al.*, 2009), *ca.* 1.5 g/100 g for zebra (Hoffman *et al.*, 2016), and *ca.* 1.7 g/100 g for springbok (Hoffman & Wiklund, 2006; Hoffman *et al.*, 2007b).

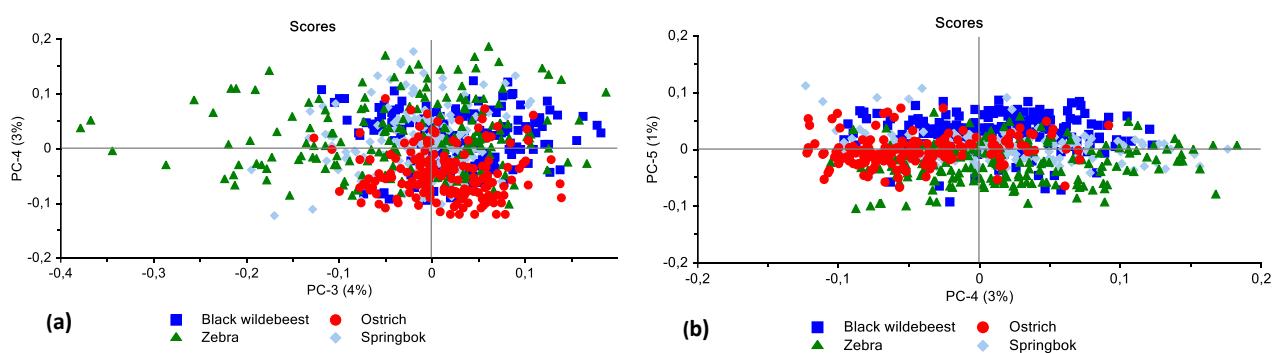


Figure 4.21 PCA analysis (SNV + detrend pre-processed) of species black wildebeest (blue), zebra (green), springbok (light blue) and ostrich (red) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC3 (4%) vs. PC4 (3%); and (b) PCA score plot of PC4 (3%) vs. PC5 (1%).

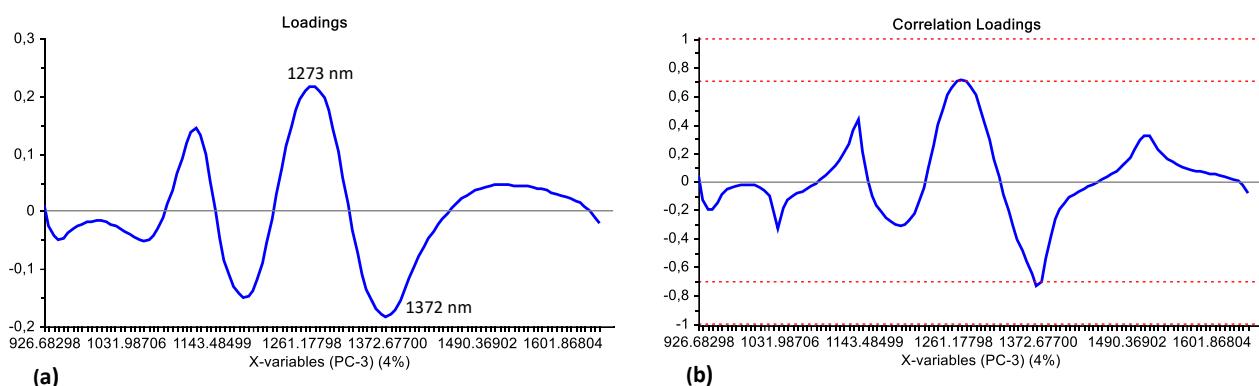


Figure 4.22 (a) PCA loadings line plot and (b) correlation loadings for PC3 (4%) with interpretable bands at 1273 and 1372 nm.

The score plots shown in **Figure 4.21a** (PC3 vs. PC4) and **Figure 4.21b** (PC4 vs. PC5) illustrated a major overlap of the four classes. Although the PCA results exhibited a lack of separation between species, the third PC, however, illustrated the intra-class variance for each species and this could be attributed to moisture, fat, protein and pH. Although PC3 accounted for minimal data variance, the loading line plot (**Figure 4.22a**) displayed interpretable absorption bands that explained the variation within each species. PC3 (4%) had a positive band at 1273 nm and a negative band at 1372 nm. These bands are associated with fat and protein (1273 nm) (Barbin *et al.*, 2013a) as well as moisture and pH (1372 nm) (ElMasry *et al.*, 2012; ElMasry *et al.*, 2013). This internal variance was likely due to animal-to-animal variance and the anatomical locations of different meat steaks from one muscle type (LTL and FF). However, the overall separation between the species was not affected by this internal variation.

The lack of separation between the different species indicates the similarity in their spectral signatures; however, there are numerous components (physical and chemical) that could differ and lead to slight spectral differences. The reasons for the differences are most likely attributed to the fluctuation of macronutrient composition. From the PCA, it was observed that the scores and loadings gave a global representation of the data. Hence the need for further exploratory data analysis to show the specific differences in the spectral data.

4.1.3.1.2 Difference spectra

The difference spectra of each species illustrated that the separation between the species was a result of numerous absorption bands (**Figure 4.23 – 4.26**). Twelve prominent and regular occurring absorption bands were exhibited at 939, 945, 976, 1062, 1118, 1149, 1162, 1341, 1366, 1403 and 1645 nm. Similar absorption bands were obtained for the difference spectra as previously discussed in **Section 4.1.2.1**. Although the results were similar, a few differences were observed with absorption bands at 945, 976, 1062, 1149 and 1341 nm. Barbin *et al.* (2013a) found that the band at 945 nm is related to the water content of pork as well as the pH and colour (Barbin *et al.*, 2012b). Therefore, the difference spectra (**Figure 4.24**) illustrates that zebra has a lower pH compared to that of springbok and ostrich as well as a higher pH compared to that of

black wildebeest, which corresponds to literature. Hoffman *et al.* (2016) reported a pH of *ca.* 5.7 for zebra which is slightly lower than that of springbok (*ca.* 5.8) (Hoffman *et al.*, 2007a) and ostrich (*ca.* 6.1) (Sales, 1996; Leygonie *et al.*, 2012b) but also higher than that of black wildebeest (*ca.* 5.6) (Hoffman *et al.*, 2009).

The band at 976 nm (**Figure 4.24** and **Figure 4.26**) relates to the second overtone of O-H stretching vibrations and was ascribed to the water content of ham (Talens *et al.*, 2013). The spectral differences between the four species can therefore be attributed to the higher moisture content of zebra (*ca.* 76.4 g/100 g) (Hoffman *et al.*, 2016) and ostrich (*ca.* 76.6 g/100 g) (Sales, 1996), in comparison to springbok (*ca.* 74.7 g/100 g) (Hoffman & Wiklund, 2006; Hoffman *et al.*, 2007b) and black wildebeest (*ca.* 74.3 g/100 g) (Hoffman *et al.*, 2009). The difference spectra of zebra (**Figure 4.24**) and ostrich (**Figure 4.26**) illustrated that there were minute differences between these two species due to the semi-straight line. This indicates that both these species have a similar moisture content and literature supports this. The band at 1149 nm can also be attributed to moisture (ElMasry *et al.*, 2013), with the difference spectra (**Figure 4.24** and **Figure 4.26**) exhibiting similar results for the four species.

The 1062 nm band represents the N-H stretch second overtone related to the RNH₂ group associated with the amino groups in proteins (Osborne *et al.*, 1993). The combination bands of C-H vibrations cause the separation observed at 1341 nm. Kamruzzaman *et al.* (2013) found that the absorption band at 1341 nm relates to the tenderness of lamb. Along with these findings, literature also states that this absorption band can be attributed to the pH (ElMasry *et al.*, 2012) and water holding capacity (ElMasry *et al.*, 2011) of beef. Previous studies have quantified meat tenderness for these specific species and concluded that black wildebeest (Hoffman *et al.*, 2009) is the most tender, followed by springbok (Hoffman *et al.*, 2007a), ostrich (Leygonie *et al.*, 2012b) and zebra. Therefore, these studies explain the variation observed in the spectral data (**Figure 4.24 – 4.26**). The difference spectra corroborate the literature, as the intensities of the bands mirrored the tenderness reported. The only exception was zebra, exhibiting a higher absorbance compared to ostrich. This was expected as both zebra and ostrich have similar shear force values that are indicative of toughness. It is thus plausible that the band at 1341 nm can be used to determine meat tenderness. The difference spectra could consequently be used to indicate the specific differences in the spectral data responsible for the separation between the four different game species.

The results from the difference spectra in the current study indicated that the physiochemical characteristics were similar to that of previous studies when using conventional analysis methods (Sales, 1996; Onyango *et al.*, 1998; Van Zyl & Ferreira, 2004; Hoffman & Wiklund, 2006; Hoffman *et al.*, 2007a; Hoffman *et al.*, 2007b; Hoffman *et al.*, 2009; Leygonie *et al.*, 2012b; Hoffman *et al.*, 2016). Therefore, these results indicate that the differences and separation between species can be explained by examining the physiochemical characteristics of the species e.g. moisture, fat, protein, pH and potentially tenderness. McElhinney *et al.* (1999), Ding and Xu (1999), Cozzolino and Murray (2004), Kamruzzaman *et al.* (2012a) and Mamani-Linares *et al.* (2012) reported that it was possible to differentiate between species based on their physiochemical characteristics. Therefore, indicating that black wildebeest, zebra, springbok and ostrich,

irrespective of the treatment (fresh or previously frozen) or the frozen period, has the potential to be successfully differentiated.

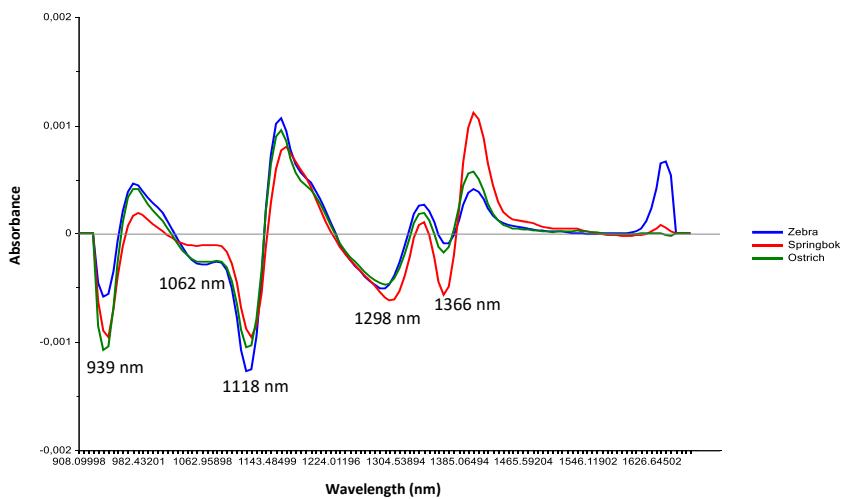


Figure 4.23 Difference spectra of black wildebeest and the remaining species, irrespective of frozen period, with absorption bands at 939, 1062, 1118, 1298, 1366 nm.

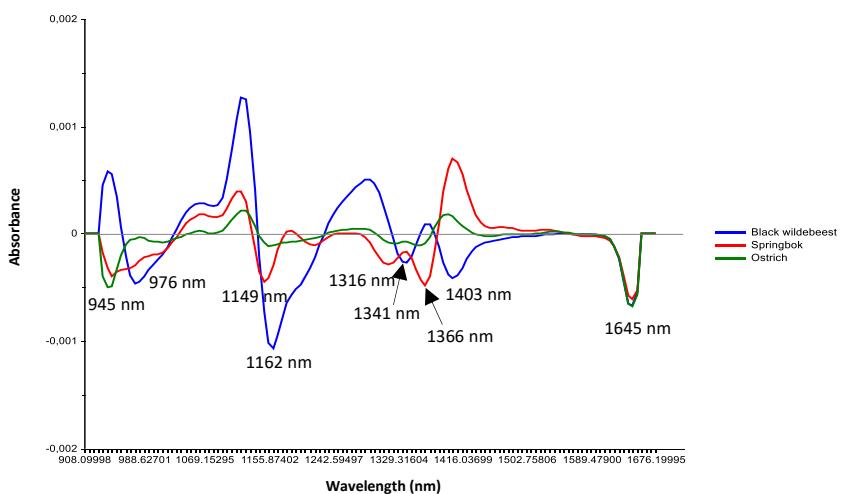


Figure 4.24 Difference spectra of zebra and the remaining species, irrespective of frozen period, with absorption bands at 945, 976, 1149, 1162, 1316, 1341, 1366, 1403, 1645 nm.

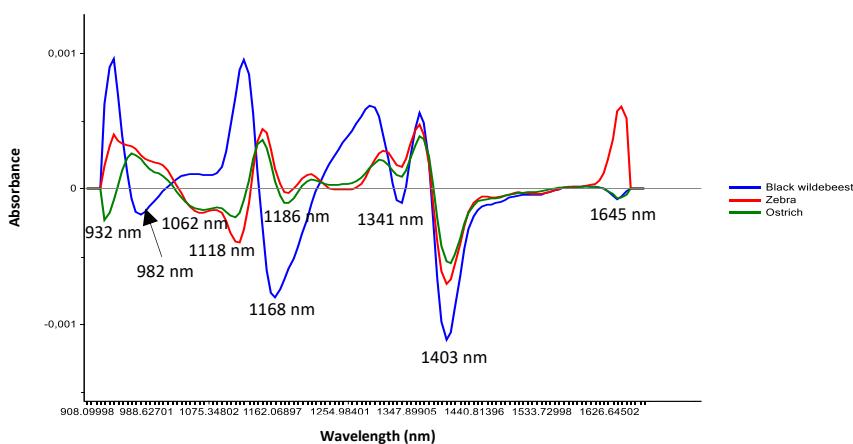


Figure 4.25 Difference spectra of springbok and the remaining species, irrespective of frozen period, with absorption bands at 932, 982, 1062, 1118, 1168, 1186, 1341, 1403, 1645 nm.

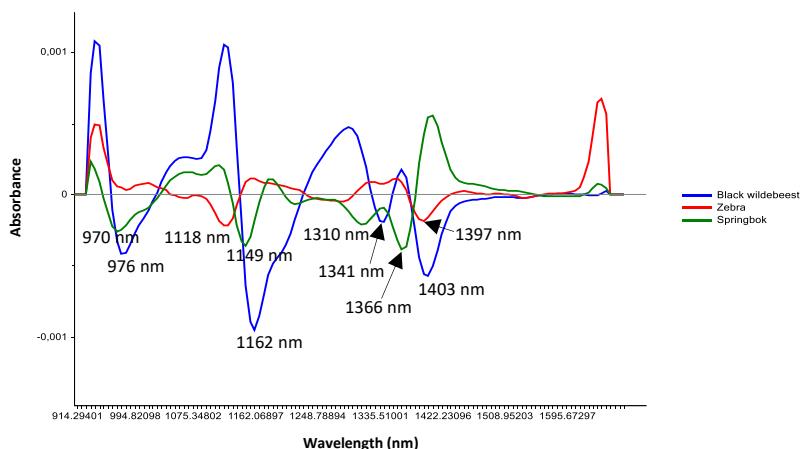


Figure 4.26 Difference spectra of ostrich and the remaining species, irrespective of frozen period, with absorption bands at 970, 976, 1118, 1149, 1162, 1310, 1341, 1366, 1397, 1403 nm.

4.1.3.2 Multivariate data analysis (MDA): Model development

4.1.3.2.1 Soft independent modelling of class analogy (SIMCA)

The SIMCA model, consisting of black wildebeest, zebra, springbok and ostrich, gave unsatisfactory classification results. The SIMCA calibration model achieved an overall classification accuracy of 45.4% and a misclassification rate of 54.6% (**Table 4.17**). The classification results for the validation model decreased, resulting in an overall classification accuracy of 34.3% and a misclassification rate of 65.8% (**Table 4.17**). The SIMCA models for both the calibration and validation resulted in a higher misclassification rate than a classification accuracy. This suggests that an accurate distinction cannot be made between the four species. These unsatisfactory classification results confirmed what was seen in the PCA score plot (**Figure 4.19a – 4.21a**). The lack of separation can be ascribed to the samples' spectral similarities, and as the SIMCA algorithm aims to classify samples based on spectral similarities, the classification accuracy was unsatisfactory.

Table 4.17 SIMCA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for species classification.

Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
Calibration	45.4	54.6
Validation	34.3	65.8

(SIMCA) Soft independent modelling od class analogy; (SNV) Standard normal variate.

The performance measures for individual SIMCA models are given in **Table 4.18** and the score plots were used to illustrate the limited classification abilities of the models (**Figure 4.27 – 4.28**). The SIMCA score plot (**Figure 4.27 – 4.28**) illustrates how well the model classified the individual samples. The black wildebeest class achieved the highest classification accuracy (65.5%), followed by springbok (62.1%), ostrich (61.9%) and zebra (60.4%) as shown in **Table 4.18**. The decreased overall classification accuracy is due to the increased misclassification of objects assigned to zebra, ostrich and springbok (**Figure 4.27b, 4.28a – b**). The PCA score plot (**Figure 4.20a**) and loadings line plot (**Figure 4.20b**) support these results as a slight separation was observed between black wildebeest and the other two species. The separation and correct classification of the black wildebeest was mainly attributed to the lower fat- and higher protein content, which accounted for the spectral differences. These properties for black wildebeest was considerably different, compared to the other species, and a correlation was observed in the difference spectra (**Figure 4.23 – 4.26**) as well as the loadings (**Figure 4.20b**), which could be compared to literature (Sales, 1996; Onyango *et al.*, 1998; Van Zyl & Ferreira, 2004; Hoffman & Wiklund, 2006; Hoffman *et al.*, 2007a; Hoffman *et al.*, 2007b; Hoffman *et al.*, 2009; Leygonie *et al.*, 2012b; Hoffman *et al.*, 2016).

The classification accuracy of springbok, ostrich and zebra was lower due to the increased false positive and false negative objects classified (**Figure 4.27 – 4.28**). The lack of separation can be ascribed to the samples' spectral similarities that corresponds to their similar physiochemical characteristics as reported in literature (Sales, 1996; Hoffman & Wiklund, 2006; Hoffman *et al.*, 2007a; Hoffman *et al.*, 2007b; Leygonie *et al.*, 2012b; Hoffman *et al.*, 2016). The sensitivity and specificity for the springbok was 62.4% and 61.1%, respectively. This suggests that the model is equally sensitive for predicting a true positive as correct (springbok), as it is specific when predicting a true negative as correct (black wildebeest, zebra and ostrich). The specificity of zebra and ostrich is 41.3% and 50.3%, respectively. In addition, the sensitivity is 78.2% and 65.5%, therefore confirming the model's ability to predict the zebra and ostrich accurately even though it struggles with the remaining species. These results suggest that spectra of zebra and ostrich are similar due to their similar physiochemical characteristics and for this reason, the SIMCA model achieved lower classification accuracies. Although the overall model is unsatisfactory, the results indicate that the black wildebeest is best distinguished from the other three species (**Table 4.18**).

McElhinney *et al.* (1999) reported similar results when discriminating between chicken, turkey, pork, beef and lamb meat. These researchers found that beef and lamb resemble each other closely as do chicken

and turkey thus their SIMCA models were less accurate and did not perform well due. Therefore, the SIMCA results in the current study indicates that these game species cannot be differentiated because of their spectral similarities.

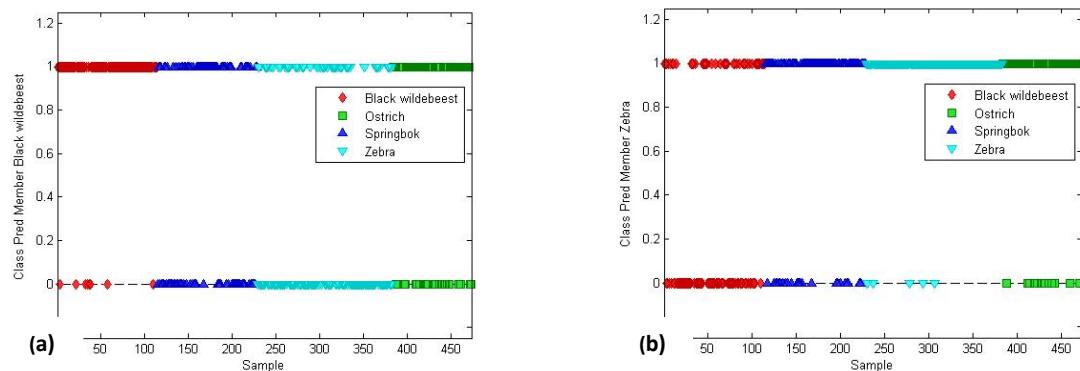


Figure 4.27 SIMCA classification (SNV + detrend pre-processed) resulting in a 65.5% and 60.4% classification accuracy for black wildebeest and zebra, respectively. SIMCA classification score plot of black wildebeest (red), ostrich (green), springbok (blue) and zebra (turquoise), illustrating the predicted objects. (a) Score plot for predicted black wildebeest (top) vs. remaining species (bottom) and (b) score plot of predicted zebra (top) vs. remaining species (bottom).

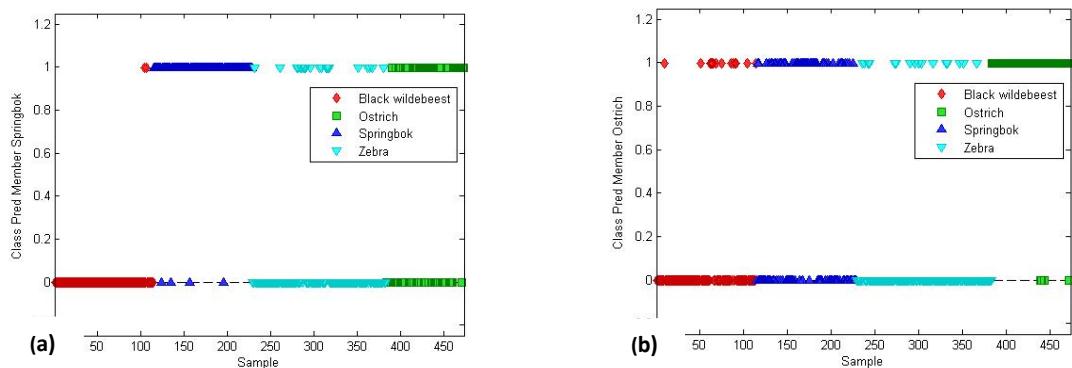


Figure 4.28 SIMCA classification (SNV + detrend pre-processed) resulting in a 62.1% and 61.9% classification accuracy for springbok and ostrich, respectively. SIMCA classification score plot of black wildebeest (red), ostrich (green), springbok (blue) and zebra (turquoise), illustrating the predicted objects. (a) Score plot for predicted springbok (top) vs. remaining species (bottom) and (b) score plot of predicted ostrich (top) vs. remaining species (bottom).

Table 4.18 The performance measures used to assess the SIMCA classification models of the four species, pre-processed with SNV + detrend.

Species	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
BWB	65.6	25.5	8.9	85.0	37.6	63.5	47.2	34.4
Zebra	60.4	28.4	11.2	78.2	41.3	64.1	50.3	39.7
Springbok	62.1	9.6	28.3	62.4	61.1	34.8	44.4	37.9
Ostrich	61.9	11.6	26.4	65.5	50.3	30.8	38.2	38.1

(SIMCA) Soft independent modelling od class analogy; (SNV) Standard normal variate; (BWB) Black wildebeest.

4.1.3.2.2 K-nearest neighbour (KNN)

The overall model accuracy of KNN(2) (68.7%), KNN(3) (73.6%) and KNN(5) (73.6%) suggested that three and five nearest neighbours would provide the best classification (**Table 4.19**). After the performance measures were examined individually, it was evident that the model with five nearest neighbours gave the best results (**Table 4.20**).

Table 4.19 KNN model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for species classification.

Number of neighbours (<i>k</i>)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
2	Calibration	68.7	31.3
	Cross-validation	73.6	26.4
	Validation	71.7	28.3
3	Calibration	73.6	26.4
	Cross-validation	74.8	25.2
	Validation	78.9	21.1
5	Calibration	73.6	26.4
	Cross-validation	75.3	24.7
	Validation	74.3	25.7

(KNN) *K*-nearest neighbours; (SNV) Standard normal variate.

The KNN(5) calibration model achieved an overall classification accuracy of 73.6%, with improved classification accuracies for the cross-validation- (75.3%) and validation model (74.3%) (**Table 4.19**). These improved accuracies are indicative of a classification model that is not over-fitted (Miller, 2005). Although this model had fairly high accuracies a few misclassified objects were observed amongst the four species (**Figure 4.29a – d**). This accounted for the misclassification rate of 26.4% for the overall model (**Table 4.19**). The sensitivity and specificity for the black wildebeest was 83.3% and 93.0%, respectively (**Table 4.20**). This indicates that the model has a high probability of correctly classifying the black wildebeest (**Figure 4.29a**), thus resulting in a classification accuracy of 90.9%. The lower sensitivity for the zebra (81.2%), springbok (78.1%) and ostrich (74.4%) reveals that the model was less suited for predicting these three species. This phenomenon can be explained by referring to the PCA score plot (**Figure 4.30**), as *k*-nearest neighbour was performed on the PC scores. The KS-calibration PCA score plot illustrates an overlap between the four classes, with a larger number of zebra samples displaying a close distance to the ostrich and springbok samples. Therefore, the zebra samples are assigned to the predominant class, ostrich and/or springbok. Hence, explaining why the model is less suited for predicting the zebra samples.

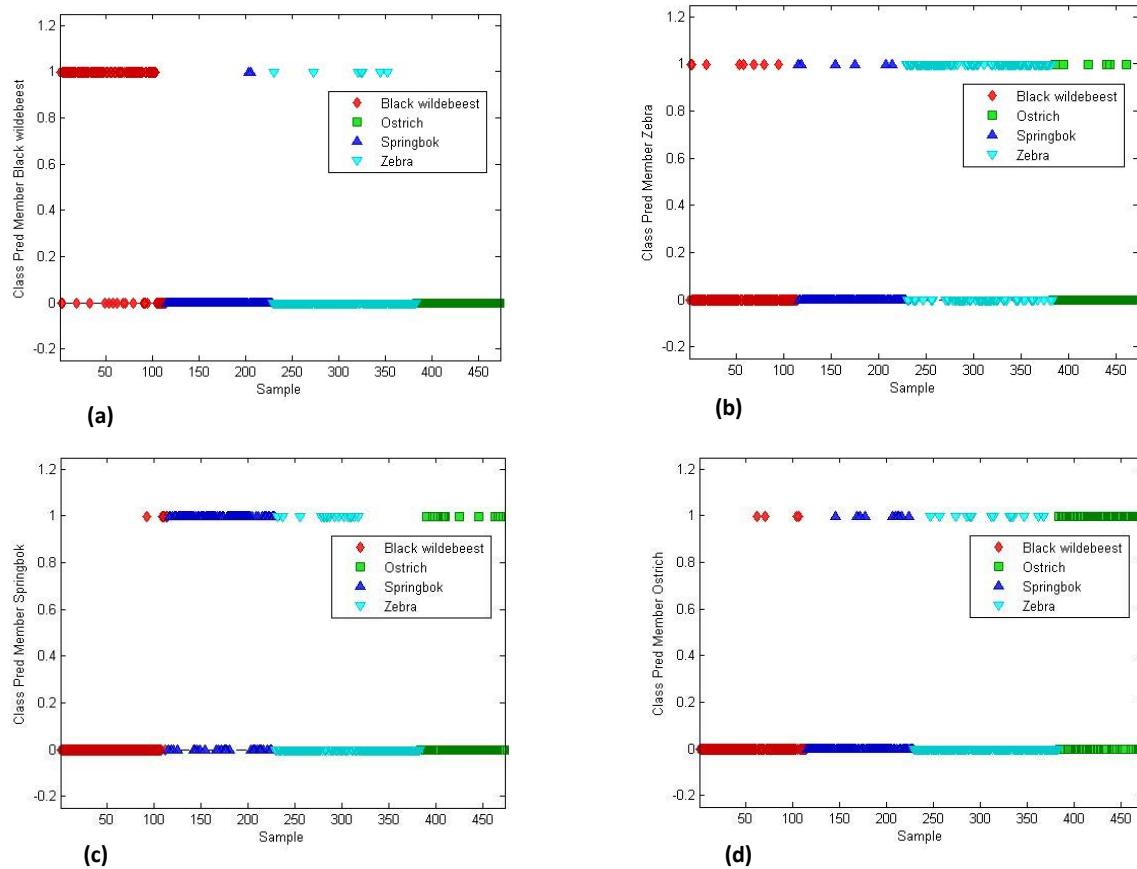


Figure 4.29 KNN ($k = 5$) classification (SNV + detrend pre-processed) resulting in an overall classification accuracy of 73.6%. KNN classification score plot of black wildebeest (red), ostrich (green), springbok (blue) and zebra (turquoise), illustrating the predicted objects. (a) Score plot for predicted black wildebeest (top) vs. remaining species (bottom), (b) score plot of predicted zebra (top) vs. remaining species (bottom), (c) score plot of predicted springbok (top) vs. remaining species (bottom) and (d) score plot of predicted ostrich (top) vs. remaining species (bottom).

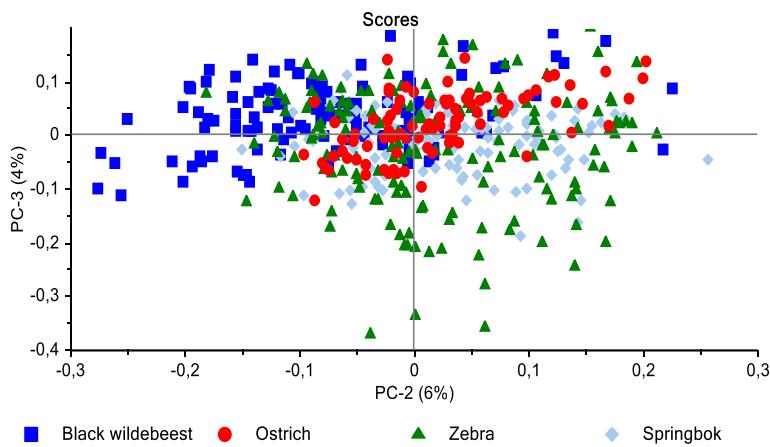


Figure 4.30 PCA analysis (SNV + detrend pre-processed) of species for black wildebeest (blue), ostrich (red), zebra (green) and springbok (light blue) classes. Minimal class separation was observed. Scores illustrated as KS-calibration PCA score plot of PC2 (6%) vs. PC3 (4%).

The decreased model accuracy of KNN(2) and decreased performance measures of KNN(3) is mainly attributed to the misclassification of objects assigned to the zebra and ostrich classes. These classes were misclassified due to the extensive overlap observed in the PCA score plots (**Figure 4.30**) which can be ascribed to the closely related protein content of zebra (ca. 22.3 g/100 g) (Onyango *et al.*, 1998; Hoffman *et al.*, 2016) and ostrich (ca. 20.9 g/100 g) (Sales, 1996). The lack of separation between these two species can also be attributed to their similar moisture contents, as illustrated by the difference spectra (**Figure 4.24** and **Figure 4.26**) and supported by literature. The sensitivity of both zebra and ostrich in KNN(2) is 47.1% and 73.3%, respectively. While the specificity is 75.5% and 91.5%, therefore confirming the model's ability to accurately predict the remaining classes, especially black wildebeest, even though it struggles with the zebra and ostrich (**Addendum A, Figure A12a – d**). The reason for this is that the protein content, as stated in literature, is higher for black wildebeest and differs from the other species, thus causing the black wildebeest samples to separate in the negative direction of PC2 (**Figure 4.30**). The same phenomenon is observed in KNN(3) (**Addendum A, Figure A13a – d**), where the sensitivity is 65.2% (zebra) and 71.1% (ostrich), and the specificity is 82.1% (zebra) and 91.6% (ostrich) (**Table 4.20**). From these results, it can be said that the black wildebeest can, with a high level of accuracy, be distinguished from the other three species with a misclassification rate below 10% (**Table 4.20**). Therefore, the improved classification of black wildebeest can be attributed to the higher protein content that was observed in the difference spectra, loadings and confirmed by literature.

The KNN results reported in this study were difficult to compare to those of McElhinney *et al.* (1999) due to differences in sample presentation (homogenised vs. whole steaks) and the use of statistical procedures, especially when referring to pre-processing [Savitzky-Golay (2nd derivative, 15 smoothing points)]. However, McElhinney *et al.* (1999) similarly reported that meat species can be distinguished based on the spectral response to compositional differences. These researchers also found that samples that resemble each other closely were difficult to differentiate, thus accounting for the greatest source of misclassification. This demonstrated that the KNN models were less accurate and did not perform well due to the close relationship between the species. Therefore, the KNN results in the current study indicates that these game species cannot be successfully differentiated because of their spectral similarities.

Table 4.20 The performance measures used to assess the KNN classification models of the four species, pre-processed with SNV + detrend.

Number of neighbours (k)	Species	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
2	BWB	87.6	3.5	8.9	88.6	94.5	75.4	81.5	12.4
	Zebra	77.9	2.4	19.7	47.1	75.5	88.0	61.3	22.1
	Springbok	81.5	7.3	11.3	74.6	89.2	65.4	69.7	18.6
	Ostrich	79.5	5.9	14.7	73.3	91.5	52.4	61.1	20.5
3	BWB	90.9	5.0	4.2	86.0	94.0	83.8	84.5	9.1
	Zebra	83.1	4.1	12.9	65.2	82.1	85.6	74.0	17.0
	Springbok	83.9	9.2	6.9	74.6	90.1	69.1	71.7	16.1
	Ostrich	81.9	12.0	6.1	71.1	91.6	55.7	62.4	18.1
5	BWB	90.9	4.2	5.0	83.3	93.0	85.6	84.4	9.1
	Zebra	81.5	10.6	5.5	81.2	82.2	62.6	71.1	18.5
	Springbok	83.5	10.6	6.0	78.1	91.2	66.9	72.1	16.6
	Ostrich	83.9	4.9	13.6	74.4	92.4	60.4	66.7	16.1

(KNN) K-nearest neighbours; (SNV) Standard normal variate; (BWB) Black wildebeest.

4.1.3.2.3 Discriminant analysis (DA)

The overall accuracy of the linear [5 PCs (72.3%); 6 PCs (74.4%)], quadratic [5 PCs (78.2); 6 PCs (80.6%)] and mahalanobis [5 PCs (75.1%); 6 PCs (78.4%)] models, suggested that a model with 6 PCs would provide better discrimination. These results can be explained by investigating the performance measures (**Table 4.22** and **Addendum A, Table A3, A4**) of the different discrimination models.

Table 4.21 DA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for species discrimination.

Number of principal components (PCs)	Method (distance calculations)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
5	Linear	Calibration	72.3	27.7
		Validation	70.5	29.5
	Quadratic	Calibration	78.2	21.8
		Validation	76.0	24.1
	Mahalanobis	Calibration	75.1	25.0
		Validation	71.3	28.7
6	Linear	Calibration	74.4	25.6
		Validation	72.6	27.4
	Quadratic	Calibration	80.6	19.5
		Validation	75.5	24.5
	Mahalanobis	Calibration	78.4	21.6
		Validation	72.2	27.9

(DA) Discriminant analysis; (SNV) Standard normal variate.

The QDA models presented the best discrimination results. The QDA (5 PCs) calibration model achieved an overall classification accuracy of 78.2% and a misclassification rate of 21.8% (**Table 4.21**). This suggests that the model can with a good accuracy distinguish between the four species. The model also revealed that the discrimination of black wildebeest was nearly perfect. The classification accuracy (92.5%), sensitivity (88.2%) and specificity (94.1%) confirmed this (**Table 4.22**).

The QDA model with 6 PCs achieved better results than that of the model with 5 PCs. The classification accuracy was improved to 80.6%, consequently lowering the misclassification rate to 19.5%. These improved model results were also observed in the improvement of the sensitivity (>66%) and the specificity (>86%) (**Table 4.22**), therefore confirming the model's ability to accurately differentiate between the four species. From both the QDA (5 PCs and 6 PCs) results it can be said that the black wildebeest can, with a high level of accuracy, be distinguished from the other three species with a misclassification rate below 8% (**Table 4.22**). Although the model exhibited an excellent overall calibration accuracy and good performance measures for each treatment (**Table 4.22**), the decreased validation accuracy suggests that the

classification model was over-fitted (Miller, 2005). Consequently, the over-fitted model has a low probability of correctly predicting the species classes when using an independent validation set.

In general, it is observed that the QDA method achieved the best classification results. The QDA models achieved better results because the algorithm was able to separate the cluster overlap, using a quadratic function to calculate a non-linear decision boundary. The lack of separation between the clusters can be attributed to their similar physiochemical characteristics, as previously discussed in **Section 4.1.3.1**. The results illustrate that QDA models are able to predict the different species, but due to over-fitting the models were not deemed effective. However, the models were able to distinguish between the four species, irrespective of the treatment (fresh or previously frozen) or the frozen period, with relatively high accuracies.

Table 4.22 The performance measures used to assess the QDA models of the four species, pre-processed with SNV + detrend.

Number of principal components (PCs)	Species	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
5	BWB	92.5	4.3	3.3	88.2	94.1	85.1	86.6	7.5
	Zebra	84.5	11.4	4.1	85.4	84.1	67.7	75.5	15.5
	Springbok	88.5	3.8	7.7	75.4	94.4	86.0	80.3	11.5
	Ostrich	86.1	4.7	9.3	63.6	93.8	77.8	70.0	14.0
6	BWB	92.5	3.9	3.6	86.7	94.7	86.0	86.3	7.5
	Zebra	86.4	9.3	4.3	85.7	86.7	73.6	79.2	13.6
	Springbok	90.5	4.3	5.2	81.4	94.1	84.2	82.8	9.5
	Ostrich	87.8	3.9	8.3	67.0	94.8	81.1	73.4	12.2

(QDA) Quadratic discriminant analysis; (SNV) Standard normal variate; (BWB) Black wildebeest.

4.1.3.2.4 Partial least squares discriminant analysis (PLS-DA)

The PLS-DA calibration model achieved an overall classification accuracy of 78.9% and a misclassification rate of 21.1% (**Table 4.23**). The PLS-DA score plots for the four species are given in **Figure 4.32a – d**. The 3D score plot [LV1 (84.72%) vs. LV2 (5.74%) vs. LV3 (3.36%)] demonstrates an overlap between the classes which is indicative of poor model calibration. The score plot of LV1 (84.72%) vs. LV3 (3.36%) (**Figure 4.31b**) exhibited the best species separation, with four clusters in the score space. The black wildebeest, springbok and ostrich exhibited the best separation in LV1, while the separation between zebra and the remaining three species was best described in LV3. The zebra was predominantly associated with the positive scores in both latent variables and the springbok with the negative scores in LV1 and positive scores in LV3. The ostrich was

predominantly associated with the negative scores in LV3 and both the positive and negative scores in LV1. The black wildebeest showed a similar clustering and was predominantly associated with the positive scores in LV1 and positive and negative scores in LV3. Therefore, the score plot (**Figure 4.31b**) exhibited that species clustering is best described in the direction of both LV1 and LV3.

Table 4.23 PLS-DA model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for species discrimination.

Number of latent variables (LVs)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
6	Calibration	78.9	21.1
	Cross-validation	77.8	22.2
	Validation	70.9	29.1

(PLS-DA) Partial least squares discriminant analysis; (SNV) Standard normal variate.

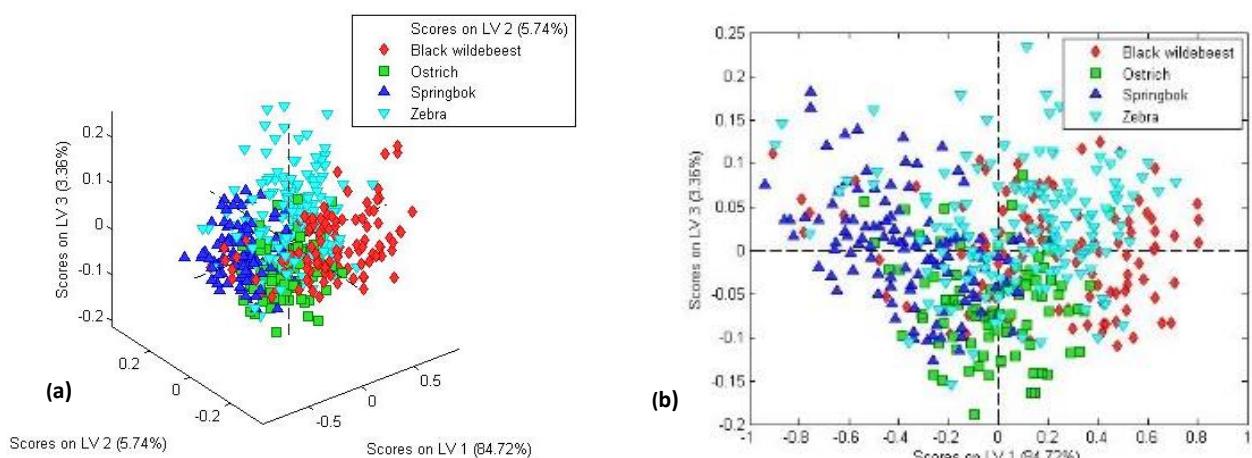
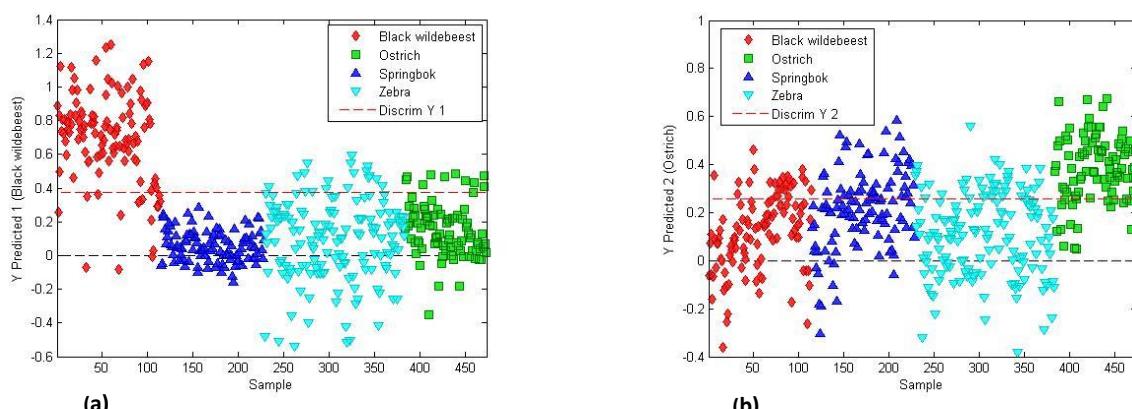


Figure 4.31 PLS-DA scores plot (SNV + detrend pre-processed) of black wildebeest (red), ostrich (green), springbok (blue) and zebra (turquoise). (a) 3D score plot of LV1 (84.72%) vs. LV2 (5.74%) vs. LV3 (3.36%), and (b) score plot of LV1 (84.72%) vs. LV3 (3.36%), colour coded per class species.



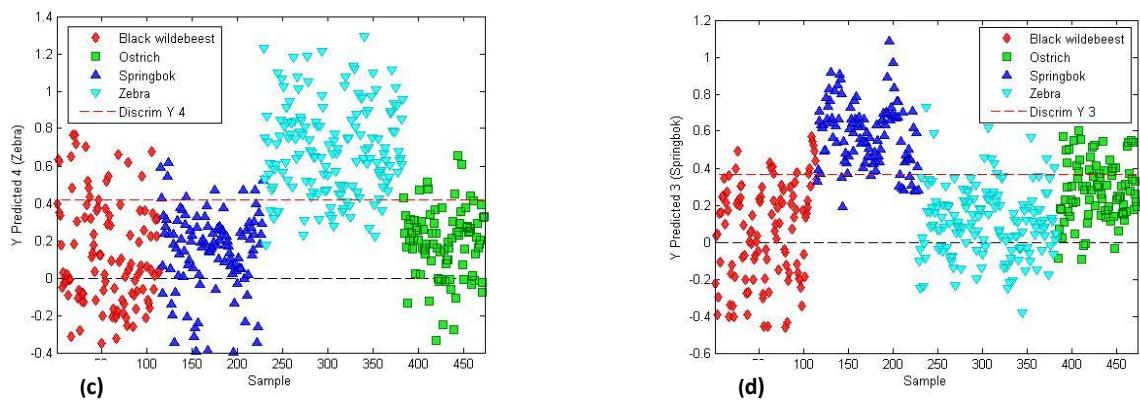


Figure 4.32 PLS-DA (6 LVs) (SNV + detrend pre-processed) resulting in an overall classification accuracy of 78.9%. PLS-DA prediction score plot of black wildebeest (red), ostrich (green), springbok (blue) and zebra (turquoise), illustrating the predicted objects. (a) Score plot of objects predicted as black wildebeest [above red line (Y_1)] vs. remaining species [below red line (Y_1)], (b) score plot of objects predicted as ostrich [above red line (Y_2)] vs. remaining species [below red line (Y_2)], (c) score plot of objects predicted as zebra [above red line (Y_4)] vs. remaining species [below red line (Y_4)] and (d) score plot of objects predicted as springbok [above red line (Y_3)] vs. remaining species [below red line (Y_3)].

The PLS-DA prediction score plots (Figure 4.32a,c,d) illustrated that the black wildebeest, zebra and springbok were nearly all correctly predicted. This was confirmed by the performance measures calculated in Table 4.24. The black wildebeest class resulted in the highest classification accuracy (91%), followed by zebra (89.2%), springbok (88%) and ostrich (84.8%). The sensitivity (85.3%), specificity (93%) and F1 score (83.4%) for black wildebeest is shown in Table 4.24. The sensitivity and specificity for zebra were above 88%, with a F1 score of 84%. These good performance measures are indicative of a model which is sufficient when predicting for these species. The spectral classification of balck wildebeest, zebra, ostrich and springbok can be attributed to their physiochemical differences as discussed in Section 4.1.3.1. Although the springbok exhibits a classification accuracy of 88%, the sensitivity (75.6%) is lower and the specificity (93%) considerably higher, therefore suggesting that the model has a higher probability of predicting the remaining species. These results confirms the model's ability to classify the black wildebeest and zebra species with little fault. The classification accuracy of ostrich (Table 4.24) was lower due to the increased false positive and false negative errors (Figure 4.32b). These increased error values can be attributed to the close spectral relationship to the species in question due to certain similar physiochemical characteridstics as discussed in Section 4.1.3.1. From these results it was evident that the model has a high probability of correctly predicting black wildebeest and zebra, with a high level of accuracy and small misclassification rate (<10%) (Table 4.24).

Even though the PLS-DA model exhibited classification accuracies of 78.9% (calibration) and 77.8% (cross-validation), indicative of an effective model, the validation accuracy (70.9%) decreased. This indicated that the model was over-fitted (Miller, 2005) and therefore not as effective as suggested by the calibration- and cross-validation accuracies.

The results reported herein were comparable to the PLS-DA results reported in previous studies by McElhinney *et al.* (1999), Cozzolino and Murray (2004) and Mamani-Linares *et al.* (2012). Therefore reinforcing that PLS-DA can be used to discriminate between black wildebeest, zebra, springbok and ostrich.

Table 4.24 The performance measures used to assess the PLS-DA models of the four species, pre-processed with SNV + detrend.

Number of latent variables (LVs)	Species	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
6	BWB	91.0	5.1	3.9	85.3	93.0	81.6	83.4	9.0
	Zebra	89.2	6.9	3.8	88.7	89.5	81.3	84.9	10.8
	Springbok	88.0	5.0	7.1	75.6	93.0	81.6	78.5	12.0
	Ostrich	84.8	6.6	8.6	61.6	91.5	67.8	64.6	15.2

(PLS-DA) Partial least squares discriminant analysis; (SNV) Standard normal variate; (BWB) Black wildebeest.

4.1.3.3 Optimal model selection

After the models [pre-processed with SNV + detrend (4th order polynomial)] were calculated to distinguish between the four different game species (black wildebeest, zebra, springbok and ostrich), regardless of the treatment (fresh or previously frozen) or frozen-period (1 – 9 months), the models were compared to identify the one with the best results. **Table 4.25** reveals that the QDA- and PLS-DA models produced the best results.

Table 4.25 An overview of the accuracies for the various classification and discrimination models, pre-processed with SNV + detrend to distinguish between species.

Model	Number of nearest neighbour (k), PCs or LVs	Classification accuracy (%)	Misclassification rate (%)
SIMCA	-	45.4	54.6
KNN	2	68.7	31.3
KNN	3	73.6	26.4
KNN	5	73.6	26.4
LDA	5	72.3	27.7
QDA	5	78.2	21.8
MDA	5	75.1	25.0
LDA	6	74.4	25.6
QDA	6	80.6	19.5
MDA	6	78.4	21.6
PLS-DA	8	78.9	21.1

(SNV) Standard normal variate; (PCs) Principal components; (LVs) Latent variables; (SIMCA) Soft independent modelling of class analogy; (KNN) K-nearest neighbour; (LDA) Linear discriminant analysis; (QDA) Quadratic discriminant analysis; (MDA) Mahalanobis discriminant analysis; (PLS-DA) Partial least squares discriminant analysis.

The data raised an important question as previously mentioned in **Section 4.1.2.3**. Therefore, the data were subjected to various pre-processing techniques [SNV; SNV + Savitzky-Golay (2nd derivative, 2nd order polynomial, 7/9 points) [SGd₂(7/9)] , SNV + detrend + Savitzky-Golay (2nd derivative, 2nd order polynomial, 7/9 points) [SGd₂(7/9)], Savitzky-Golay (1st derivative, 2nd order polynomial, 5/7 points) [SGd₁(5/7)] and Savitzky-Golay (2nd derivative, 2nd order polynomial, 7/9 points) [SGd₂(7/9)]] in different combinations. After PCA models were calculated, the data explored and outliers removed the QDA- and PLS-DA models were recalculated and evaluated.

4.1.3.4 Exploratory data analysis (EDA)

4.1.3.4.1 Principal component analysis (PCA)

The score plots of PC2 vs. PC3 (**Addendum A, Figure A14 – A22**) were used to investigate the differences between the four species, as the data shows that the variation, seen as the separation, was explained in the second component. Similar results were obtained for the various pre-processing techniques as previously discussed in **Section 4.1.2.4**. The separation was predominantly due to differences in moisture, fat, protein and pH as illustrated in the loading plots (**Addendum A, Figure A14b – A22b**). The results from the various pre-processing techniques suggests that the Savitzky-Golay first- and second derivative transformations would be sufficient and best suited for species separation. The Savitzky-Golay first- and second derivative transformations enhanced the differences in macronutrient composition of the biological samples. Even though Savitzky-Golay derivatives showed the best PCA results for separation, all the pre-processing techniques were used for further data analysis and model evaluation.

4.1.3.5 Multivariate data analysis (MDA): Optimal model development

4.1.3.5.1 Quadratic discriminant analysis (QDA)

The overall accuracy of the quadratic discriminant models, suggested that a model with 6 PCs would provide better discrimination. The QDA (6 PCs) models for the various types of pre-processing gave similar results. There are no substantial differences between the pre-processing techniques (**Table 4.26**). The QDA model, pre-processed with SNV + SGd₂(7) and SGd₁(7), presented the best discrimination results. These results can be explained by investigating the performance measures (**Table 4.27 – 4.28**).

Table 4.26 An overview of the classification rates of the QDA models with various pre-processing techniques applied for species discrimination.

Pre-processing	Number of PCs	Classification accuracy (%)	Misclassification rate (%)	Number of PCs	Classification accuracy (%)	Misclassification rate (%)
SNV		79.1	20.9		80.4	19.6
SNV+SGd ₂ (7)		80.9	19.1		85.1	14.9
SNV+SGd ₂ (9)		81.0	19.0		84.8	15.2
SNV+detrend+SGd ₂ (7)		80.4	19.6		82.5	17.5
SNV+detrend+SGd ₂ (9)	5	80.6	19.4	6	82.3	17.7
SGd ₁ (5)		80.7	19.3		82.6	17.4
SGd ₁ (7)		82.3	17.7		85.0	15.0
SGd ₂ (7)		81.6	18.4		84.9	15.1
SGd ₂ (9)		81.6	18.4		84.5	15.5

(QDA) Quadratic discriminant analysis; (PCs) Principal components; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

The QDA (6 PCs) calibration model achieved an overall classification accuracy of 85.1% for SNV + SGd₂(7) and 85% for SGd₁(7) with misclassification rates <15% (**Table 4.27 – 4.28**). Although both models had a calibration accuracy of 85%, further investigation shows that the model with an 83.1% validation accuracy is better (**Table 4.28**). Therefore, the SGd₁(7) pre-processed data resulted in a more robust model and is less affected by variation of intrinsic parameters in meat. This suggests that the model pre-processed with the Savitzky-Golay first derivative can accurately distinguish between the four species as discussed in **Section 4.1.2.5**. The model also revealed that the discrimination of black wildebeest and zebra was the most accurate. However, the probability of accurately predicting the zebra was higher and the classification accuracy (93.1%), sensitivity (94.4%) and specificity (92.5%) confirmed this (**Table 4.29**). The improved classification of black wildebeest and zebra can be further explained by referring to the PCA scores- and loadings plot (**Addendum A, Figure A20a – b**). The scores illustrate a slight clustering of the four species in the direction of PC2. This clustering and slight variation can be attributed to the negative absorption band at 1031 nm (**Addendum A, Figure A20b**). The band at 1031 nm (N-H stretch second overtone) is associated with protein (Osborne *et al.*, 1993). Therefore, the variation and slight clustering can be ascribed to the difference in protein content. The protein content has previously been quantified for these specific species and the authors have concluded that black wildebeest (Hoffman *et al.*, 2009) has the highest protein content of the species in question followed by zebra (Hoffman *et al.*, 2016), ostrich (Sales, 1996; Leygonie *et al.*, 2012b) and springbok (Hoffman & Wiklund, 2006; Hoffman *et al.*, 2007b). Therefore, these studies can explain the variation in the data (**Addendum A, Figure A20a**).

The QDA model, pre-processed with SGd₁(7), achieved better results than that of the original model pre-processed with SNV + detrend. To illustrate this, the performance measures of SGd₁(7) (**Table 4.27**) and

SNV + detrend (**Table 4.22**) are shown. The classification accuracy for differentiating between the species was 80.6% for SNV + detrend and 85% for SGd₁(7). Although both algorithms modify the data by removing the noise and corrects for the baseline shift (Barnes *et al.*, 1989). The occurring differences between these two pre-processing techniques is most likely because the Savitzky-Golay derivatives are calculated with an additional smoothing effect (Savitzky & Golay, 1964) as discussed in **Section 4.1.2.5**. Therefore, the SGd₁(7) pre-processing achieved better results as the separation due to differences in protein content were enhanced.

Table 4.27 QDA model calibration and validation results to assess the overall performance of the SNV+SGd₂(7) corrected for species discrimination.

Number of principal components (PCs)	Model [SNV + SGd ₂ (7)]	Classification accuracy (%)	Misclassification rate (%)
6	Calibration	85.1	14.9
	Validation	78.8	21.3

(QDA) Quadratic discriminant analysis; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points).

Table 4.28 QDA model calibration and validation results to assess the overall performance of the SGd₁(7) corrected data for species discrimination.

Number of principal components (PCs)	Model [SGd ₁ (7)]	Classification accuracy (%)	Misclassification rate (%)
6	Calibration	85.0	15.0
	Validation	83.1	16.9

(QDA) Quadratic discriminant analysis; [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

Table 4.29 The performance measures used to assess the QDA models of the four species, pre-processed with SGd₁(7).

Number of principal components (PCs)	Species	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
6	BWB	93.1	3.5	3.5	85.4	95.5	85.4	85.4	6.9
	Zebra	93.1	5.3	1.6	94.4	92.5	83.8	88.8	6.9
	Springbok	91.8	2.5	5.7	80.8	96.4	90.5	85.4	8.2
	Ostrich	89.8	4.9	5.4	79.8	93.4	80.5	79.8	10.2

(QDA) Quadratic discriminant analysis; (BWB) Black wildebeest; [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

4.1.3.5.2 Partial least squares discriminant analysis (PLS-DA)

The PLS-DA models for the various types of pre-processing gave similar results. The overall accuracy of the PLS-DA models, suggested that the models pre-processed with SGd₂(7) and SGd₂(9) would provide the best discrimination. There are no substantial differences between these mentioned pre-processing techniques (**Table 4.30**).

Table 4.30 An overview of the accuracies of the PLS-DA models with various pre-processing techniques applied for species discrimination.

Pre-processing	Calibration		Misclassification rate (%)	CV Classification accuracy (%)	Validation Classification accuracy (%)
	Number of LVs	Classification accuracy (%)			
SNV	4	71.4	28.6	70.4	69.2
SNV+SGd ₂ (7)	7	87.0	13.0	86.1	85.0
SNV+SGd ₂ (9)	6	86.5	13.5	83.3	82.9
SNV+detrend+SGd ₂ (7)	5	81.1	19.0	80.0	79.0
SNV+detrend+SGd ₂ (9)	4	78.7	21.3	77.7	72.3
SGd ₁ (5)	6	82.6	17.4	82.1	79.5
SGd ₁ (7)	5	78.7	21.3	78.3	79.3
SGd ₂ (7)	8	89.5	10.5	88.7	89.5
SGd ₂ (9)	8	89.8	10.3	88.9	91.6

(PLS-DA) Partial least squares discriminant analysis; (LVs) Latent variables; (CV) Cross-validation; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

The PLS-DA calibration models achieved an overall classification accuracy of 89.5% for SGd₂(7) and 89.8% for SGd₂(9) with misclassification rates <11% (**Table 4.30**). Although both models exhibited calibration- and CV accuracies of 89% and 88%, respectively, further investigation shows that the model with 91.6% validation accuracy was optimal (**Table 4.30**). Therefore, the SGd₂(9) pre-processed data (89.8%) resulted in a more robust model and is less affected by variation of intrinsic parameters in meat. This suggests that the model can accurately distinguish between the four species. These results can be explained by investigating the performance measures (**Table 4.31**).

The reason why SGd₂(9) achieved the best results is most likely attributed to the fact that the Savitzky-Golay second derivative amplified the wavebands which are important and best describes the species separation in the data (Savitzky & Golay, 1964). The model also revealed that the springbok class resulted in the highest classification accuracy (96.2%), followed by black wildebeest (95.1%), ostrich (94.7%) and zebra (92.5%) (**Table 4.31**). Therefore, the probability of accurately predicting springbok was higher as confirmed by the sensitivity (98.18%), specificity (99.38%) and F1 score (92.7%) in **Table 4.31**. The improved

classification of springbok and black wildebeest can be further explained by referring to the PCA scores- and loadings plot (**Addendum A, Figure A22a – b**). The scores illustrate a slight clustering of the species with minimal separation of springbok and black wildebeest from the other species in the positive direction of PC2. This positive separation can be attributed to pH, protein- and especially moisture content (**Addendum A, Figure A9b**) as discussed in **Section 4.1.2.4** and **Section 4.1.3.4**. The moisture content has previously been quantified for these specific species and the authors have concluded that black wildebeest (Hoffman *et al.*, 2009) and springbok (Hoffman & Wiklund, 2006; Hoffman *et al.*, 2007b) have similar but lower moisture contents than zebra (Onyango *et al.*, 1998; Hoffman *et al.*, 2016) and ostrich (Sales, 1996). Therefore, these studies can explain the variation in the data (**Addendum A, Figure A20a**). The scores plot could corroborate the literature as the clustering of the classes mirrored the moisture contents from literature.

The PLS-DA model, pre-processed with SGd₂(9), achieved better results than that of the original model pre-processed with SNV + detrend. To illustrate this, the performance measures of SGd₂(9) (**Table 4.31**) and SNV + detrend (**Table 4.24**) are shown. The overall classification accuracy for differentiating between the species was 78.9% for SNV + detrend and 89.8% for SGd₂(9). The SGd₂(9) model also exhibited better cross-validation- (88.9%) and validation (91.6%) results as well as individual species discrimination performance measures (**Table 4.31**). This indicates that the SGd₂(9) model is more robust than the original model, and has a higher predictive power for discriminating between the four species. The reason for the differences between these two pre-processing techniques were previously discussed in **Section 4.1.2.5**. Therefore, the SGd₂(9) pre-processing achieved better results as the separation due to differences in pH, protein- and especially moisture content was enhanced.

Table 4.31 The performance measures used to assess the PLS-DA models of the four species, pre-processed with SGd₂(9).

Number of latent variables (LVs)	Species	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
8	BWB	95.1	2.4	2.4	88.8	96.9	88.8	88.8	4.9
	Zebra	92.5	3.0	4.5	86.5	93.4	90.5	88.5	7.5
	Springbok	96.2	3.4	0.5	98.2	99.4	87.8	92.7	3.8
	Ostrich	94.7	2.0	3.3	87.0	95.6	91.7	89.3	5.3

(PLS-DA) Partial least squares discriminant analysis; (BWB) Black wildebeest; [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points).

4.1.3.6 Optimal model selection

After the models for the various types of pre-processing were calculated, the models with the best results were compared to identify the optimal model (**Table 4.32**). It was concluded that the PLS-DA model, pre-processed with Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points) was the best for differentiating between black wildebeest, zebra, springbok and ostrich, irrespective of the treatment (fresh or previously frozen) or the frozen period (1 – 9 months). The PLS-DA models achieved better results because PLS-DA generally outperforms QDA, when the classes are closely related, because it overcomes the collinearity problems often associated with QDA. SGd₂(9) outperformed the combined pre-processing of SNV + detrend, as the Savitzky-Golay transformation was found to have enhanced the differences in pH, protein- and especially moisture content, which was predominantly the contributors for species separation. These improved classification results correspond to what was found in literature, as previous studies reported that Savitzky-Golay 2nd derivative has the potential to improve the classification accuracy of prediction models (Ding & Xu, 1999; McElhinney *et al.*, 1999; Schmutzler *et al.*, 2015). Lastly, these results illustrate that samples that resemble each other closely were difficult to differentiate, thus accounting for the greatest source of misclassification. This suggests that greater and enhanced spectral differences between species would result in improved classification accuracies and better prediction models.

Table 4.32 An overview of the accuracies of the QDA and PLS-DA models with various pre-processing techniques applied to distinguish between species.

Model	Pre-processing	Calibration			Validation
		Number of PCs or LVs	Classification accuracy (%)	Misclassification rate (%)	Classification accuracy (%)
QDA	SNV + detrend	6	80.6	19.5	75.5
PLS-DA	SNV + detrend	8	78.9	21.1	70.9
QDA	SGd ₁ (7)	6	85.0	15.0	83.1
PLS-DA	SGd ₂ (9)	8	89.8	10.3	91.6

(QDA) Quadratic discriminant analysis; (PLS-DA) Partial least squares discriminant analysis; (PCs) Principal components; (LVs) Latent variables; (SNV) Standard normal variate; [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points).

4.1.3.7 Conclusion

NIR spectroscopy combined with MDA could accurately distinguish between the four species, irrespective of the treatment (fresh or previously frozen) or the frozen period. The PLS-DA discrimination model, data pre-processed with Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points), yielded the best results and could effectively distinguish the black wildebeest, zebra, springbok and ostrich from one another with an 89.8% accuracy. Throughout the study, the samples that resembled each other closely were difficult to differentiate. This is attributed to the samples' spectral similarities, thus accounting for the greatest source

of misclassification. It can thus be deduced that greater and enhanced spectral differences between species, would improve the classification accuracies and better the prediction models. In addition, it was found that the treatment (fresh or previously frozen) and frozen period does not influence the accuracy of the model for species discrimination.

4.2 Fresh vs. previously frozen meat determination

4.2.1 Spectral analysis

The fresh and frozen-thawed mean spectra of each species (pre-processed with SNV) was computed between 920 and 1651 nm (**Figure 4.33**) to investigate, determine and compare the chemical properties. The mean spectra of the fresh and frozen-thawed samples followed a similar trend with comparable absorption bands, however the intensity of the bands varied. The intensity differences can be attributed to the internal chemical composition. Three prominent absorption bands were exhibited at 970, 1193 and 1428 nm.

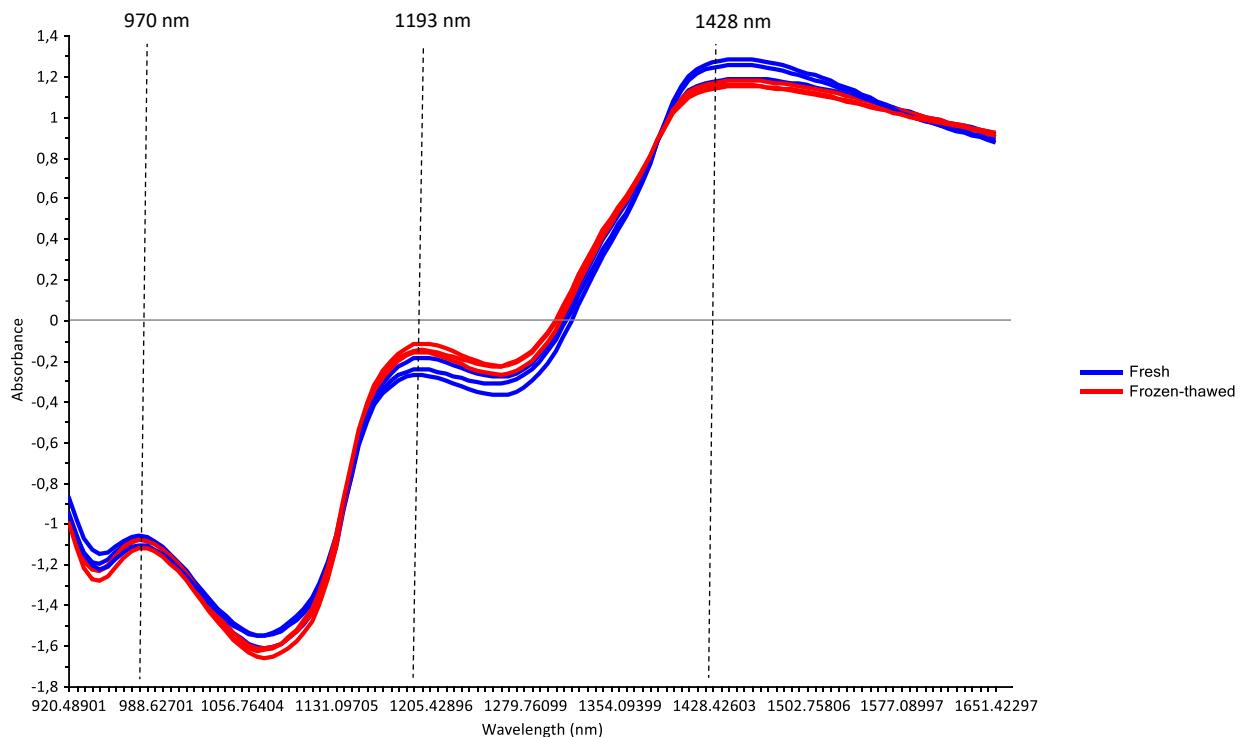


Figure 4.33 Standard normal variate (SNV) pre-processed mean spectra for fresh (**blue**) and frozen-thawed (**red**) black wildebeest, zebra, springbok and ostrich.

The band at 970 nm is related to the O-H second stretch overtone associated with water (Ding & Xu, 2000; Cozzolino & Murray, 2004; Barbin *et al.*, 2013b; Pu *et al.*, 2015; Ma *et al.*, 2016). The water band of the fresh meat has a slightly higher absorption value and this could be attributed to the possible moisture loss of the frozen-thawed meat (Leygonie *et al.*, 2012b). The 1193 nm band indicates the presence of fat (C-H stretch second overtone) as specified by Osborne *et al.* (1993). This band was higher in the frozen-thawed meat samples. It can be speculated that the fat concentration increased in the frozen-thawed samples due

to the loss of moisture (Leygonie *et al.*, 2012b). Lastly, the 1428 nm band represents the N-H stretch first overtone related to the CONH₂ group associated with the peptide bonds in proteins (Osborne *et al.*, 1993). This band was lower for the frozen-thawed meat samples, therefore suggesting that protein denaturation occurred due to the freezing process (Thyholt & Isaksson, 1997). The broad band at 1428 nm also has a contribution from moisture around 1420 and 1440 nm (O-H stretch first overtone). Therefore, the broad band at 1428 nm could be related to protein and moisture content (Osborne *et al.*, 1993). This band exhibited a higher absorption for the fresh meat samples and was indicative of the expected higher moisture content as noted by Barbin *et al.* (2013b).

4.2.2 Exploratory data analysis (EDA)

4.2.2.1 Principal component analysis (PCA)

Good class separation was observed in the PCA score plots (**Figure 4.34 – 4.35** and **Addendum B, Figure B1 – B5**) of the SNV + detrend (4th order polynomial) corrected data. This illustrates that the variation was explained in the first two components. In **Figure 4.36** a good class separation was observed between the fresh samples and the individual frozen periods for all four species, however, the frozen-thawed samples exhibited an overlap between the individual frozen periods. Therefore, lacking separation and distinct clustering for the different frozen periods.

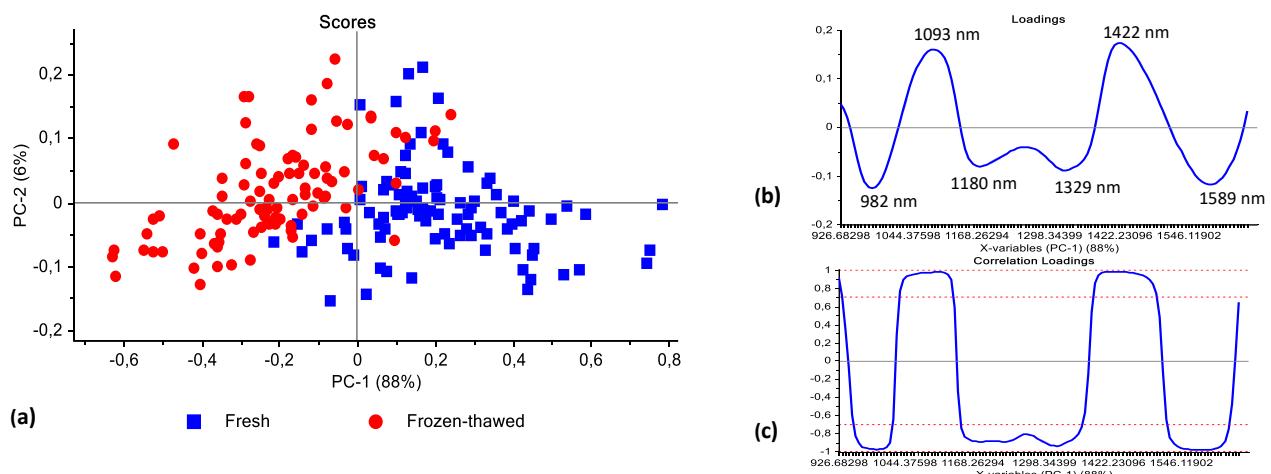


Figure 4. 34 PCA analysis (SNV + detrend pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (88%) vs. PC2 (6%). (b) PCA loadings line plot and (c) correlation loadings for PC1 with interpretable bands at 982, 1093, 1180, 1329, 1422 and 1589 nm.

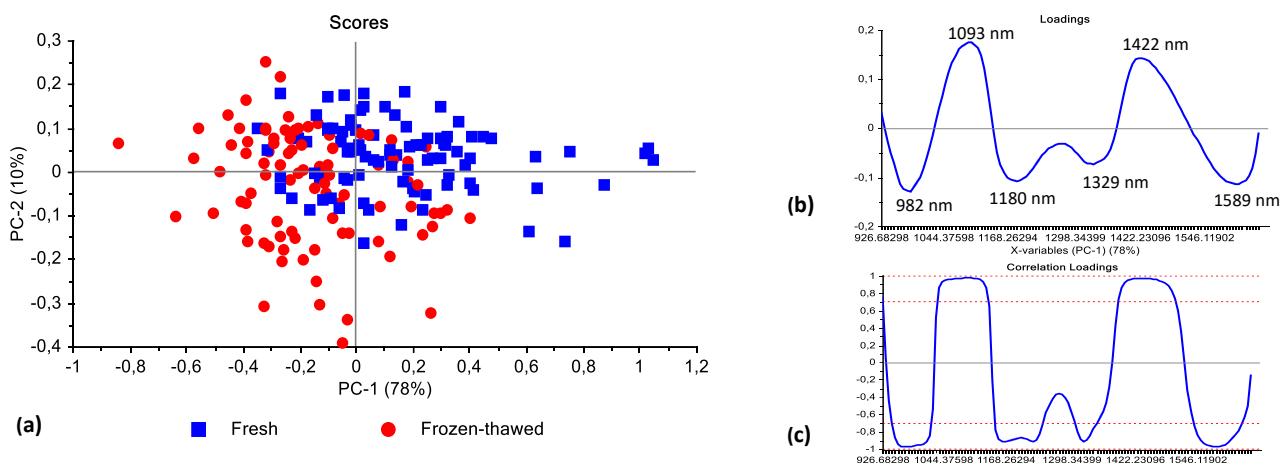


Figure 4.35 PCA analysis (SNV + detrend pre-processed) of zebra (LTL) [frozen up to 9 months] illustrating a slight overlap between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (78%) vs. PC2 (10%). (b) PCA loadings line plot and (c) correlation loadings for PC1 with interpretable bands at 982, 1093, 1180, 1329, 1422 and 1589 nm.

The variance observed in PC1 (**Figure 4.34a – 4.35a** and **Addendum B, Figure B1a – B5a**) for all four species may be attributed to the inter-class separation due to the difference between fresh and frozen-thawed meat. The loadings line plots of PC1 (**Figure 4.34b – 4.35b** and **Addendum B, Figure B1b – B5b**) exhibited interpretable positive bands at 1093 and 1422 nm and negative bands at 982, 1180, 1329 and 1589 nm. When evaluating the positive and negative loading bands of PC1, in combination with the score plot of PC1 vs. PC2 (**Figure 4.34a – 4.35a** and **Addendum B, Figure B1a – B5a**), the inter-class separation is mainly based on the positive spectral bands at 1093 nm, associated with pH (ElMasry *et al.*, 2012), and 1422 nm, which is related to moisture (Osborne *et al.*, 1993).

Water is known to be a major component in fresh meat and constitutes about 70 - 85% (Cozzolino & Murray, 2004; Prieto *et al.*, 2009). Freezing and thawing mainly influence the water fraction of meat (Leygonie *et al.*, 2012a) and due to the formation of ice crystals causes damage to the cellular structure of the meat. As reviewed by Leygonie *et al.* (2012a), the disrupted muscle fibre structure results in a reduced water-holding capacity of meat. Therefore, the main difference between fresh and frozen-thawed meat can be attributed to the loss of fluid from the meat tissue when defrosted. This decrease in moisture may cause an increase in the concentration of solutes, which consequently result in a decrease in the pH. Leygonie *et al.* (2012b) reported that the pH of previously frozen meat tends to be lower than that of fresh meat. This phenomenon is supported by the current findings, as the fresh meat samples were separated by higher loading values at 1093 nm, indicative of a higher pH (ElMasry *et al.*, 2012). The current findings also illustrated that the fresh samples were separated by higher loading values at 1422 nm, indicative of an expected increased moisture content. Barbin *et al.* (2013b) reported similar results for fresh and frozen-thawed pork meat and discriminated the samples based on physical and chemical changes.

The difference between fresh and frozen-thawed meat can also be attributed to change in the physical structure (Prieto *et al.*, 2009) caused by the formation of ice crystals (Leygonie *et al.*, 2012b). Downey and Beauchêne (1997b) reported that freezing-and-thawing alters the physical structure of the meat's surface layer, consequently changing the total reflectance spectra. Therefore a discrimination between fresh and frozen-thawed beef could be made based on the spectral baseline shift induced by freeze-thawing (Downey & Beauchêne, 1997b). In addition to the physical change in structure, other chemical changes may also occur; lipid- and protein oxidation as well as a decreased colour stability are all changes associated with previously frozen meat (Leygonie *et al.*, 2012b; reviewed by Leygonie *et al.*, 2012a). Thyholt and Isaksson (1997) successfully differentiated (90 – 100%) between frozen and unfrozen beef samples based on properties related to drip loss, irreversible denaturation and physical damage of myofibril proteins. Therefore, these results exhibit that both physical and chemical changes occurring within muscles during freeze-thawing can be used to successfully differentiate between fresh and previously frozen meat.

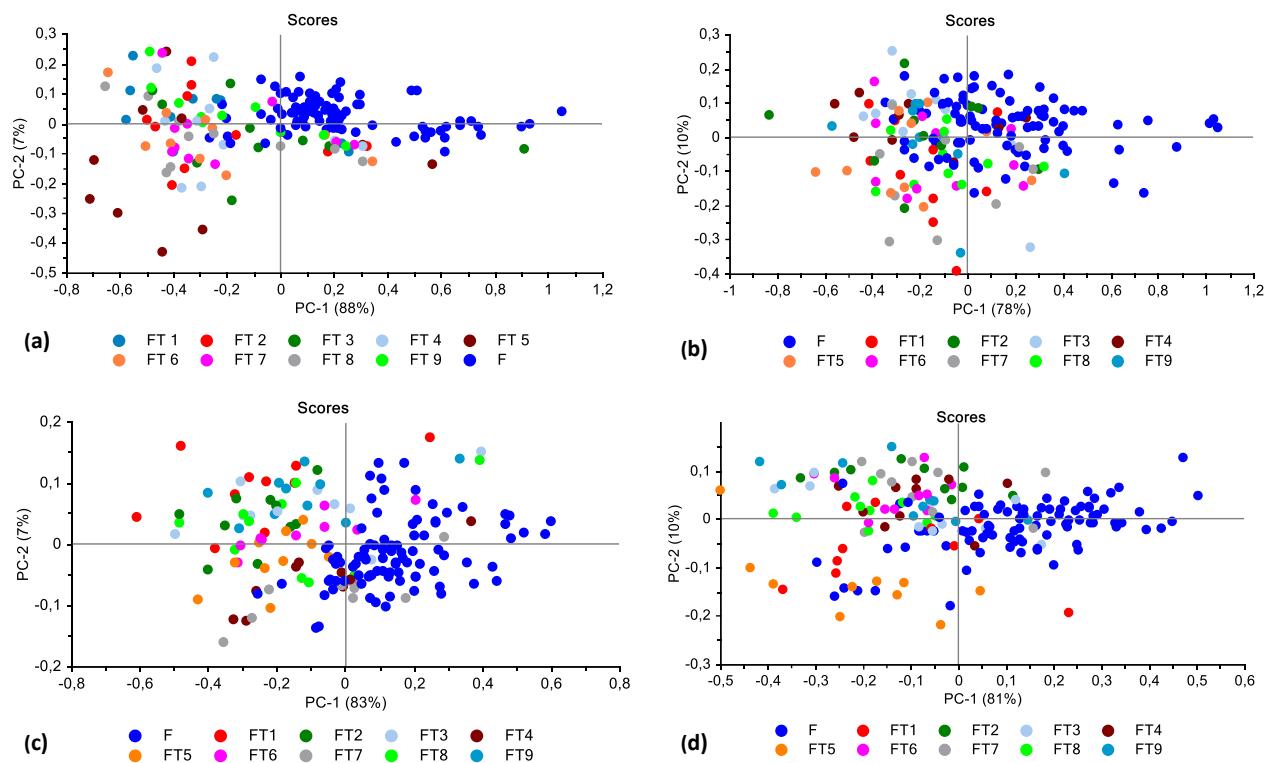


Figure 4.36 PCA analysis (SNV + detrend pre-processed) of fresh vs frozen period [frozen up to 9 months] for all four species illustrating good separation of the fresh samples with an overlap between frozen period classes. Scores illustrated as (a) PCA score plot of PC1 (88%) vs. PC2 (7%) for black wildebeest (LTL), (b) PCA score plot of PC1 (78%) vs. PC2 (10%) for zebra (LTL), (c) PCA score plot of PC1 (83%) vs. PC2 (7%) for springbok (LTL) and (d) PCA score plot of PC1 (81%) vs. PC2 (10%) for ostrich (FF).

Although a good separation was observed between the fresh and frozen-thawed samples, an overlap was exhibited between the different frozen periods (**Figure 4.36**). These figures exhibit a major overlap between the freeze-thaw treatments, thus resulting in a misclassification between classes. The lack of

separation can be ascribed to the sample's spectral similarities, as no substantial changes, between frozen periods were observed. Therefore, it can possibly be concluded that the extent of damage to cell membranes over the first month of freezing did not increase with an increase in frozen storage. From the principal component analysis, it was observed that the scores and loadings gave a global representation of the data. Hence the need for further exploratory data analysis to show the specific differences in the spectral data.

4.2.2.2 Difference spectra

The difference spectra of the fresh and frozen-thawed samples for each species followed a similar trend with comparable absorption bands, but the intensity of the bands varied. The difference spectra of black wildebeest, springbok and ostrich illustrated that the separation between fresh and frozen-thawed samples is a result of absorption bands at 982, 1162, 1341 and 1403 nm (**Addendum B, Figure B6 – B8**). On the other hand, the difference spectra of zebra (**Figure 4.37a – b**) illustrated that the separation was due to different absorption bands exhibited at 1124, 1242, 1285 and 1409 nm. The absorption band at 982 nm is related to the second overtone of O-H stretching vibrations and indicates the presence of water (Osborne *et al.*, 1993). Barbin *et al.* (2012b) also found that the absorption band at 982 nm can be used to determine the water-holding capacity of pork. The bands at 1124, 1162 and 1341 nm (C-H 2nd overtone stretch and C-H combination bands) is related to pH and is therefore, responsible for the spectral differences at these absorption bands (Barbin *et al.*, 2012b; ElMasry *et al.*, 2012; Iqbal *et al.*, 2013). The band at 1341 nm can also be associated with the water-holding capacity of beef (ElMasry *et al.*, 2011). A clear moisture band is observed due to O-H stretching and O-H bending combinations at 1403 and 1409 nm (Osborne *et al.*, 1993; Liu & Chen, 2001). These bands were also found to be related to the pH of beef (ElMasry *et al.*, 2012). Although the difference spectra for zebra exhibited different absorption bands, the separation observed was still caused by the same chemical changes due to freeze-thawing. These bands are associated with pH (1124, 1242 and 1409 nm) (ElMasry *et al.*, 2012; Iqbal *et al.*, 2013), meat tenderness (1285 nm) (ElMasry *et al.*, 2012) and moisture (1409 nm) (Liu & Chen, 2001).

The difference spectra could consequently be used to indicate the specific differences in the spectral data responsible for the separation between fresh and frozen-thawed meat. The results from the difference spectra indicated that the physical and chemical changes caused by freeze-thawing was similar to that of previous studies (Downey & Beauchêne, 1997b; Downey & Beauchêne, 1997a; Thyholt & Isaksson, 1997; Leygonie *et al.*, 2012b; Barbin *et al.*, 2013b). Therefore, indicating that the black wildebeest, zebra, springbok and ostrich species, irrespective of the muscle type or frozen period, all experience similar physical and chemical changes when subjected to frozen storage.

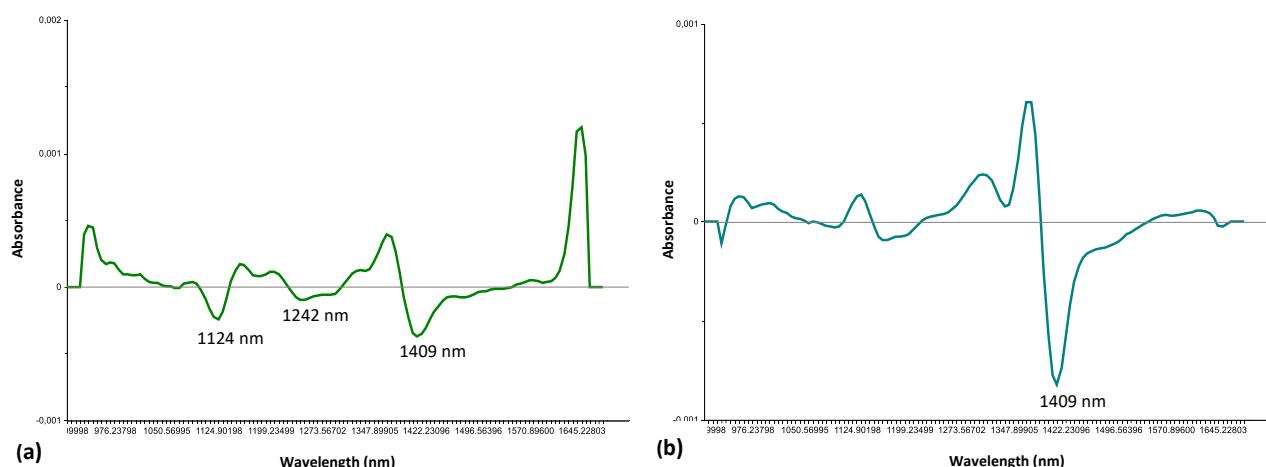


Figure 4.37 (a) Difference spectra of zebra (LTL) fresh and frozen-thawed samples, irrespective of the frozen period (1 – 9 months). (b) Difference spectra of zebra fresh and frozen-thawed samples, irrespective of the muscle type (LTL, BF, SM, ST, IS, SS, fillet).

4.2.3 Fresh vs. previously frozen meat determination irrespective of muscle type

This data set consisted of three species [zebra (*Equus quagga burchelli*), springbok (*Antidorcas marsupialis*), ostrich (*Struthio camelus*)], multiple muscle types for each species and the meat samples were frozen for one month. The aim of **Section 4.2.3** was to differentiate between the fresh and previously frozen meat samples for all three game species, irrespective of the muscle type as well as to determine whether the muscle type had an effect on the classification accuracies.

4.2.3.1 Multivariate data analysis (MDA): Model development

4.2.3.1.1 Soft independent modelling of class analogy (SIMCA)

The SIMCA model for springbok presented the best classification results. The calibration model achieved an overall classification accuracy of 100% (**Table 4.33**). This suggests that the model can accurately and with ease distinguish between fresh and frozen-thawed springbok meat. The classification accuracy (100%), sensitivity (100%) and specificity (100%) confirmed this (**Table 4.34**). The PCA score plot and loadings line plot (**Addendum B, Figure B1**) support these results as a good separation was observed between the two treatments (fresh vs. frozen-thawed). The separation and correct classification of the two classes was mainly attributed to the loss of moisture and change in pH. These properties changed during freezing and the correlation was observed in the difference spectra (**Addendum B, Figure B7b**) and loadings (**Addendum B, Figure B1b**). The irreversible damage to the cellular structure of the meat, due to the formation of the ice crystals (Thyholt & Isaksson, 1997; reviewed by Leygonie *et al.*, 2012a; Leygonie *et al.*, 2012b), could also be important predictors. Therefore, these results compare to previous studies on beef and pork meat, where the researchers found that both physical and chemical changes occurring within the meat during freezing can be used to successfully discriminate between fresh and frozen-thawed meat (Downey & Beauchêne, 1997b; Thyholt & Isaksson, 1997; Prieto *et al.*, 2009; Barbin *et al.*, 2013b).

Table 4.33 SIMCA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for fresh or previously frozen meat classification.

Species	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
Zebra	Calibration	77.5	22.5
	Validation	74.2	25.8
Springbok	Calibration	100.0	0
	Validation	100.0	0
Ostrich	Calibration	65.2	34.8
	Validation	47.4	52.6

(SIMCA) Soft independent modelling of class analogy; (SNV) Standard normal variate.

The SIMCA models for zebra and ostrich gave unsatisfactory results. These unsatisfactory classification results confirmed what was seen in the PCA score plots (**Figure 4.34a** and **Addendum B, Figure B2a**). The lack of separation can be ascribed to the samples' spectral similarities, and as the SIMCA algorithm aims to classify samples based on spectral similarities (Massart *et al.*, 1988), the classification accuracy was low. When comparing these model results, it was observed that the zebra model achieved a higher classification accuracy (77.5%) than the ostrich model (65.2%) and exhibited improved performance measures (**Table 4.34**). These results can be explained by the difference spectra (**Figure 4.37b** and **Addendum B, Figure B8b**). Although the separation between fresh and frozen-thawed meat, for both species, can mainly be attributed to the loss of moisture, the intensity of these absorption bands differed. The difference spectra for zebra (**Figure 4.37b**) exhibited a moisture band at 1409 nm, with a higher intensity compared to the moisture band exhibited in the difference spectra for ostrich (**Addendum B, Figure B8b**). This higher moisture band illustrates that there was a larger difference between the fresh and frozen-thawed samples, therefore achieving better class separation. The lower moisture band for ostrich exhibits that there is little difference between the fresh and frozen-thawed samples. This suggests that the spectra are similar and for this reason the SIMCA model for ostrich achieved a low classification accuracy. Downey and Beauchêne (1997b) reported similar results when discriminating between fresh and frozen-thawed beef meat. These researchers found that SIMCA models were less accurate and did not perform well due to the spectral similarities of the samples. Nonetheless, the SIMCA results in the current study indicates that fresh and frozen-thawed meat samples, irrespective of the muscle type, can successfully be discriminated if the spectral data is substantially different.

Table 4.34 The performance measures used to assess the SIMCA classification models for fresh vs. frozen-thawed meat of the three species, pre-processed with SNV + detrend.

Species	Treatment	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
Zebra	Fresh	77.5	10.1	12.4	74.7	76.8	78.3	76.5	22.5
	FT	77.5	12.4	10.1	80.2	78.3	76.8	78.5	22.5
Springbok	Fresh	100.0	0	0	100.0	100.0	100.0	100.0	0
	FT	100.0	0	0	100.0	100.0	100.0	100.0	0
Ostrich	Fresh	65.2	8.7	26.1	53.9	57.1	77.8	63.6	34.8
	FT	65.2	26.1	8.7	80.0	77.8	57.1	66.7	34.8

(SIMCA) Soft independent modelling of class analogy; (SNV) Standard normal variate; (FT) Frozen-thawed.

4.2.3.1.2 K-nearest neighbour (KNN)

The overall model accuracy of zebra, springbok and ostrich suggested that three nearest neighbours would provide the best classification for zebra (91.7%) and ostrich (80.0%), whereas five nearest neighbours provided the best classification for springbok (98.9%) (**Table 4.35**).

Table 4.35 KNN model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for fresh or previously frozen meat classification.

Species	Number of neighbours (k)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
Zebra	2	Calibration	90.3	9.7
		Cross-validation	91.7	8.3
		Validation	95.8	4.2
	3	Calibration	91.7	8.3
		Cross-validation	92.4	7.6
		Validation	100.0	0
	5	Calibration	91.0	9.0
		Cross-validation	91.0	9.0
		Validation	100.0	0
Springbok	2	Calibration	97.9	2.1
		Cross-validation	97.9	2.1
		Validation	100.0	0
	3	Calibration	97.9	2.1
		Cross-validation	97.9	2.1
		Validation	100.0	0
	5	Calibration	98.9	1.1
		Cross-validation	98.9	1.1
		Validation	100.0	0
Ostrich	2	Calibration	73.3	26.7
		Cross-validation	80.0	20.0
		Validation	50.0	50.0
	3	Calibration	80.0	20.0
		Cross-validation	80.0	20.0
		Validation	50.0	50.0
	5	Calibration	70.0	30.0
		Cross-validation	70.0	30.0
		Validation	50.0	50.0

(KNN) K-nearest neighbours; (SNV) Standard normal variate.

The KNN(5) calibration model for springbok achieved an overall classification- and cross-validation accuracy of 98.9%, with an improved classification accuracy for the validation model (100%) (**Table 4.35**). The excellent cross-validation- and improved validation accuracies are indicative of a classification model that is not over-fitted (Miller, 2005). Although this model had excellent accuracies, a few misclassified objects were observed (**Figure 4.38**). This accounted for the misclassification rate of 1.1% for the overall model (**Table 4.35**). The sensitivity and specificity for the frozen-thawed samples were both 100% (**Table 4.36**). This indicates that the model has a high probability of correctly classifying the frozen-thawed meat samples (**Figure 4.38**). The lower sensitivity for the fresh samples reveals that the model was less suited for predicting this class. This phenomenon can be explained by referring to the PCA score plot (**Figure 4.38**), as k -nearest neighbour was performed on the PC scores. The KS-calibration PCA score plot illustrates a slight overlap between the two classes, with a larger number of fresh samples displaying a near distance to the frozen-thawed samples. Therefore, the fresh samples are assigned to the predominant class, frozen-thawed. Hence, explaining why the model is less suited for predicting the fresh samples.

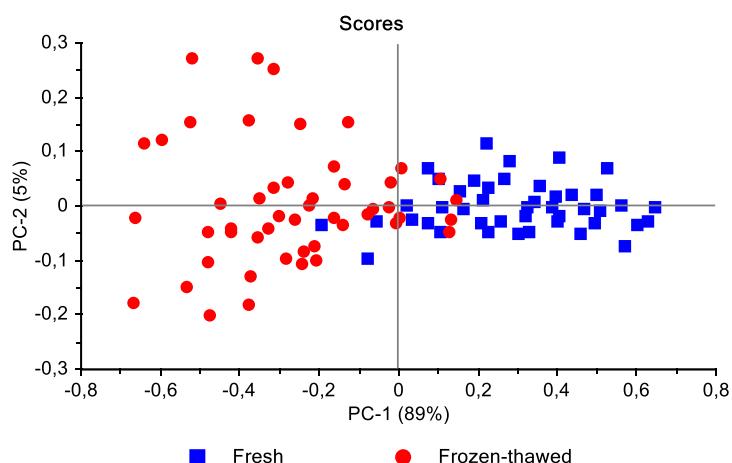


Figure 4.38 PCA analysis (SNV + detrend pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as KS-calibration PCA score plot of PC1 (89%) vs. PC2 (5%).

The decreased KNN(3) model accuracy for both zebra and ostrich is mainly attributed to the misclassification of objects assigned to the fresh and frozen-thawed classes. These classes were misclassified due to the extensive overlap observed in the PCA score plots (**Figure 4.39a – b**). The sensitivity of both fresh and frozen-thawed samples in KNN(3) for zebra is 88.2% and 94.7%, respectively. While the specificity is 90% and 93.7%, therefore confirming the model's ability to accurately predict the frozen-thawed samples even though it struggles with the fresh samples. Once again this can be ascribed to the larger number of fresh samples displaying a close distance to the frozen-thawed samples (**Figure 4.39a**), as previously discussed. A similar phenomenon was observed in KNN(3) for ostrich, except this model had a higher affinity to accurately predict the fresh samples, where the sensitivity was 92.8% (fresh) and 68.7% (frozen-thawed), and the

specificity was 91.7% (fresh) and 72.2% (frozen-thawed) (**Table 4.36**). Although the classification model exhibited good accuracies (80%) for both the calibration and cross-validation, the model achieved a validation accuracy of 50%, suggesting that the model was over-fitted (Miller, 2005).

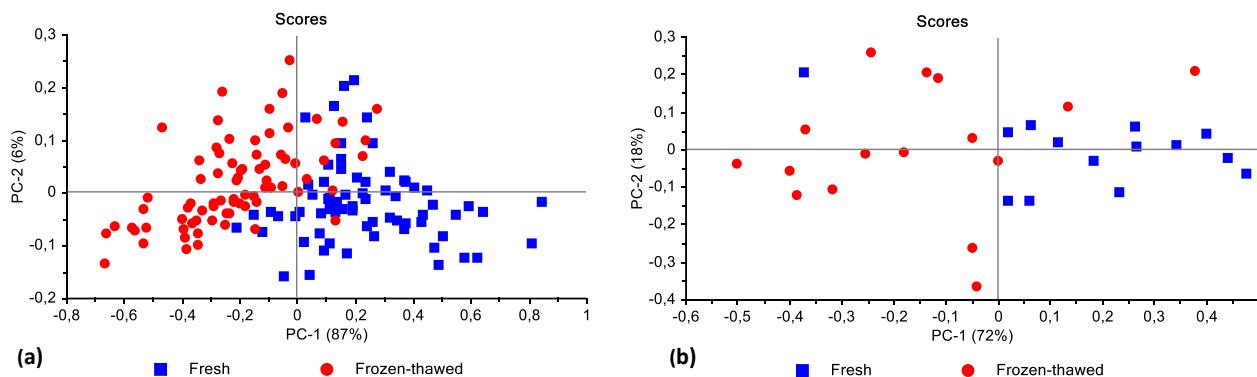


Figure 4.39 PCA analysis (SNV + detrend pre-processed) of (a) zebra (LTL, BF, SM, ST, IS, SS, fillet) and (b) ostrich (BD, FF) [frozen up to 1 month] illustrating minimal separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) KS-calibration PCA score plot of PC1 (87%) vs. PC2 (6%) for zebra, and (b) KS-calibration PCA score plot of PC1 (72%) vs. PC2 (18%) for ostrich.

The KNN model for springbok (98.9%) presented the best results, followed by zebra (91.7%) and ostrich (80%). These results are comparable to the SIMCA results, where similar findings were observed. Therefore, the improved classification of fresh and frozen-thawed springbok meat can be attributed to a greater degree of physical and chemical changes that occurred within the meat during freezing. The KNN results for springbok (98.9%) and zebra (91.7%) are comparable to the results reported by Thyholt and Isaksson (1997). These researchers classified fresh, frozen and re-frozen beef with an accuracy of 93%, concluding that NIR spectroscopy can be used for beef classification into fresh and frozen-thawed classes. The data also gave promising separation between the re-frozen meat samples (frozen once, frozen twice, and so forth) (Thyholt & Isaksson, 1997). Therefore, the KNN results in the current study indicates that fresh and frozen-thawed meat samples, irrespective of the muscle type, can successfully be discriminated.

Table 4.36 The performance measures used to assess the KNN classification models for fresh vs. frozen-thawed meat of the three species, pre-processed with SNV + detrend.

Species	Number of neighbours (<i>k</i>)	Treatment	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
Zebra	3	Fresh	91.7	2.8	5.6	88.2	90.0	93.7	90.9	8.3
		FT	91.7	5.6	2.8	94.7	93.7	90.0	92.3	8.3
Springbok	5	Fresh	98.9	0	1.1	97.7	98.0	100.0	98.9	1.1
		FT	98.9	1.1	0	100.0	100.0	98.0	99.0	1.1
Ostrich	3	Fresh	80.0	16.7	3.3	92.8	91.7	72.2	81.3	20.0
		FT	80.0	3.3	16.7	68.7	72.2	91.7	78.6	20.0

(KNN) *K*-nearest neighbours; (SNV) Standard normal variate; (FT) Frozen-thawed.

4.2.3.1.3 Discriminant analysis (DA)

The linear discriminant analysis (LDA) model for springbok (5 PCs) presented the best discrimination results. The separation between the fresh and frozen-thawed meat samples was achieved with a 100% overall accuracy (**Table 4.37**). The performance measures also revealed that the discrimination of both the fresh and frozen-thawed meat samples were all correctly predicted. The classification accuracy, sensitivity, specificity and F1 score for both classes were 100% (**Table 4.38**).

Table 4.37 Optimal DA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for fresh or previously frozen meat discrimination.

Species	Number of principal components (PCs)	Method (distance calculations)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
Zebra	6	Linear	Calibration	95.8	4.2
			Validation	97.9	2.1
		Quadratic	Calibration	97.9	2.1
			Validation	97.9	2.1
			Calibration	95.8	4.2
			Validation	97.9	2.1
	5	Mahalanobis	Calibration	100.0	0
			Validation	100.0	0
		Quadratic	Calibration	100.0	0
			Validation	100.0	0
	Springbok	Mahalanobis	Calibration	100.0	0
			Validation	100.0	0
Ostrich	6	Linear	Calibration	86.7	13.3
			Validation	70.0	30.0
		Quadratic	Calibration	96.7	3.3
			Validation	70.0	30.0
			Calibration	96.7	3.3
	5	Mahalanobis	Validation	60.0	40.0
			Calibration	96.7	3.3
		Quadratic	Validation	60.0	40.0
			Calibration	96.7	3.3
			Validation	60.0	40.0

(DA) Discriminant analysis; (SNV) Standard normal variate.

The quadratic discriminant analysis (QDA) models for zebra (6 PCs) and ostrich (5 PCs) achieved similar overall accuracies. The QDA model for zebra achieved an overall classification accuracy of 97.9% for both the calibration and validation model, with a misclassification rate of 2.1% (**Table 4.37**). These identical accuracies are indicative of a model that is not over-fitted (Miller, 2005). On the other hand, the QDA model for ostrich achieved an overall classification accuracy 96.7% for the calibration with a misclassification rate of 3.3%. Although the model exhibited an excellent overall calibration accuracy as well as good performance measures for each treatment (**Table 4.37**), the decreased validation accuracy suggests that the classification model was over-fitted (Miller, 2005). The disadvantage of a model that has been over-fitted is that it is very

specific to the exact data used to build the model (Miller, 2005), and consequently, the over-fitted model has a low probability of correctly predicting the fresh and frozen-thawed classes when using an independent validation set.

In general, it was observed that the LDA model achieved the best classification results for springbok. While, the QDA models exhibited better discrimination results for the zebra and ostrich. The LDA model for springbok achieved better results because the algorithm was able to calculate an optimal linear projection between the classes, due to the good separation illustrated in the PCA score plot (**Addendum B, Figure B1a**). The cluster overlap observed in the PCA score plots for zebra (**Figure 4.34a**) and ostrich (**Addendum B, Figure B2a**), explains why the LDA algorithm was unable to calculate an optimal linear projection between the fresh and frozen-thawed samples. Because the QDA algorithm calculates a non-linear decision boundary, using a quadratic function, it was able to separate the clusters and successfully discriminate between the treatments. These results for springbok (100%), zebra (97.9%) and ostrich (96.7%) are comparable to results reported by Thyholt and Isaksson (1997). These researchers were able to classify fresh and frozen-thawed beef with a 95.3% accuracy. The results in the current study therefore show that DA can be used to discriminate between fresh and frozen-thawed meat samples of springbok, zebra and ostrich, irrespective of the muscle type.

Table 4.38 The performance measures used to assess the DA models for fresh vs. frozen-thawed meat of the three species, pre-processed with SNV + detrend.

Species	Number of principal components (PCs)	Method	Treatment	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
Zebra	6	QDA	Fresh	97.9	2.1	0	100.0	100.0	95.8	97.8	2.1
			FT	97.9	0	2.1	96.1	95.8	100.0	98.0	2.1
Springbok	5	LDA	Fresh	100.0	0	0	100.0	100.0	100.0	100.0	0
			FT	100.0	0	0	100.0	100.0	100.0	100.0	0
Ostrich	5	QDA	Fresh	96.7	0	3.3	93.6	93.3	100.0	96.8	3.3
			FT	96.7	3.3	0	100.0	100.0	93.3	96.6	3.3

(DA) Discriminant analysis; (LDA) Linear discriminant analysis; (QDA) Quadratic discriminant analysis; (SNV) Standard normal variate; (FT) Frozen-thawed.

4.2.3.1.4 Partial least squares discriminant analysis (PLS-DA)

The PLS-DA models for zebra, springbok and ostrich gave satisfactory discrimination results. The PLS-DA calibration models achieved overall classification accuracies of 95.1% (zebra), 100% (springbok), 90% (ostrich) and misclassification rates of 4.9% (zebra), 10% (ostrich) (**Table 4.39**). The PLS-DA score plots for the three species are given in **Figure 4.41**. The score plots (LV1 vs. LV2) demonstrates minimal overlap between the treatment classes, which is indicative of satisfactory model calibration. The score plot of LV1 (89.09%) vs. LV2 (2.85%) of springbok (**Figure 4.41b**) exhibited the best separation, with two prominent clusters. The separation between the fresh and frozen-thawed treatments were best described in the direction of LV1, while LV2 accounted for little class separation. The fresh samples were predominantly associated with the negative scores in both latent variables while the frozen-thawed samples had positive scores. Therefore, the score plots (**Figure 4.40a – b**) exhibited that the separation is best described in the direction of both LV1 and LV2.

Table 4.39 PLS-DA model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for fresh or previously frozen meat discrimination.

Species	Number of latent variables (LVs)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
Zebra	3	Calibration	95.1	4.9
		Cross-validation	93.1	6.9
		Validation	100.0	0
Springbok	2	Calibration	100.0	0
		Cross-validation	100.0	0
		Validation	100.0	0
Ostrich	1	Calibration	90.0	10.0
		Cross-validation	83.3	16.7
		Validation	50.0	50.0

(PLS-DA) Partial least squares discriminant analysis; (SNV) Standard normal variate.

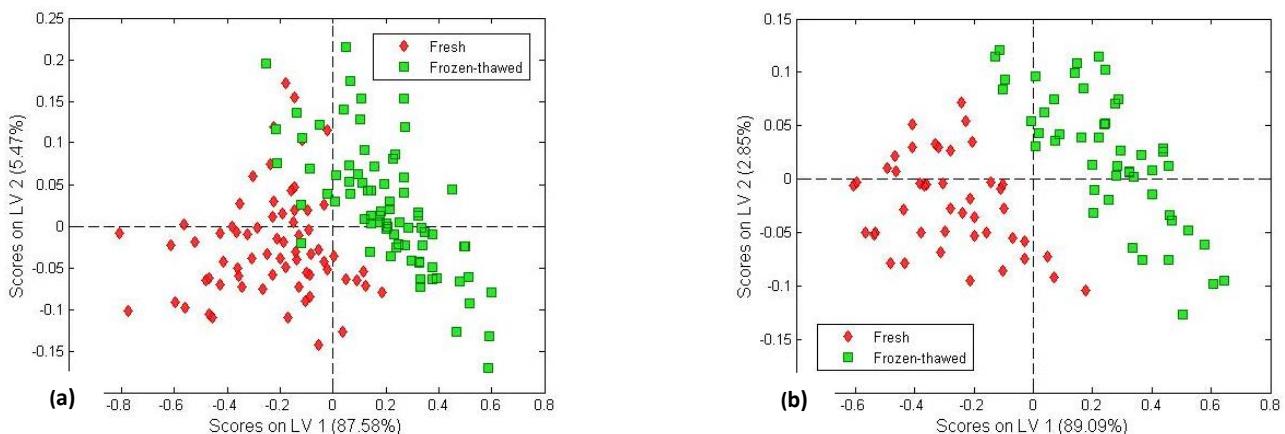


Figure 4.40 PLS-DA scores plot (SNV + detrend pre-processed) of fresh (red) and frozen-thawed (green) meat samples. (a) score plot of LV1 (87.58%) vs. LV2 (5.47%) for zebra and (b) score plot of LV1 (89.09%) vs. LV2 (2.85%) for springbok, colour coded per treatment class.

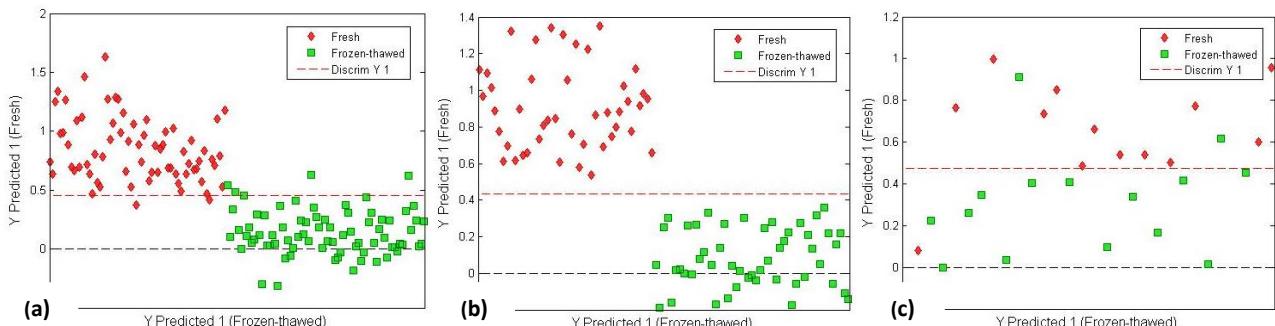


Figure 4.41 PLS-DA models (SNV + detrend pre-processed) indicating satisfactory overall classification accuracies. PLS-DA prediction score plot of (a) zebra, (b) springbok and (c) ostrich, illustrating the predicted fresh (red) and frozen-thawed (green) objects. (a) Score plot (3 LVs) for zebra (95.1%) of objects predicted as fresh [above red line (Y1)] vs. frozen-thawed [below red line (Y1)], (b) score plot (2 LVs) for springbok (100%) of objects predicted as fresh [above red line (Y1)] vs. frozen-thawed [below red line (Y1)] and (c) score plot (1 LV) for ostrich (90%) of objects predicted as fresh [above red line (Y1)] vs. frozen-thawed [below red line (Y1)].

The PLS-DA prediction score plot (**Figure 4.41b**) for springbok illustrated that the fresh and frozen-thawed treatments were all correctly predicted and the performance measures confirmed this (**Table 4.40**). On the other hand, the PLS-DA score plot for zebra (**Figure 4.41a**) and ostrich (**Figure 4.41c**) illustrated that the treatments were not all correctly predicted. This was also confirmed by the performance measures (**Table 4.40**). The springbok model resulted in the highest classification accuracy (100%) followed by zebra (95.1%) and ostrich (90%). Although both treatments for the zebra model achieved classification accuracies of 95.1%, the model exhibited a higher probability of predicting the fresh samples. This is confirmed by the higher sensitivity (97.1%), specificity (97.3%) and F1 score (95%) (**Table 4.40**). The same trend was observed for the ostrich model. These results confirm the models' ability to classify the fresh meat samples with

minimal fault. Even though the ostrich model exhibited classification accuracies of 90% (calibration) and 83.3% (cross-validation), indicative of an effective model, the validation accuracy (50%) decreased. This indicated that the model was over-fitted (Miller, 2005) and therefore not as effective as suggested by the calibration- and cross-validation accuracies. This is because the model was more specific for the calibration set, and was not a good representation of the overall data. Therefore, the validation set comprised data that was not represented by the model. This could be due to the small size of data set [calibration (30) and validation (10)]. Therefore, the validation set could possibly improve by increasing the number of samples in the calibration- as well as the validation set. The results in the current study suggest that PLS-DA can be used to discriminate between fresh and frozen-thawed game meat samples. Ropodi *et al.* (2018) reported similar results (93.3%) for fresh and frozen-thawed minced beef meat and discriminated the samples based on physical and chemical changes as discussed in **Section 4.2.2**.

Table 4.40 The performance measures used to assess the PLS-DA models for fresh vs. frozen-thawed meat of the three species, pre-processed with SNV + detrend.

Species	Number of latent variables (LVs)	Treatment	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
Zebra	3	Fresh	95.1	3.5	1.4	97.1	97.3	93.0	95.0	4.9
		FT	95.1	1.4	3.5	93.4	93.0	97.3	95.3	4.9
Springbok	2	Fresh	100.0	0	0	100.0	100.0	100.0	100.0	0
		FT	100.0	0	0	100.0	100.0	100.0	100.0	0
Ostrich	1	Fresh	90.0	6.7	3.3	92.9	93.3	86.7	89.7	10.0
		FT	90.0	3.3	6.7	87.5	86.7	93.3	90.3	10.0

(PLS-DA) Partial least squares discriminant analysis; (SNV) Standard normal variate; (FT) Frozen-thawed.

4.2.3.2 Optimal model selection

After the models [pre-processed with SNV + detrend (4th order polynomial)] were calculated to distinguish between the fresh and previously frozen meat, irrespective of muscle type, the models were compared to identify the one with the best results. **Table 4.41** reveals that the DA (QDA, LDA) and PLS-DA models produced the best results.

Table 4.41 An overview of the accuracies for the various classification and discrimination models, pre-processed with SNV + detrend to distinguish between fresh or previously frozen meat, irrespective of the muscle type.

Species	Model	Number of nearest neighbour (<i>k</i>), PCs or LVs	Classification accuracy (%)	Misclassification rate (%)
Zebra	SIMCA	-	77.5	22.5
	KNN	3	91.7	7.6
	QDA	6	97.9	2.1
	PLS-DA	3	95.1	4.9
Springbok	SIMCA	-	100.0	0
	KNN	5	98.9	1.1
	LDA	5	100.0	0
	PLS-DA	2	100.0	0
Ostrich	SIMCA	-	65.2	34.8
	KNN	3	80.0	20.0
	QDA	5	96.7	3.3
	PLS-DA	1	90.0	10.0

(SNV) Standard normal variate; (PCs) Principal components; (LVs) Latent variables; (SIMCA) Soft independent modelling of class analogy; (KNN) K-nearest neighbour; (LDA) Linear discriminant analysis; (QDA) Quadratic discriminant analysis; (PLS-DA) Partial least squares discriminant analysis.

This data was then subjected to different pre-processing techniques to determine if the classification accuracies could be improved. The following pre-processing techniques were investigated: [SNV; SNV + Savitzky-Golay (2nd derivative, 2nd polynomial order, 7/9 points) [SGd₂(7/9)], SNV + detrend + Savitzky-Golay (2nd derivative, 2nd polynomial order, 7/9 points) [SGd₂(7/9)], Savitzky-Golay (1st derivative, 2nd polynomial order, 5/7 points) [SGd₁(5/7)] and Savitzky-Golay (2nd derivative, 2nd polynomial order, 7/9 points) [SGd₂(7/9)]] in different combinations. PCA models were calculated, the data explored and outliers removed then DA- and PLS-DA models were recalculated and evaluated.

4.2.3.3 Exploratory data analysis (EDA)

4.2.3.3.1 Principal component analysis (PCA)

The score plots of PC1 vs. PC2 were used to investigate the difference between the treatments, as the data shows that the variation was best explained in the first component. Although the variance explained by PC1 worsened with the various pre-processing techniques, good (zebra, springbok) and minimal (ostrich) class separation were observed (**Addendum B, Figure B9 – B33**). The variance observed in the direction of PC2 can be attributed to the discrepancies within the species due to differences in macronutrient composition. This separation was predominantly due to differences in moisture, fat, protein and pH as illustrated in the loading plots (**Addendum B, Figure B9b – B33b**).

The PC1 score plots of the SNV treated data illustrated minimal overlap between the fresh and frozen-thawed treatments (**Addendum B, Figure B9a – B11a**) and the loading plots (**Addendum B, Figure B9b – B11b**) exhibited interpretable positive bands at 932, 1075 and 1453 nm and negative bands at 1193 and 1316 nm. These bands are associated with moisture (932 and 1453 nm)(Osborne *et al.*, 1993; ElMasry *et al.*, 2012; Kamruzzaman *et al.*, 2012b), fat (932 nm) (ElMasry *et al.*, 2012; Kamruzzaman *et al.*, 2012b) and pH (1075 and 1316 nm) (ElMasry *et al.*, 2012; Iqbal *et al.*, 2013). Although SNV removed the scatter effects, the separation was still insufficient as an overlap between treatments was observed (**Addendum B, Figure B9a – B11a**).

The score plots (PC1) of the SNV + SGd₂(7) and SNV + SGd₂(9) corrected data illustrated good separation between the treatments for the zebra and springbok models (**Addendum B, Figure B12a – B13a** and **Figure B15a – B16a**), while the score plot for ostrich illustrated an overlap (**Addendum B, Figure B14a, B17a**). The loading plots (**Addendum B, Figure B12b – B17b**) had interpretable bands at 970 nm (negative), 1075 nm (positive), 1118 nm (positive), 1155 nm (negative), 1323 nm (negative), 1366/1372 nm (negative) and 1403/1409 nm (positive). Thus suggesting that the separation between the treatments is a result of variance in the pH (ElMasry *et al.*, 2012; Iqbal *et al.*, 2013), moisture content (Osborne *et al.*, 1993; Cozzolino & Murray, 2004; Barbin *et al.*, 2013b; Ma *et al.*, 2015; Pu *et al.*, 2015) as well as the protein content (Barbin *et al.*, 2012b). The pre-processing improved the separation along PC1 for the zebra and springbok models, and though the separation in PC1 for ostrich was more prominent, an overlap between the treatments was still evident (**Addendum B, Figure B14a, B17a**). The improved separation between fresh and frozen-thawed can be ascribed to the combined pre-processing of SNV + Savitzky-Golay second derivative. The second derivative is a functional pre-processing technique and is used to amplify wavebands, in order to identify bands of importance that best describe the separation in the data (Esbensen *et al.*, 2002).

The PC1 score plot of the SNV + detrend + SGd₂(7) and SNV + detrend + SGd₂(9) treated data illustrated good separation between the treatments for the zebra and springbok models (**Addendum B, Figure B18a – B19a** and **Figure B21a – B22a**), while the score plot for ostrich illustrated an overlap (**Addendum B, Figure B20a**). The loading plots (**Addendum B, Figure B18b – B22b**) exhibited the same interpretable bands as previously discussed. Therefore, the separation is a result of variance in the pH,

moisture- and protein content. The combined pre-processing of SNV + detrend + Savitzky-Golay second derivative exhibited similar results as SNV + Savitzky-Golay second derivative, thus indicating that detrend pre-processing did not improve the separation.

The PCA results of the SGd₁(5) and SGd₁(7) treated data are shown in **Addendum B, Figure B23a – B28a**. These results suggested that Savitzky-Golay first derivative is an effective pre-processing technique as it illustrated good separation. The loading plot for both pre-processing techniques showed interpretable bands at 988 nm (positive), 1143 nm (negative), 1224 nm (negative), 1230 nm (positive), 1329 nm (negative), 1335 nm (positive), 1378 nm (negative), 1385 nm (negative) and 1539 nm (positive) (**Addendum B, Figure B23b – B28b**). The band at 988 nm is associated with fat content (Barbin *et al.*, 2013b). The band at 1143 nm is associated with the C-H stretch second overtone found in aromatic groups (Osborne *et al.*, 1993) and corresponds to the tenderness of meat (Kamruzzaman *et al.*, 2013) and the band at 1230 nm (ElMasry *et al.*, 2012). The bands at 1224 and 1329 nm are related to protein content (Barbin *et al.*, 2013a; ElMasry *et al.*, 2013). The bands at 1335 and 1539 nm can be associated with a variance in pH and moisture (Osborne *et al.*, 1993; ElMasry *et al.*, 2012; ElMasry *et al.*, 2013). Lastly, the absorption bands at 1378 and 1385 nm suggest that the separation between fresh and frozen-thawed meat is caused by the change in moisture content (ElMasry *et al.*, 2013).

The PC1 score plots of the SGd₂(7) and SGd₂(9) corrected data illustrated improved separation (**Addendum B, Figure B29a – B33a**), which was sufficient for treatment separation. The loading line plots (**Addendum B, Figure B29b – B33b**) had interpretable bands at 945, 970, 1124, 1155, 1193, 1292, 1304, 1341, 1366, 1397 and 1409 nm. As with the previous PCA results the separation between the fresh and frozen-thawed samples is due to differences in moisture (Osborne *et al.*, 1993; ElMasry *et al.*, 2011; Kamruzzaman *et al.*, 2012b; ElMasry *et al.*, 2013; Iqbal *et al.*, 2013), fat (Kamruzzaman *et al.*, 2012b), protein (Osborne *et al.*, 1993; Barbin *et al.*, 2013a; Iqbal *et al.*, 2013) and pH (Barbin *et al.*, 2012b; ElMasry *et al.*, 2012; Iqbal *et al.*, 2013).

The results from the various pre-processing techniques suggests that the Savitzky-Golay first- and second derivative transformations would be sufficient and best suited for fresh and frozen-thawed separation. The Savitzky-Golay first- and second derivative transformations enhanced the differences in macronutrient composition of the biological samples. This separation was predominantly due to differences in moisture, fat, protein and pH as illustrated in the loading plots (**Addendum B, Figure B9b – B33b**). Even though Savitzky-Golay derivatives showed the best PCA results for separation, all the pre-processing techniques were used for further data analysis and model evaluation.

4.2.3.4 Multivariate data analysis (MDA): Optimal model development

4.2.3.4.1 Discriminant analysis (DA)

The overall accuracy of the discriminant analysis models, illustrated that the springbok model provided the best discrimination. The springbok LDA models (5 PCs) for the various types of pre-processing gave similar results (**Table 4.42**). Although the LDA models, irrespective of the pre-processing, achieved identical results, when comparing the PCA score plots (**Addendum B, Figure B9 – B33**) it was observed that, when using SNV + SGd₂(7/9) and SNV + detrend + SGd₂(7/9), the separation between classes was more distinct with little overlap. The calibration model achieved an overall classification accuracy of 100% for SNV + SGd₂(7/9) and SNV + detrend + SGd₂(7/9) (**Table 4.43**). These models also exhibited excellent validation accuracies (100%), and this suggests that the models can accurately distinguish between the treatments. Therefore, SNV + SGd₂(7/9) and SNV + detrend + SGd₂(7/9) pre-processed data resulted in more robust models, and is most likely ascribed to the fact that SNV + detrend and Savitzky-Golay second derivative both eliminate noise and corrects for the baseline shift (Barnes *et al.*, 1989; Esbensen *et al.*, 2002). The derivatives were also calculated based on a smoothing method by removing unwanted features from the data, and consequently enhancing features that are advantageous to the primary analysis.

The LDA models, pre-processed with SNV + SGd₂(7/9) and SNV + detrend + SGd₂(7/9), achieved similar results to that of the original model pre-processed with SNV + detrend. To illustrate that there were minimal differences between these pre-processing techniques, the PCA score plots of SNV + SGd₂(7/9) (**Addendum B, Figure B13a, B16a**), SNV + detrend + SGd₂(7/9) (**Addendum B, Figure B19a, B22a**) and SNV + detrend (**Addendum B, Figure B1a**) are shown. The classification accuracy for differentiating between the fresh and frozen-thawed meat was 100% for all these pre-processing methods. The reason for the slight differences between the PCA results of these pre-processing techniques is most likely because the SNV + detrend and Savitzky-Golay algorithms essentially do the same thing. However, the derivatives are calculated with an additional smoothing effect that removes the unwanted features, and as a result enhances the features that are advantageous for the analysis (Savitzky & Golay, 1964). Therefore, the SNV + SGd₂(7/9) and SNV + detrend + SGd₂(7/9) pre-processing achieved better results as the separation due to the differences in moisture, fat, protein and pH were enhanced.

There are no substantial differences between the pre-processing techniques for the zebra models (**Table 4.42**). The QDA models pre-processed with SNV + SGd₂(7/9) and SGd₂(7/9) presented the best discrimination results. These results can be explained by investigating the performance measures (**Table 4.45**). The zebra QDA models (6 PCs) achieved an overall classification accuracy of 98.6% for SNV + SGd₂(7/9) and 97.9% for SGd₂(7/9) with misclassification rates <2.2% (**Table 4.44**). Although these models had calibration accuracies of approximately 98%, further investigation shows that the model with a 97.9% validation accuracy is optimal (**Table 4.4**). Therefore, the SNV + SGd₂(7) and SGd₂(7) pre-processed data resulted in models that were more robust. The performance measures also revealed that the SNV + SGd₂(7) pre-processed model achieved the best results and exhibited that the discrimination between fresh and

frozen-thawed zebra meat was nearly perfect. The classification accuracy (98.6%), sensitivity (98.6%) and specificity (98.6%) for both treatments confirmed this (**Table 4.5**).

The QDA model, pre-processed with SNV + SGd₂(7), achieved slightly better results than that of the original model pre-processed with SNV + detrend. The classification accuracy for differentiating between the treatments was 97.9% for SNV + detrend and 98.6% for SNV + SGd₂(7). The slight differences between these two pre-processing techniques is most likely because both combinations of algorithms remove the scatter effects and reduces the baseline shift and curvature in the spectroscopic data (Savitzky & Golay, 1964; Barnes *et al.*, 1989; Esbensen *et al.*, 2002). In addition, Savitzky-Golay second derivative amplifies important wavebands contributing to the separation (Savitzky & Golay, 1964; Barnes *et al.*, 1989). Therefore, the SNV + SGd₂(7) pre-processing achieved better results as the separation due to differences in moisture, protein and pH were enhanced.

There are substantial differences between the pre-processing techniques for the ostrich models (**Table 4.42**). The QDA models pre-processed with SNV + detrend + SGd₂(7) and SGd₁(5), presented the best discrimination results (**Table 4.46**). The ostrich QDA models (5 PCs) achieved an overall classification accuracy of 100% for SNV + detrend + SGd₂(7) and 96.7% for SGd₁(5) with misclassification rates of <3.3% (**Table 4.46**). The general results indicate that these models can accurately distinguish between the two treatments. The performance measures also revealed that the SNV + detrend + SGd₂(7) pre-processed model achieved the best results for ostrich and exhibited that the discrimination between the two treatment classes was perfect (**Table 4.47**).

The QDA model, pre-processed with SNV + detrend + SGd₂(7), achieved noticeably better results than that of the original model pre-processed with SNV + detrend. The classification accuracy was 96.7% for SNV + detrend and 100% for SNV + detrend + SGd₂(7). The difference between the two pre-processing techniques, as well as the model improvement, can be ascribed to the additional Savitzky-Golay second derivative transformation, as previously discussed. Therefore, the SNV + detrend + SGd₂(7) pre-processing achieved better results as the separation due to differences in moisture, protein and pH were enhanced.

Table 4.42 An overview of the classification accuracies of the DA models with various pre-processing techniques applied for fresh or previously frozen meat discrimination.

Pre-processing	Species / Model / Number of principal components (PCs)					
	Zebra		Springbok		Ostrich	
	QDA (6 PCs)	LDA (5 PCs)	QDA (5 PCs)	Misclassification rate (%)	Classification accuracy (%)	Misclassification rate (%)
SNV	96.5	3.5	98.9	1.1	90.0	10.0
SNV+SGd ₂ (7)	98.6	1.4	100.0	0	86.7	13.3
SNV+SGd ₂ (9)	98.6	1.4	100.0	0	86.7	13.3
SNV+detrend+SGd ₂ (7)	97.2	2.8	100.0	0	100.0	0
SNV+detrend+SGd ₂ (9)	97.2	2.8	100.0	0	-	-
SGd ₁ (5)	95.8	4.2	100.0	0	96.7	3.3
SGd ₁ (7)	97.2	2.8	100.0	0	93.3	6.7
SGd ₂ (7)	97.9	2.1	100.0	0	93.3	6.7
SGd ₂ (9)	97.9	2.1	100.0	0	-	-

(DA) Discriminant analysis, (QDA) Quadratic discriminant analysis; (LDA) Linear discriminant analysis; (PCs) Principal components; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

Table 4.43 Springbok LDA model calibration and validation results to assess the overall performance of the SNV + SGd₂(7/9) and SNV + detrend + SGd₂(7/9) corrected data for fresh or previously frozen meat discrimination.

Pre-processing	Number of principal components (PCs)	Model	Classification accuracy (%)	Misclassification rate (%)
SNV+SGd ₂ (7)	5	Calibration	100.0	0
SNV+SGd ₂ (9)		Validation	100.0	0
SNV+detrend+SGd ₂ (7)	5	Calibration	100.0	0
SNV+detrend+SGd ₂ (9)		Validation	100.0	0

(LDA) Linear discriminant analysis; (PCs) Principal components; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points).

Table 4.44 Zebra QDA model calibration and validation results to assess the overall performance of the SNV + SGd₂(7/9) and SGd₂(7/9) corrected data for fresh or previously frozen meat discrimination.

Pre-processing	Number of principal components (PCs)	Model	Classification accuracy (%)	Misclassification rate (%)
SNV+SGd ₂ (7)	6	Calibration	98.6	1.4
		Validation	97.9	2.1
SNV+SGd ₂ (9)	6	Calibration	98.6	1.4
		Validation	95.8	4.2
SGd ₂ (7)	6	Calibration	97.9	2.1
		Validation	97.9	2.1
SGd ₂ (9)	6	Calibration	97.2	2.1
		Validation	95.8	4.2

(QDA) Quadratic discriminant analysis; (PCs) Principal components; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points).

Table 4.45 The performance measures used to assess the optimal QDA models (6 PCs) for fresh vs. frozen-thawed zebra meat.

Pre-processing	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
SNV+SGd₂(7)								
Fresh	98.6	0.7	0.7	98.6	98.6	98.6	98.6	1.4
FT	98.6	0.7	0.7	98.6	98.6	98.6	98.6	1.4
SGd₂(7)								
Fresh	97.9	1.4	0.7	98.7	98.5	97.4	98.0	2.1
FT	97.9	0.7	1.4	97.1	97.4	98.5	97.8	2.1

(QDA) Quadratic discriminant analysis; (PCs) Principal components; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); (FT) Frozen-thawed.

Table 4.46 Ostrich QDA model calibration and validation results to assess the overall performance of the SNV + detrend + SGd₂(7) and SGd₁(5) corrected data for fresh or previously frozen meat discrimination.

Pre-processing	Number of principal components (PCs)	Model	Classification accuracy (%)	Misclassification rate (%)
SNV+detrend+SGd ₂ (7)	5	Calibration	100.0	0
		Validation	90.0	10.0
SGd ₁ (5)	5	Calibration	96.7	3.3
		Validation	90.0	10.0

(QDA) Quadratic discriminant analysis; (PCs) Principal components; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points).

Table 4.47 The performance measures used to assess the optimal QDA models (6 PCs) for fresh vs. frozen-thawed ostrich meat.

Pre-processing	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
SNV+detrend+SGd ₂ (7)								
Fresh	100.0	0	0	100.0	100.0	100.0	100.0	0
FT	100.0	0	0	100.0	100.0	100.0	100.0	0
SGd ₁ (5)								
Fresh	96.7	0	3.3	92.9	94.1	100.0	96.3	3.3
FT	96.7	3.3	0	100.0	100.0	94.1	97.0	3.3

(QDA) Quadratic discriminant analysis; (PCs) Principal components; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); (FT) Frozen-thawed.

4.2.3.4.2 Partial least squares discriminant analysis (PLS-DA)

The springbok PLS-DA models for the various types of pre-processing gave similar results. Although the PLS-DA models, irrespective of the pre-processing, achieved identical results, when comparing the PCA score plots (**Addendum B, Figure B9 – B33**) it was observed that, when using SNV + SGd₂(7/9) and SNV + detrend + SGd₂(7/9), the separation between classes was more distinct with little overlap. These results can be explained by investigating the performance measures (not shown). The calibration model achieved an overall classification accuracy of 100% for SNV + SGd₂(7/9) and SNV + detrend + SGd₂(7/9) (**Table 4.48**). These models also exhibited excellent validation accuracies (100%), and this suggests that the models can accurately distinguish between the treatments. Therefore, SNV + SGd₂(7/9) and SNV + detrend + SGd₂(7/9) pre-processed data resulted in more robust models.

Although the PLS-DA models, pre-processed with SNV + SGd₂(7/9) and SNV + detrend + SGd₂(7/9), achieved similar results to that of the original model pre-processed with SNV + detrend (100%), the separation in the PCA score plots were improved for the SNV + SGd₂(7/9) (**Addendum B, Figure B13a, B16a**) and SNV + detrend + SGd₂(7/9) (**Addendum B, Figure B19a, B22a**) treated data. Therefore, the SNV + SGd₂(7/9) and SNV + detrend + SGd₂(7/9) pre-processing achieved better PCA results as the separation due to differences in moisture, protein and pH were enhanced.

There are minimal differences between the pre-processing techniques for the zebra models (**Table 4.48**). The PLS-DA models pre-processed with SNV + SGd₂(9), SNV + detrend + SGd₂(7) and SGd₁(5), presented the best discrimination results. These results can be explained by investigating the overall model performance (**Table 4.49**). The zebra PLS-DA models achieved an overall classification accuracy of 99.3% for SNV + SGd₂(9), 100% for SNV + detrend + SGd₂(7) and 99.3% for SGd₁(5) with misclassification rates <0.7% (**Table 4.48**). Although these models had calibration accuracies of approximately 99%, further investigation shows that the model with a 100% validation accuracy is optimal (**Table 4.48**). Even though the SNV + SGd₂(9) and SNV + detrend + SGd₂(7) models exhibited classification accuracies of >99% (calibration) and >97% (cross-validation), indicative of an effective model, the validation accuracy (97%) decreased to some extent. This indicated that these models were slightly over-fitted (Miller, 2005) and therefore less effective as suggested by the calibration- and cross-validation accuracies. Therefore, the SGd₁(5) pre-processed data resulted in models that were more robust, suggesting that this model can accurately distinguish between the two treatments. The performance measures also revealed that the SGd₁(5) pre-processed model achieved excellent results and exhibited that the discrimination between fresh and frozen-thawed zebra meat was nearly perfect (99.3%). The model exhibited a higher affinity to correctly predict the previously frozen samples, as the classification accuracy (99.3%), sensitivity (100%) and specificity (100%) for the frozen-thawed treatment was higher than that of the fresh class (**Table 4.49**).

The PLS-DA model, pre-processed with SGd₁(5), achieved considerably better results than that of the original model pre-processed with SNV + detrend. To illustrate the differences between these two pre-processing techniques, the PCA score plots of the SGd₁(5) (**Addendum B, Figure B24a**) and SNV + detrend (**Figure 4.34a**) corrected data are shown. The classification accuracy for differentiating between the treatments was 95.1% for SNV + detrend and 99.3% for SGd₁(5). The SGd₁(5) pre-processing achieved better results most likely because the Savitzky-Golay transformation improved the separation by enhancing the differences in moisture and pH.

The ostrich PLS-DA models for the various types of pre-processing gave different results. The overall accuracy of the PLS-DA models suggested that those pre-processed with SNV + detrend + SGd₂(7), SGd₁(5) and SGd₁(7) would provide the best discrimination. The PLS-DA calibration models achieved an overall classification accuracy of 93.3% for SNV + detrend + SGd₂(7), 90% for SGd₁(5) and 90% for SGd₁(7) with misclassification rates <10% (**Table 4.48**). Even though the SNV + detrend + SGd₂(7) model had a higher calibration accuracy (93.3%) and good cross-validation (86.7%) results, the validation accuracy (80%) was

considerably lower and can be indicative of a model that has been overfitted (Miller, 2005). Therefore, the SGd₁(5) and SGd₁(7) pre-processed data resulted in more robust models and is less affected by variation of intrinsic parameters in meat. This suggests that the model can accurately distinguish between the treatments and these results can be explained by investigating the performance measures (**Table 4.50**). Both models exhibited a 90% classification accuracy for both treatments, however, further investigation of the performance measures revealed that the probability of accurately predicting fresh was higher as confirmed by the sensitivity (92.9%) and specificity (93.3%) in **Table 4.50**. Although both the SGd₁(5) and SGd₁(7) models exhibited calibration and validation accuracies of 90%, indicative of a model with a good fit (Miller, 2005), further inspection of the cross-validation results illustrated that SGd₁(5) (86.7%) achieved a slightly higher cross-validation accuracy than SGd₁(7) (80%). This suggests that the SGd₁(5) model is slightly more effective for differentiating between the two treatments. These decreased accuracies suggest that the training set comprised of samples that were more complex with regards to chemical composition. The validation set however had samples that were simpler with less variation and therefore the prediction accuracy of the model was compromised. This phenomenon is one drawback of the leave-one-out cross-validation method as the results for the sub-validations can be overly pessimistic, where ‘edge’ samples were excluded from the calibration set (Miller, 2005). These results are even more pessimistic if any of the ‘edge’ samples are very unique in their responses.

The PLS-DA model, pre-processed with SGd₁(5), resulted in a similar classification accuracy compared to the original model pre-processed with SNV + detrend (90%). Although the SNV + detrend model had similar accuracies (**Table 4.39**) and individual performance measures (**Table 4.40**), the CV (83.3%) and validation (50%) results suggests that the model has a good predictive power for the training set, but not for the independent validation set. Hence, indicating that the calibration model has been overfitted (Miller, 2005). Therefore, the PLS-DA model pre-processed with SGd₁(5) is considered to be better than the original SNV + detrend model. The SGd₁(5) model achieved the same classification (90%) and similar CV (86.7%) results, but the independent validation set achieved a higher accuracy (90%). This indicates that the model is more robust and not overfitted to the training set. Although both algorithms essentially do the same thing, the SGd₁(5) pre-processing achieved better results most likely because the Savitzky-Golay transformation improved the treatment separation by enhancing the differences in moisture, fat, protein and pH.

Table 4.48 An overview of the classification accuracies of the PLS-DA models with various pre-processing techniques applied for fresh vs. frozen-thawed meat.

Pre-processing	Species									
	Zebra			Springbok			Ostrich			Validation accuracy (%)
	Calibration accuracy (%)	CV (%)	Validation accuracy (%)	Classification accuracy (%)	CV (%)	Validation accuracy (%)	Classification accuracy (%)	CV (%)	Validation accuracy (%)	
Pre-processing	Misclassification rate (%)			Misclassification rate (%)			Misclassification rate (%)			
SNV	97.2 2.8	97.2	95.8	100.0 0	100.0	76.8	79.7 23.3	76.7	70.0	
SNV+SGd ₂ (7)	97.9 2.1	97.9	93.8	100.0 0	100.0	100.0	80.0 20.0	76.7	70.0	
SNV+SGd ₂ (9)	99.3 0.7	98.6	97.9	100.0 0	100.0	100.0	80.0 20.0	76.7	70.0	
SNV+detrend+SGd ₂ (7)	100.0 0	97.9	97.9	100.0 0	100.0	100.0	93.3 6.7	86.7	80.0	
SNV+detrend+SGd ₂ (9)	99.3 0.7	97.9	95.8	100.0 0	100.0	100.0	- -	-	-	
SGd ₁ (5)	99.3 0.7	96.5	100.0	100.0 0	100.0	100.0	90.0 10.0	86.7	90.0	
SGd ₁ (7)	97.2 2.8	95.8	97.9	100.0 0	100.0	100.0	90.0 10.0	80.0	90.0	
SGd ₂ (7)	97.2 2.8	97.2	97.9	100.0 0	100.0	100.0	86.7 13.3	86.7	90.0	
SGd ₂ (9)	95.8 4.2	95.8	97.9	100.0 0	100.0	100.0	- -	-	-	

(PLS-DA) Partial least squares discriminant analysis; (LVs) Latent variables; (CV) Cross-validation; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

Table 4.49 The performance measures used to assess the optimal PLS-DA (5LVs) model for fresh vs. frozen-thawed zebra meat.

Pre-processing	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
SGd ₁ (5)								
Fresh	99.3	0	0.7	98.7	98.6	100.0	98.0	0.7
FT	99.3	0.7	0	100.0	100.0	98.6	97.8	0.7

(PLS-DA) Partial least squares discriminant analysis; (LVs) Latent variables; (SNV) Standard normal variate; [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); (FT) Frozen-thawed.

Table 4.50 The performance measures used to assess the optimal PLS-DA (3LVs) model for fresh vs. frozen-thawed ostrich meat.

Pre-processing	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
SGd ₁ (5)								
Fresh	90.0	6.7	3.3	92.9	93.3	86.7	89.7	10.0
FT	90.0	3.3	6.7	87.5	86.7	93.3	90.3	10.0
SGd ₁ (7)								
Fresh	90.0	6.7	3.3	92.9	93.3	86.7	89.7	10.0
FT	90.0	3.3	6.7	87.5	86.7	93.3	90.3	10.0

(PLS-DA) Partial least squares discriminant analysis; (LVs) Latent variables; (SNV) Standard normal variate; [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points); (FT) Frozen-thawed.

4.2.3.5 Optimal model selection

After the models for the various types of pre-processing were calculated, the models with the best results were compared to identify the optimal one (**Table 4.51**). It was concluded that the PLS-DA models, pre-processed with Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points) was the best for differentiating between fresh and frozen-thawed zebra and ostrich meat, irrespective of the muscle type. The results for the springbok models (LDA and PLS-DA) suggest that all the pre-processing methods would be sufficient for differentiating between fresh and previously frozen meat. However, the PCA score plots for the SNV + detrend (**Addendum B, Figure B1a**) and SNV + detrend + SG₂(9) (**Addendum B, Figure B22a**) pre-processed data illustrates that these methods improved the separation between treatments along PC1. Hence, making

it possible to conclude that these two pre-processing techniques would be the best for enhancing the separation between fresh and frozen-thawed samples. The PLS-DA models achieved better results as it generally outperforms LDA and QDA, when classes are closely related, because it overcomes the collinearity problems often associated with LDA and QDA. SGd₁(5) outperformed the combined pre-processing of SNV + detrend for the zebra and ostrich, as the Savitzky-Golay transformation was found to have enhanced the differences in moisture, protein and pH, which was predominantly the contributors for fresh and frozen-thawed separation.

Table 4.51 An overview of the classification accuracies of the DA and PLS-DA models with various pre-processing techniques applied for fresh or previously frozen meat discrimination.

	Model	Pre-processing	Number of PCs or LVs	Calibration		Validation
				Classification accuracy (%)	Misclassification rate (%)	
Zebra	QDA	SNV + detrend	6	97.9	2.1	97.9
	PLS-DA	SNV + detrend	3	95.1	4.9	100.0
	QDA	SNV + SGd ₂ (7)	6	98.6	1.4	97.9
	PLS-DA	SGd ₁ (5)	5	99.3	0.7	100.0
Springbok	LDA	SNV + detrend	5	100.0	0	100.0
	PLS-DA	SNV + detrend	2	100.0	0	100.0
	LDA	SNV + SGd ₂ (7)	5	100.0	0	100.0
		SNV + SGd ₂ (9)				
		SNV+DT+SGd ₂ (7)	5	100.0	0	100.0
		SNV+DT+SGd ₂ (9)				
	PLS-DA	SNV + SGd ₂ (7)	2	100.0	0	100.0
		SNV + SGd ₂ (9)				
Ostrich	QDA	SNV + detrend	5	96.7	3.3	70.0
	PLS-DA	SNV + detrend	1	90.0	10.0	50.0
	QDA	SNV+DT+ SGd ₂ (7)	5	100.0	0	90.0
	PLS-DA	SGd ₁ (5)	3	90.0	10.0	90.0

(DA) Discriminant analysis; (QDA) Quadratic discriminant analysis; (LDA) Linear discriminant analysis; (PLS-DA) Partial least squares discriminant analysis; (PCs) Principal components; (LVs) Latent variables; (SNV) Standard normal variate; (DT) Detrend; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points).

4.2.3.6 Conclusion

NIR spectroscopy combined with MDA could accurately distinguish between the fresh and previously frozen meat samples for zebra, springbok and ostrich, irrespective of the muscle type. The PLS-DA discrimination models yielded the best results and could effectively distinguish the fresh and previously frozen springbok meat from one another with an accuracy of 100%. The zebra and ostrich models could also distinguish between the fresh and previously frozen meat and yielded accuracies of 99.3 and 90%, respectively. In addition, it was found that the muscle type does not influence the accuracy of the model for fresh or previously frozen meat discrimination.

4.2.4 Fresh vs. previously frozen meat determination as well as frozen period

This data set consisted of four species [black wildebeest (*Connochaetes gnou*), zebra (*Equus quagga burchelli*), springbok (*Antidorcas marsupialis*), ostrich (*Struthio camelus*)], one muscle type for each species and the meat samples were frozen for a period of nine months. The aim of **Section 4.2.4** was to differentiate between the fresh and previously frozen meat samples for all four game species, irrespective of the frozen period (1 – 9 months), as well as to determine whether the different frozen periods had an effect on the classification accuracies. The aim was also to determine whether it is possible to predict the different frozen periods.

4.2.4.1 Multivariate data analysis (MDA): Model development

4.2.4.1.1 Soft independent modelling of class analogy (SIMCA)

Table 4.52 SIMCA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen period classification.

Species	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
BWB	Calibration	45.2	54.8
	Validation	40.9	59.1
Zebra	Calibration	67.9	32.1
	Validation	64.4	35.6
Springbok	Calibration	44.6	55.4
	Validation	36.2	63.8
Ostrich	Calibration	68.4	31.6
	Validation	69.3	30.7

(SIMCA) Soft independent modelling of class analogy; (SNV) Standard normal variate; (BWB) Black wildebeest.

The classification accuracies shown in **Table 4.52** provides a general overview of the models' performance. The table also illustrates the misclassification of meat samples subjected to different frozen periods (1 – 9 months). Freezing is known to change the physical structure of meat (Prieto *et al.*, 2009). Ice

crystal formation occurs within the cytoplasm of the cells and over time they enlarge. This can lead to the ice crystals rupturing the cell membranes and upon thawing, cytoplasm and moisture can leach out of the cells resulting in a lower moisture content and compromised cell integrity (Leygonie *et al.*, 2012b). Prolonged frozen storage does not seem to have a further effect on these physical and chemical changes. This phenomenon can be confirmed by studying the PCA score plots (**Figure 4.35a** and **Addendum B, Figure B3 – B5**) of the different freeze-thaw treatments. These figures exhibit a major overlap between the freeze-thaw treatments, thus resulting in a misclassification between classes. The lack of separation can be ascribed to the sample's spectral similarities, as no substantial changes between frozen periods were observed. Therefore, it can be concluded that the specific changes occurring during freezing after one month did not increase with prolonged frozen storage under the same conditions. For this reason, the SIMCA model could not accurately distinguish between the different frozen periods, as the algorithm aims to classify samples based on spectral similarities (Wold & Sjöström, 1977). Downey and Beauchêne (1997b) reported similar results when attempting to discriminate between different freeze-thaw cycles. When evaluating the results in the current study alongside those reported in literature, it is possible to conclude that the differences in the spectra for fresh and frozen-thawed meat can be attributed to the initial damage caused by freezing. Any additional freeze-thaw cycles or increased period of freezing does not account for any further changes, as the damage done to the microstructures within meat had already been achieved to a very large extent.

Given the minimal accurate classification achieved between the frozen periods, it was deemed important to examine the success of a two-group classification, as such a model might be closer to actual commercial requirements. SIMCA models were developed to distinguish between fresh and frozen-thawed meat, irrespective of the frozen period.

Table 4.53 SIMCA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen-thawed meat classification, irrespective of frozen period.

Species	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
BWB	Calibration	75.8	24.2
	Validation	57.3	42.7
Zebra	Calibration	71.9	28.1
	Validation	57.8	42.2
Springbok	Calibration	90.0	10.0
	Validation	72.5	27.5
Ostrich	Calibration	80.4	19.6
	Validation	72.8	27.2

(SIMCA) Soft independent modelling of class analogy; (SNV) Standard normal variate; (BWB) Black wildebeest.

The overall SIMCA model accuracy of black wildebeest (75.8%), zebra (71.9%), springbok (90%) and ostrich (80.4%) improved, illustrating that the springbok and ostrich models gave good results. The SIMCA

model for springbok presented the best classification results. The calibration model achieved an overall classification accuracy of 90% (**Table 4.53**). Although both treatments for the springbok model achieved classification accuracies of 90%, the model exhibited a higher probability of predicting for frozen-thawed. This is confirmed by the higher sensitivity (100%), specificity (100%) and F1 score (91.6%) (**Table 4.54**). The same trend was observed for the ostrich model (80.4%). The sensitivity, specificity and F1 score for frozen-thawed were 87.8%, 81.1% and 83.7%, respectively. These results confirm the models' ability to classify the frozen-thawed meat samples with little fault. The PCA score plots and loading line plots (**Addendum B, Figure B4 – B5**) support these results as a good separation was observed between the two treatments (fresh vs. frozen-thawed). The separation and correct classification of the two classes was mainly attributed to the loss of moisture and change in pH. These properties changed during freezing and a correlation was observed in the difference spectra (**Addendum B, Figure B7a – B8a**) and loadings (**Addendum B, Figure B4b – B5b**).

The SIMCA models for black wildebeest and zebra achieved lower classification accuracies. The low classification results confirmed what was seen in the PCA score plots (**Figure 4.35** and **Addendum B, Figure B3a**). The lack of separation and misclassification can be ascribed to the samples' spectral similarities. When comparing these model results it is observed that the black wildebeest model achieved a slightly higher classification accuracy (75.8%) than the zebra model (71.9%), and also exhibited improved performance measures (**Table 4.54**). These results are supported and can be explained by the difference spectra (**Figure 4.37a** and **Addendum B, Figure B6**). Although the separation between fresh and frozen-thawed meat, for both species, can mainly be attributed to the loss of moisture and change in pH, the intensity of these absorption bands differed. The difference spectra for black wildebeest (**Addendum B, Figure B6**) exhibited absorption bands, attributed to moisture and pH, with higher intensities compared to the bands exhibited in the difference spectra for zebra (**Figure 4.37a**). The higher absorption bands for black wildebeest illustrates that there was a larger difference between the fresh and frozen-thawed meat, therefore achieving better class separation. On the other hand, the lower absorption bands for zebra suggests that the spectra are similar because of little differences between the fresh and frozen-thawed meat. Therefore, the SIMCA model for zebra achieved a lower classification accuracy.

Even though all four species exhibited calibration accuracies between 71.9 and 90%, which are indicative of effective models, the validation accuracies decreased (57.3 – 72.8%). This indicates that the models were over-fitted (Miller, 2005) and therefore not as effective as suggested by the calibration accuracies. The results in the current study illustrates that SIMCA models were not effective for predicting the frozen periods, due to the spectral similarities. However, the models were able to distinguish between fresh and frozen-thawed meat, irrespective of the frozen period, with relatively high accuracies. Downey and Beauchêne (1997b) reported similar results, also concluding that a two-group classification improved the SIMCA model accuracies.

Table 4.54 The performance measures used to assess the SIMCA classification models for fresh vs. frozen-thawed meat of the four species, pre-processed with SNV + detrend.

Species	Treatment	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
BWB	Fresh	75.8	1.2	23.0	55.4	67.3	95.8	70.2	24.2
	FT	75.8	23.0	1.2	97.4	95.8	67.3	79.6	24.2
Zebra	Fresh	71.9	11.7	16.4	65.4	71.4	72.6	68.8	28.1
	FT	71.9	16.4	11.7	77.8	72.6	71.4	74.5	28.1
Springbok	Fresh	90.0	0	10.0	78.0	84.5	100.0	87.6	10.0
	FT	90.0	10.0	0	100.0	100.0	84.5	91.6	10.0
Ostrich	Fresh	80.4	7.0	12.6	70.5	80.0	81.1	75.4	19.6
	FT	80.4	12.6	7.0	87.8	81.1	80.0	83.7	19.6

(SIMCA) Soft independent modelling of class analogy; (SNV) Standard normal variate; (BWB) Black wildebeest; (FT) Frozen-thawed.

4.2.4.1.2 K-nearest neighbour (KNN)

The classification accuracies shown in **Table 4.55** provides a general overview of the optimal models' performance. The overall model accuracy for the four species suggested that three nearest neighbours would provide the best classification for black wildebeest (50.8%) and zebra (45.2%) frozen period prediction. While two nearest neighbours provided the best classification for springbok (55.8%) and ostrich (54.2%). The table also illustrates the misclassification of meat samples subjected to different frozen periods (1 – 9 months). The misclassification of the treatments can be explained by looking at the PCA score plots (**Figure 4.36**) of the different frozen periods. These figures exhibit a major overlap between the freeze-thaw treatments (as previously explained), thus resulting in a misclassification between classes which might be considered as nearest neighbours. When evaluating the accuracies of the models, it is evident that the models were under-fitted, resulting in improved validation accuracies (Miller, 2005). This suggests that these models are not effective when attempting to classify the different frozen periods. For this reason, it was deemed important to examine the success of a two-group classification model. KNN models were developed to discriminate between fresh and frozen-thawed meat, irrespective of the frozen period, as such a model was considered to be closer to commercial requirements.

Table 4.55 Optimal KNN model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen period classification.

Species	Number of neighbours (<i>k</i>)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
BWB	3	Calibration	50.8	49.2
		Validation	81.5	18.5
Zebra	3	Calibration	45.2	54.8
		Validation	70.4	29.6
Springbok	2	Calibration	55.8	44.2
		Validation	83.3	16.7
Ostrich	2	Calibration	54.2	45.8
		Validation	76.7	23.3

(KNN) *K*-nearest neighbours; (SNV) Standard normal variate; (BWB) Black wildebeest.

The overall KNN model accuracies (**Table 4.56** and **Addendum B, Table B3**) improved, suggesting that three nearest neighbours would provide the best classification for the black wildebeest (94.4%), zebra (84.1%), springbok (94.2%) and ostrich (78.3%) models. The KNN(3) calibration model for black wildebeest achieved an overall classification- and cross-validation accuracy of 94.4%, with an improved classification accuracy for the validation (96.3%) (**Table 4.56**). The excellent cross-validation- and improved validation accuracies are indicative of a classification model that is not over-fitted (Miller, 2005). Although this model had excellent accuracies, a few misclassified objects were observed amongst the fresh and frozen-thawed meat samples (**Figure 4.42**). This accounted for the low misclassification rate of 5.6% for the overall model

(Table 4.56). The sensitivity and specificity for the fresh samples was 97.8% and 98.7%, respectively (Table 4.57). This indicates that the model has a high probability of correctly classifying the fresh meat samples. The lower sensitivity (92.5%) and specificity (88.2%) for the frozen-thawed samples reveals that the model was less suited for predicting this class. This phenomenon can be explained by referring to the PCA score plot (**Figure 4.42**), as k -nearest neighbour was performed on the PC scores. The KS-calibration PCA score plot illustrates a slight overlap between the two classes, with a larger number of frozen-thawed samples displaying a close distance to the fresh samples. Therefore, the frozen-thawed samples are assigned to the predominant class, fresh. Hence, explaining why the model is less suited for predicting the frozen-thawed samples.

Table 4.56 Optimal KNN model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen-thawed meat classification, irrespective of frozen period.

Species	Number of neighbours (k)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
BWB	3	Calibration	94.4	5.6
		Cross-validation	94.4	5.6
		Validation	96.3	3.7
Zebra	3	Calibration	84.1	15.9
		Cross-validation	84.1	15.9
		Validation	88.9	11.1
Springbok	3	Calibration	94.2	5.8
		Cross-validation	94.2	5.8
		Validation	98.3	1.7
Ostrich	3	Calibration	78.3	21.7
		Cross-validation	79.2	20.8
		Validation	91.7	8.3

(KNN) K -nearest neighbours; (SNV) Standard normal variate; (BWB) Black wildebeest.

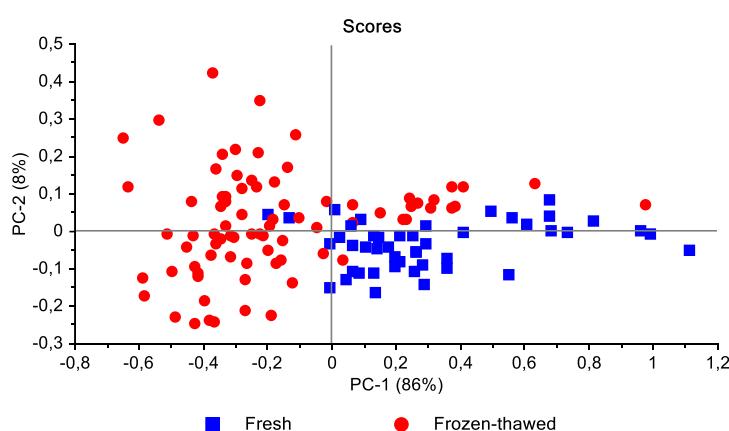


Figure 4.42 PCA analysis (SNV + detrend pre-processed) of black wildebeest (LTL) [frozen up to 9 months] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as KS-calibration PCA score plot of PC1 (86%) vs. PC2 (8%).

The KNN(3) calibration model for springbok achieved similar results, exhibiting an overall classification- and cross-validation accuracy of 94.2%, with an improved classification accuracy for the validation (98.3%) (**Table 4.56**). This model also exhibited a few misclassified (5.8%) objects amongst the fresh and frozen-thawed meat samples (**Figure 4.43**). The performance measures in **Table 4.57** indicate that the model has a high probability of correctly classifying the fresh meat samples, hence suggesting that the model was less suited for predicting the frozen-thawed class. This phenomenon can be explained by referring to the PCA score plot (**Figure 4.43**), as discussed previously.

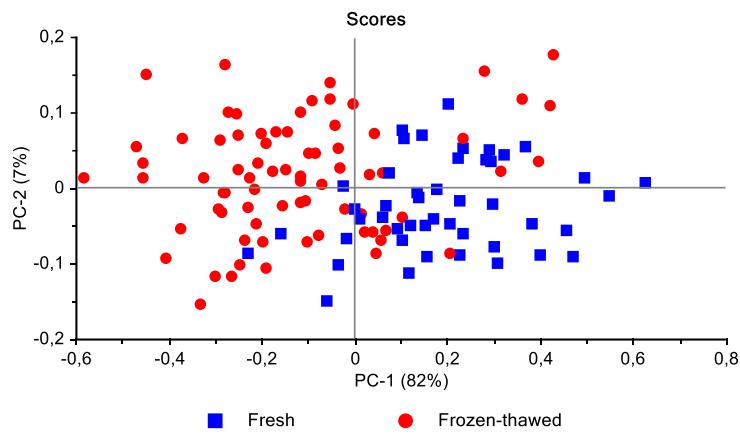


Figure 4.43 PCA analysis (SNV + detrend pre-processed) of springbok (LTL) [frozen up to 9 months] illustrating good separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as KS-calibration PCA score plot of PC1 (82%) vs. PC2 (7%).

The decreased KNN(3) model accuracy for both zebra and ostrich is mainly attributed to the misclassification of objects assigned to the fresh and frozen-thawed classes. These classes were misclassified due to the extensive overlap observed in the PCA score plots (**Figure 4.44a – b**). The sensitivity of both fresh and frozen-thawed samples in KNN(3) for zebra is 81.5% and 86.1%, respectively. While the specificity is 86.1% and 81.5%, therefore confirming the model's ability to accurately predict the frozen-thawed samples even though it struggles with the fresh samples. Once again this can be ascribed to the larger number of fresh samples displaying a close distance to the frozen-thawed samples (**Figure 4.44a**), as previously discussed. A similar phenomenon is observed in KNN(3) for ostrich, except this model had a higher affinity to accurately predict the fresh samples, where the sensitivity is 82.6% (fresh) and 75.7% (frozen-thawed), and the specificity is 87.5% (fresh) and 67.9% (frozen-thawed) (**Table 4.57**). Although the classification model exhibited good accuracies for both the calibration (78.3%) and cross-validation (79.2%), the model achieved a validation accuracy of 91.7%, suggesting that the model is under-fitted (Miller, 2005).

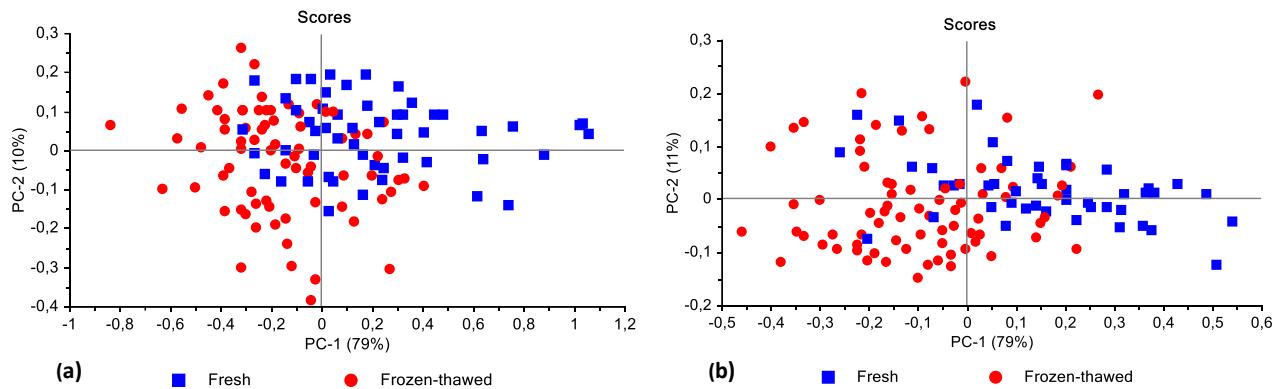


Figure 4.44 PCA analysis (SNV + detrend pre-processed) of (a) zebra (LTL) and (b) ostrich (FF) [frozen up to 9 months] illustrating minimal separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) KS-calibration PCA score plot of PC1 (79%) vs. PC2 (10%) for zebra, and (b) KS-calibration PCA score plot of PC1 (79%) vs. PC2 (11%) for ostrich.

The KNN model for black wildebeest (94.4%) presented the best results, followed by springbok (94.2%), zebra (84.1%) and ostrich (78.3%). These results are comparable to the SIMCA results, where similar findings were observed. Therefore, the improved classification of fresh and frozen-thawed black wildebeest and springbok meat can be attributed to a greater degree of physical and chemical changes that occurred within the meat during freezing. The KNN results for black wildebeest (94.4%) and springbok (94.2%) are comparable to the results reported by Thyholt and Isaksson (1997). These researchers classified fresh, frozen and re-frozen beef with a 93% accuracy. The data also gave promising separation between the re-frozen meat samples (Thyholt & Isaksson, 1997). The results in the current study illustrates that KNN models were not effective for predicting the frozen periods, due to the major overlap of the frozen treatments. However, the models were able to distinguish between fresh and frozen-thawed meat, irrespective of the frozen period, with relatively high accuracies. When considering these results alongside those reported in literature, it is possible to conclude that a two-group classification would result in improved KNN model accuracies.

Table 4.57 The performance measures used to assess the KNN classification models for fresh vs. frozen-thawed meat of the four species, pre-processed with SNV + detrend.

Species	Number of neighbours (<i>k</i>)	Treatment	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
BWB	3	Fresh	94.4	4.8	0.8	97.8	98.7	88.2	92.8	5.6
		FT	94.4	0.8	4.8	92.5	88.2	98.7	95.5	5.6
Zebra	3	Fresh	84.1	7.9	7.9	81.5	86.1	81.5	81.5	15.9
		FT	84.1	7.9	7.9	86.1	81.5	86.1	86.1	15.9
Springbok	3	Fresh	94.2	4.2	1.7	95.7	97.2	89.8	92.6	5.8
		FT	94.2	1.7	4.2	93.2	89.8	97.2	95.2	5.8
Ostrich	3	Fresh	78.3	15.0	6.7	82.6	87.5	67.9	74.5	21.7
		FT	78.3	6.7	15.0	75.7	67.9	87.5	81.2	21.7

(KNN) *K*-nearest neighbours; (SNV) Standard normal variate; (BWB) Black wildebeest; (FT) Frozen-thawed.

4.2.4.1.3 Discriminant analysis (DA)

Table 4.58 Optimal DA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen period classification.

Species	Number of principal components (PCs)	Method (distance calculations)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
Black wildebeest	6	Linear	Calibration	67.5	32.5
			Validation	83.3	16.7
		Quadratic	Calibration	90.5	9.5
			Validation	85.2	14.8
			Calibration	92.9	7.1
			Validation	83.3	16.7
	6	Mahalanobis	Calibration	76.2	23.8
			Validation	72.2	27.8
		Quadratic	Calibration	93.7	6.3
			Validation	69.8	30.2
			Calibration	96.0	4.0
			Validation	70.4	29.6
Zebra	6	Linear	Calibration	83.3	16.7
			Validation	83.3	16.7
		Quadratic	Calibration	94.2	5.8
			Validation	81.7	18.3
			Calibration	94.2	5.8
			Validation	80.0	20.0
	6	Mahalanobis	Calibration	80.8	19.2
			Validation	78.3	21.7
		Quadratic	Calibration	94.2	5.8
			Validation	76.7	23.3
			Calibration	93.3	6.7
			Validation	76.7	23.3
Springbok	6	Linear	Calibration	83.3	16.7
			Validation	83.3	16.7
		Quadratic	Calibration	94.2	5.8
			Validation	81.7	18.3
			Calibration	94.2	5.8
			Validation	80.0	20.0
	6	Mahalanobis	Calibration	80.8	19.2
			Validation	78.3	21.7
		Quadratic	Calibration	94.2	5.8
			Validation	76.7	23.3
			Calibration	93.3	6.7
			Validation	76.7	23.3
Ostrich	6	Linear	Calibration	80.8	19.2
			Validation	78.3	21.7
		Quadratic	Calibration	94.2	5.8
			Validation	76.7	23.3
			Calibration	93.3	6.7
			Validation	76.7	23.3
	6	Mahalanobis	Calibration	80.8	19.2
			Validation	78.3	21.7
		Quadratic	Calibration	94.2	5.8
			Validation	76.7	23.3
			Calibration	93.3	6.7
			Validation	76.7	23.3

(DA) Discriminant analysis; (SNV) Standard normal variate.

The classification accuracies shown in **Table 4.58** provides a general overview of the optimal models' performance. The overall model accuracy for the four species suggested that Quadratic discriminant analysis (QDA) with six principal components would provide the best classification for black wildebeest (90.5%), zebra (93.7%), springbok (94.2%) and ostrich (94.2%) frozen period prediction. The table also illustrates the misclassification of meat samples subjected to different frozen periods (1 – 9 months). The misclassification of the treatments can be explained by studying the PCA score plots (**Figure 4.36**) of the different frozen periods. These figures exhibit a major overlap between the freeze-thaw treatments (as previously explained), thus resulting in a misclassification between the frozen periods. However, because the QDA algorithm calculates a non-linear decision boundary, using a quadratic function, it was able to separate the clusters and

successfully discriminate between the treatments. Although these models exhibited excellent overall calibration accuracies as well as good performance measures for each treatment (not shown), the decreased validation accuracies suggests that the classification models were all over-fitted (Miller, 2005). Consequently, the over-fitted model has a low probability of correctly predicting the different frozen periods when using an independent validation set. For this reason, it was deemed important to examine success of a two-group discrimination model. DA models were developed to discriminate between fresh and frozen-thawed meat, irrespective of the frozen period.

Table 4.59 Optimal DA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen-thawed meat classification, irrespective of frozen period.

Species	Number of principal components (PCs)	Method (distance calculations)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
Black wildebeest	6	Linear	Calibration	93.0	7.0
			Validation	89.1	10.9
		Quadratic	Calibration	96.7	3.3
			Validation	94.2	5.8
			Calibration	89.0	11.0
			Validation	76.0	24.0
	6	Linear	Calibration	91.0	9.0
			Validation	91.0	9.0
		Quadratic	Calibration	94.5	5.5
			Validation	89.1	10.9
			Calibration	93.7	6.3
Zebra	6	Mahalanobis	Validation	82.6	17.4
			Calibration	98.3	1.7
		Linear	Validation	100.0	0
			Calibration	99.2	0.8
			Validation	98.3	1.7
	6	Quadratic	Calibration	99.2	0.8
			Validation	99.2	0.8
		Mahalanobis	Calibration	98.3	1.7
			Validation	98.3	1.7
Springbok	6	Linear	Calibration	90.0	10.0
			Validation	91.7	8.3
		Quadratic	Calibration	95.8	4.2
			Validation	91.7	8.3
			Calibration	88.3	11.7
	6	Mahalanobis	Validation	81.7	18.3
			Calibration	99.2	0.8

(DA) Discriminant analysis; (SNV) Standard normal variate.

The Quadratic discriminant analysis (QDA) model for springbok (6 PCs) presented the best discrimination results. The separation between the fresh and frozen-thawed meat samples was achieved with a 99.2% overall accuracy (**Table 4.59**). The performance measures also revealed that the discrimination

of both the fresh and frozen-thawed meat samples were nearly all (99.2%) correctly predicted. The classification accuracy, sensitivity, specificity and F1 score for both classes were above 97% (**Table 4.60**), which means that the model can accurately distinguish between the two treatments, irrespective of the frozen period.

The QDA models for black wildebeest (6 PCs), zebra (6 PCs) and ostrich (6 PCs) achieved similar overall accuracies. The QDA model for black wildebeest achieved an overall classification accuracy of 96.7% and a misclassification rate of 3.3% (**Table 4.59**). The classification results for the validation model decreased, resulting in an overall classification accuracy of 94.2% and a misclassification rate of 5.8% (**Table 4.59**). The sensitivity and specificity for the fresh samples were both 100% (**Table 4.60**). This indicates that the model has a high probability of correctly classifying the fresh meat samples. While the lower sensitivity and specificity for the frozen-thawed samples reveals that the model was less suited for predicting this class. The frozen-thawed samples were misclassified due to the overlap observed in the PCA score plot (**Addendum B, Figure B3a**). The separation and correct classification of the two classes was mainly attributed to the loss of moisture and change in pH. These properties changed during freezing and the correlation was observed in the difference spectra (**Addendum B, Figure B6**) and loadings (**Addendum B, Figure B3b**). Thus, the lack of separation suggests that the samples have similar spectra, and therefore it is possible to conclude that freezing did not damage the microstructures within these meat samples to the same extent.

The decreased model accuracy of zebra (94.5%) and ostrich (95.8%) is mainly attributed to the misclassification of objects assigned to the fresh and frozen-thawed classes. Once again this can be ascribed to the larger number of fresh samples displaying a close distance to the frozen-thawed samples and vice versa, as previously discussed. These classes were misclassified due to the extensive overlap observed in the PCA score plots (**Figure 4.35a** and **Addendum B, Figure B5a**).

Although these models exhibited excellent overall calibration accuracies as well as good performance measures (**Table 4.60**), which are indicative of effective models, the validation accuracies decreased. This indicates that the models were slightly over-fitted (Miller, 2005) and therefore not as effective as suggested by the calibration accuracies.

In general, it is observed that the QDA method achieved the best classification results for all four species. The QDA models achieved better results because the algorithm was able to separate the cluster overlap, using a quadratic function to calculate a non-linear decision boundary. The results in the current study illustrates that QDA models are able to predict the frozen periods, but, due to over-fitting, the models were not deemed as effective as what the calibration models suggested. However, the models were able to distinguish between fresh and frozen-thawed meat, irrespective of the frozen period, with relatively high accuracies.

Table 4.60 The performance measures used to assess the DA models for fresh vs. frozen-thawed meat of the four species, pre-processed with SNV + detrend.

Species	Number of principal components (PCs)	Method	Treatment	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
BWB	6	QDA	Fresh	96.7	3.2	0	100.0	100.0	93.0	96.4	3.3
			FT	96.7	0	3.2	94.4	93.0	100.0	97.1	3.3
Zebra	6	QDA	Fresh	94.5	2.4	3.2	93.2	94.1	94.8	94.0	5.5
			FT	94.5	3.2	2.4	95.5	94.8	94.1	94.8	5.5
Springbok	6	QDA	Fresh	99.2	0.8	0	100.0	100.0	97.9	98.9	0.8
			FT	99.2	0	0.8	98.7	97.9	100.0	99.3	0.8
Ostrich	6	QDA	Fresh	95.8	3.3	0.8	97.8	98.6	91.8	94.7	4.2
			FT	95.8	0.8	3.3	94.6	91.8	98.6	96.6	4.2

(DA) Discriminant analysis; (SNV) Standard normal variate; (BWB) Black wildebeest; (QDA) Quadratic discriminant analysis; (FT) Frozen-thawed.

4.2.4.1.4 Partial least squares discriminant analysis (PLS-DA)

Table 4.61 PLS-DA model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen period classification.

Species	Number of latent variables (LVs)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
BWB	4	Calibration	68.3	31.8
		Cross-validation	45.2	54.8
		Validation	78.4	23.6
Zebra	4	Calibration	66.7	33.3
		Cross-validation	57.9	42.1
		Validation	77.8	22.2
Springbok	3	Calibration	60.0	40.0
		Cross-validation	50.0	50.0
		Validation	63.3	36.7
Ostrich	4	Calibration	70.8	29.2
		Cross-validation	48.3	51.7
		Validation	68.8	31.2

(PLS-DA) Partial least squares discriminant analysis; (SNV) Standard normal variate; (BWB) Black wildebeest.

The classification accuracies shown in **Table 4.61** provides a general overview of the optimal models' performance. The table also illustrates the misclassification of meat samples subjected to different frozen periods (1 – 9 months). The misclassification of the treatments can be explained by studying the PCA score plots (**Figure 4.36**) of the different frozen periods. These figures exhibit a major overlap between the freeze-thaw treatments, as previously explained, thus resulting in a misclassification between the frozen periods. When looking at the accuracies of the models, it is evident that the black wildebeest and zebra models were under-fitted, resulting in improved validation accuracies (Miller, 2005). In contrast, the ostrich model exhibited a decreased validation accuracy, thus suggesting that the model was over-fitted. The classification rates for the springbok model achieved similar accuracies which are indicative of a model with a good fit. The results also revealed some possible model implications as the cross-validation results were considerably lower than the calibration- and validation accuracies, thus resulting in a model with an unknown fit (Miller, 2005). This phenomenon is one drawback of the leave-one-out cross-validation method as previously discussed in **Section 4.2.3.4**. This suggests that these models are not effective when trying to classify the different frozen periods. For this reason, it was deemed important to examine success of a two-group discrimination model. PLS-DA models were developed to discriminate between fresh and frozen-thawed meat, irrespective of the frozen period.

Table 4.62 PLS-DA model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen-thawed meat classification, irrespective of frozen period.

Species	Number of latent variables (LVs)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
BWB	4	Calibration	96.0	4.0
		Cross-validation	96.0	4.0
		Validation	96.0	4.0
Zebra	2	Calibration	88.9	11.1
		Cross-validation	87.3	12.7
		Validation	83.3	16.7
Springbok	4	Calibration	97.5	2.5
		Cross-validation	97.5	2.5
		Validation	98.3	1.7
Ostrich	3	Calibration	91.7	8.3
		Cross-validation	88.3	11.7
		Validation	96.7	3.3

(PLS-DA) Partial least squares discriminant analysis; (SNV) Standard normal variate; (BWB) Black wildebeest.

The PLS-DA models for black wildebeest, zebra, springbok and ostrich gave good discrimination results. The PLS-DA calibration models achieved overall classification accuracies of 96% (black wildebeest), 88.9% (zebra), 97.5% (springbok), 91.7% (ostrich) and misclassification rates of 4% (black wildebeest), 11.1% (zebra), 2.5% (springbok), 8.3% (ostrich) (**Table 4.62**). The PLS-DA prediction score plots for the four species are given in **Figure 4.46**. The score plots (LV1 vs. LV2) (**Figure 4.45a – d**) demonstrates minimal overlap between the treatment classes which is indicative of satisfactory model calibration. The score plot of LV1 vs. LV2 of black wildebeest and springbok (**Figure 4.45a,c**) exhibited the best treatment separation, with two prominent clusters. The separation between the fresh and frozen-thawed treatments were best described in LV1, while LV2 accounted for little class separation. The fresh samples were predominantly associated with the positive scores in LV1 along with the negative score in LV2. Alternatively, the frozen-thawed samples associated with the negative scores in LV1 and positive scores in LV2. Therefore, the score plots (**Figure 4.5a – d**) exhibited that treatment separation is best described in the direction of both LV1 and LV2.

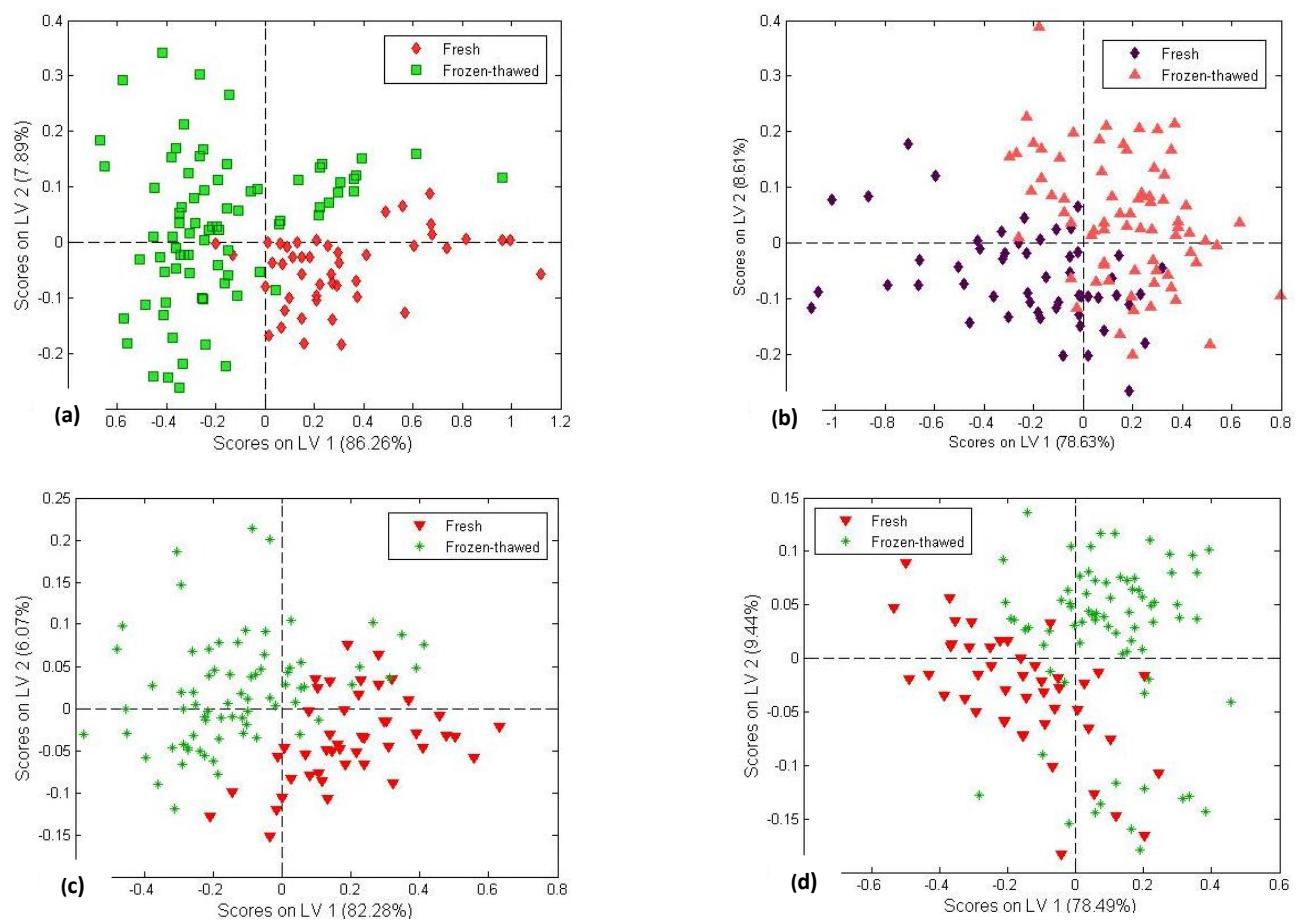
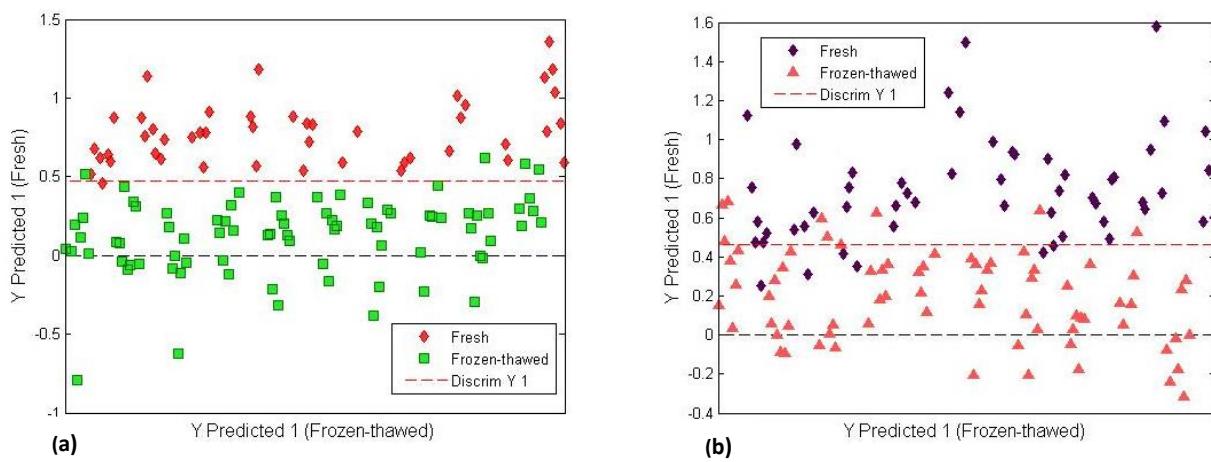


Figure 4.45 PLS-DA score plots (SNV + detrend pre-processed) of fresh (red/purple) and frozen-thawed (green/peach) meat samples. (a) score plot of LV1 (86.26%) vs. LV2 (7.89%) for black wildebeest, (b) score plot of LV1 (78.63%) vs. LV2 (8.61%) for zebra, (c) score plot of LV1 (82.28%) vs. LV2 (6.07%) for springbok and (d) score plot of LV1 (78.49%) vs. LV2 (9.44%) for ostrich, colour coded per treatment class.



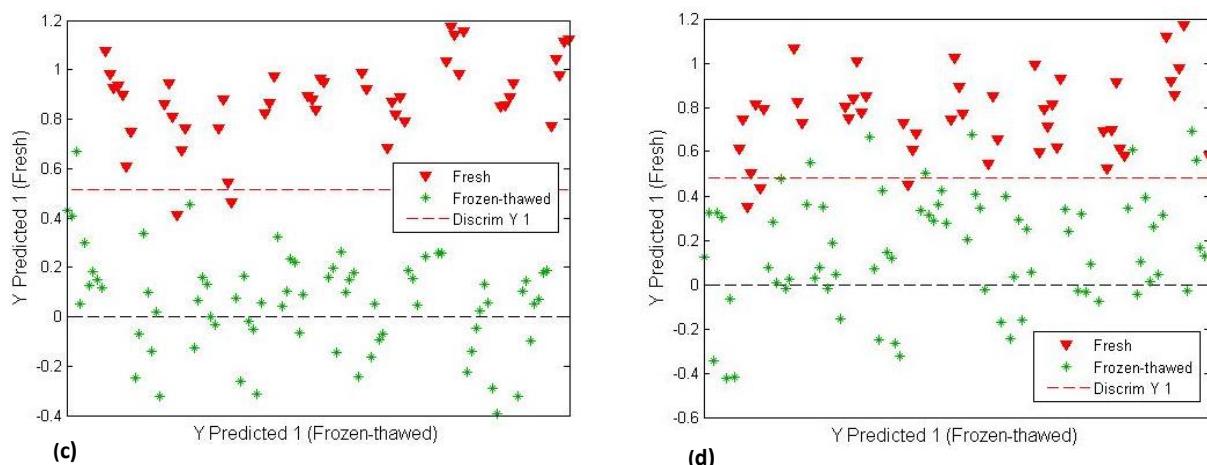


Figure 4.46 PLS-DA models (SNV + detrend pre-processed) resulting in satisfactory overall classification accuracies. PLS-DA prediction score plot of (a) black wildebeest (b) zebra, (c) springbok and (d) ostrich, illustrating the predicted fresh (red/purple) and frozen-thawed (green/peach) objects. (a) Score plot (4 LVs) for black wildebeest (96%) of objects predicted as fresh [above red line (Y1)] vs. frozen-thawed [below red line (Y1)], (b) score plot (2 LVs) for zebra (88.9%) of objects predicted as fresh [above red line (Y1)] vs. frozen-thawed [below red line (Y1)], (c) score plot (4 LVs) for springbok (97.5%) of objects predicted as fresh [above red line (Y1)] vs. frozen-thawed [below red line (Y1)] and (d) score plot (3 LVs) for ostrich (91.7%) of objects predicted as fresh [above red line (Y1)] vs. frozen-thawed [below red line (Y1)].

The PLS-DA prediction score plots (**Figure 4.46a,c**) for black wildebeest and springbok illustrated that the fresh and frozen-thawed treatments were nearly all correctly predicted and the performance measures in **Table 4.63** confirmed this. The PLS-DA prediction score plots for zebra (**Figure 4.46b**) and ostrich (**Figure 4.46d**) illustrated a larger overlap between the treatments, resulting in a higher misclassification. This was also confirmed by the performance measures (**Table 4.63**). The springbok model resulted in the highest classification accuracy (97.5%), followed by black wildebeest (96%), ostrich (91.7%) and zebra (88.9%). Although both treatments for the springbok model achieved classification accuracies of 97.5%, the model exhibited a higher probability of predicting for frozen-thawed. This is confirmed by the higher sensitivity (98.7%), specificity (97.8%) and F1 score (98%) (**Table 4.63**). An opposite trend was observed for the black wildebeest, zebra and ostrich models. These performance measures in **Table 4.63** confirm the models' ability to classify the fresh meat samples with little fault. Even though the zebra model exhibited classification accuracies of 88.9% (calibration) and 87.3% (cross-validation), indicative of an effective model, the validation accuracy (83.3%) decreased. This indicated that the model was over-fitted (Miller, 2005) and therefore not as effective as suggested by the calibration- and cross-validation accuracies. The results in the current study suggest that PLS-DA can be used to discriminate between fresh and frozen-thawed game meat samples. Ropodi *et al.* (2018) reported similar results (93.3%) for fresh and frozen-thawed minced beef meat and discriminated the samples based on physical and chemical changes as discussed in **Section 4.2.2**.

Table 4.63 The performance measures used to assess the PLS-DA models for fresh vs. frozen-thawed meat of the four species, pre-processed with SNV + detrend.

Species	Number of latent variables (LVs)	Treatment	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
BWB	4	Fresh	96.0	3.2	0.8	97.8	98.7	91.8	94.7	4.0
		FT	96.0	0.8	3.2	95.0	91.8	98.7	96.8	4.0
Zebra	2	Fresh	88.9	6.4	4.8	88.9	91.4	85.7	87.3	11.1
		FT	88.9	4.8	6.4	88.9	85.7	91.4	90.1	11.1
Springbok	4	Fresh	97.5	0.8	1.7	95.7	97.3	97.8	96.7	2.5
		FT	97.5	1.7	0.8	98.7	97.8	97.3	98.0	2.5
Ostrich	3	Fresh	91.7	5.8	2.5	93.5	95.7	86.0	89.6	8.3
		FT	91.7	2.5	5.8	90.5	86.0	95.7	93.1	8.3

(PLS-DA) Partial least squares discriminant analysis; (SNV) Standard normal variate; (BWB) Black wildebeest; (FT) Frozen-thawed.

4.2.4.2. Optimal model selection

After the models [pre-processed with SNV + detrend (4th order polynomial)] were calculated to distinguish between the fresh and previously frozen meat, irrespective of the frozen period, as well as determine the frozen period, the models were compared to identify the one with the best results. **Table 4.64** reveals that the QDA- and PLS-DA models produced the best results.

Table 4.64 An overview of the accuracies for the various classification and discrimination models, pre-processed with SNV + detrend to distinguish between fresh or previously frozen meat, irrespective of the frozen period, as well as determine the frozen period.

Species	Model	Fresh vs. frozen period			Fresh vs. frozen-thawed, irrespective of frozen period		
		Number of nearest neighbour (<i>k</i>), PCs or LVs	Classification accuracy (%)	Misclassification rate (%)	Number of nearest neighbour (<i>k</i>), PCs or LVs	Classification accuracy (%)	Misclassification rate (%)
BWB	SIMCA	-	45.2	54.8	-	75.8	24.2
	KNN	3	50.8	49.2	3	94.4	5.6
	QDA	6	90.5	9.5	6	96.7	3.3
	PLS-DA	4	68.3	31.8	4	96.0	4.0
Zebra	SIMCA	-	67.9	32.1	-	71.9	28.1
	KNN	3	45.2	54.8	3	84.1	15.9
	QDA	6	93.7	6.3	6	94.5	5.5
	PLS-DA	4	66.7	33.3	4	88.9	11.1
Springbok	SIMCA	-	44.6	55.4	-	90.0	10.0
	KNN	2	55.8	44.2	3	94.2	5.8
	QDA	6	94.2	5.8	6	99.2	0.8
	PLS-DA	3	60.0	40.0	4	97.5	2.5
Ostrich	SIMCA	-	68.4	31.6	-	80.4	19.6
	KNN	2	54.2	45.8	3	78.3	21.7
	QDA	6	94.2	5.8	6	95.8	4.2
	PLS-DA	4	70.8	29.2	3	91.7	8.3

(SNV) Standard normal variate; (PCs) Principal components; (LVs) Latent variables; (SIMCA) Soft independent modelling of class analogy; (KNN) K-nearest neighbour; (QDA) Quadratic discriminant analysis; (PLS-DA) Partial least squares discriminant analysis.

This data was then subjected to different pre-processing techniques, as discussed in **Section 4.2.3.2**, to determine if the classification accuracies could be improved. PCA models were calculated, the data explored and outliers removed, then QDA- and PLS-DA models were recalculated and evaluated.

4.2.4.3 Exploratory data analysis (EDA)

4.2.4.3.1 Principal component analysis (PCA)

The score plots of PC1 vs. PC2 (**Addendum B, Figure B34a,c – B65a,c**) were used to investigate the difference between the treatments, as the data shows that the variation was best explained in the first component. Similar results were obtained for the various pre-processing techniques as previously discussed in **Section 4.2.3.3**. The separation between treatments was predominantly due to differences in moisture, fat, protein and pH as illustrated in the loading plots (**Addendum B, Figure B34b – B65b**). The results from the various pre-processing techniques suggests that the Savitzky-Golay first- and second derivative transformations would be sufficient and best suited for fresh and frozen-thawed separation as well as frozen period prediction. The Savitzky-Golay first- and second derivative transformations enhanced the differences in macronutrient composition of the biological samples. Even though Savitzky-Golay derivatives showed the best PCA results for separation, all the pre-processing techniques were used for further data analysis and model evaluation.

4.2.4.4 Multivariate data analysis (MDA): Optimal model development

4.2.4.4.1 Quadratic discriminant analysis (QDA)

4.2.4.4.1.1 Discrimination of fresh vs. frozen period

The overall accuracy of the discriminant analysis models, illustrated that the springbok model provided the best discrimination. The springbok QDA models (6 PCs) for the various types of pre-processing gave similar results (**Table 4.65**). The QDA models, pre-processed with SNV, SNV + SGd₂(7), SNV + detrend + SGd₂(7) and SGd₂(9), presented the best discrimination results. Although there was no substantial difference between the calibration results of the mentioned pre-processing techniques, further investigation of the validation accuracies and performance measures (not shown) suggests that the model pre-processed with SNV + detrend + SGd₂(7) would be optimal. The calibration model achieved an overall classification accuracy of 98.3% with a misclassification rate of 1.7% (**Table 4.65**). This model also exhibited excellent performance measures (not shown), thus the sensitivity (98.4%) and specificity (99.8%) suggests that the model can accurately distinguish between the treatments. Therefore, SNV + detrend + SGd₂(7) pre-processed data resulted in more robust models, and is most likely ascribed to the fact that SNV + detrend and Savitzky-Golay second derivative both eliminate noise and corrects for the baseline shift (Barnes *et al.*, 1989; Esbensen *et al.*, 2002).

The QDA models, pre-processed with SNV + detrend + SGd₂(7), achieved better results than that of the original model pre-processed with SNV + detrend. To illustrate that there was a difference between these two pre-processing techniques, the PCA score plots of SNV + detrend + SGd₂(7) (**Addendum B, Figure B51c**) and SNV + detrend (**Figure 4.63c**) are shown. The pre-processing improved the separation in PC1 as well as clustering of the frozen periods, and though the separation and clustering was more prominent, an overlap between the frozen periods was still evident (**Addendum B, Figure B51c**). The reason for the differences

between these pre-processing techniques was previously discussed in **Section 4.2.3.4**. Therefore, the SNV + detrend + SGd₂(7) pre-processing achieved better results as the separation due to differences in moisture, fat, protein and pH were enhanced.

The overview of the classification accuracies (**Table 4.65**) and performance measures (not shown) illustrated that the QDA model (6 PCs) pre-processed with SNV gave the best results for black wildebeest (95.2%) and the QDA (6 PCs) models pre-processed with SNV + detrend + SGd₂(7) and SNV + detrend + SGd₂(9) gave the best results for zebra (92.9%) and ostrich (95.8%), respectively. The zebra QDA model, pre-processed with SNV + detrend + SGd₂(7) (92.9%), achieved slightly lower results than that of the original model pre-processed with SNV + detrend (93.7%). To illustrate that there were differences between these two pre-processing techniques, the PCA score plots of SNV + detrend + SGd₂(7) (**Addendum B, Figure B42c**) and SNV + detrend (**Figure 4.36b**) are shown. The SNV + detrend score plot of PC1 exhibited an improved explained variance of 78%, with less overlap between frozen periods. Therefore, the original SNV + detrend model performed better due to the improved separation as a result of the more prominent differences in moisture, protein and pH, enhanced by the pre-processing. The QDA models, pre-processed with SNV (black wildebeest) and SNV + detrend + SGd₂(9) (ostrich), gave better results than that of the original models pre-processed with SNV + detrend. The classification accuracy for differentiating between fresh vs. frozen period was 95.2% for black wildebeest [SNV] and 95.8% for ostrich [SNV + detrend + SGd₂(9)]. Both the black wildebeest and ostrich models improved with the mentioned techniques because both pre-processing methods minimised the overlap between the frozen periods and improved the clustering of the classes. Consequently, improving separation between classes as well as the predictive power of the model. Therefore, the SNV (black wildebeest) and SNV + detrend + SGd₂(9) (ostrich) pre-processing achieved better results as the separation due to differences in moisture, fat, protein and pH were enhanced.

Although the use of different pre-processing techniques improved the calibration accuracies and performance measures of the models, the validation accuracies decreased. This indicated that the models were over-fitted (Miller, 2005) and therefore not as effective as suggested by the calibration accuracies. The results in the current study suggest that the discrimination of the frozen periods was less sufficient due to the lack of separation between the frozen periods. This phenomenon can be ascribed to the samples' spectral similarities, hence causing a misclassification between classes. Therefore, it can be concluded that the extent of damage to the cell membranes over the first month of freezing did not increase with an increase in frozen storage.

4.2.4.4.1.2 Discrimination of fresh vs. previously frozen, irrespective of frozen period

The overall accuracy of the discriminant analysis models, illustrated that the springbok model provided the best discrimination. The springbok QDA models (6 PCs) for the various types of pre-processing gave similar results (**Table 4.66**). Although the QDA models, irrespective of the pre-processing, achieved identical results, when comparing the PCA score plots (**Addendum B, Figure B48a – B56a**) it was observed that, when using

SNV + SGd₂(7/9) and SNV + detrend + SGd₂(7/9), the separation between classes was more distinct with little overlap. The calibration models achieved an overall classification accuracy of 99.2% for SNV + SGd₂(7/9) and SNV + detrend + SGd₂(7/9) with misclassification rates of 0.8% (**Table 4.66**). These models also exhibited excellent validation accuracies (100%), and this suggests that the models can accurately distinguish between the treatments. Therefore, SNV + SGd₂(7/9) and SNV + detrend + SGd₂(7/9) pre-processed data resulted in more robust models. The classification accuracy (99.2%), sensitivity (99.3%) and specificity (99%) confirmed this (not shown).

The QDA models, pre-processed with SNV + SGd₂(7/9) and SNV + detrend + SGd₂(7/9), achieved similar results to that of the original model pre-processed with SNV + detrend. The classification accuracy for differentiating between the fresh and frozen-thawed meat was 99.2% for all these pre-processing methods. Although the SNV + detrend model gave the same calibration accuracy (99.2%) and sensitivity (99.3%), the validation accuracy (98.3%) and specificity (98.9%) was slightly lower, suggesting that the models pre-processed with SNV + SGd₂(7/9) and SNV + detrend + SGd₂(7/9) would provide better discrimination between the fresh and frozen-thawed meat. The reason for the slight differences between these pre-processing techniques was previously discussed in **Section 4.2.3.4**. Therefore, the SNV + SGd₂(7/9) (**Addendum B, Figure B49a, B50a**) and SNV + detrend + SGd₂(7/9) (**Addendum B, Figure B51a, B52a**) pre-processing achieved better results as the separation due to differences in moisture, fat, protein and pH were enhanced.

The overview of the classification accuracies (**Table 4.66**) and performance measures (not shown) illustrated that the QDA model (6 PCs) pre-processed with SNV gave the best results for black wildebeest (96%) and the QDA (6 PCs) models pre-processed with SGd₁(7) and SNV + detrend + SGd₂(7/9) gave the best results for zebra (97.6%) and ostrich (98.3%), respectively. The black wildebeest QDA model, pre-processed with SNV (96%), achieved slightly lower results than that of the original model pre-processed with SNV + detrend (96.7%). The SNV + detrend score plot of PC1 (**Addendum B, Figure B3a**) exhibited an improved explained variance of 88%, with less overlap between fresh and frozen-thawed meat. Therefore, the original SNV + detrend model performed better due to the improved separation because of the more prominent differences in moisture, fat, protein and pH, enhanced by the pre-processing. The QDA models, pre-processed with SGd₁(7) (zebra) and SNV + detrend + SGd₂(7/9) (ostrich), gave better results than that of the original models pre-processed with SNV + detrend. The classification accuracy was 97.6% for zebra [SGd₁(7)] and 98.3% for ostrich [SNV + detrend + SGd₂(7/9)]. Both the black zebra and ostrich models improved with the mentioned techniques because both pre-processing methods minimised the overlap and improved the clustering of the classes. Consequently, improving the separation between classes as well as the predictive power of the model. Therefore, the SGd₁(7) (zebra) and SNV + detrend + SGd₂(7/9) (ostrich) pre-processing achieved better results as the separation due to differences in moisture, fat, protein and pH were enhanced.

Table 4.65 An overview of the classification accuracies of the QDA models (6 PCs) for fresh vs. frozen period with various pre-processing techniques applied.

Pre-processing	Species							
	Black wildebeest		Zebra		Springbok		Ostrich	
	Calibration accuracy (%)	Validation accuracy (%)						
Pre-processing	Misclassification rate (%)							
SNV	95.2 4.8	74.1	70.6 29.4	61.1	98.3 1.7	78.3	87.5 12.5	56.7
SNV+SGd ₂ (7)	-	-	80.2 19.8	61.1	97.5 2.5	71.7	88.3 11.7	73.3
SNV+SGd ₂ (9)	86.5 13.5	70.4	73.0 27.0	63.0	90.8 9.2	71.7	91.7 8.3	61.7
SNV+detrend+SGd ₂ (7)	67.5 32.5	70.4	92.9 7.1	75.9	98.3 1.7	81.7	92.5 7.5	76.7
SNV+detrend+SGd ₂ (9)	86.5 13.5	66.7	90.5 9.5	74.1	93.3 6.7	82.0	95.8 4.2	80.0
SGd ₁ (5)	-	-	91.3 8.7	64.8	93.3 6.7	75.0	83.3 11.7	63.3
SGd ₁ (7)	69.0 31.0	59.3	92.1 7.9	66.7	90.0 10.0	73.3	79.2 20.8	55.0
SGd ₂ (7)	-	-	-	-	91.7 8.3	70.0	77.5 22.5	48.3
SGd ₂ (9)	80.2 19.8	66.7	89.7 10.3	70.4	97.5 2.5	73.3	90.8 9.2	61.7

(QDA) Quadratic discriminant analysis; (PCs) Principal components; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

Table 4.66 An overview of the classification rates of the QDA models (6 PCs) for fresh vs. previously frozen meat, irrespective of the frozen period, with various pre-processing techniques applied.

Pre-processing	Species							
	Black wildebeest		Zebra		Springbok		Ostrich	
	Calibration accuracy (%)	Validation accuracy (%)						
	Misclassification rate (%)		Misclassification rate (%)		Misclassification rate (%)		Misclassification rate (%)	
SNV	96.0 4.0	100.0	89.7 10.3	88.9	99.2 0.8	96.7	92.5 7.5	90.0
SNV+SGd ₂ (7)	-	-	91.3 8.7	92.5	98.3 1.7	100.0	95.0 5.0	95.0
SNV+SGd ₂ (9)	90.5 9.5	92.6	89.7 10.3	94.4	99.2 0.8	100.0	95.8 4.2	96.7
SNV+detrend+SGd ₂ (7)	92.1 7.9	88.9	87.3 12.7	87.0	99.2 0.8	100.0	98.3 1.7	96.7
SNV+detrend+SGd ₂ (9)	92.1 7.9	96.3	87.3 12.7	83.3	99.2 0.8	100.0	98.3 1.7	96.7
SGd ₁ (5)	-	-	95.2 4.8	92.6	99.2 0.8	100.0	98.3 1.7	93.3
SGd ₁ (7)	95.2 4.8	87.0	97.6 2.4	88.9	99.2 0.8	100.0	97.5 2.5	95.0
SGd ₂ (7)	-	-	-	-	99.2 0.8	100.0	96.7 3.3	95.0
SGd ₂ (9)	85.7 14.3	83.3	94.4 4.6	94.4	99.2 0.8	96.7	96.7 3.3	95.0

(QDA) Quadratic discriminant analysis; (PCs) Principal components; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

4.2.4.4.2 Partial least squares discriminant analysis (PLS-DA)

4.2.4.4.2.1 Discrimination of fresh vs. frozen period

The overall accuracy of the discriminant analysis models, illustrated that the ostrich model provided the best discrimination. The ostrich PLS-DA models for the various types of pre-processing gave different results (**Table 4.67**). The PLS-DA models, pre-processed with SNV + detrend + SGd₂(7) (90.8%) and SNV + detrend + SGd₂(9) (91.7%), presented the best discrimination results. Although there was no substantial difference between the calibration results of the mentioned pre-processing techniques, further investigation of the cross-validation-, validation accuracies (**Table 4.67**) and performance measures (not shown) suggests that the model pre-processed with SNV + detrend + SGd₂(7) would be optimal. The calibration model achieved an overall classification accuracy of 90.8% with a misclassification rate of 9.2% (**Table 4.67**). This model also exhibited excellent performance measures (not shown), thus the sensitivity (92.1%) and specificity (98.7%) suggests that the model can accurately distinguish between the treatments. Therefore, SNV + detrend + SGd₂(7) pre-processed data resulted in a more robust model.

The PLS-DA models, pre-processed with SNV + detrend + SGd₂(7) (90.8%), achieved better results than that of the original model pre-processed with SNV + detrend (70.8%). To illustrate this, the PCA score plots of SNV + detrend + SGd₂(7) (**Addendum B, Figure B60c**) and SNV + detrend (**Figure 4.36d**) are shown. The reason for the differences between these pre-processing techniques was previously discussed in **Section 4.2.3.4**. Therefore, the SNV + detrend + SGd₂(7) pre-processing achieved better results as the separation due to differences in moisture, fat, protein and pH were enhanced.

There are substantial differences between the various pre-processing techniques for the zebra, springbok and black wildebeest models (**Table 4.67**). The overview of the classification accuracies and performance measures (not shown) illustrated that the PLS-DA model pre-processed with SNV gave the best results for zebra (75.4%) and the models pre-processed with SNV + detrend + SGd₂(7) and SNV + detrend + SGd₂(9) gave the best results for springbok (86.7%) and black wildebeest (86.5%), respectively. The zebra PLS-DA model, pre-processed with SNV (75.4%), achieved slightly higher results compared to the original model pre-processed with SNV + detrend (66.7%). The PLS-DA models, pre-processed with SNV + detrend + SGd₂(7) (springbok) and SNV + detrend + SGd₂(9) (black wildebeest), gave considerably better results than that of the original models pre-processed with SNV + detrend. The classification accuracy for differentiating between fresh vs. frozen period was 86.7% for springbok [SNV + detrend + SGd₂(7)] and 86.5% for black wildebeest [SNV + detrend + SGd₂(9)]. The original models resulted in accuracies of 66% for springbok and 68.3% for black wildebeest. All the models improved with the mentioned techniques because the pre-processing methods minimised the overlap between the frozen periods and improved the clustering of the classes. Consequently, improving the separation between classes as well as the predictive power of the model. Therefore, the SNV (zebra), SNV + detrend + SGd₂(7) (springbok) and SNV + detrend + SGd₂(9) (black wildebeest) pre-processing achieved better results as the separation due to differences in moisture, fat, protein and pH were enhanced.

Although the use of different pre-processing techniques improved the calibration accuracies and performance measures of the models, the validation accuracies decreased. This indicated that the models were over-fitted (Miller, 2005) and therefore not as effective as suggested by the calibration accuracies.

4.2.4.4.2.2 Discrimination of fresh vs. previously frozen, irrespective of frozen period

The PLS-DA models for the various types of pre-processing gave similar results. The overall accuracy of the discriminant analysis models, illustrated that the springbok model provided the best discrimination. The springbok PLS-DA models exhibited no substantial difference between the various types of pre-processing (**Table 4.68**). Although the PLS-DA models, irrespective of the pre-processing, achieved similar results, when comparing the PCA score plots (**Addendum B, Figure B48a – B56a**) it was observed that, when using SNV + SGd₂(7) and SGd₂(9), the separation between classes was more distinct with little overlap. The calibration models achieved an overall classification accuracy of 100% for both the SNV + SGd₂(7) and SGd₂(9) corrected data (**Table 4.68**). These models also exhibited excellent validation accuracies (98.3 and 100%), and this suggests that the models can accurately distinguish between the treatments. Although there was no substantial difference between the calibration results of the mentioned pre-processing techniques, further investigation of the cross-validation-, validation accuracies and performance measures (not shown) suggests that the model pre-processed with SGd₂(9) would be optimal. The models also revealed that the classification between fresh and previously frozen meat was perfect. The classification accuracy (100%), sensitivity (100%) and specificity (100%) confirmed this (not shown). Therefore, SGd₂(9) pre-processed data resulted in more robust models.

The PLS-DA models, pre-processed with SGd₂(9) (100%), achieved better results than that of the original model pre-processed with SNV + detrend (97.5%). To illustrate that there were minimal differences between these two pre-processing techniques, the PCA score plots of SGd₂(9) (**Addendum B, Figure B56a**) and SNV + detrend (**Addendum B, Figure B4a**) are shown. The reason for the slight differences between these two pre-processing techniques is most likely because the derivatives enhanced the differences in moisture, fat, protein and pH and due to the enhancement improved the class separation. Therefore, the SGd₂(9) pre-processing achieved better results.

The overview of the classification accuracies (**Table 4.68**) and performance measures (not shown) illustrated that the PLS-DA model pre-processed with SGd₁(5) gave the best results for zebra (94.4%) and the models pre-processed with SNV + detrend + SGd₂(9) gave the best results for black wildebeest (99.2%) and ostrich (98.3%), respectively. The PLS-DA models, pre-processed with SGd₁(5) (zebra) and SNV + detrend + SGd₂(9) (black wildebeest and ostrich), gave better results than that of the original models pre-processed with SNV + detrend. To illustrate that there were minimal differences between these two pre-processing techniques, the PCA score plots of SGd₁(5) (**Addendum B, Figure B45a**), SNV + detrend + SGd₂(9) (**Addendum B, Figure B37a, B61a**) and SNV + detrend (**Figure 4.35a, Addendum B, Figure B3a, B5a**) are shown. The mentioned pre-processing methods improved the separation as well as the clustering of the fresh vs. frozen-

thawed treatments by minimising the overlap. The use of derivatives made it possible to minimise the overlap, as the derivatives removed the unwanted features and by doing so enhanced the features advantageous for the analysis (Savitzky & Golay, 1964). Consequently, this improved separation and clustering results in a model with an improved predictive power. Therefore, the SGd₁(5) (zebra), SNV + detrend + SGd₂(9) (black wildebeest and ostrich) pre-processing achieved better results as the separation due to differences in moisture, fat, protein and pH were enhanced.

These results exhibited that the models' improvement for all the species was predominantly due to the enhancement of the absorption band related to fat. The SNV + detrend pre-processing rarely exhibited absorption bands related to fat. Therefore, the conclusion can be made that better separation results were obtained due to the enhancement of differences in the macronutrient composition such as moisture-, fat-, protein content as well as changes in pH. These results also indicate that these models are not effective when trying to classify the different frozen periods. This suggests that it is unnecessary to determine the frozen period of meat as the discrimination between fresh and previously frozen meat, irrespective of the frozen period, is more sufficient and closer to commercial requirements.

Table 4.67 An overview of the classification accuracies of the PLS-DA models for fresh vs. frozen period with various pre-processing techniques applied.

Pre-processing	Species														
	Black wildebeest				Zebra				Springbok				Ostrich		
	Calibration accuracy (%)	CV (%)	Validation accuracy (%)	Misclassification rate (%)	Calibration accuracy (%)	CV (%)	Validation accuracy (%)	Misclassification rate (%)	Calibration accuracy (%)	CV (%)	Validation accuracy (%)	Misclassification rate (%)	Calibration accuracy (%)	CV (%)	Validation accuracy (%)
SNV	64.3 35.7	64.3	68.5		75.4 24.6	70.5	66.7		62.5 37.5	50.0	76.7	80.8 19.2	61.7	73.3	
SNV+SGd ₂ (7)	-		-		35.7 64.3	34.0	42.6		79.2 20.8	61.7	85.0	80.0 20.0	64.2	80.0	
SNV+SGd ₂ (9)	69.1 30.9	69.0	70.4		39.7 60.3	38.6	42.6		80.0 20.0	63.3	81.7	77.5 22.5	65.8	75.0	
SNV+detrend+SGd ₂ (7)	67.5 32.5	65.3	70.4		34.9 65.1	30.1	51.9		86.7 13.3	69.2	80.0	90.8 9.2	68.3	76.7	
SNV+detrend+SGd ₂ (9)	86.5 13.5	70.1	66.7		37.3 62.7	35.7	48.2		81.7 18.3	70.0	81.7	91.7 8.3	67.5	71.7	
SGd ₁ (5)	-	-	-		50.8 49.2	47.3	40.7		59.2 40.8	46.7	56.7	60.0 40.0	52.5	50.0	
SGd ₁ (7)	69.0 31.0	58.3	59.3		47.6 52.4	47.6	51.9		59.2 40.8	43.3	56.7	74.2 25.8	60.0	65.0	
SGd ₂ (7)	-	-	-		-	-	-		55.0 45.0	41.7	46.7	80.0 20.0	61.7	76.7	
SGd ₂ (9)	33.3 66.7	30.2	33.3		65.4 34.7	63.8	60.4		75.0 25.0	52.9	65.0	77.5 22.5	59.2	76.7	

(PLS-DA) Partial least squares discriminant analysis; (CV) Cross-validation; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

Table 4.68 An overview of the classification accuracies of the PLS-DA models for fresh vs. previously frozen meat, irrespective of the frozen period, with various pre-processing techniques applied.

Pre-processing	Species															
	Black wildebeest				Zebra				Springbok				Ostrich			
	Calibration accuracy (%)		CV (%)	Validation accuracy (%)	Calibration accuracy (%)		CV (%)	Validation accuracy (%)	Calibration accuracy (%)		CV (%)	Validation accuracy (%)	Calibration accuracy (%)		CV (%)	Validation accuracy (%)
	Misclassification rate (%)	CV (%)			Misclassification rate (%)	CV (%)			Misclassification rate (%)	CV (%)			Misclassification rate (%)	CV (%)		
SNV	94.4 5.6	92.4	96.3	-	94.4 5.6	92.4	96.3	-	97.5 2.5	96.7	96.7	-	95.8 4.2	93.3	93.3	
SNV+SGd ₂ (7)	-	-	-	-	92.1 7.9	90.5	88.9	-	100.0 0	98.3	98.3	-	92.5 7.7	94.2	98.3	
SNV+SGd ₂ (9)	98.4 1.6	98.2	98.2	-	94.4 5.6	94.4	94.4	-	98.3 1.7	98.3	100.0	-	93.3 6.7	92.5	98.3	
SNV+detrend+SGd ₂ (7)	99.2 0.8	98.2	98.2	-	88.9 11.1	87.3	98.2	-	99.2 0.8	99.2	100.0	-	96.7 3.3	95.0	96.7	
SNV+detrend+SGd ₂ (9)	99.2 0.8	98.2	98.2	-	92.1 7.9	90.4	88.9	-	96.7 3.3	96.7	98.3	-	98.3 1.7	96.7	98.3	
SGd ₁ (5)	-	-	-	-	94.4 5.6	94.4	96.3	-	99.2 0.8	99.2	100.0	-	98.3 1.7	95.8	96.7	
SGd ₁ (7)	98.4 1.6	96.3	96.3	-	94.4 5.6	90.3	92.6	-	99.2 0.8	99.2	100.0	-	97.5 2.5	95.8	98.3	
SGd ₂ (7)	-	-	-	-	-	-	-	-	99.2 0.8	99.2	100.0	-	96.7 3.3	96.7	95.0	
SGd ₂ (9)	97.6 2.4	97.0	96.3	-	90.5 9.5	89.2	88.9	-	100.0 0	99.2	100.0	-	96.7 3.3	96.7	95.0	

(PLS-DA) Partial least squares discriminant analysis; (CV) Cross-validation; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

4.2.4.5 Optimal model selection

After the models for the various types of pre-processing were calculated, the models with the best results were compared to identify the optimal one (**Table 4.69** and **Table 4.70**). The results for the frozen period models (**Table 4.69**) suggest that the use of QDA models would be best suited for the differentiation between the different frozen periods. Although these QDA models achieved excellent calibration accuracies the validation accuracies were considerably lower, indicating that the models were over-fitted and not as effective as suggested by the calibration models. Therefore, it was possible to conclude that discrimination between fresh and previously frozen meat, irrespective of the frozen period, would be more sufficient.

It was concluded that the PLS-DA models, pre-processed with Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points) and Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points) was the best for differentiating between fresh and frozen-thawed zebra and springbok meat, irrespective of the frozen period. The results for the black wildebeest and ostrich models (PLS-DA) suggests that the SNV + detrend + SGd₂(9) pre-processing method would be sufficient for differentiating between fresh and previously frozen meat. The PCA score plots for the above-mentioned pre-processed data illustrates that these methods improved the separation between treatments in PC1. Hence, making it possible to conclude that these pre-processing techniques would be the best for enhancing the separation between fresh and frozen-thawed samples for these specific species. The PLS-DA models achieved better results as it generally outperforms QDA, when classes are closely related, because it overcomes the collinearity problems often associated with QDA. SGd₁(5), SGd₂(7) and SNV + detrend + SGd₂(9) outperformed the combined pre-processing of SNV + detrend for the black wildebeest, zebra, springbok and ostrich, as the Savitzky-Golay transformation was found to have enhanced the differences in moisture, fat, protein and pH, which was predominantly the contributors for fresh and frozen-thawed separation.

Table 4.69 An overview of the classification rates of the DA and PLS-DA models for fresh vs. frozen period with various pre-processing techniques applied.

Species	Model	Pre-processing	Calibration		Validation	
			Number of PCs or LVs	Classification accuracy (%)	Misclassification rate (%)	Classification accuracy (%)
BWB	QDA	SNV + detrend	6	90.5	9.5	85.2
	PLS-DA	SNV + detrend	4	68.3	31.8	78.4
	QDA	SNV	6	95.2	4.8	74.1
	PLS-DA	SNV+DT+SGd ₂ (9)	4	86.5	13.5	66.7
Zebra	QDA	SNV + detrend	6	93.7	6.3	69.8
	PLS-DA	SNV + detrend	4	66.7	33.3	77.8
	QDA	SNV+DT+SGd ₂ (7)	6	92.9	7.1	75.9
	PLS-DA	SNV	6	75.4	24.6	66.7
Springbok	QDA	SNV + detrend	6	94.2	5.8	81.7
	PLS-DA	SNV + detrend	3	60.0	40.0	63.3
	QDA	SNV+DT+SGd ₂ (7)	6	98.3	1.7	81.7
	PLS-DA	SNV+DT+SGd ₂ (7) SNV+DT+SGd ₂ (9)	5	86.7	13.3	80.0
Ostrich	QDA	SNV + detrend	6	94.2	5.8	76.7
	PLS-DA	SNV + detrend	4	70.8	29.2	68.8
	QDA	SNV+DT+SGd ₂ (9)	6	95.8	4.2	80.0
	PLS-DA	SNV+DT+SGd ₂ (7)	7	90.8	9.2	76.7

(DA) Discriminant analysis; (QDA) Quadratic discriminant analysis; (PLS-DA) Partial least squares discriminant analysis; (PCs) Principal components; (LVs) Latent variables; (BWB) Black wildebeest; (SNV) Standard normal variate; (DT) Detrend; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points).

Table 4.70 An overview of the classification rates of the DA and PLS-DA models for fresh vs. previously frozen meat, irrespective of the frozen period, with various pre-processing techniques applied.

	Model	Pre-processing	Calibration		Validation
			Number of PCs or LVs	Classification accuracy (%)	Misclassification rate (%)
BWB	QDA	SNV + detrend	6	96.7	3.3
	PLS-DA	SNV + detrend	4	96.0	4.0
	QDA	SNV	6	96.0	4.0
	PLS-DA	SNV+DT+SGd ₂ (9)	4	99.2	0.8
Zebra	QDA	SNV + detrend	6	94.5	5.5
	PLS-DA	SNV + detrend	2	88.9	11.1
	QDA	SGd ₁ (7)	6	97.6	2.4
	PLS-DA	SGd ₁ (5)	5	94.4	5.6
Springbok	QDA	SNV + detrend	6	99.2	0.8
	PLS-DA	SNV + detrend	4	97.5	2.5
	QDA	SNV + SGd ₂ (7)	6	99.2	0.8
		SNV + SGd ₂ (9)			
Ostrich	QDA	SNV+DT+SGd ₂ (7)	6	98.3	1.7
		SNV+DT+SGd ₂ (9)			
	PLS-DA	SGd ₂ (9)	5	100.0	0
	QDA	SNV + detrend	6	95.8	4.2
	PLS-DA	SNV + detrend	3	91.7	8.3
	QDA	SNV+DT+SGd ₂ (7)	6	98.3	1.7
		SNV+DT+SGd ₂ (9)			
	PLS-DA	SNV+DT+SGd ₂ (9)	3	98.3	1.7

(DA) Discriminant analysis; (QDA) Quadratic discriminant analysis; (PLS-DA) Partial least squares discriminant analysis; (PCs) Principal components; (LVs) Latent variables; (BWB) Black wildebeest; (SNV) Standard normal variate; (DT) Detrend; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

4.2.4.6 Conclusion

NIR spectroscopy combined with MDA could accurately distinguish between the fresh and previously frozen muscle samples of black wildebeest, zebra, springbok and ostrich, irrespective of the frozen period. The QDA models yielded the best results for differentiating between the different frozen periods. Although these QDA models achieved excellent calibration accuracies the validation accuracies decreased, indicating that the models were over-fitted and not as effective as suggested by the calibration models. Therefore, the results suggested that the discrimination between fresh and previously frozen meat, irrespective of the frozen period, would be better. Thus, improving the classification accuracies of the prediction models. The PLS-DA results showed that it was possible to differentiate between the fresh and previously frozen meat of black wildebeest, zebra, springbok and ostrich, irrespective of the frozen period. The black wildebeest (99.2%), zebra (94.4%), springbok (100%) and ostrich (98.3%) models achieved good overall accuracies. In addition, it was found that the frozen period does not influence the accuracy of the model for fresh or previously frozen meat discrimination.

4.3 Muscle type determination

This data set consisted of three species [zebra (*Equus quagga burchelli*), springbok (*Antidorcas marsupialis*), ostrich (*Struthio camelus*)], multiple muscle types for each species and the meat samples were frozen for one month. The aim of **Section 4.3** was to differentiate between the different types of muscles for each species, irrespective of the treatment (fresh or previously frozen), as well as to determine whether the treatments had an effect on the muscle classification accuracies.

4.3.1 Spectral analysis

The mean spectrum for the different muscles of each species (pre-processed with SNV) was computed between 920 and 1651 nm (**Figure 4.47 – 4.49**) to investigate, determine and compare the chemical properties. The mean spectra of the different muscles followed a similar trend with comparable absorption bands, however the intensity of the bands varied. The intensity differences displayed between the spectra of the muscles can be attributed to the internal chemical composition. Three prominent absorption bands were exhibited at 970, 1193 and 1428 nm. These bands are related to moisture- (Ding & Xu, 2000; Cozzolino & Murray, 2004; Barbin *et al.*, 2013b; Pu *et al.*, 2015; Ma *et al.*, 2016), fat- and protein content (Osborne *et al.*, 1993) as previously discussed in **Section 4.1.1**.

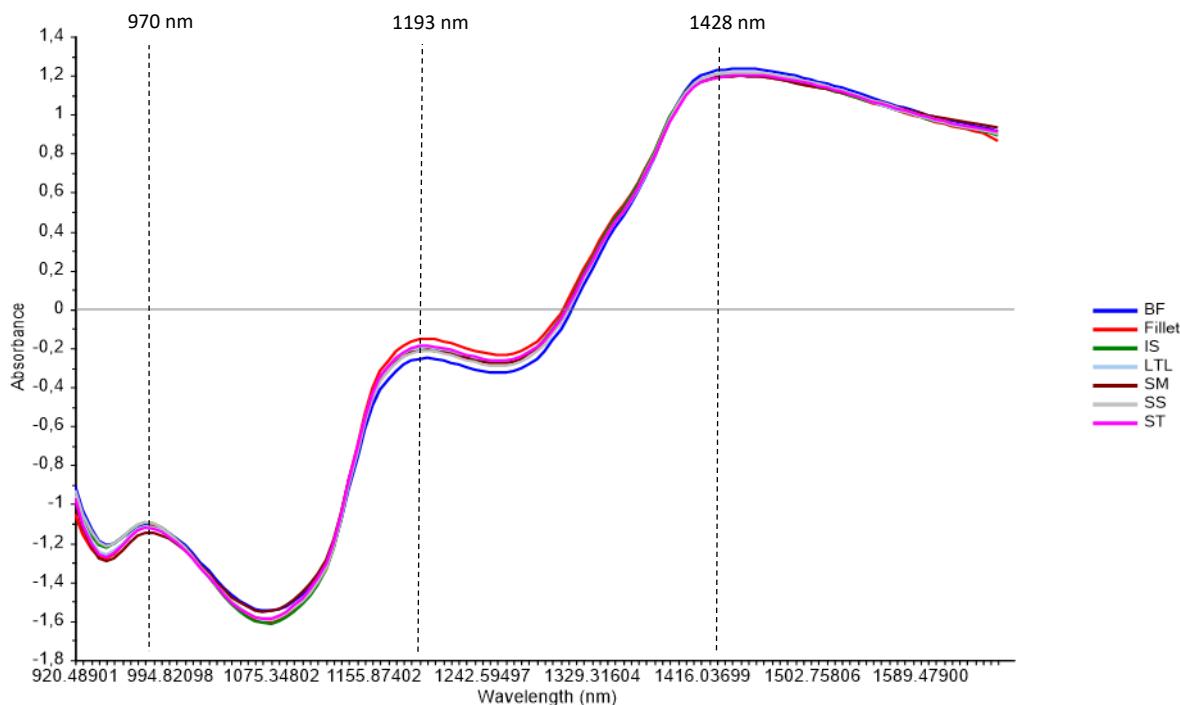


Figure 4.47 Standard normal variate (SNV) pre-processed mean spectra for various zebra muscle types [(BF) biceps femoris (blue), (Fillet) psoas major (red), (IS) infraspinatus (green), (LTL) longissimus thoracis et lumborum (light blue), (SM) semimembranosus (maroon), (SS) supraspinatus (grey), (ST) semitendinosus (purple)], irrespective of the meat being fresh or previously frozen.

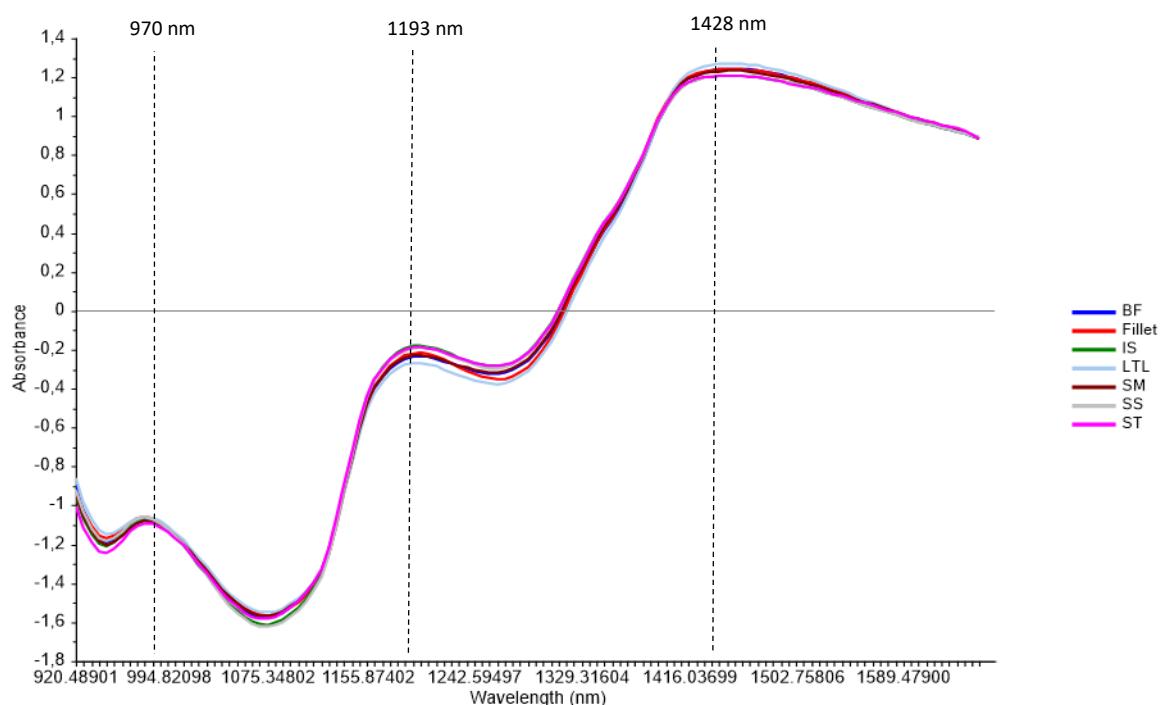


Figure 4.48 Standard normal variate (SNV) pre-processed mean spectra for various springbok muscle types [(BF) biceps femoris (blue), (Fillet) psoas major (red), (IS) infraspinatus (green), (LTL) longissimus thoracis et lumborum (light blue), (SM) semimembranosus (maroon), (SS) supraspinatus (grey), (ST) semitendinosus (purple)], irrespective of the meat being fresh or previously frozen.

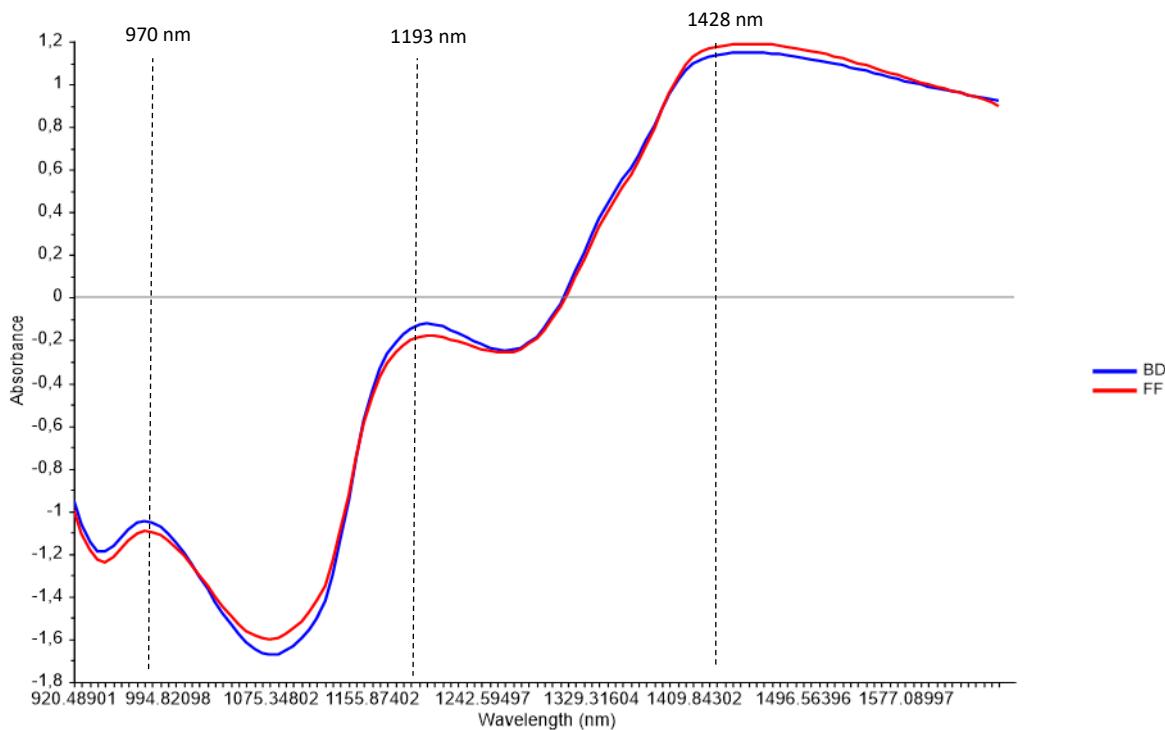


Figure 4.49 Standard normal variate (SNV) pre-processed mean spectra for various ostrich muscle types [(BD) *gastrocnemius* (blue) and (FF) *iliofibularis* (red)], irrespective of the meat being fresh or previously frozen.

4.3.2 Exploratory data analysis (EDA)

4.3.2.1 Principal component analysis (PCA)

Minimal class separation between the different muscle samples was observed in the PCA score plots (**Figure 4.50a – 4.52a**) of the SNV + detrend (4th order polynomial) corrected data. In **Figure 4.50a – 4.52a** a trend was observed in the direction of PC2 (6%; 4%; 16%), illustrating a slight class separation between the different muscles for all three species, however, the samples exhibited an overlap between the individual muscles. Therefore, lacking separation and distinct clustering for the different muscle types.

Lawrie and Ledward (2006) stated that the differences between muscles is very complex, therefore making it difficult to differentiate between the different muscle types. Literature also states that both the dynamic (biochemical) and static (chemical) characteristics of the muscles' composition are intricate (Lawrie & Ledward, 2006). The muscle variability is known to be influenced by a large number of *intrinsic*- (species, breed, sex, age, anatomical location, training/exercise, inter-animal variability) and *extrinsic* factors (food, fatigue, fear, pre-slaughter manipulation, environmental conditions at slaughter), therefore making it difficult to classify muscles based on fixed factors affecting the different muscles' composition (Lawrie & Ledward, 2006).

The variation of muscles within and between species is vast, owing to the biological nature of the samples. However, differences between muscles, concerning moisture-, protein-, fat content, pH and meat tenderness have previously been quantified (Mellet, 1985; Sales, 1996; Onyango *et al.*, 1998; Viljoen, 2003;

Du Buisson, 2006; Hoffman & Wiklund, 2006; Hoffman *et al.*, 2009; Engelbrecht, 2013; North, 2015; Neethling, 2016). Yet, various authors found that different parameters/factors contributed to the observed differences, even when investigating the same species or specific muscles among different species. Therefore, the following results are not substantiated by the literature, as there is no real consensus regarding standard chemical parameters for specific muscles within a species or among species.

The PC2 score plot of zebra accounted for 6% of the variance in the data (**Figure 4.50a**) and the loadings plot (**Figure 4.50b**) exhibited positive bands at 1075, 1366, and 1589 nm as well as negative bands at 963, 1162, 1453 nm. These were associated with moisture- and fat content of lamb (963 nm) (Kamruzzaman *et al.*, 2012b), fat content of beef (1075 nm) (ElMasry *et al.*, 2013), protein content of ham (1162 nm) (Talens *et al.*, 2013) as well as the pH of pork (Barbin *et al.*, 2012b), pH of beef (1366 nm) (ElMasry *et al.*, 2012), meat tenderness of lamb (1453 nm) (Kamruzzaman *et al.*, 2013) and moisture (1589 nm) (Osborne *et al.*, 1993). When investigating the positive and negative loading bands of PC2, in combination with the score plot of PC1 (88%) vs. PC2 (6%) (**Figure 4.50a**), it is evident that the intra-species muscle separation is mainly based on the variation in the moisture-, protein-, fat content and pH as well as meat tenderness.

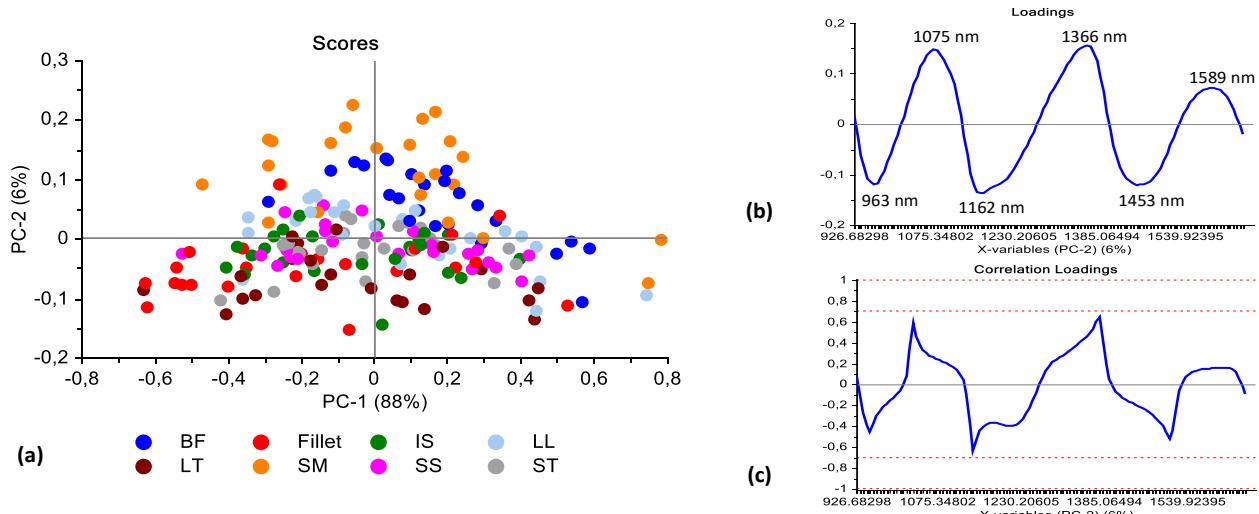


Figure 4.50 PCA analysis (SNV + detrend pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating minimal separation between classes. Scores illustrated as (a) PCA score plot of PC1 (88%) vs. PC2 (6%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with absorption bands at 963, 1075, 1162, 1366, 1453 and 1589 nm.

The PC2 score plot of springbok accounted for 4% of the data variance (**Figure 4.51a**) and the loadings plot (**Figure 4.51b**) exhibited interpretable bands at 1484 nm (negative) and 1639 nm (positive). The band at 1484 nm represents the N-H stretch, first overtone related to the CONH₂ group associated with the peptide bonds in proteins (Osborne *et al.*, 1993). This band explains the muscle separation in the negative direction of PC2. The band at 1639 nm represents the C-H stretch first overtone (Osborne *et al.*, 1993) and corresponds to the pH of meat as stated by Iqbal *et al.* (2013), and explains the muscle separation in the positive direction

of PC2. Although the different muscles exhibit a major overlap, the minimal muscle separation observed within springbok can be attributed to the variation in protein content and pH.

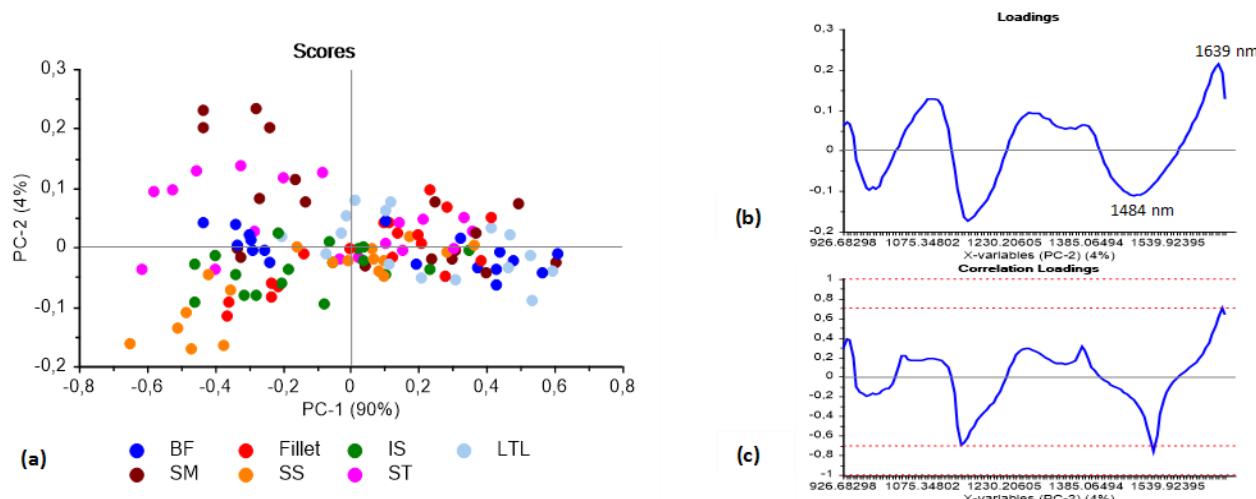


Figure 4.51 PCA analysis (SNV + detrend pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [fresh or previously frozen samples] illustrating minimal separation between classes. Scores illustrated as (a) PCA score plot of PC1 (90%) vs. PC2 (4%). (b) PCA loadings line plot and (c) correlation loadings for PC1 with interpretable bands at 1484 and 1639 nm.

Lastly, the PCA score plot of ostrich (**Figure 4.52a**) in combination with the loadings plot (**Figure 4.52b**) exhibited interpretable positive and negative absorption bands at 1279 and 1372 nm, respectively. The positive band (1279 nm) is associated with fat content (Barbin *et al.*, 2013a) and illustrates the separation of the *iliofibularis* (FF) muscle in the positive direction of PC2. The negative band (1372 nm) illustrates the separation of the *gastrocnemius* (BD) muscle in the negative direction of PC2 and is associated with moisture content (ElMasry *et al.*, 2013) as well as pH (ElMasry *et al.*, 2012).

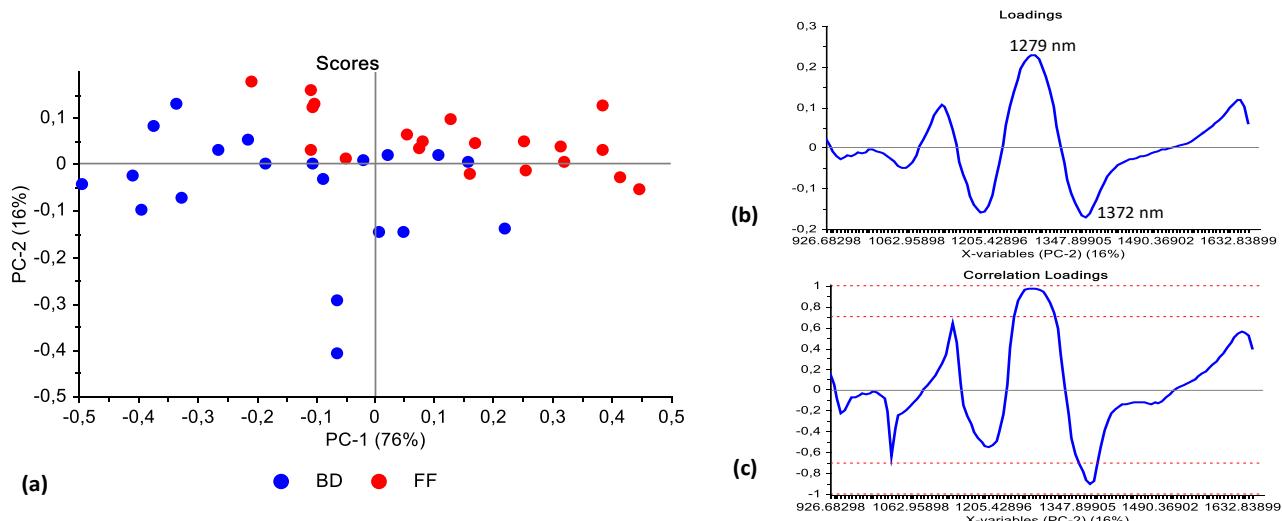


Figure 4.52 PCA analysis (SNV + detrend pre-processed) of ostrich (BD, FF) [fresh and previously frozen samples] illustrating a slight overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (76%) vs. PC2 (16%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 1279 and 1372 nm.

From these results, it was possible to conclude that the internal variation (muscle types) can be attributed to differences in moisture-, fat-, protein content and pH as well as meat tenderness. However, no correlation could be drawn between the specific factors and different muscles, within a species and among species. This can be attributed to the multitude of factors contributing to the separation for each species and muscle type. PCA gave a global representation of the data, thus the need for further exploratory data analysis to show the specific differences in the spectral data.

4.3.2.2 Difference spectra

The Savitzky-Golay treated spectra of the different muscle samples for each species were found to follow a similar trend with comparable absorption bands, but the intensity of the bands varied. The difference spectra of zebra and springbok illustrated that the separation between muscle types is a result of absorption bands at 976, 1162, 1335, 1403 and 1626 nm (**Figure 4.53 – 4.54**). The difference spectra of ostrich (**Figure 4.55**) illustrated that the separation was due to similar and different absorption bands exhibited at 976, 1162, 1341, 1403 and 1639 nm.

The band at 976 nm (**Figure 4.53 – 4.55**) represents the second overtone of O-H stretching vibrations (Osborne *et al.*, 1993) and was ascribed to the water content of ham (Talens *et al.*, 2013). The spectral differences between the muscles can therefore be attributed to the higher moisture content of fillet (zebra and springbok), ST (springbok) and BD (ostrich). The mean spectra of springbok also illustrate that there was little difference between the moisture contents of fillet and ST due to similar intensities of the absorption bands. This indicates that both muscles have similar moisture contents; however, it was difficult to support this with literature. The springbok spectra illustrate that the moisture content of the LTL and BF muscles are similar, but when compared to other muscles, the moisture is seen to be lower. This phenomenon was supported by Du Buisson (2006) who reported similar findings. The increased moisture content observed for the SM and SS muscles also followed a similar trend (Du Buisson, 2006). Relevant literature regarding zebra, and the specific muscles in question, is sparse, making it problematic to explain the observed differences. Because zebra and horse form part of the same *Equidae* family known as the genera *Equus*, it was assumed that their muscles would have similar physical and chemical characteristics, therefore the zebra spectra were compared to literature on horse meat. The spectra for zebra (**Figure 4.53**) illustrate that the BF and SM muscles had the lowest moisture contents. These results were supported by two studies on different horse breeds. Tateo *et al.* (2008) (Italian Heavy Draft) as well as Lorenzo and Pateiro (2013) (Galician Mountain), both reported that the BF and SM muscles had the lowest moisture contents, therefore attributing to some of the differences observed between the muscles. Lastly, the difference observed between the BD and FF muscles for ostrich (**Figure 4.55**) illustrates that the BD muscle has a higher moisture content than the FF muscle. Sales (1996) supports this, as his results also reported a higher moisture content for BD than FF.

Barbin *et al.* (2012b) and ElMasry *et al.* (2012) found that the bands at 1162, 1335 and 1341 nm (second and first stretch overtones and C-H combination/deformation bands) can be related to pH and is

therefore, responsible for the spectral differences at these absorption bands. The spectral differences observed in **Figure 4.53** and **Figure 4.54** corresponds to the pH differences as reported by Du Buisson (2006) (springbok) and Lorenzo *et al.* (2013) (horse), with the exception of ostrich (Sales, 1996). The spectra illustrates that the pH of BD is higher than that of FF, however, this does not correspond to literature as Sales (1996) reported the FF muscle to have a higher pH. For the zebra spectra it was observed that the fillet and IS muscles resulted in the highest pH and the SM and LTL muscles in the lowest. This trend in pH is supported by literature, as Lorenzo *et al.* (2013) observed similar pH differences for the different muscles of horse (Galician Mountain). Also, the pH differences observed in the spectra of springbok follow a similar trend as reported by Du Buisson (2006). The band at 1341 nm can also be associated with the water-holding capacity of beef (ElMasry *et al.*, 2011), but no literature could support this for these species and muscle types.

A clear moisture band is observed, due to O-H stretching and O-H bending combinations, at 1403 nm (Osborne *et al.*, 1993; Liu & Chen, 2001) and was linked to the pH of beef (ElMasry *et al.*, 2012). Lastly, the bands at 1626 and 1639 nm represents the C-H first overtone stretch and is associated with lamb meat tenderness (Kamruzzaman *et al.*, 2013), as well as protein content of beef (ElMasry *et al.*, 2013). Although both spectra for zebra and springbok illustrated differences between the muscles at this waveband (1626 nm), it was not possible to relate these differences to specific values reported in literature. The differences observed in the ostrich spectra at 1639 nm, attributed to protein, corresponded with the results reported by Mellet (1985), but differed from the results reported by Sales (1996). Mellet (1985) reported the BD muscle to have a slightly higher protein content than that of FF, therefore matching the increased absorption intensity of BD exhibited in **Figure 4.55**. The tenderness could however not be corroborated with literature. Sales (1996) reported that the BD muscle was more tender than the FF muscle, whereas the spectra in the current study illustrates the opposite.

The results from the spectra indicated that the physiochemical characteristics of some of the species and muscles were similar to that of previous studies (Mellet, 1985; Sales, 1996; Onyango *et al.*, 1998; Du Buisson, 2006; Tateo *et al.*, 2008; Lorenzo & Pateiro, 2013; Lorenzo *et al.*, 2013; reviewed by Lorenzo *et al.*, 2014). Therefore, these results indicate that the differences and separation between different muscle types can be explained by examining the physiochemical characteristics of the species and specific muscles in question e.g. moisture, protein, pH and potentially meat tenderness (Fumière *et al.*, 2000; Alomar *et al.*, 2003; Kamruzzaman *et al.*, 2011). Thus, indicating that the specific muscles in question, irrespective of the treatment (fresh or previously frozen), has the potential to be successfully differentiated within each species.

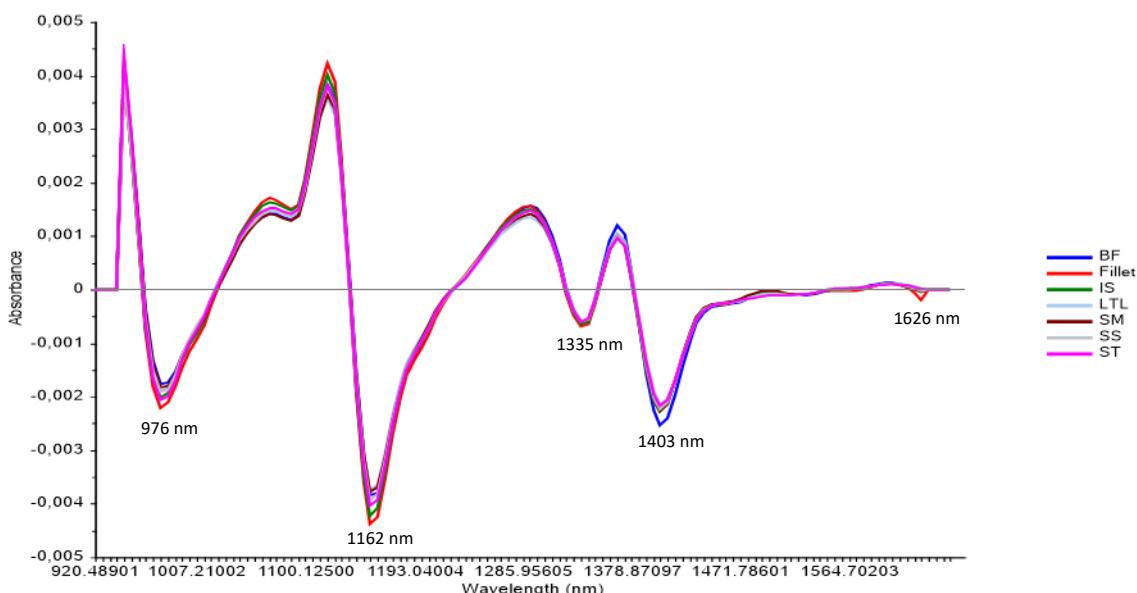


Figure 4.53 Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 smoothing points) treated mean spectra of zebra illustrating the spectral differences between the six muscle types [(BF) *biceps femoris* (blue), (Fillet) *psoas major* (red), (IS) *infraspinatus* (green), (LTL) *longissimus thoracis et lumborum* (light blue), (SM) *semimembranosus* (maroon), (SS) *supraspinatus* (grey), (ST) *semitendinosus* (purple)], with absorption bands at 976, 1162, 1335, 1403 and 1626 nm.

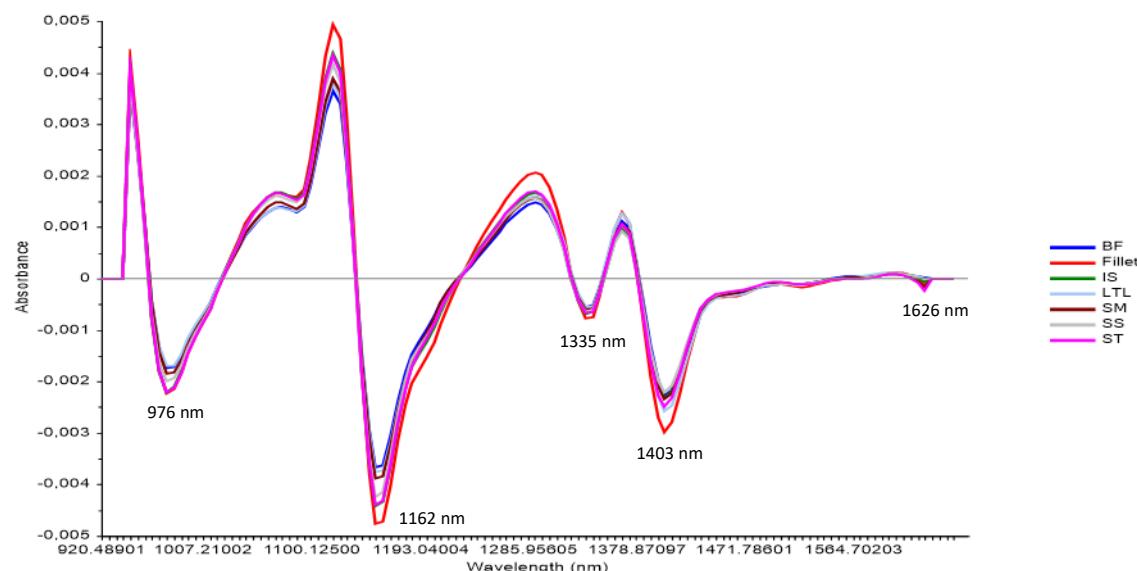


Figure 4.54 Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 smoothing points) treated mean spectra of springbok illustrating the spectral differences between the six muscle types [(BF) *biceps femoris* (blue), (Fillet) *psoas major* (red), (IS) *infraspinatus* (green), (LTL) *longissimus thoracis et lumborum* (light blue), (SM) *semimembranosus* (maroon), (SS) *supraspinatus* (grey), (ST) *semitendinosus* (purple)], with absorption bands at 976, 1162, 1335, 1403 and 1626 nm.

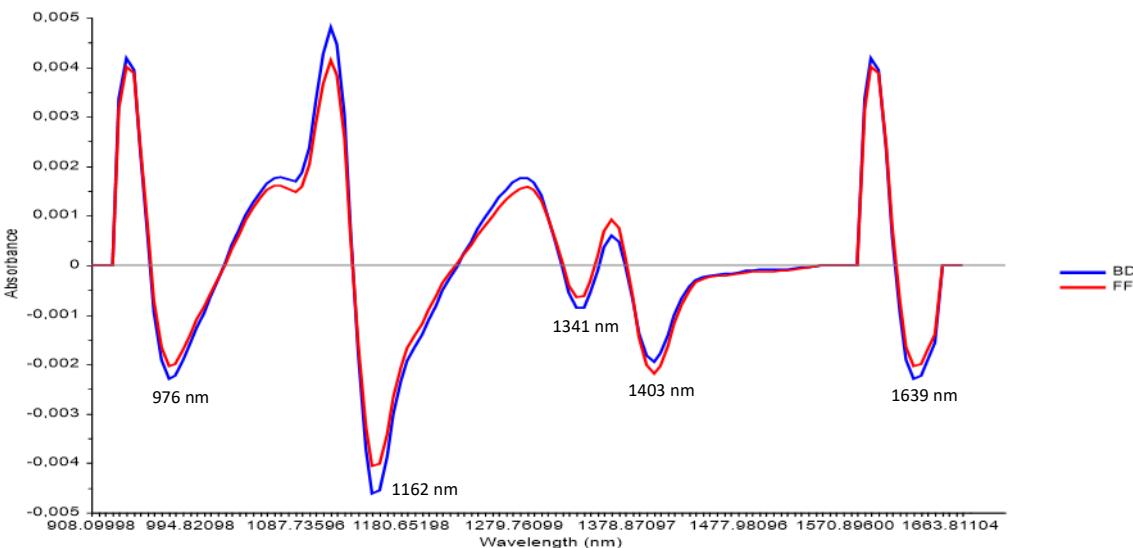


Figure 4.55 Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 smoothing points) treated mean spectra of ostrich illustrating the spectral differences between the two muscle types [(BD) *gastrocnemius* (blue) and (FF) *iliofibularis* (red)], with absorption bands at 976, 1162, 1341, 1403 and 1639 nm.

4.3.3 Multivariate data analysis (MDA): Model development

4.3.3.1 Soft independent modelling of class analogy (SIMCA)

The overall SIMCA model accuracy (**Table 4.71**) of zebra (27.3%), springbok (38.4%) and ostrich (88.2%) illustrated that the ostrich model gave excellent results. These results can be explained by investigating the performance measures (**Table 4.72**) of the different classification models.

Table 4.71 SIMCA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for muscle type classification.

Species	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
Zebra	Calibration	27.3	72.7
	Validation	21.3	78.7
Springbok	Calibration	38.4	61.6
	Validation	25.0	27.0
Ostrich	Calibration	88.2	11.8
	Validation	83.3	16.7

(SIMCA) Soft independent modelling of class analogy; (SNV) Standard normal variate.

The SIMCA model for ostrich presented the best classification results. The calibration model achieved an overall classification accuracy of 88.2% (**Table 4.71**). This suggests that the model can accurately distinguish between the two ostrich muscles. The model also revealed that the classification of the BD and FF muscles was all predicted with a high level of accuracy. The classification accuracy (88.2%), sensitivity (>82%) and specificity (>84%) in **Table 4.72** confirmed this. The PCA score plot and loadings line plot (**Figure**

4.52a – b) support these results, as good separation with a slight overlap was observed between the two muscles (BD and FF). The separation and correct classification of the two classes was mainly attributed to moisture, fat, protein and pH. These properties differed for each muscle and a correlation was observed in the mean spectra (**Figure 4.55**) and loadings (**Figure 4.52b**). The physical and chemical characteristics of the different muscle types can thus be used as important predictors. Therefore, these results compare to previous studies done on chicken, beef and lamb meat, where researchers found that both the physical and chemical differences can be used to successfully discriminate between different muscle types or meat cuts (Mitsumoto *et al.*, 1991; Fumière *et al.*, 2000; Alomar *et al.*, 2003; Kamruzzaman *et al.*, 2011).

The SIMCA models for zebra and springbok gave unsatisfactory results and confirmed what was seen in the PCA score plots (**Figure 4.50a** and **Figure 4.51a**). The lack of separation can be ascribed to the samples' spectral similarities, and as the SIMCA algorithm aims to classify samples based on spectral similarities (Massart *et al.*, 1988), the classification accuracy was low. Therefore, a small distance means a high similarity. When comparing these model results, it is observed that the springbok model achieved a higher classification accuracy (38.4%) than the zebra model (27.3%) and also exhibited improved performance measures (**Table 4.72**). These results are supported and can be explained by the mean spectra (**Figure 4.53 – 4.54**). Although the separation between the different muscles, for both species, is mainly attributed to the moisture content and pH, the intensity of these absorption bands differed. The mean spectra for springbok (**Figure 4.54**) exhibited a larger difference between the absorption bands of the different muscle types, with higher intensities, compared to the bands exhibited in the spectra for zebra (**Figure 4.53**). This increased waveband separation illustrates that there was a larger difference between these muscles, therefore achieving better class separation. The decreased waveband separation between the zebra muscles illustrates that there is little difference between these muscles. This suggests that the spectra are similar and for this reason, the SIMCA model for zebra achieved a lower classification accuracy. Therefore, the SIMCA results for zebra and springbok indicates that the different muscles cannot be differentiated because of their spectral similarities.

Table 4.72 The performance measures used to assess the SIMCA classification models for different muscle types of the three species, pre-processed with SNV + detrend.

Species	Treatment	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
Zebra	BF	63.2	11.0	25.9	24.4	67.9	43.2	31.2	36.8
	Fillet	64.8	16.1	19.3	34.9	73.8	39.0	36.8	35.4
	IS	66.1	11.0	22.9	27.5	71.4	44.2	33.9	33.9
	LTL	56.2	17.6	26.2	20.7	65.4	28.8	23.8	43.8
	SM	60.3	33.1	6.7	55.6	88.6	20.2	29.6	39.8
	SS	57.4	27.9	14.7	27.5	77.8	16.7	20.7	42.6
	ST	58.1	23.8	18.2	23.7	74.3	19.2	21.2	41.9
Springbok	BF	72.7	10.2	17.2	35.3	78.6	48.0	40.7	27.3
	Fillet	72.1	10.1	17.8	42.5	76.8	56.7	48.6	27.9
	IS	68.9	12.6	18.5	32.4	76.4	41.4	36.4	31.1
	LTL	61.6	19.2	19.2	34.1	72.9	34.1	34.1	38.4
	SM	82.3	4.4	13.3	44.4	84.4	70.6	54.6	17.7
	SS	58.9	31.0	10.1	36.0	84.0	15.5	21.7	41.1
	ST	68.9	17.0	14.1	45.7	80.2	41.0	43.2	31.1
Ostrich	BD	88.2	8.8	2.9	94.1	93.3	84.2	88.9	11.8
	FF	88.2	2.9	8.8	82.4	84.2	93.3	87.5	11.8

(SIMCA) Soft independent modelling of class analogy; (SNV) Standard normal variate; (BF) *biceps femoris*; (Fillet) *psoas major*; (IS) *infraspinatus*; (LTL) *longissimus thoracis et lumborum*; (SM) *semimembranosus*; (SS) *supraspinatus*; (ST) *semitendinosus*; (BD) *gastrocnemius*; (FF) *iliofibularis* muscles.

4.3.3.2 K-nearest neighbour (KNN)

Table 4.73 KNN model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for muscle type classification.

Species	Number of neighbours (<i>k</i>)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
Zebra	2	Calibration	43.1	56.9
		Cross-validation	40.2	59.8
		Validation	31.3	68.8
	3	Calibration	42.4	57.6
		Cross-validation	39.2	60.8
		Validation	33.3	66.7
	5	Calibration	41.7	58.3
		Cross-validation	40.1	59.9
		Validation	37.5	62.5
Springbok	2	Calibration	40.9	59.1
		Cross-validation	40.5	59.5
		Validation	38.7	61.3
	3	Calibration	37.6	62.4
		Cross-validation	37.6	62.4
		Validation	38.7	61.3
	5	Calibration	45.2	54.8
		Cross-validation	39.5	60.5
		Validation	35.5	64.5
Ostrich	2	Calibration	70.0	30.0
		Cross-validation	68.8	31.2
		Validation	70.0	30.0
	3	Calibration	76.7	23.3
		Cross-validation	72.8	27.2
		Validation	70.0	30.0
	5	Calibration	83.3	16.7
		Cross-validation	82.1	17.9
		Validation	80.0	20.0

(KNN) *K*-nearest neighbours; (SNV) Standard normal variate.

The overall model accuracy of zebra, springbok and ostrich suggested that two nearest neighbours would provide the best classification for zebra (43.1%), whereas five nearest neighbours provided the best classification for springbok (45.2%) and ostrich (83.3%) (Table 4.73). The KNN(5) calibration model for ostrich achieved an overall classification- and cross-validation accuracy of 83.3% and 82.1%, with a slightly decreased classification accuracy for the validation model (80%) (Table 4.73). The decreased validation accuracy is indicative of a classification model that is over-fitted (Miller, 2005). Although this model had good accuracies, a few misclassified objects were observed amongst the different muscles. This accounted for a misclassification rate of 16.7% for the overall model (Table 4.73). The sensitivity and specificity for the FF

muscle were both >85% (**Table 4.74**). This indicates that the model has a high probability of correctly classifying the FF muscle (**Figure 4.74**). The lower sensitivity (81.3%) for the BD muscle reveals that the model was less suited for predicting this class. This phenomenon can be explained by referring to the PCA score plot (**Figure 4.56**), as *k*-nearest neighbour was performed on the PC scores. The KS-calibration PCA score plot illustrates a slight overlap between the two classes, with a larger number of BD muscles displaying a close distance to the FF muscles. Therefore, the BD muscles are assigned to the predominant class, FF. Hence, explaining why the model is less suited for predicting the BD muscles.

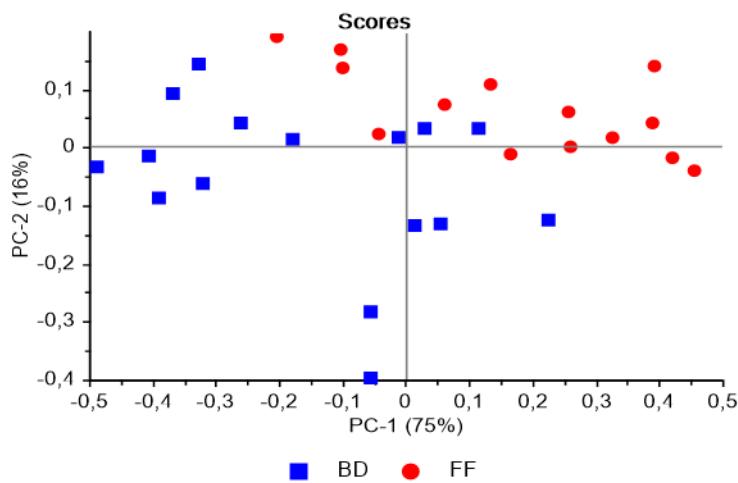


Figure 4.56 PCA analysis (SNV + detrend pre-processed) of ostrich muscles [fresh and previously frozen samples] illustrating a slight overlap between BD (blue) and FF (red) classes. Scores illustrated as KS-calibration PCA score plot of PC1 (75%) vs. PC2 (16%).

The decreased model accuracy for both zebra [KNN(2)] and springbok [KNN(5)] is mainly attributed to the misclassification of objects assigned to the different muscle classes. These classes were misclassified due to the extensive overlap observed in the PCA score plots (**Figure 4.57a – b**). The KNN(2) model for zebra exhibited that the BF muscle achieved the highest classification accuracy. This is confirmed by the sensitivity (89.5%) and specificity (95.7%). On the other hand, the performance measures for the remaining muscles exhibits sensitivity values between 7 and 65%, while the specificity is between 80 and 85%. Therefore, confirming the model's ability to accurately predict the BF muscle even though it struggles with the remaining muscles. Once again this can be ascribed to the larger number of the remaining muscles displaying a close distance to the BF muscles (**Figure 4.57a**), as previously discussed. A similar phenomenon is observed in KNN(5) for springbok, except this model had a higher affinity to accurately predict the LTL muscle, where the sensitivity is 60% (LTL) and 8 - 58% (remaining muscles), and the specificity is 84.6% (LTL) and 78 - 87% (remaining muscles) (**Table 4.57b**). The classification models for zebra and springbok exhibited higher accuracies for both the calibration and cross-validation compared to the lower validation accuracies, suggesting that the models are over-fitted (Miller, 2005).

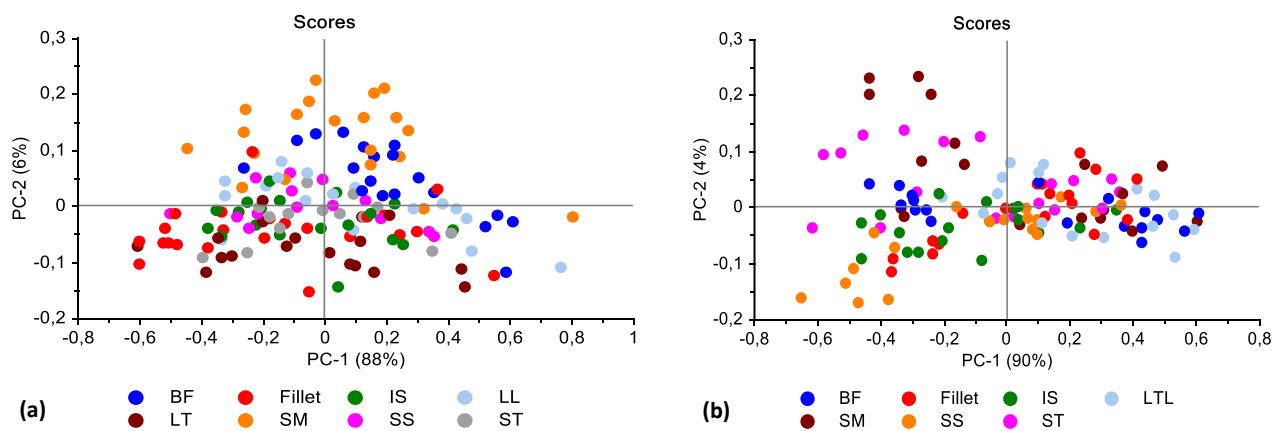


Figure 4.57 PCA analysis (SNV + detrend pre-processed) of (a) zebra (LTL, BF, SM, ST, IS, SS, fillet) and (b) springbok (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating minimal separation between the different muscle classes. Scores illustrated as (a) KS-calibration PCA score plot of PC1 (88%) vs. PC2 (6%) for zebra, and (b) KS-calibration PCA score plot of PC1 (90%) vs. PC2 (4%) for ostrich.

The KNN model for ostrich (83.3%) presented the best results followed by springbok (45.2%) and zebra (43.1%). These results compare to the SIMCA results, where similar findings were observed. Therefore, the improved classification of the different muscle types for ostrich can be attributed to a greater degree of physical and chemical differences that was observed within the muscles. Unfortunately, these KNN results cannot be compared to literature because limited research has been done on the differentiation between muscle types and the few studies available have only considered LDA and PLS-DA. These KNN results for zebra and springbok are in fact worse than the results reported in literature (Mitsumoto *et al.*, 1991; Fumi  re *et al.*, 2000; Alomar *et al.*, 2003; Kamruzzaman *et al.*, 2011). However, it was not possible to conclude whether the results herein achieved worse accuracies due to the classification method used, or due to the number of muscle types used, as previous studies have focused on differentiating between a maximum of three muscle types. If the number of muscle types used did however influence the results, it would explain why the ostrich achieved sufficient accuracies as only two muscle types were analysed. The literature also indicated that differences in moisture-, fat-, protein content, pH and meat tenderness (shear force value) contributed to the successful differentiation and prediction of different muscle types (Mitsumoto *et al.*, 1991; Fumi  re *et al.*, 2000; Alomar *et al.*, 2003; Prieto *et al.*, 2009; Kamruzzaman *et al.*, 2011). Therefore, the KNN results in the current study indicates that different ostrich muscles, irrespective of the treatment (fresh or previously frozen), can successfully be discriminated because of the substantial difference between these muscles. On the other hand, the KNN results for zebra and springbok indicates that their different muscles cannot be differentiated because of spectral similarities.

Table 4.74 The performance measures used to assess the KNN classification models for different muscle types of the three species, pre-processed with SNV + detrend.

Species	Number of neighbours (k)	Treatment	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
Zebra	2	BF	72.9	24.7	2.4	89.5	95.7	44.7	59.7	27.1
		Fillet	73.8	16.7	9.5	65.2	85.5	51.7	57.7	26.2
		IS	69.7	18.0	12.4	42.1	83.1	33.3	37.2	30.3
		LTL	73.3	13.6	13.1	37.3	83.5	40.7	37.0	26.7
		SM	80.5	1.3	18.2	30.0	80.0	85.7	44.4	19.5
		SS	78.5	5.1	16.5	7.1	82.4	20.0	10.5	21.5
Springbok	5	BF	66.7	23.8	9.5	50.0	85.7	28.6	36.4	33.3
		Fillet	79.3	7.6	13.2	58.8	82.1	71.4	64.5	20.8
		IS	73.7	17.5	8.8	58.3	87.5	41.2	48.3	26.3
		LTL	79.3	9.4	11.3	60.0	84.6	64.3	62.1	20.8
		SM	73.7	7.0	19.3	8.3	78.9	20.0	11.8	26.3
		SS	80.8	5.8	13.5	22.2	85.1	40.0	28.6	19.2
		ST	68.9	16.4	14.8	43.8	79.6	41.2	42.4	31.2
Ostrich	5	BD	83.3	6.7	10.0	81.3	80.0	86.7	83.9	16.7
		FF	83.3	10.0	6.7	85.7	86.7	80.0	82.8	16.7

(KNN) K-nearest neighbours; (SNV) Standard normal variate; (BF) *biceps femoris*; (Fillet) *psoas major*; (IS) *infraspinatus*; (LTL) *longissimus thoracis et lumborum*; (SM) *semimembranosus*; (SS) *supraspinatus*; (ST) *semitendinosus*; (BD) *gastrocnemius*; (FF) *iliofibularis* muscles.

4.3.3.3 Discriminant analysis (DA)

The linear discriminant analysis (LDA) model for ostrich (6 PCs) presented the best discrimination results. The separation between the different muscle types was achieved with a 100% overall accuracy (**Table 4.75**). The performance measures also revealed that the discrimination of both the BD and FF muscles were all correctly predicted. The classification accuracy, sensitivity, specificity and F1 score for both classes were 100% (**Table 4.76**), which means that the model can accurately and with ease distinguish between the two muscle types.

Table 4.75 Optimal DA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for muscle type discrimination.

Species	Number of principal components (PCs)	Method (distance calculations)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
Zebra	5	Linear	Calibration	63.2	36.8
			Validation	50.0	50.0
		Quadratic	Calibration	72.2	27.8
			Validation	52.1	47.9
			Calibration	68.8	31.3
			Validation	50.0	50.0
		Mahalanobis	Calibration	50.5	49.5
			Validation	12.9	87.1
			Calibration	81.7	18.3
			Validation	41.9	58.1
			Calibration	72.0	28.0
			Validation	32.3	67.7
Springbok	5	Linear	Calibration	100.0	0
			Validation	100.0	0
		Quadratic	Calibration	100.0	0
			Validation	90.0	10.0
			Calibration	100.0	0
			Validation	90.0	10.0
		Mahalanobis	Calibration	100.0	0
			Validation	100.0	0
			Calibration	100.0	0
			Validation	100.0	0
			Calibration	100.0	0
Ostrich	6	Linear	Calibration	100.0	0
			Validation	100.0	0
		Quadratic	Calibration	100.0	0
			Validation	90.0	10.0
			Calibration	100.0	0
			Validation	90.0	10.0
		Mahalanobis	Calibration	100.0	0
			Validation	100.0	0
			Calibration	100.0	0
			Validation	100.0	0
			Calibration	100.0	0

(DA) Discriminant analysis; (SNV) Standard normal variate.

The quadratic discriminant analysis (QDA) models for zebra (5 PCs) and springbok (5 PCs) achieved good overall accuracies. The QDA model for zebra achieved an overall classification accuracy of 72.2% for the calibration and 52.1% for the validation model, with misclassification rates of 27.8 and 47.9% (**Table 4.75**). The decreased validation accuracy is indicative of a model that is over-fitted (Miller, 2005). On the other hand, the QDA model for springbok achieved an overall classification accuracy 81.7% for the calibration set with a misclassification rate of 18.3%. Although the model exhibited a good overall calibration accuracy as well as good performance measures for each treatment (**Table 4.76**), the decreased validation accuracy

(41.9%) suggests that the model was over-fitted (Miller, 2005). Consequently, the over-fitted models have a low probability of correctly predicting the different muscle classes when using an independent validation set.

In general, it is observed that the LDA model achieved the best classification results for ostrich. While, the QDA models exhibited better discrimination results for the zebra and springbok. The LDA model for ostrich achieved better results because the algorithm was able to calculate an optimal linear projection between the two classes, due to the good separation illustrated in the PCA score plot (**Figure 4.52a**). The cluster overlap observed in the PCA score plots for zebra (**Figure 4.50a**) and springbok (**Figure 4.51a**), explains why the LDA algorithm was unable to calculate an optimal linear projection between the multiple different muscle types. Moreover, because the QDA algorithm calculates a non-linear decision boundary, using a quadratic function, it was able to separate the clusters and successfully discriminate between the muscles. These results for ostrich (100%), zebra (72.2%) and springbok (81.7%) are comparable to results reported by Kamruzzaman *et al.* (2011). These researchers were able to discriminate between three types of lamb muscles [*semitendinosus* (ST), *longissimus dorsi* (LD), *psoas major* (PM)] with an 81.9 – 100% accuracy. The results in the current study therefore show that DA can be used to discriminate between different muscle types of springbok, zebra and ostrich, irrespective of the treatment (fresh or previously frozen).

Table 4.76 The performance measures used to assess the DA models for different muscle types of the three species, pre-processed with SNV + detrend.

Species	Number of principal components (PCs)	Method	Treatment	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
Zebra	5	QDA	BF	90.4	6.1	3.5	79.0	95.7	68.2	73.2	9.6
			Fillet	92.9	1.8	5.4	73.9	93.6	89.5	81.0	7.1
			IS	93.7	3.6	2.7	84.2	96.7	80.0	82.1	6.3
			LTL	89.8	5.6	4.7	69.0	94.5	64.6	66.7	10.2
			SM	92.0	2.7	5.3	70.0	93.8	82.4	75.7	8.0
			SS	92.0	2.7	5.3	57.1	94.1	72.7	64.0	8.0
			ST	89.7	6.9	3.5	71.4	95.9	55.6	62.5	10.3
Springbok	5	QDA	BF	91.6	8.4	0	100.0	100.0	63.2	77.4	8.4
			Fillet	96.2	1.3	2.5	88.2	96.8	93.8	90.9	3.8
			IS	98.7	1.3	0	100.0	100.0	92.3	96.0	1.3
			LTL	89.4	4.7	5.9	66.7	93.0	71.4	69.0	10.6
			SM	93.8	1.2	4.9	66.7	94.4	88.9	76.2	6.2
			SS	93.8	2.5	3.7	66.7	95.9	75.0	70.6	6.2
			ST	95.0	1.3	3.8	81.3	95.5	92.9	86.7	5.0
Ostrich	6	LDA	BD	100.0	0	0	100.0	100.0	100.0	100.0	0
			FF	100.0	0	0	100.0	100.0	100.0	100.0	0

(DA) Discriminant analysis; (SNV) Standard normal variate; (QDA) Quadratic discriminant analysis; (LDA) Linear discriminant analysis; (BF) *biceps femoris*; (Fillet) *psoas major*; (IS) *infraspinatus*; (LTL) *longissimus thoracis et lumborum*; (SM) *semimembranosus*; (SS) *supraspinatus*; (ST) *semitendinosus*; (BD) *gastrocnemius*; (FF) *iliofibularis* muscles.

4.3.3.4 Partial least squares discriminant analysis (PLS-DA)

The PLS-DA model for ostrich gave excellent discrimination results, while the models for zebra and springbok were unsatisfactory. The PLS-DA calibration models achieved overall classification accuracies of 67.4% (zebra), 50.5% (springbok), 100% (ostrich) and misclassification rates of 32.6% and 49.5% for zebra and springbok, respectively (**Table 4.77**). The PLS-DA prediction score plots for the three species are given in **Figure 4.58**. The 3D score plots for zebra [LV1 (86.8%) vs. LV2 (6.84%) vs. LV3 (2.87%)] and springbok [LV1 (89.09%) vs. LV2 (4.45%) vs. LV3 (3.06%)] (**Figure 4.58a – b**) exhibits a major overlap between the muscle classes, which is indicative of an unsatisfactory model calibration. While the 3D score plot for ostrich [LV1 (73.49%) vs. LV2 (18.01%) vs. LV3 (4.6%)] (**Figure 4.58c**) demonstrates minimal overlap between the muscle classes, indicative of a satisfactory model calibration. The score plot of LV1 (73.49%) vs. LV2 (18.01%) vs. LV3 (4.6%) of ostrich (**Figure 4.58c**) exhibited the best muscle separation, with two prominent clusters. The separation between the BD and FF muscles were best described in LV1, while LV2 and LV3 accounted for little class separation. The BD muscles were predominantly associated with the negative scores in LV1 along with the positive and negative score in LV2 and LV3. On the other hand, the FF muscles associated with the positive scores in LV1 and positive scores in LV2 and LV3. Therefore, the score plots (**Figure 4.58a – c**) exhibited that muscle separation is best described in the direction of both LV1, LV2 and LV3.

Table 4.77 PLS-DA model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for muscle type discrimination.

Species	Number of latent variables (LVs)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
Zebra	6	Calibration	67.4	32.6
		Cross-validation	60.5	39.5
		Validation	52.1	47.9
Springbok	4	Calibration	50.5	49.5
		Cross-validation	38.4	61.6
		Validation	22.6	77.4
Ostrich	5	Calibration	100.0	0
		Cross-validation	100.0	0
		Validation	100.0	0

(PLS-DA) Partial least squares discriminant analysis; (SNV) Standard normal variate.

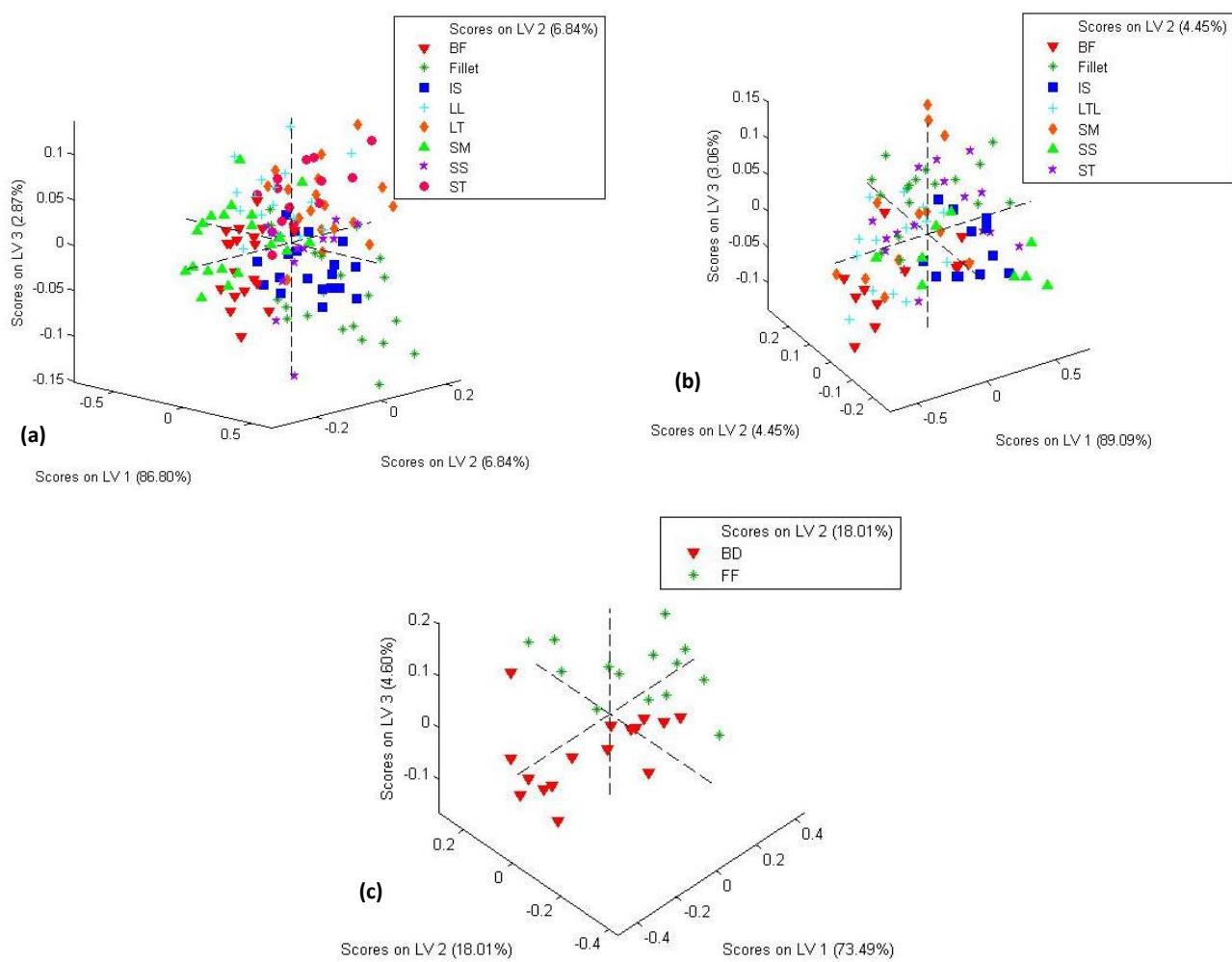


Figure 4.58 PLS-DA (SNV + detrend pre-processed) (a) score plot of LV1 (86.80%) vs. LV2 (6.84%) vs. LV3 (2.87%) for zebra, (b) score plot of LV1 (89.09%) vs. LV2 (4.45%) vs. LV3 (3.06%) for springbok, (c) score plot of LV1 (73.49%) vs. LV2 (18.01%) vs. LV3 (4.6%) for ostrich, colour coded per muscle type class.

The PLS-DA prediction score plot (**Figure 4.59 c**) for ostrich illustrates that the BD and FF muscles are all correctly predicted and the performance measures in **Table 4.78** confirms this. The PLS-DA prediction score plots for zebra (**Figure 4.59a** and **Addendum C, Figure C1a – g**) and springbok (**Figure 4.59b** and **Addendum C, Figure C2a – f**) illustrate a large overlap between the muscles, resulting in less correctly predicted muscle types. This was also confirmed by the performance measures calculated in **Table 4.78**. The ostrich model resulted in the highest classification accuracy (100%), followed by zebra (67.4%) and springbok (50.5%). The ostrich model exhibited excellent performance measures for both muscle types (BD and FF) (**Table 4.78**). The spectral classification of the different muscle types for zebra, springbok and ostrich can be attributed to their physiochemical differences as discussed in **Section 4.3.2**. The individual muscle prediction accuracies for zebra were very similar, with the BF muscle achieving the highest accuracy and best performance measures (**Table 4.78**). Although the zebra BF muscle exhibits a classification accuracy of 91.5%, the sensitivity (79%) and specificity (95.4%) suggests that the model has a higher probability of predicting the remaining muscles, due to increased specificity and decreased sensitivity when compared to

the classification accuracy. This trend was observed in all the muscle types. The overlap observed between the muscles (**Figure 4.59a** and **Addendum C, Figure C1a – g**) can be attributed to the close spectral relationship, due to certain similar physiochemical characteristics as discussed in **Section 4.3.2**. From these results it was evident that the model has a high probability of correctly predicting the individual muscles, with a high level of accuracy and minor misclassification rate (<12%) (**Table 4.78**). The individual muscle prediction accuracies and performance measures for springbok, suggested that the model was more suited to predict the fillet muscle. This muscle achieved a classification accuracy of 79.7% and as with the zebra, the sensitivity (76.5%) is lower and the specificity (89.5%) higher compared to the classification accuracy. The same was observed for all the muscle types, therefore suggesting that the model has a higher probability of predicting the remaining muscles rather than the specific one in question. As previously mentioned, the overlap between the muscles (**Figure 4.59b** and **Addendum C, Figure C2a – f**) can be attributed to the close spectral relationship of the muscles due to certain similar physiochemical characteristics as discussed in **Section 4.3.2**. From these results it was possible to conclude that the model has a lower probability of correctly predicting the individual muscles, with a decreased level of accuracy (<85%) and an increased misclassification rate (<30%).

Even though the PLS-DA models for zebra and springbok exhibited sufficient individual muscle classification accuracies, the overall model accuracies were unsatisfactory. These models also exhibited decreased cross-validation- and validation accuracies, therefore indicating that these models were over-fitted (Miller, 2005) and not as effective as suggested by the calibration accuracy.

The results achieved for the ostrich model (100%) were comparable to that obtained by Alomar *et al.* (2003). These researchers were able to discriminate between three types of bovine muscles [*semitendinosus* (ST), *longissimus thoracis et lumborum* (LTL), *supraspinatus* (SS)] with an accuracy of 94.5%. The results for zebra (67.4%) and springbok (50.5%) are worse than that reported in literature (Alomar *et al.*, 2003). Although the same discrimination method was used in both studies, the pre-processing methods and amount of muscles analysed differed. Therefore, it was not possible to conclude whether the results herein achieved inferior accuracies due to the pre-processing methods used, or due to the amount of muscle types used. This study also indicated that the differences in moisture-, fat-, protein content, pH and meat tenderness (shear force value) contributed to the successful differentiation and prediction of different muscle types (Alomar *et al.*, 2003). These physiochemical differences were similar to the differences observed in the current study, thereby reinforcing that PLS-DA can be used to discriminate between different muscle types of springbok, zebra and ostrich, irrespective of the treatment (fresh or previously frozen).

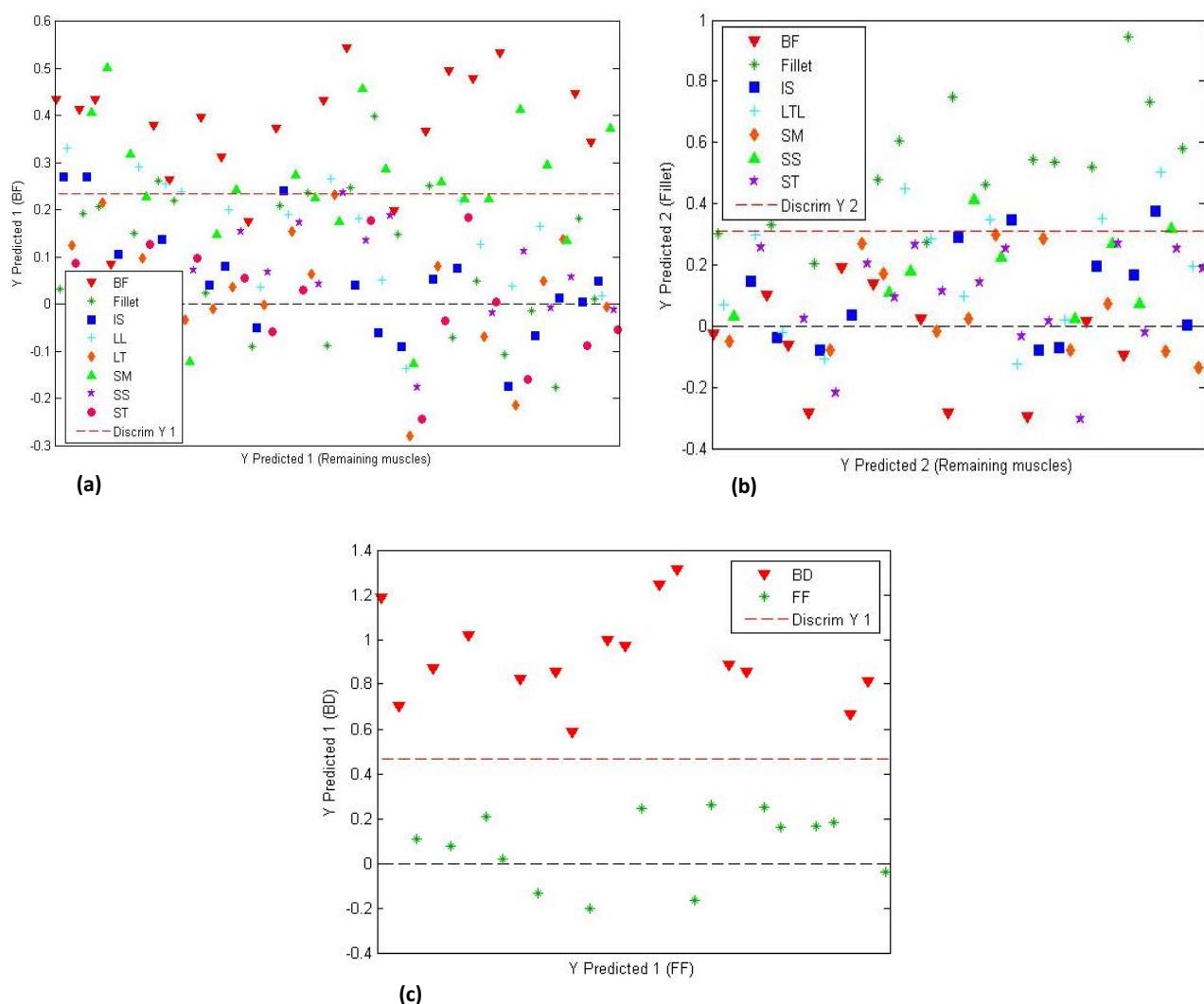


Figure 4.59 PLS-DA models (SNV + detrend pre-processed) for muscle discrimination, resulting in satisfactory individual classification accuracies. PLS-DA prediction score plot of (a) zebra (b) springbok and (c) ostrich, illustrating the predicted objects. (a) Score plot (6 LVs) for zebra (91.5%) of objects predicted as BF [above red line (Y1)] vs. remaining muscles [below red line (Y1)], (b) score plot (4 LVs) for springbok (79.7%) of objects predicted as Fillet [above red line (Y2)] vs. remaining muscles [below red line (Y2)], (c) score plot (5 LVs) for ostrich (100%) of objects predicted as BD [above red line (Y1)] vs. FF [below red line (Y1)].

Table 4.78 The performance measures used to assess the PLS-DA models for different muscle types of the three species, pre-processed with SNV + detrend.

Species	Number of latent variables (LVS)			Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
		Treatment									
Zebra	6	BF		91.5	4.7	3.8	79.0	95.4	75.0	76.9	8.5
		Fillet		87.4	4.5	8.1	60.9	90.2	73.7	66.7	12.6
		IS		88.2	7.3	4.6	73.7	94.3	63.6	68.3	11.8
		LTL		87.8	7.7	4.5	71.2	94.4	59.7	64.8	12.2
		SM		91.5	1.9	6.6	65.0	92.3	86.7	74.3	8.5
		SS		88.2	3.6	8.2	35.7	91.1	55.6	43.5	11.8
		ST		91.5	5.7	2.8	78.6	96.6	64.7	71.0	8.5
Springbok	4	BF		77.1	13.1	9.8	50.0	87.2	42.9	46.2	23.0
		Fillet		79.7	13.6	6.8	76.5	89.5	61.9	68.4	20.3
		IS		82.5	8.8	8.8	58.3	88.9	58.3	58.3	17.5
		LTL		70.2	11.9	17.9	20.0	78.6	27.3	23.1	29.9
		SM		74.6	12.7	12.7	33.3	84.3	33.3	33.3	25.4
		SS		85.5	5.5	9.1	44.4	89.6	57.1	50.0	14.6
		ST		79.7	10.2	10.2	62.5	86.1	62.5	62.5	20.3
Ostrich	5	BD		100.0	0	0	100.0	100.0	100.0	100.0	0
		FF		100.0	0	0	100.0	100.0	100.0	100.0	0

(PLS-DA) Partial least squares discriminant analysis; (SNV) Standard normal variate; (BF) *biceps femoris*; (Fillet) *psoas major*; (IS) *infraspinatus*; (LTL) *longissimus thoracis et lumborum*; (SM) *semimembranosus*; (SS) *supraspinatus*; (ST) *semitendinosus*; (BD) *gastrocnemius*; (FF) *iliofibularis* muscles.

4.3.4 Optimal model selection

After the models [pre-processed with SNV + detrend (4th order polynomial)] were calculated to distinguish between the different muscle types, irrespective of the treatment (fresh or previously frozen), the models were compared to identify the model with the best results. **Table 4.79** reveals that the DA (QDA, LDA) and PLS-DA models produced the best results.

Table 4.79 An overview of the accuracies for the various classification and discrimination models, pre-processed with SNV + detrend to distinguish between the different muscle types.

Species	Model	Number of nearest neighbour (<i>k</i>), PCs or LVs	Classification accuracy (%)	Misclassification rate (%)
Zebra	SIMCA	-	27.3	72.7
	KNN	2	43.1	56.9
	QDA	5	72.2	27.8
	PLS-DA	6	67.4	32.6
Springbok	SIMCA	-	38.4	61.6
	KNN	5	45.2	54.8
	QDA	5	81.7	18.3
	PLS-DA	4	50.5	49.5
Ostrich	SIMCA	-	88.2	11.8
	KNN	5	83.3	16.7
	LDA	6	100.0	0
	PLS-DA	5	100.0	0

(SNV) Standard normal variate; (PCs) Principal components; (LVs) Latent variables; (SIMCA) Soft independent modelling of class analogy; (KNN) K-nearest neighbour; (QDA) Quadratic discriminant analysis; (LDA) Linear discriminant analysis; (PLS-DA) Partial least squares discriminant analysis.

This data was then subjected to different pre-processing techniques, as previously discussed, to determine if the classification accuracies could be improved. PCA models were calculated, the data explored and outliers removed, then DA- and PLS-DA models were recalculated and evaluated.

4.3.5 Exploratory data analysis (EDA)

4.3.5.1 Principal component analysis (PCA)

The score plots of PC2 vs. PC3 (**Addendum C, Figure C3 – C27**) were used to investigate the difference between the muscles, as the data shows that the variation was explained in the second component. Although the data variance explained by PC2 improved when using the various pre-processing techniques; good (ostrich) and minimal (zebra, springbok) class separation between the different muscle types was still observed in the PCA score plots (**Addendum C, Figure C3 – C27**). The variance observed in PC2 can be

attributed to the variances within the species due to differences in the macronutrient composition. This separation was predominantly due to differences in moisture, fat, protein and pH as illustrated in the loading plots (**Addendum C, Figure C3b – C27b**).

The PC2 score plots of the SNV treated data illustrated major overlap (zebra and springbok) as well as good separation (ostrich) between the different muscle types (**Addendum C, Figure C3a, C12a, C21a**) and the loading plots (**Addendum C, Figure C3b, C12b, C21b**) exhibited interpretable bands at 1118, 1124 and 1131 nm. These bands are associated with the pH (1118 and 1124 nm)(ElMasry *et al.*, 2012; Iqbal *et al.*, 2013), moisture- and fat content (1131 nm) (Kamruzzaman *et al.*, 2012b). Although SNV removed the scatter effects, the separation was still insufficient for zebra and springbok, as an overlap between the different muscles was observed (**Addendum C, Figure C3b, C12b, C21b**).

The score plots (PC2) of the SNV + SGd₂(7) and SNV + SGd₂(9) treated data illustrated an overlap between the different muscles for all three species (**Addendum C, Figure C4a – C5a, C13a – C14a, C22a – C23a**). The loading plots (**Addendum C, Figure C4b – C5b, C13b – C14b, C22b – C23b**) had interpretable bands at 1137, 1143, 1205, 1236, 1248, 1279, 1298, 1304, 1366 and 1577 nm. Thus suggesting that the separation between the treatments is a result of variance in the pH (ElMasry *et al.*, 2012; Iqbal *et al.*, 2013), moisture content (Osborne *et al.*, 1993; Cozzolino & Murray, 2004; Barbin *et al.*, 2013b; Ma *et al.*, 2015; Pu *et al.*, 2015), protein content (Barbin *et al.*, 2012b; Barbin *et al.*, 2013a; Talens *et al.*, 2013), fat content (Barbin *et al.*, 2013a) as well as meat tenderness (ElMasry *et al.*, 2012; Kamruzzaman *et al.*, 2013). The pre-processing did not improve the separation in the direction of PC2 for the zebra and springbok models, and though the separation in PC2 for ostrich was more prominent, an overlap between the muscles was still evident (**Addendum C, Figure C4a – C5a, C13a – C14a, C22a – C23a**). The improved separation between muscle types can be ascribed to the combined pre-processing of SNV + Savitzky-Golay second derivative.

The PC2 score plot of the SNV + detrend + SGd₂(7) and SNV + detrend + SGd₂(9) corrected data illustrated insufficient separation between the different muscles for the zebra and springbok models (**Addendum C, Figure C6a – C7a, C15a – C16a**), while the score plot for ostrich illustrated good separation with a slight overlap between the different muscles (**Addendum C, Figure C24a**). The loading plots (**Addendum C, Figure C6b – C7b, C15b – C16b, C24b**) exhibited the same interpretable bands as previously discussed, with the exception of a few additional bands at 1062, 1211, 1372 and 1409 nm. Therefore, the separation between the different muscles types is a result of variance in the pH, moisture-, protein-, fat content as well as meat tenderness. The combined pre-processing of SNV + detrend + Savitzky-Golay second derivative exhibited similar results as SNV + Savitzky-Golay second derivative, thus indicating that detrend pre-processing did not improve the separation.

The PCA results of the SGd₁(5) and SGd₁(7) treated data are shown in **Addendum C, Figure C8a – C9a, C17a – C18a, C25a – C26a**. These results suggested that Savitzky-Golay first derivative would be an effective pre-processing technique as it illustrated improved/good separation. The loading plot for both pre-processing techniques showed interpretable bands at 957, 1025, 1031, 1137 and 1385 nm (**Addendum C**,

Figure C8b – C9b, C17b – C18b, C25b – C26b. The band at 957 nm is associated with moisture- and fat content (Kamruzzaman *et al.*, 2012a; Kamruzzaman *et al.*, 2012b). The band at 1025 nm is associated with the N-H stretch second overtone found in aromatic groups (Osborne *et al.*, 1993) and is related to protein content. The 1031 and 1137 nm absorption bands are related to meat tenderness (ElMasry *et al.*, 2012; ElMasry *et al.*, 2013). The band at 1137 nm can be associated with a variance in protein (Barbin *et al.*, 2013a; Talens *et al.*, 2013). Lastly, the absorption band 1385 nm suggest that the separation between the different muscle types is caused by changes in moisture content (ElMasry *et al.*, 2013).

The PC2 score plots of the SGd₂(7) and SGd₂(9) corrected data illustrated improved separation (**Addendum C, Figure C10a – C11a, C19a – C20a, C27a**), which was sufficient for muscle type cluster analysis. The loading line plots (**Addendum C, Figure C10a – C11a, C19a – C20a, C27a**) had interpretable bands at 957, 970, 976, 1069, 1118, 1155, 1310, 1329, 1341, 1366, 1372 and 1403 nm. The separation between the different muscle types were attributable to similar differences as discussed in the previous pre-treatment.

The results from the various pre-processing techniques suggests that the Savitzky-Golay first- and second derivative transformations would be sufficient and best suited for muscle type separation. The Savitzky-Golay first- and second derivative transformations enhanced the differences in macronutrient composition of the biological samples. This separation was predominantly due to differences in moisture, fat, protein, pH and meat tenderness as illustrated in the loading plots. Even though Savitzky-Golay derivatives showed the best PCA results for separation, all the pre-processing techniques were used for further data analysis and model evaluation.

4.3.6 Multivariate data analysis (MDA): Optimal model development

4.3.6.1 Discriminant analysis (DA)

The overall accuracy of the discriminant analysis models, illustrated that the ostrich model provided the best discrimination. The ostrich LDA models (6 PCs) for the various forms of pre-processing gave similar results (**Table 4.80**). The LDA model, pre-processed with SGd₁(5), presented the best discrimination results and the separation between classes was more distinct, with little overlap (**Addendum C, Figure C25a**). The calibration model achieved an overall classification accuracy of 100% for SGd₁(5) (**Table 4.80**). This model also exhibited excellent validation accuracies (100%), and this suggests that the model can accurately distinguish between the two muscle types. The classification accuracy (100%), sensitivity (100%) and specificity (100%) confirmed this (not shown). Therefore, the SGd₁(5) pre-processed data resulted in a more robust model and the reason why the SGd₁(5) pre-treatment achieved the best results was discussed in **Section 4.1.2.5**.

The LDA model, pre-processed with SGd₁(5), achieved identical results compared to that of the original model pre-processed with SNV + detrend. To further illustrate that there was a minimal difference between these two pre-processing techniques, the PCA score plots of SGd₁(5) (**Addendum C, Figure C25a**) and SNV + detrend (**Figure 4.52a**) are shown. The classification accuracy for differentiating between the BD and FF muscles were 100% for both these pre-processing methods. The reason for the slight differences

between these pre-processing techniques was discussed in **Section 4.1.2.5**. However, the derivatives are calculated with an additional smoothing effect that removes the unwanted features and as a result enhances the features that are advantageous for the analysis (Savitzky & Golay, 1964). Therefore, the SGd₁(5) pre-processing would be the better model as the separation due to differences in moisture content were enhanced. The results discussed in **Section 4.3.2** and literature (Sales, 1996) supports this phenomenon that moisture is the main contributor for differentiating between the BD and FF muscles .

The overview of the classification accuracies (**Table 4.80**) and performance measures (not shown) illustrated that the QDA model (5 PCs) pre-processed with SNV gave the best results for zebra (79.9%), and the QDA (5 PCs) model pre-processed with SGd₁(7) gave the best results for springbok (77.4%). The zebra QDA model, pre-processed with SNV (79.9%), achieved slightly higher results than that of the original model pre-processed with SNV + detrend (72.2%). To illustrate that there were differences between these two pre-processing techniques, the PCA score plots of SNV (**Addendum C, Figure C3a**) and SNV + detrend (**Figure 4.50b**) are shown. The SNV score plot of PC2 exhibited an improved explained variance of 9%, with improved clustering of different muscle types. Therefore, the improved SNV model performance can be attributed to prominent differences in pH, enhanced by the pre-processing. The QDA model, pre-processed with SGd₁(7) (springbok), gave slightly lower results than that of the original models pre-processed with SNV + detrend. The classification accuracy for differentiating between the different muscle types was 77.4% for SGd₁(7) and 81.7% for SNV + detrend. When comparing these results to the Savitzky-Golay mean spectra, it is observed that the SNV + detrend pre-processing enhanced the differences in protein content and pH. The protein content and pH were illustrated to be the main contributors for differentiating between the different springbok muscle types (**Section 4.3.2**). Although the separation in the scores plot of the SGd₁(7) treated (**Addendum C, Figure C18a**) appeared to be improved, the pre-processing amplified other physiochemical characteristics, some of which were not deemed important for springbok muscle differentiation. These amplified characteristics include moisture, fat, protein and meat tenderness. Although these physiochemical characteristics were observed in the difference spectra, the differences were either, not substantial between the different muscle types or not supported by literature. Therefore, the springbok model improved with the SNV + detrend pre-processing because this method enhanced the important physiochemical characteristics (protein and pH), which exhibited the largest difference between the different muscles. Consequently, improving separation between classes as well as the predictive power of the model.

Although the use of different pre-processing techniques improved the calibration accuracies and performance measures of the zebra and springbok models, the validation accuracies decreased. This indicated that the models were over-fitted (Miller, 2005) and therefore not as effective as suggested by the calibration accuracies. The results in the current study suggest that the discrimination of the different muscle types was less sufficient due to the lack of separation between the muscles. This phenomenon can be ascribed to the samples' spectral similarities, hence causing confusion between classes.

Table 4.80 An overview of the accuracies of the DA models with various pre-processing techniques applied for muscle type discrimination.

Pre-processing	Species / Model / Number of principal components (PCs)								
	Zebra			Springbok			Ostrich		
	QDA (5 PCs)		Validation accuracy (%)	QDA (5 PCs)		Validation accuracy (%)	LDA (6 PCs)		Validation accuracy (%)
Classification accuracy (%)	Misclassification rate (%)	Validation accuracy (%)	Classification accuracy (%)	Misclassification rate (%)	Validation accuracy (%)	Classification accuracy (%)	Misclassification rate (%)	Validation accuracy (%)	Validation accuracy (%)
SNV	79.9	20.1	41.7	69.9	30.1	41.9	100.0	0	40.0
SNV+SGd ₂ (7)	75.7	24.3	45.8	69.9	30.1	54.8	93.3	6.7	80.0
SNV+SGd ₂ (9)	73.6	26.4	37.5	75.3	24.7	48.4	86.7	13.3	80.0
SNV+detrend+SGd ₂ (7)	68.8	31.3	56.3	76.3	23.7	32.3	100.0	0	90.0
SNV+detrend+SGd ₂ (9)	65.3	34.7	62.5	76.3	23.7	32.3	-	-	-
SGd ₁ (5)	72.2	28.8	56.3	77.4	22.6	54.8	100.0	0	100.0
SGd ₁ (7)	69.4	30.6	56.3	77.4	22.6	51.6	96.7	3.3	100.0
SGd ₂ (7)	71.5	28.5	45.8	74.2	25.8	48.4	96.7	3.3	100.0
SGd ₂ (9)	68.1	31.9	47.9	77.4	22.6	58.1	-	-	-

(DA) Discriminant analysis; (QDA) Quadratic discriminant analysis; (LDA) Linear discriminant analysis; (PCs) Principal components; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

4.3.6.2 Partial least squares discriminant analysis (PLS-DA)

The overall accuracy of the discriminant analysis models, illustrated that the ostrich model provided the best discrimination. The ostrich PLS-DA models for the various pre-processing techniques yielded similar results (**Table 4.81**). The PLS-DA models, pre-processed with SNV, SNV + detrend + SGd₂(7), SGd₁(5), SGd₁(7) and SGd₂(7), presented the best discrimination results. Although there was no substantial difference between the calibration results, further investigation of the cross-validation-, validation accuracies (**Table 4.81**) and performance measures (not shown) suggests that the model pre-processed with SGd₂(7) would be optimal. The calibration model achieved an overall classification accuracy of 100% (**Table 4.81**). This model also exhibited excellent performance measures (not shown), thus the sensitivity (100%) and specificity (100%) suggests that the model can accurately distinguish between the BD and FF muscles.

Although the PLS-DA models, pre-processed with SGd₂(7) (100%), achieved identical results to that of the original model pre-processed with SNV + detrend (100%), the separation was improved for the SGd₂(7) treated data. To further illustrate that there were minimal differences between the PCA results of these two pre-processing techniques, the score plots of SGd₂(7) (**Addendum C, Figure C27a**) and SNV + detrend (**Figure 4.52a**) are shown. The reason for the differences between these pre-processing techniques is most likely because of the additional smoothing effect of the derivatives removes unwanted features and as a result enhances features that are advantageous for the analysis (Savitzky & Golay, 1964). Therefore, the SGd₂(7) pre-processing achieved better PCA results as the separation due to differences in moisture, fat, pH and meat tenderness were enhanced. These physiochemical characteristics were found to be the main contributors for differentiating between the BD and FF muscles as these characteristics illustrated a greater difference between the muscle samples as discussed in **Section 4.3.2**. Therefore, by enhancing these differences the muscle separation and model prediction was improved.

The overview of the classification accuracies (**Table 4.81**) exhibits substantial differences between the various pre-processing techniques for the zebra and springbok models. The overview of the classification accuracies (**Table 4.81**) and performance measures (not shown) illustrated that the PLS-DA model pre-processed with SNV + detrend + SGd₂(9) gave the best results for zebra (77.8%) and the model pre-processed with SNV + detrend + SGd₂(7) gave the best results for springbok (79.6%). The zebra PLS-DA model, pre-processed with SNV + detrend + SGd₂(9) (77.8%), achieved slightly higher results than that of the original model pre-processed with SNV + detrend (67.4%). To illustrate that there were differences between these two pre-processing techniques, the PCA score plots of SNV + detrend + SGd₂(9) (**Addendum C, Figure C7a**) and SNV + detrend (**Figure 4.50a**) are shown. The SNV + detrend + SGd₂(9) score plot of PC2 exhibited an improved explained variance of 6%, with improved clustering of different muscle types. Therefore, the SNV + detrend + SGd₂(9) model performed better because of the improved separation as a result of the more prominent differences in moisture, pH and meat tenderness, enhanced by the pre-processing. The PLS-DA model, pre-processed with SNV + detrend + SGd₂(7) (springbok), gave better results than that of the original models pre-processed with SNV + detrend. The classification accuracy for differentiating between the

different muscle types was 79.6% for SNV + detrend + SGd₂(7) and 50.5% for SNV + detrend. When comparing these results, it is observed that the SNV + detrend + SGd₂(7) pre-processing enhanced the differences in fat-, protein content and meat tenderness. The protein content and meat tenderness were illustrated, in **Section 4.3.2**, to be the main contributors for differentiating between the different springbok muscle types. Therefore, the springbok model improved with the SNV + detrend + SGd₂(7) pre-processing because this method enhanced the important physiochemical characteristics (protein content and meat tenderness) which exhibited the largest difference between the different muscles.

All the models improved with the mentioned techniques because the pre-processing methods minimised the overlap between the different muscle types and improved the clustering of the classes. Consequently, improving separation between classes as well as the predictive power of the model. Therefore, the SNV + detrend + SGd₂(9) (zebra), SNV + detrend + SGd₂(7) (springbok) and SGd₂(7) (ostrich) pre-processing achieved better results as the separation due to differences in moisture-, fat-, protein content pH as well as meat tenderness were enhanced. Although the use of different pre-processing techniques improved the calibration accuracies and performance measures of the zebra and springbok models, the validation accuracies decreased. This indicated that the models were over-fitted (Miller, 2005) and therefore not as effective as suggested by the calibration accuracies. This is because the models were more specific for the calibration set, and was not a good representation of the overall data. Therefore, the validation set comprised data that was not represented by the model. This could also be because more muscles were investigated, with some of them having large variation within the muscle causing their "finger print" to overlap with another muscle's. Thus, resulting in the misclassification of muscle samples. Therefore, the validation set could possibly improve by decreasing the amount of muscles investigated in the data set, or by grouping the muscles with similar physiochemical characteristics/"finger prints".

Table 4.81 An overview of the classification accuracies of the PLS-DA models with various pre-processing techniques applied for muscle type discrimination.

Pre-processing	Species										
	Zebra			Springbok			Ostrich			CV (%)	Validation accuracy (%)
	Calibration accuracy (%)	CV (%)	Validation accuracy (%)	Classification accuracy (%)	CV (%)	Validation accuracy (%)	Classification accuracy (%)	CV (%)	Validation accuracy (%)		
Pre-processing	Misclassification rate (%)			Misclassification rate (%)			Misclassification rate (%)				
SNV	57.6 42.4	46.5	45.8	63.4 36.6	46.2	48.4	100.0 0	96.7	100.0		
SNV+SGd ₂ (7)	75.7 24.3	69.4	54.2	50.5 49.5	37.6	35.5	90.0 10.0	83.3	70.0		
SNV+SGd ₂ (9)	74.3 25.7	70.1	52.1	52.7 47.3	36.6	38.7	90.0 10.0	83.3	70.0		
SNV+detrend+SGd ₂ (7)	72.9 27.1	60.4	64.6	79.6 20.4	66.7	45.2	100.0 0	86.7	80.0		
SNV+detrend+SGd ₂ (9)	77.8 22.2	61.1	64.6	71.0 29.0	54.8	45.2	-	-	-		
SGd ₁ (5)	52.8 47.2	38.9	45.8	63.4 36.6	43.0	48.4	100.0 0	90.0	100.0		
SGd ₁ (7)	70.8 29.2	56.9	64.6	65.6 34.4	39.8	48.4	100.0 0	96.7	100.0		
SGd ₂ (7)	69.4 30.6	56.9	66.7	75.3 24.7	59.1	54.8	100.0 0	100.0	100.0		
SGd ₂ (9)	61.8 38.2	47.9	53.1	64.5 35.5	39.8	51.6	-	-	-		

(PLS-DA) Partial least squares discriminant analysis; (CV) Cross-validation; (SNV) Standard normal variate; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

4.3.7 Optimal model selection

After the models for the various types of pre-processing were calculated, the models with the best results were compared to identify the optimal one (**Table 4.82**). It was concluded that the QDA model, pre-processed with SNV was the best for discriminating between the different zebra muscles. Whilst the QDA model pre-processed with SNV + detrend was the best for discriminating between the different springbok muscles, irrespective of the treatment (fresh or previously frozen) for both species. Although the use of different pre-processing techniques improved the calibration accuracies and performance measures of the models, the validation accuracies decreased. This indicated that the models were over-fitted (Miller, 2005) and therefore not as effective as suggested by the calibration accuracies. The results in the current study suggest that the discrimination of the different muscle types was less sufficient due to the lack of separation between the different muscles. This phenomenon can be ascribed to the samples' spectral similarities, hence causing confusion between classes. The misclassification of muscles mostly occurred between those that are located near to one another anatomically. Consequently, it can be concluded that these muscles are closely related and that the differences in their physiochemical characteristics are therefore negligible. For this reason, it was deemed important to examine the success of a two-group discrimination model. The muscles were grouped together based on their anatomical location. PLS-DA and QDA models were developed to discriminate between the forequarters (IS, SS) and hindquarters (BF, Fillet, LTL, SM, ST), irrespective of the treatment (fresh or previously frozen).

The results for the ostrich models (LDA and PLS-DA) suggests that all the pre-processing methods would be sufficient for differentiating between the two muscle types. However, the PCA score plots for the SG₁(5) (**Addendum C, Figure C25a**) and SG₂(7) (**Addendum C, Figure C27a**) pre-processed data illustrates that these methods improved the separation between the muscles. Hence, making it possible to conclude that these two pre-processing techniques would be the best for enhancing the separation between the different muscle types. The ostrich PLS-DA and LDA models achieved similar results. The reason why LDA performed just as well as PLS-DA, can be ascribed to the fact that the classes are not closely related, therefore not resulting in collinearity problems often associated with LDA. The models pre-treated with SGd₁(5) and SGd₂(7) outperformed the combined pre-processing of SNV + detrend for the ostrich, as the Savitzky-Golay transformation was found to have enhanced the differences in moisture, fat, pH and meat tenderness, which was predominantly the contributors for muscle separation.

Table 4.82 An overview of the classification accuracies of the DA and PLS-DA models with various pre-processing techniques applied to distinguish between the muscle types.

Species	Model	Pre-processing	Calibration		Validation	
			Number of PCs or LVs	Classification accuracy (%)	Misclassification rate (%)	Classification accuracy (%)
Zebra	QDA	SNV + detrend	5	72.2	27.8	52.1
	PLS-DA	SNV + detrend	6	67.4	32.6	52.1
	QDA	SNV	5	79.9	20.1	41.7
	PLS-DA	SNV+DT+SGd ₂ (9)	8	77.8	22.2	64.6
Springbok	QDA	SNV + detrend	5	81.7	18.3	41.9
	PLS-DA	SNV + detrend	4	50.5	49.5	22.6
	QDA	SGd ₁ (7)	5	77.4	22.6	51.6
	PLS-DA	SNV+DT+ SGd ₂ (7)	7	79.6	20.4	45.2
Ostrich	LDA	SNV + detrend	6	100.0	0	100.0
	PLS-DA	SNV + detrend	5	100.0	0	100.0
	LDA	SGd ₁ (5)	6	100.0	0	100.0
	PLS-DA	SGd ₂ (7)	5	100.0	0	100.0

(DA) Discriminant analysis; (QDA) Quadratic discriminant analysis; (LDA) Linear discriminant analysis; (PLS-DA) Partial least squares discriminant analysis; (PCs) Principal components; (LVs) Latent variables; (SNV) Standard normal variate; (DT) Detrend; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

4.3.8 Two-group discrimination model evaluation

Table 4.83 An overview of the classification rates of the two-group distribution QDA and PLS-DA models for zebra and springbok with various pre-processing techniques applied.

Species	Model	Pre-processing	Calibration		Validation	
			Number of PCs or LVs	Classification accuracy (%)	Misclassification rate (%)	Classification accuracy (%)
Zebra	QDA	SNV + detrend	5	75.7	24.3	70.8
	PLS-DA	SNV + detrend	4	81.9	18.1	81.3
	QDA	SNV	5	79.9	20.1	66.7
	PLS-DA	SNV+DT+SGd ₂ (9)	6	90.3	9.7	85.4
Springbok	QDA	SNV + detrend	5	87.1	12.9	74.2
	PLS-DA	SNV + detrend	3	83.9	16.1	61.3
	QDA	SGd ₁ (7)	5	87.1	12.9	77.4
	PLS-DA	SNV+DT+ SGd ₂ (7)	5	97.9	2.1	96.8

(DA) Discriminant analysis; (QDA) Quadratic discriminant analysis; (PLS-DA) Partial least squares discriminant analysis; (PCs) Principal components; (LVs) Latent variables; (SNV) Standard normal variate; (DT) Detrend; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points).

After the zebra and springbok models were re-calculated, the new models were compared to identify the best one (**Table 4.83**). The results for the two-group discrimination models illustrates that all the model's calibration- and validation accuracies improved. Hence, making it possible to conclude that the two-group discrimination improved the clustering and separation of the classes. Further examination of the results exhibited that the PLS-DA models pre-processed with SNV + detrend + SGd₂(9) for zebra (90.3%) and SNV + detrend + SGd₂(7) for springbok (97.9%) was the best for differentiating between the muscles grouped as forequarters and hindquarters. The PLS-DA models achieved better results as it generally outperforms QDA, because it overcomes the collinearity problems often associated with QDA. The models pre-treated with SNV + detrend + SGd₂(9) for zebra and SNV + detrend + SGd₂(7) for springbok outperformed the combined pre-processing of SNV + detrend for both species, as the additional Savitzky-Golay transformation was found to have enhanced the differences in protein-, fat content, pH and meat tenderness, which was predominantly the contributors for muscle separation.

The results in the current study suggest that grouping samples with spectral similarities together, can be used to improve the discrimination as well as the overall model accuracies and performance measures. McElhinney *et al.* (1999) reported similar observations when attempting to differentiate between species. These researchers found that their model accuracies improved when chicken and turkey were combined into a single poultry class, as it was difficult for the various models to accurately discriminate between chicken and turkey meats. In the current study, the results show that it is possible to differentiate between the BD and FF muscles of ostrich with a 100% accuracy, irrespective of the meat being fresh or previously frozen. The results also suggest that a two-group discrimination of samples with similar spectra would be better. Therefore, the results show that it is possible to differentiate between the forequarters and hindquarters of the zebra and springbok muscles, irrespective of the treatment (fresh or previously frozen). The zebra and springbok model achieved a 90.3% and 97.9% overall accuracy, respectively.

4.3.9 Conclusion

NIR spectroscopy combined with MDA could accurately distinguish between the different muscle types for ostrich, as well as the forequarters and hindquarters for zebra and springbok, irrespective of the treatment (fresh or previously frozen). The PLS-DA discrimination models yielded the best results and could effectively distinguish the ostrich muscles (BD and FF) from one another with an accuracy of 100%. Throughout the study, the samples that resembled each other closely were difficult to differentiate. This is attributed to the samples' spectral similarities, thus accounting for the greatest source of misclassification. Therefore, the results suggested that a two-group discrimination of samples with similar spectra would be better, thus improving the classification accuracies of the prediction models. The PLS-DA results showed that it was possible to differentiate between the forequarters and hindquarters of the zebra and springbok muscles, irrespective of the treatment (fresh or previously frozen). The zebra and springbok models achieved good

overall accuracies of 90.3% and 97.9%, respectively. In addition, it was found that the treatment (fresh or previously frozen) does not influence the accuracy of the model for muscle type discrimination.

4.4 Hierarchical model development

Due to the complexity of the data it was not possible to classify the multiple classes (e.g. species, fresh vs. frozen-thawed, frozen period and muscle type), using a single classification and/or discrimination model. Therefore, to solve the classification/discrimination problems and handle the increased detail of the data, it had to be divided into sub-groups with individual models. After all the models for the multiple classes were examined and the optimal models were selected, a hierarchical model was constructed in which the data of the models are organised into a tree-like structure (**Addendum D, Figures D1 – D3**).

Two multilevel hierarchical models were constructed. The first hierarchical model (**Addendum D, Figure D1**) differentiated between three species (zebra, springbok, ostrich), determined whether the meat was fresh or previously frozen, and the muscle type for each species. The model for ostrich differentiated between the BD- and FF muscles, while the models for the ungulates (zebra and springbok) differentiated between the forequarters (IS, SS) and the hindquarters (BF, Fillet, LTL, SM, ST). This was done as grouping the muscles based on their anatomical location improved the model accuracies and performance measures. The second hierarchical model (**Addendum D, Figures D2 – D3**) differentiated between four species (black wildebeest, zebra, springbok, ostrich) using one muscle type [LTL (ungulates) and FF (ostrich)], determined whether the meat was fresh or previously frozen for each species and the frozen period (1 – 9 months) of the meat.

The results for the first multilevel hierarchical model (**Table 4.84**) illustrates that the model developed for species determination was able to identify almost all samples of zebra (48 of 52) and springbok (31 of 32), whilst accurately identifying the ostrich samples (9 of 9). The overview of the model classification accuracies therefore exhibits that ostrich (100%) achieved the highest correctly classified samples followed by springbok (96.9%) and zebra (92.3%). Different models were then selected for each species in order to differentiate between the fresh and previously frozen meat as well as to determine the different muscle types. The models for zebra and springbok could accurately distinguish between the fresh and frozen-thawed meat samples. For the ostrich on the other hand, all the frozen-thawed samples were assigned to the correct treatment (100%), while one fresh sample was incorrectly assigned to the frozen-thawed treatment, resulting in an accuracy of 80% (fresh meat determination). Lastly, the model for ostrich could differentiate between the BD- and FF muscles with a 100% accuracy. The springbok model achieved slightly lower accuracies, with the model able to classify all the forequarters (15 of 15) and almost all the hindquarters (15 of 16). The zebra model was, to a lesser extent, able to differentiate between the forequarters and hindquarters. The model correctly classified 80.9% of the forequarters (17 of 21) and 88.9% of the hindquarters (24 of 27). Therefore, this hierarchical model exhibits the possibility and ability of multiple models used in combination to classify

multiple classes. This is preferred as complex multiple classes can be classified using a single multilevel model, therefore simplifying the discrimination process.

In the second multilevel hierarchical model (**Table 4.85**), the model developed for species determination was able to identify almost all the springbok (60 of 63) and black wildebeest (54 of 58), and to a lesser extent the ostrich (58 of 65) and zebra (54 of 66). The model overview therefore exhibits that springbok (95.2%) achieved the highest correctly classified samples followed by black wildebeest (93.1%), ostrich (89.2%) and zebra (81.8%). Different models were then selected for each species in order to differentiate between the fresh and previously frozen meat as well as the different individual frozen periods. The model for springbok could accurately distinguish between all the fresh and frozen-thawed samples. The model for ostrich could accurately determine all the frozen-thawed samples and 97.7% of the fresh samples (43 of 44). For the black wildebeest model all the fresh samples were assigned to the correct treatment, while one frozen-thawed sample was incorrectly assigned to the fresh treatment, resulting in an accuracy of 94.4% (frozen-thawed meat determination). Lastly, the model for zebra could accurately distinguish between all the frozen-thawed samples and 92.9% of the fresh samples (26 of 28). As can be seen in **Table 4.85**, it was not possible to obtain suitable discrimination models for most frozen periods. This indicates that although the different models for each species could accurately determine the previously frozen meat, with the exception of black wildebeest, the models were not able to determine the different frozen periods, with sufficient accuracies. This suggests that it is unnecessary to determine the frozen period of meat as the discrimination between fresh and previously frozen meat, irrespective of the frozen period, is sufficient and closer to commercial requirements. Therefore, this hierarchical model exhibits the possibility and ability of multiple models used in combination to classify multiple classes.

The use of multilevel hierarchical modelling therefore enables one to simultaneously identify the species, treatment (irrespective of the frozen period) and different muscle types, without having to use separate individual models for each class classification. Therefore, the hierarchical model provides a holistic viewpoint of the meat samples' characteristics. The hierarchical model also highlighted the fact that different types of models (data pre-treatment and number of latent variables) may be suitable for each class discrimination step. Therefore, confirming that it is not always possible to use a single model (same data pre-treatment and number of latent variables) to classify complex data with multiple classes.

Table 4.84 An overview of the multilevel hierarchical model classification accuracies of the PLS-DA models with various pre-processing techniques applied for determining the species, fresh vs. previously frozen meat as well as the muscle type.

Data pre-treatment / LVs			Correct			Data pre-treatment / LVs			Correct			Data pre-treatment / LVs			Correct		
Species classification			classification			Fresh vs. frozen-thawed			classification			Muscle type			classification		
SGd ₁ (7) / 8 LVs			n	n	%	SGd ₁ (5) / 5 LVs			n	n	%	SNV + DT + SGd ₂ (9) / 6 LVs			n	n	%
Zebra	52	48	92.3	Fresh	21	21	100	Forequarters	21	17	80.9						
				Frozen-thawed	27	27	100	Hindquarters	27	24	88.9						
Springbok	32	31	96.9	Fresh	18	18	100	Forequarters	15	15	100						
				Frozen-thawed	13	13	100	Hindquarters	16	15	93.8						
Ostrich	9	9	100	Fresh	5	4	80	BD	4	4	100						
				Frozen-thawed	4	4	100	FF	4	4	100						

(PLS-DA) Partial least squares discriminant analysis; (LVs) Latent variables; (SNV) Standard normal variate; (DT) Detrend; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points); (BD) *gastrocnemius*; (FF) *iliofibularis*.

Table 4.85 An overview of the multilevel hierarchical model classification accuracies of the PLS-DA models with various pre-processing techniques applied for determining the species, fresh vs. previously frozen meat as well as the frozen period.

Data pre-treatment / LVs			Correct classification			Data pre-treatment / LVs			Correct classification			Data pre-treatment / LVs			Correct classification		
Species classification						Fresh vs. frozen-thawed						Frozen period					
SGd2(9) / 8 LVs			n	n	%	SNV + DT + SGd ₂ (9) / 4 LVs			n	n	%	SNV + DT + SGd ₂ (9) / 4 LVs			n	n	%
BWB	58	54	93.1	Fresh	36	36	100					FT1	3	0	-		
				Frozen-thawed	18	17	94.4					FT2	2	0	-		
Zebra	66	54	81.8	Fresh	28	26	92.9					FT3	1	1	100		
				Frozen-thawed	26	26	100					FT4	2	1	50		

Table 4.85 (Continued)

Data pre-treatment / LVs				Correct classification			Data pre-treatment / LVs				Correct classification			Data pre-treatment / LVs				Correct classification		
Species classification				Fresh vs. frozen-thawed						Species classification					Frozen period					
SGd2(9) / 8 LVs		n	n	%	SGd2(9) / 5 LVs			n	n	%	SNV + DT + SGd2(7) / 5 LVs			n	n	%				
Springbok	63	60	95.2		Fresh	40	40	100						FT1	3	2	66.7			
					Frozen-thawed	20	20	100						FT2	2	2	100			
Ostrich	65	58	89.2	Fresh	44	43	97.7						FT3	1	1	100				
					Frozen-thawed	16	16	100					FT4	1	1	100				

(PLS-DA) Partial least squares discriminant analysis; (LVs) Latent variables; (SNV) Standard normal variate; (DT) Detrend; [SGd₂(7)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 7 points); [SGd₂(9)] Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points); [SGd₁(5)] Savitzky-Golay (1st derivative, 2nd order polynomial, 5 points); [SGd₁(7)] Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points); (FT1 – FT9) Frozen-thawed 1 – 9 months.

4.5 References

- Alomar, D., Gallo, C., Castaneda, M. & Fuchslocher, R. (2003). Chemical and discriminant analysis of bovine meat by near infrared reflectance spectroscopy (NIRS). *Meat Science*, **63**, 441-450.
- Barbin, D., Elmasry, G., Sun, D.-W. & Allen, P. (2012a). 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. (2012b). Predicting quality and sensory attributes of pork using near-infrared hyperspectral imaging. *Analytica Chimica Acta*, **719**, 30-42.
- Barbin, D.F., ElMasry, G., Sun, D.-W. & Allen, P. (2013a). Non-destructive determination of chemical composition in intact and minced pork using near-infrared hyperspectral imaging. *Food Chemistry*, **138**, 1162-1171.
- Barbin, D.F., Sun, D.-W. & Su, C. (2013b). NIR hyperspectral imaging as non-destructive evaluation tool for the recognition of fresh and frozen-thawed porcine *longissimus dorsi* muscles. *Innovative Food Science & Emerging Technologies*, **18**, 226-236.
- 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.
- Cozzolino, D. & Murray, I. (2004). Identification of animal meat muscles by visible and near infrared reflectance spectroscopy. *LWT-Food Science and Technology*, **37**, 447-452.
- Ding, H. & Xu, R. (1999). Differentiation of Beef and Kangaroo Meat by Visible/Near-Infrared Reflectance Spectroscopy. *Journal of Food Science*, **64**, 814-817.
- Ding, H. & Xu, R. (2000). Near-infrared Spectroscopic Technique for Detection of Beef Hamburger Adulteration. *Journal of Agricultural and Food Chemistry*, **48**, 2193-2198.
- Downey, G. & Beauchêne, D. (1997a). Authentication of fresh vs. frozen-then-thawed beef by near-infrared reflectance spectroscopy of dried drip juice. *LWT-Food Science and Technology*, **30**, 721-726.
- Downey, G. & Beauchêne, D. (1997b). Discrimination between fresh and frozen-then-thawed beef *m. longissimus dorsi* by combined visible-near infrared reflectance spectroscopy: A feasibility study. *Meat Science*, **45**, 353-363.
- Du Buisson, P.-M. (2006). Improving the meat quality of blesbok (*Damaliscus dorcas phillipsi*) and springbok (*Antidorcas marsupialis*) through enhancement with inorganic salts. (MSc Thesis). South Africa: University of Stellenbosch.
- ElMasry, G., Sun, D.-W. & Allen, P. (2011). Non-destructive determination of water-holding capacity in fresh beef by using NIR hyperspectral imaging. *Food Research International*, **44**, 2624-2633.
- ElMasry, G., Sun, D.-W. & Allen, P. (2012). Near-infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef. *Journal of Food Engineering*, **110**, 127-140.
- ElMasry, G., Sun, D.-W. & Allen, P. (2013). Chemical-free assessment and mapping of major constituents in beef using hyperspectral imaging. *Journal of Food Engineering*, **117**, 235-246.

- Engelbrecht, A. (2013). Establishing genetic and environmental parameters for ostrich (*Struthio camelus domesticus*) growth and slaughter characteristics. (PhD Thesis). South Africa: University of Stellenbosch.
- Esbensen, K.H., Guyot, D., Westad, F. & Houmoller, L.P. (2002). *Multivariate data analysis in practice: An introduction to multivariate data analysis and experimental design*. Pp. 598. Oslo: CAMO Process AS.
- Fumière, O., Sinnaeve, G. & Dardenne, P. (2000). Attempted authentication of cut pieces of chicken meat from certified production using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **8**, 27-34.
- Hoffman, L.C., Geldenhuys, G. & Cawthorn, D.M. (2016). Proximate and fatty acid composition of zebra (*Equus quagga burchellii*) muscle and subcutaneous fat. *Journal of the Science of Food and Agriculture*, **96**, 3922-3927.
- Hoffman, L.C., Kroucamp, M. & Manley, M. (2007a). Meat quality characteristics of springbok (*Antidorcas marsupialis*). 1: Physical meat attributes as influenced by age, gender and production region. *Meat Science*, **76**, 755-761.
- Hoffman, L.C., Kroucamp, M. & Manley, M. (2007b). Meat quality characteristics of springbok (*Antidorcas marsupialis*). 2: Chemical composition of springbok meat as influenced by age, gender and production region. *Meat Science*, **76**, 762-767.
- Hoffman, L.C., Schalkwyk, S.v. & Muller, N. (2009). Effect of season and gender on the physical and chemical composition of black wildebeest (*Connochaetus gnou*) meat. *South African Journal of Wildlife Research*, **39**, 170-174.
- Hoffman, L.C. & Wiklund, E. (2006). Game and venison-meat for the modern consumer. *Meat Science*, **74**, 197-208.
- Iqbal, A., Sun, D.-W. & Allen, P. (2013). Prediction of moisture, color and pH in cooked, pre-sliced turkey hams by NIR hyperspectral imaging system. *Journal of Food Engineering*, **117**, 42-51.
- Kamruzzaman, M., Barbin, D., ElMasry, G., Sun, D.-W. & Allen, P. (2012a). Potential of hyperspectral imaging and pattern recognition for categorization and authentication of red meat. *Innovative Food Science & Emerging Technologies*, **16**, 316-325.
- Kamruzzaman, M., ElMasry, G., Sun, D.-W. & Allen, P. (2011). Application of NIR hyperspectral imaging for discrimination of lamb muscles. *Journal of Food Engineering*, **104**, 332-340.
- Kamruzzaman, M., ElMasry, G., Sun, D.-W. & Allen, P. (2012b). 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.
- Kamruzzaman, M., ElMasry, G., Sun, D.-W. & Allen, P. (2013). Non-destructive assessment of instrumental and sensory tenderness of lamb meat using NIR hyperspectral imaging. *Food Chemistry*, **141**, 389-396.

- Lawrie, R.A. & Ledward, D.A. (2006). *Lawrie's meat science*. Pp. 1-417. Cambridge, England: Woodhead Publishing Limited.
- Leygonie, C., Britz, T.J. & Hoffman, L.C. (2012a). Impact of freezing and thawing on the quality of meat: Review. *Meat Science*, **91**, 93-98.
- Leygonie, C., Britz, T.J. & Hoffman, L.C. (2012b). Meat quality comparison between fresh and frozen/thawed ostrich *M. iliofibularis*. *Meat Science*, **91**, 364-368.
- Liu, Y. & Chen, Y.-R. (2001). Two-dimensional visible/near-infrared correlation spectroscopy study of thawing behavior of frozen chicken meats without exposure to air. *Meat Science*, **57**, 299-310.
- Lorenzo, J.M. & Pateiro, M. (2013). Influence of type of muscles on nutritional value of foal meat. *Meat Science*, **93**, 630-638.
- Lorenzo, J.M., Pateiro, M. & Franco, D. (2013). Influence of muscle type on physicochemical and sensory properties of foal meat. *Meat Science*, **94**, 77-83.
- Lorenzo, J.M., Sarriés, M.V., Tateo, A., Polidori, P., Franco, D. & Lanza, M. (2014). Carcass characteristics, meat quality and nutritional value of horsemeat: A review. *Meat Science*, **96**, 1478-1488.
- Ma, J., Pu, H., Sun, D.-W., Gao, W., Qu, J.-H. & Ma, K.-Y. (2015). Application of Vis-NIR hyperspectral imaging in classification between fresh and frozen-thawed pork *Longissimus Dorsi* muscles. *International Journal of Refrigeration*, **50**, 10-18.
- Ma, J., Sun, D.-W. & Pu, H. (2016). Spectral absorption index in hyperspectral image analysis for predicting moisture contents in pork longissimus dorsi muscles. *Food Chemistry*, **197**, 848-854.
- Mamani-Linares, L., Gallo, C. & Alomar, D. (2012). Identification of cattle, llama and horse meat by near infrared reflectance or transreflectance spectroscopy. *Meat Science*, **90**, 378-385.
- Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y. & Kaufman, L. (1988). *Chemometrics: A Textbook*. Pp. 477. Amsterdam: Elsevier Science Publishers.
- McElhinney, J., Downey, G. & Fearn, T. (1999). Chemometric processing of visible and near infrared reflectance spectra for species identification in selected raw homogenised meats. *Journal of Near Infrared Spectroscopy*, **7**, 145-154.
- Mellet, F.D. (1985). *The ostrich as meat animal - anatomical and muscle characteristics*. (MSc Thesis). South Africa: University of Stellenbosch.
- Miller, C.E. (2005). Chemometrics in Process Analytical Chemistry In: *Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries*. Pp. 226-324. Oxford, UK: Blackwell Publishing Ltd.
- Mitsumoto, M., Maeda, S., Mitsuhashi, T. & Ozawa, S. (1991). Near-infrared spectroscopy determination of physical and chemical characteristics in beef cuts. *Journal of Food Science*, **56**, 1493-1496.
- Neethling, J. (2016). *Factors influencing the flavour of the meat derived from South African game species*. (PhD Thesis). South Africa: University of Stellenbosch.

- North, M.K. (2015). The conditioning of springbok (*Antidorcas marsupialis*) meat: changes in texture and mechanisms involved. (MSc Thesis).South Africa: University of Stellenbosch.
- Onyango, C.A., Izumimoto, M. & Kutima, P.M. (1998). Comparison of some physical and chemical properties of selected game meats. *Meat Science*, **49**, 117-125.
- Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*. Essex, England: Longman Scientific and Technical.
- Prieto, N., Roehe, R., Lavin, P., Batten, G. & Andres, S. (2009). Application of near infrared reflectance spectroscopy to predict meat and meat products quality: A review. *Meat Sci*, **83**, 175-186.
- Pu, H., Sun, D.-W., Ma, J. & Cheng, J.-H. (2015). Classification of fresh and frozen-thawed pork muscles using visible and near infrared hyperspectral imaging and textural analysis. *Meat Science*, **99**, 81-88.
- Ropodi, A.I., Panagou, E.Z. & Nychas, G.J.E. (2018). Rapid detection of frozen-then-thawed minced beef using multispectral imaging and Fourier transform infrared spectroscopy. *Meat Science*, **135**, 142-147.
- Sales, J. (1996). Histological, biophysical, physical and chemical characteristics of different ostrich muscles. *Journal of the Science of Food and Agriculture*, **70**, 109-114.
- Savitzky, A. & Golay, M.J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, **36**, 1627-1639.
- Schmutzler, M., Beganovic, A., Böhler, G. & Huck, C.W. (2015). Methods for detection of pork adulteration in veal product based on FT-NIR spectroscopy for laboratory, industrial and on-site analysis. *Food Control*, **57**, 258-267.
- Talens, P., Mora, L., Morsy, N., Barbin, D.F., ElMasry, G. & Sun, D.-W. (2013). Prediction of water and protein contents and quality classification of Spanish cooked ham using NIR hyperspectral imaging. *Journal of Food Engineering*, **117**, 272-280.
- Tateo, A., De Palo, P., Ceci, E. & Centoducati, P. (2008). Physicochemical properties of meat of Italian Heavy Draft horses slaughtered at the age of eleven months. *Journal of Animal Science*, **86**, 1205-1214.
- Thyholt, K. & Isaksson, T. (1997). Differentiation of frozen and unfrozen beef using near-infrared spectroscopy. *Journal of the Science of Food and Agriculture*, **73**, 525-532.
- Van Zyl, L. & Ferreira, A.V. (2004). Physical and chemical carcass composition of springbok (*Antidorcas marsupialis*), blesbok (*Damaliscus dorcas phillipsi*) and impala (*Aepyceros melampus*). *Small Ruminant Research*, **53**, 103-109.
- Viljoen, M. (2003). *The use of near infrared reflectance spectroscopy (NIRS) for the chemical analysis of meat and feedstuffs*. (MSc Thesis).South Africa: University of Stellenbosch.
- Wold, S. & Sjöström, M. (1977). SIMCA: a method for analyzing chemical data in terms of similarity and analogy. In: *Chemometrics: Theory and Application*. Pp. 243 - 282. ACS Publications.

Chapter 5

General Discussion and Conclusion

Meat authenticity and traceability are important issues in modern society (Premanandh, 2013), as incidences regarding meat adulteration and fraud have become more sophisticated and mainstream (Cawthorn *et al.*, 2013). Food fraud is a collective term used for the deliberate and intentional substitution, addition, tampering or misrepresentation of food for economic gain (Spink & Moyer, 2011). Therefore, the main driving force of meat adulteration can be attributed to the ever-increasing prices of commercial meat products, the globalisation of food trade and the increased processing of meat into value-added products (Cawthorn *et al.*, 2013). Typical cases involve the intentional substitution of high value raw ingredients with inferior species or materials, the addition of non-declared proteins from several origins, or the marketing of frozen-thawed meat as fresh (reviewed by Ballin & Lametsch, 2008; Alamprese *et al.*, 2016). This type of food fraud concerns consumers in terms of economic loss, food allergies, religious compliance, and food safety (Dean *et al.*, 2006). This study aimed to investigate a feasible alternative to the manual, tedious and time-consuming conventional analytical methods used for meat differentiation and authentication that could provide the meat industry with a rapid, non-destructive, accurate and reliable automated solution in the near future.

Near-infrared (NIR) spectroscopy combined with multivariate data analysis (MDA) techniques were used to rapidly differentiate between South African game species, irrespective of the meat being fresh, previously frozen or the muscle type and determine these individual classes (fresh; previously frozen; frozen period; muscle type) per species. The study included two data sets. The first consisted of three species [zebra (*Equus quagga burchelli*), springbok (*Antidorcas marsupialis*), ostrich (*Struthio camelus*)], multiple muscle types for each species and the meat samples were frozen for one month. The second data set consisted of four species [black wildebeest (*Connochaetes gnou*), zebra (*Equus quagga burchelli*), springbok (*Antidorcas marsupialis*), ostrich (*Struthio camelus*)], one muscle type for each species and the meat samples were frozen for a period of nine months. With the first data set, the aim was to differentiate between three game species, irrespective of the treatment (fresh or previously frozen) or the muscle type, as well as determine the different muscle types and the effect thereof on the species class classification accuracies. The aim for the second data set was to differentiate between four game species, using one muscle type, irrespective of the treatment (fresh or previously frozen) or the frozen period (1 – 9 months), as well as predict the different frozen periods and determine the effect thereof on the classification accuracies. In both data sets the classification of fresh or previously frozen meat was also investigated as well as the effect thereof on the classification accuracies of the individual classes, e.g. species and muscle type.

The results from this study clearly showed that it was possible to differentiate between South African game species, irrespective of the meat being fresh, previously frozen or the muscle type. The partial least

squares discriminant analysis (PLS-DA) model, pre-processed with Savitzky-Golay (1st derivative, 2nd order polynomial, 7 points) was successful and achieved an accuracy of 93.2% for differentiating between zebra, springbok and ostrich. The spectral classification of the species was attributed to their physiochemical differences. The SGd₁(7) pre-processing achieved better results, as the separation due to differences in protein content and meat tenderness were enhanced. On the other hand, the PLS-DA model pre-processed with Savitzky-Golay (2nd derivative, 2nd order polynomial, 9 points) was successful for differentiating between black wildebeest, zebra, springbok and ostrich, achieving an 89.8% classification accuracy. SGd₂(9) outperformed the other pre-processing techniques investigated, as the Savitzky-Golay transformation enhanced the differences in pH, protein- and especially moisture content, which were the predominant contributors for species separation. The results from both data sets also illustrated that samples that resemble each other closely, due to certain similar physiochemical characteristics, were difficult to differentiate, thus accounting for the greatest source of misclassification. This suggests that enhanced spectral differences between species would result in improved classification accuracies and better prediction models.

These classification results correspond to what was found in literature, as previous studies reported that spectral derivatives have the potential to improve the classification accuracy of prediction models (Ding & Xu, 1999; McElhinney *et al.*, 1999; Schmutzler *et al.*, 2015). The results reported herein were also comparable to that described by McElhinney *et al.* (1999). The researchers suggested that the discrimination of species were based on compositional chemical differences. Cozzolino and Murray (2004) reported similar results (96%) for the identification of beef, lamb, pork and chicken meat and a discrimination was made between the species based on intra-muscular fat, fatty acids and moisture. Lastly, Mamani-Linares *et al.* (2012) differentiated between beef, llama and horse meat with accuracies between 89 and 100%. The results from the previously mentioned studies were comparable to that of the current study; therefore reinforcing that PLS-DA can be used to discriminate between the game meat species in question.

This technique was also used to distinguish between fresh and previously frozen meat as well as determine the frozen period. The principal component analysis (PCA) score plots illustrated good separation between the fresh and frozen-thawed samples, however, the frozen-thawed samples exhibited an overlap between the individual frozen periods. Therefore, lacking separation and distinct clustering for the different frozen periods. Two prominent bands were observed at 1093 nm, associated with pH (ElMasry *et al.*, 2012b), and 1422 nm, which is related to moisture (Osborne *et al.*, 1993), and these were considered to be responsible for the differences and separation between the fresh and frozen-thawed meat samples. Water is known to be a major component in fresh meat and constitutes about 70 - 85% (Cozzolino & Murray, 2004; Prieto *et al.*, 2009). Freezing and thawing mainly influences the water fraction of meat (Leygonie *et al.*, 2012a) and due to the formation of ice crystals causes damage to the cellular structure of the meat. As reviewed by Leygonie *et al.* (2012a), the disrupted muscle fibre structure results in a reduced water-holding capacity of meat. Therefore, the main difference between fresh and frozen-thawed meat can be attributed

to the loss of fluid from the meat tissue when being defrosted. This decrease in moisture may cause an increase in the concentration of solutes, which consequently results in a decrease in the pH. Leygonie *et al.* (2012b) reported that the pH of previously frozen meat tends to be lower than that of fresh meat. This phenomenon is supported by the current findings as the fresh meat samples were separated by higher loadings at 1093 nm, indicative of a higher pH (ElMasry *et al.*, 2012b). The current findings also illustrated that the fresh samples were separated by higher loadings at 1422 nm, indicative of an expected increased moisture content. Barbin *et al.* (2013) reported similar results for fresh and frozen-thawed pork meat and discriminated the samples based on complex physical and chemical changes caused by freezing.

The difference between fresh and frozen-thawed meat can also be attributed to changes in the physical structure (Prieto *et al.*, 2009), caused by the formation of ice crystals (Leygonie *et al.*, 2012b). Downey and Beauchêne (1997b) reported that freezing-and-thawing alters the physical structure of the meat's surface layer, consequently changing the total reflectance spectra. Therefore a discrimination between fresh and frozen-thawed beef could be made based on the spectral baseline shift induced by freeze-thawing (Downey & Beauchêne, 1997b). In addition to the physical change in structure, other chemical changes may also occur. Lipid- and protein oxidation as well as a decreased colour stability are all changes associated with previously frozen meat (Leygonie *et al.*, 2012b; reviewed by Leygonie *et al.*, 2012a). Thyholt and Isaksson (1997) successfully differentiated (90 – 100%) between frozen and unfrozen beef samples based on properties related to drip loss, irreversible denaturation and physical damage of myofibril proteins. Therefore, these results exhibit that both physical and chemical changes occurring within muscles during freeze-thawing can be used to successfully differentiate between fresh and previously frozen meat. The results in the current study also indicated that the physical and chemical changes caused by freeze-thawing were similar to that of previous studies (Downey & Beauchêne, 1997b; Downey & Beauchêne, 1997a; Thyholt & Isaksson, 1997; Leygonie *et al.*, 2012b; Barbin *et al.*, 2013). Therefore, indicating that the black wildebeest, zebra, springbok and ostrich species, irrespective of the muscle type or frozen period, all experience the same physical and chemical changes when subjected to frozen storage.

Although a good separation was observed between the fresh and frozen-thawed samples, an overlap was exhibited between the different frozen periods, thus resulting in a major misclassification between classes. The lack of separation can be ascribed to the samples' spectral similarities, as no substantial changes between the frozen periods were observed. Therefore, prolonged frozen storage does not seem to have a further effect on the previously mentioned physical and chemical changes. These results also indicated that the MDA models were not effective when trying to classify the different frozen periods. Downey and Beauchêne (1997b) reported similar results when attempting to discriminate between different freeze-thaw cycles. When considering the results in the current study alongside those reported in literature, it is possible to conclude that the differences in the spectra for fresh and frozen-thawed meat can be attributed to the initial damage done by freezing. Any additional freeze-thaw cycles or increased period of freezing does not account for further changes, as the damage done to the microstructures within meat has already been

achieved. This suggests that it is unnecessary to determine the frozen period of meat as the discrimination between fresh and previously frozen meat, irrespective of the frozen period, is sufficient.

Lastly, this technique was used to determine whether it is possible to differentiate between different muscle types. However, this study found it to be a challenging task. Lawrie and Ledward (2006) stated that the differences between muscles are very complex, therefore making it difficult to differentiate between the different muscle types. Literature also states that both the biochemical and chemical characteristics of the muscles' composition are intricate (Lawrie & Ledward, 2006). The muscle variability is known to be influenced by a large number of *intrinsic*- and *extrinsic* factors, therefore making it difficult to classify muscles based on fixed factors affecting the different muscles' composition (Lawrie & Ledward, 2006). However, differences between muscles, concerning moisture-, protein-, fat content, pH and meat tenderness have previously been quantified (Mallet, 1985; Sales, 1996; Onyango *et al.*, 1998; Viljoen, 2003; Du Buisson, 2006; Hoffman & Wiklund, 2006; Hoffman *et al.*, 2009; Engelbrecht, 2013; North, 2015; Neethling, 2016). Yet, various authors found that several parameters/factors contributed to the observed differences, even when investigating the same species or specific muscles among different species. Consequently, there is no real consensus regarding standard chemical parameters for specific muscles within a species or among species.

From the PCA results (score plots and loadings line plots), it was possible to conclude that the variation between muscle types can be attributed to differences in moisture-, fat-, protein content and pH as well as meat tenderness, as reported in literature. However, no correlation could be drawn between the specific factors and different muscles, within a species and among species, as the factors contributing to the separation differed for each species and muscle type. Although the PCA score plots illustrated minimal class separation between the different muscle samples, it was still necessary to apply classification algorithms to determine whether it would be possible to distinguish between the different muscle types.

This study found that the BD (*gastrocnemius*) and FF (*iliofibularis*) muscles for ostrich could be distinguished with a 100% accuracy, irrespective of the meat being fresh or previously frozen. Furthermore, the results suggested that discrimination of the different muscle types for both zebra and springbok was less sufficient due to the lack of separation between the different muscles. This phenomenon was ascribed to the samples' spectral similarities, hence causing misclassification between the classes. Misclassification mostly occurred between muscles that are anatomically close to one another. Consequently, it can be concluded that these muscles are closely related and that the differences in their physiochemical characteristics are therefore negligible. This suggested that the MDA models were not effective when trying to classify the different muscles. For this reason, it was deemed important to examine the success of a two-group discrimination model. The muscles were grouped together based on their anatomical location. Partial least squares discriminant analysis (PLS-DA) and quadratic discriminant analysis (QDA) models were developed to discriminate between the forequarters [IS (*infraspinatus*), SS (*supraspinatus*)] and hindquarters [BF (*biceps femoris*), fillet (*psoas major*), LTL (*longissimus thoracis et lumborum*), SM (*semimembranosus*), ST (*semitendinosus*)], irrespective of the meat being fresh or previously frozen.

The two-group class discrimination improved the overall accuracies of the PLS-DA and QDA models. Therefore, the results suggested that grouping samples with spectral similarities could be used to improve the discrimination of the samples as well as the overall model accuracies and performance measures. The results showed that it was possible to differentiate between the forequarters and hindquarters of the zebra (90.3%) and springbok (97.9%) muscles, irrespective of the meat being fresh or previously frozen. Further examination of the model results showed that the PLS-DA models were the best for differentiating between the muscles grouped as forequarters and hindquarters. The PLS-DA models achieved better results as it generally outperforms QDA, when classes are closely related, because it overcomes the collinearity problems often associated with QDA.

Due to the complexity of the data, it was not possible to classify the multiple classes (e.g. species, fresh vs. frozen-thawed, frozen period and muscle type) using a single classification and/or discrimination model. To solve this problem and handle the increased detail of the data, it was divided into sub-groups with individual models. This was achieved with multilevel hierarchical modelling of the individual PLS-DA models, in which the data were organised into a tree-like structure. This enabled the simultaneous identification of the species, treatment (fresh or frozen-thawed) and different muscle types, without having to use separate models for each class classification. The hierarchical model also highlighted the fact that different types of models (data pre-treatment and number of latent variables) may be suitable for each class discrimination step. Therefore, confirming that it was not possible to use a single model (same data pre-treatment and number of latent variables) to classify a complex multiple class classification.

The results showed NIR spectroscopy's potential as a rapid and non-destructive method for species identification, fresh and previously frozen meat differentiation as well as muscle type determination. However, conventional NIR spectroscopy has one shortcoming: it only provides spectral information about an entire sample in one spectrum. Thus, no spatial information is gleaned from conventional NIR spectroscopy. Therefore this work should be furthered by investigating the use of NIR hyperspectral imaging, as it incorporates both spatial and spectral information to overcome this problem (Feng *et al.*, 2013). In so doing, differentiation would potentially be improved so that results that are more accurate can be achieved. NIR hyperspectral imaging has shown promise in predicting the quality of meat and meat products (reviewed by ElMasry *et al.*, 2012a; ElMasry *et al.*, 2012b; Kamruzzaman *et al.*, 2012b; Kamruzzaman *et al.*, 2013a; Kamruzzaman *et al.*, 2016; Ma *et al.*, 2016), and in meat science it was successfully applied for authentication purposes (Kamruzzaman *et al.*, 2011; Kamruzzaman *et al.*, 2012a; Kamruzzaman *et al.*, 2013b; Kamruzzaman *et al.*, 2015; Pu *et al.*, 2015). These studies demonstrated that hyperspectral imaging, paired with multivariate image analysis and image processing, shows high potential as a rapid and objective method to classify different animal species (Kamruzzaman *et al.*, 2012a; Kamruzzaman *et al.*, 2013b) and muscle types (Kamruzzaman *et al.*, 2011), as well as fresh and frozen-thawed meat (Barbin *et al.*, 2013; Ma *et al.*, 2015; Pu *et al.*, 2015). By optimising wavelength selection, it is also possible to produce a cheap multispectral imaging

instrument that can be used as rapid and reliable alternative to traditional analytical methods (Kamruzzaman *et al.*, 2015).

Furthermore, this work should be furthered with the focus placed on extending the database by including more game species from different regions to develop a robust classification model that can be used for meat authentication and origin detection. Future work should also investigate the classification and differentiation potential of game meat emulsions containing different adulterants at varying concentrations. This should consider other species (game and non-game), adulterants often used for meat substitution (inferior species, protein and fat) and non-meat ingredient addition (water and additives), as well as determine the limits of detection for each of the adulterants in question. In conjunction with the proposed recommendations, more laboratory based compositional data must be acquired using wet chemistry techniques to quantify the differences due to the vast amount of variation within the samples. In this dissertation the optimal temperature for scanning the meat was chosen to be *ca.* 23° C and the meat was left to bloom for *ca.* 30 – 60 min prior to scanning. Thus, future research should consider other temperatures, particularly lower temperatures typically used in industry, as well as compare the effect of blooming and the absence thereof on the spectra and model accuracies. Nonetheless, the results presented in this study serve as a baseline for future work.

This study showed that a portable NIR spectrophotometer, with a spectral range of 908 – 1676 nm, is a suitable instrument to use for distinguishing game meat samples. Therefore, this method holds ample potential for the authentication of game meat, as it can be used as a rapid alert system against the adulteration or contamination of meat products. And in so doing, NIR spectroscopy can prevent retailers from offering fraudulent meat products and consequently protect consumers from adulteration and/or harmful frauds. On the other hand, the majority of the research performed make use of benchtop NIR devices, which are mostly suitable for laboratory setup. Thus, making this research, using a portable NIR spectrophotometer, unique as it is the first such work done on these exotic and alternative species. Portable handheld devices make it possible to take measurements at any time and place, and obtain immediate results. Considering the cost of the instrument, ease of measurement and flexibility of the instrumental set-up, it is probable that the game meat industry would, in the near future, have a rapid on-site measurement tool to authenticate game meat in mislabelling attempts.

In conclusion, NIR spectroscopy combined with MDA offers an accurate and reliable technique for the rapid identification and authentication of species and muscles in, fresh as well as frozen-thawed game meat products. This study has the potential of providing the South African game meat industry with an alternative technique to the current manual, destructive and time-consuming methods used to detect fraud, thus contributing to the authenticity and fair-trade of game meat locally and internationally.

5.1 References

- Alamprese, C., Amigo, J.M., Casiraghi, E. & Engelsen, S.B. (2016). Identification and quantification of turkey meat adulteration in fresh, frozen-thawed and cooked minced beef by FT-NIR spectroscopy and chemometrics. *Meat Science*, **121**, 175-181.
- Ballin, N.Z. & Lametsch, R. (2008). Analytical methods for authentication of fresh vs. thawed meat - A review. *Meat Science*, **80**, 151-158.
- Barbin, D.F., Sun, D.-W. & Su, C. (2013). NIR hyperspectral imaging as non-destructive evaluation tool for the recognition of fresh and frozen-thawed porcine *longissimus dorsi* muscles. *Innovative Food Science & Emerging Technologies*, **18**, 226-236.
- Cawthorn, D.-M., Steinman, H.A. & Hoffman, L.C. (2013). A high incidence of species substitution and mislabelling detected in meat products sold in South Africa. *Food Control*, **32**, 440-449.
- Cozzolino, D. & Murray, I. (2004). Identification of animal meat muscles by visible and near infrared reflectance spectroscopy. *LWT-Food Science and Technology*, **37**, 447-452.
- Dean, N., Murphy, T.B. & Downey, G. (2006). Using unlabelled data to update classification rules with applications in food authenticity studies. *Journal of the Royal Statistical Society: Applied Statistics*, **55**, 1-14.
- Ding, H. & Xu, R. (1999). Differentiation of Beef and Kangaroo Meat by Visible/Near-Infrared Reflectance Spectroscopy. *Journal of Food Science*, **64**, 814-817.
- Downey, G. & Beauchêne, D. (1997a). Authentication of fresh vs. frozen-then-thawed beef by near-infrared reflectance spectroscopy of dried drip juice. *LWT-Food Science and Technology*, **30**, 721-726.
- Downey, G. & Beauchêne, D. (1997b). Discrimination between fresh and frozen-then-thawed beef *m. longissimus dorsi* by combined visible-near infrared reflectance spectroscopy: A feasibility study. *Meat Science*, **45**, 353-363.
- Du Buisson, P.-M. (2006). Improving the meat quality of blesbok (*Damaliscus dorcas phillipsi*) and springbok (*Antidorcas marsupialis*) trough enhancement with inorganic salts. (MSc Thesis).South Africa: University of Stellenbosch.
- ElMasry, G., Kamruzzaman, M., Sun, D.-W. & Allen, P. (2012a). Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: a review. *Critical Reviews in Food Science and Nutrition*, **52**, 999-1023.
- ElMasry, G., Sun, D.-W. & Allen, P. (2012b). Near-infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef. *Journal of Food Engineering*, **110**, 127-140.
- Engelbrecht, A. (2013). Establishing genetic and environmental parameters for ostrich (*Struthio camelus domesticus*) growth and slaughter characteristics. (PhD Thesis).South Africa: University of Stellenbosch

- Feng, Y., ElMasry, G., Sun, D., 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.
- Hoffman, L.C., Schalkwyk, S.v. & Muller, N. (2009). Effect of season and gender on the physical and chemical composition of black wildebeest (*Connochaetus gnou*) meat. *South African Journal of Wildlife Research*, **39**, 170-174.
- Hoffman, L.C. & Wiklund, E. (2006). Game and venison—meat for the modern consumer. *Meat Science*, **74**, 197-208.
- Kamruzzaman, M., Barbin, D., ElMasry, G., Sun, D.-W. & Allen, P. (2012a). Potential of hyperspectral imaging and pattern recognition for categorization and authentication of red meat. *Innovative Food Science & Emerging Technologies*, **16**, 316-325.
- Kamruzzaman, M., ElMasry, G., Sun, D.-W. & Allen, P. (2011). Application of NIR hyperspectral imaging for discrimination of lamb muscles. *Journal of Food Engineering*, **104**, 332-340.
- Kamruzzaman, M., ElMasry, G., Sun, D.-W. & Allen, P. (2012b). 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.
- Kamruzzaman, M., ElMasry, G., Sun, D.-W. & Allen, P. (2013a). Non-destructive assessment of instrumental and sensory tenderness of lamb meat using NIR hyperspectral imaging. *Food Chemistry*, **141**, 389-396.
- Kamruzzaman, M., Makino, Y. & Oshita, S. (2016). Hyperspectral imaging for real-time monitoring of water holding capacity in red meat. *LWT-Food Science and Technology*, **66**, 685-691.
- Kamruzzaman, M., Makino, Y., Oshita, S. & Liu, S. (2015). Assessment of visible near-infrared hyperspectral imaging as a tool for detection of horsemeat adulteration in minced beef. *Food and Bioprocess Technology*, **8**, 1054-1062.
- Kamruzzaman, M., Sun, D.-W., ElMasry, G. & Allen, P. (2013b). Fast detection and visualization of minced lamb meat adulteration using NIR hyperspectral imaging and multivariate image analysis. *Talanta*, **103**, 130-136.
- Lawrie, R.A. & Ledward, D.A. (2006). *Lawrie's meat science*. Pp. 1-417. Cambridge, England: Woodhead Publishing Limited.
- Leygonie, C., Britz, T.J. & Hoffman, L.C. (2012a). Impact of freezing and thawing on the quality of meat: Review. *Meat Science*, **91**, 93-98.
- Leygonie, C., Britz, T.J. & Hoffman, L.C. (2012b). Meat quality comparison between fresh and frozen/thawed ostrich *M. iliofibularis*. *Meat Science*, **91**, 364-368.
- Ma, J., Pu, H., Sun, D.-W., Gao, W., Qu, J.-H. & Ma, K.-Y. (2015). Application of Vis–NIR hyperspectral imaging in classification between fresh and frozen-thawed pork *Longissimus Dorsi* muscles. *International Journal of Refrigeration*, **50**, 10-18.

- Ma, J., Sun, D.-W. & Pu, H. (2016). Spectral absorption index in hyperspectral image analysis for predicting moisture contents in pork longissimus dorsi muscles. *Food Chemistry*, **197**, 848-854.
- Mamani-Linares, L., Gallo, C. & Alomar, D. (2012). Identification of cattle, llama and horse meat by near infrared reflectance or transreflectance spectroscopy. *Meat Science*, **90**, 378-385.
- McElhinney, J., Downey, G. & Fearn, T. (1999). Chemometric processing of visible and near infrared reflectance spectra for species identification in selected raw homogenised meats. *Journal of Near Infrared Spectroscopy*, **7**, 145-154.
- Mellet, F.D. (1985). *The ostrich as meat animal - anatomical and muscle characteristics*. (MSc Thesis). South Africa: University of Stellenbosch.
- Neethling, J. (2016). *Factors influencing the flavour of the meat derived from South African game species*. (PhD Thesis). South Africa: University of Stellenbosch.
- North, M.K. (2015). The conditioning of springbok (*Antidorcas marsupialis*) meat: changes in texture and mechanisms involved. (MSc Thesis). South Africa: University of Stellenbosch.
- Onyango, C.A., Izumimoto, M. & Kutima, P.M. (1998). Comparison of some physical and chemical properties of selected game meats. *Meat Science*, **49**, 117-125.
- Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*. Essex, England: Longman Scientific and Technical.
- Premanandh, J. (2013). Horse meat scandal – A wake-up call for regulatory authorities. *Food Control*, **34**, 568-569.
- Prieto, N., Roehe, R., Lavin, P., Batten, G. & Andres, S. (2009). Application of near infrared reflectance spectroscopy to predict meat and meat products quality: A review. *Meat Sci*, **83**, 175-186.
- Pu, H., Sun, D.-W., Ma, J. & Cheng, J.-H. (2015). Classification of fresh and frozen-thawed pork muscles using visible and near infrared hyperspectral imaging and textural analysis. *Meat Science*, **99**, 81-88.
- Sales, J. (1996). Histological, biophysical, physical and chemical characteristics of different ostrich muscles. *Journal of the Science of Food and Agriculture*, **70**, 109-114.
- Schmutzler, M., Beganovic, A., Böhler, G. & Huck, C.W. (2015). Methods for detection of pork adulteration in veal product based on FT-NIR spectroscopy for laboratory, industrial and on-site analysis. *Food Control*, **57**, 258-267.
- Spink, J. & Moyer, D.C. (2011). Defining the public health threat of food fraud. *Journal of Food Science*, **76**, 157-163.
- Thyholt, K. & Isaksson, T. (1997). Differentiation of frozen and unfrozen beef using near-infrared spectroscopy. *Journal of the Science of Food and Agriculture*, **73**, 525-532.
- Viljoen, M. (2003). *The use of near infrared reflectance spectroscopy (NIRS) for the chemical analysis of meat and feedstuffs*. (MSc Thesis). South Africa: University of Stellenbosch.

Addendum A

Supplementary information pertaining to Chapter 4:

4.1 Species determination

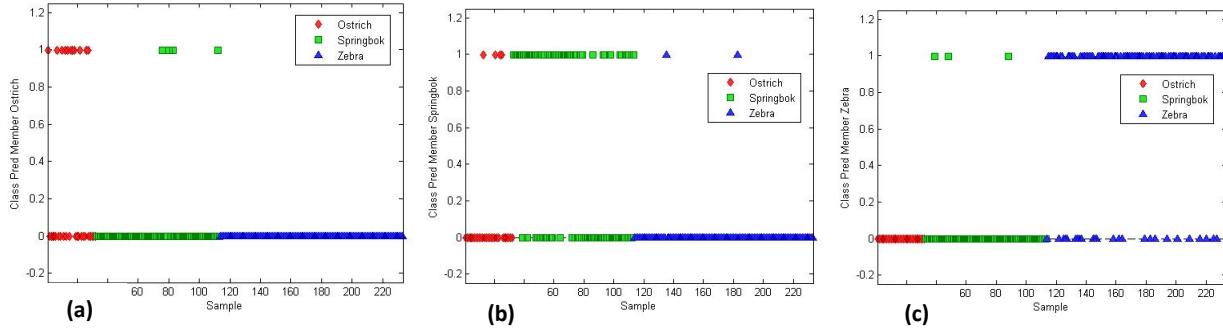


Figure A1 KNN ($k = 2$) classification (SNV + detrend pre-processed) resulting in an overall classification accuracy of 74.7%. KNN classification score plot of ostrich (red), springbok (green) and zebra (blue), illustrating the predicted objects. (a) Score plot of predicted ostrich (top) vs. remaining species (bottom), (b) score plot of predicted springbok (top) vs. remaining species (bottom) and (c) score plot of predicted zebra (top) vs. remaining species (bottom).

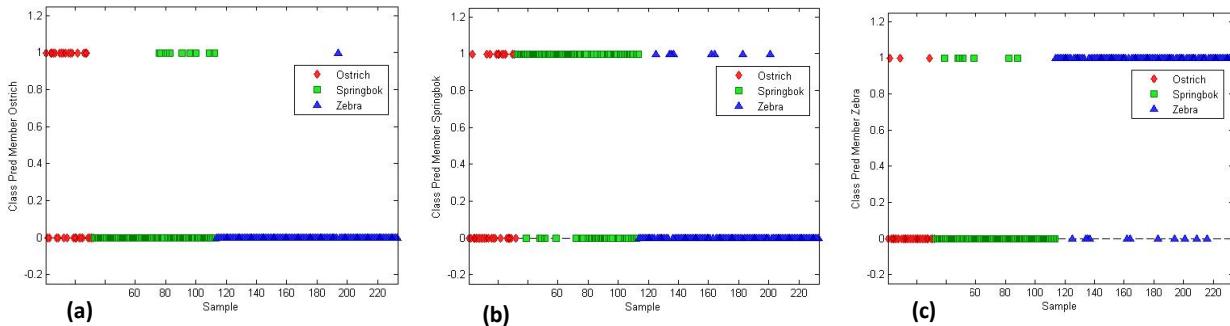


Figure A2 KNN ($k = 3$) classification (SNV + detrend pre-processed) resulting in an overall classification accuracy of 78.1%. KNN classification score plot of ostrich (red), springbok (green) and zebra (blue), illustrating the predicted objects. (a) Score plot of predicted ostrich (top) vs. remaining species (bottom), (b) score plot of predicted springbok (top) vs. remaining species (bottom) and (c) score plot of predicted zebra (top) vs. remaining species (bottom).

Table A1 The performance measures used to assess the LDA classification models of the three species, pre-processed with SNV + detrend.

Number of principal components (PCs)	Species	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
5	Zebra	89.4	7.5	3.1	93.6	85.3	85.8	89.6	10.6
	Springbok	86.7	6.0	7.3	79.8	90.6	82.7	81.2	13.3
	Ostrich	96.7	0	3.4	82.1	100.0	100.0	90.1	3.4
6	Zebra	90.7	6.2	3.1	93.8	87.5	88.3	91.0	9.3
	Springbok	87.6	6.4	6.0	82.5	90.2	81.5	82.0	12.5
	Ostrich	96.2	0	3.8	80.0	100.0	100.0	89.0	3.8

(LDA) Linear discriminant analysis; (SNV) Standard normal variate.

Table A2 The performance measures used to assess the QDA classification models of the three species, pre-processed with SNV + detrend.

Number of principal components (PCs)	Species	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
5	Zebra	91.6	6.2	2.2	95.5	87.8	88.3	91.8	8.4
	Springbok	89.2	4.3	6.5	82.6	93.2	87.7	85.0	10.8
	Ostrich	96.3	0.9	2.8	83.3	98.9	93.8	88.2	3.7
6	Zebra	90.8	6.1	3.1	93.8	87.8	88.3	91.0	9.2
	Springbok	88.8	4.3	6.9	81.6	93.2	87.7	84.5	11.2
	Ostrich	97.6	0.9	1.4	90.9	98.9	93.8	92.3	2.4

(QDA) Quadratic discriminant analysis; (SNV) Standard normal variate.

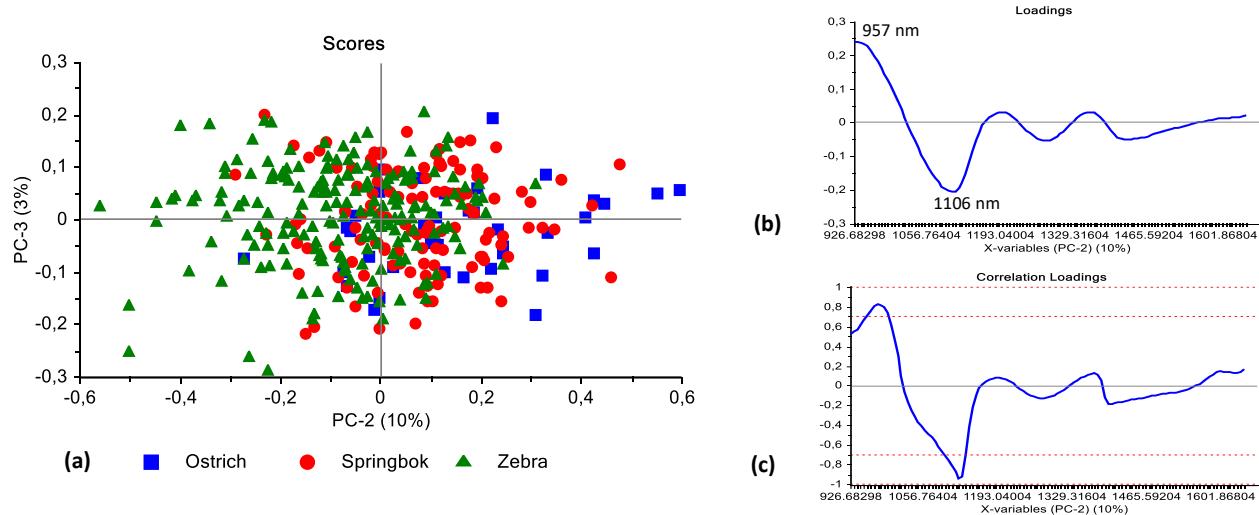


Figure A3 PCA analysis (SNV pre-processed) of species for ostrich (blue), springbok (red) and zebra (green) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (10%) vs. PC3 (3%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 957 and 1106 nm.

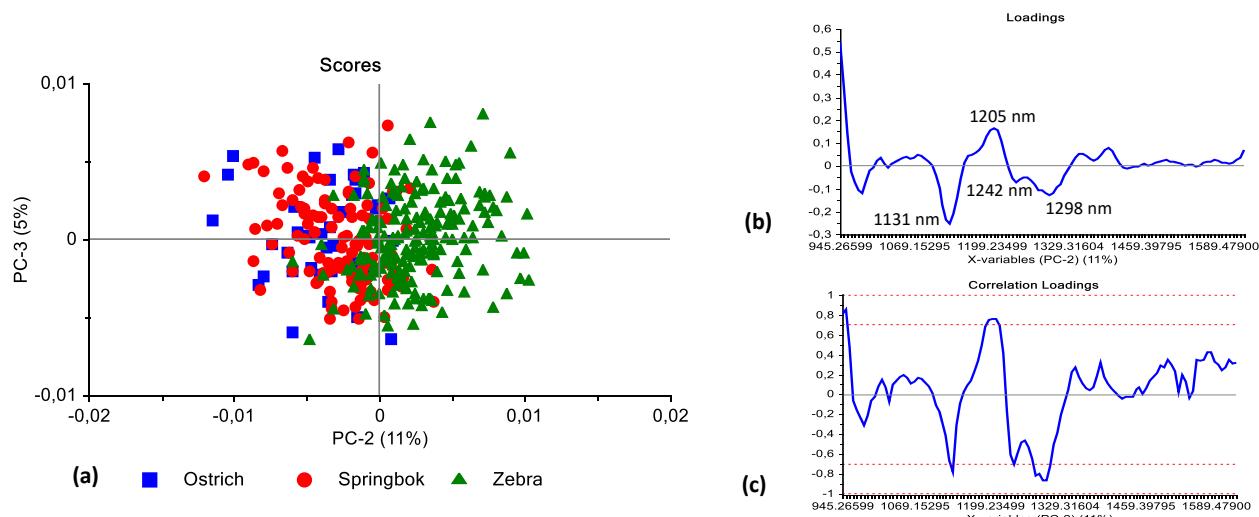


Figure A4 PCA analysis (SNV + SGd₂(7) pre-processed) of species for ostrich (blue), springbok (red) and zebra (green) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (11%) vs. PC3 (5%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 1131, 1205, 1242 and 1298 nm.

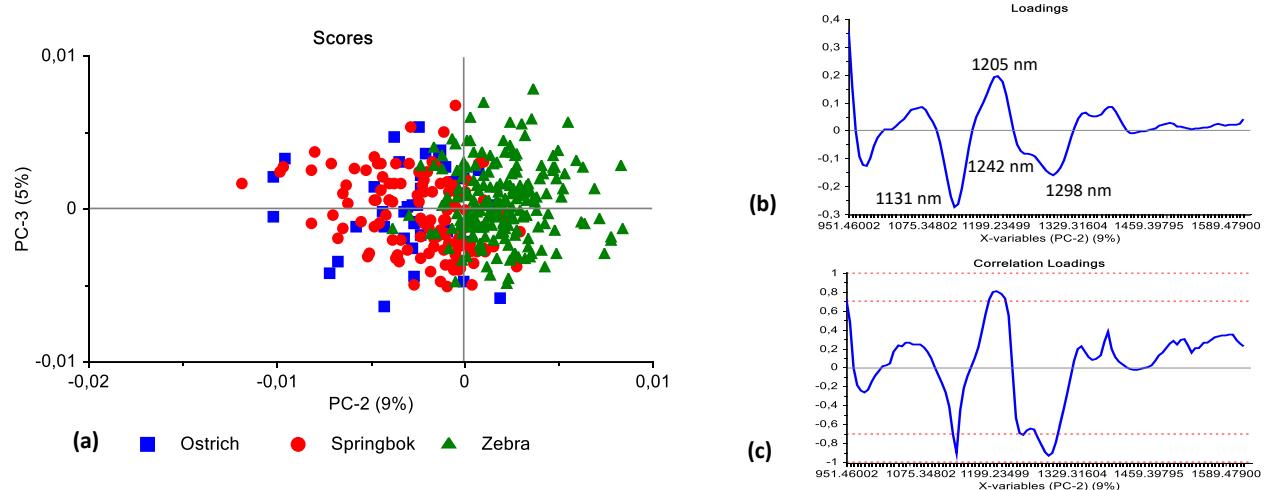


Figure A5 PCA analysis (SNV + SGd₂(9) pre-processed) of species for ostrich (blue), springbok (red) and zebra (green) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (9%) vs. PC3 (5%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 1131, 1205, 1242 and 1298 nm.

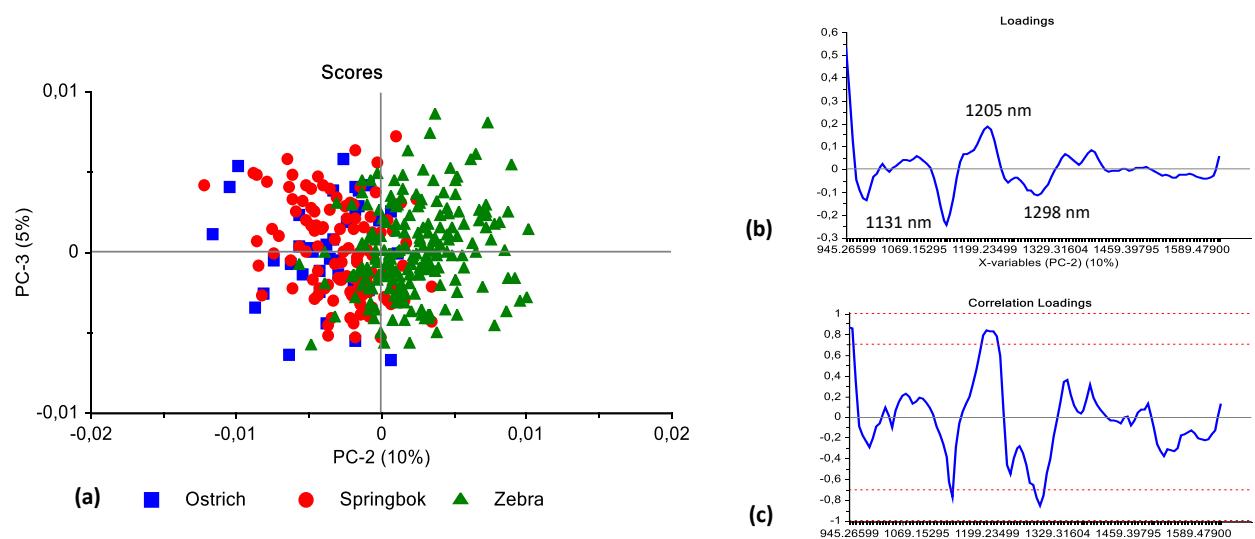


Figure A6 PCA analysis (SNV + detrend + SGd₂(7) pre-processed) of species for ostrich (blue), springbok (red) and zebra (green) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (10%) vs. PC3 (5%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 1131, 1205 and 1298 nm.

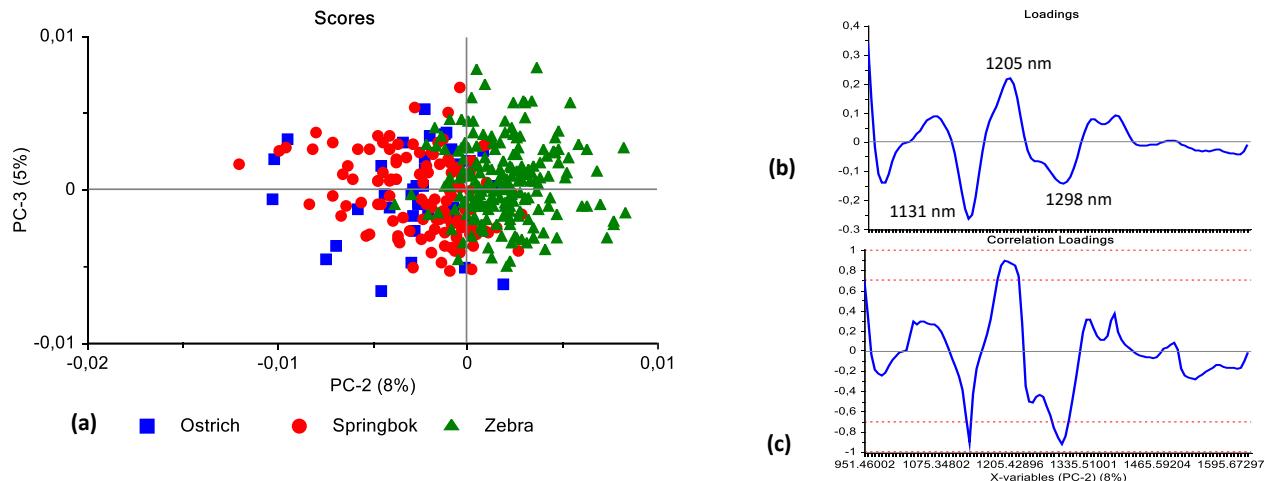


Figure A7 PCA analysis (SNV + detrend + SGd₂(9) pre-processed) of species for ostrich (blue), springbok (red) and zebra (green) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (8%) vs. PC3 (5%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 1131, 1205 and 1298 nm.

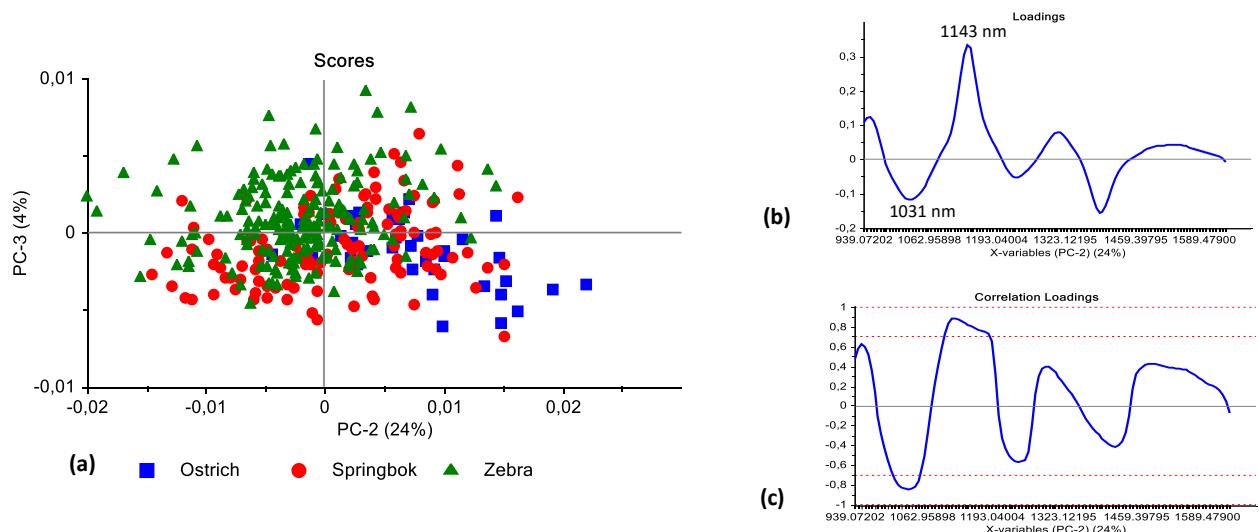


Figure A8 PCA analysis (SGd₁(5) pre-processed) of species for ostrich (blue), springbok (red) and zebra (green) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (24%) vs. PC3 (4%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 1031 and 1143 nm.

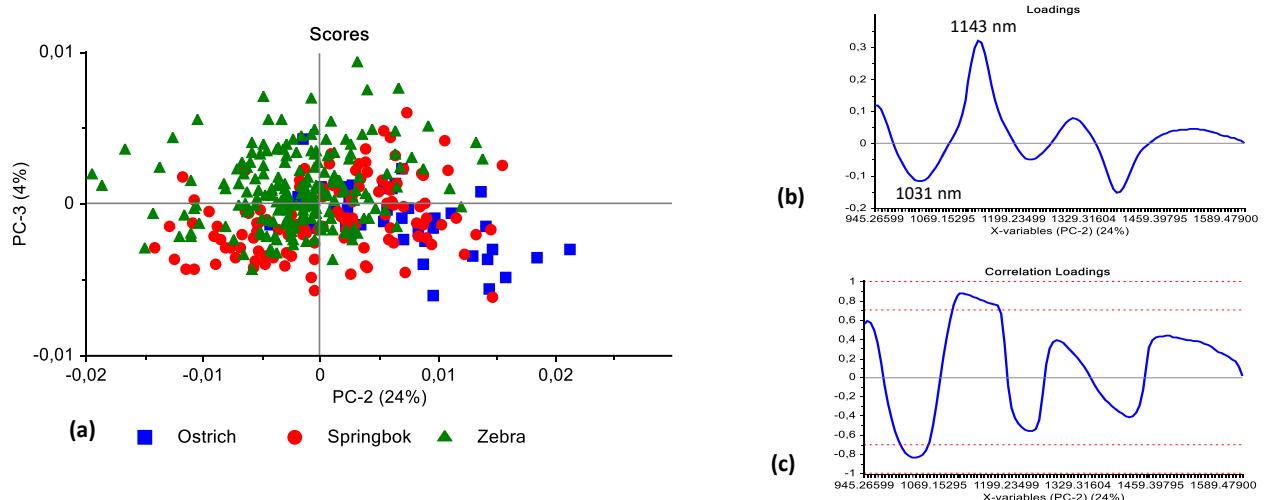


Figure A9 PCA analysis (SGd₁(7) pre-processed) of species for ostrich (blue), springbok (red) and zebra (green) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (24%) vs. PC3 (4%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 1031 and 1143 nm.

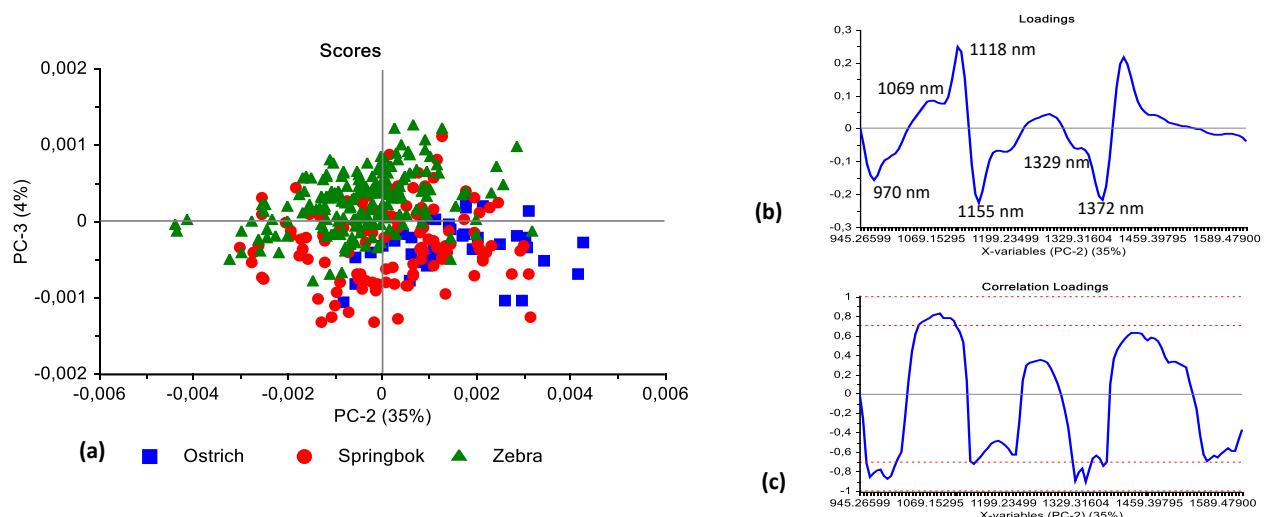


Figure A10 PCA analysis (SGd₂(7) pre-processed) of species for ostrich (blue), springbok (red) and zebra (green) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (35%) vs. PC3 (4%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 970, 1069, 1118, 1155, 1329 and 1372 nm.

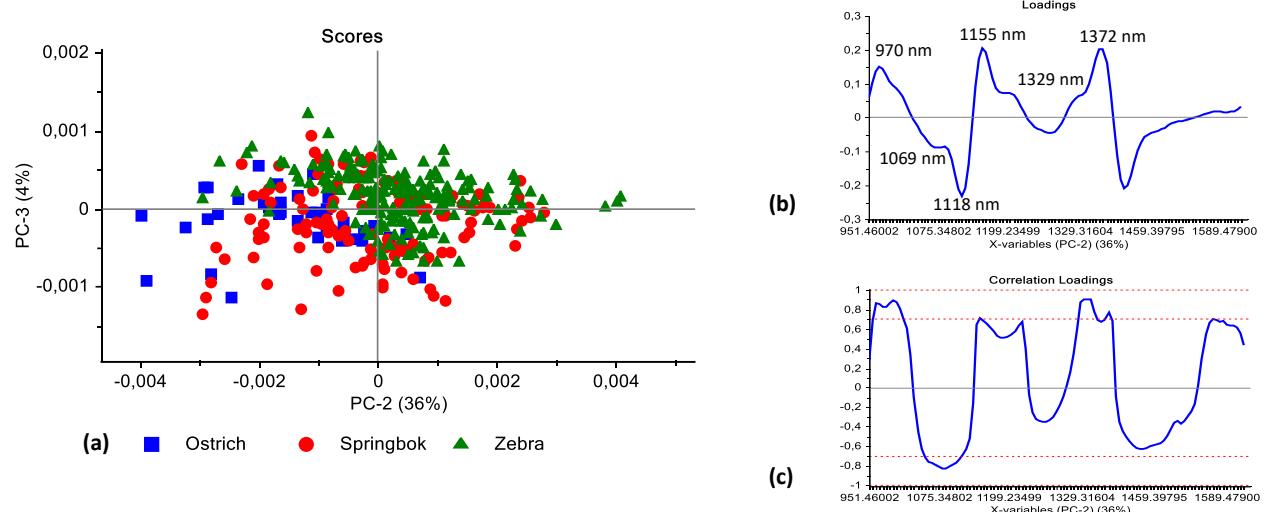


Figure A11 PCA analysis ($\text{SGd}_2(9)$ pre-processed) of species for ostrich (blue), springbok (red) and zebra (green) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (36%) vs. PC3 (4%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 970, 1069, 1118, 1155, 1329 and 1372 nm.

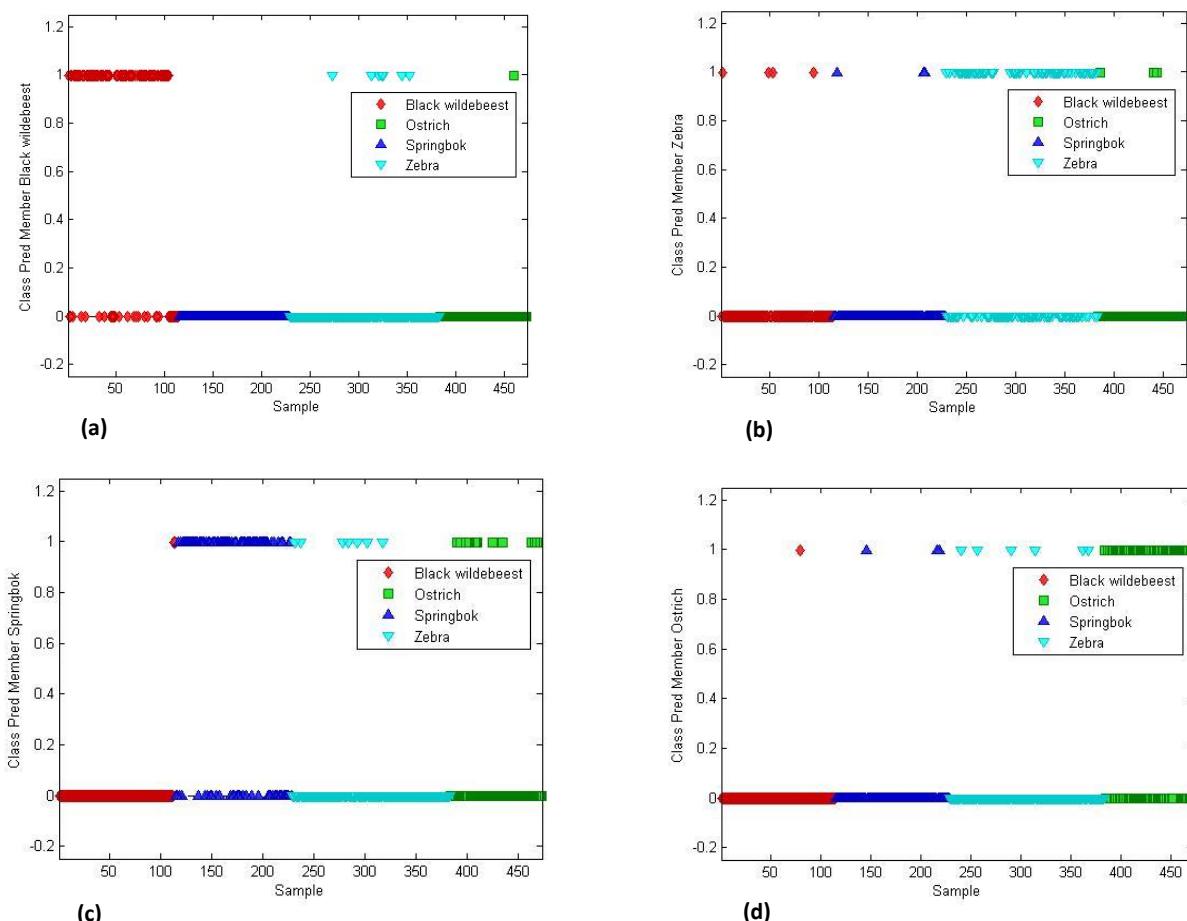


Figure A12 KNN ($k = 2$) classification (SNV + detrend pre-processed) resulting in an overall classification accuracy of 68.7%. KNN classification score plot of black wildebeest (red), ostrich (green), springbok (blue) and zebra (turquoise),

illustrating the predicted objects. (a) Score plot for predicted black wildebeest (top) vs. remaining species (bottom), (b) score plot of predicted zebra (top) vs. remaining species (bottom), (c) score plot of predicted springbok (top) vs. remaining species (bottom) and (d) score plot of predicted ostrich (top) vs. remaining species (bottom).

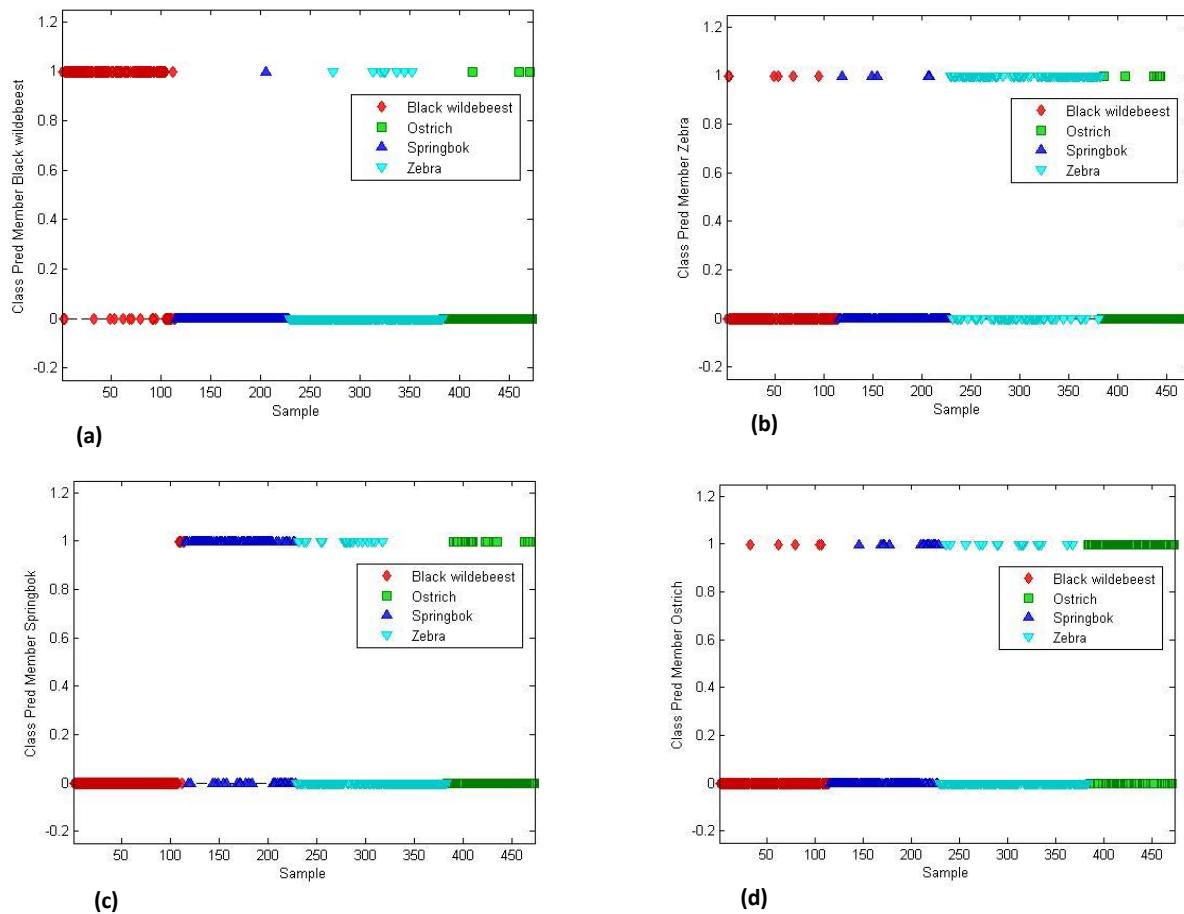


Figure A13 KNN ($k = 3$) classification (SNV + detrend pre-processed) resulting in an overall classification accuracy of 73.6%. KNN classification score plot of black wildebeest (red), ostrich (green), springbok (blue) and zebra (turquoise), illustrating the predicted objects. (a) Score plot for predicted black wildebeest (top) vs. remaining species (bottom), (b) score plot of predicted zebra (top) vs. remaining species (bottom), (c) score plot of predicted springbok (top) vs. remaining species (bottom) and (d) score plot of predicted ostrich (top) vs. remaining species (bottom).

Table A3 The performance measures used to assess the LDA models of the four species, pre-processed with SNV + detrend.

Number of principal components (PCs)	Species	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
5	BWB	89.8	5.8	4.5	84.7	91.9	80.7	82.5	10.2
	Zebra	82.2	13.7	4.1	85.2	81.1	63.2	72.6	17.8
	Springbok	83.6	6.4	10.0	68.2	90.7	77.2	72.4	16.4
	Ostrich	80.7	6.1	13.2	53.3	91.5	71.1	61.0	19.3
6	BWB	91.2	5.7	3.1	88.5	92.2	80.7	84.4	8.8
	Zebra	82.8	12.9	4.2	84.8	82.1	64.5	73.3	17.2
	Springbok	84.6	6.5	8.9	70.2	90.8	76.3	73.1	15.4
	Ostrich	83.2	4.0	12.8	57.5	94.3	81.1	67.3	16.8

(LDA) Linear discriminant analysis; (SNV) Standard normal variate; (BWB) Black wildebeest.

Table A4 The performance measures used to assess the LDA [mahalanobis distance] models of the four species, pre-processed with SNV + detrend.

Number of principal components (PCs)	Species	Classification accuracy (%)	False positive error (%)	False negative error (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	Misclassification rate (%)
5	BWB	91.0	5.1	3.9	86.2	92.9	82.5	84.3	9.0
	Zebra	79.8	4.9	15.3	66.2	91.0	85.8	74.7	20.2
	Springbok	87.2	8.4	4.4	81.6	89.0	70.2	75.5	12.8
	Ostrich	85.8	10.1	4.1	73.9	88.0	53.3	61.9	14.3
6	BWB	92.5	4.2	3.2	88.2	94.2	85.1	86.6	7.5
	Zebra	83.0	4.3	12.8	70.5	92.5	87.7	78.2	17.0
	Springbok	89.4	7.2	3.4	85.7	90.5	73.7	79.3	10.6
	Ostrich	87.3	8.5	4.2	75.0	89.8	60.0	66.7	12.7

(LDA) Linear discriminant analysis [mahalanobis distance]; (SNV) Standard normal variate; (BWB) Black wildebeest.

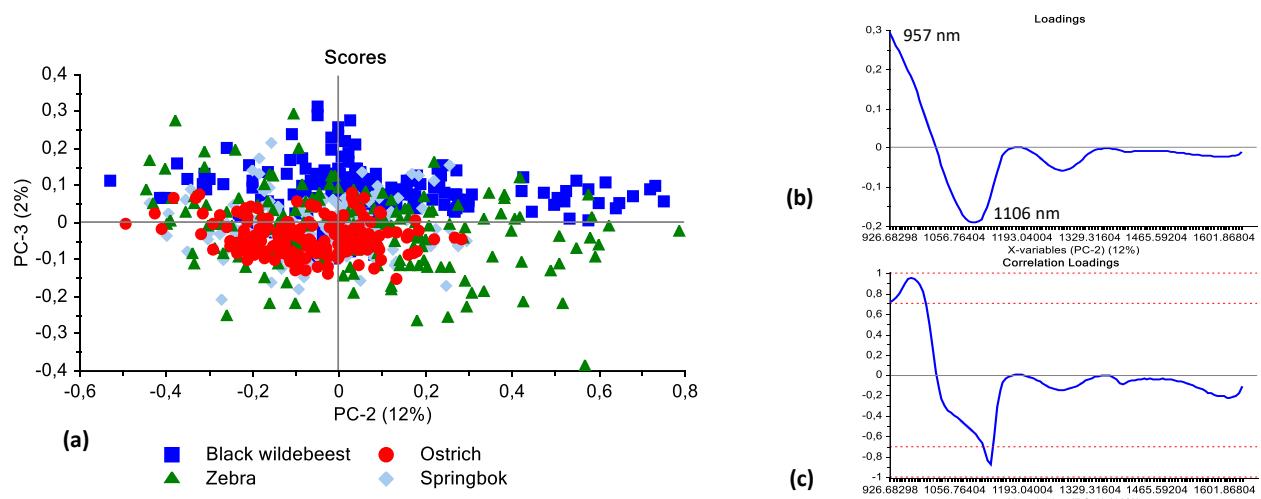


Figure A14 PCA analysis (SNV pre-processed) of species for black wildebeest (blue), zebra (green), springbok (light blue) and ostrich (red) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (12%) vs. PC3 (2%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 957 and 1106 nm.

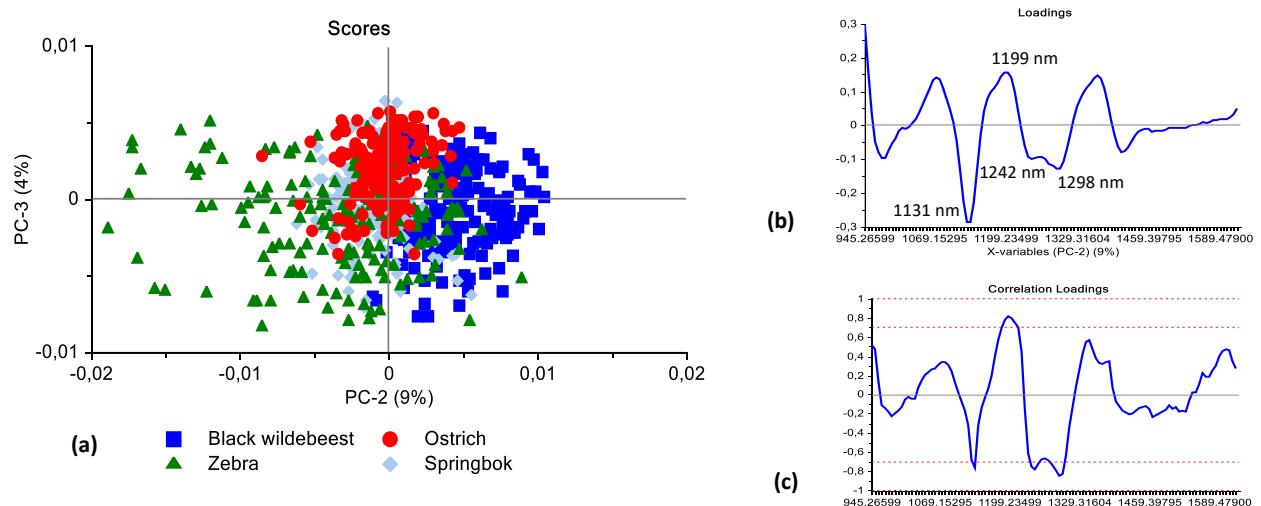


Figure A15 PCA analysis (SNV + SGd₂(7) pre-processed) of species for black wildebeest (blue), zebra (green), springbok (light blue) and ostrich (red) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (9%) vs. PC3 (4%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 1131, 1199, 1242 and 1298 nm.

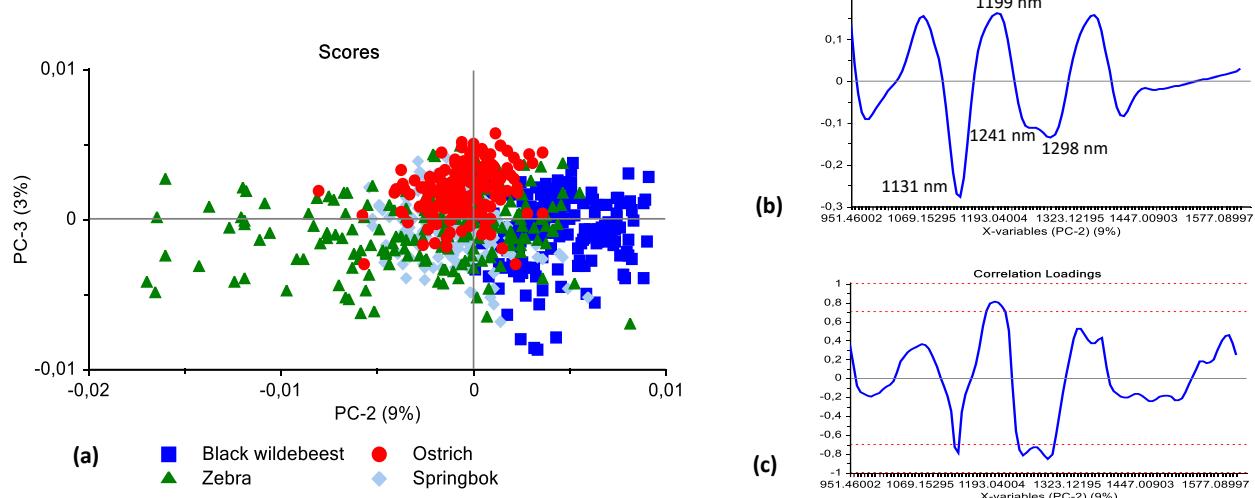


Figure A16 PCA analysis (SNV + SGd₂(9) pre-processed) of species for black wildebeest (**blue**), zebra (**green**), springbok (**light blue**) and ostrich (**red**) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (9%) vs. PC3 (3%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 1131, 1199, 1242 and 1298 nm.

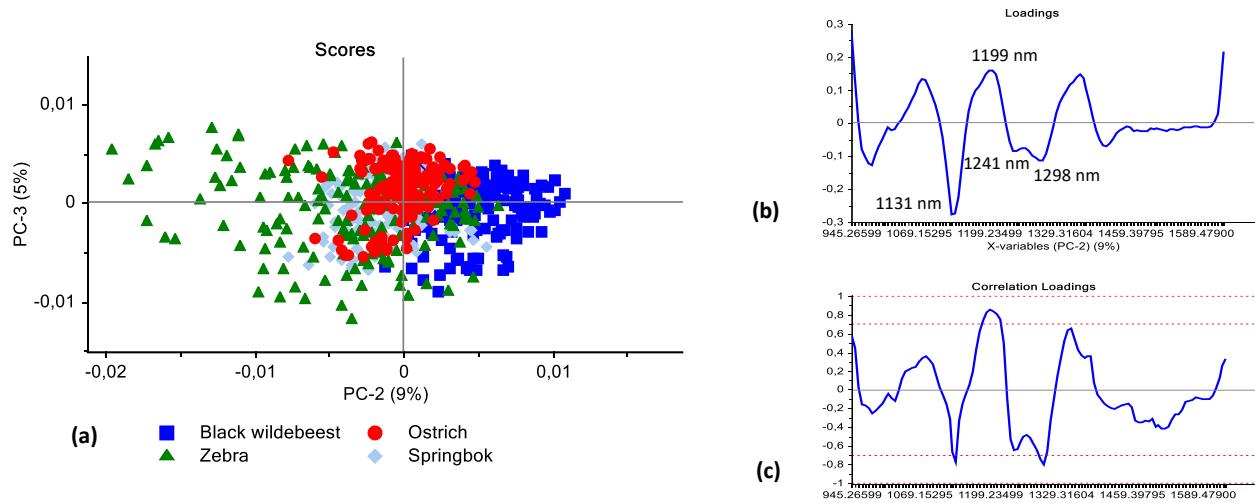


Figure A17 PCA analysis (SNV + detrend + SGd₂(7) pre-processed) of species for black wildebeest (**blue**), zebra (**green**), springbok (**light blue**) and ostrich (**red**) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (9%) vs. PC3 (5%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 1131, 1199 and 1298 nm.

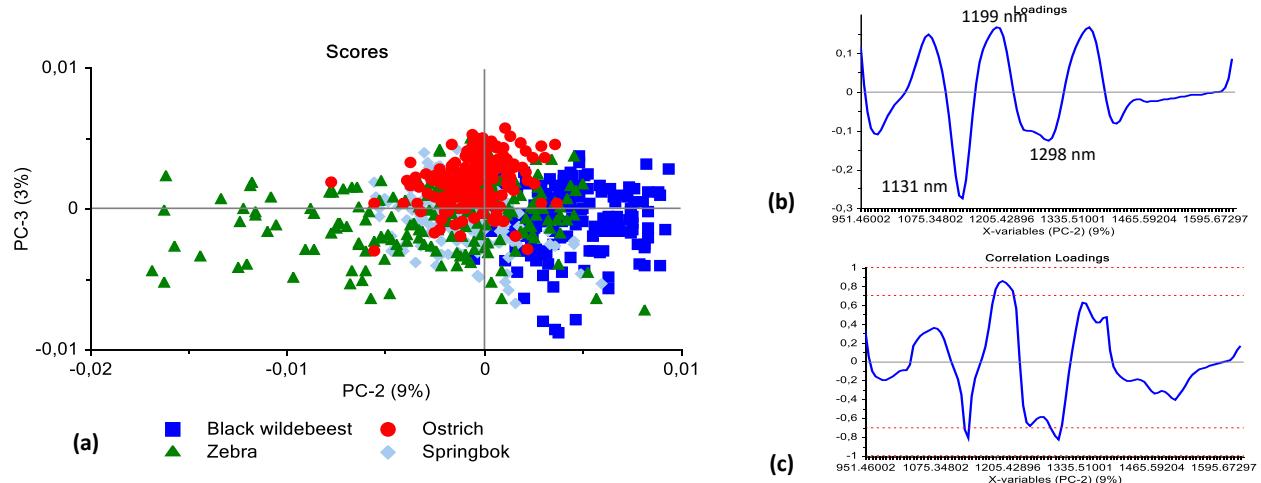


Figure A18 PCA analysis (SNV + detrend + SGd₂(7) pre-processed) of species for black wildebeest (**blue**), zebra (**green**), springbok (**light blue**) and ostrich (**red**) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (9%) vs. PC3 (3%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 1131, 1199 and 1298 nm.

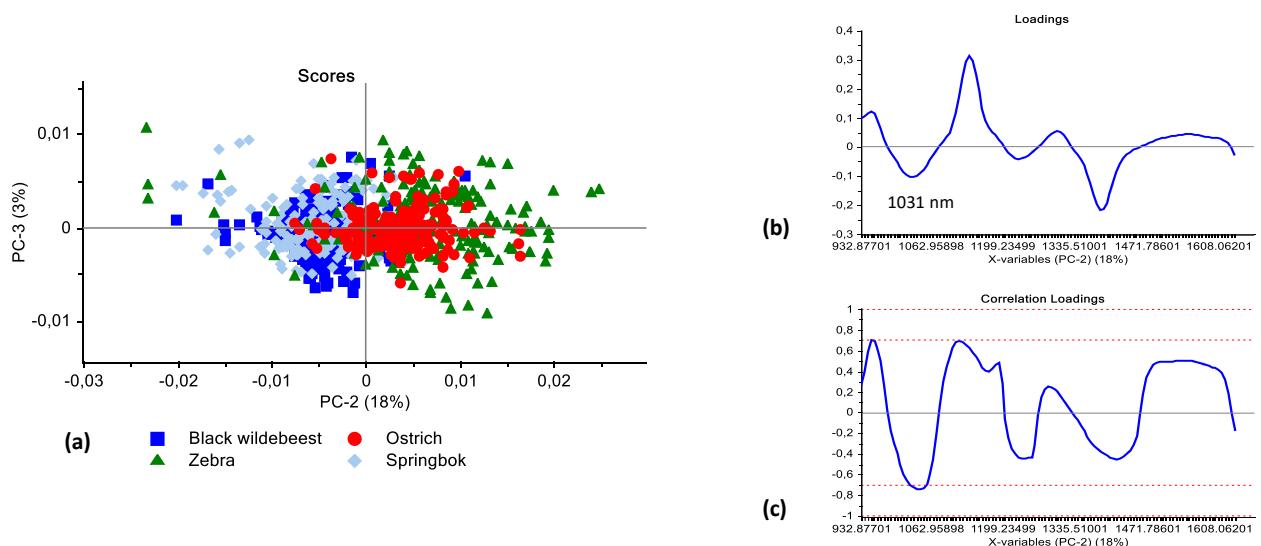


Figure A19 PCA analysis (SGd₁(5) pre-processed) of species for black wildebeest (**blue**), zebra (**green**), springbok (**light blue**) and ostrich (**red**) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (18%) vs. PC3 (3%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with an interpretable band at 1031 nm.

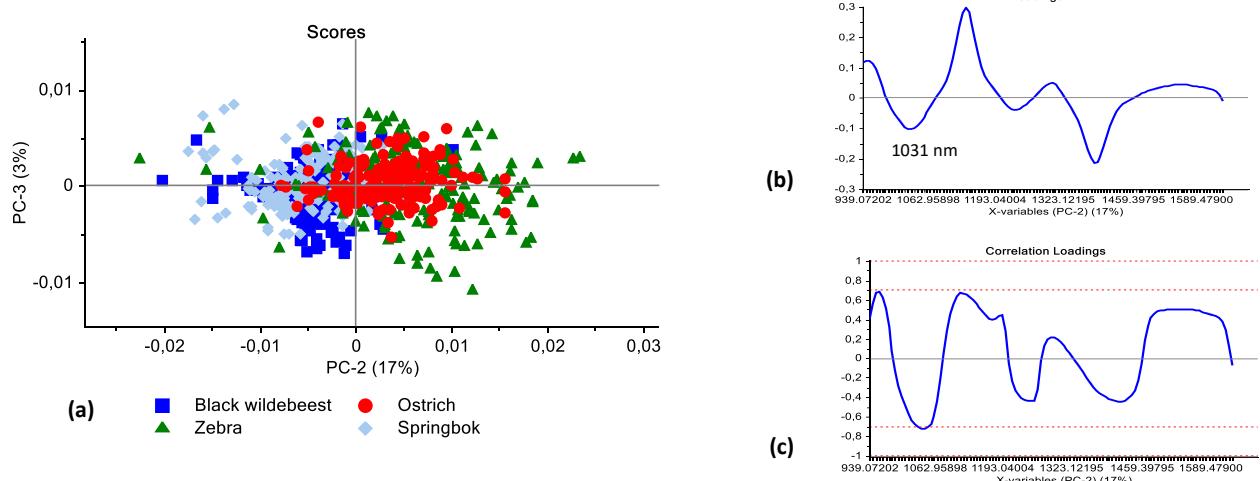


Figure A20 PCA analysis (SGd₁(7) pre-processed) of species for black wildebeest (blue), zebra (green), springbok (light blue) and ostrich (red) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (17%) vs. PC3 (3%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with an interpretable band at 1031 nm.

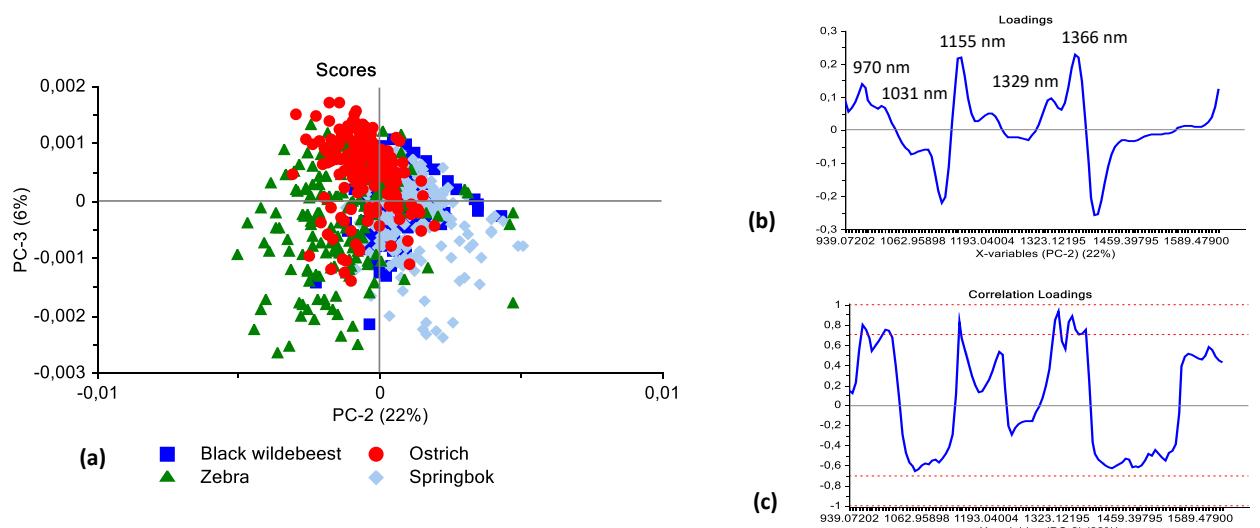


Figure A21 PCA analysis (SGd₂(7) pre-processed) of species for black wildebeest (blue), zebra (green), springbok (light blue) and ostrich (red) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (22%) vs. PC3 (6%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 970, 1013, 1155, 1329 and 1366 nm.

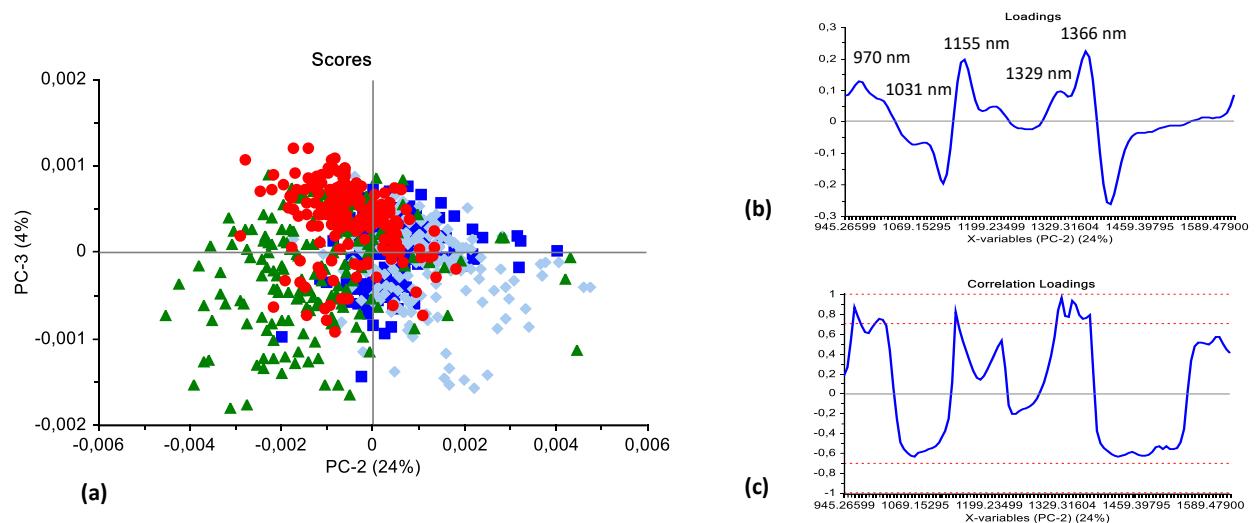


Figure A22 PCA analysis ($\text{SGd}_2(9)$ pre-processed) of species for black wildebeest (blue), zebra (green), springbok (light blue) and ostrich (red) classes. Minimal class separation was observed. Scores illustrated as (a) PCA score plot of PC2 (24%) vs. PC3 (4%). (b) PCA loadings line plot and (c) correlation loadings for PC2 with interpretable bands at 970, 1013, 1155, 1329 and 1366 nm.

Addendum B

Supplementary information pertaining to Chapter 4:

4.2 Fresh vs. previously frozen meat determination

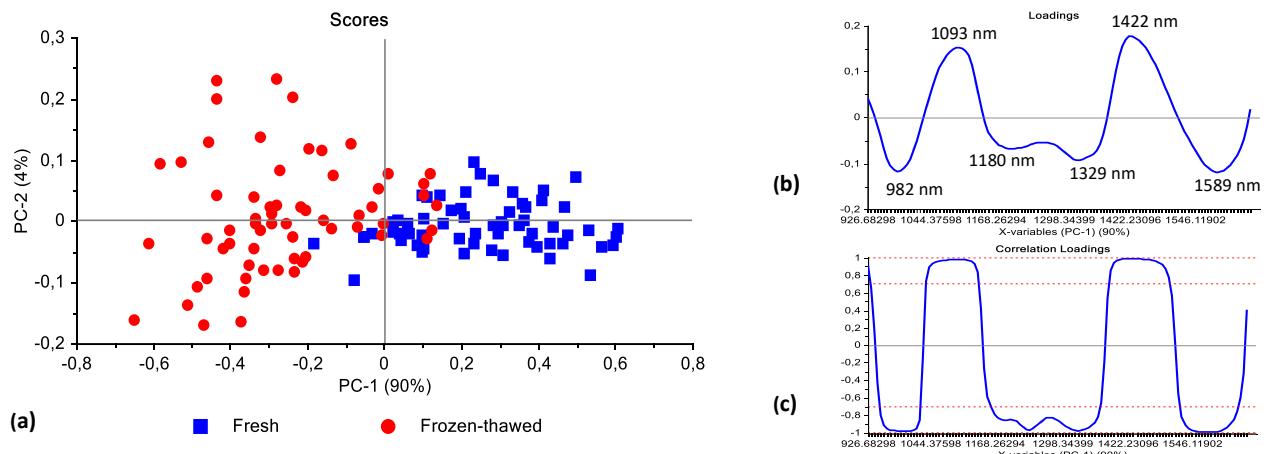


Figure B1 PCA analysis (SNV + detrend pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (90%) vs. PC2 (4%). (b) PCA loadings line plot and (c) correlation loadings for PC1 with interpretable bands at 982, 1093, 1180, 1329, 1422 and 1589 nm.

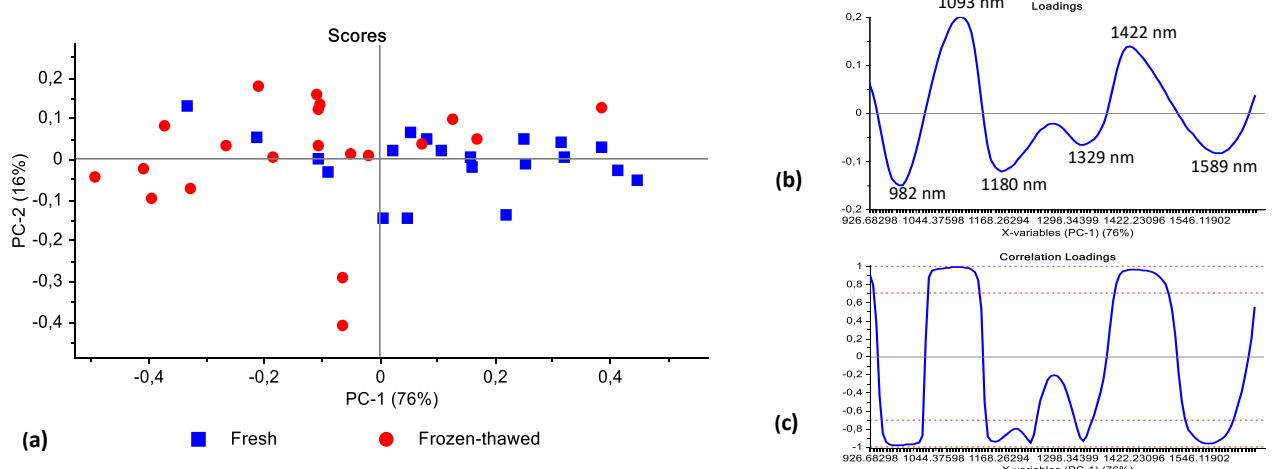


Figure B2 PCA analysis (SNV + detrend pre-processed) of ostrich (BD, FF) [frozen up to 1 month] illustrating minimal separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (76%) vs. PC2 (16%). (b) PCA loadings line plot and (c) correlation loadings for PC1 with interpretable bands at 982, 1093, 1180, 1329, 1422 and 1589 nm.

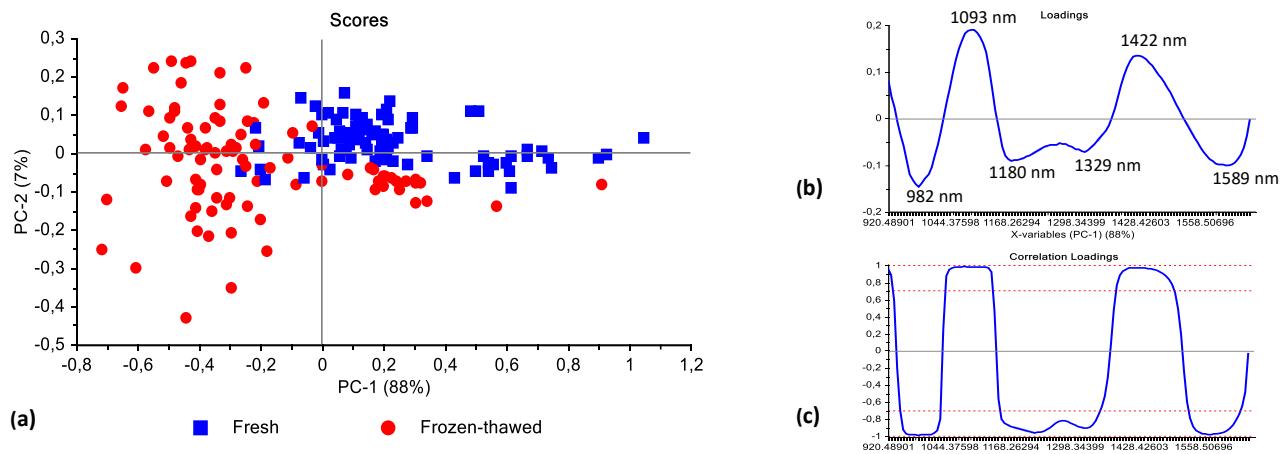


Figure B3 PCA analysis (SNV + detrend pre-processed) of black wildebeest (LTL) [frozen up to 9 months] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (88%) vs. PC2 (7%). (b) PCA loadings line plot and (c) correlation loadings for PC1 with interpretable bands at 982, 1093, 1180, 1329, 1422 and 1589 nm.

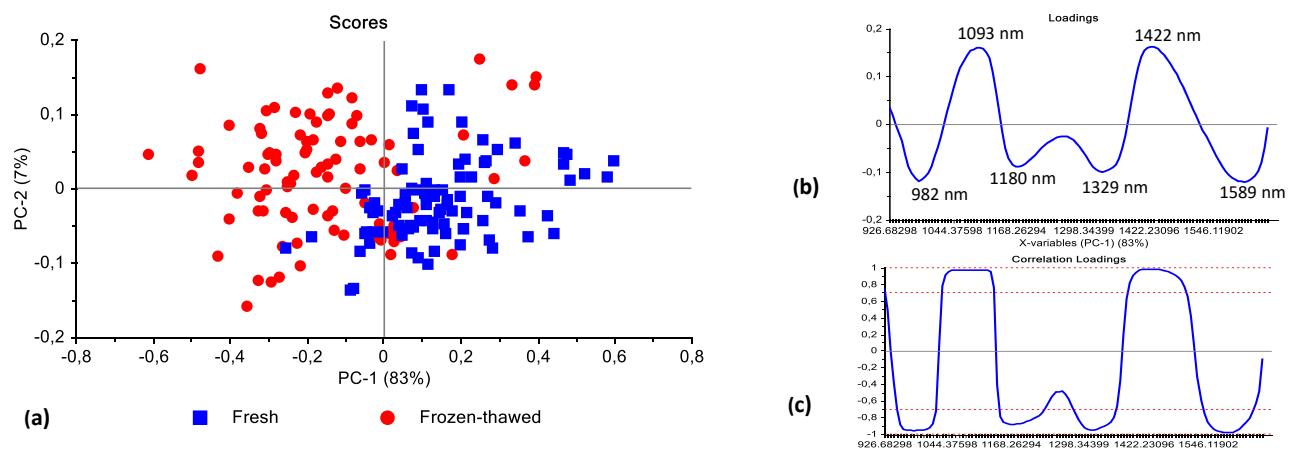


Figure B4 PCA analysis (SNV + detrend pre-processed) of springbok (LTL) [frozen up to 9 months] illustrating good separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (83%) vs. PC2 (7%). (b) PCA loadings line plot and (c) correlation loadings for PC1 with interpretable bands at 982, 1093, 1180, 1329, 1422 and 1589 nm.

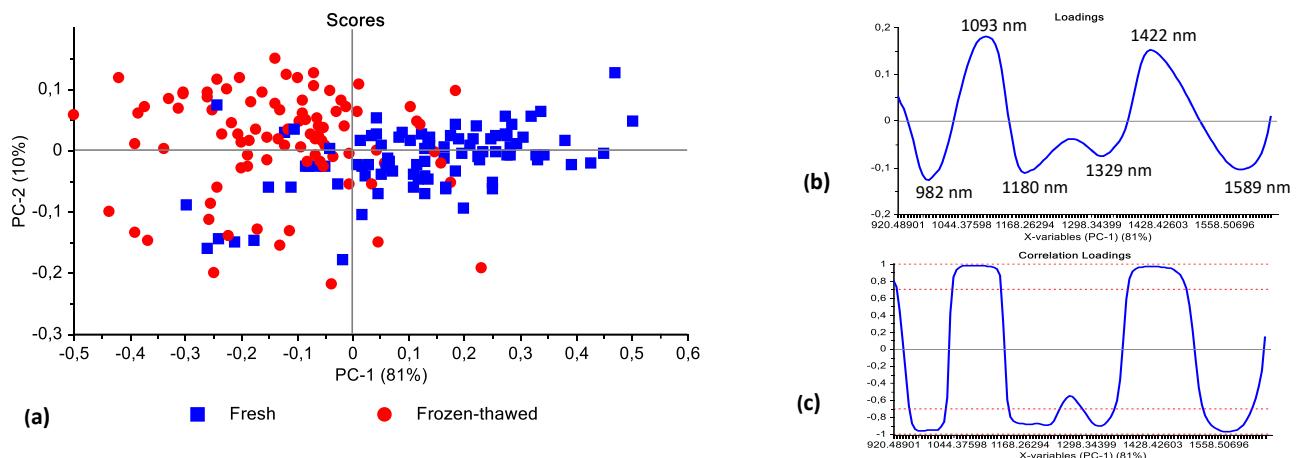


Figure B5 PCA analysis (SNV + detrend pre-processed) of ostrich (FF) [frozen up to 9 months] illustrating good separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (83%) vs. PC2 (7%). (b) PCA loadings line plot and (c) correlation loadings for PC1 with interpretable bands at 982, 1093, 1180, 1329, 1422 and 1589 nm.

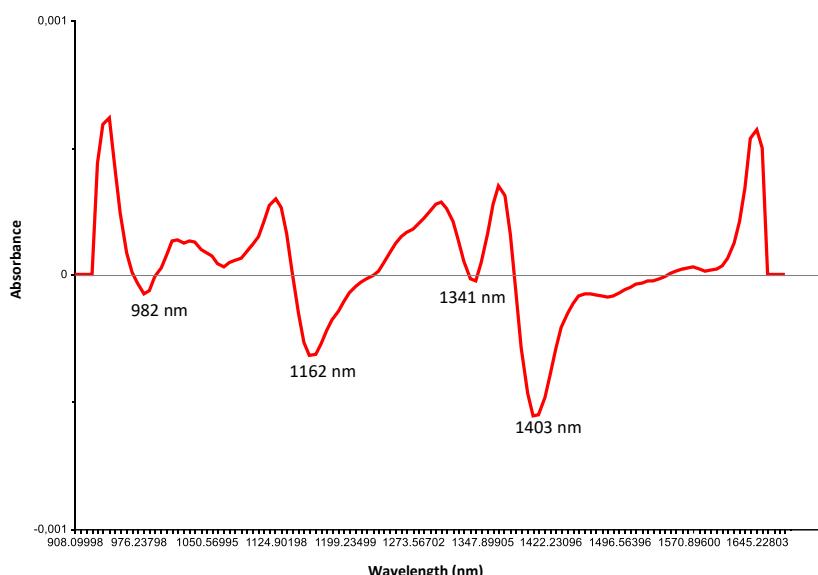


Figure B6 Difference spectra of black wildebeest (LTL) fresh and frozen-thawed samples, irrespective of the frozen period (1 – 9 months).

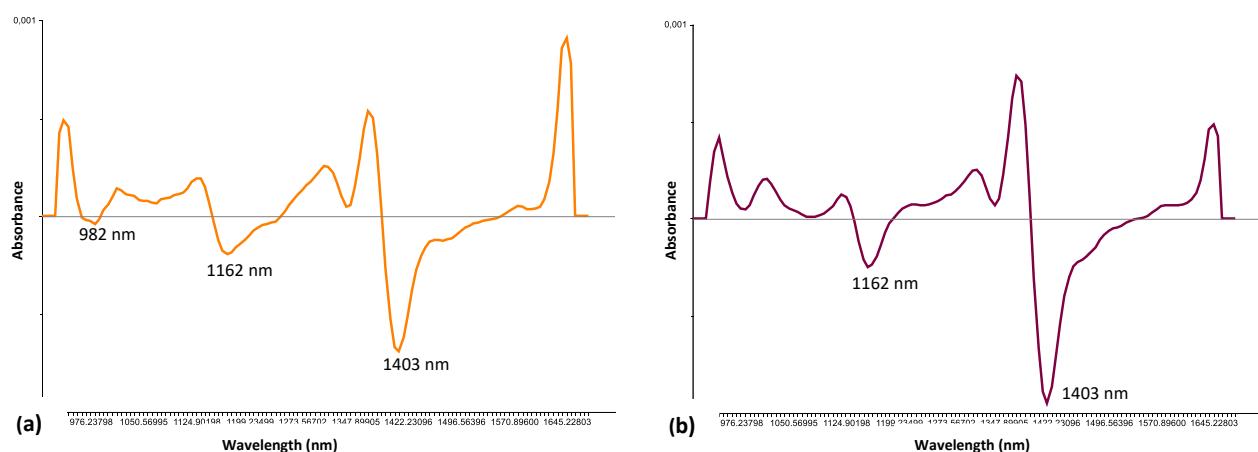


Figure B7 (a) Difference spectra of springbok (LTL) fresh and frozen-thawed samples, irrespective of the frozen period (1 – 9 months). (b) Difference spectra of springbok fresh and frozen-thawed samples, irrespective of the muscle type (LTL, BF, SM, ST, IS, SS, fillet).

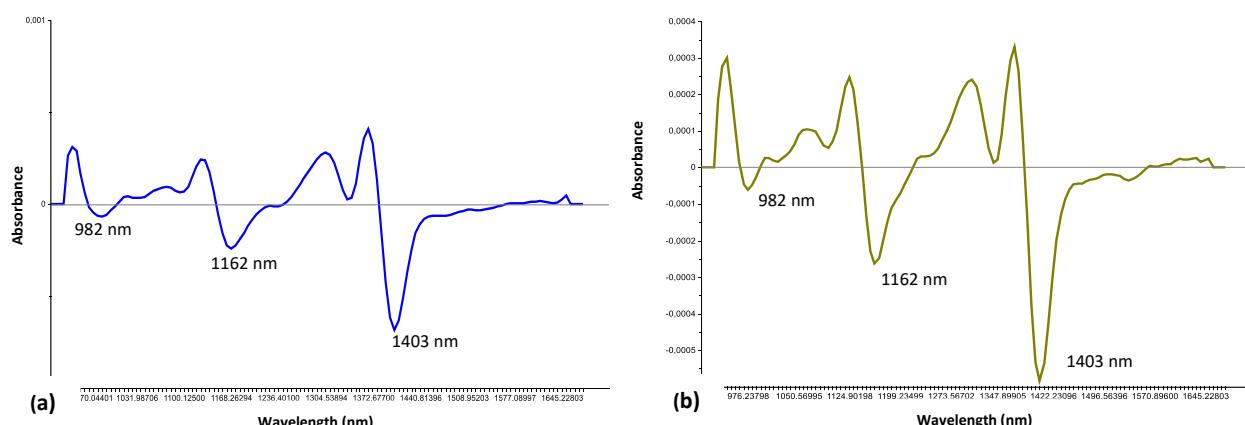
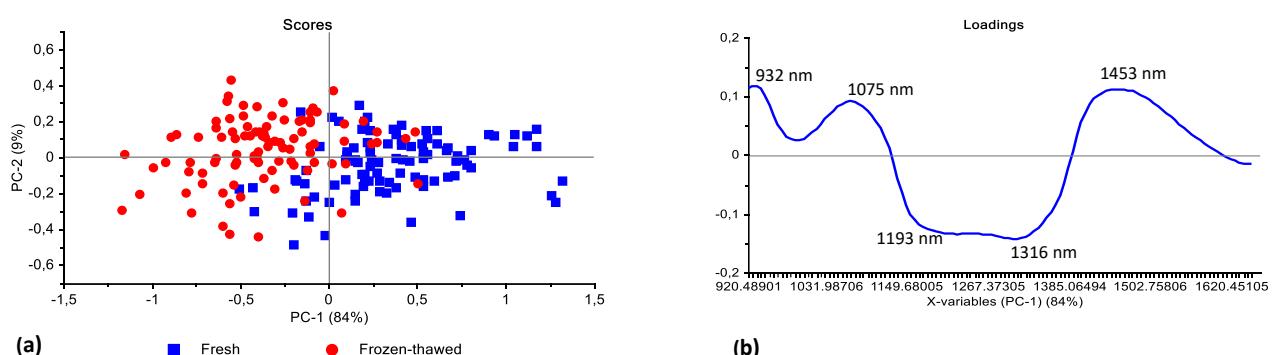


Figure B8 (a) Difference spectra of ostrich (FF) fresh and frozen-thawed samples, irrespective of the frozen period (1 – 9 months). (b) Difference spectra of ostrich fresh and frozen-thawed samples, irrespective of the muscle type (BD, FF).

Table B1 DA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for fresh or previously frozen meat discrimination.

Species	Number of principal components (PCs)	Method (distance calculations)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
Zebra	5	Linear	Calibration	92.4	7.6
			Validation	100.0	0
		Quadratic	Calibration	93.8	6.2
			Validation	97.9	2.1
		Mahalanobis	Calibration	93.1	6.9
			Validation	97.9	2.1
			Calibration	98.9	1.1
	4	Linear	Validation	100.0	0
			Calibration	96.8	3.2
		Quadratic	Validation	93.6	6.4
			Calibration	97.9	2.1
		Mahalanobis	Validation	100.0	0
Springbok	6	Linear	Calibration	86.7	13.3
			Validation	70.0	30.0
		Quadratic	Calibration	96.7	3.3
			Validation	40.0	60.0
		Mahalanobis	Calibration	100.0	0
			Validation	60.0	40.0
			Calibration	98.9	1.1
			Validation	100.0	0
		Calibration	96.8	3.2	
		Validation	93.6	6.4	
Ostrich	7	Linear	Calibration	97.9	2.1
			Validation	100.0	0
		Quadratic	Calibration	97.9	2.1
			Validation	100.0	0
		Mahalanobis	Calibration	98.9	1.1
			Validation	100.0	0
			Calibration	98.9	1.1
	8	Linear	Validation	70.0	30.0
			Calibration	96.7	3.3
		Quadratic	Validation	40.0	60.0
			Calibration	100.0	0
		Mahalanobis	Validation	60.0	40.0

(DA) Discriminant analysis; (SNV) Standard normal variate.

**Figure B9** PCA analysis (SNV pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating minimal overlap between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (84%) vs. PC2 (9%) and (b) PCA loadings line plot for PC1 with interpretable bands at 932, 1075, 1193, 1316 and 1453 nm.

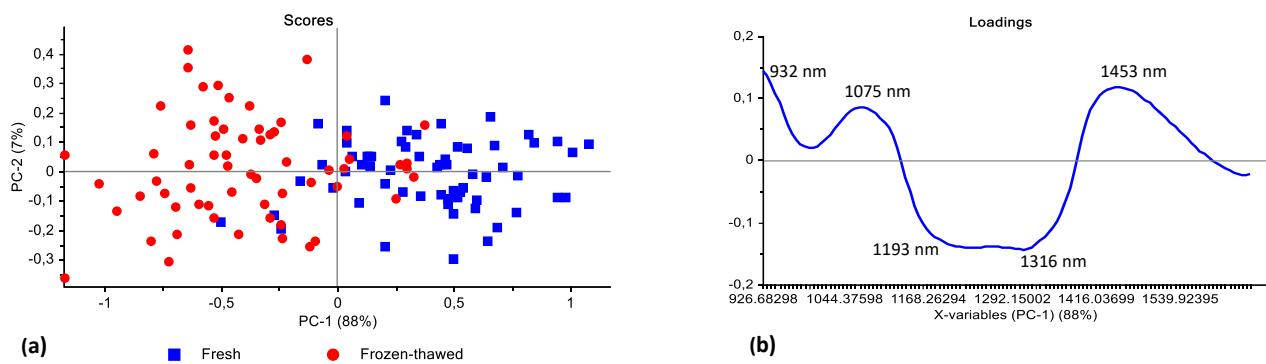


Figure B10 PCA analysis (SNV pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating minimal overlap between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (88%) vs. PC2 (7%) and (b) PCA loadings line plot for PC1 with interpretable bands at 932, 1075, 1193, 1316 and 1453 nm.

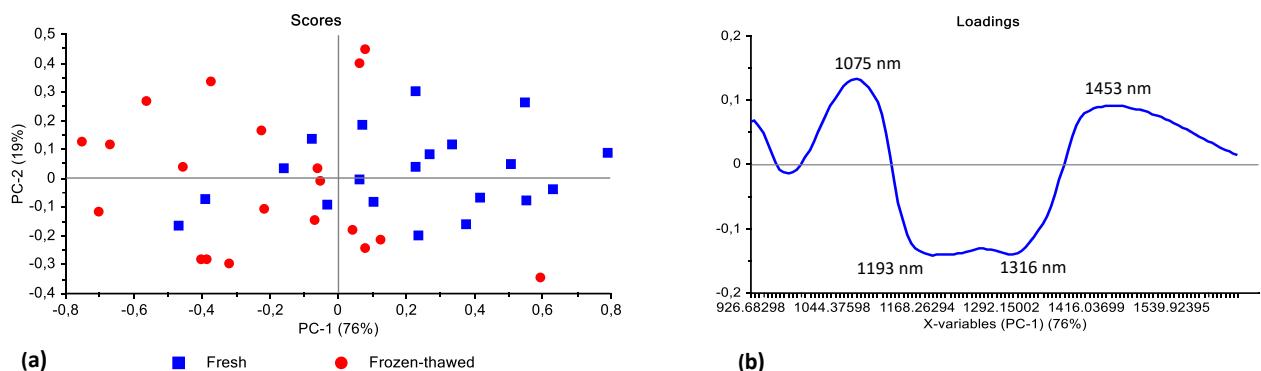


Figure B11 PCA analysis (SNV pre-processed) of ostrich (BD, FF) [frozen up to 1 month] illustrating minimal separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (76%) vs. PC2 (19%) and (b) PCA loadings line plot for PC1 with interpretable bands at 1075, 1193, 1316 and 1453 nm.

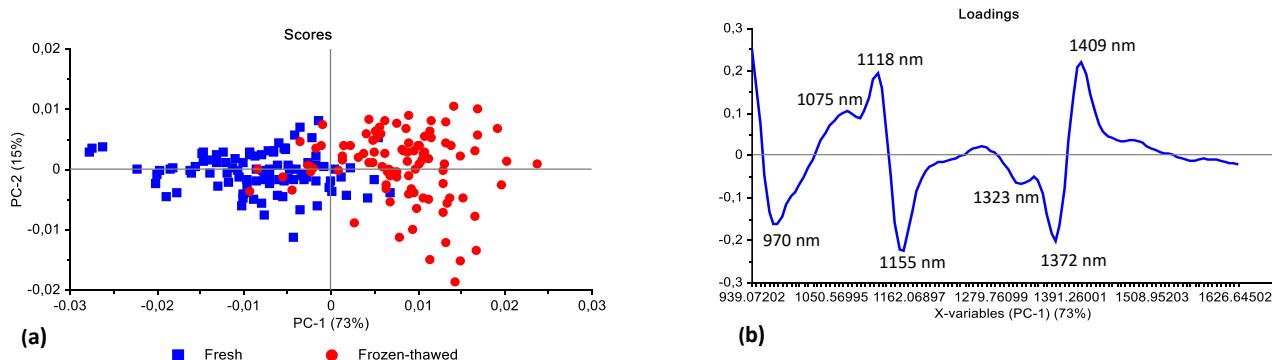


Figure B12 PCA analysis (SNV + SGd₂(7) pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (73%) vs. PC2 (15%) and (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1075, 1118, 1155, 1323, 1372 and 1409 nm.

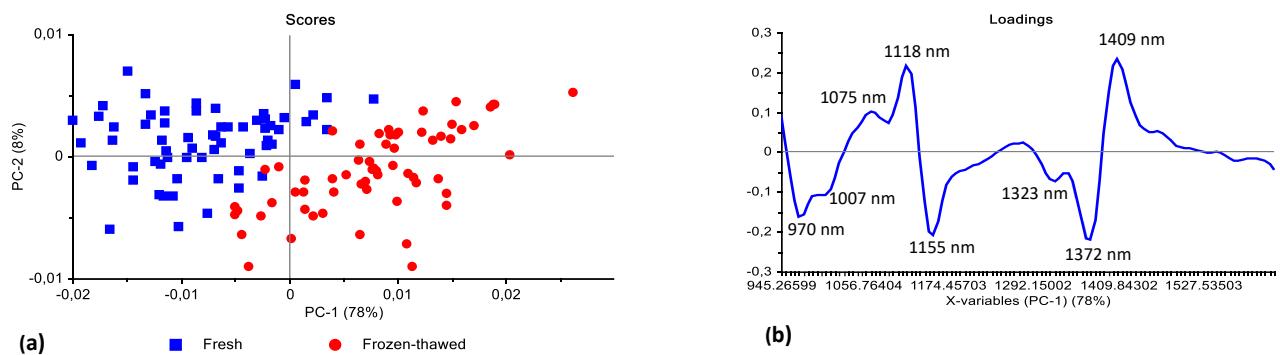


Figure B13 PCA analysis (SNV + SGd₂(7) pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating excellent separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (78%) vs. PC2 (8%) and (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1007, 1075, 1118, 1155, 1323, 1372 and 1409 nm.

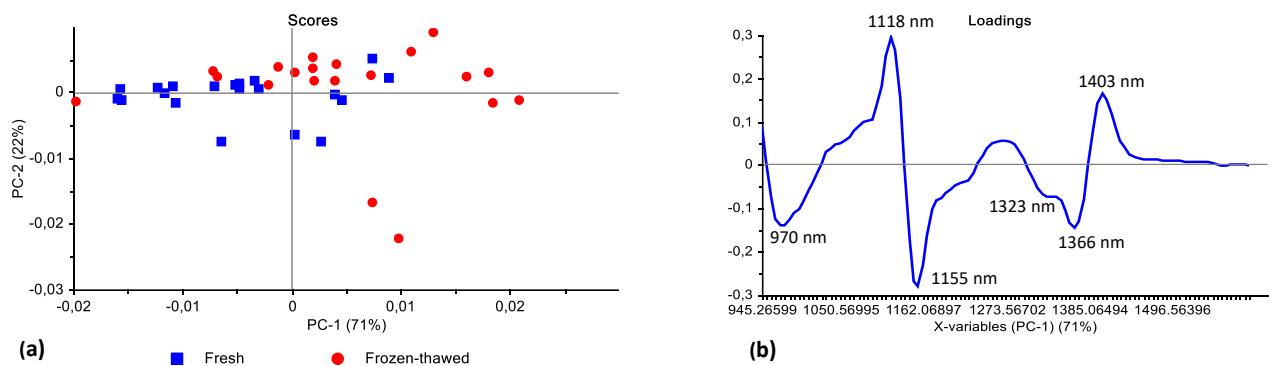


Figure B14 PCA analysis (SNV + SGd₂(7) pre-processed) of ostrich (BD, FF) [frozen up to 1 month] illustrating an overlap between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (71%) vs. PC2 (22%) and (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1118, 1155, 1323, 1366 and 1403 nm.

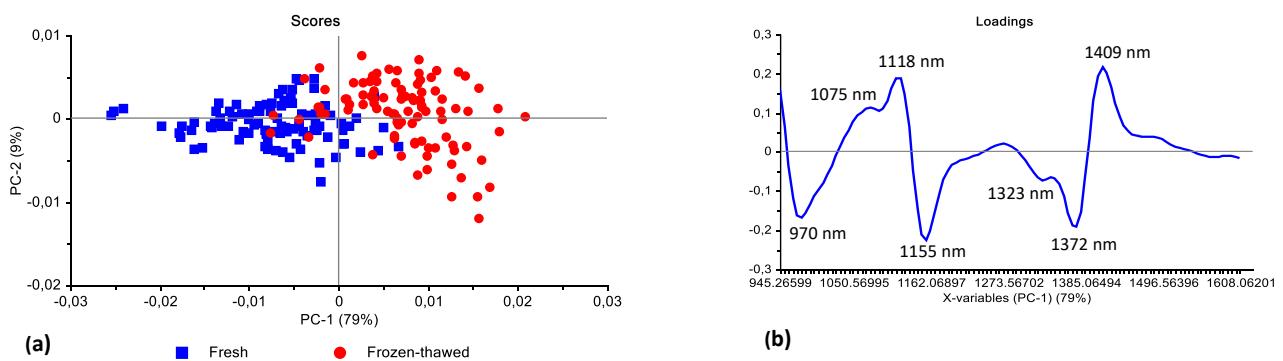


Figure B15 PCA analysis (SNV + SGd₂(9) pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (79%) vs. PC2 (9%) and (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1075, 1118, 1155, 1323, 1372 and 1409 nm.

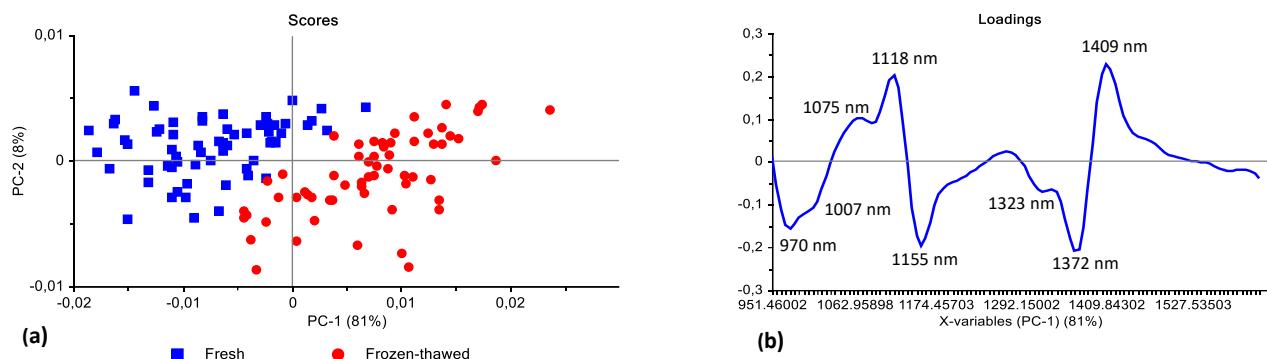


Figure B16 PCA analysis (SNV + SGd₂(9) pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating excellent separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (81%) vs. PC2 (8%) and (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1007, 1075, 1118, 1155, 1323, 1372 and 1409 nm.

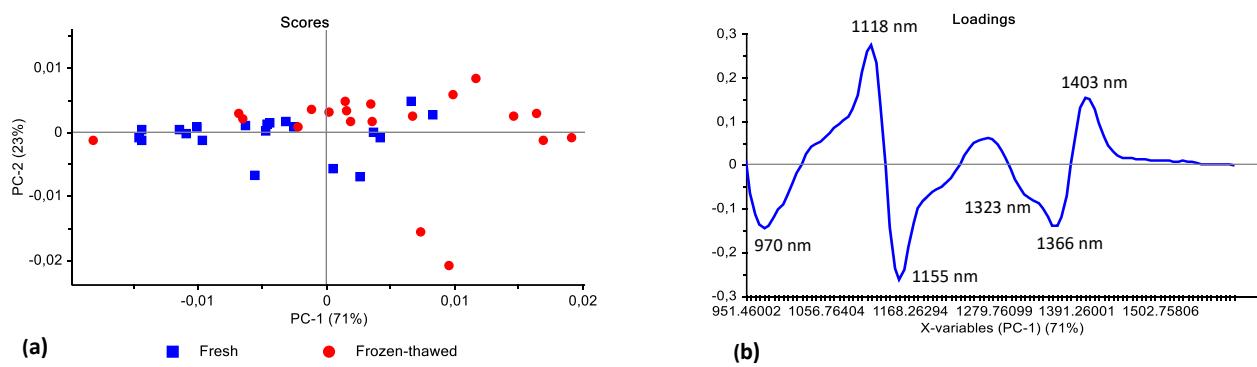


Figure B17 PCA analysis (SNV + SGd₂(9) pre-processed) of ostrich (BD, FF) [frozen up to 1 month] illustrating an overlap between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (71%) vs. PC2 (23%) and (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1118, 1155, 1323, 1366 and 1403 nm.

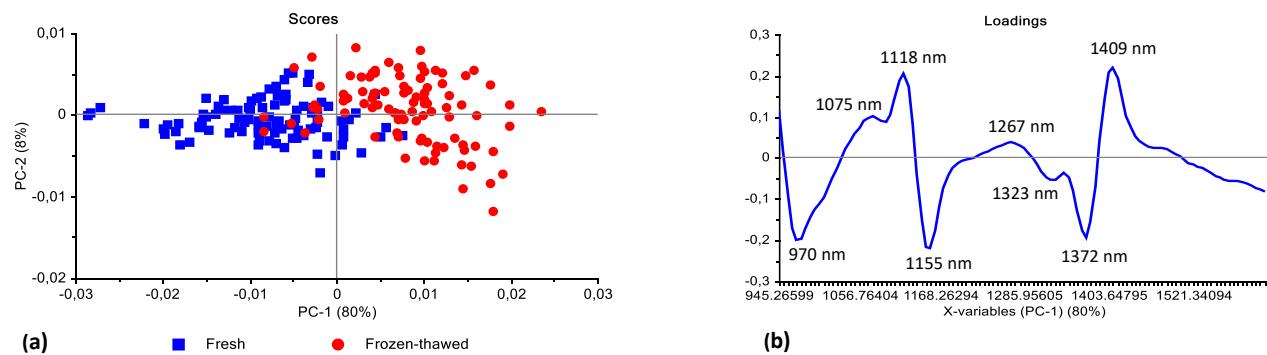


Figure B18 PCA analysis (SNV + detrend + SGd₂(7) pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (80%) vs. PC2 (8%) and (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1075, 1118, 1155, 1267, 1323, 1372 and 1409 nm.

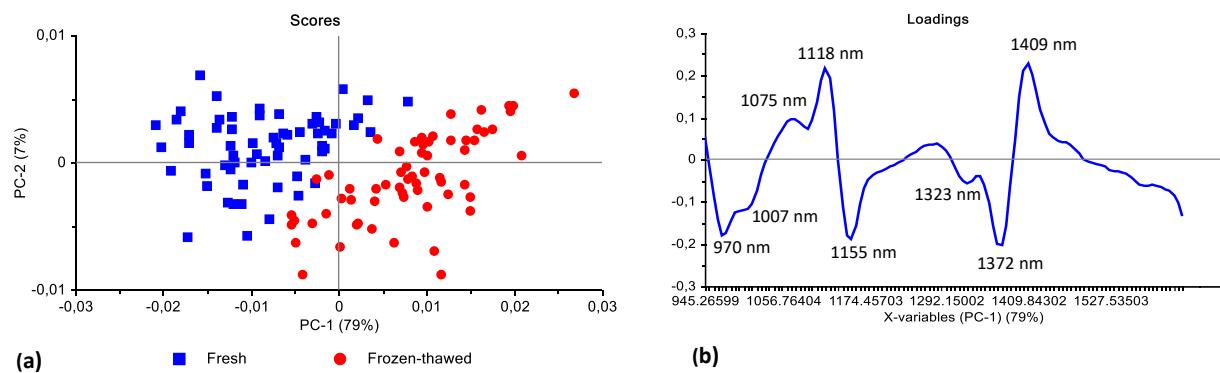


Figure B19 PCA analysis (SNV + detrend + SGd₂(7) pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating excellent separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (79%) vs. PC2 (7%) and (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1007, 1075, 1118, 1155, 1323, 1372 and 1409 nm.

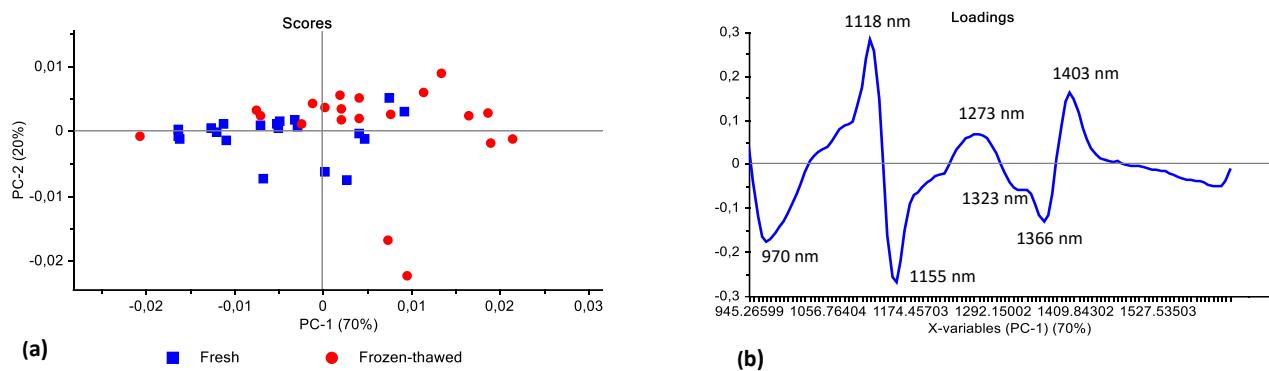


Figure B20 PCA analysis (SNV + detrend + SGd₂(7) pre-processed) of ostrich (BD, FF) [frozen up to 1 month] illustrating an overlap between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (70%) vs. PC2 (20%) and (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1118, 1155, 1273, 1323, 1366 and 1403 nm.

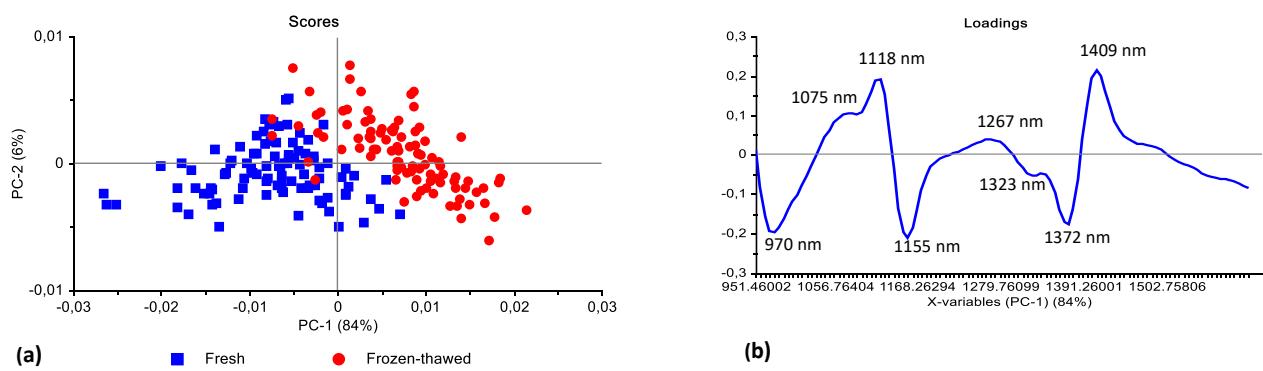


Figure B21 PCA analysis (SNV + detrend + SGd₂(9) pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (84%) vs. PC2 (6%) and (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1075, 1118, 1155, 1267, 1323, 1372 and 1409 nm.

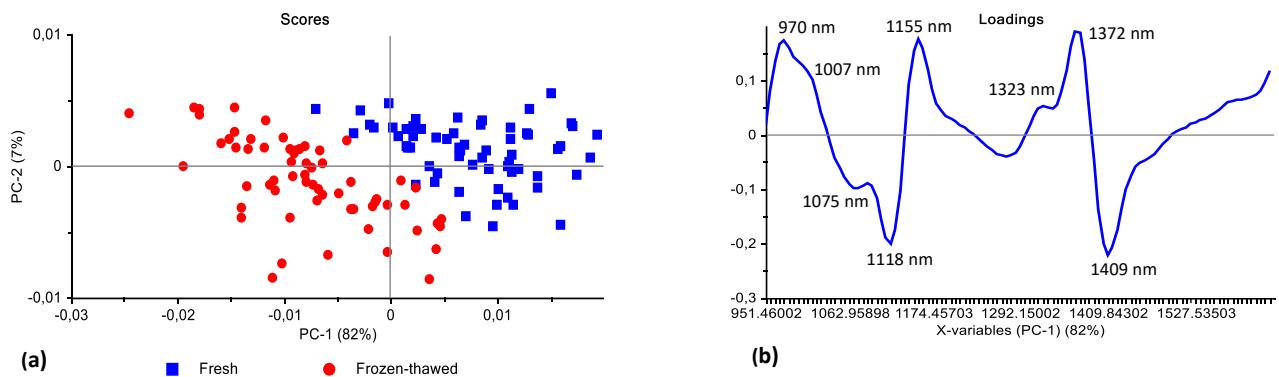


Figure B22 PCA analysis (SNV + detrend + SGd₂(9) pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating excellent separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (82%) vs. PC2 (7%) and (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1007, 1075, 1118, 1155, 1323, 1372 and 1409 nm.

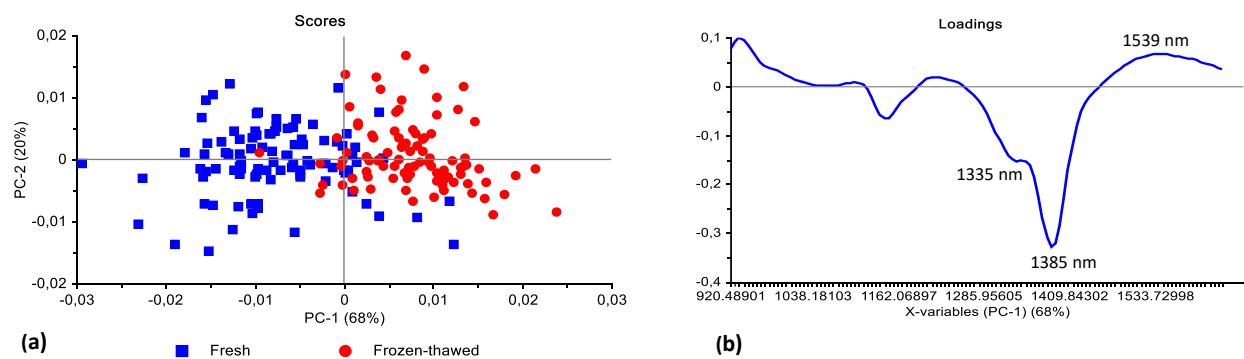


Figure B23 PCA analysis (SGd₁(5) pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (68%) vs. PC2 (20%) and (b) PCA loadings line plot for PC1 with interpretable bands at 1335, 1385 and 1539 nm.

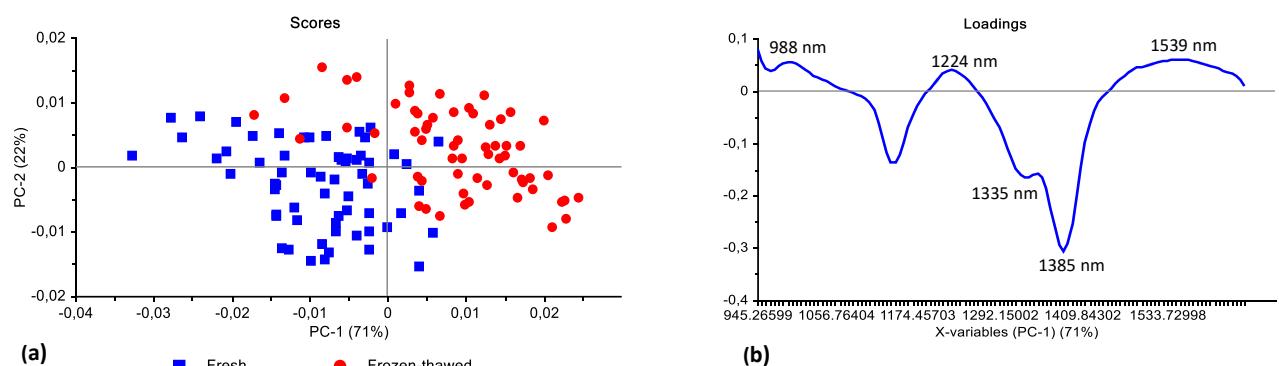


Figure B24 PCA analysis (SGd₁(5) pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (71%) vs. PC2 (22%) and (b) PCA loadings line plot for PC1 with interpretable bands at 988, 1224, 1335, 1385 and 1539 nm.

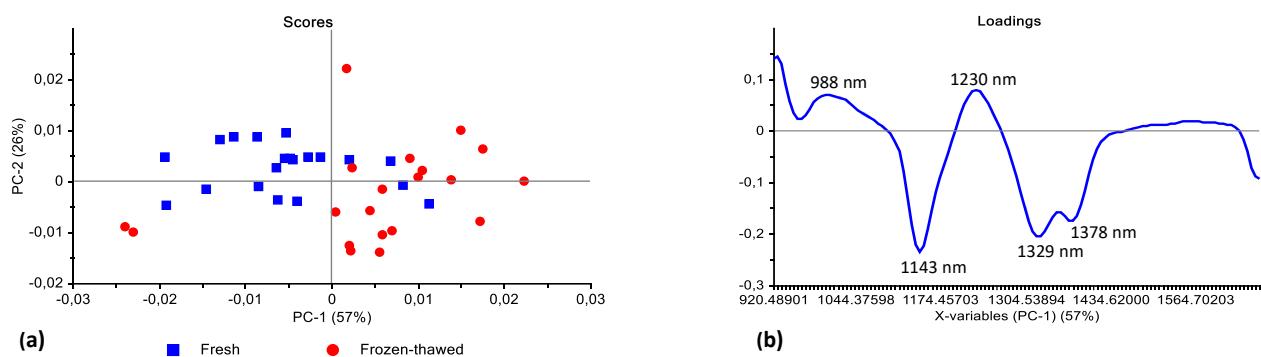


Figure B25 PCA analysis (SGd₁(5) pre-processed) of ostrich (BD, FF) [frozen up to 1 month] illustrating an overlap between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (57%) vs. PC2 (26%) and (b) PCA loadings line plot for PC1 with interpretable bands at 988, 1143, 1230, 1329 and 1378 nm.

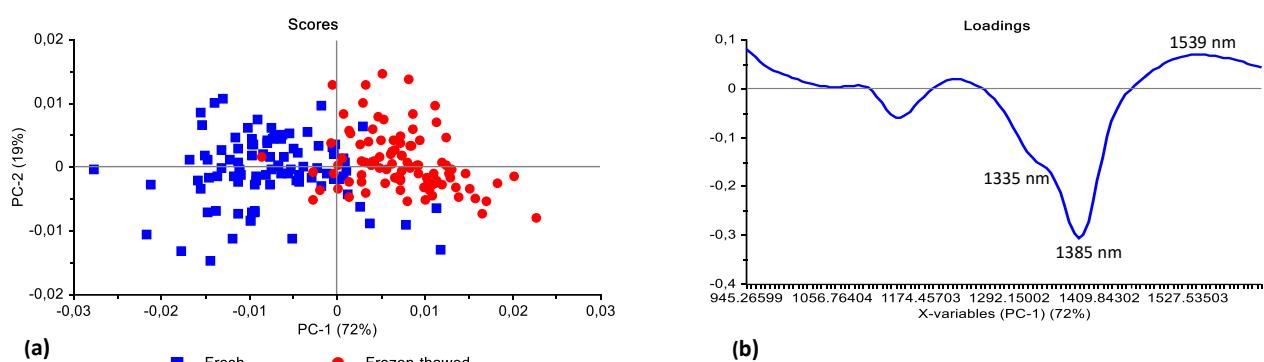


Figure B26 PCA analysis (SGd₁(7) pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (72%) vs. PC2 (19%) and (b) PCA loadings line plot for PC1 with interpretable bands at 1335, 1385 and 1539 nm.

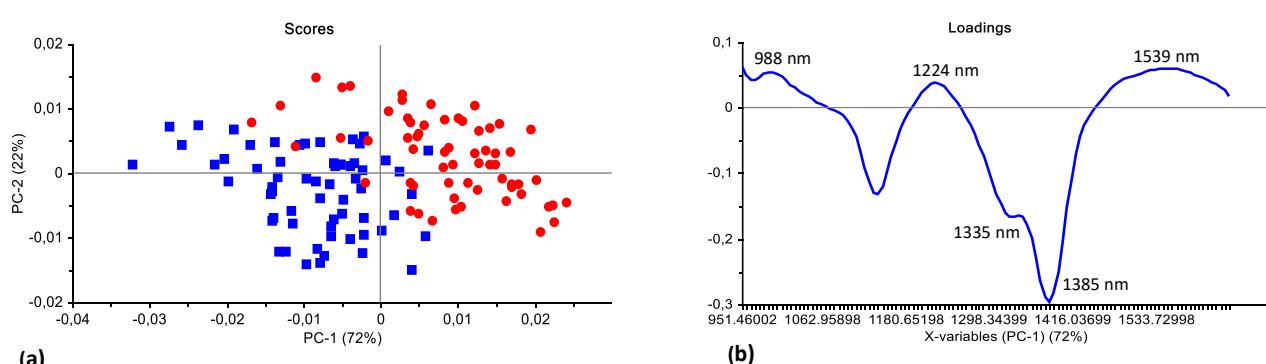


Figure B27 PCA analysis (SGd₁(7) pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (72%) vs. PC2 (22%) and (b) PCA loadings line plot for PC1 with interpretable bands at 988, 1224, 1335, 1385 and 1539 nm.

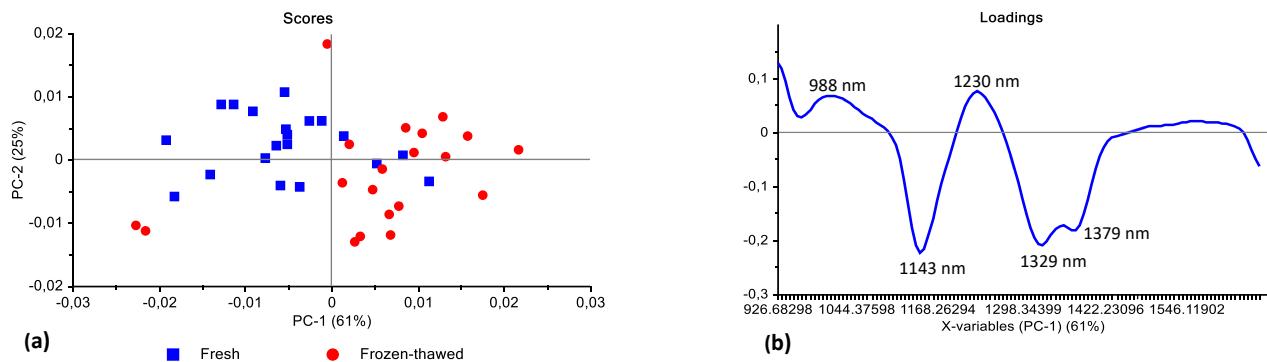


Figure B28 PCA analysis (SGd₁(7) pre-processed) of ostrich (BD, FF) [frozen up to 1 month] illustrating an overlap between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (61%) vs. PC2 (25%) and (b) PCA loadings line plot for PC1 with interpretable bands at 988, 1143, 1230, 1329 and 1378 nm.

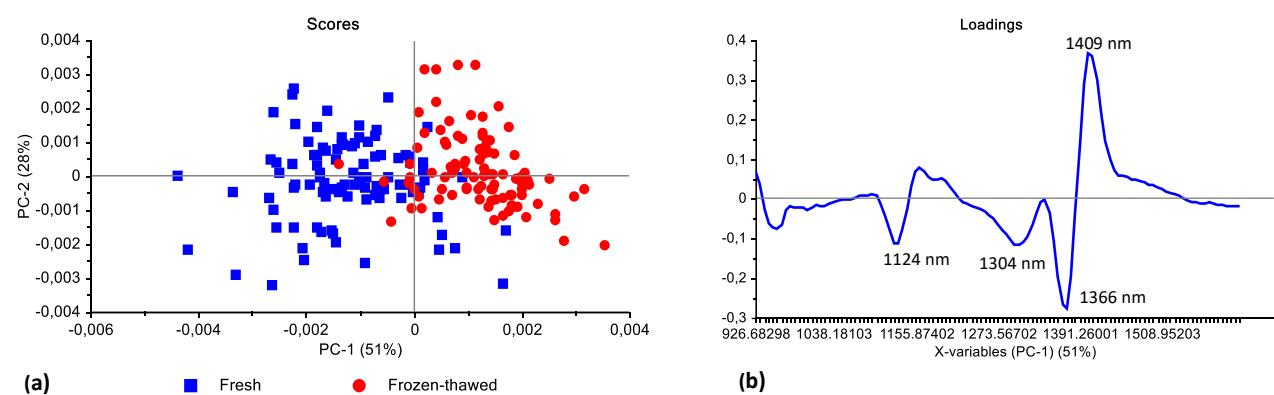


Figure B29 PCA analysis (SGd₂(7) pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (51%) vs. PC2 (28%) and (b) PCA loadings line plot for PC1 with interpretable bands at 1124, 1304, 1366 and 1409 nm.

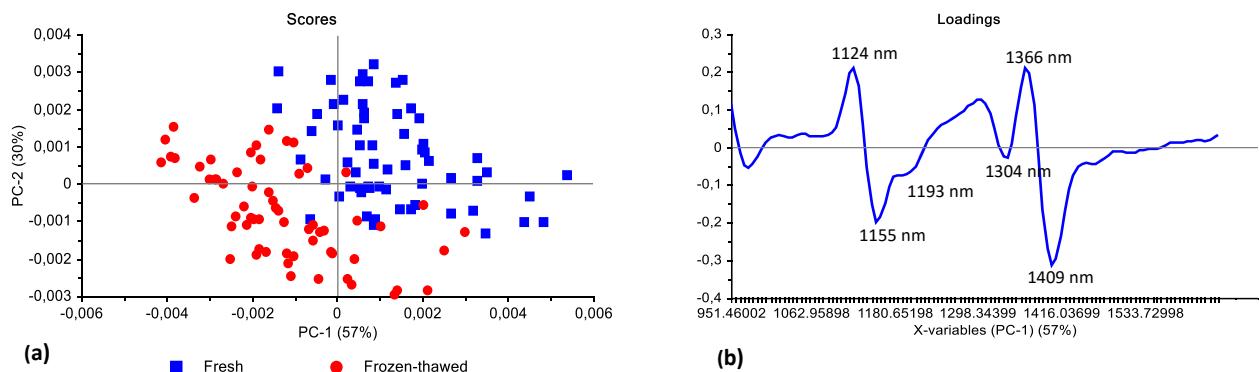


Figure B30 PCA analysis (SGd₂(7) pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (57%) vs. PC2 (30%) and (b) PCA loadings line plot for PC1 with interpretable bands at 1124, 1155, 1193, 1304, 1366 and 1409 nm.

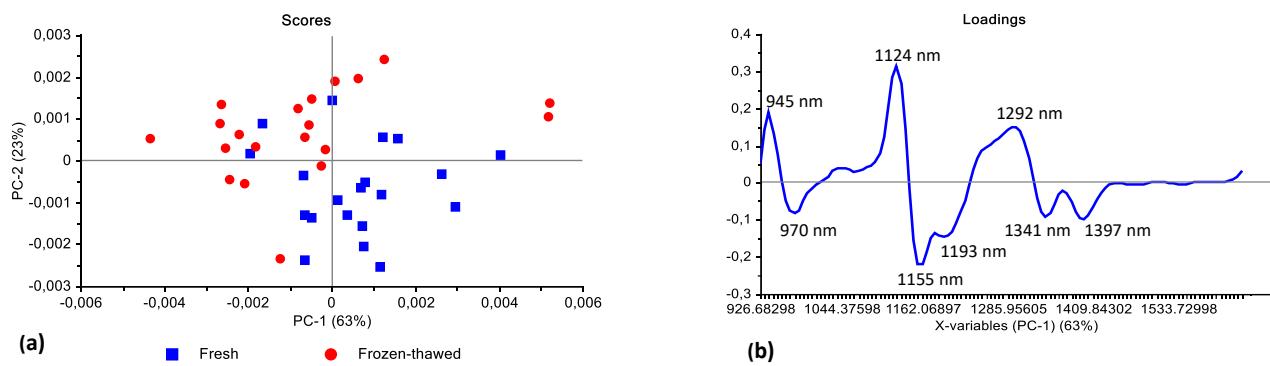


Figure B31 PCA analysis (SGd₂(7) pre-processed) of ostrich (BD, FF) [frozen up to 1 month] illustrating an overlap between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (63%) vs. PC2 (23%) and (b) PCA loadings line plot for PC1 with interpretable bands at 945, 970, 1124, 1155, 1193, 1292, 1341 and 1397 nm.

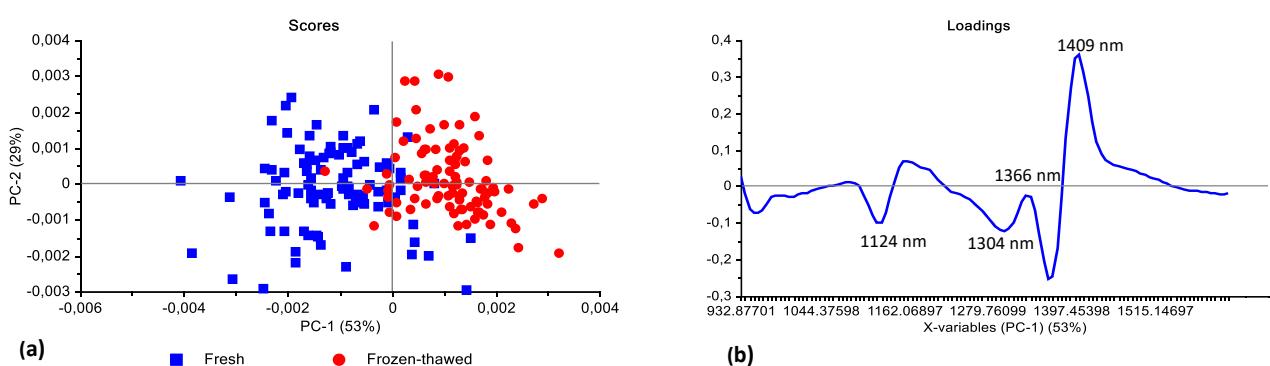


Figure B32 PCA analysis (SGd₂(9) pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (53%) vs. PC2 (29%) and (b) PCA loadings line plot for PC1 with interpretable bands at 1124, 1304, 1366 and 1409 nm.

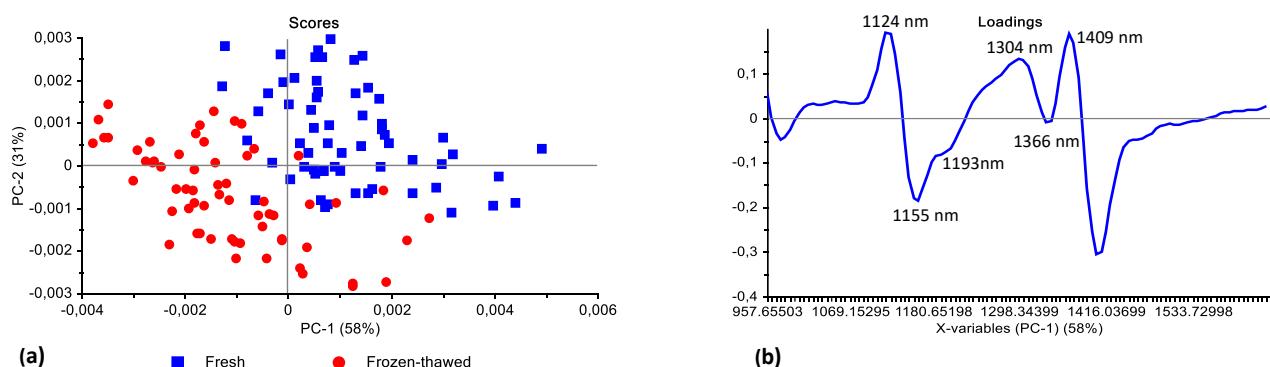


Figure B33 PCA analysis (SGd₂(9) pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [frozen up to 1 month] illustrating good separation between fresh (blue) and frozen-thawed (red) classes. Scores illustrated as (a) PCA score plot of PC1 (58%) vs. PC2 (31%) and (b) PCA loadings line plot for PC1 with interpretable bands at 1124, 1155, 1193, 1304, 1366 and 1409 nm.

Table B2 KNN model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen period classification.

Species	Number of neighbours (<i>k</i>)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
Black wildebeest	2	Calibration	46.8	53.2
		Validation	83.3	16.7
	5	Calibration	46.0	54.0
		Validation	77.8	22.2
Zebra	2	Calibration	42.1	57.9
		Validation	68.5	31.5
	5	Calibration	44.4	55.6
		Validation	70.4	29.6
Springbok	3	Calibration	54.2	45.8
		Validation	78.3	21.7
	5	Calibration	51.7	48.3
		Validation	76.7	23.3
Ostrich	3	Calibration	50.0	50.0
		Validation	75.0	25.0
	5	Calibration	49.2	50.8
		Validation	78.3	21.7

(KNN) *K*-nearest neighbours; (SNV) Standard normal variate.**Table B3** KNN model calibration, cross-validation and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen-thawed meat classification, irrespective of frozen period.

Species	Number of neighbours (<i>k</i>)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
Black wildebeest	2	Calibration	92.9	7.1
		Cross-validation	94.4	5.6
		Validation	96.3	3.7
	5	Calibration	93.7	6.4
		Cross-validation	93.7	6.4
		Validation	96.3	3.7
Zebra	2	Calibration	76.2	23.8
		Cross-validation	75.4	24.6
		Validation	83.3	16.7
	5	Calibration	83.3	16.7
		Cross-validation	83.3	16.7
		Validation	81.5	18.5
Springbok	2	Calibration	93.3	6.7
		Cross-validation	93.3	6.7
		Validation	100.0	0
	5	Calibration	91.7	8.3
		Cross-validation	92.5	7.5
		Validation	100.0	0
Ostrich	2	Calibration	75.8	24.2
		Cross-validation	85.8	14.2
		Validation	91.7	8.3
	5	Calibration	78.3	21.7
		Cross-validation	77.5	22.5
		Validation	93.3	6.7

(KNN) *K*-nearest neighbours; (SNV) Standard normal variate.

Table B4 DA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen period classification.

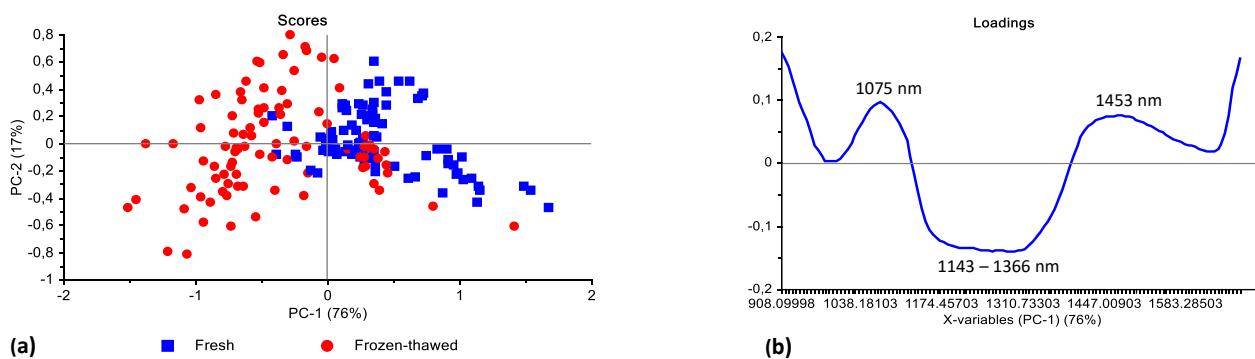
Species	Number of principal components (PCs)	Method (distance calculations)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
Black wildebeest	5	Linear	Calibration	69.1	30.9
			Validation	72.2	27.8
		Quadratic	Calibration	88.9	11.1
			Validation	72.2	27.8
			Calibration	84.9	15.1
			Validation	72.2	27.8
	5	Mahalanobis	Calibration	61.9	38.1
			Validation	59.3	40.7
		Quadratic	Calibration	79.4	20.6
			Validation	48.2	51.8
			Calibration	81.8	18.2
			Validation	64.8	35.2
Zebra	5	Linear	Calibration	74.2	25.8
			Validation	75.0	25.0
		Quadratic	Calibration	90.0	10.0
			Validation	80.0	20.0
			Calibration	88.3	11.7
			Validation	80.0	20.0
	5	Mahalanobis	Calibration	75.0	25.0
			Validation	80.0	20.0
		Quadratic	Calibration	84.2	15.8
			Validation	71.7	28.3
			Calibration	89.2	10.8
			Validation	68.3	31.7
Springbok	5	Linear	Calibration	75.0	25.0
			Validation	80.0	20.0
		Quadratic	Calibration	90.0	10.0
			Validation	80.0	20.0
			Calibration	88.3	11.7
			Validation	80.0	20.0
	5	Mahalanobis	Calibration	75.0	25.0
			Validation	80.0	20.0
		Quadratic	Calibration	84.2	15.8
			Validation	71.7	28.3
			Calibration	89.2	10.8
			Validation	68.3	31.7
Ostrich	5	Linear	Calibration	75.0	25.0
			Validation	80.0	20.0
		Quadratic	Calibration	84.2	15.8
			Validation	71.7	28.3
			Calibration	89.2	10.8
			Validation	68.3	31.7
	5	Mahalanobis	Calibration	75.0	25.0
			Validation	80.0	20.0
		Quadratic	Calibration	84.2	15.8
			Validation	71.7	28.3
			Calibration	89.2	10.8
			Validation	68.3	31.7

(DA) Discriminant analysis; (SNV) Standard normal variate.

Table B5 DA model calibration and validation results to assess the overall performance of the SNV + detrend corrected data for fresh vs. frozen-thawed meat classification, irrespective of frozen period.

Species	Number of principal components (PCs)	Method (distance calculations)	Model (SNV + detrend)	Classification accuracy (%)	Misclassification rate (%)
Black wildebeest	5	Linear	Calibration	91.8	8.2
			Validation	87.1	12.9
		Quadratic	Calibration	95.2	4.8
			Validation	93.0	7.0
		Mahalanobis	Calibration	87.5	12.5
			Validation	85.3	14.7
	5	Linear	Calibration	88.7	11.3
			Validation	91.0	9.0
		Quadratic	Calibration	87.0	13.0
			Validation	93.0	7.0
		Mahalanobis	Calibration	90.4	9.6
			Validation	88.6	11.4
Zebra	5	Linear	Calibration	90.8	9.2
			Validation	93.3	6.7
		Quadratic	Calibration	91.7	8.3
			Validation	96.7	3.3
		Mahalanobis	Calibration	95.0	5.0
			Validation	95.0	5.0
	5	Linear	Calibration	80.8	19.2
			Validation	91.7	8.3
		Quadratic	Calibration	89.2	10.8
			Validation	91.7	8.3
		Mahalanobis	Calibration	85.8	14.2
			Validation	81.7	18.3

(DA) Discriminant analysis; (SNV) Standard normal variate.



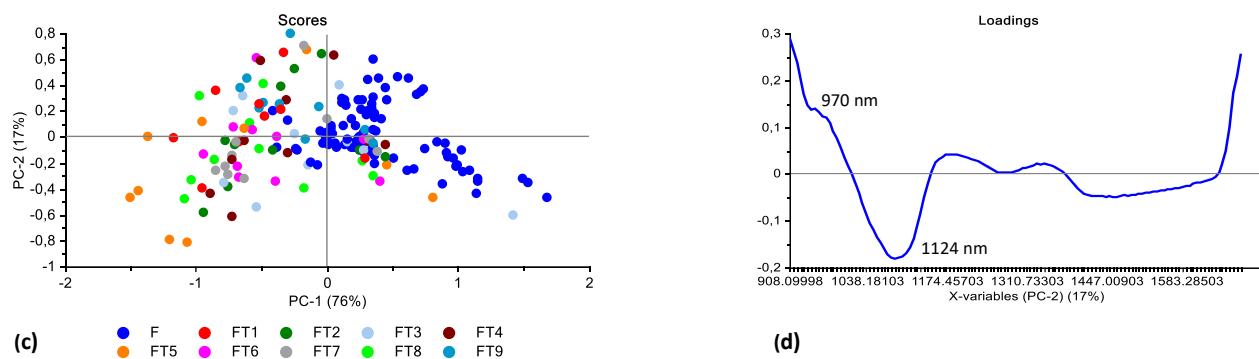


Figure B34 PCA analysis (SNV pre-processed) of black wildebeest (LTL) [frozen up to 9 months] illustrating a good separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (d) PCA score plot of PC1 (76%) vs. PC2 (17%), (b) PCA loadings line plot for PC1 with interpretable bands at 1075, 1143 – 1366 and 1453 nm and (d) PCA loadings line plot for PC2 with interpretable bands at 970 and 1124 nm.

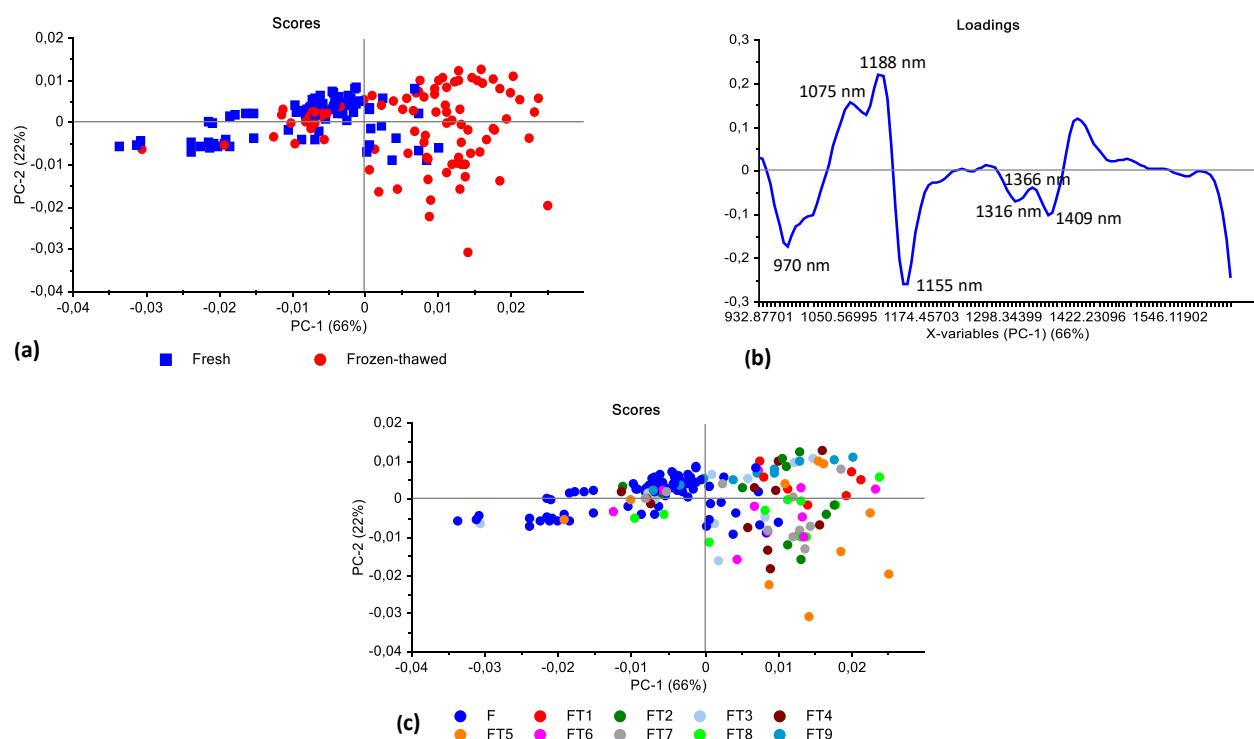


Figure B35 PCA analysis (SNV + SGd₂(9) pre-processed) of black wildebeest (LTL) [frozen up to 9 months] illustrating a good separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (66%) vs. PC2 (22%), (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1075, 1118, 1155, 1316, 1366 and 1409 nm.

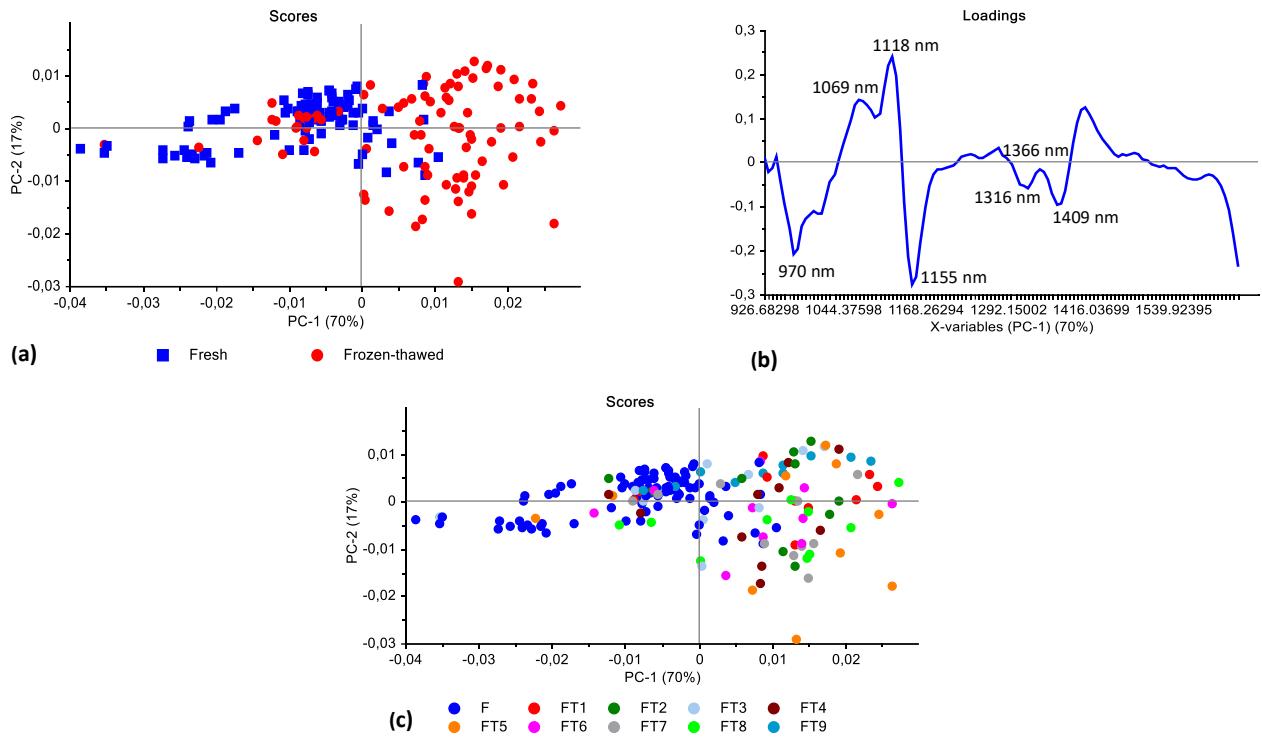


Figure B36 PCA analysis (SNV + detrend + SGd₂(7) pre-processed) of black wildebeest (LTL) [frozen up to 9 months] illustrating a good separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (70%) vs. PC2 (17%), (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1069, 1118, 1155, 1316, 1366 and 1409 nm.

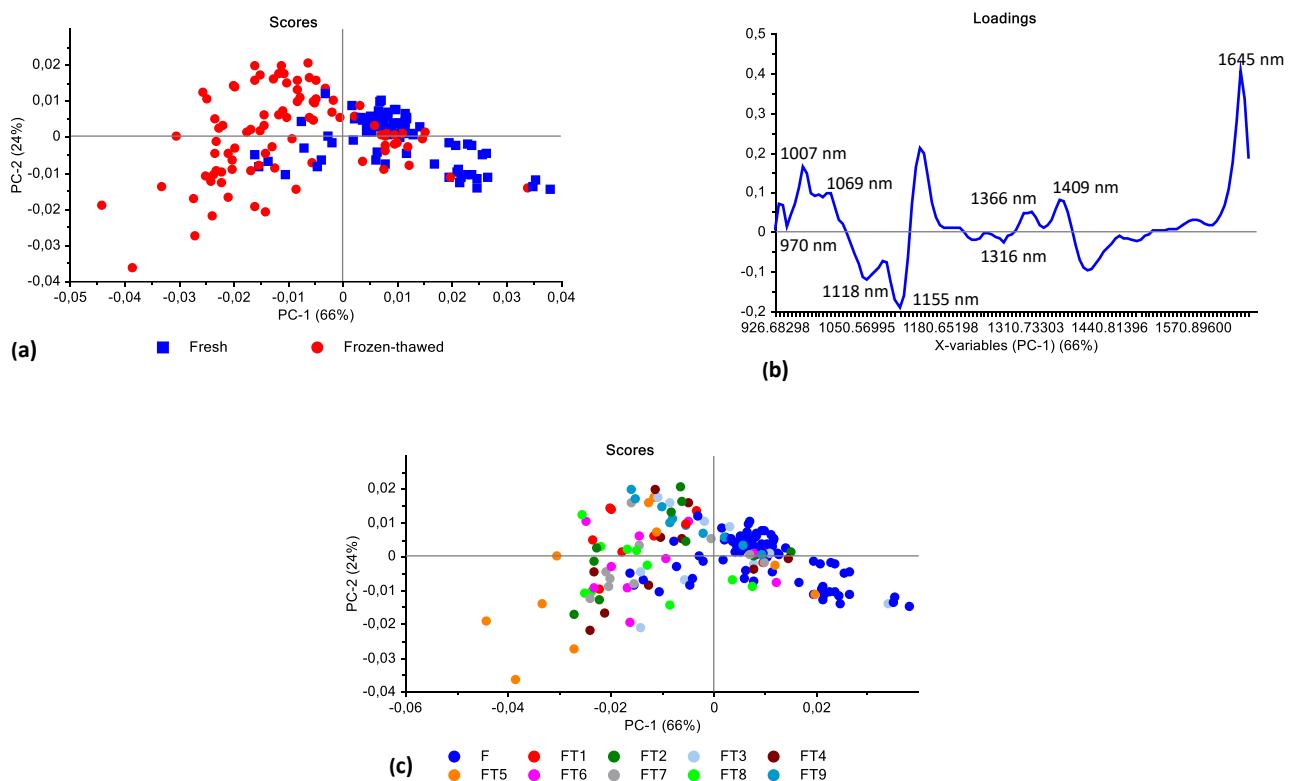


Figure B37 PCA analysis (SNV + detrend + SGd₂(9) pre-processed) of black wildebeest (LTL) [frozen up to 9 months] illustrating a good separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (70%) vs. PC2 (17%), (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1007, 1069, 1118, 1155, 1316, 1366, 1409 and 1645 nm.

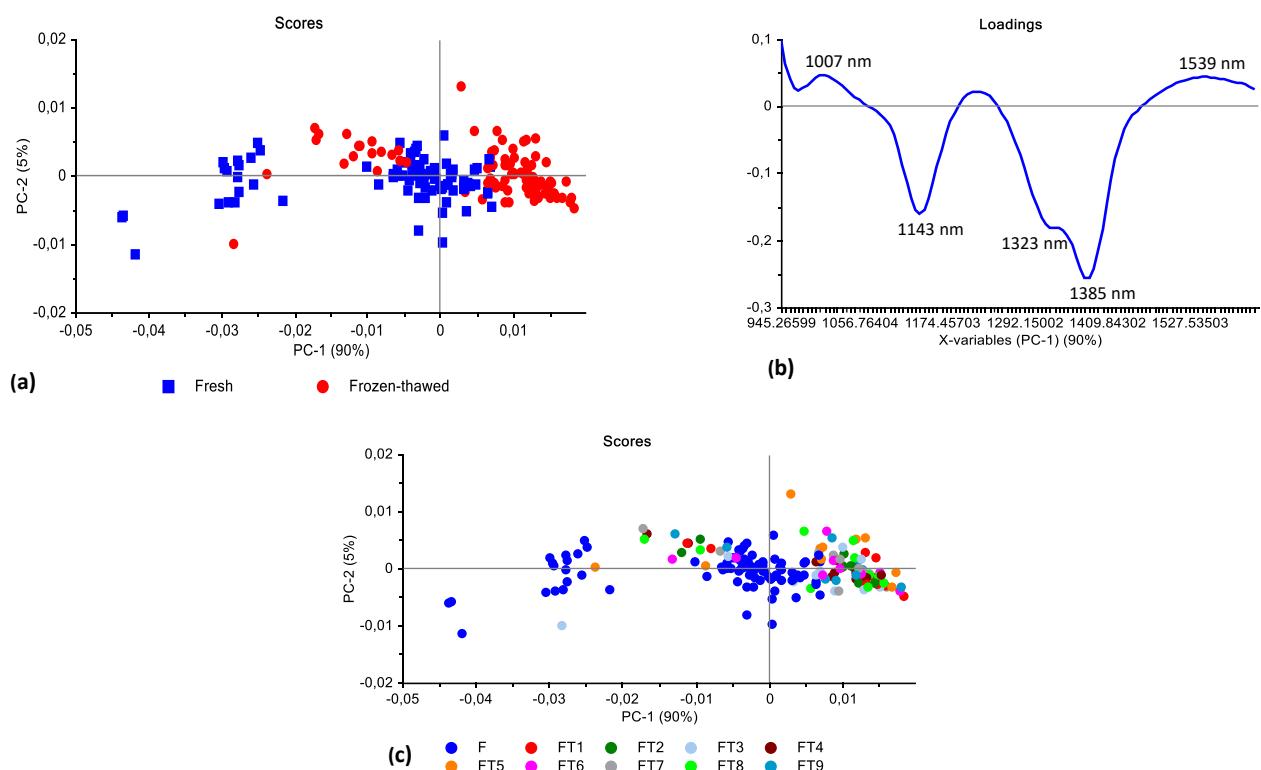
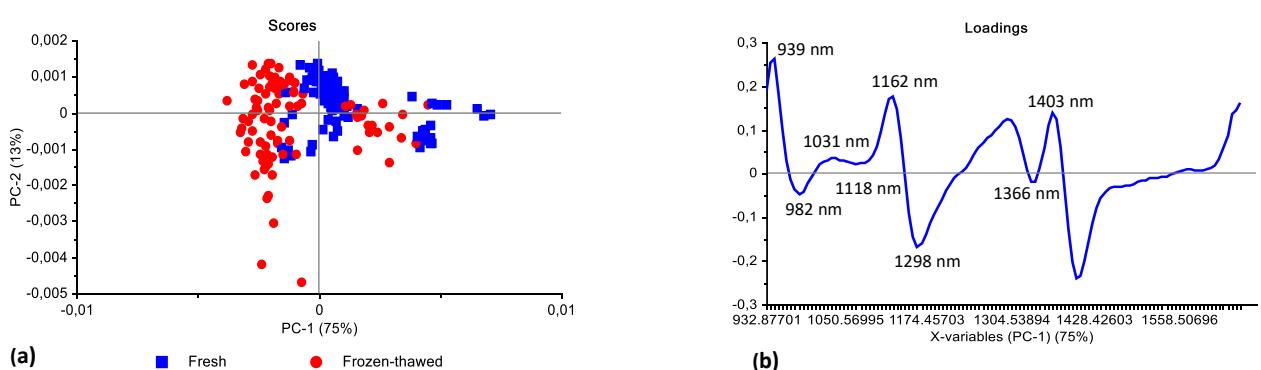


Figure B38 PCA analysis (SGd₁(7) pre-processed) of black wildebeest (LTL) [frozen up to 9 months] illustrating an unsatisfactory separation with an overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (90%) vs. PC2 (5%), (b) PCA loadings line plot for PC1 with interpretable bands at 1007, 1143, 1323, 1385 and 1539 nm.



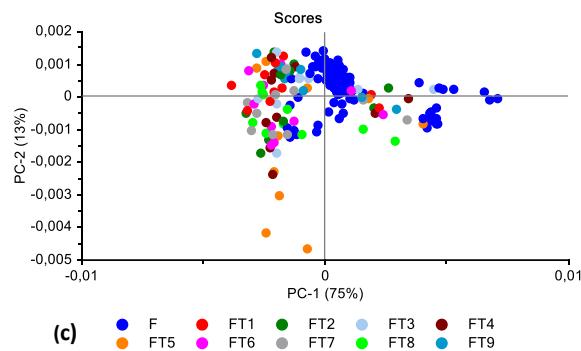


Figure B39 PCA analysis (SGd₂(9) pre-processed) of black wildebeest (LTL) [frozen up to 9 months] illustrating an unsatisfactory separation with an overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (75%) vs. PC2 (13%), (b) PCA loadings line plot for PC1 with interpretable bands at 939, 982, 1031, 1118, 1162, 1298, 1366 and 1403 nm.

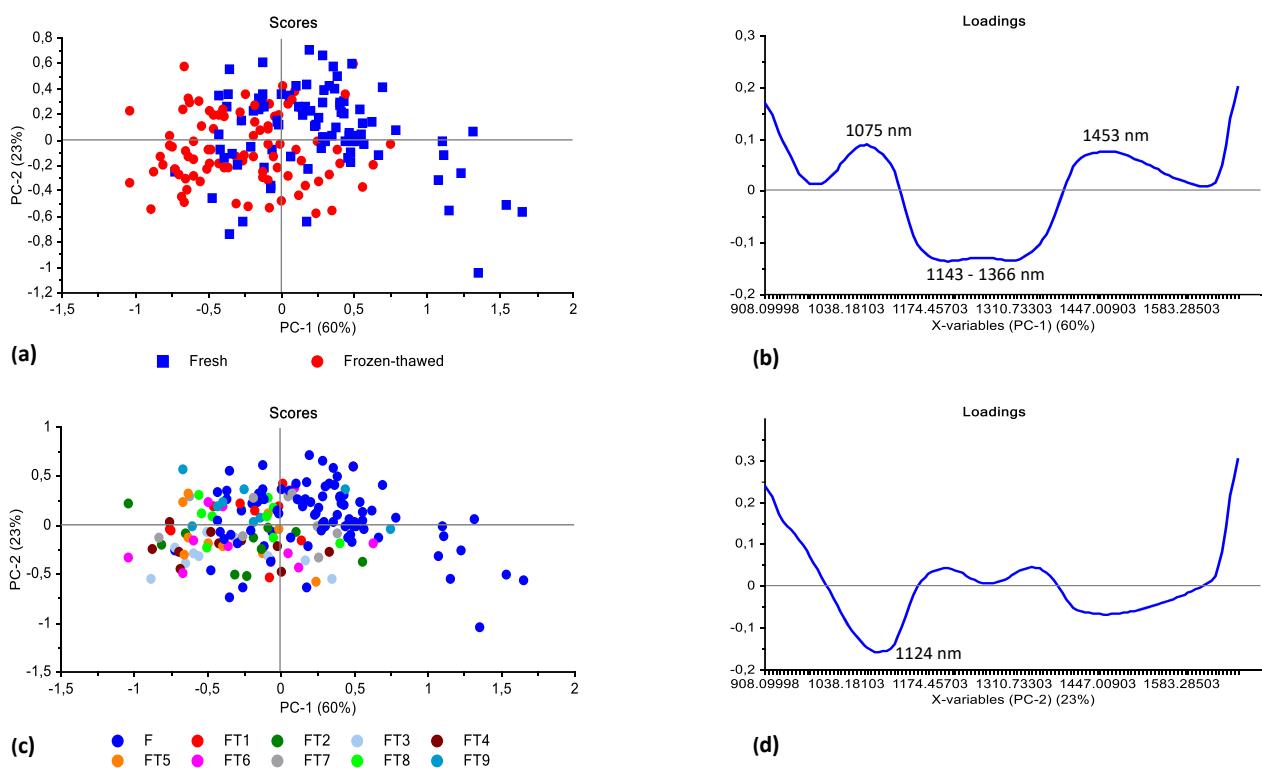


Figure B40 PCA analysis (SNV pre-processed) of zebra (LTL) [frozen up to 9 months] illustrating an unsatisfactory separation with a major overlap between fresh (blue) and frozen-thawed (red) classes as well as between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (60%) vs. PC2 (23%), (b) PCA loadings line plot for PC1 with interpretable bands at 1075, 1143 – 1366 and 1453 nm and (d) PCA loadings line plot for PC2 with an interpretable band at 1124 nm.

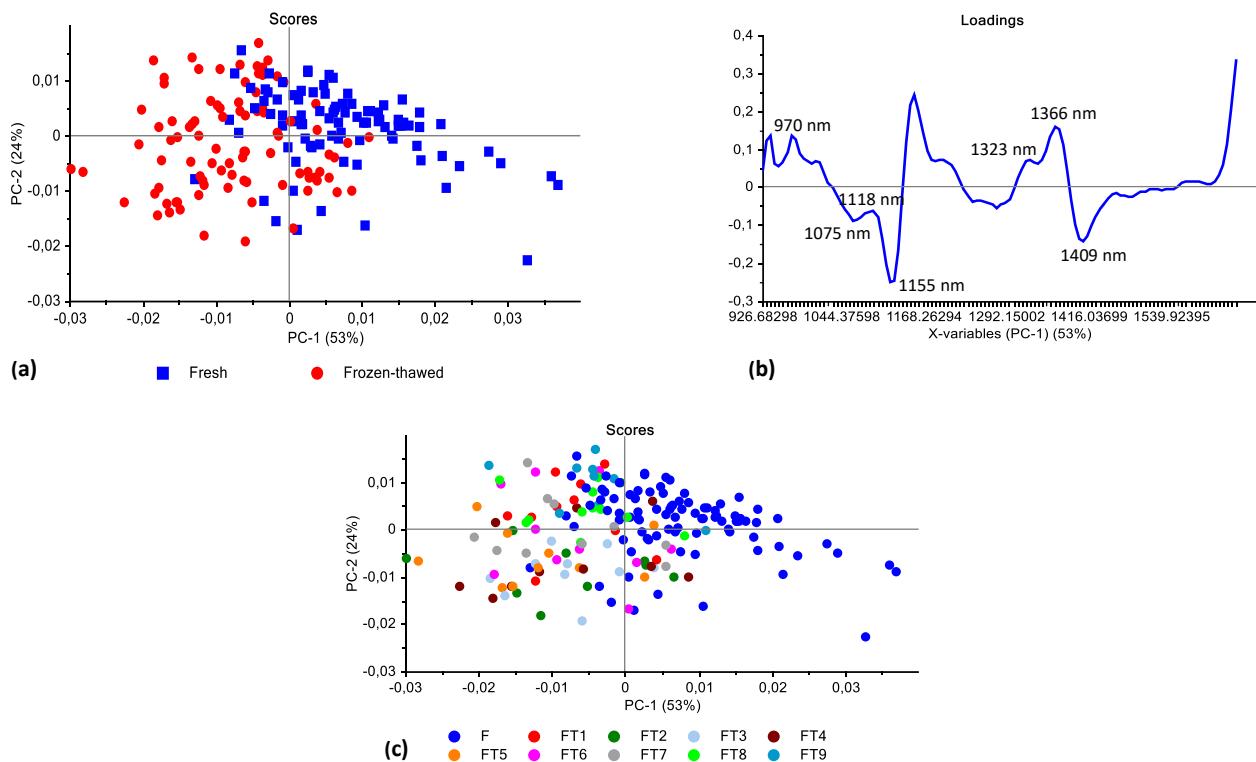


Figure B41 PCA analysis (SNV + SGd₂(7) pre-processed) of zebra (LTL) [frozen up to 9 months] illustrating a good separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (53%) vs. PC2 (24%), (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1075, 1118, 1155, 1323, 1366 and 1409 nm.

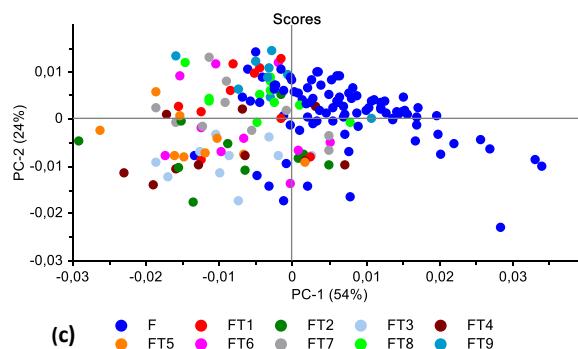
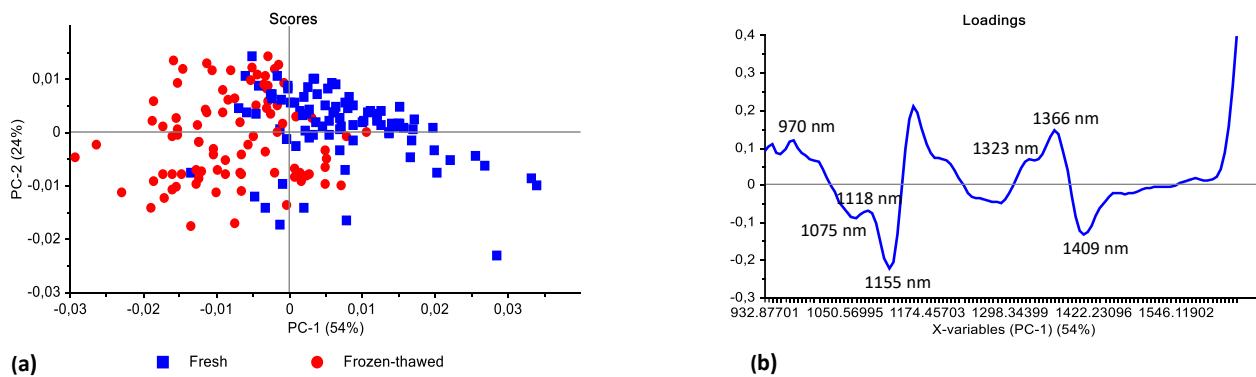


Figure B42 PCA analysis (SNV + SGd₂(9) pre-processed) of zebra (LTL) [frozen up to 9 months] illustrating a good separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (53%) vs. PC2 (24%), (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1075, 1118, 1155, 1323, 1366 and 1409 nm.

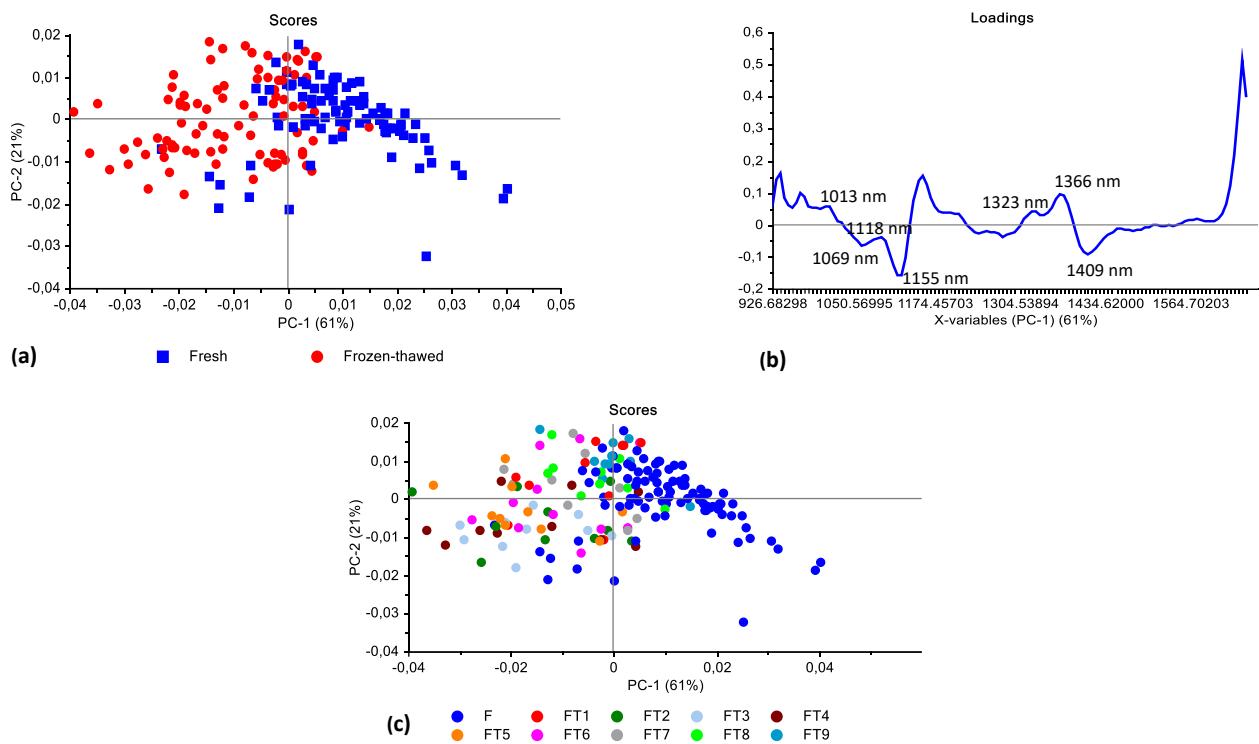
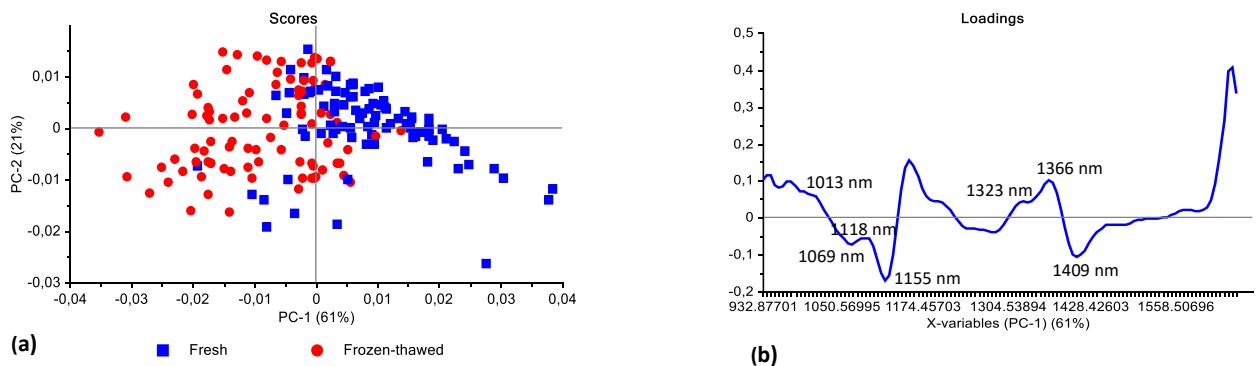


Figure B43 PCA analysis (SNV + detrend + SGd₂(7) pre-processed) of zebra (LTL) [frozen up to 9 months] illustrating a good separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (61%) vs. PC2 (21%), (b) PCA loadings line plot for PC1 with interpretable bands at 1013, 1069, 1118, 1155, 1323, 1366 and 1409 nm.



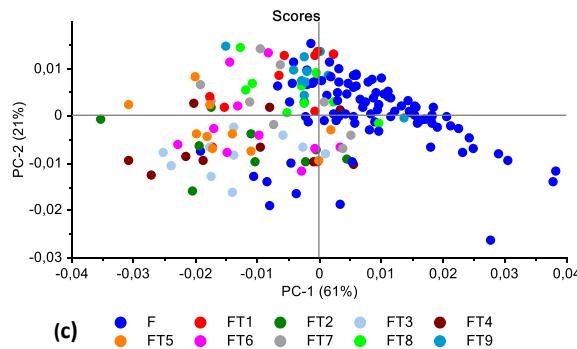


Figure B44 PCA analysis (SNV + detrend + SGd₂(9) pre-processed) of zebra (LTL) [frozen up to 9 months] illustrating a good separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (61%) vs. PC2 (21%), (b) PCA loadings line plot for PC1 with interpretable bands at 1013, 1069, 1118, 1155, 1323, 1366 and 1409 nm.

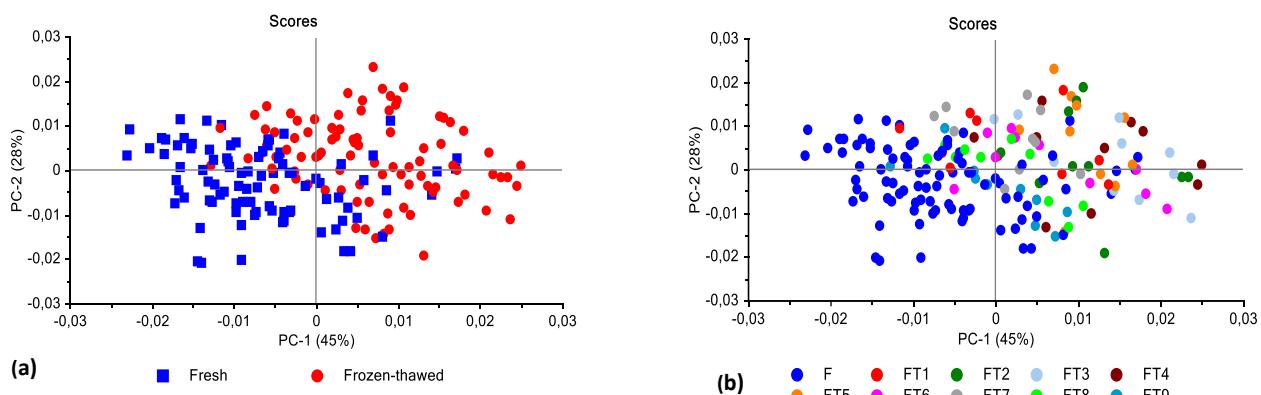


Figure B45 PCA analysis (SGd₁(5) pre-processed) of zebra (LTL) [frozen up to 9 months] illustrating an unsatisfactory separation with an overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (b) PCA score plot of PC1 (45%) vs. PC2 (28%). PCA loadings line plot for PC1, PC2, PC4 and PC5 (not shown) with interpretable bands at 939, 970, 1093, 1143, 1230, 1323 and 1385 nm.

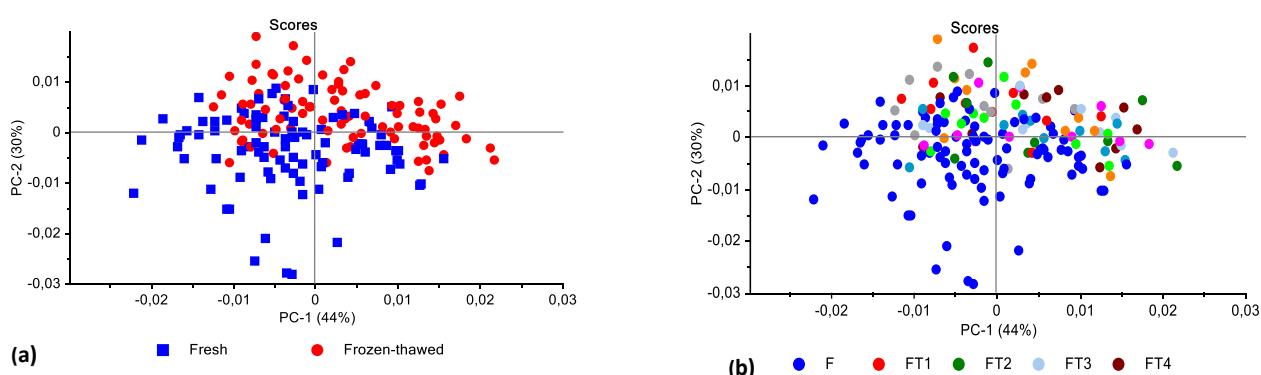


Figure B46 PCA analysis (SGd₁(7) pre-processed) of zebra (LTL) [frozen up to 9 months] illustrating an unsatisfactory separation with an overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (b) PCA score plot of PC1 (44%) vs. PC2 (30%). PCA loadings line plot for PC1, PC2, PC4 and PC5 (not shown) with interpretable bands at 939, 970, 1031, 1093, 1143, 1230, 1323 and 1385 nm.

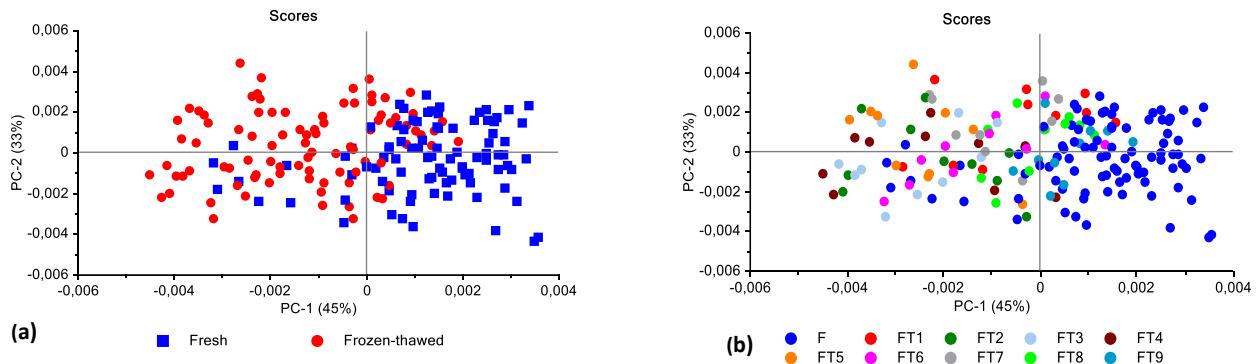


Figure B47 PCA analysis (SGd₂(9) pre-processed) of zebra (LTL) [frozen up to 9 months] illustrating an unsatisfactory separation with an overlap between fresh (blue) and frozen-thawed (red) and a major overlap between frozen period classes. Scores illustrated as (a) + (b) PCA score plot of PC1 (44%) vs. PC2 (30%). PCA loadings line plot for PC1, PC2 and PC3 (not shown) with interpretable bands at 945, 970, 1155, 1193, 1292, 1304, 1341, 1385, 1409 nm.

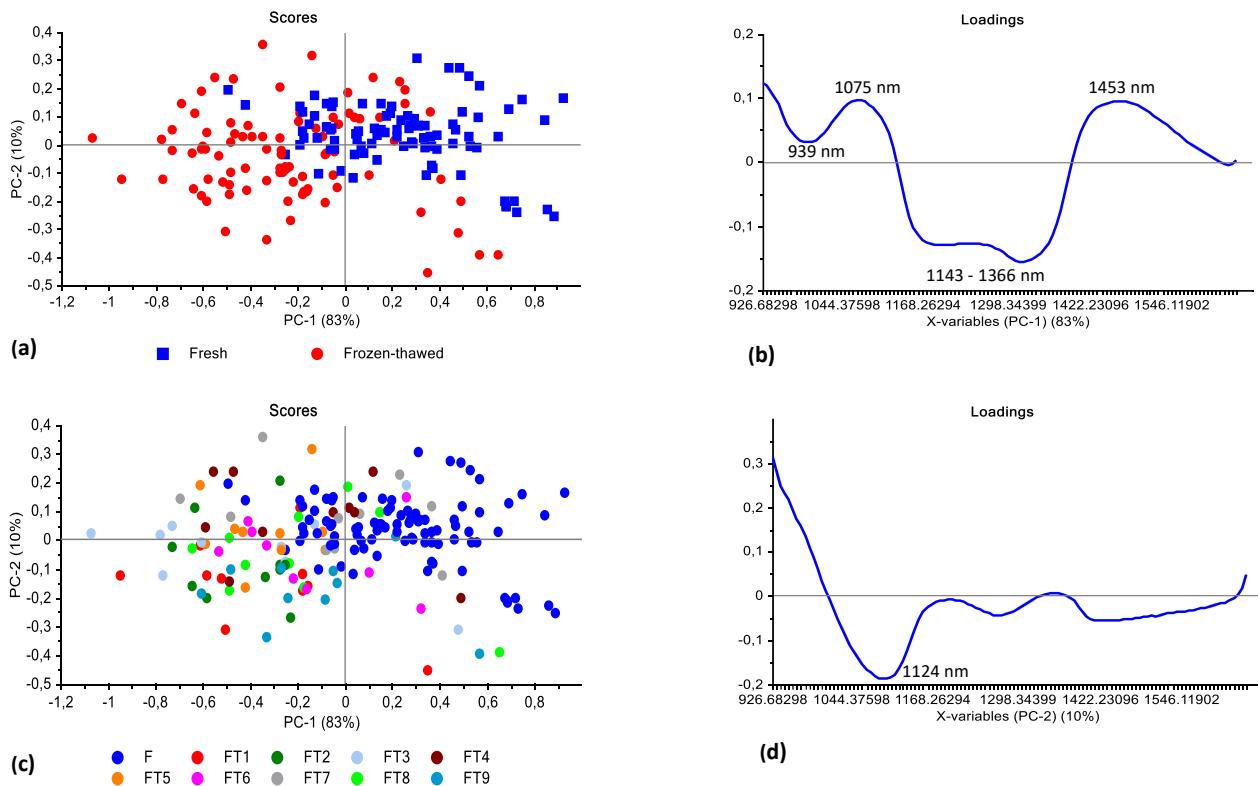


Figure B48 PCA analysis (SNV pre-processed) of springbok (LTL) [frozen up to 9 months] illustrating a medium separation with an overlap between fresh (blue) and frozen-thawed (red) classes as well as between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (83%) vs. PC2 (10%), (b) PCA loadings line plot for PC1 with interpretable bands at 939, 1075, 1143 – 1366 and 1453 nm and (d) PCA loadings line plot for PC2 with interpretable bands at 970 and 1124 nm.

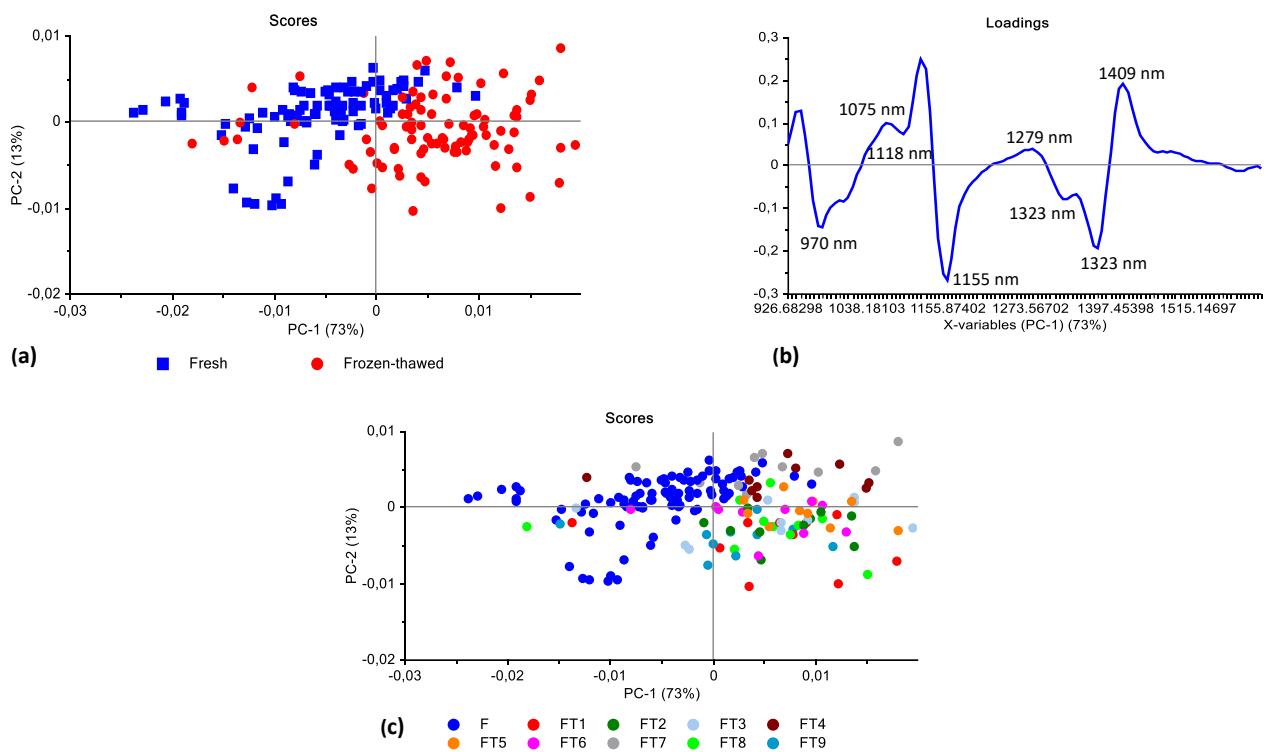


Figure B49 PCA analysis (SNV + SGd₂(7) pre-processed) of springbok (LTL) [frozen up to 9 months] illustrating a good separation with a slight overlap between fresh (blue) and frozen-thawed (red) and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (73%) vs. PC2 (13%), (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1075, 1118, 1155, 1279, 1323, 1366 and 1409 nm.

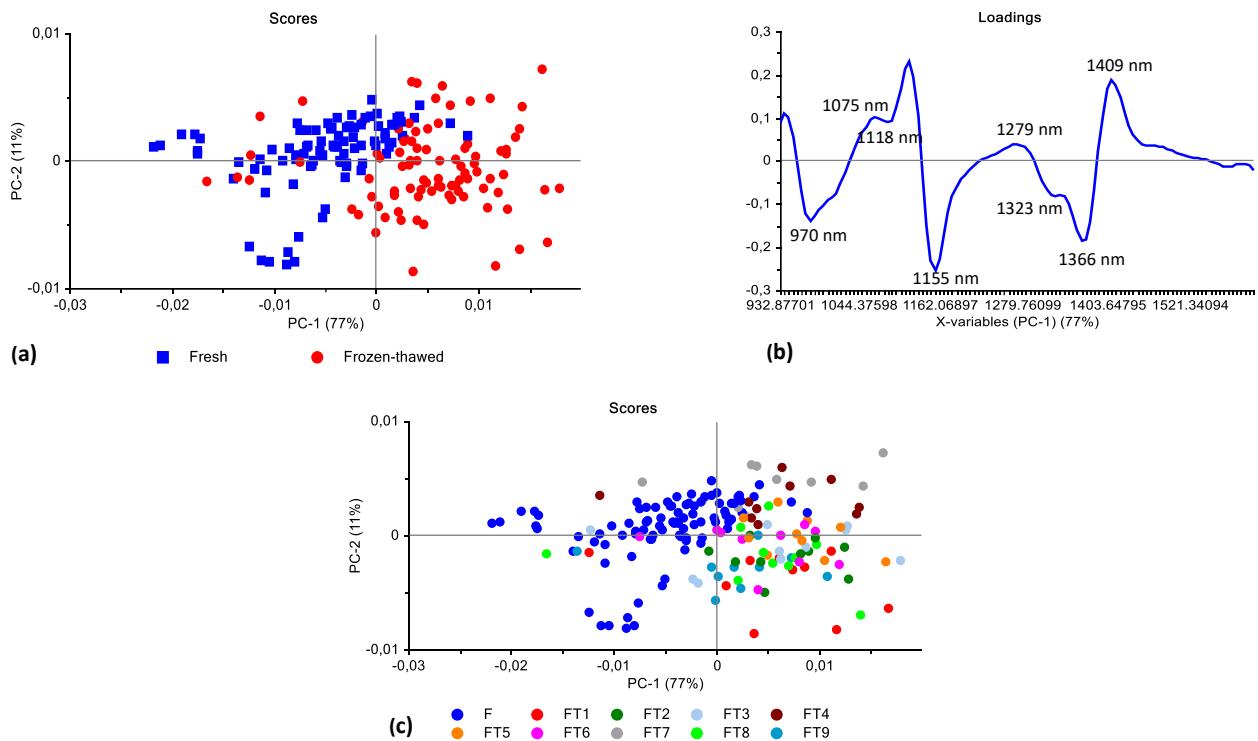


Figure B50 PCA analysis (SNV + SGd₂(9) pre-processed) of springbok (LTL) [frozen up to 9 months] illustrating a good separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (77%) vs. PC2 (11%), (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1075, 1118, 1155, 1279, 1323, 1366 and 1409 nm.

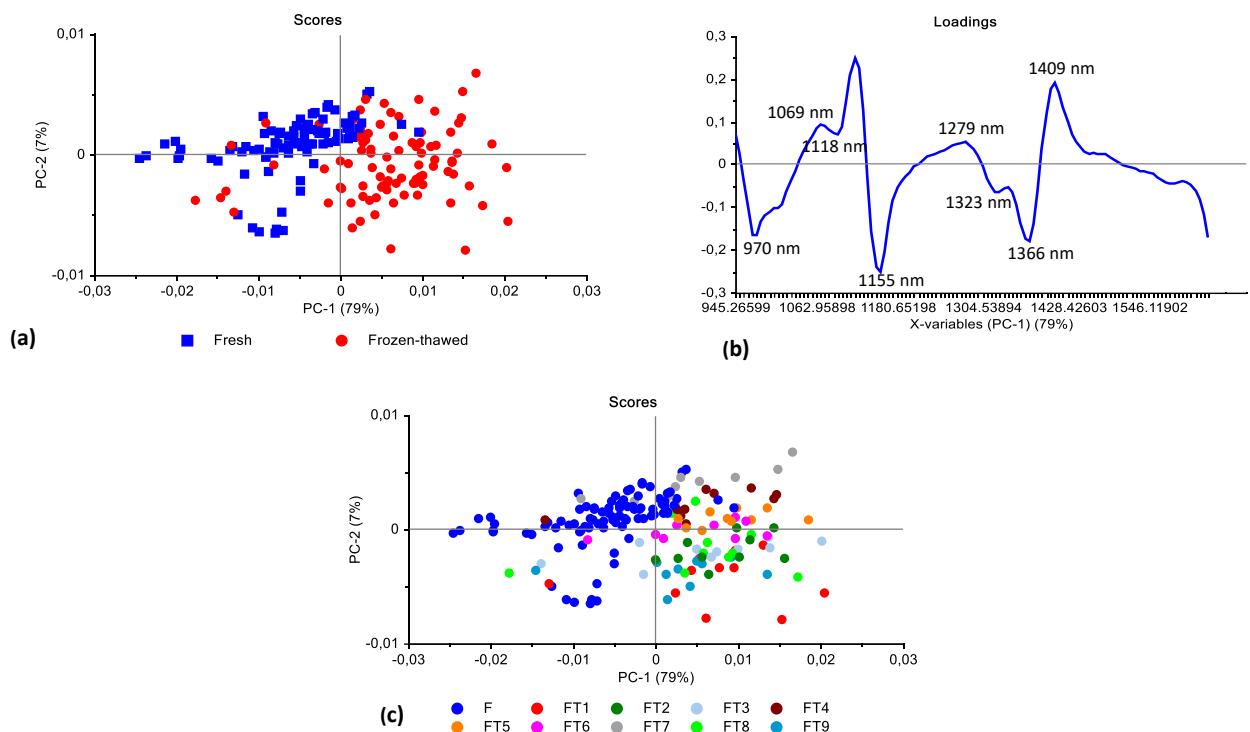


Figure B51 PCA analysis (SNV + detrend + SGd₂(7) pre-processed) of springbok (LTL) [frozen up to 9 months] illustrating a good separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (79%) vs. PC2 (7%), (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1069, 1118, 1155, 1279, 1323, 1366 and 1409 nm.

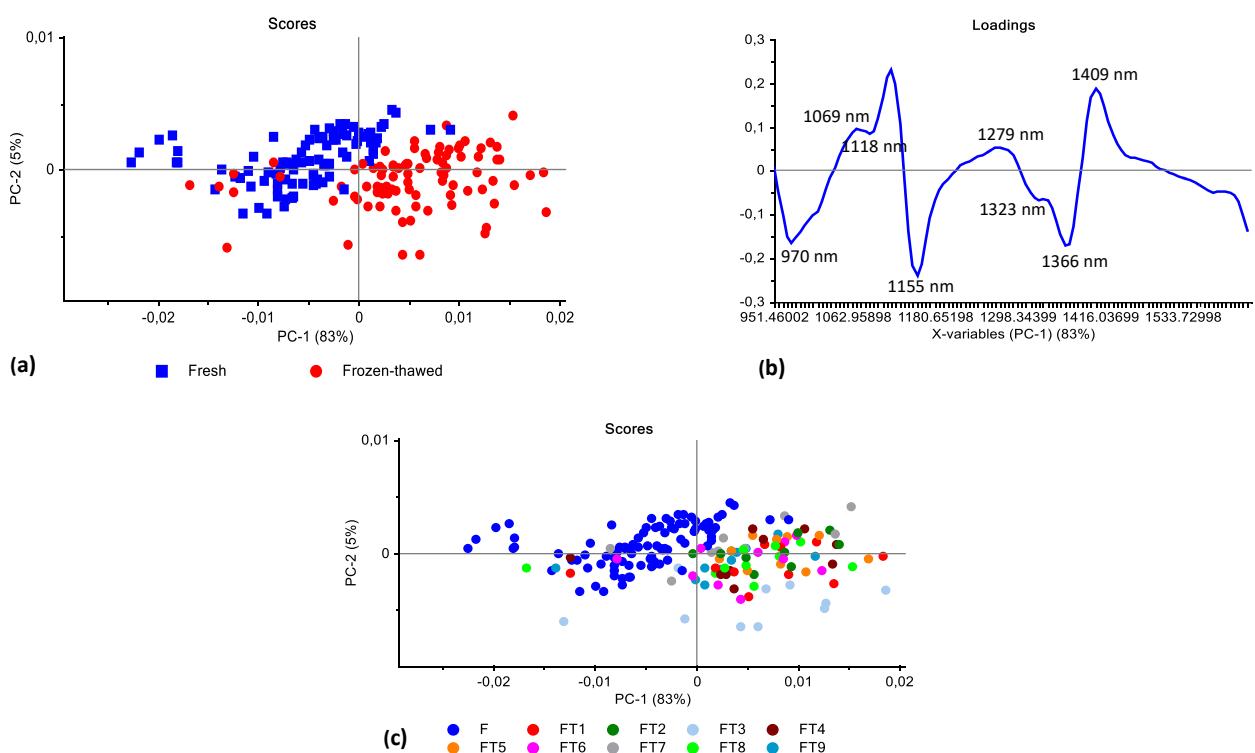


Figure B52 PCA analysis (SNV + detrend + SGd₂(9) pre-processed) of springbok (LTL) [frozen up to 9 months] illustrating a good separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (83%) vs. PC2 (5%), (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1069, 1118, 1155, 1279, 1323, 1366 and 1409 nm.

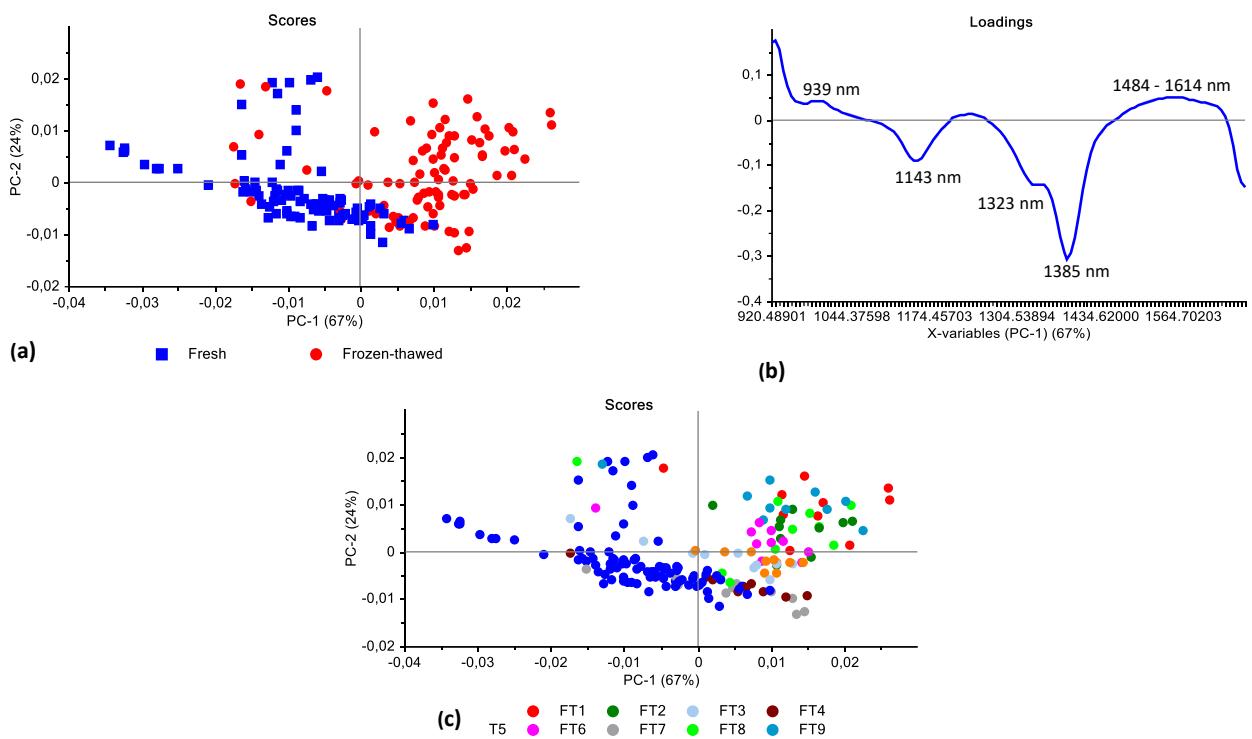


Figure B53 PCA analysis (SGd₁(5) pre-processed) of springbok (LTL) [frozen up to 9 months] illustrating a good separation with an overlap between fresh (blue) and frozen-thawed (red) classes and a slight overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (67%) vs. PC2 (24%), (b) PCA loadings line plot for PC1 with interpretable bands at 939, 1143, 1323, 1385 and 1484 – 1614 nm.

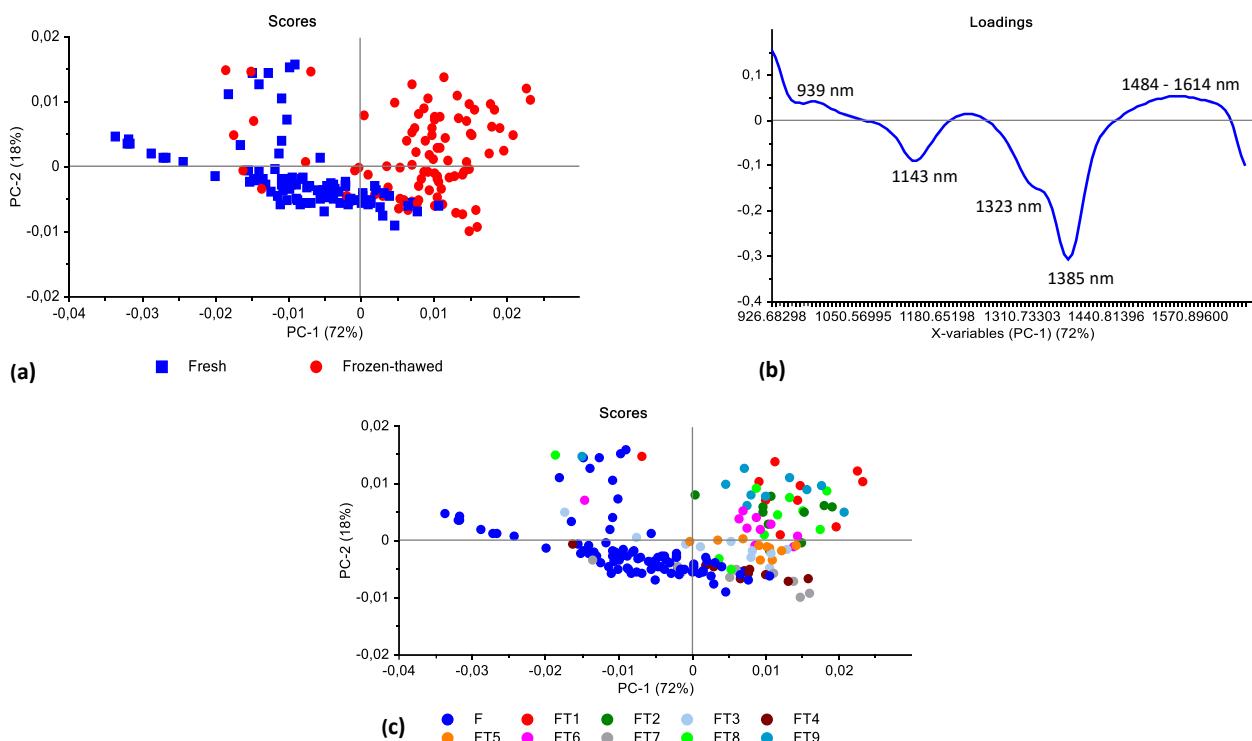


Figure B54 PCA analysis (SGd₁(7) pre-processed) of springbok (LTL) [frozen up to 9 months] illustrating a good separation with an overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (72%) vs. PC2 (18%), (b) PCA loadings line plot for PC1 with interpretable bands at 939, 1143, 1323, 1385 and 1484 – 1614 nm.

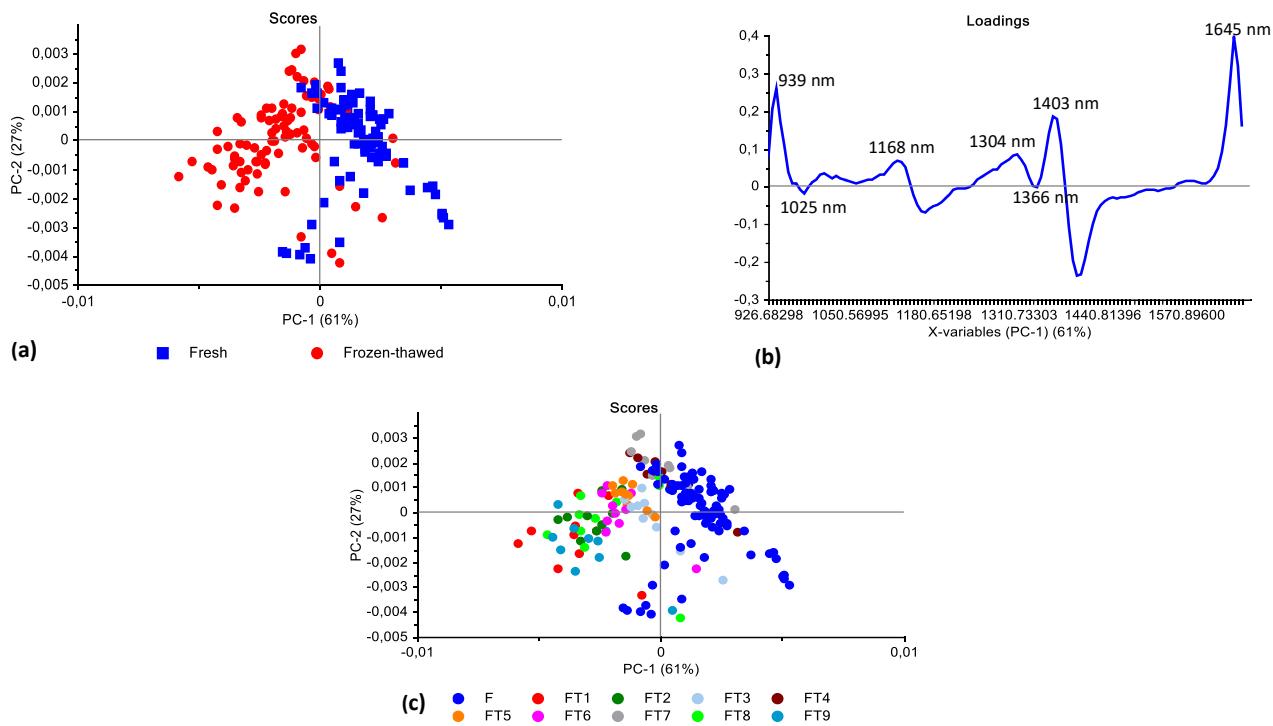


Figure B55 PCA analysis (SGd₂(7) pre-processed) of springbok (LTL) [frozen up to 9 months] illustrating a good separation with an overlap between fresh (blue) and frozen-thawed (red) classes and a slight overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (61%) vs. PC2 (27%), (b) PCA loadings line plot for PC1 with interpretable bands at 939, 1025, 1168, 1304, 1366, 1403 and 1645 nm.

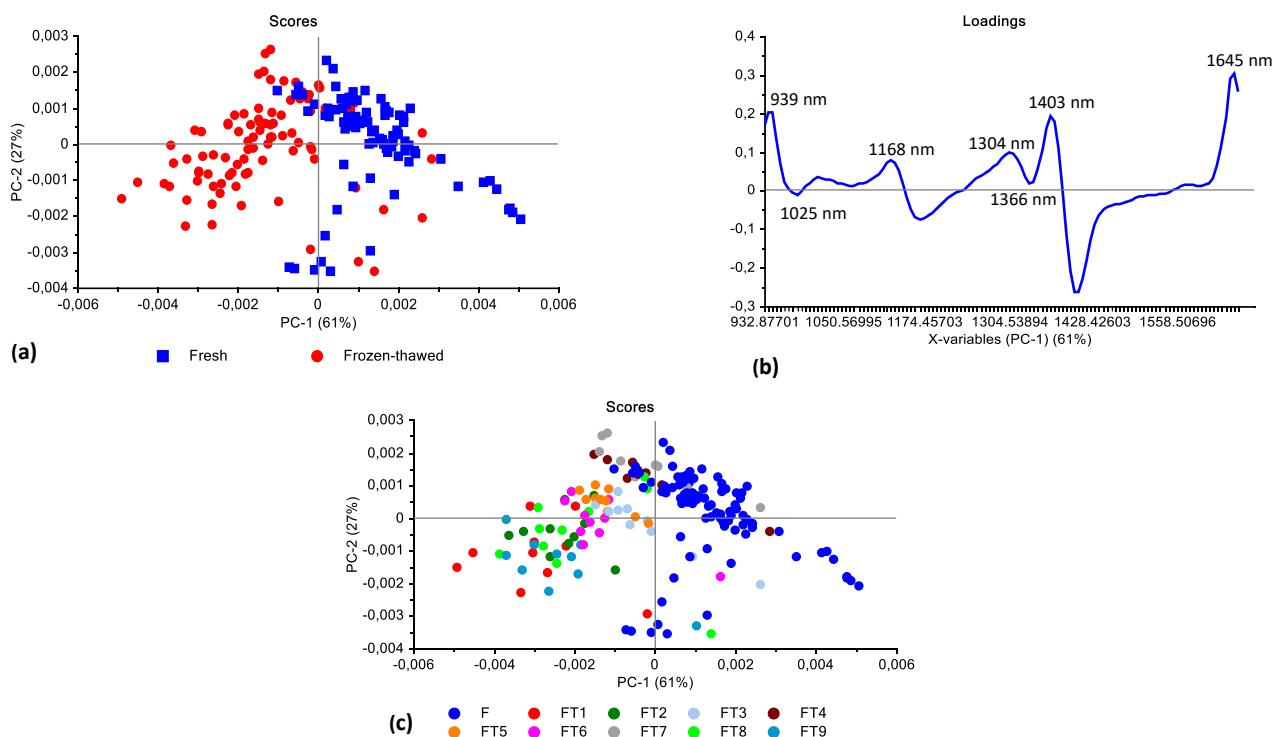


Figure B56 PCA analysis (SGd₂(9) pre-processed) of springbok (LTL) [frozen up to 9 months] illustrating a good separation with an overlap between fresh (blue) and frozen-thawed (red) classes and a slight overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (61%) vs. PC2 (27%), (b) PCA loadings line plot for PC1 with interpretable bands at 939, 1025, 1168, 1304, 1366, 1403 and 1645 nm.

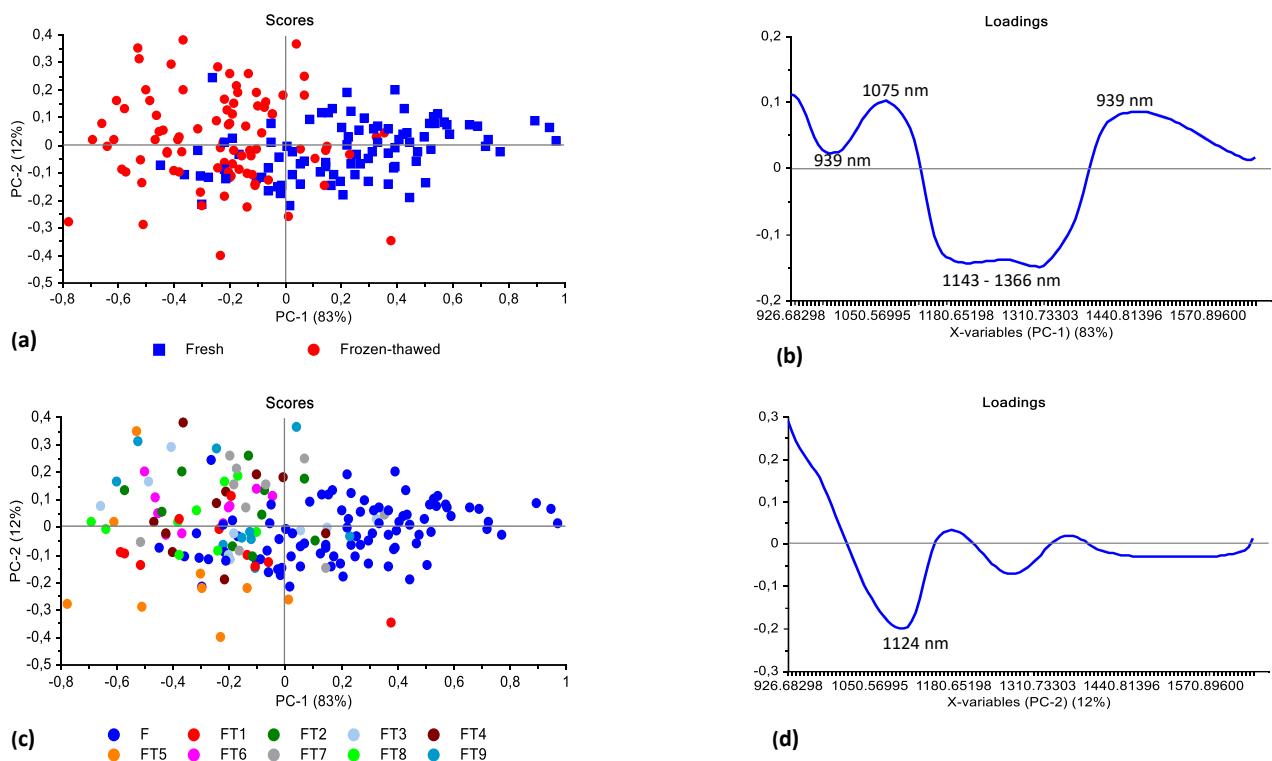


Figure B57 PCA analysis (SNV pre-processed) of ostrich (FF) [frozen up to 9 months] illustrating a medium separation with an overlap between fresh (blue) and frozen-thawed (red) classes as well as between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (83%) vs. PC2 (12%), (b) PCA loadings line plot for PC1 with interpretable bands at 939, 1075, 1143 – 1366 and 1453 nm and (d) PCA loadings line plot for PC2 with interpretable bands at 970 and 1124 nm.

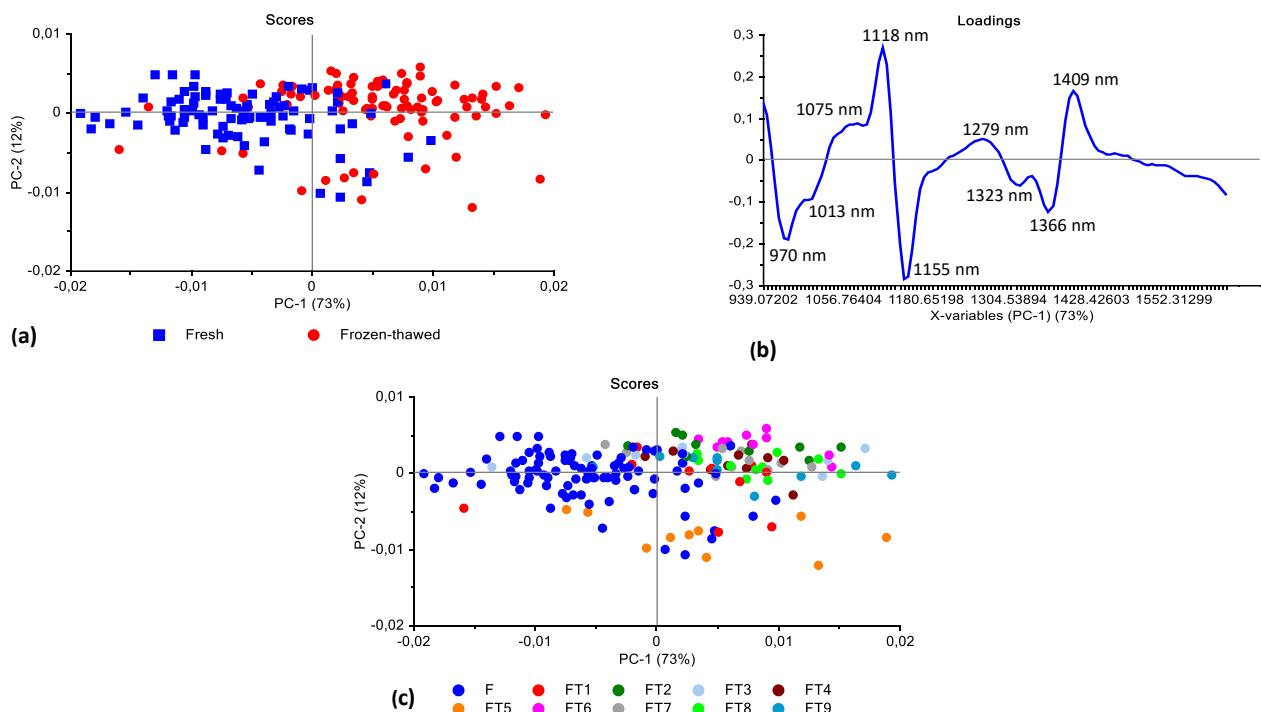


Figure B58 PCA analysis (SNV + SGd₂(7) pre-processed) of ostrich (FF) [frozen up to 9 months] illustrating a good separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between

frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (73%) vs. PC2 (12%), (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1013, 1075, 1118, 1155, 1279, 1323, 1366 and 1409 nm.

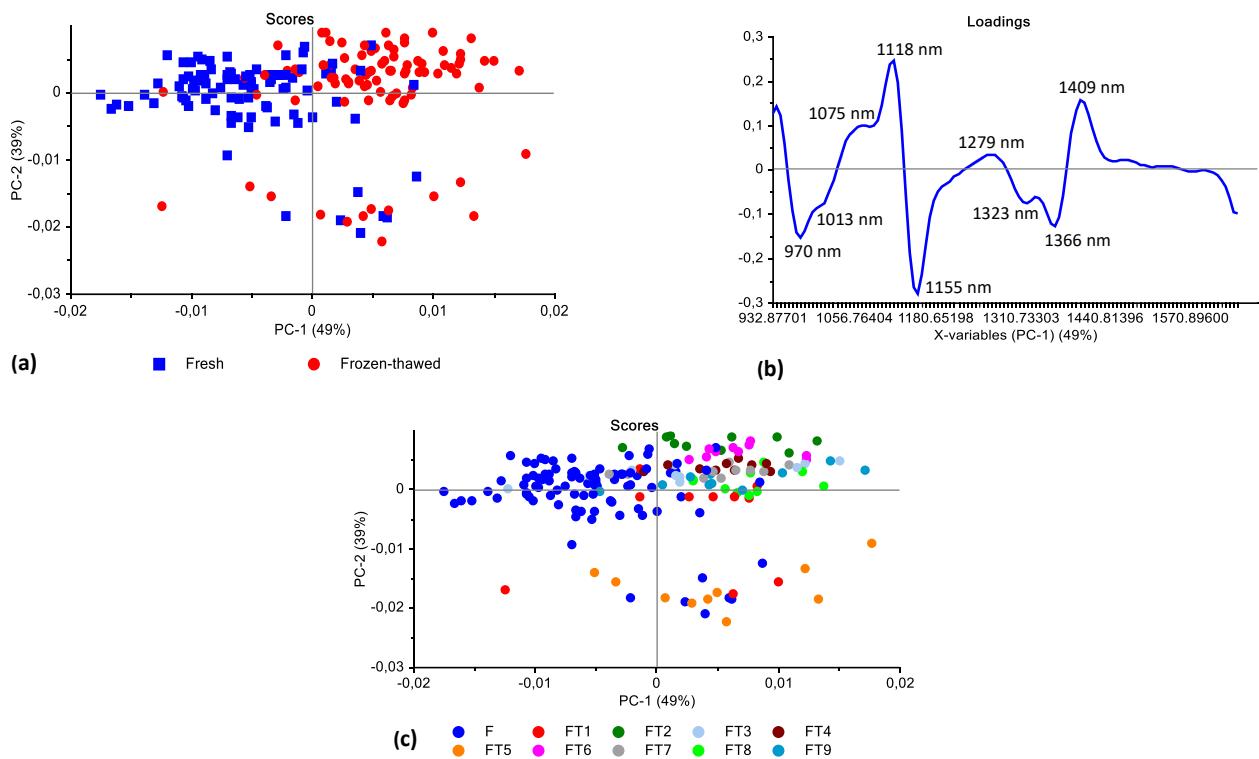
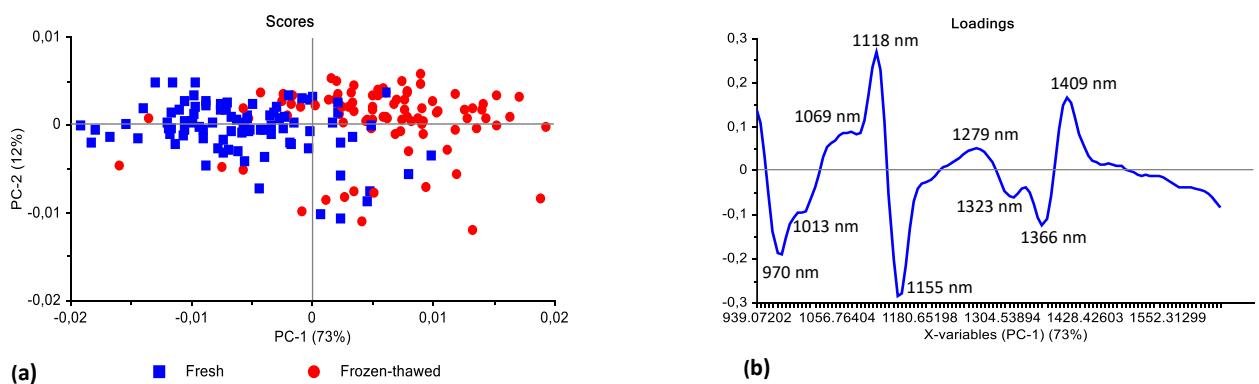


Figure B59 PCA analysis (SNV + SGd₂(9) pre-processed) of ostrich (FF) [frozen up to 9 months] illustrating an unsatisfactory separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (49%) vs. PC2 (39%), (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1013, 1075, 1118, 1155, 1279, 1323, 1366 and 1409 nm.



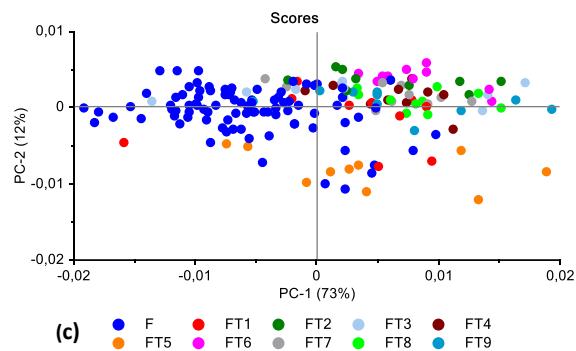


Figure B60 PCA analysis (SNV + detrend + SGd₂(7) pre-processed) of ostrich (FF) [frozen up to 9 months] illustrating a good separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (73%) vs. PC2 (12%), (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1013, 1069, 1118, 1155, 1279, 1323, 1366 and 1409 nm.

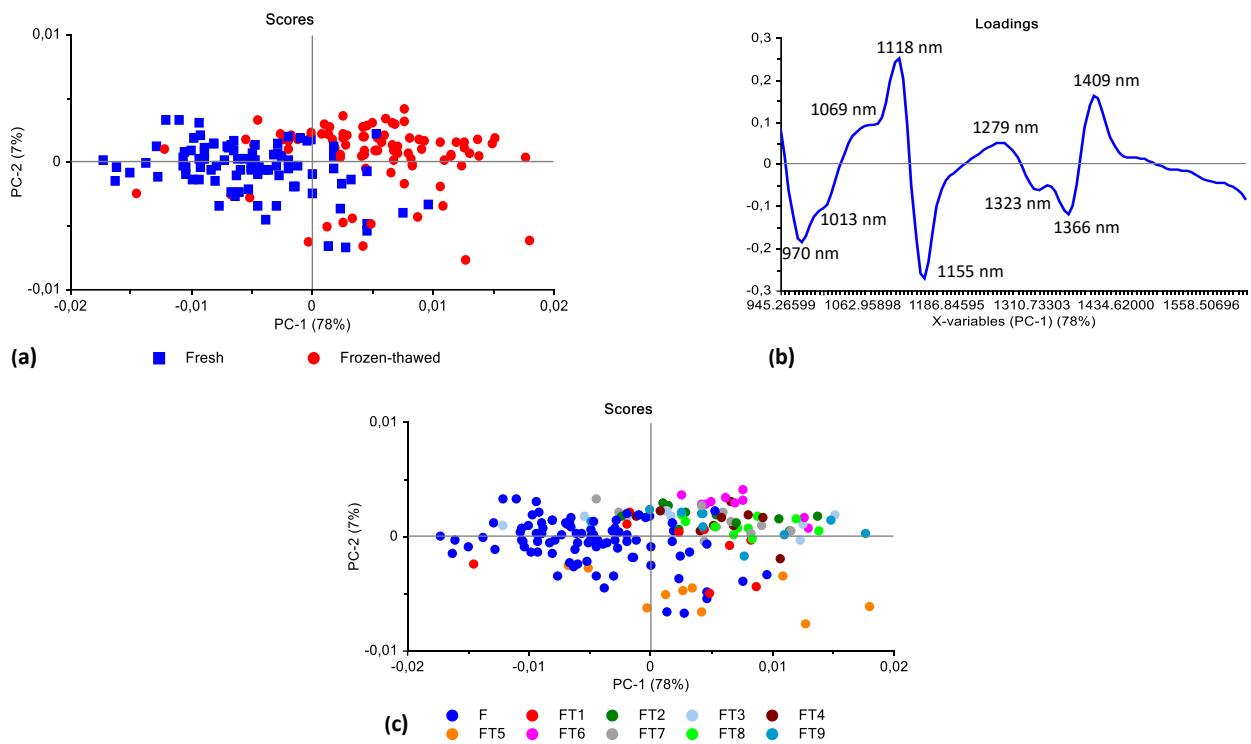


Figure B61 PCA analysis (SNV + detrend + SGd₂(9) pre-processed) of ostrich (FF) [frozen up to 9 months] illustrating a good separation with a slight overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (78%) vs. PC2 (7%), (b) PCA loadings line plot for PC1 with interpretable bands at 970, 1013, 1069, 1118, 1155, 1279, 1323, 1366 and 1409 nm.

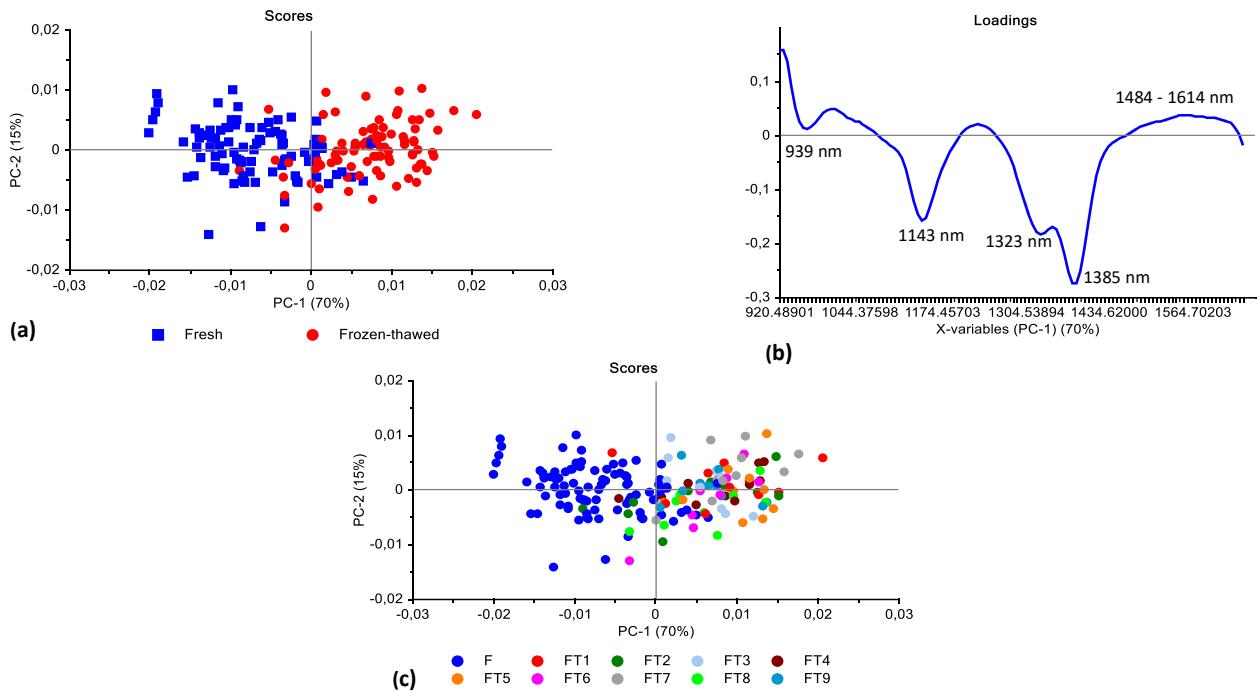


Figure B62 PCA analysis (SGd₁(5) pre-processed) of ostrich (FF) [frozen up to 9 months] illustrating a good separation with an overlap between fresh (blue) and frozen-thawed (red) classes and a slight overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (70%) vs. PC2 (15%), (b) PCA loadings line plot for PC1 with interpretable bands at 939, 1143, 1323, 1385 and 1484 – 1614 nm.

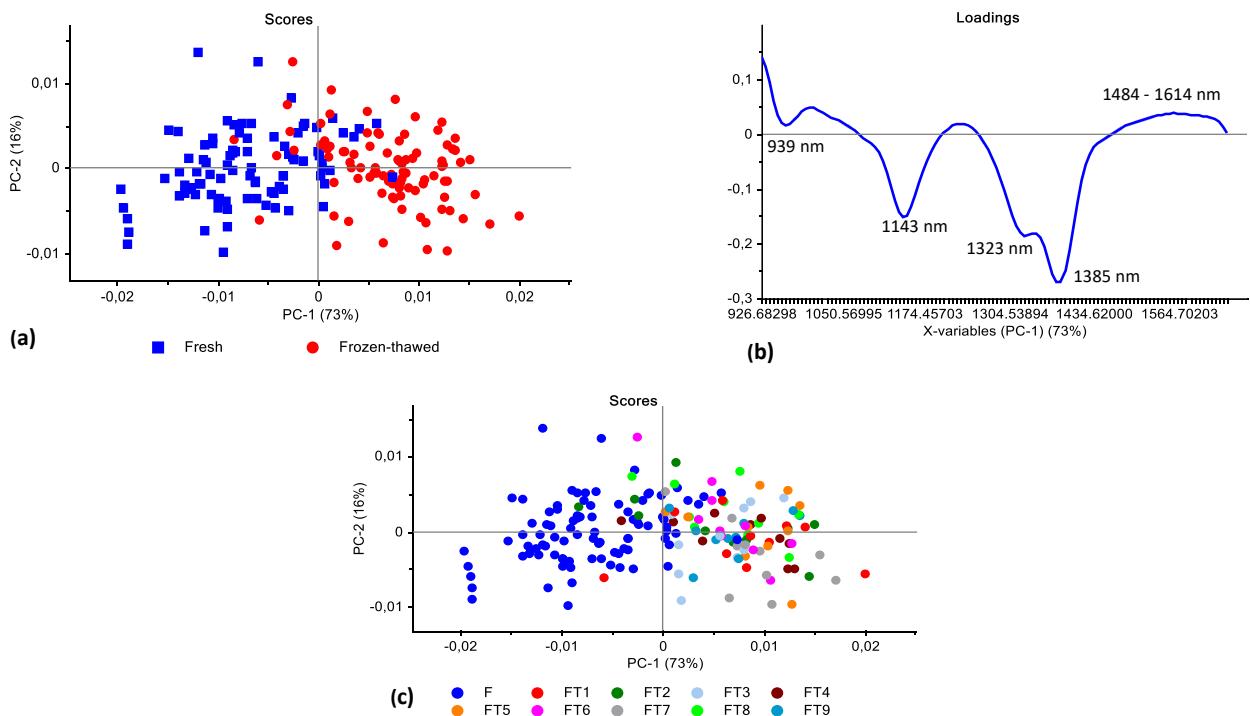


Figure B63 PCA analysis (SGd₁(7) pre-processed) of ostrich (FF) [frozen up to 9 months] illustrating a good separation with an overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (73%) vs. PC2 (16%), (b) PCA loadings line plot for PC1 with interpretable bands at 939, 1143, 1323, 1385 and 1484 – 1614 nm.

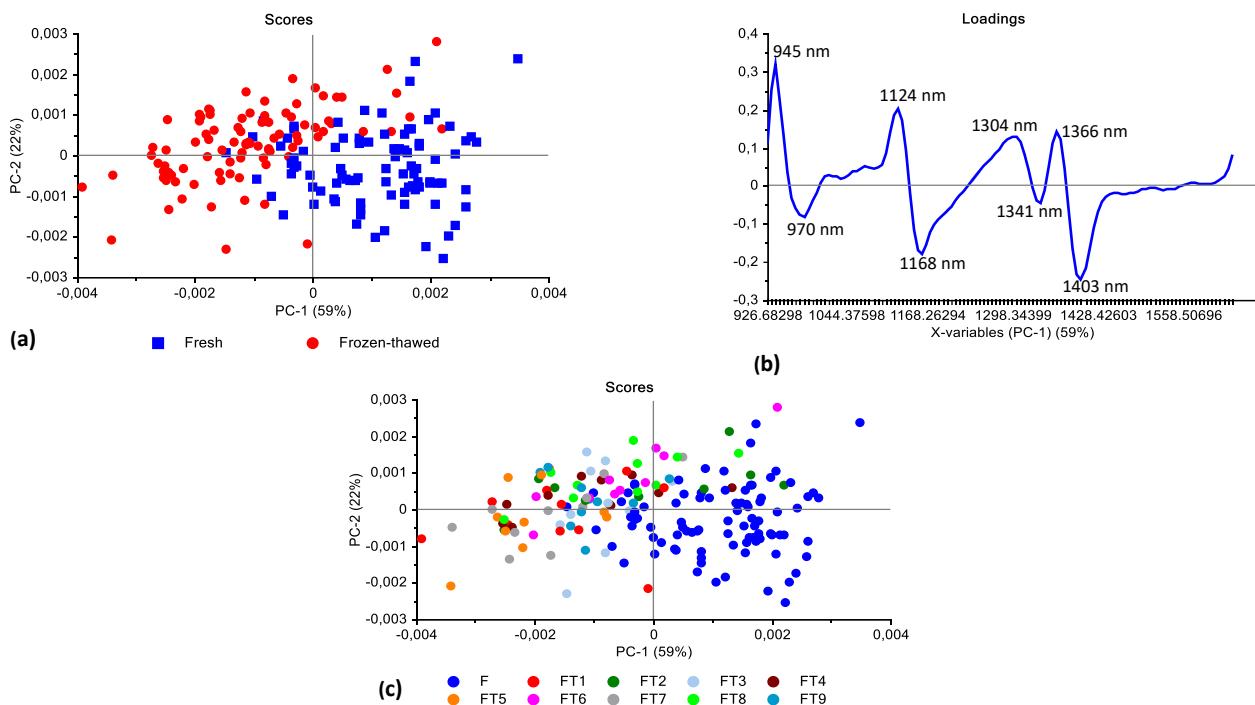


Figure B64 PCA analysis (SGd₂(7) pre-processed) of ostrich (FF) [frozen up to 9 months] illustrating a good separation with an overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (59%) vs. PC2 (22%), (b) PCA loadings line plot for PC1 with interpretable bands at 945, 970, 1124, 1168, 1304, 1341, 1366 and 1403 nm.

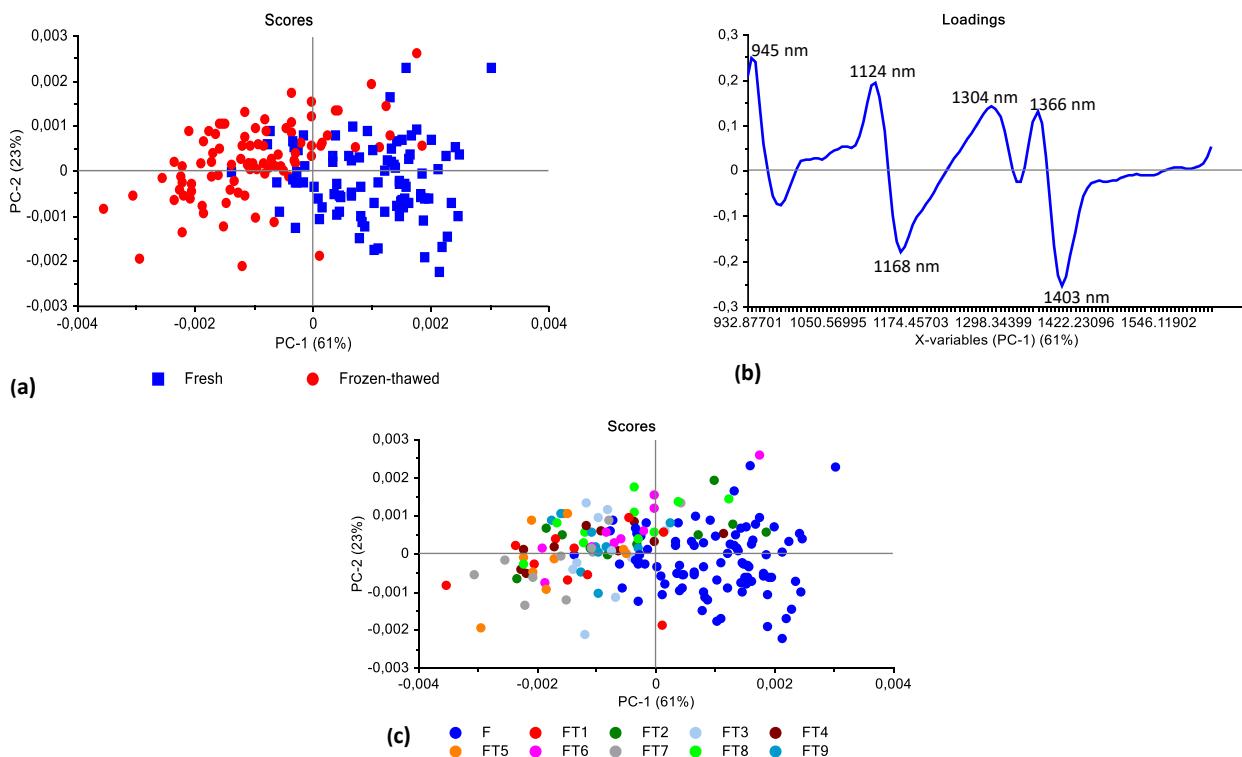
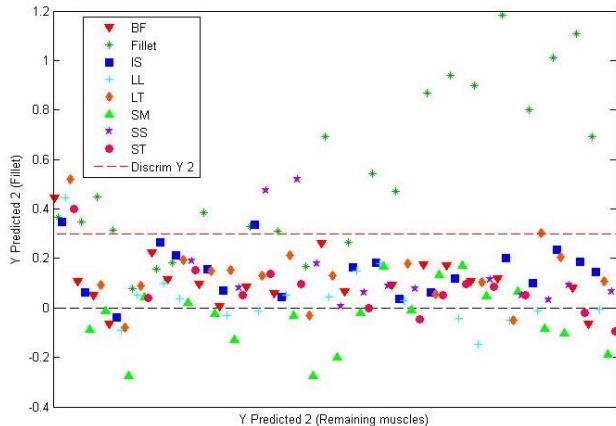


Figure B65 PCA analysis (SGd₂(9) pre-processed) of springbok (LTL) [frozen up to 9 months] illustrating a good separation with an overlap between fresh (blue) and frozen-thawed (red) classes and a major overlap between frozen period classes. Scores illustrated as (a) + (c) PCA score plot of PC1 (61%) vs. PC2 (23%), (b) PCA loadings line plot for PC1 with interpretable bands at 945, 1124, 1168, 1304, 1366 and 1403 nm.

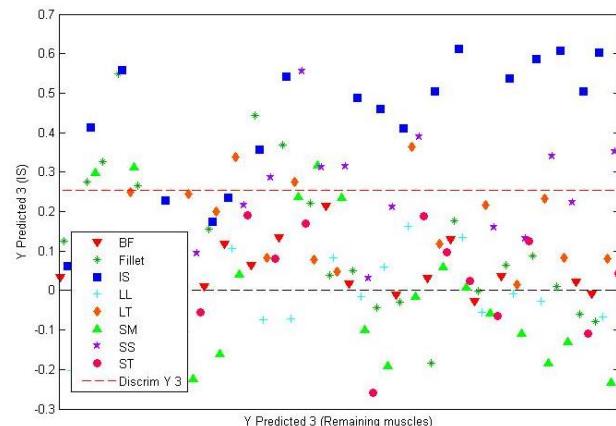
Addendum C

Supplementary information pertaining to Chapter 4:

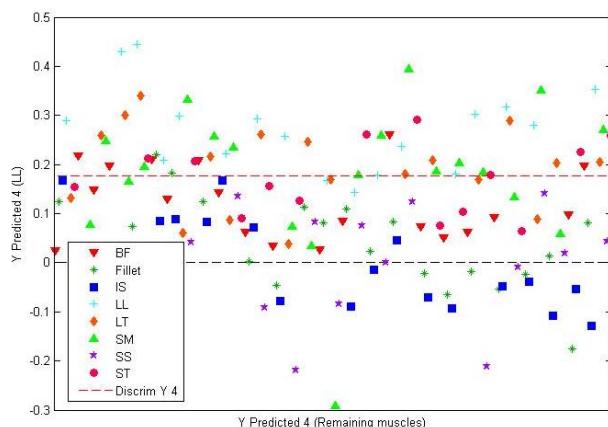
4.3 Muscle type determination



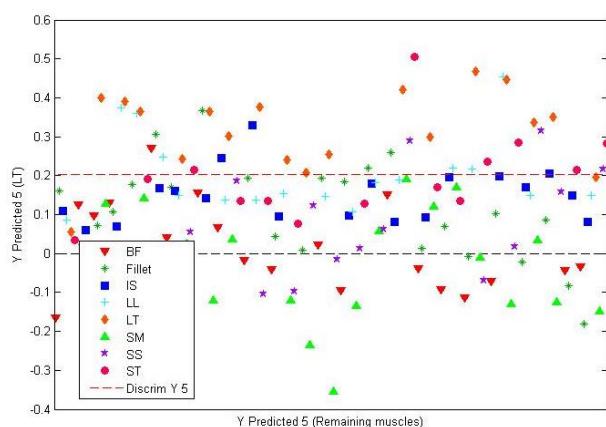
(a)



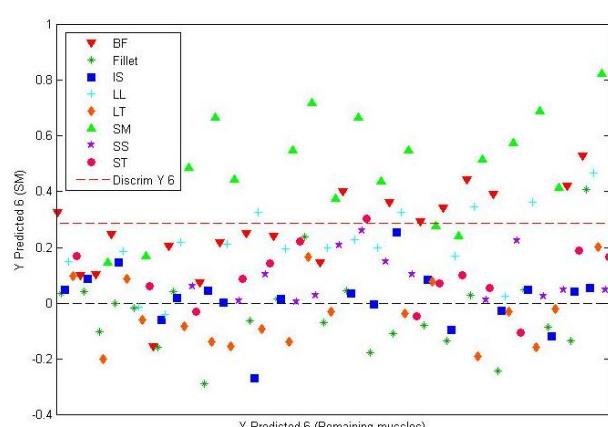
(b)



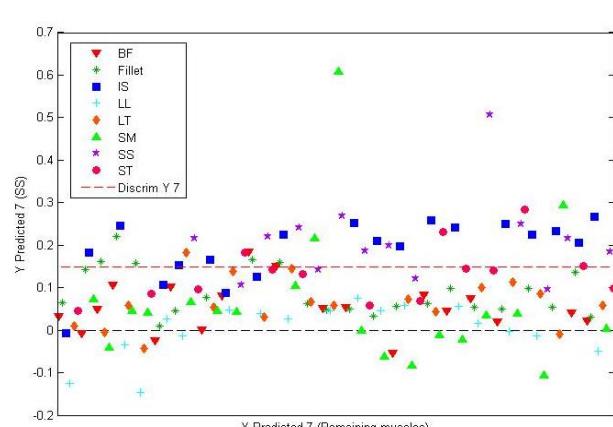
(c)



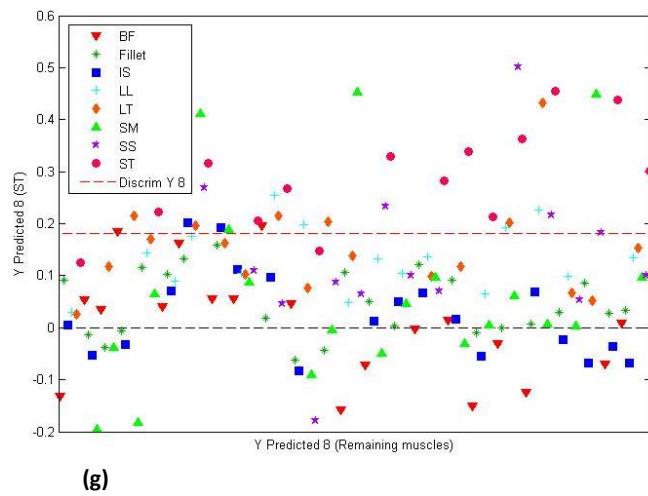
(d)



(e)

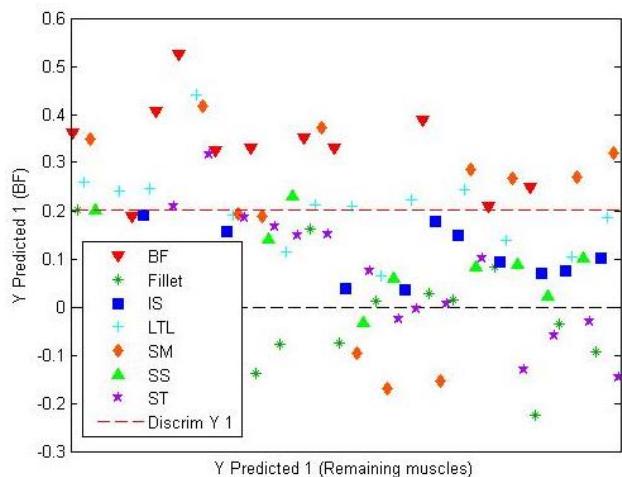


(f)

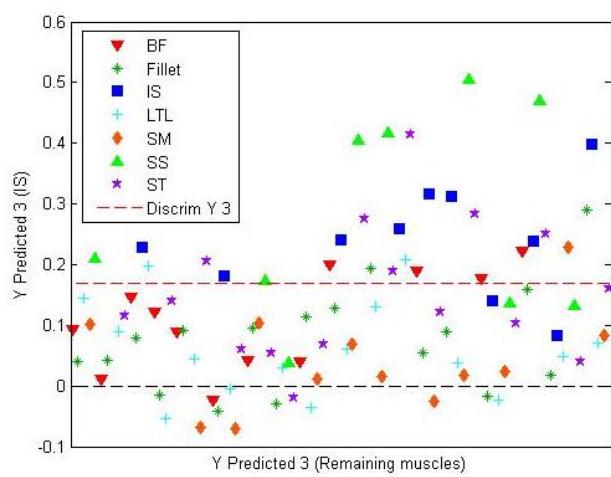


(g)

Figure C1 PLS-DA models (6 LV's) (SNV + detrend pre-processed) for zebra muscles discrimination, resulting in satisfactory individual classification accuracies. PLS-DA prediction score plots illustrating the predicted objects. (a) Score plot for Fillet (87.4%) of objects predicted as Fillet [above red line (Y2)] vs. remaining muscles [below red line (Y2)], (b) score plot for IS (88.2%) of objects predicted as IS [above red line (Y3)] vs. remaining muscles [below red line (Y3)], (c) score plot for LL (88.2%) of objects predicted as LL [above red line (Y4)] vs. remaining muscles [below red line (Y4)], (d) score plot for LT (87.4%) of objects predicted as LT [above red line (Y5)] vs. remaining muscles [below red line (Y5)], (e) score plot for SM (91.5%) of objects predicted as SM [above red line (Y6)] vs. remaining muscles [below red line (Y6)], (f) score plot for SS (88.2%) of objects predicted as SS [above red line (Y7)] vs. remaining muscles [below red line (Y7)], (g) score plot for ST (91.5%) of objects predicted as ST [above red line (Y8)] vs. remaining muscles [below red line (Y8)].



(a)



(b)

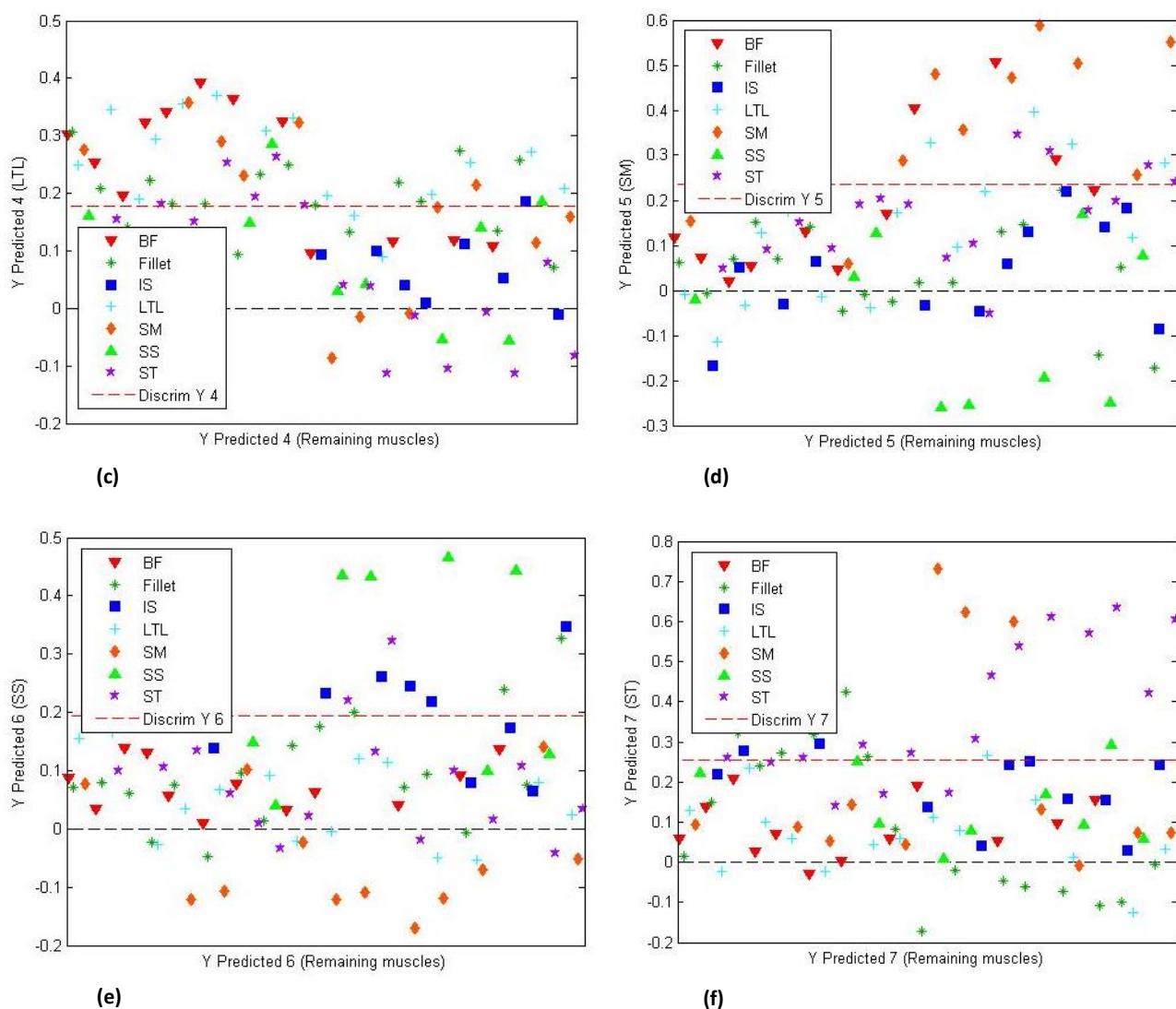


Figure C2 PLS-DA models (4 LV's) (SNV + detrend pre-processed) for springbok muscles discrimination, resulting in satisfactory individual classification accuracies. PLS-DA prediction score plots illustrating the predicted objects. (a) Score plot for BF (77.1%) of objects predicted as BF [above red line (Y1)] vs. remaining muscles [below red line (Y1)], (b) score plot for IS (82.5%) of objects predicted as IS [above red line (Y3)] vs. remaining muscles [below red line (Y3)], (c) score plot for LTL (70.2%) of objects predicted as LTL [above red line (Y4)] vs. remaining muscles [below red line (Y4)], (d) score plot for SM (74.6%) of objects predicted as SM [above red line (Y5)] vs. remaining muscles [below red line (Y5)], (e) score plot for SS (85.5%) of objects predicted as SS [above red line (Y6)] vs. remaining muscles [below red line (Y6)], (f) score plot for ST (79.7%) of objects predicted as ST [above red line (Y7)] vs. remaining muscles [below red line (Y7)].

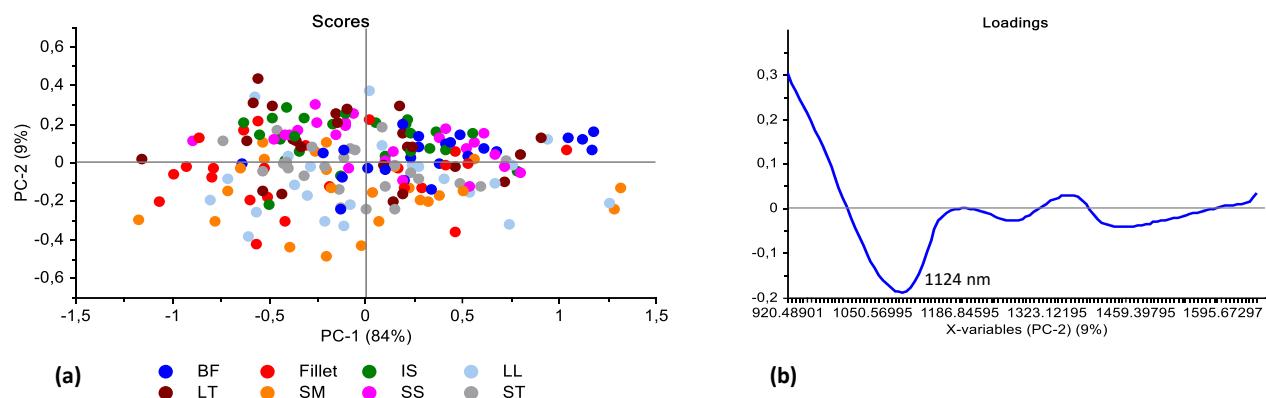


Figure C3 PCA analysis (SNV pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a major overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (84%) vs. PC2 (9%). (b) PCA loadings line plot for PC2 with an interpretable absorption band at 1124 nm.

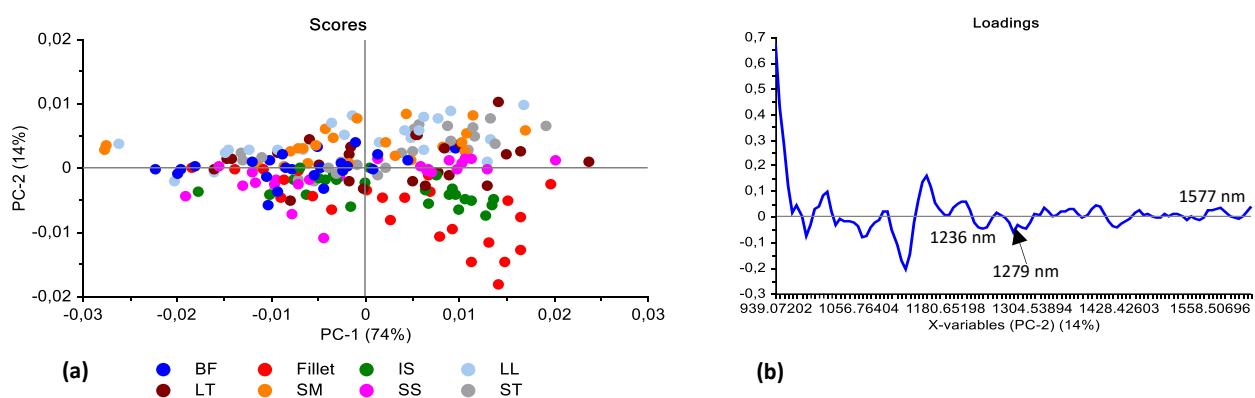


Figure C4 PCA analysis (SNV + SGd₂(7) pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a major overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (74%) vs. PC2 (14%). (b) PCA loadings line plot for PC2 with interpretable absorption bands at 1236, 1279 and 1577 nm.

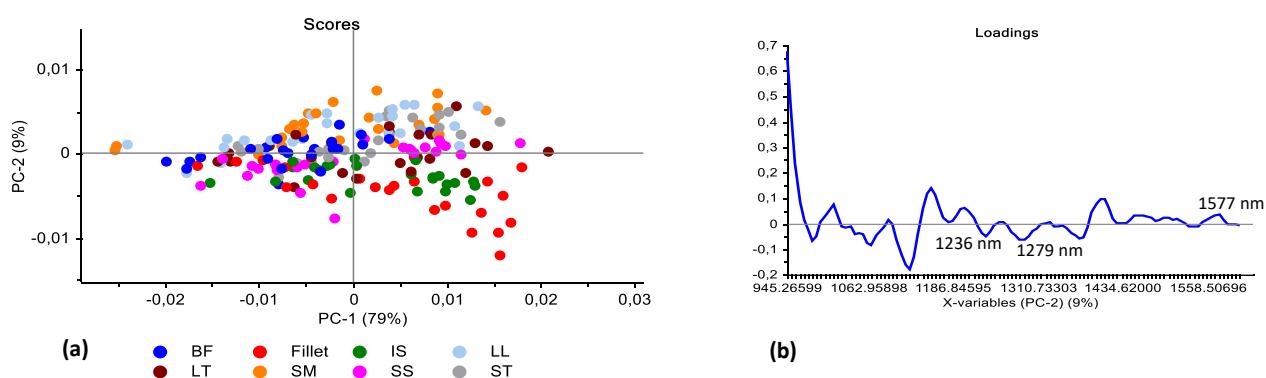


Figure C5 PCA analysis (SNV + SGd₂(9) pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a major overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (79%) vs. PC2 (9%). (b) PCA loadings line plot for PC2 with interpretable absorption bands at 1236, 1279 and 1577 nm.

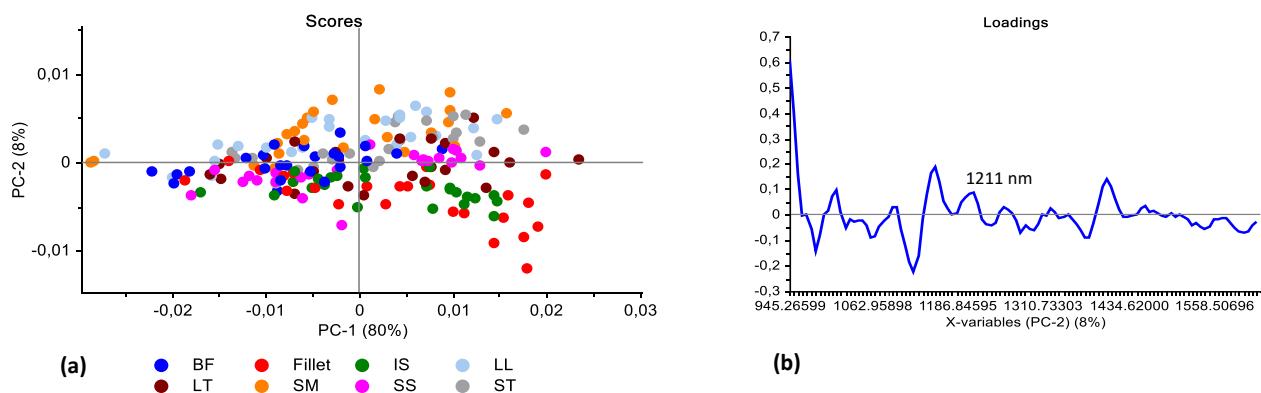


Figure C6 PCA analysis (SNV + detrend + SGd₂(7) pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a major overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (80%) vs. PC2 (8%). (b) PCA loadings line plot for PC2 with an interpretable absorption bands at 1211 nm.

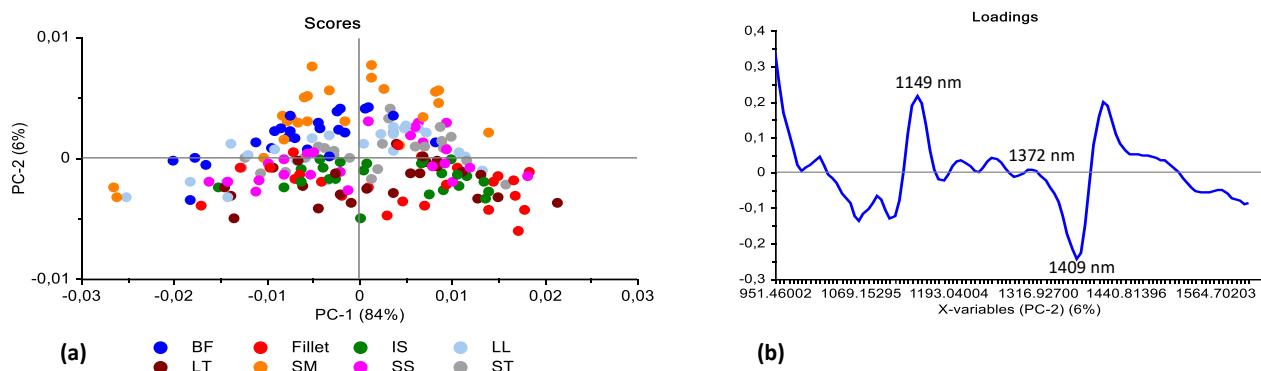


Figure C7 PCA analysis (SNV + detrend + SGd₂(9) pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a major overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (84%) vs. PC2 (6%). (b) PCA loadings line for PC2 with absorption bands at 1149, 1372 and 1409 nm.

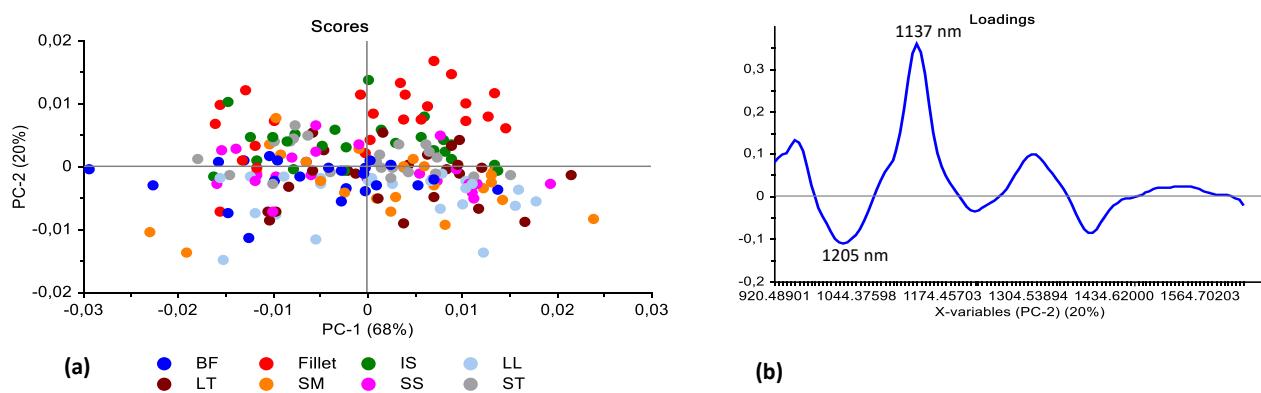


Figure C8 PCA analysis (SGd₁(5) pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a major overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (68%) vs. PC2 (20%). (b) PCA loadings line plot for PC2 with interpretable absorption bands at 1025 and 1137 nm.

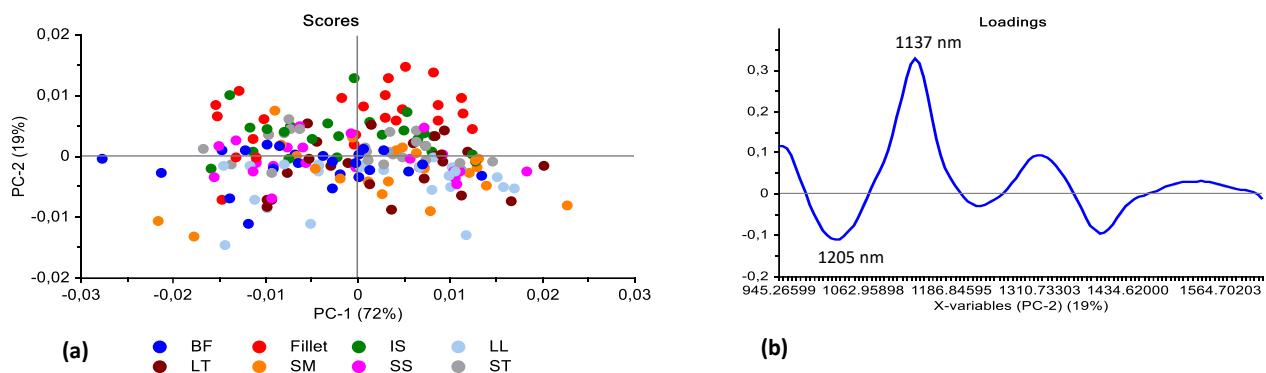


Figure C9 PCA analysis (SGd₁(7) pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a major overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (72%) vs. PC2 (19%). (b) PCA loadings line plot for PC2 with interpretable absorption bands at 1025 and 1137 nm.

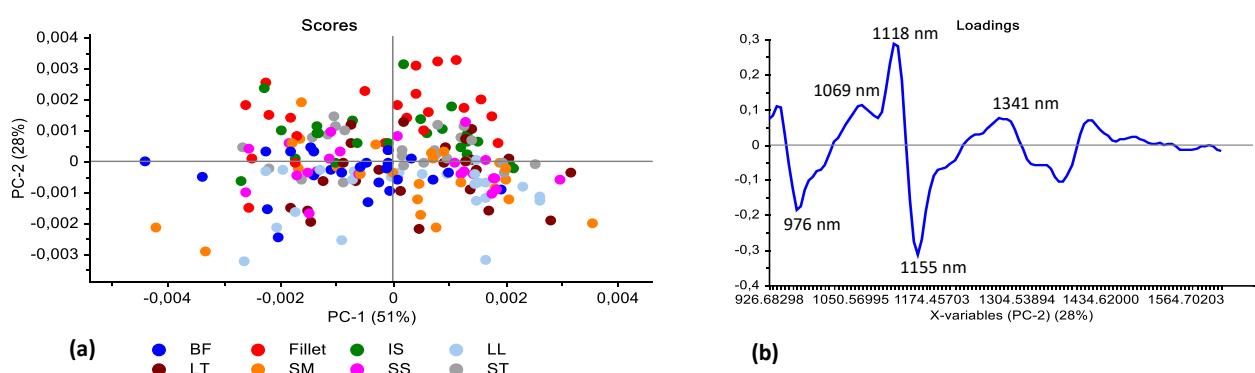


Figure C10 PCA analysis (SGd₂(7) pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a major overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (51%) vs. PC2 (28%). (b) PCA loadings line plot for PC2 with interpretable absorption bands at 976, 1069, 1118, 1155 and 1341 nm.

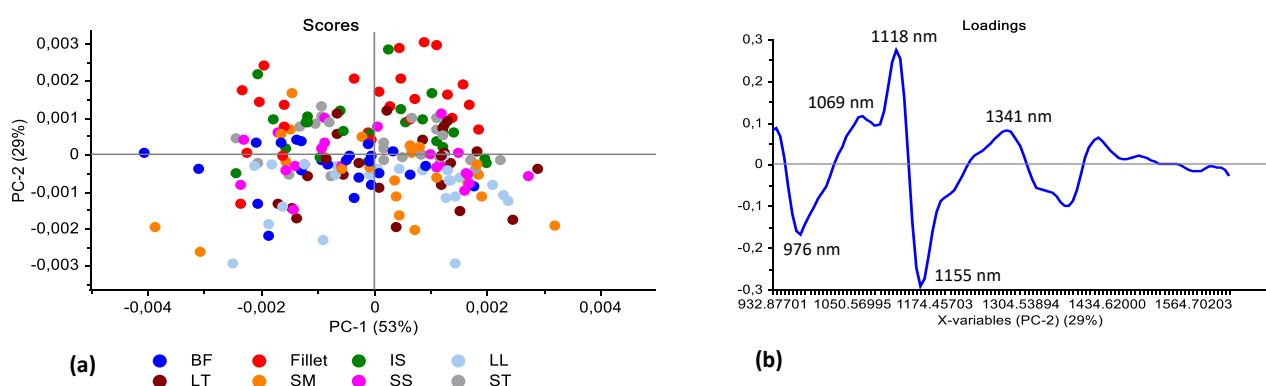


Figure C11 PCA analysis (SGd₂(9) pre-processed) of zebra (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a major overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (53%) vs. PC2 (29%). (b) PCA loadings line plot for PC2 with interpretable absorption bands at 976, 1069, 1118, 1155 and 1341 nm.

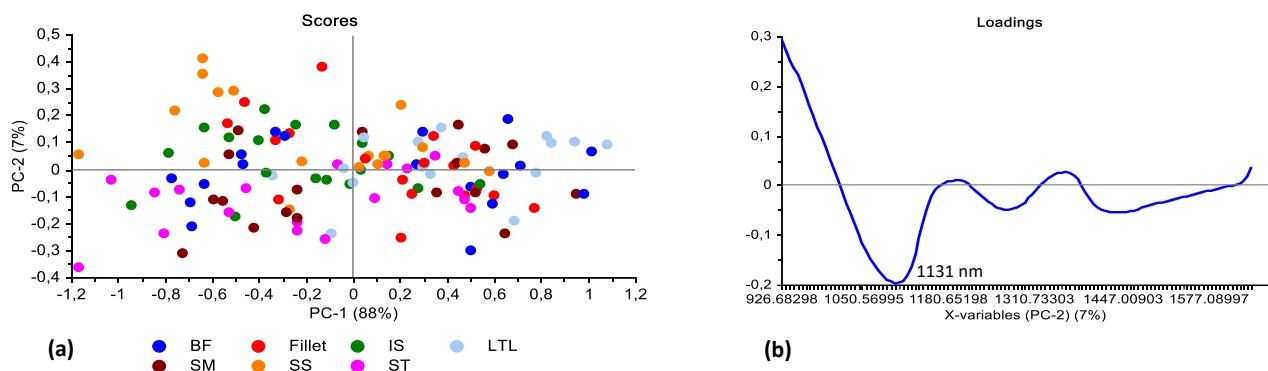


Figure C12 PCA analysis (SNV pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a major overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (88%) vs. PC2 (7%). (b) PCA loadings line plot for PC2 with an interpretable absorption band at 1131 nm.

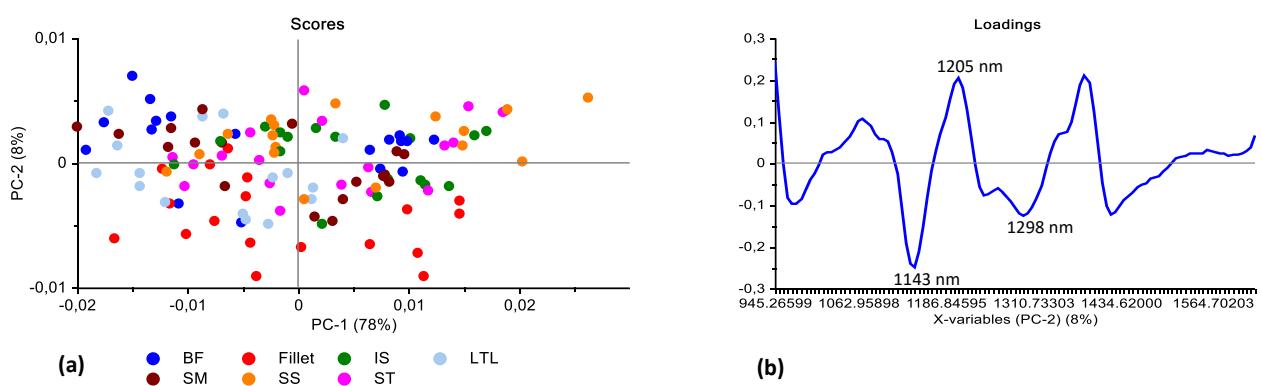


Figure C13 PCA analysis (SNV + SGd₂(7) pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a major overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (78%) vs. PC2 (8%). (b) PCA loadings line plot for PC2 with interpretable absorption bands at 1143, 1205 and 1298 nm.

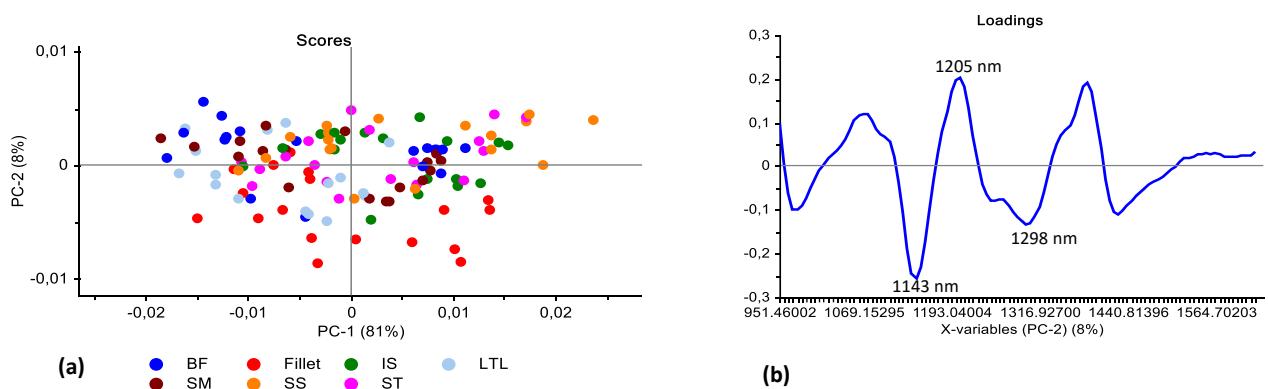


Figure C14 PCA analysis (SNV + SGd₂(9) pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a major overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (81%) vs. PC2 (8%). (b) PCA loadings line plot for PC2 with interpretable absorption bands at 1143, 1205 and 1298 nm.

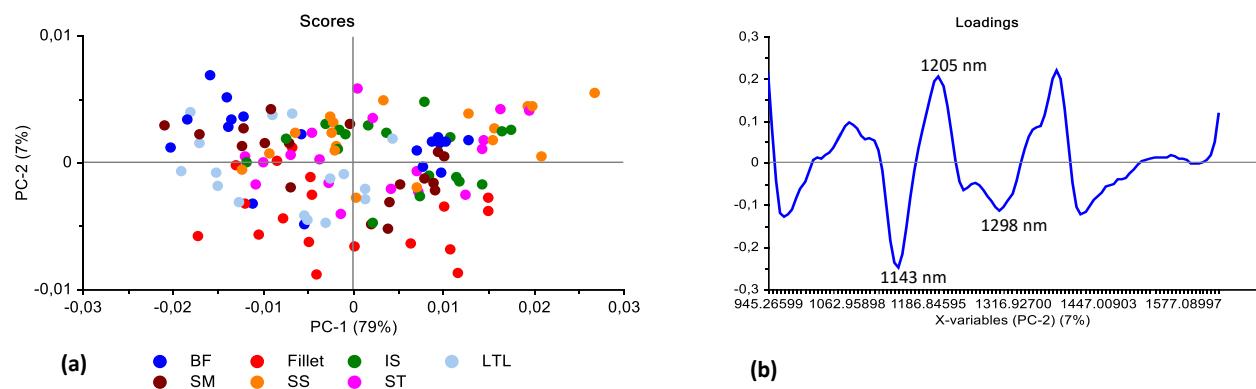


Figure C15 PCA analysis (SNV + detrend + SGd₂(7) pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a major overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (79%) vs. PC2 (7%). (b) PCA loadings line plot for PC2 with interpretable absorption bands at 1143, 1205 and 1298 nm.

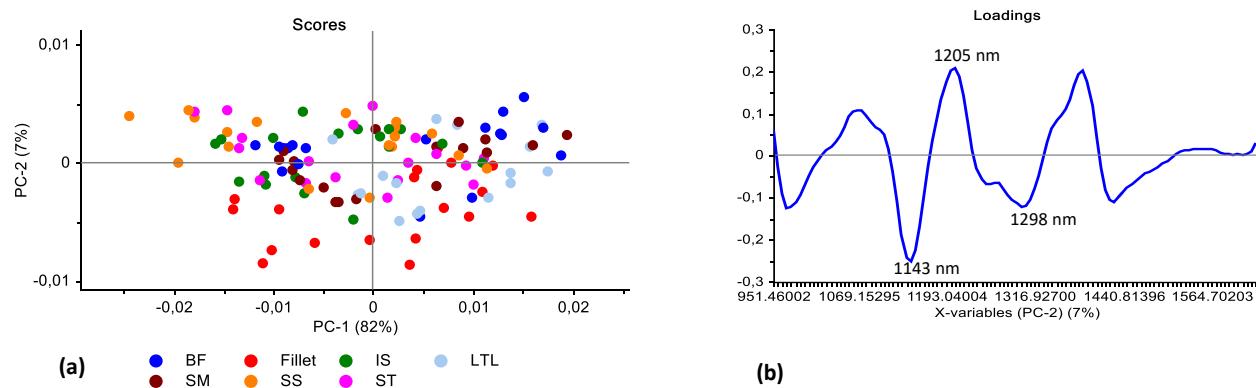


Figure C16 PCA analysis (SNV + detrend + SGd₂(9) pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a major overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (82%) vs. PC2 (7%). (b) PCA loadings line plot for PC2 with interpretable absorption bands at 1143, 1205 and 1298 nm.

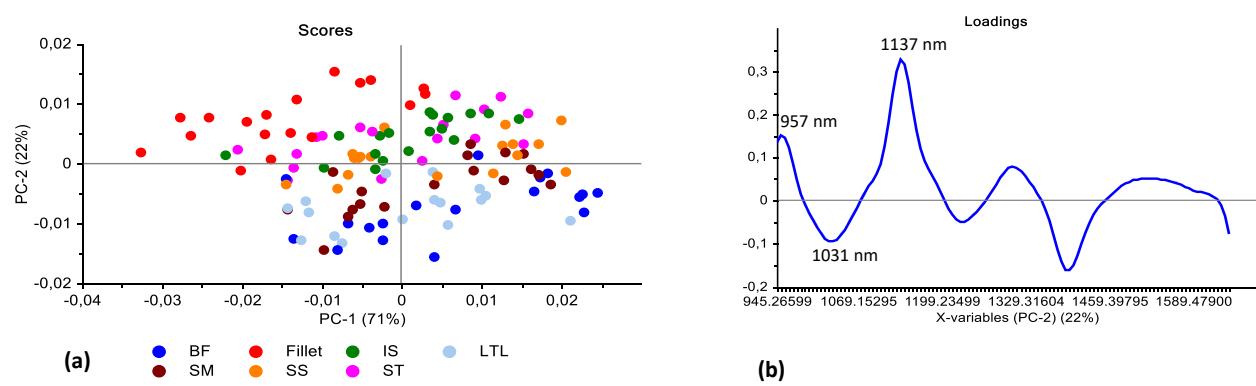


Figure C17 PCA analysis (SGd₁(5) pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a minimal separation between classes. Scores illustrated as (a) PCA score plot of PC1 (71%) vs. PC2 (22%). (b) PCA loadings line plot for PC2 with interpretable absorption bands at 957, 1031 and 1137 nm.

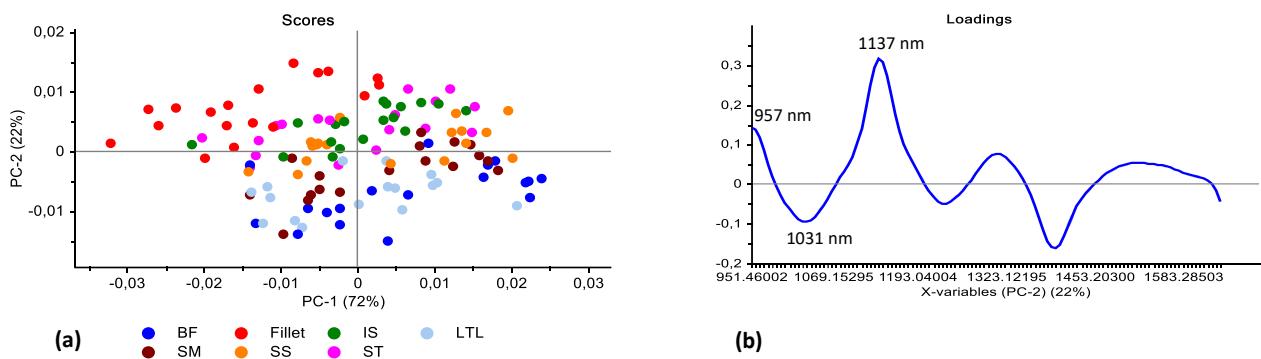


Figure C18 PCA analysis (SGd₁(7) pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a minimal separation between classes. Scores illustrated as (a) PCA score plot of PC1 (72%) vs. PC2 (22%). (b) PCA loadings line plot for PC2 with interpretable absorption bands at 957, 1031 and 1137 nm.

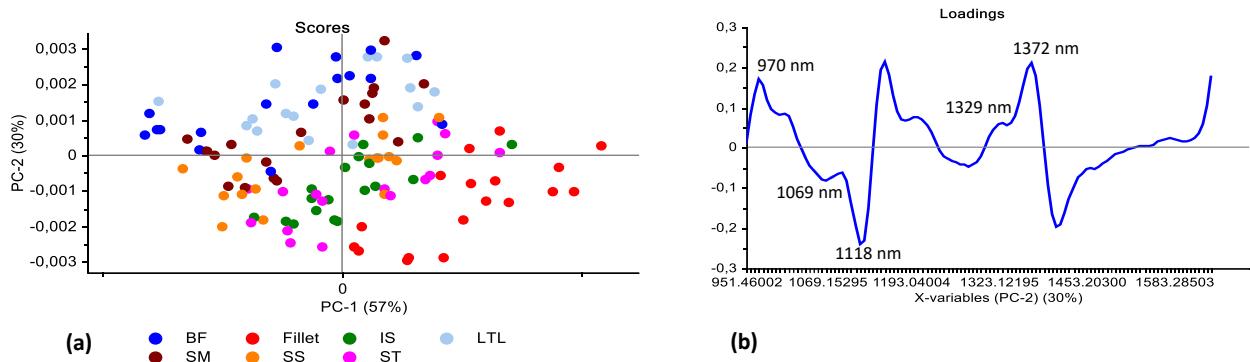


Figure C19 PCA analysis (SGd₂(7) pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a minimal separation between classes. Scores illustrated as (a) PCA score plot of PC1 (57%) vs. PC2 (30%). (b) PCA loadings line plot for PC2 with interpretable absorption bands at 970, 1069, 1118, 1329 and 1372 nm.

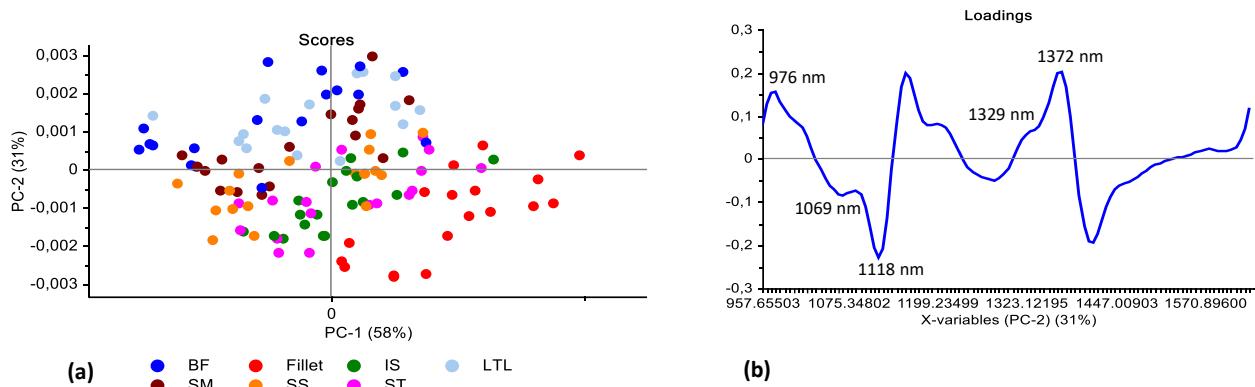


Figure C20 PCA analysis (SGd₂(9) pre-processed) of springbok (LTL, BF, SM, ST, IS, SS, fillet) [fresh and previously frozen samples] illustrating a minimal separation between classes. Scores illustrated as (a) PCA score plot of PC1 (58%) vs. PC2 (31%). (b) PCA loadings line plot for PC2 with interpretable absorption bands at 976, 1069, 1118, 1329 and 1372 nm.

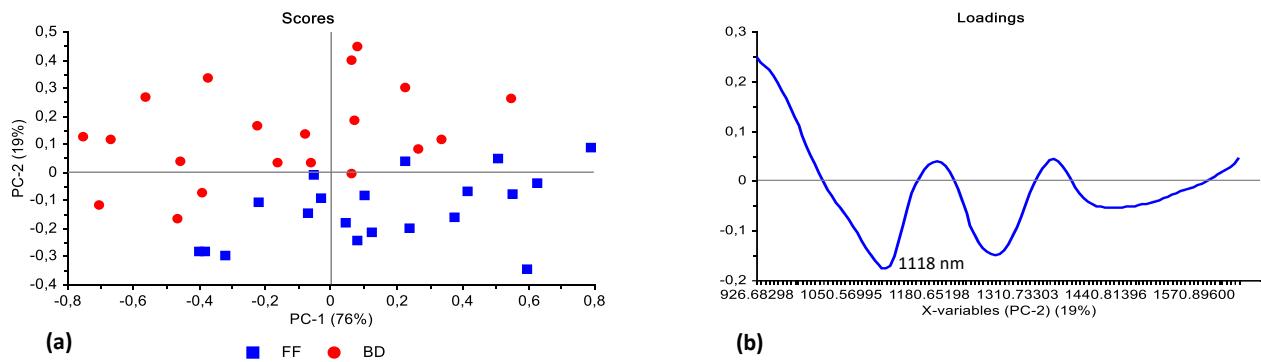


Figure C21 PCA analysis (SNV pre-processed) of ostrich (BD, FF) [fresh and previously frozen samples] illustrating a good separation between classes. Scores illustrated as (a) PCA score plot of PC1 (76%) vs. PC2 (19%). (b) PCA loadings line plot for PC2 with an interpretable band at 1118 nm.

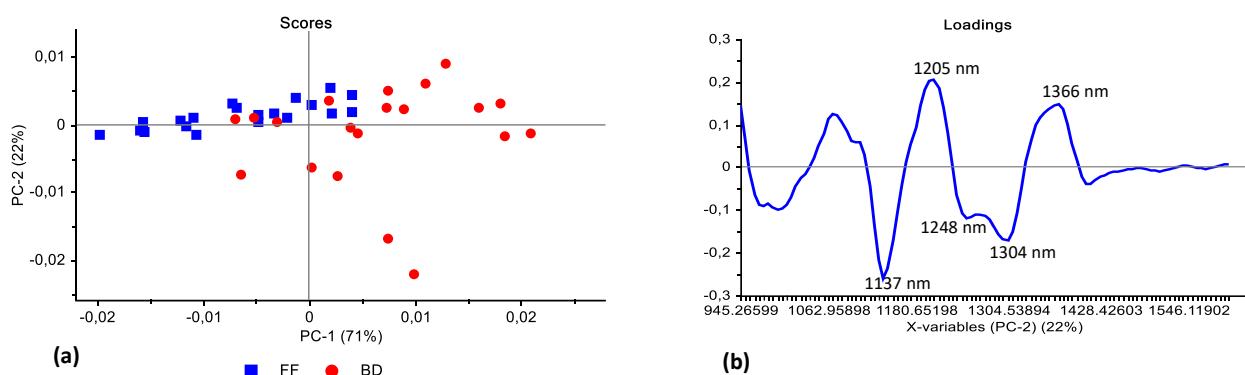


Figure C22 PCA analysis (SNV + SGd₂(7) pre-processed) of ostrich (BD, FF) [fresh and previously frozen samples] illustrating a good separation with a slight overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (71%) vs. PC2 (22%). (b) PCA loadings line plot for PC2 with interpretable bands at 1137, 1205, 1248, 1304 and 1366 nm.

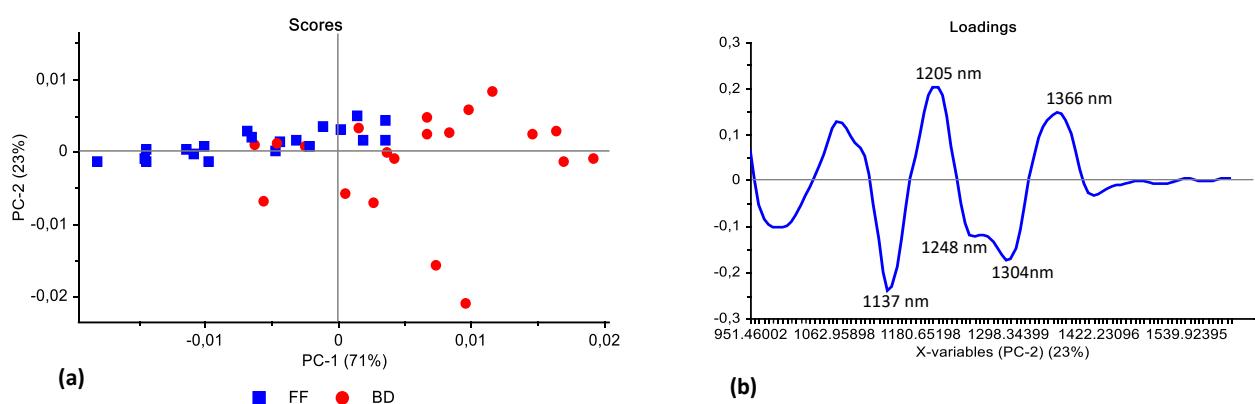


Figure C23 PCA analysis (SNV + SGd₂(9) pre-processed) of ostrich (BD, FF) [fresh and previously frozen samples] illustrating a good separation with a slight overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (71%) vs. PC2 (23%). (b) PCA loadings line plot for PC2 with interpretable bands at 1137, 1205, 1248, 1304 and 1366 nm.

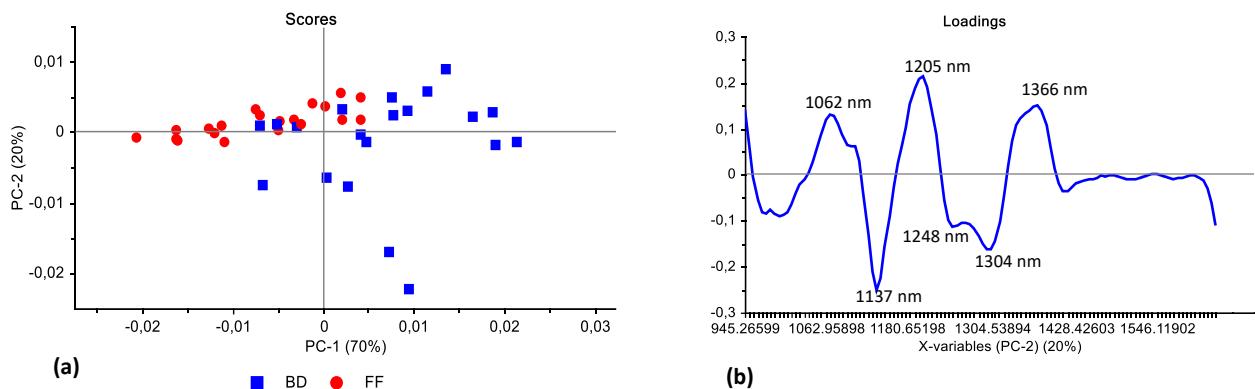


Figure C24 PCA analysis (SNV + detrend + SGd₂(7) pre-processed) of ostrich (BD, FF) [fresh and previously frozen samples] illustrating a good separation with a slight overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (70%) vs. PC2 (20%). (b) PCA loadings line plot for PC2 with interpretable bands at 1062, 1137, 1205, 1248, 1304 and 1366 nm.

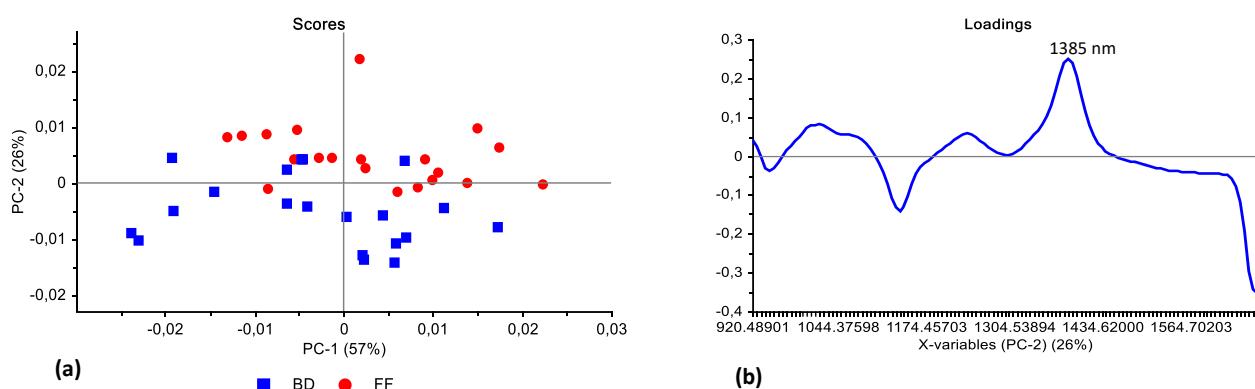


Figure C25 PCA analysis (SGd₁(5) pre-processed) of ostrich (BD, FF) [fresh and previously frozen samples] illustrating a good separation with a slight overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (57%) vs. PC2 (26%). (b) PCA loadings line plot for PC2 with an interpretable band at 1385 nm.

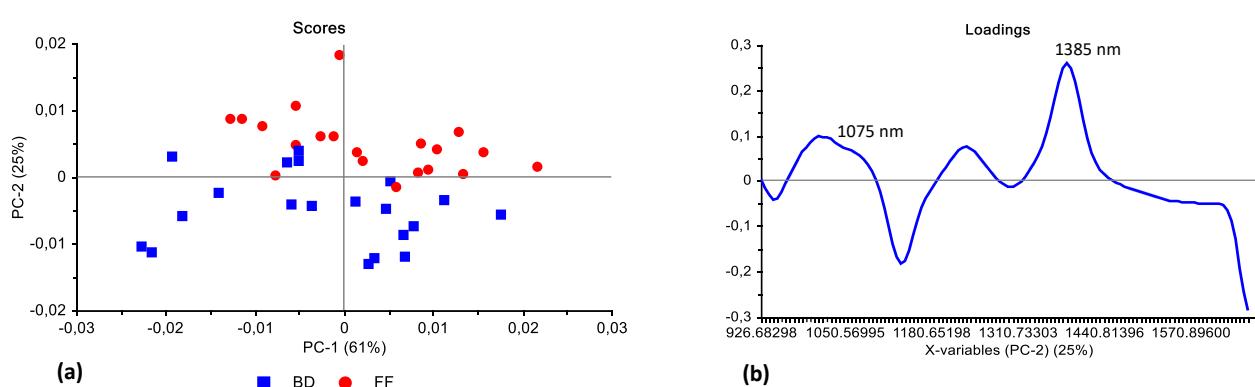


Figure C26 PCA analysis (SGd₁(7) pre-processed) of ostrich (BD, FF) [fresh and previously frozen samples] illustrating a good separation with a slight overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (61%) vs. PC2 (25%). (b) PCA loadings line plot for PC2 with interpretable bands at 1075 and 1385 nm.

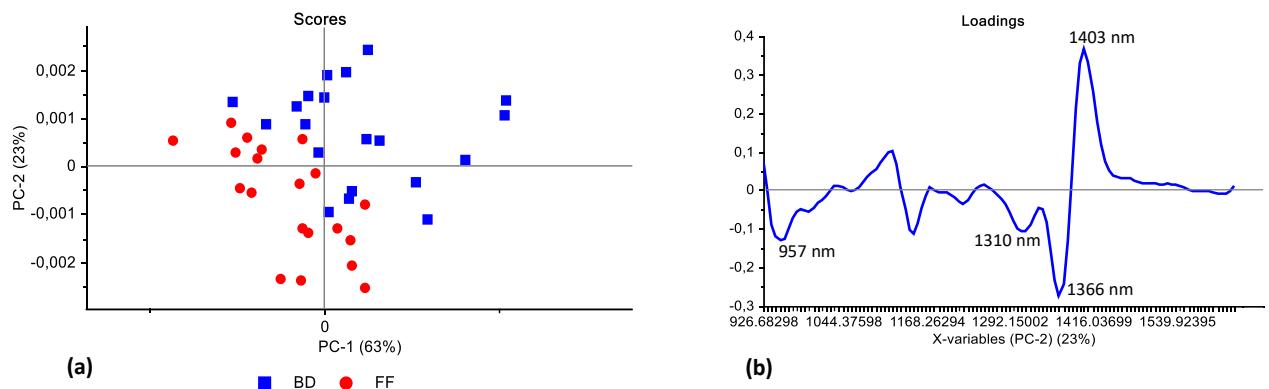


Figure C27 PCA analysis ($\text{SGd}_2(7)$ pre-processed) of ostrich (BD, FF) [fresh and previously frozen samples] illustrating a good separation with a slight overlap between classes. Scores illustrated as (a) PCA score plot of PC1 (63%) vs. PC2 (23%). (b) PCA loadings line plot for PC2 with interpretable bands at 957, 1310, 1366 and 1403 nm.

Addendum D

Supplementary information pertaining to Chapter 4:

4.4 Hierarchical model development

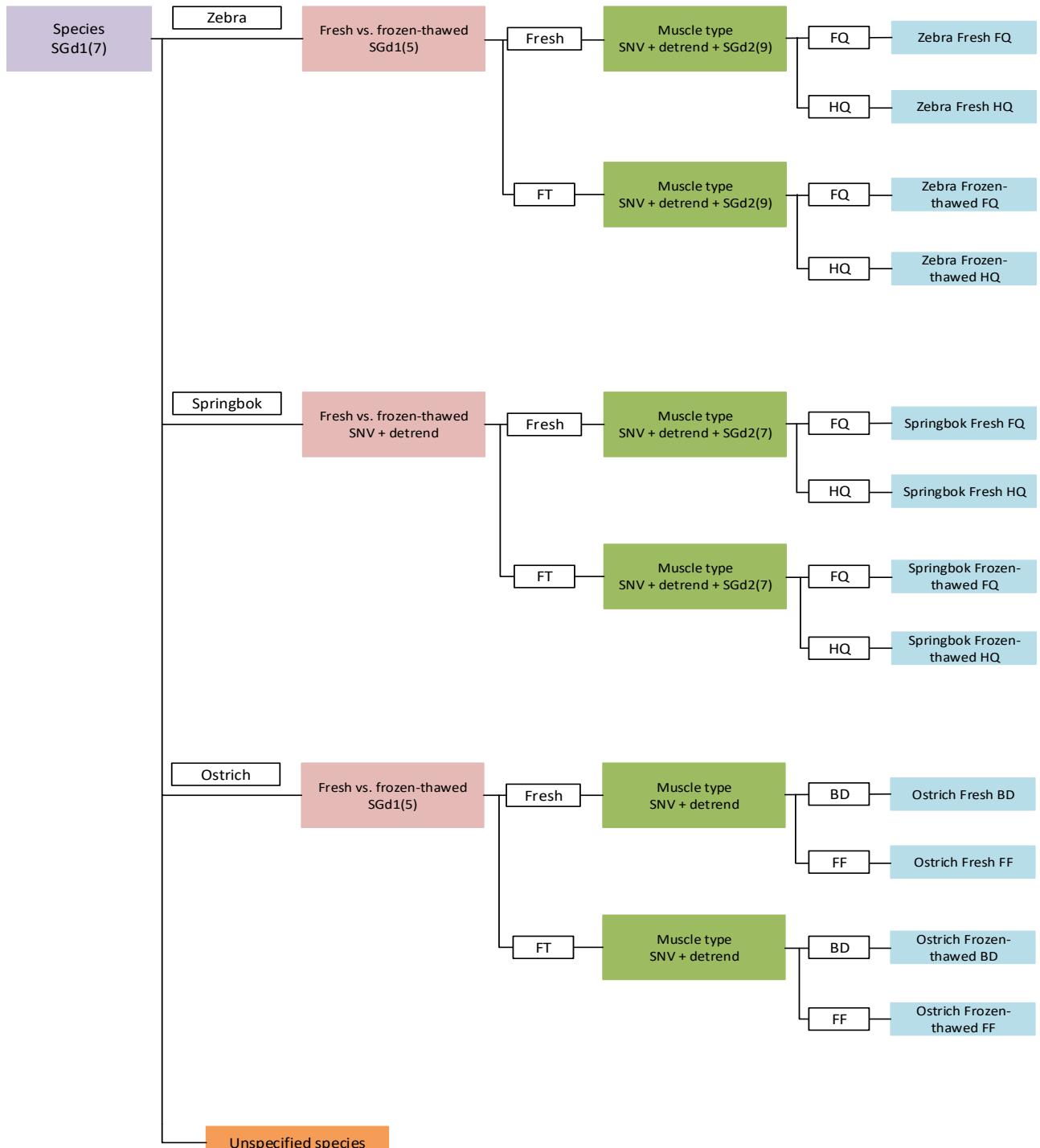
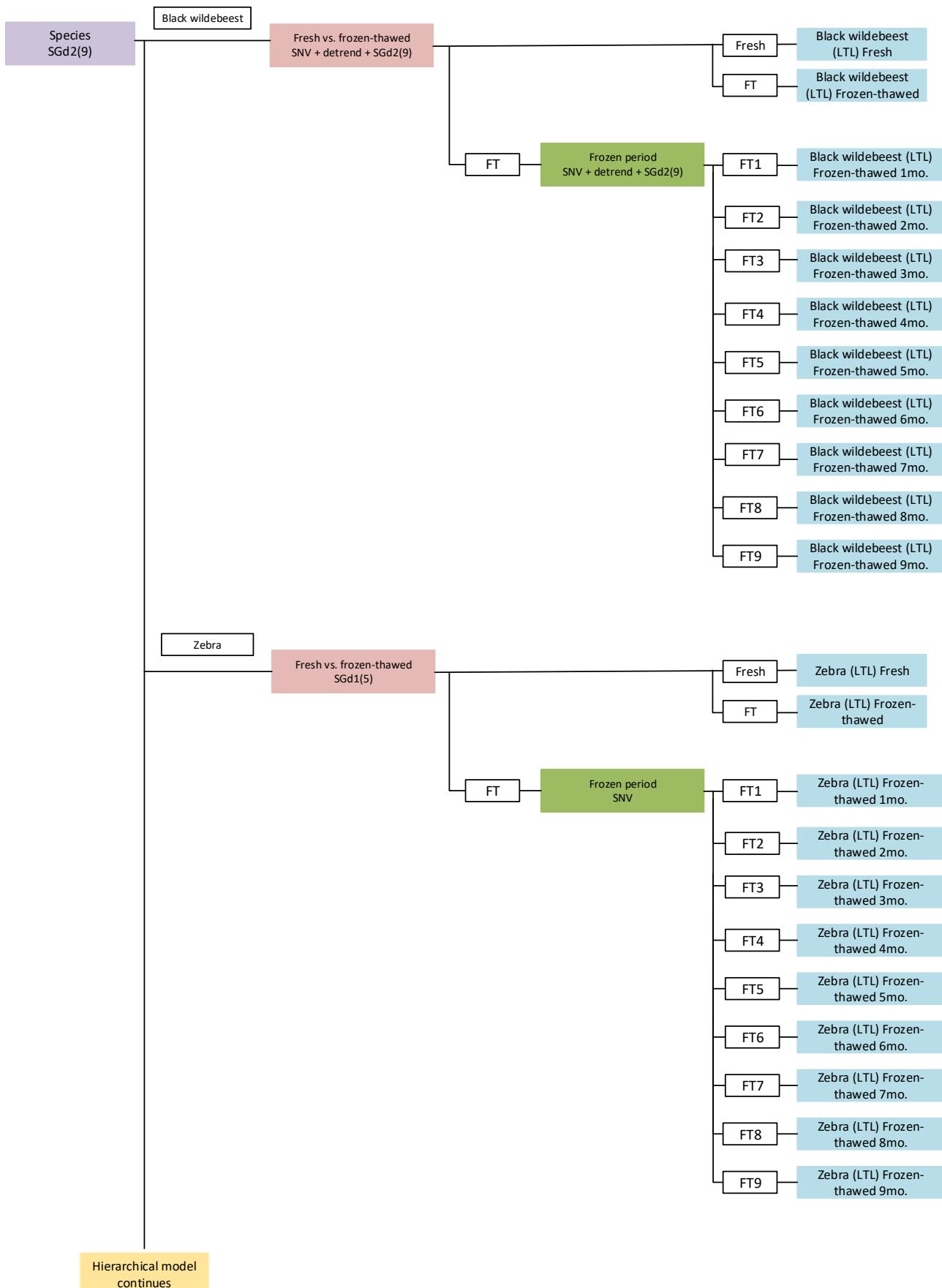


Figure D1 Schematic of the hierarchical model exhibiting the PLS-DA models and pre-processing used for differentiating between zebra, springbok and ostrich, fresh and previously meat as well as the muscle types.



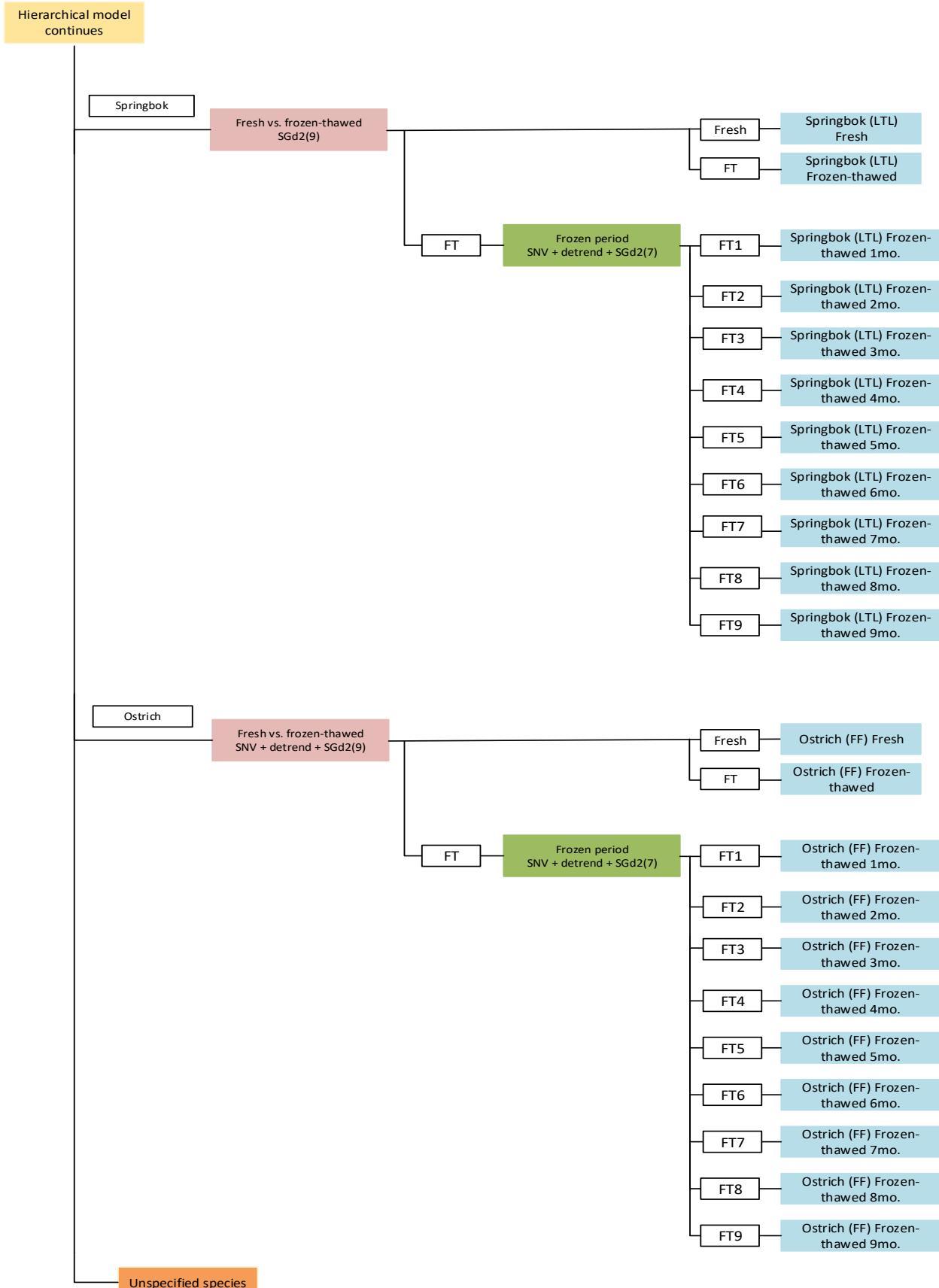


Figure D2 Schematic of the hierarchical model exhibiting the PLS-DA models and pre-processing used for differentiating between black wildebeest, zebra, springbok and ostrich, fresh and previously meat as well as the frozen period.

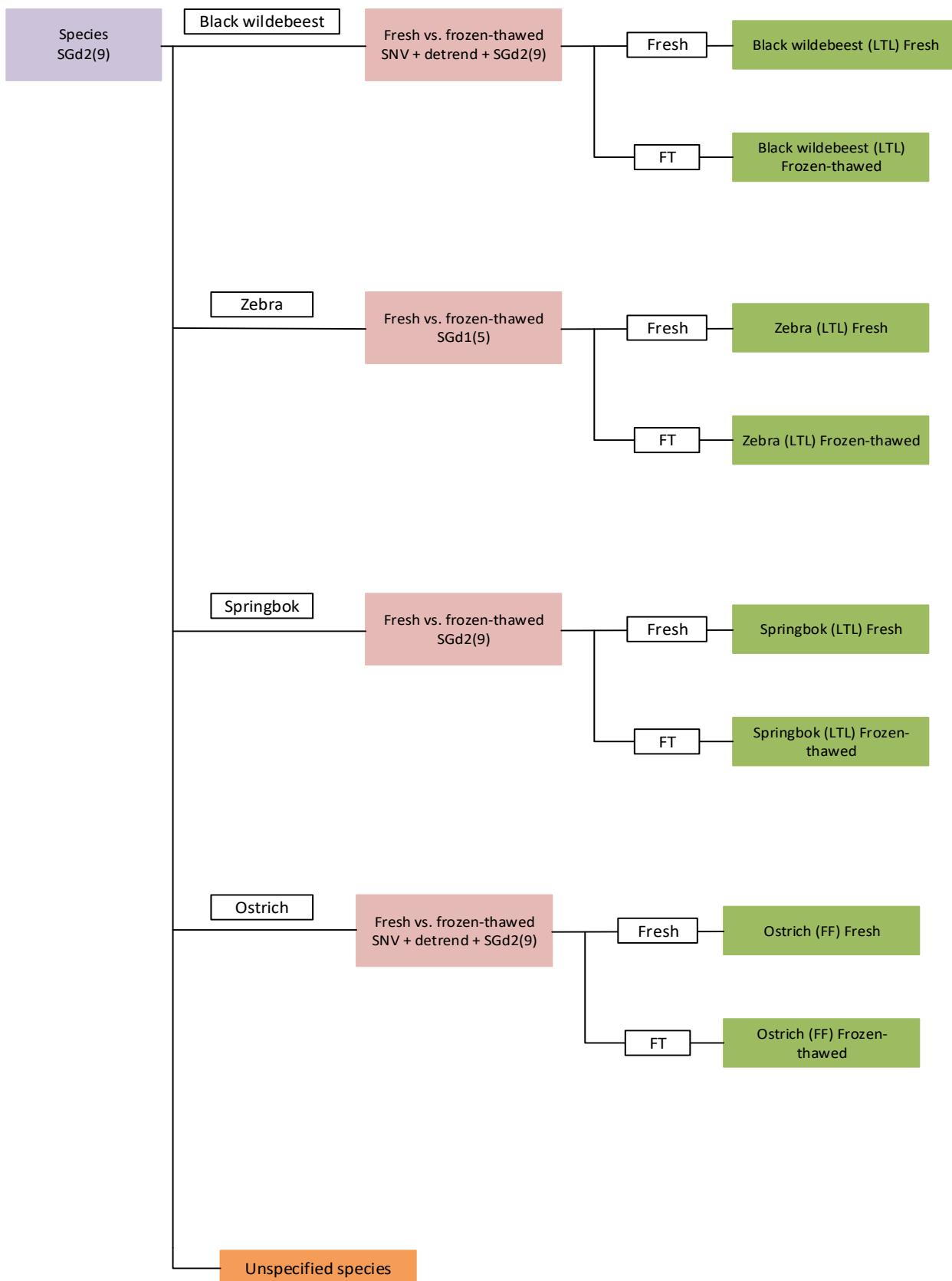


Figure D3 Schematic of the hierarchical model exhibiting the PLS-DA models and pre-processing used for differentiating between black wildebeest, zebra, springbok and ostrich as well as fresh and previously meat.