

Near Infrared Hyperspectral Imaging: A Rapid Method For The Differentiation Of Maize Ear Rot Pathogens On Growth Media

by
Cenette Bezuidenhout



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

Master of Science in Food Science



at

Stellenbosch University

Department of Food Science, Faculty of AgriSciences

Supervisor: Dr Paul J Williams

Co-supervisor: Dr Lindy J Rose

March 2018

DECLARATION

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

Date: March 2018

SUMMARY

Maize grain is highly susceptible to the toxin-producing fungal pathogens *Fusarium* spp. and *Stenocarpella* spp. Infection of grain with these species leads to maize ear rots but of greater concern is their ability to produce mycotoxins, which can promote cancer, among other diseases, in humans and animals. By combining microbiology, plant pathology and chemistry disciplines to create rapid screening methods that can accurately distinguish fungal pathogens, an initial step in ensuring food safety can be achieved, which can potentially be applied to maize grain in future. This thesis aimed to distinguish between the most important maize ear rot pathogens namely *Fusarium verticillioides*, *F. graminearum* species complex (FGSC) and *Stenocarpella maydis* that cause Fusarium-, Gibberella- and Diplodia ear rot, respectively. Furthermore, pathogen isolates from the same species were also distinguished. This was done with near infrared (NIR) hyperspectral imaging, a technology that offers the ability for rapid sample measurement that provides data containing both spatial and spectral information. Through multivariate analysis such as principal component analysis, primarily used for data exploration, and partial least square discriminant analysis, hyperspectral images was used to build classification models that accurately distinguish between the maize ear rot pathogens. All the major ear rot pathogens were distinguished with reasonable accuracy on day 3 of growth, with the isolates from the same species showing higher accuracy on day 5 of growth.

OPSOMMING

Mieliegraan is hoogs vatbaar vir die toksien-produiserende swam-patogene naamlik *Fusarium* spp. en *Stenocarpella* spp. Infeksie van graan met hierdie spesies lei tot mielie-kopvrot, en wat groter kommer wek, is dat hulle oor die vermoë beskik om mikotoksiene te produseer, wat siektes soos kanker by mense en diere bevorder. Deur die kombinasie van mikrobiologie, plantpatologie en chemie dissiplines om vinnige siftingsmetodes te skep wat die swam-patogene akkuraat kan onderskei, kan 'n aanvanklike stap in die versekering van voedselveiligheid bereik word, wat moontlik in die toekoms op mieliegraan toegepas kan word. Hierdie proefskrif het ten doel om te onderskei tussen die belangrikste mielie-kopvrot-patogene, naamlik *Fusarium verticillioides*, *Fusarium graminearum* spesie kompleks (FGSK) en *Stenocarpella maydis* wat onderskeidelik Fusarium-, Gibberella- en Diplodia-kopvrot veroorsaak. Verder is patogeen-isolate van dieselfde spesie ook onderskei. Dit is gedoen met naby infrarooi (NIR) hiperspektrale beelding, 'n tegnologie wat die vermoë bied vir vinnige monstermeting wat data bevat wat beide ruimtelike en spektrale inligting bevat. Deur middel van meerveranderlike-beeldontleding soos hoofkomponent-analise, hoofsaaklik vir dataverkenning, en pasiële kleinste waarde diskriminant-analise, is hiperspektrale beelde gebruik om klassifikasie modelle te bou wat akkuraat tussen die mielie-kopvrot-patogene onderskei. Al die hoof kopvrot-patogene is met redelike akkuraatheid op dag 3 van groei onderskei, met die isolate van dieselfde spesie wat hoër akkuraatheid op dag 5 van groei toon.

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Dr Paul Williams and Dr Lindy Rose, for the tremendous support of my research. Their guidance, motivation, knowledge and understanding 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 Alvaro Viljoen and Dr Ilze Vermaak (Tshwane University of Technology, Pretoria) for the use of their Sisuchema hyperspectral imaging system and laboratory, as well as their advice while introducing me to the instrument;

Department of Plant Pathology, Stellenbosch University for supplying fungal cultures and assisting with culture preparation;

All the staff and post-graduate students in the Department of Food Science, Stellenbosch University for providing a friendly working environment and support when needed. A special thanks to Prof. Sigge and Prof. Gouws for providing guidance and support on my extra activities (S'coolBeans); and

The National Research Foundation (NRF) [Thutuka Grant Holder Bursary, 2016 and 2017] 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).

I would also like to thank Elda Theron, Johané Botes, Shannon Howell, Elisma Ackermann and all my friends and family for always listening to me, supporting me and providing me with endless laughter and memories;

Thank you to my family, without whom I would not have been able to complete this study. My parents, Rudolf and Annemarie and my sister Danielle have been my primary support structure throughout this journey and their love, encouragement and support is undying.

Lastly, to my better half Leon Boshoff, thank you for always encouraging me to aim higher and providing continuous love and support while I chase my dreams.

CONTENTS

DECLARATION	i
ACKNOWLEDGEMENTS	ii
SUMMARY	iii
OPSOMMING	iv
LIST OF FIGURES	vi
LIST OF TABLES	xvii
LIST OF ABBREVIATIONS	xviii
CHAPTER 1 Introduction.....	1
CHAPTER 2 The potential application of NIR hyperspectral imaging in the maize industry: A review	8
CHAPTER 3 The differentiation of maize ear rot pathogens, on growth media, using near infrared hyperspectral imaging.....	41
CHAPTER 4 The influence of colony age on the accuracy of ear rot pathogen differentiation, on growth media, using NIR hyperspectral imaging.....	73
CHAPTER 5 General Discussion and Conclusion.....	99
ADDENDUM A Supplementary information pertaining to Chapter 3	104
ADDENDUM B Supplementary information pertaining to Chapter 4	107

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 is an individual entity and some repetition between chapters has, therefore, been unavoidable.

LIST OF FIGURES

Figure 2.1 Fusarium ear rot (FER) illustrating the random kernel infection on a maize ear (Pannar, 2017).....	10
Figure 2.2 Gibberella ear rot (GER) illustrating the pink mould formation on a maize ear (Pannar, 2017).....	11
Figure 2.3 Diplodia ear rot (DER) illustrating the white mould formation in between the kernels on a maize ear (Pannar, 2017).....	12
Figure 2.4 (a) An NIR hyperspectral image is a data cube that consists of two spatial (x and y) and one spectral (λ) dimension, where each pixel in the image range can be depicted as (b) spectra in the NIR region of the electromagnetic spectrum. (c) These can then be depicted as the mean spectra per object in the image.....	16
Figure 2.5 NIR hyperspectral processing steps which includes (a) multivariate image acquisition, (b) pre-processing, (c) image analysis and (d) prediction [adapted from Dorrepaal <i>et al.</i> (2016)].....	16
Figure 2.6 NIR hyperspectral imaging instrumentation consisting of a light source, filter, camera, detector and computer, which can be positioned in either (a) reflectance or (b) transmittance mode.	17
Figure 2.7 Three different scanning modes (a) point scan or whiskbroom, (b) line scan or pushbroom and (c) plane scan or stare down.....	19
Figure 2.8 (a) The data cube that consists of two spatial (x and y) and one spectral (λ) dimension is unfolded by multiplying the two spatial locations to yield (b) a two-dimensional dataset that can be further processed.	20
Figure 3.1 Schematic illustrating the inoculation of PDA glass plates with the pathogen isolates.....	46
Figure 3.2 Digital images of fungal isolates of <i>F. verticillioides</i> , <i>F. graminearum</i> s.s., <i>F. boothii</i> and <i>S. maydis</i> isolates, after 5 days, used in the study.....	46
Figure 3.3 Step-wise approach used for the pixel wise analysis of the various models illustrating (a) the RGB image of the mosaic of <i>F. verticillioides</i> , <i>F. graminearum</i> s.s. and <i>S. maydis</i> triplicates, (b) the uncleaned PC score image before background removal, (c) the PC score image of the isolates and (d) the PLS-DA score image illustrating the calibration and validation sets.....	48

Figure 3.4 Step-wise analysis used for the object wise analysis of the various models illustrating (a) the RGB image of the mosaic of *F. verticillioides*, *F. graminearum* s.s. and *S. maydis* triplicates, (b) the uncleaned PC score image before background removal, (c) the PC score image before the subdivision, (d) the PC score image of the subdivided isolates and (e) the PLS-DA score image illustrating the calibration and validation sets.48

Figure 3.5 Mean spectra of the pathogen isolates, depicting *Fusarium verticillioides* MRC826 (dark blue), *Fusarium verticillioides* MRC8267 (teal), *Fusarium verticillioides* MRC8559 (pink), *F. boothii* M0010 (purple), *F. boothii* M0010 (light blue), *F. boothii* M0010 (yellow), *F. graminearum* s.s. M14-55 (green) and *Stenocarpella maydis* SM8 (orange), used in the four different mosaics.52

Figure 3.6 Pixel wise PCA of mosaic one illustrating good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (80.2%) vs. PC2 (14.7%) and (b) PCA score plot of PC2 (14.7%) vs PC3 (1%) shaded per the density of the points. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. graminearum* s.s. M14-55 (middle) and *S. maydis* SM8 (right). (d) PCA score plot of PC1 vs. PC2 illustrating the separation between the *F. verticillioides* MRC826 (blue), *F. graminearum* s.s. M14-55 (green) and *S. maydis* SM8 (orange) isolates.54

Figure 3.7 Object wise PCA of mosaic one illustrating good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (82.2%) vs. PC2 (16%) and (b) PCA score plot of PC2 (16%) vs PC3 (0.896%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. graminearum* s.s.M14-55 (middle) and *S. maydis* SM8 (right).....54

Figure 3.8 PCA loadings plots of mosaic one for both the pixel and object wise analysis is given for (a) PC1 with bands at 1260, 1400 and 1930 nm, (b) PC2 with bands at 1430 and 1930 nm and (c) PC3 with a band at 1200 nm.....55

Figure 3.9 Pixel wise PCA of mosaic two illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (58.5%) vs. PC2 (26.8%) and (b) PCA score plot of PC2 (26.8%) vs PC3 (3.5%) shaded per the density of the points. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. boothii* M0002 (middle) and *S. maydis* SM8 (right) is also given. PCA score plot of PC1 vs. PC2 (d) colour coded per pathogen isolate as seen in the accompanying legend is also given.56

- Figure 3.10** Object wise PCA of mosaic two illustrates separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (67.4%) vs. PC2 (28.1%) and (b) PCA score plot of PC2 (28.1%) vs PC3 (3.16%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. boothii* M14-55 (middle) and *S. maydis* SM8 (right).....57
- Figure 3.11** PCA loadings plots of mosaic two for both the pixel and object wise analysis is given for (a) PC1 with bands at 1260, 1400 and 1930 nm, (b) PC2 with bands at 1450 and 1930 nm and (c) PC3 with a band at 1200 and 1880 nm.....58
- Figure 3.12** Object wise PCA of mosaic three illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (83.2%) vs. PC2 (10.8%) and (b) PCA score plot of PC2 (10.8%) vs PC3 (4.72%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. verticillioides* MRC8267 (middle) and *F. verticillioides* MRC8559 (right).....59
- Figure 3.13** PCA loadings plots of mosaic three is given for (a) PC1 with bands at 1260, 1400 and 1930 nm, (b) PC2 with bands at 1430 and 1880 nm and (c) PC3 with a few low intensity bands at 1380, 1400 and 1930 nm.59
- Figure 3.14** Object wise PCA of mosaic four illustrates the separation between *F. boothii* isolates. Scores illustrated by (a) PCA score plot of PC1 (92.4%) vs. PC2 (5.39%) and (b) PCA score plot of PC2 (5.39%) vs PC3 (1.43%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. boothii* M0002 (left), *F. boothii* M0010 (middle) and *F. boothii* M0100 (right).....60
- Figure 3.15** PCA loadings plots of mosaic four is given for (a) PC1 with bands at 1260, 1400 and 1930 nm, (b) PC2 with bands at 1430 and 1880 nm and (c) PC3 with a band at 1400 nm.....60
- Figure 3.16** Pixel wise PLS-DA of mosaic one illustrating the (a) score plot of PLS factor 1 (80.2%) vs. factor 2 (14.7%) colour coded per pathogen isolate, the (b) score plot of factor 2 (14.7%) vs factor 3 (0.777%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. graminearum* s.s. M14-55 (green) and *S. maydis* SM8 (orange), illustrating the predicted pixels, with the pathogen categories provided in the legend, where the first row was used as validation and the second and third as calibration.....62

Figure 3.17 Object wise PLS-DA of mosaic one illustrating the (a) score plot of PLS factor 1 (82.2%) vs factor 2 (16%) colour coded per pathogen isolate, the (b) score plot of factor 2 (16%) vs factor 3 (0.878%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. graminearum* s.s. M14-55 (green) and *S. maydis* SM8 (orange), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used as validation and the second and third as calibration.....63

Figure 3.18 Pixel wise PLS-DA of mosaic two illustrating the (a) score plot of PLS factor 1 (55.3%) vs factor 2 (30%) colour coded per pathogen isolate, the (b) score plot of factor 2 (30%) vs factor 3 (3.41%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. boothii* M0002 (purple) and *S. maydis* SM8 (orange), illustrating the predicted pixels with the pathogen categories provided in the legend, where the first row was used as validation and the second and third as calibration.64

Figure 3.19 Object wise PLS-DA of mosaic two illustrating the (a) score plot of PLS factor 1 (65.6%) vs factor 2 (29.6%) colour coded per pathogen isolate, the (b) score plot of factor 2 (29.6%) vs factor 3 (3.23%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. boothii* M0002 (purple) and *S. maydis* SM8 (orange), illustrating the predicted objects with the pathogen categories provided in the legend, where the first row was used as validation and the second and third as calibration.65

Figure 3.20 Object wise PLS-DA of mosaic three illustrating the (a) score plot of PLS factor 1 (82.9%) vs factor 2 (8.71%) colour coded per pathogen isolate, the (b) score plot of factor 2 (8.71%) vs factor 3 (7.03%) colour coded per pathogen isolate and the (c) classification image of *Fusarium verticillioides* MRC826 (blue), *F. verticillioides* MRC8267 (teal) and *F. verticillioides* MRC8559 (pink), illustrating the predicted objects with the pathogen categories provided in the legend, where the first row was used as validation and the second and third as calibration.....65

Figure 3.21 Object wise PLS-DA of mosaic four illustrating the (a) score plot of PLS factor 1 (92.4%) vs factor 2 (4.23%) colour coded per pathogen isolate, the (b) score plot of factor 2 (4.23%) vs factor 3 (2.57%) colour coded per pathogen isolate and the (c) classification image of *Fusarium boothii* M0002 (purple), *F. boothii* M0010 (light blue) and *F. boothii* M0100 (yellow), illustrating the predicted objects with the pathogen categories provided in the legend, where the first row was used as validation and the second and third as calibration.66

- Figure 4.1** Step-wise approach used for the analysis of the various models. (a) RGB image of the mosaic, illustrating the triplicates of three pathogen isolates, (b) the uncleaned PC score image before the subdivision, (c) the PC score image of the subdivided isolates and (d) the PLS-DA score image illustrating the calibration and validation sets. 78
- Figure 4.2** Mean spectra of the pathogen isolates on day 3 of growth, depicting *Fusarium verticillioides* MRC826 (blue), *F. verticillioides* MRC8267 (teal), *F. verticillioides* MRC8559 (pink), *F. boothii* M0010 (purple), *F. boothii* M0010 (light blue), *F. boothii* M0010 (yellow), *F. graminearum* s.s. M14-55 (green) and *Stenocarpella maydis* SM8 (orange), used in the different mosaics..... 81
- Figure 4.3** Mean spectra of the pathogen isolates on day 5 of growth, depicting *Fusarium verticillioides* MRC826 (blue), *F. verticillioides* MRC8267 (teal), *F. verticillioides* MRC8559 (pink), *F. boothii* M0010 (purple), *F. boothii* M0010 (light blue), *F. boothii* M0010 (yellow), *F. graminearum* s.s. M14-55 (green) and *Stenocarpella maydis* SM8 (orange), used in the different mosaics..... 82
- Figure 4.4** Mean spectra of the pathogen isolates on day 7 of growth, depicting *Fusarium verticillioides* MRC826 (blue), *F. verticillioides* MRC8267 (teal), *F. verticillioides* MRC8559 (pink), *F. boothii* M0010 (purple), *F. boothii* M0010 (light blue), *F. boothii* M0010 (yellow), *F. graminearum* s.s. M14-55 (green) and *Stenocarpella maydis* SM8 (orange), used in the different mosaics..... 83
- Figure 4.5** Object wise PCA of mosaic one (Day 3) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (94.6%) vs. PC2 (4.32%) and (b) PCA score plot of PC2 (4.32%) vs PC3 (0.553%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. graminearum* s.s. M14-55 (middle) and *S. maydis* SM8 (right). The loading plots is given for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1930 nm..... 84
- Figure 4.6** PLS-DA of mosaic one (Day 3) illustrating the (a) score plot of factor 1 (94.6%) vs factor 2 (4.31%) colour coded per pathogen isolate, the (b) score plot of factor 2 (4.31%) vs factor 3 (0.433%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. graminearum* s.s. M14-55 (green) and *S. maydis* SM8 (orange), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration. 87

Figure 4.7 PLS-DA of mosaic two (Day 5) illustrating the (a) score plot of factor 1 (82.2%) vs factor 2 (16%) colour coded per pathogen isolate, the (b) score plot of factor 2 (16%) vs factor 3 (0.878%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. graminearum* s.s. M14-55 (green) and *S. maydis* SM8 (orange), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.87

Figure 4.8 PLS-DA of mosaic three (Day 7) illustrating the (a) score plot of factor 1 (69.4%) vs factor 2 (29.8%) colour coded per pathogen isolate, the (b) score plot of factor 2 (29.8%) vs factor 3 (0.387%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. graminearum* s.s. M14-55 (green) and *S. maydis* SM8 (orange), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.88

Figure 4.9 PLS-DA of mosaic four (Day 3) illustrating the (a) score plot of factor 1 (82.2%) vs factor 2 (12.8%) colour coded per pathogen isolate, the (b) score plot of factor 2 (12.8%) vs factor 3 (3.12%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. boothii* M0002 (purple) and *S. maydis* SM8 (orange), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.89

Figure 4.10 PLS-DA of mosaic five (Day 5) illustrating the (a) score plot of factor 1 (65.6%) vs factor 2 (29.6%) colour coded per pathogen isolate, the (b) score plot of factor 2 (29.6%) vs factor 3 (3.23%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. boothii* M0002 (purple) and *S. maydis* SM8 (orange), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.89

Figure 4.11 PLS-DA of mosaic six (Day 7) illustrating the (a) score plot of factor 1 (53.9%) vs factor 2 (43.8%) colour coded per pathogen isolate, the (b) score plot of factor 2 (43.8%) vs factor 3 (1.19%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. boothii* M0002 (purple) and *S. maydis* SM8 (orange), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.89

Figure 4.12 PLS-DA of mosaic seven (Day 3) illustrating the (a) score plot of factor 1 (82.3%) vs factor 2 (14.4%) colour coded per pathogen isolate, the (b) score plot of factor 2 (14.4%) vs factor 3 (1.07%) colour coded per pathogen isolate and the (c) classification image of *F. boothii* M0002 (purple), *F. boothii* M0010 (light blue) and *F. boothii* M0100 (yellow), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.90

Figure 4.13 PLS-DA of mosaic eight (Day 5) illustrating the (a) score plot of factor 1 (92.4%) vs factor 2 (4.23%) colour coded per pathogen isolate, the (b) score plot of factor 2 (4.23%) vs factor 3 (2.57%) colour coded per pathogen isolate and the (c) classification image of *F. boothii* M0002 (purple), *F. boothii* M0010 (light blue) and *F. boothii* M0100 (yellow), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.91

Figure 4.14 PLS-DA of mosaic nine (Day 7) illustrating the (a) score plot of factor 1 (86.6%) vs factor 2 (9.21%) colour coded per pathogen isolate, the (b) score plot of factor 2 (9.21%) vs factor 3 (2.08%) colour coded per pathogen isolate and the (c) classification image of *F. boothii* M0002 (purple), *F. boothii* M0010 (light blue) and *F. boothii* M0100 (yellow), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.91

Figure 4.15 PLS-DA of mosaic ten (Day 3) illustrating the (a) score plot of factor 1 (86.3%) vs factor 2 (8.11%) colour coded per pathogen isolate, the (b) score plot of factor 2 (8.11%) vs factor 3 (2.51%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. verticillioides* MRC8267 (teal) and *F. verticillioides* MRC8559 (pink), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.92

Figure 4.16 PLS-DA of mosaic eleven (Day 5) illustrating the (a) score plot of factor 1 (82.9%) vs factor 2 (8.71%) colour coded per pathogen isolate, the (b) score plot of factor 2 (8.71%) vs factor 3 (7.03%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. verticillioides* MRC8267 (teal) and *F. verticillioides* MRC8559 (pink), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.92

Figure 4.17 PLS-DA of mosaic twelve (Day 7) illustrating the (a) score plot of factor 1 (90.9%) vs factor 2 (7.87%) colour coded per pathogen isolate, the (b) score plot of factor 2 (7.87%) vs factor 3 (0.493%) colour coded per pathogen isolate and the (c) classification image of <i>F. verticillioides</i> MRC826 (blue), <i>F. verticillioides</i> MRC8267 (teal) and <i>F. verticillioides</i> MRC8559 (pink), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.	93
Figure A1 Difference spectra of <i>F. verticillioides</i> MRC826 and the remaining isolates. ...	104
Figure A2 Difference spectra of <i>F. verticillioides</i> MRC8267 and remaining isolates.	104
Figure A3 Difference spectra of <i>F. verticillioides</i> MRC8559 and the remaining isolates.	104
Figure A4 Difference spectra of <i>F. boothii</i> M0002 and the remaining isolates.	105
Figure A5 Difference spectra of <i>F. boothii</i> M0010 and the remaining isolates.	105
Figure A6 Difference spectra of <i>F. boothii</i> M0100 and the remaining isolates.	105
Figure A7 Difference spectra of <i>F. graminearum</i> M14-55 and the remaining isolates. ...	106
Figure A8 Difference spectra of <i>S. maydis</i> SM8 and the remaining isolates.	106
Figure B1 Pathogen diagram illustrating the isolates used in this study, together with the incubation times and number of replicates for each isolate, where maize ear rot types are distinguished by colour, i.e. FER (blue), GER (pink) and DER (green).	107
Figure B2 Digital images of the eight <i>F. verticillioides</i> , <i>F. graminearum</i> s.s., <i>F. boothii</i> and <i>S. maydis</i> isolates, on day 3 of growth, used in the study.	108
Figure B3 Digital images of the eight <i>F. verticillioides</i> , <i>F. graminearum</i> s.s., <i>F. boothii</i> and <i>S. maydis</i> isolates, on day 5 of growth, used in the study.	108
Figure B4 Digital images of the eight <i>F. verticillioides</i> , <i>F. graminearum</i> s.s., <i>F. boothii</i> and <i>S. maydis</i> isolates, on day 7 of growth, used in the study.	108
Figure B5 Mean spectra of <i>F. verticillioides</i> MRC826 on day 3, 5 and 7 of growth.	109
Figure B6 Mean spectra of <i>F. verticillioides</i> MRC8267 on day 3, 5 and 7 of growth.	109
Figure B7 Mean spectra of <i>F. verticillioides</i> MRC8559 on day 3, 5 and 7 of growth.	109
Figure B8 Mean spectra of <i>F. boothii</i> M0002 on day 3, 5 and 7 of growth.	110
Figure B9 Mean spectra of <i>F. boothii</i> M0010 on day 3, 5 and 7 of growth.	110
Figure B10 Mean spectra of <i>F. boothii</i> M0100 on day 3, 5 and 7 of growth.	110

- Figure B11** Mean spectra of *F. graminearum* s.s. M14-55 on day 3, 5 and 7 of growth. 111
- Figure B12** Mean spectra of *S. maydis* SM8 on day 3, 5 and 7 of growth. 111
- Figure B13** Object wise PCA of mosaic two (Day 5) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (82.2%) vs. PC2 (16%) and (b) PCA score plot of PC2 (16%) vs PC3 (0.896%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. graminearum* s.s. M14-55 (middle) and *S. maydis* SM8 (right). The loading plots is given for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1930 nm..... 112
- Figure B14** Object wise PCA of mosaic three (Day 7) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (69.7%) vs. PC2 (29.4%) and (b) PCA score plot of PC2 (29.4%) vs PC3 (0.398%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. graminearum* s.s. M14-55 (middle) and *S. maydis* SM8 (right). The loading plots for (d) PC1 with bands at 1260, 1400, 1430 and 1930 nm and (e) PC2 with bands at 1260, 1430 and 1880 nm..... 112
- Figure B15** Object wise PCA of mosaic four (Day 3) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (82.5%) vs. PC2 (12.6%) and (b) PCA score plot of PC2 (12.6%) vs PC3 (3.06%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. boothii* M0002 (middle) and *S. maydis* SM8 (right). The loading plots for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1930 nm. 113
- Figure B16** Object wise PCA of mosaic five (Day 5) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (67.4%) vs. PC2 (28.1%) and (b) PCA score plot of PC2 (28.1%) vs PC3 (3.16%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. boothii* M0002 (middle) and *S. maydis* SM8 (right). The loading plots for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1930 nm. 113
- Figure B17** Object wise PCA of mosaic six (Day 7) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (55.5%) vs. PC2 (42.2%) and (b) PCA score plot of PC2 (42.2%) vs PC3 (1.18%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left),

F. boothii M0002 (middle) and *S. maydis* SM8 (right). The loading plots for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1930 nm. 114

Figure B18 Object wise PCA of mosaic seven (Day 3) illustrates poor separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (86.5%) vs. PC2 (10.3%) and (b) PCA score plot of PC2 (10.3%) vs PC3 (1.45%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. boothii* M0002 (left), *F. boothii* M0010 (middle) and *F. boothii* M0100 (right). The loading plots for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1930 nm. 114

Figure B19 Object wise PCA of mosaic eight (Day 5) illustrates separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (92.4%) vs. PC2 (5.39%) and (b) PCA score plot of PC2 (5.39%) vs PC3 (1.43%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. boothii* M0002 (left), *F. boothii* M0010 (middle) and *F. boothii* M0100 (right). The loading plots for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1930 nm. 115

Figure B20 Object wise PCA of mosaic nine (Day 7) illustrates poor separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (86.6%) vs. PC2 (9.86%) and (b) PCA score plot of PC2 (9.86%) vs PC3 (1.47%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. boothii* M0002 (left), *F. boothii* M0010 (middle) and *F. boothii* M0100 (right). The loading plots for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1900 nm. 115

Figure B21 Object wise PCA of mosaic ten (Day 3) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (86.4%) vs. PC2 (8.36%) and (b) PCA score plot of PC2 (8.36%) vs PC3 (2.34%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. verticillioides* MRC8267 (middle) and *F. verticillioides* MRC8559 (right). The loading plots for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1260, 1430 and 1930 nm. 116

Figure B22 Object wise PCA of mosaic eleven (Day 5) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (83.2%) vs. PC2 (10.8%) and (b) PCA score plot of PC2 (10.8%) vs PC3 (4.72%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. verticillioides* MRC8267 (middle) and *F. verticillioides* MRC8559 (right). The loading plots for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1930 nm. 116

Figure B23 Object wise PCA of mosaic twelve (Day 7) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (90.9%) vs. PC2 (7.87%) and (b) PCA score plot of PC2 (7.87%) vs PC3 (0.493%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. verticillioides* MRC8267 (middle) and *F. verticillioides* MRC8559 (right). The loading plots for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1400, 1900 and 2100 nm..... 117

LIST OF TABLES

Table 2.1 A summary of the NIR hyperspectral imaging applications focussing on microbial and pathogenic applications.	29
Table 3.1 The origin of the <i>Fusarium</i> spp. and <i>Stenocarpella maydis</i> isolates investigated.	45
Table 3.2 The performance measures used to assess the overall pixel and object wise PLS-DA models of the pathogen isolates.	61
Table 3.3 The performance measures used to assess the pixel wise PLS-DA models of the pathogen isolates used in the four mosaics.	67
Table 3.4 The performance measures used to assess the object wise PLS-DA models of the pathogen isolates used in the four mosaics.	67
Table 4.1 Performance measures used to assess the overall PLS-DA models of the pathogen isolates in the twelve different mosaics.	93
Table 4.2 Performance measures used to assess the PLS-DA models of the three pathogen isolates in mosaic 1, 2 and 3.	94
Table 4.3 Performance measures used to assess the PLS-DA models of the three pathogen isolates in mosaic 4, 5 and 6.	94
Table 4.4 Performance measures used to assess the PLS-DA models of the three <i>F. boothii</i> isolates in mosaic 7, 8 and 9.	95
Table 4.5 Performance measures used to assess the PLS-DA models of the <i>F. verticillioides</i> isolates in mosaic 10, 11 and 12.	95

LIST OF ABBREVIATIONS

2D	Two Dimensional
3D	Three Dimensional
ANN	Artificial Neural Networks
AOTF	Acousto-Optic Tunable Filters
CCD	Charge Coupled Device
CMOS	Complementary Metal Oxide Semiconductors
DA	Discriminant Analysis
DER	Diplodia Ear Rot
DIP	Diplodia
DON	Deoxynivalenol
ELISA	Enzyme-Linked Immunosorbent Assay
EMCCD	Electron Multiplying Charge Coupled Device
FER	Fusarium Ear Rot
FGSC	<i>Fusarium graminearum</i> species complex
FT	Fourier Transform
FUM	Fumonisin
GER	Gibberella Ear Rot
HSI	Hyperspectral Imaging
InGaAs	Indium Gallium Arsenide
K-M	Kubelka-Munk
LCTF	Liquid Crystal Tunable Filters
LDA	Linear Discriminant Analysis
LED	Light Emitting Diodes
MCT	Mercury-Cadmium-Telluride
MIA	Multivariate Image Analysis
MLR	Multiple Linear Regression
MON	Moniliformin
MPLS	Modified Partial Least Squares
MRC	Medical Research Council (Tygerberg, South Africa)
MSC	Multiplicative Scatter Correction
MVA	Multivariate Analysis
NAS	Net Analyte Signal
NIR	Near Infrared

OSC	Orthogonal Signal Correction
PbS	Lead Sulphide
PC	Principal Component
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PCR	Principal Component Regression
PDA	Potato Dextrose Agar
PGP	Prism-Grating-Prisms
PLS-DA	Partial Least Squares Discriminant Analysis
PLSR	Partial Least Squares Regression
Q ²	Cross-Validated Coefficient of Determination
R ²	Coefficient of Determination
RNV	Robust Normal Variate
ROI	Region of Interest
SAM	Spectral Angle Mapping
SIMCA	soft independent modelling of class analogy
SNV	Standard Normal Variate
SVM	Support Vector Machines
SWIR	Short Wave Infrared
TH	Tungsten-Halogen
Vis-NIR	Visible-Near Infrared
WT	Wavelet Transforms
WM	White Maize
YM	Yellow Maize

CHAPTER 1

Introduction

Maize (*Zea mays* L.) is one of the most important cereal grains produced in South Africa and Africa as it serves as a staple food for many people and comprises a significant percentage of animal feed. Furthermore, it may also serve as an ingredient in specialised foods (Hellin *et al.*, 2012; Williams *et al.*, 2012). The 2017 South African commercial maize crop had a yield upwards of 14 million tons (Department of Agriculture, Forestry and Fisheries, 2017), which illustrates the vital role maize plays in the South African agricultural sector. Crop losses caused by physical plant damage or disease, a result of external factors such as weather changes and maize crop infestation, should be avoided to ensure maximum yield from each maize crop.

Maize is known to be highly susceptible to numerous micro-organisms, specifically the mycotoxin-producing fungi belonging to the *Fusarium* and *Stenocarpella* genus, which leads to three maize ear rots in South Africa, namely Fusarium ear rot (FER), Gibberella ear rot (GER) and Diplodia ear rot (DER). *Fusarium verticillioides* (syn: *F. moniliforme*) produces the fumonisin mycotoxin and results in the formation of FER (Munkvold, 2003; Boutigny *et al.*, 2012). The fumonisin toxin has been associated with oesophageal cancer in humans and rats (Yao *et al.*, 2008), as well as leukoencephalomalacia in horses and pulmonary edema in pigs (Gelderblom *et al.*, 2001). Boutigny *et al.* (2012) found that *Fusarium boothii* is the leading cause of GER in maize ears, while *F. graminearum* s.s. was found to be the most abundant *Fusarium graminearum* species complex (FGSC) species on maize ears undergoing crop rotation with wheat (Gokul, 2015). Deoxynivalenol (DON) is one of the main toxins produced by the FGSC and has been found to suppress the immune system and cause nausea and vomiting in humans (Peraica *et al.*, 1999). Another plant pathogen, that results in DER, of maize is *Stenocarpella maydis*, which affects the yield and quality of maize (Rossouw *et al.*, 2002) and is known to produce the diplonin toxin. The maize ears are gradually infected, starting at the base and moving upwards to form a white mould between the kernels (EPPO, 2017). These four ear rot pathogens have led to serious maize crop losses, placing unnecessary strain on the maize industry (Battilani *et al.*, 2008). It is important to not only distinguish between *Stenocarpella* and *Fusarium*, but also species and isolates within the *Fusarium* spp. as they produce different mycotoxins. Furthermore, the taxonomy of the *Fusarium* genus is similar and it is therefore difficult to distinguish between species, as well as isolates (Snyder & Hansen, 1940; Williams, 2013).

As microbiological testing plays a vital role in food manufacturing and processing to ensure food safety, it is necessary to investigate rapid alternative methods to those currently used in the laboratory. The current methods employed by microbiology laboratories for the enumeration and identification of pathogens include both molecular and immunological detection such as the polymerase chain reaction (PCR) (Strausbaugh *et al.*, 2005) and enzyme-linked immunosorbent assay (ELISA) (ThermoFischer Scientific, 2017). These multi-step methods are known to be tedious and time-consuming. Furthermore, specialised growth media, supplementary staining and microscopy may also be required, which extends the identification timeframe (Velusamy *et al.*, 2010; Turner *et al.*, 2015). These methods allow the fate of a cereal batch to only be determined once the identification thereof has been completed, which could take over two weeks. The potential of toxin build-up within a batch will increase within this timeframe, as well as spread the pathogens to other samples (Velusamy *et al.*, 2010).

Therefore, there is a need for a rapid, non-destructive scanning method that can accurately distinguish between the pathogens present in maize (Yao *et al.*, 2008). Maize grain needs to be tested rapidly and accurately to ensure that it adheres to the food safety regulations set by governing bodies, as it is such an important part of the South African diet. This can be achieved using near infrared (NIR) hyperspectral imaging, a technique that enables rapid sample measurement.

NIR hyperspectral imaging produces three dimensional (3D) images known as hypercubes that contain one spectral (λ) and two spatial dimensions (x and y) (Burger, 2006). The images consist of numerous contiguous wavelengths within the NIR region of the electromagnetic spectrum between 780 and 2500 nm (Geladi *et al.*, 2004). Each pixel, or data point, has a spectrum, which can act as a fingerprint during the identification or differentiation of a sample. The unique signature can be used to estimate the chemical composition of the sample and thus classify the constituents present (Burger & Gowen, 2011). This is all made possible with multivariate image analysis (MIA), a powerful combination of chemometrics and image analysis, used to obtain sample characteristics from the images (Prats-Montalbán *et al.*, 2011).

Principal component analysis (PCA) is a popular, versatile technique that can be employed for the analysis of various datasets (Cowe & McNicol, 1985; Esbensen *et al.*, 2002; Reich, 2005) and is often referred to as the workhorse of MIA. PCA can be used as an essential initial step in reducing data dimensionality and permitting exploratory analysis, an essential process during hyperspectral image analysis. PCA is applied before other multivariate techniques as it showcases the main variability sources by breaking the data

down into scores and loadings (Gowen *et al.*, 2007; Amigo *et al.*, 2013). Furthermore, PCA illustrates vital information with regards to the chemical and physical properties of a sample and how these properties are distributed. While PCA provides an overview of the samples studied, classification algorithms are used to identify and predict unknown samples. Partial least squares discriminant analysis (PLS-DA) is one such algorithm.

PLS-DA is frequently used to qualitatively predict objects within a sample (Gowen *et al.*, 2007). The algorithm is based on finding a straight line that can divide the data into two or more regions to separate it into groups (Musumarra *et al.*, 2007; Amigo *et al.*, 2013; Brereton & Lloyd, 2014; ElMasry & Nakauchi, 2016). Both PCA and PLS-DA can be performed using one of two approaches: pixel wise or object wise (Williams & Kucheryavskiy, 2016). The main difference between the two approaches lies in the number of data points used during the analysis. The pixel wise approach utilises each pixel, and subsequent spectrum, during the analysis of samples. The object wise approach clusters pixels into pre-determined groups and uses a mean spectra for the analysis, thus decreasing the data points (Kucheryavskiy, 2013).

NIR hyperspectral imaging studies have proven that it is possible to distinguish pathogens on growth media (Yao *et al.*, 2008; Del Fiore *et al.*, 2010; Williams *et al.*, 2012b; Foca *et al.*, 2016; Kammies *et al.*, 2016; Seo *et al.*, 2016). However, these studies did not compare isolates from the same species or investigate different growth ages during the differentiation thereof. There is thus a need to expand the pathogens investigated by the previous studies, as well as investigate the effect of growth age on the differentiation accuracy. This study was aimed at investigating the potential of using near infrared hyperspectral imaging as a rapid screening method to distinguish between maize ear rot pathogens on growth media, as well as determine the age of the colony best suited to distinguish between these pathogens. Two approaches to multivariate data analysis were also compared, namely pixel wise and object wise analysis. Specific objectives were established to:

- differentiate maize ear rot pathogens, on growth media, using near infrared hyperspectral imaging and
- determine the influence of colony growth age on the accuracy of ear rot pathogen differentiation, on growth media, using NIR hyperspectral imaging.

1.1. REFERENCES

- Amigo, J.M., Martí, I. & Gowen, A. (2013). Hyperspectral Imaging and Chemometrics: A Perfect Combination for the Analysis of Food Structure, Composition and Quality. In: *Data Handling in Science and Technology*. Pp. 343–370.
- Bacon, C.W., Porter, J.K., Norred, W.P. & Leslie, J.F. (1996). Production of fusaric acid by *Fusarium* species. *Applied and Environmental Microbiology*, **62**, 4039–4043.
- Balconi, C., Berardo, N., Locatelli, S., Lanzanova, C., Torri, A. & Redaelli, R. (2014). Evaluation of ear rot (*Fusarium verticillioides*) resistance and fumonisin accumulation in Italian maize in. *Phytopathologia Mediterranea*, **53**, 14–26.
- Battilani, P., Pietri, A., Barbano, C., Scandolara, A., Bertuzzi, T. & Marocco, A. (2008). Logistic regression modeling of cropping systems to predict fumonisin contamination in maize. *Journal of Agricultural and Food Chemistry*, **56**, 10433–10438.
- Boutigny, A.L., Beukes, I., Small, I., Zühlke, S., Spiteller, M., Van Rensburg, B.J., Flett, B. & Viljoen, A. (2012). Quantitative detection of *Fusarium* pathogens and their mycotoxins in South African maize. *Plant Pathology*, **61**, 522–531.
- Brereton, R.G. & Lloyd, G.R. (2014). Partial least squares discriminant analysis: taking the magic away. *Journal of Chemometrics*, **28**, 213–225.
- Burger, J. & Gowen, A. (2011). Data handling in hyperspectral image analysis. *Chemometrics and Intelligent Laboratory Systems*, **108**, 13–22.
- Burger, J.E. (2006). *Hyperspectral NIR Image Analysis*. PhD Thesis. Swedish University of Agricultural Sciences, Umeå.
- Cotten, T.K. & Munkvold, G.P. (1998). Survival of *Fusarium moniliforme*, *F. proliferatum*, and *F. subglutinans* in maize stalk residue. *Phytopathology*, **88**, 550–555.
- Cowe, I., & McNicol, J.W. (1985). The Use of Principal Components in the Analysis of Near-Infrared Spectra. *Applied Spectroscopy*, **39**, 257–266.
- Department of Agriculture, Forestry and Fisheries. (2017). Summer Crops: Third Production Forecast (2017) and Winter Cereals: Intentions to plant (2017). [Internet document] URL <http://www.daff.gov.za/daffweb3/Home/Crop-Estimates>. Accessed 24/07/2017.
- Del Fiore, A., Reverberi, M., Ricelli, A., Pinzari, F., Serranti, S., Fabbri, A.A., Bonifazi, G. & Fanelli, C. (2010). Early detection of toxigenic fungi on maize by hyperspectral imaging analysis. *International Journal of Food Microbiology*, **144**, 64–71.
- EIMasry, G.M. & Nakauchi, S. (2016). Image analysis operations applied to hyperspectral images for non-invasive sensing of food quality - A comprehensive review. *Biosystems Engineering*, **142**, 53–82.

- EPPO. (2017). *Stenocarpella macrospora* and *Stenocarpella maydis* [Internet document] URL https://www.eppo.int/QUARANTINE/data_sheets/.../DIPDSP_ds.pdf. Accessed 24/07/2017.
- Esbensen, K.H., Guyot, D., Westad, F. & Houmoller, L.P. (2002). Principal Component Analysis (PCA). In: *Multivariate data analysis in practice: an introduction to multivariate data analysis and experimental design*. 5th edn. CAMO Process AS.
- Flett, B.C., McLaren, N.W. & Wehner, F.C. (1998). Incidence of Ear Rot Pathogens Under Alternating Corn Tillage Practices. *Plant Disease*, **82**, 781–784.
- Foca, G., Ferrari, C., Ulrici, A., Sciutto, G., Prati, S., Morandi, S., Brasca, M., Lavermicocca, P., Lanteri, S. & Oliveri, P. (2016). The potential of spectral and hyperspectral-imaging techniques for bacterial detection in food: A case study on lactic acid bacteria. *Talanta*, **153**, 111–119.
- Geladi, P., Burger, J. & Lestander, T. (2004). Hyperspectral imaging: Calibration problems and solutions. *Chemometrics and Intelligent Laboratory Systems*, **72**, 209–217.
- Gelderblom, W.C.A., Seier, J.V., Snijman, P.W., Van Schalkwyk, D.J., Shephard, G.S. & Marasas, W.F.O. (2001). Toxicity of culture material of *Fusarium verticillioides* strain MRC 826 to nonhuman primates. *Environmental Health Perspectives*, **109**, 267–276.
- Gokul, A. (2015). *Fusarium graminearum* species complex (FGSC) composition in South African wheat and maize grown in rotation. MSc Thesis. Stellenbosch University, Stellenbosch.
- Gowen, A.A., Feng, Y., Gaston, E. & Valdramidis, V. (2015). Recent applications of hyperspectral imaging in microbiology. *Talanta*, **137**, 43–54.
- Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Downey, G. & Frias, J.M. (2007). Hyperspectral imaging - an emerging process analytical tool for food quality and safety control. *Trends in Food Science and Technology*, **18**, 590–598.
- Hellin, J., Shiferaw, B., Cairns, J.E., Reynolds, M.P., Ortiz-monasterio, I., Banziger, M., Sonder, K. & La Rovere, R. (2012). Climate Change and Food Security in the Developing World: Potential of Maize and Wheat Research to Expand Options for Adaptation and Mitigation. *Journal of Development and Agricultural Economics*, **4**, 311–321.
- Kammies, T., Manley, M., Gouws, P.A. & Williams, P.J. (2016). Differentiation of foodborne bacteria using NIR hyperspectral imaging and multivariate data analysis. *Applied Microbiology and Biotechnology*, **100**, 9305–9320.
- Kucheryavskiy, S. (2013). A new approach for discrimination of objects on hyperspectral images. *Chemometrics and Intelligent Laboratory Systems*, **120**, 126–135.

- Lanubile, A., Ferrarini, A., Maschietto, V., Delledonne, M., Marocco, A. & Bellin, D. (2014). Functional genomic analysis of constitutive and inducible defense responses to *Fusarium verticillioides* infection in maize genotypes with contrasting ear rot resistance. *BMC Genomics*, **15**, 710.
- Munkvold, G.P. (2003). Cultural and genetic approaches to managing mycotoxins in maize. *Annual Review of Phytopathology*, **41**, 99–116.
- Musumarra, G., Trovato-Salinaro, A., Scirè, S., Foti, A., Barresi, V., Fortuna, C.G., Strazzulla, G., Condorelli, D.F. (2007). Identification of genes involved in radiation-induced G1 arrest. *Journal of Chemometrics*, 398–405.
- Peraica, M., Radić, B., Lucić, A. & Pavlović, M. (1999). Toxic effects of mycotoxins in humans. *Bulletin of the World Health Organization*, **77**, 754–766.
- Prats-Montalbán, J.M., De Juan, A. & Ferrer, A. (2011). Multivariate image analysis: A review with applications. *Chemometrics and Intelligent Laboratory Systems*, **107**, 1–23.
- Reich, G. (2005). Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Advanced Drug Delivery Reviews*, **57**, 1109–1143.
- Rossouw, J.D., Van Rensburg, J.B.J. & Van Deventer, C.S. (2002). Breeding for resistance to ear rot of maize, caused by *Stenocarpella maydis* (Berk) Sutton. 1. Evaluation of selection criteria. *South African Journal of Plant and Soil*, **19**, 182–187.
- Seo, Y., Park, B., Hinton, A., Yoon, J.S. & Lawrence, K.C. (2016). Identification of *Staphylococcus* species with hyperspectral microscope imaging and classification algorithms. *Journal of Food Measurement and Characterization*, **10**, 253–263.
- Snyder, W.C. & Hansen, H.N. (1940). The Species Concept in *Fusarium*. *American Journal of Botany*, **27**, 64–67.
- Strausbaugh, C.A., Overturf, K. & Koehn, A.C. (2005). Pathogenicity and real-time PCR detection of *Fusarium* spp. in wheat and barley roots. *Canadian Journal of Plant Pathology*, **27**, 430–438.
- Sydenham, E.W., Gelderblom, W.C.A., Thiel, P.G. & Marasas, W.F.O. (1990). Evidence for the natural occurrence of fumonisin B₁, a mycotoxin produced by *Fusarium moniliforme*, in corn. *Journal of Agricultural and Food Chemistry*, **38**, 285–290.
- ThermoFischer Scientific. (2017). Overview of ELISA [Internet document]. URL <https://www.thermofisher.com/za/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-elisa.html>. Accessed 27/09/2017.

- Turner, N.W., Bramhmbhatt, H., Szabo-Vezse, M., Poma, A., Coker, R. & Piletsky, S.A. (2015). Analytical methods for determination of mycotoxins: An update (2009-2014). *Analytica Chimica Acta*, **901**, 12–33.
- Velusamy, V., Arshak, K., Korostynska, O., Oliwa, K. & Adley, C. (2010). An overview of foodborne pathogen detection: In the perspective of biosensors. *Biotechnology Advances*, **28**, 232–254.
- Wagacha, J.M. & Muthomi, J.W. (2008). Mycotoxin problem in Africa: Current status, implications to food safety and health and possible management strategies. *International Journal of Food Microbiology*, **124**, 1–12.
- 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. Stellenbosch University, Stellenbosch.
- Williams, P.J., Geladi, P., Britz, T.J. & Manley, M. (2012a). Investigation of fungal development in maize kernels using NIR hyperspectral imaging and multivariate data analysis. *Journal of Cereal Science*, **55**, 272–278.
- Williams, P.J., Geladi, P., Britz, T.J. & Manley, M. (2012b). Growth characteristics of three *Fusarium* species evaluated by near-infrared hyperspectral imaging and multivariate image analysis. *Applied Microbiology and Biotechnology*, **96**, 803–813.
- Williams, P.J. & Kucheryavskiy, S. (2016). Classification of maize kernels using NIR hyperspectral imaging. *Food Chemistry*, **209**, 131–138.
- Yao, H., Hruska, Z., Kincaid, R., Brown, R.L. & Cleveland, T.E. (2008). Differentiation of toxigenic fungi using hyperspectral imagery. *Sensing and Instrumentation for Food Quality and Safety*, **2**, 215–224.
- Yazar, S. & Omurtag, G.Z. (2008). Fumonisin, Trichothecenes and Zearalenone in Cereals. *International Journal of Molecular Sciences*, **9**, 2062–2090.

CHAPTER 2

The potential application of NIR hyperspectral imaging in the maize industry: A review

2.1. INTRODUCTION

Maize (*Zea mays* L.) is one of the most important cereal crops in Africa, as many African households rely on it as a dietary staple food. Maize is known to be highly susceptible to fungal contamination, especially by *Fusarium* spp. and *Stenocarpella* spp., which is of great concern to the industry as they produce a number of mycotoxins (Wagacha & Muthomi, 2008; Lanubile *et al.*, 2014).

The *Fusarium* genus is known to be difficult to distinguish due to the taxonomy thereof (Snyder & Hansen, 1940; Williams, 2013). However, it is important to be able to easily distinguish between the species as each of the pathogens produce different mycotoxins. Microbiological testing is important for food processing to ensure products are safe for consumption. Molecular and immunological detection methods such as the polymerase chain reaction (PCR) (Strausbaugh *et al.*, 2005) and enzyme-linked immunosorbent assay (ELISA) have been used for the enumeration and identification of the pathogens. These multi-step methods are tedious, time-consuming and can require specialised growth media, supplementary staining techniques and microscopy (Turner *et al.*, 2015). The identification can take more than two weeks, where the fate of a cereal batch is only determined once identification is complete, thus increasing the chance of toxin build up within the batch (Velusamy *et al.*, 2010). The need for a non-invasive, rapid screening method that can be implemented in microbiology laboratories, therefore exists. Food products need to be tested rapidly and accurately to ensure it adheres to the food safety regulations set by governing bodies. This can be achieved using spectroscopic techniques such as near infrared (NIR) hyperspectral imaging.

NIR hyperspectral imaging is a chemical imaging technique that provides chemical and structural information from a given sample (Burger, 2006). NIR hyperspectral imaging offers rapid sample measurement and is capable of producing three dimensional (3D) images comprised of spectral (wavelength) and spatial (pixel coordinates) resolution (Yao *et al.*, 2008). It captures images at many wavelength bands in the NIR region between 780 nm and 2500 nm, where each pixel contains the spectrum of that exact spatial location. Each image contains millions of data points, across a range of wavelengths that is obtained within seconds of imaging. The main goal of NIR hyperspectral imaging is either classification or

quantification of material, which can be achieved using multivariate analysis (MVA) and chemometric tools. The NIR hyperspectral images can be analysed using techniques such as principal component analysis (PCA) for exploratory analysis and partial least squares discriminant analysis (PLS-DA) or support vector machines (SVM) for classification (Gowen *et al.*, 2007; Prats-Montalbán *et al.*, 2011).

This review aims to address the potential application of NIR hyperspectral imaging in the maize industry. This technique, once optimised, would enable by focussing on the maize ear rot pathogens, *Fusarium* spp. and *Stenocarpella* spp. It also addresses the conventional methods used and offers a spectroscopic alternative, namely NIR hyperspectral imaging.

2.2. MAIZE EAR ROT

Maize is used for human and animal consumption, since it is considered to be one of the main dietary staple foods in underdeveloped countries, as well as an ingredient in specialised foods (Wagacha & Muthomi, 2008). The maize industry produced approximately 66% of the South African cereal in 2016 and it is estimated that the 2017 crop will yield more than 14 million tons of commercial maize, which includes both white and yellow maize (Department of Agriculture, Forestry and Fisheries, 2017). These statistics illustrate the importance of maize in South African agriculture, thus ensuring its sustained production is vital. Crop losses occur due to physical plant damage or disease, which can be attributed to various external factors such as weather changes and organisms, such as fungi, that infect maize.

Environmental factors such as high humidity, temperature and moisture, as well as unseasonal rainfall enable fungi to proliferate and produce mycotoxins. South Africa and the larger Africa commonly experience these weather conditions. Socio-economic factors also contribute, as the consumption of maize increases with the ever-growing population (Wagacha & Muthomi, 2008).

Maize is known to be highly susceptible to various forms of microorganism contamination, specifically to the toxin-producing fungi such as *Fusarium* spp. The maize ear rot specifically caused by *Fusarium* species includes Fusarium- and Gibberella ear rot of maize while *Stenocarpella* spp. cause Diplodia ear rot of maize in South Africa. These maize ear rots constitute the greatest economical concern to farmers in South Africa (Boutigny *et al.*, 2012; Schoeman & Flett, 2012). Recently, the regulations governing the tolerance for fungus-produced toxins in foodstuffs in South Africa has been amended (South African Gazette, 2016).

2.2.1. Fusarium Ear Rot (FER)

Fusarium verticillioides (synonym, *Fusarium moniliforme* Sheldon; teleomorph, *Gibberella moniliformis*) is a fungal pathogen most commonly associated with Fusarium ear rot in South African maize, along with *F. subglutinans* and *F. proliferatum* (Boutigny *et al.*, 2011). The fungus is known to infect physically damaged kernels, groups of kernels or random kernels (Figure 2.1), resulting in pink or white mould formation (Munkvold, 2003).



Figure 2.1 Fusarium ear rot (FER) illustrating the random kernel infection on a maize ear (Pannar, 2017).

The fungus systematically infects kernels by entering through the pedicle and moves towards the endosperm, where it sporulates in the presence of moisture (Bacon *et al.*, 1996; Williams, 2009). Initially the fungus is prone to infect smaller, developing kernels, allowing it time to grow without presenting any visual symptoms (Woloshuk & Wise, 2014). This results in the possible production of mycotoxins without any visible damage (Bacon *et al.*, 1996).

Fusarium verticillioides is present worldwide in a large variety of plants and soils in both sub-tropical and tropical zones, where warmer, humid conditions are found (Gelderblom *et al.*, 2001). The fungal growth on media appears cream-beige, pale orange or violet in colour with white mycelium covering it (Joffe, 1986). *Fusarium verticillioides* is a mycotoxigenic fungus that produces a variety of mycotoxins of which fumonisins are the most important. They are harmful to both animal and human health. The first outbreak of *F. moniliforme* in South Africa, was reported by Van der Walt and Steyn (1943) and was associated with leukoencephalomalacia (LEM) syndrome in horses accompanied by brain lesions, edema and liver damage due to the production of fumonisins by the fungus. This

mycotoxins has also been associated with infertility among dairy cattle, and chicken weight loss and death (Joffe, 1986). The presence of fumonisins in maize has also been related to pulmonary edema in pigs, as well as tumour formation in rats (Gelderblom *et al.*, 1988). The level of fumonisins in maize grain for processing has been determined at 4 000 µg/kg fumonisins while products for human consumption should not contain more than 2 000 µg/kg of fumonisins (South African Gazette, 2016).

2.2.2. Gibberella Ear Rot (GER)

Fusarium graminearum (synonym, *Fusarium roseum*; teleomorph, *Gibberella zeae*) is one of the main fungi responsible for Gibberella ear rot (GER) on maize. *Fusarium graminearum* is readily referred to as part of the *Fusarium graminearum* species complex (FGSC), which consists of 16 phylogenetically distinct species (Aoki *et al.*, 2012). *Fusarium boothii* has also been commonly associated with GER in South Africa (Boutigny *et al.*, 2011), while *F. graminearum* sensu stricto was found to be the most abundant FGSC species in a survey of wheat and maize produced within crop rotation systems (Gokul, 2015).



Figure 2.2 Gibberella ear rot (GER) illustrating the pink mould formation on a maize ear (Pannar, 2017).

The FGSC infect maize ears from the tip downwards, gradually covering the entire ear in mycelia (Munkvold, 2003; Boutigny *et al.*, 2012). Initially, the symptoms present as a white mould that gradually turns dark pink or red with time (Figure 2.2). Early infection results in the whole maize ear being covered in mould and rotting, with individual kernels also being shrivelled and broken (Gokul, 2015). Furthermore, the fungal growth on culture

media appears white-rose, peach or grey in colour, where the microconidia are absent and the mycelium appears white in colour and chlamydospores are present (Joffe, 1986).

Apart from visible mould formation, kernels may also be contaminated with mycotoxins, specifically the Type B trichothecenes which consists of deoxynivalenol (DON) and nivalenol. DON is also known as vomitoxin due to the effect it has on animals and people after consumption of contaminated grain. DON has been linked to vomiting and anorexia in pigs and nausea and diarrhoea in humans (Beukes *et al.*, 2017). The toxins produced by *Fusarium graminearum* s.s., namely nivalenol and DON (Beukes *et al.*, 2017), has been associated with maize, wheat and barley and found by Fischer *et al.* (1967) to cause staggering and low milk production in dairy cattle (Fischer *et al.* 1967, cited by Joffe, 1986). It has also been associated with fertility problems, udder inflammation and irregular heat in cattle, hyperestrogenism in pigs and abortions in goats (Joffe, 1986). The concentration of DON in South African grains including maize, before and after processing, has been limited to 2 000 and 1 000 µg/kg DON, respectively (South African Gazette, 2016).

2.2.3. Diplodia Ear Rot (DER)

Diplodia ear rot, another prominent ear rot found in South Africa, is most commonly caused by *Stenocarpella maydis* (basionym, *Diplodia maydis*). The fungus forms white or grey mould in between the kernels that spread throughout the maize ear until it is completely covered (Figure 2.3).



Figure 2.3 Diplodia ear rot (DER) illustrating the white mould formation in between the kernels on a maize ear (Pannar, 2017).

Stenocarpella maydis is one of two *Stenocarpella* species, the other being *S. macrospora*, that is a known cause of maize ear rot (Wicklow *et al.*, 2011). It is considered one of the most important maize pathogens as it causes ear rot in all countries where maize is grown, especially humid areas (Rossouw *et al.*, 2009). *Stenocarpella maydis* has also been linked to diplodiosis, a common nervous disorder called neuro-mycotoxicosis, with outbreaks found in sheep and cattle that graze on maize crops in Southern Africa (Kellerman *et al.*, 1985) and Argentina (Odriozola *et al.*, 2005). *Stenocarpella maydis* is known to produce diplonin, a toxin that leads to diplodia in maize, and has been shown to affect the grading quality thereof (Rossouw *et al.*, 2002; Sendin, 2017).

2.2.4. Current Industry Practices

2.2.4.1. Maize Grading

The South African maize industry produces two types of maize that are utilized for different purposes. White maize is generally used for human consumption and therefore has stricter guidelines compared to yellow maize that is used as animal feed (Gokul, 2015). Currently the maize industry uses a grading system for all the maize produced in South Africa. This legislation is based on grouping the maize into three grades, for example the top grade of white maize (WM) would be WM1, followed by WM2 and WM3. The same is seen in the yellow maize (YM), with YM1, YM2 and lastly YM3 (Department of Agriculture, 2009). There are four categories of material that ideally should not be present in the maize. These include defective maize, maize of other colour, pinked maize and foreign matter. Fungal damage is one of the main attributes regarded as defective maize, with both *Fusarium* and *Diplodia* damage being included in legislation (Sendin, 2017). Top grade white maize (WM1) should have a maximum of 7% fungal damage present, where yellow maize (YM1) is allowed 9% (Department of Agriculture, 2009). This is worrisome as the grading process is done visually, and fungal presence and subsequent mycotoxin contamination, may occur without symptoms.

2.2.4.2. Conventional Testing Methods

As food safety is one of the major concerns for both the maize industry and food industry in general, adequate testing methods must be applied to ensure that food is safe for consumption. Various techniques have been employed to aid in both the detection and identification of bacteria and fungi. The techniques may start with general culture media to obtain an initial understanding of what is present in/on the sample, where after selective

media may be used to isolate the organisms present (Williams, 2013). Identifying microorganisms based on morphological traits may be possible at this stage, however, this requires specialised training and experience and it may be difficult to discriminate between closely related species. Apart from culture media, molecular and immunological methods have also been used to detect and identify pathogenic species (Strausbaugh *et al.*, 2005). Polymerase chain reaction (PCR), a frequently used molecular method, uses the DNA of an organism as a means of identification. The method follows numerous steps that include DNA extraction, PCR method design and electrophoresis (Schilling *et al.*, 1996). An immunological method used for the identification and characterisation of the pathogens is enzyme-linked immunosorbent assay (ELISA). ELISA uses the antibodies, peptides and proteins present in a pathogen. The enzymes are linked to the antibodies, after which the specific conjugated enzyme activity observed allows for identification of a pathogen (ThermoFischer Scientific, 2017). Apart from quantitative realtime PCR, where the fungi are quantified directly in the grain, other PCR-based and ELISA methods require the isolation of the pathogen before proceeding with the detection, which is time-consuming. The main drawback of quantitative realtime PCR is that it is extremely expensive.

Therefore, the need for a non-destructive, rapid screening method that can accurately detect the fungal pathogens present in a sample and that can be implemented in microbiology laboratories has risen. This can be achieved using a technique such as near infrared (NIR) hyperspectral imaging, which offers rapid sample measurement. NIR hyperspectral imaging is based on the same principles of its predecessor, NIR spectroscopy. Therefore, a brief introduction NIR spectroscopy is necessary before elaborating on the hyperspectral imaging technique.

2.3. NEAR INFRARED SPECTROSCOPY

In recent years, the use of near infrared (NIR) spectroscopy as a non-destructive, rapid analytical technique has evolved, especially in the food industry (Cen & He, 2007). The NIR region was first discovered by Herschel in the early 1800's, when he found that, after separating the electromagnetic spectrum with a prism, temperature increased beyond the red of the visible spectrum (Herschel, 1800). As stated in Reich (2005), the NIR region is defined by the American Society of Testing and Materials as the wavelength range of 780 - 2526 nm in the electromagnetic spectrum and corresponds to the wave number range of 12820 - 3959 cm^{-1} . Bunsen and Kirchhoff conducted spectroscopic analyses in the 1860's using a prism spectroscope, which is the base of many analytical spectroscopic techniques (Scotter, 1997). It was only in the 1960's that the contribution of Norris on the evaluation of

food quality allowed for the widespread practical implementation of NIR spectroscopy in the food industry (Cen & He, 2007).

Spectroscopic analysis relies on the interaction between electromagnetic radiation and atoms or molecules to obtain quantitative and qualitative information contained within the spectrum of energy that is emitted or absorbed. Thus the data obtained from NIR spectroscopy is either called an absorbance or reflectance spectrum (Scotter, 1997). With an NIR spectrum representing numerous absorption bands, it can provide structural information of a given sample (Cen & He, 2007). The absorption bands in an NIR spectrum represent the overtones and vibrations that originate from the hydrogen, carbon, oxygen and nitrogen bonds [C-C, C-H, C=O, N-H, O-H], present in a sample, which undergo a change in energy when it is irradiated by NIR frequencies. These are a result of the change in vibrational energy, which include both stretching and bending vibrations. The assignment of peaks representing functional groups began in the early 1900's, with Coblentz (1905) being the first to observe the C-H bond. Lanza and Li (1984) reported that absorption bands at 1450, 1790 and 2266 nm are related to glucose, fructose and sucrose, whereas Cozzolino *et al.* (2003) suggested that white wine had an absorption band at 1450 nm relating to the second overtone of O-H of water and ethanol.

NIR spectroscopy can only provide spectral information about a sample, where one spectrum represents the entire sample. The lack of spatial information can be easily overcome with the use of hyperspectral imaging, a spectroscopic technique that incorporates imaging to provide both spectral and spatial information (Feng *et al.*, 2013).

2.4. NIR HYPERSPECTRAL IMAGING

Hyperspectral imaging was first used by Goetz *et al.* (1985) as a new technique to survey the earth's surface by means of remote sensing. After which the technology has evolved and been applied in various fields. Agriculture is one of the fields where this technology has been proven to be effective in food composition (Cozzolino *et al.*, 2003; Mehl *et al.*, 2004; Mahesh *et al.*, 2008; Manley *et al.*, 2009; Menesatti *et al.*, 2009; Molto *et al.*, 2010), defect detection (Ariana *et al.*, 2006; Wang *et al.*, 2006; Ariana & Lu, 2010; Menesatti *et al.*, 2010; Wu *et al.*, 2013; Ravikanth *et al.*, 2016) and microbial detection (Berardo *et al.*, 2005; Saccon *et al.*, 2005; Yao *et al.*, 2008; Del Fiore *et al.*, 2010; Peng *et al.*, 2011; Senthilkumar *et al.*, 2012) to name just a few.

Hyperspectral imaging produces three-dimensional images that can consist of hundreds of different wavelengths representing the NIR spectrum in a pre-selected range. These images are three-dimensional multivariate data cubes (Figure 2.4), also known as

hypercubes, that contain two spatial dimensions (X and Y) and one spectral dimension (λ) (Burger, 2006). Each pixel, or data point, is related to a spectrum, which acts as a fingerprint of a sample with a unique signature that can be used to estimate the composition of the material or classify the components present. This makes it superior to that of conventional NIR spectroscopy as it provides spatial information, which allows for the visualisation of the given sample (Burger & Gowen, 2011).

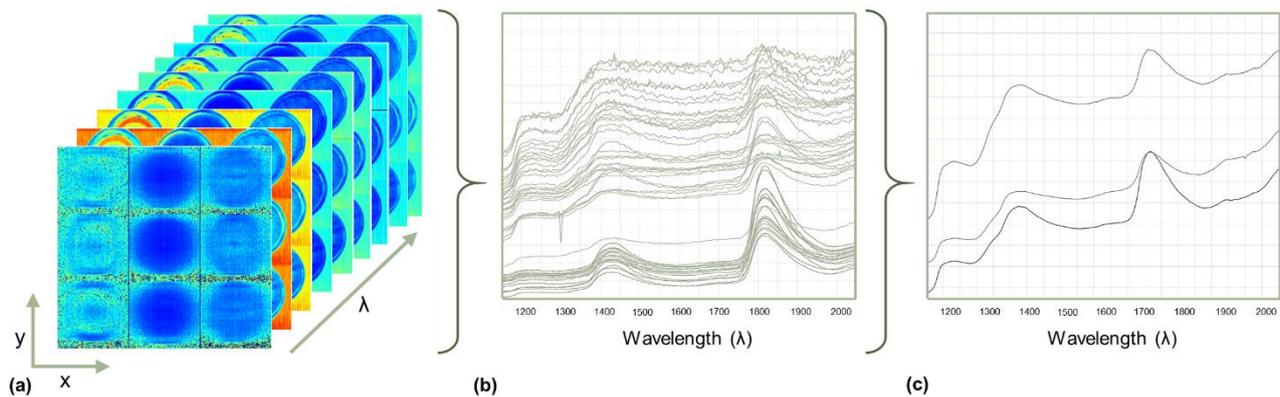


Figure 2.4 (a) An NIR hyperspectral image is a data cube that consists of two spatial (x and y) and one spectral (λ) dimension, where each pixel in the image range can be depicted as (b) spectra in the NIR region of the electromagnetic spectrum. (c) These can then be depicted as the mean spectra per object in the image.

In general, data are obtained and processed following four basic steps (1) image acquisition, (2) pre-processing, (3) multivariate image analysis and (4) prediction shown in Figure 2.5 (Cen & He, 2007; Dorrepaal *et al.*, 2016).

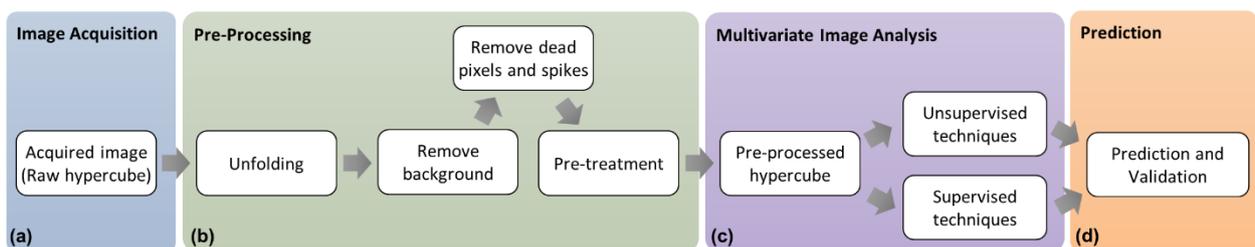


Figure 2.5 NIR hyperspectral image processing steps which includes (a) multivariate image acquisition, (b) pre-processing, (c) image analysis and (d) prediction [adapted from Dorrepaal *et al.* (2016)]

2.4.1. Image Acquisition and Instrumentation

Hyperspectral imaging systems are based on basic instrumentation, which are important for acquiring reliable, high quality images (Wu *et al.*, 2013). The instrumentation is comprised of four basic components: (1) a light source, (2) monochromator or filtering system, (3) detector and (4) controlling software.

2.4.1.1. Light source

There are two main light source configurations (Figure 2.6) used to obtain images, namely reflectance and transmittance (Burger, 2006). These configurations refer to the positioning of the light source, sample and camera. In reflectance spectroscopy, the light source is above the sample and the light is then reflected from the sample surface towards the camera (Burger, 2006). On the other hand, transmittance spectroscopy refers to the light being passed through the sample, where the light source is positioned underneath the sample (Siska & Hurburgh, 1995).

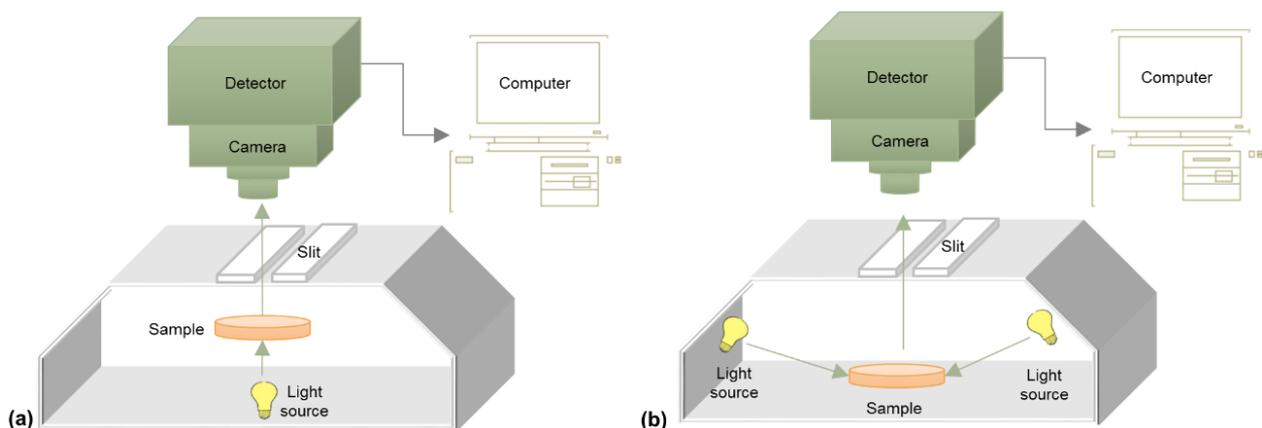


Figure 2.6 NIR hyperspectral imaging instrumentation consisting of a light source, filter, camera, detector and computer, where the light source can be positioned in either (a) reflectance or (b) transmittance mode.

Illumination or excitation sources are generally the two categories of light sources used in spectral imaging systems (Qin *et al.*, 2013). Illumination sources are generally broadband lights like tungsten-halogen (TH). TH light sources are used in many NIR imaging systems due to their high intensity in the NIR range of 900 – 2500 nm, continuous spectral outputs and low cost. The disadvantage of TH light sources is the generation of heat, which is not ideal when working with live biological or microbial samples. The spectral output may also drift with time and the operational lifetime is limited (Cen & He, 2007; Gowen *et al.*, 2015).

This is overcome by using fibre optic line lights to place distance between the sample and the TH light bulb. Lasers are used as excitation sources for hyperspectral fluorescence imaging (Kim *et al.*, 2003) as it provides monochromatic light generated in a cavity filled with gas, dye or a semi-conductor. Other excitation sources include ultraviolet (UV) and low-pressure metal vapour lamps (Qin *et al.*, 2013).

Light emitting diodes (LED), which cover a wide spectral range, have also been used as an alternative illumination light source (Gowen *et al.*, 2015). Broadband LEDs are being used in food applications as they have a low cost, low heat generation, small size and long lifespan (Qin *et al.*, 2013). Narrowband LEDs, on the other hand, can be used as excitation sources as stated by (Qin *et al.*, 2011).

2.4.1.2. Monochromator

The monochromator or filtering system is considered the core of the light dispersion set-up as it provides the ability to select, separate and measure wavelengths, as well as affect the reliability, sensitivity, implementation and calibration transferability of the final image (Reich, 2005; Burger, 2006; Qin *et al.*, 2013). These filtering systems pass light to the detector or camera (Gowen *et al.*, 2007; Boldrini *et al.*, 2012). The devices include imaging spectrographs, tunable filters and bandpass filters that separate the broadband light, required for the detection of light changes, into the different wavelengths (Gowen *et al.*, 2015). Monochromators are generally prism-grating-prisms (PGP), liquid crystal tunable filters (LCTF), acousto-optic tunable filters (AOTF), Fourier transform (FT) imaging spectrometers or single shot imagers (Wu *et al.*, 2013). The imaging spectrograph differs from the traditional spectra obtained from NIR spectroscopy as it also provides spatial information together with the spectral information (Qin *et al.*, 2013). Optical resolution is pre-determined by the monochromatic element, whereas the pixel resolution is defined by the number of pixels found in the wavelength range specified by the instrumentation. The correct choice of the optical resolution of the spectrograph and the pixel resolution of the detector is key for an optimal imaging system. These resolutions should be adapted to each specific task as to achieve the best results (Boldrini *et al.*, 2012).

2.4.1.3. Detector

The camera of a hyperspectral imaging (HSI) system generally has one of three detectors, namely: Mercury-cadmium-telluride (MCT), lead sulphide (PbS) or Indium Gallium Arsenide (InGaAs), which are sensitive in the NIR region of 1100 – 2600 nm (Gowen *et al.*, 2007; Boldrini *et al.*, 2012). A detector operating in the Visible-Near Infrared (Vis-NIR) can be a

complementary metal oxide semiconductors (CMOS) or a charge coupled device (CCD) as it works best in the range 300-1000 nm (Gómez-Sanchis *et al.*, 2008a; Gowen *et al.*, 2015). Electron multiplying CCD (EMCCD) detectors are used for microbiological samples that may be heat or photo-degradation sensitive as they decrease the image acquisition time by improving the signal to noise ratio (Gowen *et al.*, 2015).

2.4.1.4. Camera Configurations

The spatial information of a sample is obtained by controlling the sample positions or measuring the sample through the spectrometer optics. There are three basic camera configurations used to acquire a hypercube as shown in Figure 2.7.

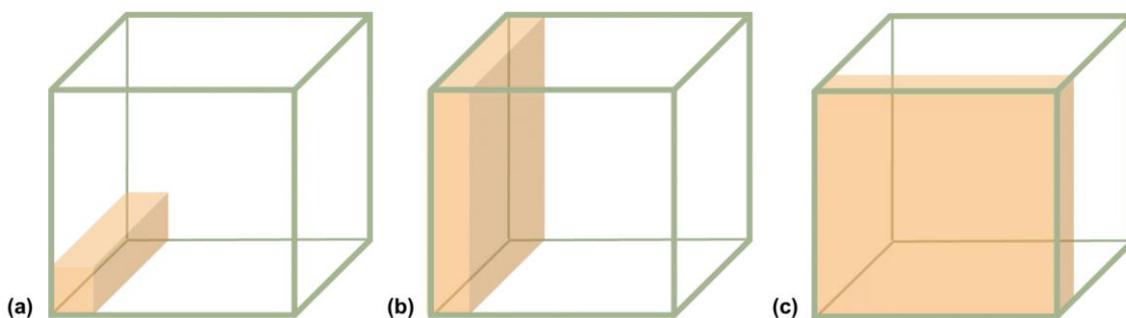


Figure 2.7 Three different scanning modes (a) point scan or whiskbroom, (b) line scan or pushbroom and (c) plane scan or stare down.

2.4.1.4.1. Point scan or “Whiskbroom”

The sample is imaged one pixel at a time, in other words a spectrum is obtained at a single spatial location (Burger, 2006). After a pixel has been imaged, the sample is repositioned underneath the detector through automatic movement of the translation stage. This is done at each spatial location until the entire sample is scanned (Gowen *et al.*, 2015). This is generally applied to Raman hyperspectral imaging and microscopic imaging. This method is the most time consuming due to the amount of spatial locations within a given sample.

2.4.1.4.2. Line scan or “Pushbroom”

Line scanning is one of the most used configurations. The system images one entire line of pixels at a time through movement between the sample and the detector (Gowen *et al.*, 2015). It is generally used in Vis-NIR and SWIR hyperspectral imaging and is faster than point scanning. This method is also ideal to implement in conveyor belt systems for in-line/on-line safety and quality checks (Gowen *et al.*, 2007).

2.4.1.4.3. Plane scan or “stare down”

The stare down imaging system collects a sequence of images of the entire sample one wavelength at a time. This is the only configuration where the sample is not repositioned for further imaging (Gowen *et al.*, 2015). As stated by Gowen *et al.* (2015), plane scanning is considered to be the only “true” imaging technique as it captures an entire image at each wavelength.

2.4.2. Pre-processing Techniques

Hyperspectral images contain sample information, as well as unwanted background noise and information (Cen & He, 2007). As a pre-cursor to multivariate image analysis (MIA), hyperspectral images first undergo pre-treatment to remove unwanted pixels, noise and unwanted background information. This is done on an unfolded two-dimensional (2D) dataset, where a three-dimensional (3D) data cube \mathbf{X} is unfolded by multiplying the x and y spatial variables and plotting them against the spectral z (λ) variable (Figure 2.8).

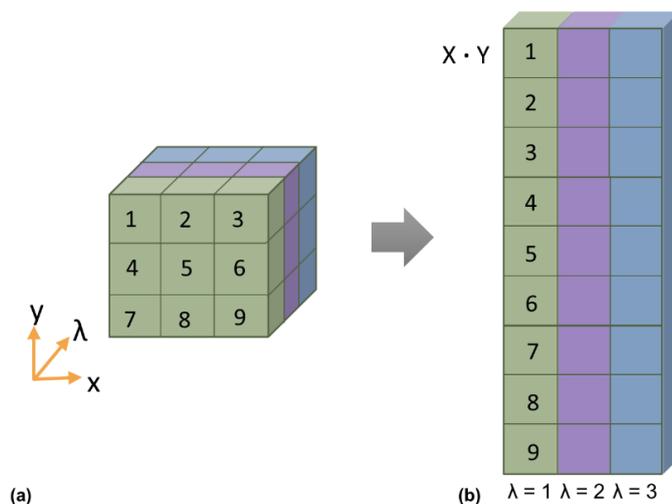


Figure 2.8 (a) The data cube that consists of two spatial (x and y) and one spectral (λ) dimension is unfolded by multiplying the two spatial locations to yield (b) a two-dimensional dataset that can be further processed.

The pre-treatment of the matrix involves selecting a region of interest (ROI), removing dead pixels and spiked points/wavelengths and spectral pre-processing (Amigo *et al.*, 2013). Region of interest (ROI) selection is the most critical part of data analysis, as it determines which data will be extracted (EIMasry & Nakauchi, 2016). Outlier and dead pixel detection is also a key component of the analysis that needs to be addressed to prevent any unnecessary errors that may negatively influence the calibration models (Burger, 2006).

Various factors can affect the spectra obtained during NIR hyperspectral imaging, such as the particle size of the sample, the light scattering, surface roughness and even the instrumentation (Vidal & Amigo, 2012). As stated by Esquerre *et al.* (2012), the morphological effect that is caused by natural variability in the size and shape of samples especially biological samples, can lead to light scattering and other spectral noise. To combat these effects, pre-processing is applied to the spectra, which reduces the spatial and spectral effects (Amigo, 2010).

There are numerous pre-processing methods available to enhance or remove certain spectral features, such as baseline correction, derivatives (Rinnan *et al.*, 2009), Fourier transforms (FT) (Young & Games, 1994; Kos *et al.*, 2002), standard normal variate (SNV) transform (Barnes *et al.*, 1989; Gowen *et al.*, 2007; Fearn *et al.*, 2009; Rinnan *et al.*, 2009; Taghizadeh *et al.*, 2010), robust normal variate (RNV) (Guo *et al.*, 1999; Zeaiter *et al.*, 2005; Rinnan *et al.*, 2009), multiplicative scatter correction (MSC) (Geladi *et al.*, 1985; Martens & Stark, 1991; Fearn *et al.*, 2009; Rinnan *et al.*, 2009), orthogonal signal correction (OSC) (Wold *et al.*, 1998; Roggo *et al.*, 2003; Luypaert *et al.*, 2004; Pizarro *et al.*, 2004), wavelet transforms (WT) (Cocchi *et al.*, 2005; Fu *et al.*, 2005) and net analyte signal (NAS) (Brás *et al.*, 2005; Cen & He, 2007). Although, the most frequently used techniques include linear transformation, scatter correction (SNV and MSC) and derivatives as they yield the best results when working with NIR hyperspectral imaging.

2.4.2.1.1. Linear Transformation

Linearization transformations, as stated by Esquerre *et al.* (2012), are often used in NIR hyperspectral imaging to increase the linear relationship between the sample concentration and the spectral data. This is done with a logarithmic transformation of the reflectance spectra, in order to obtain a better resolution. There are also other linearization transformations available, such as the Kubelka-Munk (K-M) shown in equation 2.1 (Švedas, 2004; Esquerre *et al.*, 2012).

$$K/S = (1 - R_{\infty})^2 / 2R_{\infty} \quad (\text{eq. 2.1})$$

Where:

K = K-M absorption coefficient determined by the sample's chemical composition

S = K-M scattering coefficient of the sample's physical and optical non-homogeneity

R_{∞} = Diffuse reflectance

Gómez-Sanchis *et al.* (2008) proposed that Lambertian transformation (equation 2.2) can be used as a technique to effectively correct the light scattering on curved samples and thus improve the imaging sensitivity (EIMasry & Nakauchi, 2016).

$$\rho(\lambda) = \frac{\rho_{xy}(\lambda)}{[\alpha_D \cos(\varphi) + (1 - \alpha_D)]} \quad (\text{eq. 2.2})$$

Where:

$\rho_{xy}(\lambda)$ = Reflectance of a sample at a specific point (x, y) and wavelength (λ)

α_D = Coefficient of the percentage direct and diffuse light arriving at a given sample

$\cos(\varphi)$ = Amount of direct reflected light at each point with a specific incidence angle (φ)

2.4.2.1.2. Scatter Correction

Scatter correction, a form of multiplicative transformation, is applied in order to correct the light scattering that occurs when imaging a given sample (Vidal & Amigo, 2012). It is one of the main drawbacks of imaging due to uneven sample surfaces. There are a number of scatter correction techniques, such as scaling, RNV, SNV and MSC, where SNV and MSC are the two most popular techniques (Esquerre *et al.*, 2012). SNV and MSC are applied to the spectral data in order to remove the interferences of particle size, scatter, slope variations and the light distance changes that might have occurred (Cen & He, 2007).

Scaling is based on taking an attribute, such as the minimum, maximum or mean, and dividing the individual spectra by the selected attribute in order to suppress the effects of multiplicative scattering (Esquerre *et al.*, 2012; Rinnan *et al.*, 2009). Martens and Naes (1984) first introduced MSC transformation, which is based on creating a line of best fit for the individual spectra to a reference spectrum, after which the obtained spectra are adjusted with the slope of the line of best fit. This is done in order to remove imperfections before the data is modelled. The MSC algorithm is illustrated in equation 2.3 (Fearn *et al.*, 2009).

$$z_i = (x_i - a) / b \quad (\text{eq. 2.3})$$

Where:

z_i = The transformed spectrum ($z_1, z_2, z_3, \dots, z_p$) with a given number of wavelengths (p)

x_i = The original spectrum ($x_1, x_2, x_3, \dots, x_p$) with a given number of wavelengths (p)

a = The intercept of the least-squares regression calculated with the corresponding values of the reference spectrum ($r_1, r_2, r_3, \dots, r_p$) based on the mean of all the available spectra

b = The slope of the least-squares regression calculated with the corresponding values of the reference spectrum ($r_1, r_2, r_3, \dots, r_p$) based on the mean of all the available spectra

The SNV transformation is also aimed at removing the multiplicative effects through applying mean centering to the individual spectra and then scaling the spectra by the standard deviation (Rinnan *et al.*, 2009). Taghizadeh *et al.* (2010) and Williams *et al.* (2012) both showed the use of SNV transformation improves the performance of PLS models. The SNV algorithm is illustrated in equation 2.4 (Fearn *et al.*, 2009).

$$z_i = (x_i - m) / s \quad (\text{eq. 2.4})$$

Where:

z_i = The transformed spectrum ($z_1, z_2, z_3, \dots, z_p$) with a given number of wavelengths (p)

x_i = The original spectrum ($x_1, x_2, x_3, \dots, x_p$) with a given number of wavelengths (p)

m = The calculated mean of the values expressed by $x_1, x_2, x_3, \dots, x_p$

s = The calculated standard deviation of the values expressed by $x_1, x_2, x_3, \dots, x_p$

2.4.2.1.3. Derivatives

Derivative transformation, a form of additive transformation, is used to enhance the spectral features, which emphasizes the noise level and affects the baseline and thus partially removes the unwanted noise from the spectra (Rinnan *et al.*, 2009). Smoothing techniques are generally used to reduce the effects of the baseline of the spectra, whereas the first derivative removes the additive baseline, the second derivative removes the additive and linear baseline (Rinnan *et al.*, 2009; Vidal & Amigo, 2012).

$$x'_i = x_i - x_{i-1} \quad (\text{eq. 2.5})$$

$$x''_i = x'_i - x'_{i-1} = x_{i-1} - 2 \cdot x_i + x_{i+1} \quad (\text{eq. 2.6})$$

Where:

x'_i = First derivative at a specific wavelength (i)

x''_i = Second derivative at a specific wavelength (i)

The first (equation 2.5) and second (equation 2.6) derivatives are also used to enhance the spectral resolution and effectively remove the background (Cen & He, 2007). However, it is vital that the correct parameters, i.e. polynomial order and window size, are selected. The polynomial order used, for second derivative, is generally the second or third order, with the window size being higher than the polynomial order chosen. Factors such as the spectral resolution, noise and absorption band amplitudes should also be taken into account when choosing a window size as small window sizes still have high noise levels, whereas

big window sizes potentially removes important structural information (Esquerre *et al.*, 2012). Although this method is relatively easy to apply as a form of pre-processing, it has been stated by Rinnan *et al.* (2009) that it is not always feasible due to an increase in the noise of the spectra, which should ideally be avoided in practice.

Smoothing techniques can also be applied to partially remove the unwanted noise from the spectra. One of the main smoothing algorithms is the Savitzky-Golay technique (Savitzky & Golay, 1964; Rinnan *et al.*, 2009; Vidal & Amigo, 2012). When paired with other pre-processing techniques, it can produce an optimum signal-to-noise ratio for the given spectra (Wang *et al.*, 2006; Cen & He, 2007). Once all the unwanted background information and noise have been removed, the data can be processed further by means of multivariate image analysis.

2.4.3. Multivariate Image Analysis

A hyperspectral image may be comprised of 580 x 580 pixels, which results in 336400 spectra per image (Burger, 2006). The data tends to show a high correlation in both spectral and spatial dimensions. The vast amount of data obtained from these images may deem difficult to analyse, thus various analysis tools exist to ease the process and extract useful information (Burger, 2006). Chemometric techniques can be used to explore and interpret large amounts of data. One of the advantages of hyperspectral images are the large volume of data that can be obtained and used for calibration models and training sets (Burger, 2006). This is also a disadvantage as the large volume creates opportunity for errors in either the exploration or calibration of the data. Various algorithms can be used to explore and reduce the vast amount of data into more manageable amounts, these include both unsupervised and supervised methods.

2.4.3.1. Analysis Approach

Multivariate image analysis (MIA) can be performed using two different approaches, namely pixel wise and object wise (Kucheryavskiy, 2013; Williams & Kucheryavskiy, 2016). The main differences between these two approaches are the number of data points (spectra) used during the calculations of the MIA methods. The object wise approach groups the pixels of an image into separate clusters taking the mean, or another statistic like the median, of the spectra as a data point, compared to the pixel wise approach where each pixel represents a data point with a spectrum (Kucheryavskiy, 2013).

2.4.3.2. Unsupervised Methods

Principal component analysis is one of the most popular and versatile unsupervised techniques used for the analysis of various datasets (Esbensen *et al.*, 2002a; Reich, 2005). PCA is generally used before other multivariate techniques are applied, because it has the ability to showcase the main sources of variability and distribution of elements on a single image (Gowen *et al.*, 2007; Amigo *et al.*, 2013). In simpler terms, the variables within the data set are reduced to principal component (PC) scores and loadings which explain the variance in descending order of components (Gowen *et al.*, 2007). PCA thus takes a unfolded 2D dataset (λ ; $x \cdot y$) and reconstructs the data into scores and loadings using equation 2.8, where $\mathbf{T}(x \cdot y)$ is the scores matrix, $\mathbf{P}(\lambda)$ is the loadings matrix and \mathbf{E} refers to the residual matrix (Amigo *et al.*, 2013).

$$X = TP' + E \quad (\text{eq. 2.8})$$

PCA can therefore illustrate vital information regarding the chemical compounds and physical properties and how these properties are distributed within the measured sample. Although, according to Amigo *et al.* (2013), the main drawback of using PCA is its inability to take spatial information into consideration, yet it remains one of the most popular and widespread exploratory techniques (Esbensen *et al.*, 2002a; Reich, 2005).

2.4.3.3. Supervised Methods

Supervised methods are grouped into regression and classification, where the main difference lies in the format of the output variable. Classification methods classifies a variable based on a pre-determined category, whereas regression methods yield a real value as an output, like the protein content for example.

Regression algorithms enable the user to predict the concentrations of the constituents in a sample at a pixel level. This can lead to visually mapping certain components in a sample for better understanding thereof. Some of the regression methods include partial least squares regression (PLSR), principal component regression (PCR), multiple linear regression (MLR) and stepwise linear regression (Williams, 2013).

Classification algorithms identify the objects present in a sample with the most similarity, based on their key characteristics. This is achieved by using the spectral and spatial information of each pixel of a sample image (Burger, 2006). Partial least squares

discriminant analysis (PLS-DA) and linear discriminant analysis (LDA) are two prominent classification methods used for hyperspectral image analysis.

PLS-DA can be used to qualitatively predict a set of variables within a larger dataset (Gowen *et al.*, 2007). It uses an algorithm (equation 2.9) that finds a straight line that divides a given space into two regions, thus separating the data into groups. This can also be applied to multi-class models, where separation is achieved in higher spatial planes (Musumarra *et al.*, 2007; Amigo *et al.*, 2013; Brereton & Lloyd, 2014; EIMasry & Nakauchi, 2016).

$$y = X\beta + f \quad (\text{eq. 2.9})$$

Where:

X = Unfolded datacube, or matrix, obtained from the hyperspectral image

β = Beta coefficient

f = Residual or “error” matrix

LDA is also based on finding linearity to separate objects/pixels into two or more classes, thus maximising the variance between the classes. The distance between the individual objects and the centre of the newly formed classes is then determined in order to place them in the nearest class (Singh *et al.*, 2012; Karuppiah *et al.*, 2016).

Another supervised method that is being used more frequently is support vector machines (SVM), a pattern recognition method developed by Cortes and Vapnik (1995) that is vastly different to the chemometric methods generally used in prediction/classification studies. As explained by Jiang *et al.* (2010), SVM is based on finding a hyperplane, with maximum margins between classes, to efficiently separate the largest part of the data. It was originally designed for binary classification, but multiclass SVM is now possible (Hsu & Lin, 2002). The study by Zhang *et al.* (2016) illustrated that combining SVM with NIR spectroscopy allows for the rapid identification of different teas. However, Stumpe *et al.* (2012) stated good classification results can only be obtained with SVM if the sample sizes are large and there is low variability within the classes.

Other unsupervised methods include artificial neural networks (ANN) (Mahesh *et al.*, 2008), soft independent modelling of class analogy (SIMCA) (Cocchi *et al.*, 2005; Stumpe *et al.*, 2012) and spectral angle mapping (SAM) (Bauriegel *et al.*, 2011).

2.4.4. Prediction and Validation

Prediction and validation is the last step in processing NIR hyperspectral images. By using a new image or a portion of data that has not been analysed previously, classification algorithms can be validated and unknown samples can be predicted.

2.5. APPLICATIONS OF HYPERSPECTRAL IMAGING

The applications of NIR hyperspectral imaging are broad and cover numerous fields of study. It has been increasingly used for the detection of fungal pathogens on various sample matrices, especially cereal crops (Table 2.1). Research on growth media has also been rising as it is better to initially determine whether separation between organisms is possible with more organisms being present on growth media, in comparison to food samples.

A study by Del Fiore *et al.* (2010) on the discrimination of *Aspergillus* spp., *Fusarium* spp. and *Penicillium* spp. on maize kernels and potato dextrose agar (PDA) media was done using a vis-NIR hyperspectral imaging system in the range of 400 – 1000 nm. The study showed that rapid discrimination of *Aspergillus* spp. was possible on maize kernels. The method was able to discriminate between commercial and toxigenic maize within 48 hours of inoculation. The discrimination between the different fungal species was possible using PCA and discriminant analysis (DA), however the study used plastic petri dishes, which is known to absorb in the NIR region. This could then potentially interfere with the fungal spectra obtained from a NIR hyperspectral imaging system.

A previous study by Yao *et al.* (2008) also showed that the differentiation of the toxigenic fungi *Fusarium verticillioides*, *Aspergillus* spp., *Penicillium chrysogenum* and *Trichoderma viride* was possible on PDA, using NIR hyperspectral imaging. The study, however, focussed on fungal isolates from vastly different fungal species that were not linked to a specific food crop, but rather food pathogens in general. At a later stage Yao *et al.* (2013) investigated the differentiation of *Aspergillus flavus* infection on maize kernels as a matrix. It was found that the presence of *A. flavus* infection was too low <100 ppb to yield good classification accuracy, although samples with >100 ppb showed promising results.

Williams *et al.* (2012c) found that different *Fusarium* spp. could be distinguished from one another on PDA media. The data analysis included both PCA and PLS-DA, which showed *F. verticillioides* isolates could only be distinguished with less than 47% accuracy. The study only focussed on FER of maize and not the other two types of maize ear rot commonly found in South Africa. Another study by Williams *et al.* (2012a) was done on maize kernels by investigating the fungal development of *Fusarium verticillioides* with

hyperspectral imaging. By applying both PCA and PLSR methods without pre-processing it was possible to track the development of the fungal samples over time.

Another study that used maize as a matrix is Berardo *et al.* (2005) that differentiated *Fusarium verticillioides* using modified partial least squares (MPLS). Studies on wheat were also done with Bauriegel *et al.* (2011) proving that separation between *Fusarium* spp., *Aspergillus* spp. and *Penicillium* spp. is possible with > 67% accuracy. Singh *et al.* (2012) also showed differentiation between *Aspergillus* spp. and *Penicillium* spp. on wheat kernels was possible using PCA, LDA and quadratic discriminant analysis (QDA) analysis methods. *Aspergillus* spp. and *Penicillium* spp. was also distinguished on both canola seeds (Senthilkumar *et al.*, 2015) and legumes (Karuppiyah *et al.*, 2016).

Other recent studies done on culture media with NIR hyperspectral imaging include the identification of *Lactobacillus* spp. (Foca *et al.*, 2016), the differentiation of *Bacillus cereus*, *Escherichia coli* and *Staphylococcus* spp. (Kammies *et al.*, 2016) and the classification of *Staphylococcus* spp. (Seo *et al.*, 2016).

Table 2.1 A summary of the NIR hyperspectral imaging applications focussing predominantly on fungal, as well as microbial applications

Sample matrix	Pathogen	Chemometric Technique	Analytical Technique	Wavelength range (nm)	Reference
Maize	<i>Fusarium verticillioides</i>	MPLS ^a	Vis-NIRS	400 - 1100	(Berardo <i>et al.</i> , 2005)
Maize / PDA ¹	<i>Fusarium</i> spp. and <i>Aspergillus</i> spp.	PCA ^b , DA ^c	Vis-NIR HSI	400 - 1000	(Del Fiore <i>et al.</i> , 2010)
Wheat	<i>Fusarium</i> spp., <i>Aspergillus</i> spp. and <i>Penicillium</i> spp.	PCA ^b , SAM ^d	Vis-NIR HSI	400 - 1000	(Bauriegel <i>et al.</i> , 2011)
Wheat	<i>Aspergillus</i> spp. and <i>Penicillium</i> spp.	PCA ^b , LDA ^e , QDA ^f	NIR HSI	700 - 1100	(Singh <i>et al.</i> , 2012)
Maize	<i>Fusarium verticillioides</i>	PCA ^b , PLSR ^g	NIR HSI	1000 - 2500	(Williams <i>et al.</i> , 2012a)
Wheat	Ergot (<i>Claviceps purpurea</i>)	PCA ^b , SIMCA ^h	NIR HSI	970 - 2500	(Vermeulen <i>et al.</i> , 2013)
Maize	<i>Aspergillus flavus</i>	DA ^c	Vis-NIRS	400 - 1700	(Yao <i>et al.</i> , 2013)
Canola Seeds	<i>Aspergillus</i> spp. and <i>Penicillium</i> spp.	PCA ^b , LDA ^e , QDA ^f	NIR HSI	960 - 1700	(Senthilkumar <i>et al.</i> , 2015)
Legumes / Beans	<i>Aspergillus</i> spp. and <i>Penicillium</i> spp.	QDA ^f , LDA ^e	NIR HSI	900 - 1700	(Karuppiah <i>et al.</i> , 2016)
PDA ¹	<i>Aspergillus flavus</i>	PCA ^b , SVM ⁱ	Vis-NIRS	400 - 1000	(Jin <i>et al.</i> , 2009)
PDA ¹	<i>Fusarium</i> spp.	PCA ^b , PLS-DA ^j	NIR HSI	900 - 1700	(Williams <i>et al.</i> , 2012b)
TSB ²	<i>Staphylococcus</i> spp.	PLS-DA ^j , SVM ^h	Vis-NIRS	450 - 800	(Seo <i>et al.</i> , 2016)
Nutrient agar	<i>Salmonella enteritidis</i> , <i>Bacillus cereus</i> and <i>Staphylococcus aureus</i>	PCA ^b , PLS-DA ^j	NIR HSI	1100 - 2500	(Kammies <i>et al.</i> , 2016)
MRS agar ³	<i>Lactobacillus</i> spp.	PCA ^b	NIR HSI	900 - 1700	(Foca <i>et al.</i> , 2016)

¹ Potato dextrose agar; ² Trypticase soy broth; ³ De Man, Rogosa and Sharpe agar; ^a Modified partial least squares; ^b Principal component analysis; ^c Discriminant analysis; ^d Spectral angle mapping; ^e Linear discriminant analysis; ^f Quadratic discriminant analysis; ^g Partial least squares regression; ^h Soft independent modelling of class analogy; ⁱ Support vector machines; ^j Partial least squares discriminant analysis.

2.6. CONCLUSION

The high demand placed on food safety by consumers the current maize industry practices in South African could benefit from the use of rapid screening alternatives. NIR hyperspectral imaging has the potential to act as an alternative method to speed up the identification and differentiation of pathogens on growth media. It can also be used as a non-destructive screening method for maize grain during the maize grading process. NIR hyperspectral imaging provides data rich images that, when analysed with multivariate image analysis, could be used to distinguish the major maize ear rot pathogens from one another. These include *Fusarium verticillioides*, *F. graminearum* species complex and *Stenocarpella maydis*, which lead to Fusarium-, Gibberella- and Diplodia ear rot respectively.

2.7. REFERENCES

- Amigo, J.M. (2010). Practical issues of hyperspectral imaging analysis of solid dosage forms. *Analytical and Bioanalytical Chemistry*, **398**, 93–109.
- Amigo, J.M., Martí, I. & Gowen, A. (2013). Hyperspectral Imaging and Chemometrics: A Perfect Combination for the Analysis of Food Structure, Composition and Quality. *Data Handling in Science and Technology*, **28**, 343–370.
- Aoki, T., Ward, T.J., Kistler, H.C. & O'Donnell, K. (2012). Systematics, Phylogeny and Trichothecene Mycotoxin Potential of Fusarium Head Blight Cereal Pathogens. *Mycotoxins*, **62**, 91–102.
- Ariana, D.P. & Lu, R. (2010). Hyperspectral Imaging for Defect Detection of Pickling Cucumbers. In: *Hyperspectral Imaging for Food Quality Analysis and Control* (edited by D. Sun). Pp. 431–447. Elsevier Inc.
- Ariana, D.P., Lu, R. & Guyer, D.E. (2006). Near-infrared hyperspectral reflectance imaging for detection of bruises on pickling cucumbers. *Computers and Electronics in Agriculture*, **53**, 60–70.
- Bacon, C.W., Porter, J.K., Norred, W.P. & Leslie, J.F. (1996). Production of fusaric acid by *Fusarium* species. *Applied and Environmental Microbiology*, **62**, 4039–4043.
- Barnes, R.J., Dhanoa, M.S. & Lister, S.J. (1989). Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, **43**, 772–777.

- Bauriegel, E., Giebel, A., Geyer, M., Schmidt, U. & Herppich, W.B. (2011). Early detection of *Fusarium* infection in wheat using hyperspectral imaging. *Computers and Electronics in Agriculture*, **75**, 304–312.
- Berardo, N., Pisacane, V., Battilani, P., Scandolara, A., Pietri, A. & Marocco, A. (2005). Rapid detection of kernel rots and mycotoxins in maize by near-infrared reflectance spectroscopy. *Journal of Agricultural and Food Chemistry*, **53**, 8128–8134.
- Beukes, I., Rose, L.J., Shephard, G.S., Flett, B.C. & Viljoen, A. (2017). Mycotoxigenic *Fusarium* species associated with grain crops in South Africa – A review. *South African Journal of Science*, **113**, 1–12.
- Boldrini, B., Kessler, W., Rebner, K. & Kessler, R.W. (2012). Hyperspectral imaging: A review of best practice, performance and pitfalls for in-line and on-line applications. *Journal of Near Infrared Spectroscopy*, **20**, 483–508.
- Boutigny, A., Ward, T.J., Van Coller, G.J., Flett, B., Lamprecht, S.C., O'Donnell, K. & Viljoen, A. (2011). Analysis of the *Fusarium graminearum* species complex from wheat, barley and maize in South Africa provides evidence of species-specific differences in host preference. *Fungal Genetics and Biology*, **48**, 914–920.
- Boutigny, A., Beukes, I., Small, I., Zühlke, S., Spiteller, M., Van Rensburg, B.J., Flett, B. & Viljoen, A. (2012). Quantitative detection of *Fusarium* pathogens and their mycotoxins in South African maize. *Plant Pathology*, **61**, 522–531.
- Brás, L.P., Bernardino, S.A., Lopes, J.A. & Menezes, J.C. (2005). Multiblock PLS as an approach to compare and combine NIR and MIR spectra in calibrations of soybean flour. *Chemometrics and Intelligent Laboratory Systems*, **75**, 91–99.
- Brereton, R.G. & Lloyd, G.R. (2014). Partial least squares discriminant analysis: taking the magic away. *Journal of Chemometrics*, **28**, 213–225.
- Burger, J. & Gowen, A. (2011). Data handling in hyperspectral image analysis. *Chemometrics and Intelligent Laboratory Systems*, **108**, 13–22.
- Burger, J.E. (2006). Hyperspectral NIR Image Analysis. PhD Thesis. Swedish University of Agricultural Sciences, Umeå.
- 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.
- Coblentz, W.W. (1905). Infra-Red Absorption Spectra: I. Gases. In: *The Physical Review*. Pp. 273–291.
- Cocchi, M., Corbellini, M., Foca, G., Lucisano, M., Pagani, M.A., Tassi, L. & Ulrici, A. (2005). Classification of bread wheat flours in different quality categories by a wavelet-based

- feature selection/classification algorithm on NIR spectra. *Analytica Chimica Acta*, **544**, 100–107.
- Cortes, C. & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, **20**, 273–297.
- Cozzolino, D., Smyth, H.E. & Gishen, M. (2003). Feasibility study on the use of visible and near-infrared spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins. *Journal of Agricultural and Food Chemistry*, **51**, 7703–7708.
- Del Fiore, A., Reverberi, M., Ricelli, A., Pinzari, F., Serranti, S., Fabbri, A.A., Bonifazi, G. & Fanelli, C. (2010). Early detection of toxigenic fungi on maize by hyperspectral imaging analysis. *International Journal of Food Microbiology*, **144**, 64–71.
- Department of Agriculture (2009). Regulations relating to the grading, packing and marking of maize intended for sale in the Republic of South Africa. In: Agricultural Product Standards Act (Act No. 119 of 1990).
- Department of Agriculture, Forestry and Fisheries. (2017). Summer Crops: Third Production Forecast (2017) and Winter Cereals: Intentions to plant (2017). [Internet document] URL <http://www.daff.gov.za/daffweb3/Home/Crop-Estimates>. Accessed 24/07/2017.
- Dorrepaal, R., Malegori, C. & Gowen, A. (2016). Tutorial: Time series hyperspectral image analysis. *Journal of Near Infrared Spectroscopy*, **24**, 89–107.
- EIMasry, G.M. & Nakauchi, S. (2016). Image analysis operations applied to hyperspectral images for non-invasive sensing of food quality - A comprehensive review. *Biosystems Engineering*, **142**, 53–82.
- Esbensen, K.H., Guyot, D., Westad, F. & Houmøller, L.P. (2002). Principal Component Analysis (PCA) - Introduction. In: *Multivariate data analysis in practice: an introduction to multivariate data analysis and experimental design*. 5th edition. CAMO Process AS.
- Esquerre, C., Gowen, A.A., Burger, J., Downey, G. & O'Donnell, C.P. (2012). Suppressing sample morphology effects in near infrared spectral imaging using chemometric data pre-treatments. *Chemometrics and Intelligent Laboratory Systems*, **117**, 129–137.
- Fearn, T., Riccioli, C., Garrido-Varo, A. & Guerrero-Ginel, J.E. (2009). On the geometry of SNV and MSC. *Chemometrics and Intelligent Laboratory Systems*, **96**, 22–26.
- 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.
- Foca, G., Ferrari, C., Ulrici, A., Sciutto, G., Prati, S., Morandi, S., Brasca, M., Lavermicocca, P., Lanteri, S. & Oliveri, P. (2016). The potential of spectral and hyperspectral-imaging

- techniques for bacterial detection in food: A case study on lactic acid bacteria. *Talanta*, **153**, 111–119.
- Fu, X., Yan, G., Chen, B. & Li, H. (2005). Application of wavelet transforms to improve prediction precision of near infrared spectra. *Journal of Food Engineering*, **69**, 461–466.
- Geladi, P., MacDougall, D. & Martens, H. (1985). Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. *Applied Spectroscopy*, **39**, 491–500.
- Gelderblom, W.C.A., Jaskiewicz, K., Marasas, W.F.O., Thiel, P.G., Horak, R.M., Vlegaar, R. & Kriek, N.P.J. (1988). Fumonisin - novel mycotoxins with cancer-promoting activity produced by *Fusarium moniliforme*. *Applied and Environmental Microbiology*, **54**, 1806–1811.
- Gelderblom, W.C.A., Seier, J. V., Snijman, P.W., Schalkwyk, D.J. Van, Shephard, G.S. & Marasas, W.F.O. (2001). Toxicity of culture material of *Fusarium verticillioides* strain MRC 826 to nonhuman primates. *Environmental Health Perspectives*, **109**, 267–276.
- Goetz, A.F.H., Vane, G., Solomon, J.E. & Rock, B.N. (1985). Imaging Spectrometry for Earth Remote Sensing. *Science*, **228**, 1147–1153.
- Gokul, A. (2015). *Fusarium graminearum species complex (FGSC) composition in South African wheat and maize grown in rotation*. MSc Thesis. Stellenbosch University, Stellenbosch.
- Gómez-Sanchis, J., Gómez-Chova, L., Aleixos, N., Camps-Valls, G., Montesinos-Herrero, C., Moltó, E. & Blasco, J. (2008a). Hyperspectral system for early detection of rotteness caused by *Penicillium digitatum* in mandarins. *Journal of Food Engineering*, **89**, 80–86.
- Gómez-Sanchis, J., Moltó, E., Camps-Valls, G., Gómez-Chova, L., Aleixos, N. & Blasco, J. (2008b). Automatic correction of the effects of the light source on spherical objects. An application to the analysis of hyperspectral images of citrus fruits. *Journal of Food Engineering*, **85**, 191–200.
- Gowen, A.A., Feng, Y., Gaston, E. & Valdramidis, V. (2015). Recent applications of hyperspectral imaging in microbiology. *Talanta*, **137**, 43–54.
- Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Downey, G. & Frias, J.M. (2007). Hyperspectral imaging - an emerging process analytical tool for food quality and safety control. *Trends in Food Science and Technology*, **18**, 590–598.
- Guo, Q., Wu, W. & Massart, D.L. (1999). The robust normal variate transform for pattern recognition with near-infrared data. *Analytica Chimica Acta*, **382**, 87–103.

- 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.
- Hsu, C.-W. & Lin, C. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, **13**, 415–425.
- Jiang, L., Zhu, B. & Tao, Y. (2010). Hyperspectral Image Classification Methods. In: *Hyperspectral Imaging for Food Quality Analysis and Control* (edited by D. Sun). Pp. 79–98. Elsevier Inc.
- Jin, J., Tang, L., Hruska, Z. & Yao, H. (2009). Classification of toxigenic and atoxigenic strains of *Aspergillus flavus* with hyperspectral imaging. *Computers and Electronics in Agriculture*, **69**, 158–164.
- Joffe, A.Z. (1986). *Fusarium species: Their Biology and Toxicology*. New York: Wiley.
- Kammies, T., Manley, M., Gouws, P.A. & Williams, P.J. (2016). Differentiation of foodborne bacteria using NIR hyperspectral imaging and multivariate data analysis. *Applied Microbiology and Biotechnology*, **100**, 9305–9320.
- Karuppiah, K., Senthilkumar, T., Jayas, D.S. & White, N.D.G. (2016). Detection of fungal infection in five different pulses using near-infrared hyperspectral imaging. *Journal of Stored Products Research*, **65**, 13–18.
- Kellerman, T.S., Rabie, C.J., Van der Westhuizen, G.C.A., Kriek, N.P.J. & Prozesky, L. (1985). Induction of diplodiosis, a neuromycotoxicosis, in domestic ruminants with cultures of indigenous and exotic isolates of *diplodia maydis*. *Onderstepoort Journal of Veterinary Research*, **52**, 35–42.
- Kim, M.S., Lefcourt, A.M. & Chen, Y.R. (2003). Multispectral laser-induced fluorescence imaging system for large biological samples. *Applied Optics*, **42**, 3927–3934.
- Kos, G., Lohninger, H. & Krska, R. (2002). Fourier transform mid-infrared spectroscopy with attenuated total reflection (FT-IR/ATR) as a tool for the detection of *Fusarium* fungi on maize. *Vibrational Spectroscopy*, **29**, 115–119.
- Kucheryavskiy, S. (2013). A new approach for discrimination of objects on hyperspectral images. *Chemometrics and Intelligent Laboratory Systems*, **120**, 126–135.
- Lanubile, A., Ferrarini, A., Maschietto, V., Delledonne, M., Marocco, A. & Bellin, D. (2014). Functional genomic analysis of constitutive and inducible defense responses to *Fusarium verticillioides* infection in maize genotypes with contrasting ear rot resistance. *BMC Genomics*, **15**, 710.

- Lanza, E. & Li, B.W. (1984). Application for near infrared spectroscopy for predicting the sugar content of fruit juices. *Journal of Food Science*, **49**, 995–998.
- Luybaert, J., Heuerding, S., Van der Heyden, Y. & Massart, D.L. (2004). The effect of preprocessing methods in reducing interfering variability from near-infrared measurements of creams. *Journal of Pharmaceutical and Biomedical Analysis*, **36**, 495–503.
- Mahesh, S., Manickavasagan, A., Jayas, D.S., Paliwal, J. & White, N.D.G. (2008). Feasibility of near-infrared hyperspectral imaging to differentiate Canadian wheat classes. *Biosystems Engineering*, **101**, 50–57.
- Manley, M., Williams, P., Nilsson, D. & Geladi, P. (2009). Near infrared hyperspectral imaging for the evaluation of endosperm texture in whole yellow maize (*Zea mays* L.) kernels. *Journal of Agricultural and Food Chemistry*, **57**, 8761–8769.
- Martens, H. & Naes, T. (1984). Multivariate calibration. I. Concepts and distinctions. *Trends in Analytical Chemistry*, **3**, 204–210.
- Martens, H. & Stark, E. (1991). Extended multiplicative signal correction and spectral interference subtraction: New preprocessing methods for near infrared spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis*, **9**, 625–635.
- Mehl, P.M., Chen, Y.R., Kim, M.S. & Chan, D.E. (2004). Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. *Journal of Food Engineering*, **61**, 67–81.
- Menesatti, P., Costa, C. & Aguzzi, J. (2010). Quality evaluation of fish by hyperspectral imaging. In: *Hyperspectral Imaging for Food Quality Analysis and Control* (edited by D. Sun). Pp. 273–294. Elsevier Inc.
- Menesatti, P., Zanella, A., D'Andrea, S., Costa, C., Paglia, G. & Pallottino, F. (2009). Supervised multivariate analysis of hyperspectral NIR images to evaluate the starch index of apples. *Food and Bioprocess Technology*, **2**, 308–314.
- Molto, E., Blasco, J. & Gomez-Sanchis, J. (2010). Analysis of hyperspectral images of citrus fruits. In: *Hyperspectral Imaging for Food Quality Analysis and Control* (edited by S. De Saeger). Pp. 321–348. Elsevier Inc.
- Munkvold, G.P. (2003). Cultural and genetic approaches to managing mycotoxins in maize. *Annual Review of Phytopathology*, **41**, 99–116.
- Musumarra, G., Trovato-Salinaro, A., Scirè, S., Foti, A., Barresi, V., Fortuna, C.G., Strazzulla, G., Condorelli, D.F. (2007). Identification of genes involved in radiation-induced G1 arrest. *Journal of Chemometrics*, **21**, 398–405.

- Odrizola, E., Odeón, A, Canton, G., Clemente, G. & Escande, A. (2005). *Diplodia maydis*: a cause of death of cattle in Argentina. *New Zealand Veterinary Journal*, **53**, 160–161.
- Pannar. (2017). Maize Ear Rot [Internet document] URL <http://www.pannar.com/diseases#>. Accessed 27/11/2017.
- Peng, Y., Zhang, J., Wang, W., Li, Y., Wu, J., Huang, H., Gao, X. & Jiang, W. (2011). Potential prediction of the microbial spoilage of beef using spatially resolved hyperspectral scattering profiles. *Journal of Food Engineering*, **102**, 163–169.
- 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.
- Prats-Montalbán, J.M., De Juan, A. & Ferrer, A. (2011). Multivariate image analysis: A review with applications. *Chemometrics and Intelligent Laboratory Systems*, **107**, 1–23.
- Qin, J., Chao, K., Kim, M.S., Kang, S., Cho, B. & Jun, W. (2011). Detection of organic residues on poultry processing equipment surfaces by LED-induced fluorescence imaging. *Applied Engineering in Agriculture*, **27**, 153–161.
- Qin, J., Chao, K., Kim, M.S., Lu, R. & Burks, T.F. (2013). Hyperspectral and multispectral imaging for evaluating food safety and quality. *Journal of Food Engineering*, **118**, 157–171.
- Ravikanth, L., Chelladurai, V., Jayas, D.S. & White, N.D.G. (2016). Detection of broken kernels content in bulk wheat samples using near-infrared hyperspectral imaging. *Agricultural Research*, **5**, 285–292.
- Reich, G. (2005). Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Advanced Drug Delivery Reviews*, **57**, 1109–1143.
- Rinnan, A., Van Den Berg, F. & Engelsen, S.B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, **28**, 1201–1222.
- Roggo, Y., Duponchel, L., Ruckebusch, C. & Huvenne, J.P. (2003). Statistical tests for comparison of quantitative and qualitative models developed with near infrared spectral data. *Journal of Molecular Structure*, **654**, 253–262.
- Rossouw, J.D., Pretorius, Z.A., Silva, H.D. & Lamkey, K.R. (2009). Breeding for Resistance to *Stenocarpella* Ear Rot in Maize. In: *Plant Breeding Reviews*. Pp. 223–245.
- Rossouw, J.D., Rensburg, J.B.J. van & Deventer, C.S. van. (2002). Breeding for resistance to ear rot of maize, caused by *Stenocarpella maydis* (Berk) Sutton. 1. Evaluation of selection criteria. *South African Journal of Plant and Soil*, **19**, 182–187.

- Saccon, F.A.M., Elrewainy, A., Parcey, D., Paliwal, J. & Sherif, S.S. (2005). Detection of *Fusarium* on wheat using near infrared hyperspectral imaging, *Photonics North*, **43**, 3118.
- Savitzky, A. & Golay, M.J.E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, **36**, 1627–1639.
- Scotter, C.N.G. (1997). Non-destructive spectroscopic techniques for the measurement of food quality. *Trends in Food Science & Technology*, **8**, 285–292.
- Schoeman, A. and Flett, B.C. (2012). Diplodia and stalk rot of maize in the spotlight. SA Grain. Online publication: www.grainsa.co.za/diplodia-ear-and-stalk-rot-of-maize-in-the-spotlight. (20 December 2016).
- Sendin, K. (2017). *Characterisation of whole white maize kernels*. MSc Thesis. Stellenbosch University, Stellenbosch.
- Senthilkumar, T., Jayas, D.S. & White, N.D.G. (2015). Detection of different stages of fungal infection in stored canola using near-infrared hyperspectral imaging. *Journal of Stored Products Research*, **63**, 80–88.
- Senthilkumar, T., Singh, C.B., Jayas, D.S. & White, N.D.G. (2012). Detection of Fungal Infection in Canola using Near-Infrared Hyperspectral Imaging, *Journal of Agricultural Engineering*, **49**, 21–27.
- Seo, Y., Park, B., Hinton, A., Yoon, J.S. & Lawrence, K.C. (2016). Identification of Staphylococcus species with hyperspectral microscope imaging and classification algorithms. *Journal of Food Measurement and Characterization*, **10**, 253–263.
- Singh, C.B., Jayas, D.S., Paliwal, J. & White, N.D.G. (2012). Fungal Damage Detection in Wheat Using Short-Wave Near-Infrared Hyperspectral and Digital Colour Imaging. *International Journal of Food Properties*, **15**, 11–24.
- Siska, J. & Hurburgh, C.R. (1995). Corn density measurement by near-infrared transmittance. *Transactions of the ASAE*, **38**, 1821–1824.
- Snyder, W.C. & Hansen, H.N. (1940). The Species Concept in *Fusarium*. *American Journal of Botany*, **27**, 64–67.
- Strausbaugh, C.A., Overturf, K. & Koehn, A.C. (2005). Pathogenicity and real-time PCR detection of *Fusarium* spp. in wheat and barley roots. *Canadian Journal of Plant Pathology*, **27**, 430–438.
- Stumpe, B., Engel, T., Steinweg, B. & Marschner, B. (2012). Application of PCA and SIMCA statistical analysis of FT-IR spectra for the classification and identification of different slag types with environmental origin. *Environmental Science and Technology*, **46**, 3964–3972.

- South African Gazette. Foodstuffs, Cosmetics and Disinfectants Act (54/1972): Regulations governing tolerances for fungus-produced toxins in foodstuffs: Amendment [document on the Internet]. c2016 [cited 2016 Dec 07]. Available from: https://www.greengazette.co.za/documents/national-gazette-40250-of-05-september-2016-vol-615_20160905-GGN-40250.pdf
- Švedas, V. (2004). Spectral pre-treatment for diffuse transmittance linearity improvement. *Journal of Near Infrared Spectroscopy*, **12**, 347–358.
- Taghizadeh, M., Gowen, A., Ward, P. & O'Donnell, C.P. (2010). Use of hyperspectral imaging for evaluation of the shelf-life of fresh white button mushrooms (*Agaricus bisporus*) stored in different packaging films. *Innovative Food Science and Emerging Technologies*, **11**, 423–431.
- ThermoFischer Scientific. (2017). Overview of ELISA [Internet document]. URL <https://www.thermofisher.com/za/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-elisa.html>. Accessed 27/09/2017.
- Turner, N.W., Bramhmbhatt, H., Szabo-Vezse, M., Poma, A., Coker, R. & Piletsky, S.A. (2015). Analytical methods for determination of mycotoxins: An update (2009-2014). *Analytica Chimica Acta*, **901**, 12–33.
- Velusamy, V., Arshak, K., Korostynska, O., Oliwa, K. & Adley, C. (2010). An overview of foodborne pathogen detection: In the perspective of biosensors. *Biotechnology Advances*, **28**, 232–254.
- Vermeulen, P., Fernández Pierna, J.A., Van Egmond, H.P., Zegers, J., Dardenne, P. & Baeten, V. (2013). Validation and transferability study of a method based on near-infrared hyperspectral imaging for the detection and quantification of ergot bodies in cereals. *Analytical and Bioanalytical Chemistry*, **405**, 7765–7772.
- Vidal, M. & Amigo, J.M. (2012). Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemometrics and Intelligent Laboratory Systems*, **117**, 138–148.
- Wagacha, J.M. & Muthomi, J.W. (2008). Mycotoxin problem in Africa: Current status, implications to food safety and health and possible management strategies. *International Journal of Food Microbiology*, **124**, 1–12.
- Van Der Walt, S. & Steyn, D.G. (1943). Recent Investigations into the Toxicity of Plants, etc., XIII. *Onderstepoort Journal of Veterinary Science*, **18**, 207–224.

- Wang, L., Lee, F.S.C., Wang, X. & He, Y. (2006). Feasibility study of quantifying and discriminating soybean oil adulteration in camellia oils by attenuated total reflectance MIR and fiber optic diffuse reflectance NIR. *Food Chemistry*, **95**, 529–536.
- Wicklow, D.T., Rogers, K.D., Dowd, P.F. & Gloer, J.B. (2011). Bioactive metabolites from *Stenocarpella maydis*, a stalk and ear rot pathogen of maize. *Fungal Biology*, **115**, 133–142.
- Williams, P.J. (2009). *Near infrared (NIR) hyperspectral imaging for evaluation of whole maize kernels: Chemometrics for exploration and classification*. MSc Thesis. Stellenbosch University, 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. Stellenbosch University, Stellenbosch.
- Williams, P.J., Geladi, P., Britz, T.J. & Manley, M. (2012a). Investigation of fungal development in maize kernels using NIR hyperspectral imaging and multivariate data analysis. *Journal of Cereal Science*, **55**, 272–278.
- Williams, P.J., Geladi, P., Britz, T.J. & Manley, M. (2012b). Growth characteristics of three *Fusarium* species evaluated by near-infrared hyperspectral imaging and multivariate image analysis. *Applied Microbiology and Biotechnology*, **96**, 803–813.
- Williams, P.J., Geladi, P., Britz, T.J. & Manley, M. (2012c). Near-infrared (NIR) hyperspectral imaging and multivariate image analysis to study growth characteristics and differences between species and strains of members of the genus *Fusarium*. *Analytical and Bioanalytical Chemistry*, **404**, 1759–1769.
- Williams, P.J. & Kucheryavskiy, S. (2016). Classification of maize kernels using NIR hyperspectral imaging. *Food Chemistry*, **209**, 131–138.
- Wold, S., Antti, H., Lindgren, F. & Öhman, J. (1998). Orthogonal signal correction of near-infrared spectra. *Chemometrics and Intelligent Laboratory Systems*, **44**, 175–185.
- Woloshuk, C. & Wise, K. (2014). Diseases of Corn: *Fusarium* Ear Rot. *Purdue Extension*, 2–4.
- Wu, D., Wang, S., Wang, N., Nie, P., He, Y., Sun, D.W. & Yao, J. (2013). Application of time series hyperspectral imaging (TS-HSI) for determining water distribution within beef and spectral kinetic analysis during dehydration. *Food and Bioprocess Technology*, **6**, 2943–2958.
- Yao, H., Hruska, Z., Kincaid, R., Brown, R.L., Bhatnagar, D. & Cleveland, T.E. (2013). Detecting maize inoculated with toxigenic and atoxigenic fungal strains with fluorescence hyperspectral imagery. *Biosystems Engineering*, **115**, 125–135.

- Yao, H., Hruska, Z., Kincaid, R., Brown, R.L. & Cleveland, T.E. (2008). Differentiation of toxigenic fungi using hyperspectral imagery. *Sensing and Instrumentation for Food Quality and Safety*, **2**, 215–224.
- Young, J.C. & Games, D.E. (1994). Analysis of *Fusarium* mycotoxins by gas chromatography-Fourier transform infrared spectroscopy. *Journal of Chromatography A*, **663**, 211–218.
- Zeaiter, M., Roger, J.M. & Bellon-Maurel, V. (2005). Robustness of models developed by multivariate calibration. Part II: The influence of pre-processing methods. *Trends in Analytical Chemistry*, **24**, 437–445.
- Zhang, Z., Hu, X., Zhang, Q. & Li, P. (2016). Determination for multiple mycotoxins in agricultural products using HPLC-MS/MS via a multiple antibody immunoaffinity column. *Journal of Chromatography B*, **1021**, 145–152.

CHAPTER 3

The differentiation of maize ear rot pathogens, on growth media, using near infrared hyperspectral imaging

3.1. ABSTRACT

Near infrared (NIR) hyperspectral imaging can be used as a rapid screening tool that distinguishes fungal pathogens in the laboratory. This study focussed on determining the possibility of differentiating the major maize ear rot pathogens from one another, on growth media. The fungal isolates of four pathogens, commonly associated with maize grain [*Fusarium verticillioides*, *F. graminearum* s.s., *F. boothii* and *Stenocarpella maydis*], were plated on potato dextrose agar, in triplicate, and incubated at 25°C for 5 days. Images were collected with a SisuChema short-wave infrared (SWIR) pushbroom hyperspectral imaging system in the range of 1000 to 2500 nm. Pixel and object wise classification algorithms were compared to determine the best classification method as determined by principal component analysis and partial least squares discriminant analysis. The pathogens were all distinguished using three-way classification models that were validated with new images. The object wise classification proved more effective in distinguishing ear rot pathogens with a higher average overall classification accuracy (93.75 %) for the pathogen models. The object wise classification models achieved two 100 % classification accuracies from the four pathogen models, while none of the pixel wise classification models were error free. The F1 score, together with the specificity and sensitivity further highlighted the superiority of object wise classification. The high accuracy of object wise classification (> 86 %) can be attributed to the representation of the mean spectra per object. NIR hyperspectral imaging can thus accurately distinguish between the major maize ear rot pathogens, with object wise classification proving to be the optimal method.

3.2. INTRODUCTION

In South Africa, maize (*Zea mays* L.) is considered the most important cereal grain, as it is one of the main dietary staple foods with consumption as high as 400 to 500 g per person per day (Shephard, 2008). However, it is prone to infection by pathogenic organisms, which reduce the overall crop yields and/or contaminate the maize with mycotoxins, leading to various human and animal diseases.

The maize industry contributed 66% of the total South African cereals produced in 2016. The maize yield for 2017 is estimated at more than 14 million tons of commercial maize, which includes both white and yellow maize (Department of Agriculture, Forestry and Fisheries, 2017). These statistics illustrate the importance of maize in South African agriculture, which means that the highest possible yield should be obtained from each maize crop. Crop losses occur due to physical plant damage or disease, which can be attributed to various external factors such as weather changes and infestation, to name a few.

As stated by Munkvold (2003), there are two diseases, predominantly found in South Africa and caused by *Fusarium* spp., that affect the yield and quality of maize, namely Fusarium ear rot (FER) and Gibberella ear rot (GER). FER is primarily caused by *Fusarium verticillioides*, as well as *F. subglutinans* and *F. proliferatum* (Munkvold, 2003; Boutigny *et al.*, 2012). These pathogens result in light pink or white mould formation on the maize ears (Beukes *et al.*, 2017). *Fusarium verticillioides* (syn: *F. moniliforme*) produces fumonisins, a mycotoxin that has been associated with oesophageal cancer in humans and rats (Yao *et al.*, 2008), leukoencephalomalacia in horses and pulmonary edema in pigs (Gelderblom *et al.*, 2001). The pathogen, *F. graminearum*, which is a species complex that includes *F. graminearum* s.s. and *F. boothii*, among others, is the leading cause of GER in maize. Boutigny *et al.* (2012) determined that *F. boothii* were exclusively associated with maize ears showing symptoms of GER, while Gokul (2015) found *F. graminearum* s.s. to be the most abundant species of the *Fusarium graminearum* species complex (FGSC) on maize ears where maize and wheat were rotated. GER results in pink or red mould, initiating at the tip of the maize ear and spreading until it covers the entire ear (Munkvold, 2003; Beukes *et al.*, 2017). GER, caused by the FGSC, predominantly produces the deoxynivalenol (DON) toxin, a member of the trichothecene group which suppresses the immune system and leads to nausea and vomiting (Peraica *et al.*, 1999). *Stenocarpella maydis* is another plant pathogen that affects the yield and quality of maize. The pathogen causes Diplodia ear rot (DER) in maize (Rossouw *et al.*, 2002). DER infects the maize ear at the base and gradually

moves upwards forming a white mould between the kernels and turning the ear a greyish-brown colour (EPPO, 2017).

The *Fusarium* genus is known to be difficult to distinguish due to the taxonomy thereof (Snyder & Hansen, 1940; Williams, 2013). It is important to be able to easily distinguish between the species due to maize being highly susceptible to these mycotoxin-producing pathogens. Molecular and immunological detection methods such as the polymerase chain reaction (PCR) (Strausbaugh *et al.*, 2005) and enzyme-linked immunosorbent assay (ELISA) have been used for the identification and characterisation of the pathogens. PCR uses the DNA of a given sample in order to identify it through a series of steps, which include DNA extraction, PCR primer and probe design, as well as PCR and electrophoresis (Schilling *et al.*, 1996). A sample is initially plated on an agar plate and incubated for five days under controlled conditions, after which the DNA is extracted. ELISA, on the other hand, relies on antibodies, peptides and proteins for the identification of samples. Enzymes are linked to sample antibodies and identification is based on the conjugated enzyme activity observed during this process (ThermoFischer Scientific, 2017).

These multi-step methods are tedious, time-consuming and can require specialised growth media, supplementary staining techniques and microscopy (Turner *et al.*, 2015). The identification can take more than two weeks, where the fate of a cereal batch is only determined once identification is complete, thus increasing the chance of toxin build-up within the batch (Velusamy *et al.*, 2010).

Ensuring the safety of food consumed by humans and animals can be achieved by combining microbiology, plant pathology and chemistry disciplines to create a rapid pathogen screening method, which accurately distinguishes the major ear rot pathogens on growth media. Near infrared (NIR) spectroscopy, as well as hyperspectral imaging offers the ability to achieve this goal.

NIR hyperspectral imaging offers rapid sample measurement. The technique is used for chemical analysis, as it can extract specific chemical and structural information from a given sample (Burger, 2006). Hyperspectral imaging is capable of producing images with both spectral and spatial resolution and captures it at many wavelength bands in the NIR region, between 780 and 2500 nm (Yao *et al.*, 2008). The images are composed of different wavebands with spatial positions, where each pixel contains the spectrum of that exact spatial location, thus forming a three-dimensional (3D) datacube, also known as a hypercube. Each image contains millions of data points (spectra), across a range of wavelengths that is obtained within seconds of imaging. Although the analysis of these images are problematic, due to the size of the datasets, multivariate image analysis (MIA)

has been proven to remedy the problem (Prats-Montalbán *et al.*, 2011). MIA is used for images that contain more than one measurement per pixel and can be applied to classification, detection or prediction models.

Principal component analysis (PCA) is used as an initial step to reduce dimensionality and analyse the data (Esbensen *et al.*, 2002b). As hyperspectral images are hypercubes, the data needs to be unfolded by multiplying the x and y data points and plotting them against the wavelengths. PCA can then be used to take the two-dimensional (2D) dataset (λ ; x.y) and reconstruct the data into scores and loadings. Partial least squares discriminant analysis (PLS-DA) is a supervised classification technique that can be used to qualitatively predict a set of variables within a larger dataset (Gowen *et al.*, 2007). The analysis methods, PCA and PLS-DA, enable rapid screening of samples for the detection of pathogens by building calibrated prediction models that can predict unknown samples. Both PCA and PLS-DA can be performed using a pixel or object wise approach (Williams & Kucheryavskiy, 2016). The main differences between the two approaches are the number of data points (spectra) used for the calculations. The object wise approach groups the pixels of an image into separate clusters taking the mean, or other statistics such as the median of the spectra as a data point, compared to the pixel wise approach where each pixel represents a data point with a spectrum (Kucheryavskiy, 2013).

NIR hyperspectral imaging was used investigate fungal pathogens on both growth media and other matrices. Yao *et al.* (2008) showed that the differentiation of toxigenic fungi is possible on growth media, using NIR hyperspectral imaging with a classification accuracy of 97.7%. The study, however, focussed on fungal isolates from vastly different fungal species, including *Penicillium* spp., *Fusarium* spp. and *Aspergillus* spp. A study on the discrimination of *Aspergillus* spp., *Fusarium* spp. and *Penicillium* spp. on PDA media was done by Del Fiore *et al.* (2010). They showed that discrimination between the different fungal species was possible using PCA and discriminant analysis (DA), however they used plastic petri dishes, which is known to absorb in the NIR region. Williams *et al.* (2012a) found that *Fusarium* spp. could be distinguished from one another on PDA media. The data analysis included both PCA and PLS-DA, which showed *F. verticillioides* isolates could only be distinguished with less than 47% accuracy. The study only focussed on FER of maize and not the other two types of maize ear rot pathogens. Other recent studies done on culture media with NIR hyperspectral imaging include the identification of *Lactobacillus* spp. (Foca *et al.*, 2016), the differentiation of *Bacillus cereus*, *Escherichia coli* and *Staphylococcus* spp. (Kammies *et al.*, 2016) and the classification of *Staphylococcus* spp. (Seo *et al.*, 2016).

The aim of this study was thus to differentiate between the major maize ear rot pathogens on growth media, using NIR hyperspectral imaging, to decrease laboratory testing time and deduce what possible mycotoxins may be present within a cereal batch.

3.3. MATERIALS AND METHODS

3.3.1. Sample Preparation

Isolates of the fungal pathogens, *Fusarium* spp. and *Stenocarpella* spp., most commonly associated with maize ear rot were provided by the Department of Plant Pathology at Stellenbosch University, South Africa. *Fusarium* species are stored at -80 °C in 15% glycerol while *S. maydis* are stored in water at room temperature. The strains used, include *Fusarium verticillioides* MRC826, MRC8267 and MRC8559, *F. graminearum* s.s. M14-55, *F. boothii* M0002, M0010 and M0100, as well as *Stenocarpella maydis* SM8 (Table 3.1).

Table 3.1 The origin of the *Fusarium* spp. and *Stenocarpella maydis* isolates investigated.

Isolate	Species	Province	Year	Mycotoxin*	Reference
MRC826	<i>F. verticillioides</i>	Transkei	2005	FUM	Marasas <i>et al.</i> (1986)
MRC8267	<i>F. verticillioides</i>	North Benin	-	FUM	Williams (2013)
MRC8559	<i>F. verticillioides</i>	California, USA	-	FUM	Nirenberg & O'Donnell (1998)
M14-55	<i>F. graminearum</i> s.s.	Limpopo	2014	DON	Gokul (2015)
M0002	<i>F. boothii</i>	Free State	2008	DON	Boutigny <i>et al.</i> (2011)
M0010	<i>F. boothii</i>	Mpumalanga	2008	DON	Boutigny <i>et al.</i> (2011)
M0100	<i>F. boothii</i>	North West	2010	DON	Boutigny <i>et al.</i> (2011)
SM8	<i>S. maydis</i>	Potchefstroom	2012	DIP	Rabie <i>et al.</i> (1977)

* FUM = Fumonisin; DON = Deoxynivalenol; DIP = Diplodiatoxin

The fungi were plated onto potato dextrose agar in individual glass Petri dishes. The plates were inoculated by placing a piece of previously inoculated PDA in the centre of the agar filled plates (Figure 3.1).

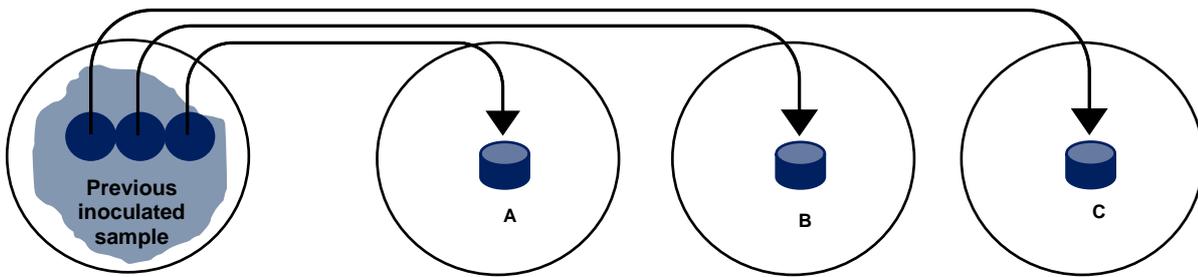


Figure 3.1 Schematic illustrating the inoculation of PDA glass plates with the pathogen isolates.

The pathogen isolates were plated out in triplicate, and incubated at 25 °C for five days as this is the time generally used in industry. Digital images of the pathogen isolates were taken on day 5 for comparison during the modelling (Figure 3.2).

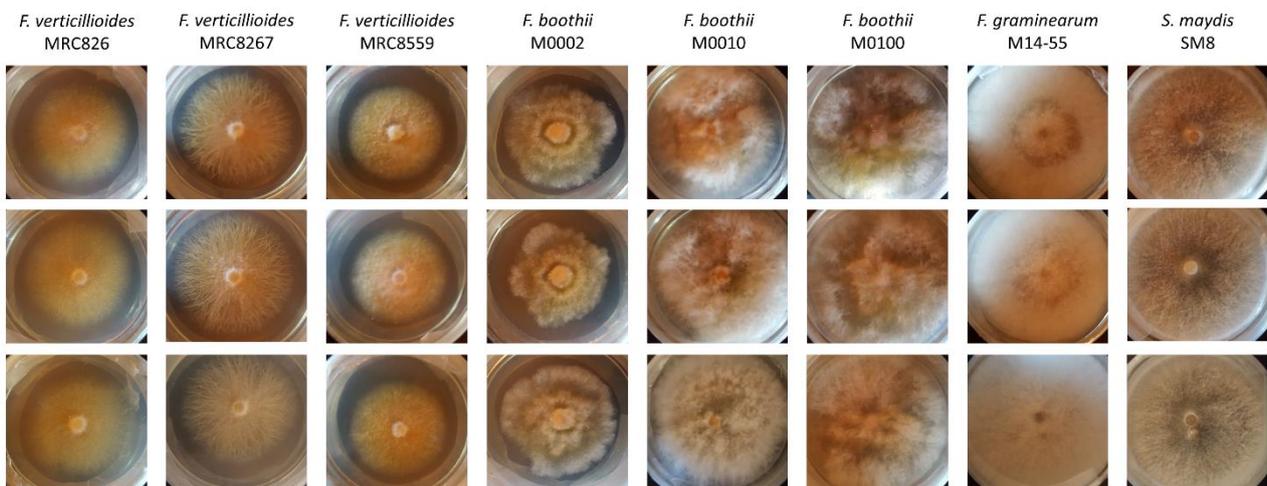


Figure 3.2 Digital images of fungal isolates of *F. verticillioides*, *F. graminearum* s.s., *F. boothii* and *S. maydis* isolates, after 5 days, used in the study.

3.3.2. NIR Hyperspectral Imaging System

The hyperspectral images were collected with a pushbroom SisuChema short-wave infrared (SWIR) imaging system (Specim Spectral Imaging Ltd, Oulu, Finland) equipped with the Chemadaq version 3.62.183.19 controlling software. The hyperspectral imaging system consisted of a spectrograph connected to a 2-D array mercury-cadmium-telluride (HgCdTe) detector with quartz halogen lamps. The camera was equipped with a 31 mm lens giving a 100 mm field of view. A spatial resolution of 300 µm and spectral range of 920 – 2514 nm was used, with a 10 nm resolution and 6.3 nm spectral sampling per pixel. The frame rate was 100 Hz at an exposure of 3.5 ms and 30% scanning rate. The entire Petri dish was imaged without removing the lid after the 5-day growth period. Before each sample was

imaged, 100% reflectance (white) and closed shutter (dark) reference standards were obtained. The final images consisted of 320 spatial pixels on the x-axis and 256 spectral bands.

3.3.3. Hyperspectral Image Analysis

The hyperspectral images were analysed with Prediktera Evince v.2.7.0 (Umeå, Sweden), a hyperspectral image analysis software. Selected pathogen isolates were mosaicked, in other words merged together to form the calibration and validation sets used during the study. The mosaics were analysed individually, either by pixel or object wise analysis.

3.3.3.1. Mosaic Construction

Individual images, within the wavelength range 920 to 2483 nm, of selected pathogen isolates were merged to form mosaics. The first mosaic consisted of *Fusarium verticillioides* MRC826, *F. graminearum* s.s. M14-55 and *Stenocarpella maydis* SM8, which represents the three major ear rots, namely FER, GER and DER. The second mosaic was comprised of *F. verticillioides* MRC826, *F. boothii* M0002 and *S. maydis* SM8, where the GER-causing pathogen was replaced. Mosaic three consisted of the three *F. verticillioides* isolates, MRC826, MRC8267 and MRC8559. This was done to establish whether it was possible to distinguish between isolates from the same species. The same was done in mosaic four, where the *F. boothii* isolates were investigated, namely M0002, M0010 and M0100.

3.3.3.2. Image Correction and Cleaning

The image calibration and absorbance correction was done per equation 3.1 (Sendin, 2017) in the Evince software package.

$$I_{\lambda,n} = -\log_{10} \left[\left(\frac{S_{\lambda,n} - B_{\lambda,n}}{W_{\lambda,n} - B_{\lambda,n}} \right) \right] \quad (\text{eq. 3.1})$$

Where:

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

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

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

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

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

The data was mean-centred, prior to the calculation of the principal components. The score images and score plots were interactively used, together with the digital images (Figure 3.2), to select the regions of interest. The background removal was an iterative process. For the pixel wise analysis, a circular region was selected and the agar plug used for inoculation was removed from the images as it contributes to unnecessary noise and light scattering. The images used during the object wise analysis were each further segmented into 12 circular areas as to increase the calibration and validation sets. The PCA was then recalculated on the segmented images. The image correction and cleaning stages for both the pixel and object wise analysis are shown in Figure 3.3 and Figure 3.4, respectively.

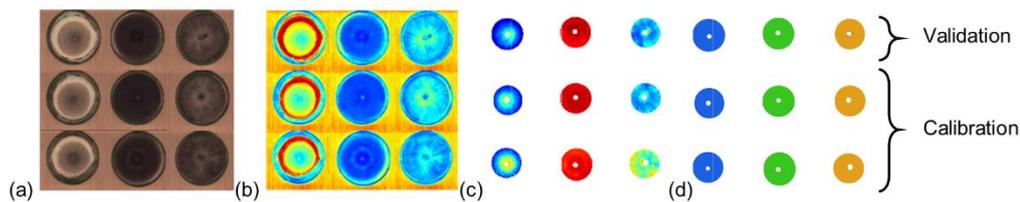


Figure 3.3 Step-wise approach used for the pixel wise analysis of the various models illustrating (a) the RGB image of the mosaic of *F. verticillioides*, *F. graminearum* s.s. and *S. maydis* triplicates, (b) the uncleaned PC score image before background removal, (c) the PC score image of the isolates and (d) the PLS-DA score image illustrating the calibration and validation sets.

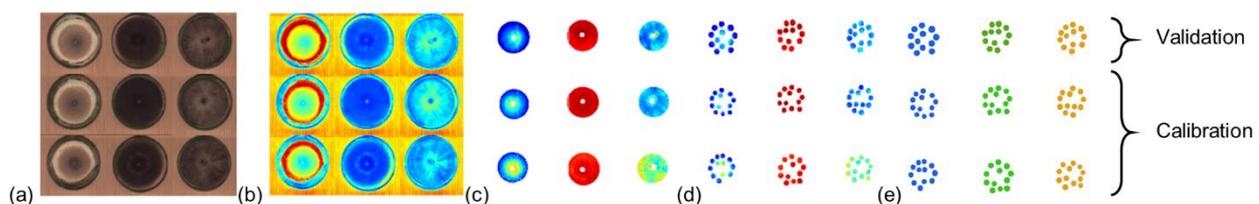


Figure 3.4 Step-wise analysis used for the object wise analysis of the various models illustrating (a) the RGB image of the mosaic of *F. verticillioides*, *F. graminearum* s.s. and *S. maydis* triplicates, (b) the uncleaned PC score image before background removal, (c) the PC score image before the subdivision, (d) the PC score image of the subdivided isolates and (e) the PLS-DA score image illustrating the calibration and validation sets.

3.3.3.3. Spectral Analysis

The raw spectra of the hyperspectral images were investigated to determine the chemical properties of the pathogen isolates. The mean spectrum of each pathogen isolate was computed between 1103 and 2464 nm and plotted on one graph for comparison.

3.3.3.4. Principal Component Analysis (PCA)

Several pre-processing techniques were evaluated to determine which would yield the best results. Apart from mean-centering, standard normal variate (SNV) transformation was applied to the data (Rinnan *et al.*, 2009). The wavelengths from 920 to 1097 nm and 2477 to 2514 nm were removed as it corresponded to noisy areas and were deemed unnecessary for further analysis.

PCA takes a two-dimensional dataset (λ ; $x \cdot y$) and reconstructs the data into scores and loadings using equation 3.2, where \mathbf{T} is the scores matrix, \mathbf{P}' is the loadings matrix and \mathbf{E} refers to the residual matrix. The principal components (PC), made up from the scores and loadings, are used to explain the variance of the data in descending order of components. The variance captured in the first PC is not always related to the properties responsible for the separation between samples and may be misleading. It is therefore necessary to also look at the lower PC's in order to determine which properties are truly responsible for the separation of samples (Manley, 2014). The loadings provide information about specific wavelengths contributing to the variance.

$$X = TP' + E \quad (\text{eq. 3.2})$$

The PCA of the first two mosaics were calculated using both pixel and object wise analysis in order to compare the two methods. Thereafter the last two mosaics were calculated using the best method as determined by the first two mosaics. For consistency, all the PCA's were calculated with three principal components (PC).

The pixel wise analysis of the mosaicked images, uses the spectrum of each pixel in the PCA calculation, whereas the object wise analysis first calculates the average spectrum per cluster of pixels, that is selected by the user, and then calculates the PCA based on the set of average spectra. The score images, score plots and loading plots were used to examine the PCA results. The score images illustrate the similarity between pathogens resulting from characteristics such as chemical composition or morphology. When applying a heat map to these images the similarity can be seen where similar colours represent closely related pathogens. On a score plot each point represents either a pixel or single unit of measurement.

3.3.3.5. Partial Least Squares Discriminant Analysis (PLS-DA)

The fungal isolates were distinguished from one another by calculating PLS-DA models for each of the four mosaicked models. SNV was again applied to all the models as a pre-processing step. PLS-DA uses an algorithm (equation 3.3) that aims to find a straight line, which divides a given space into two regions, thus separating the data into groups (Musumarra *et al.*, 2007; Amigo *et al.*, 2013; Brereton & Lloyd, 2014; EIMasry & Nakauchi, 2016). It qualitatively predicts a set of variables within a larger dataset (Gowen *et al.*, 2007). Furthermore the relationship between the spectral data and the response variables is thus modelled (Amigo *et al.*, 2013). This can also be applied to multi-class models, where separation is obtained in more spatial directions (Brereton & Lloyd, 2014).

$$y = X\beta + f \quad (\text{eq. 3.3})$$

Where:

X = Unfolded hypercube, or data matrix, obtained from the hyperspectral image

β = Beta coefficient

f = Residual or “error” matrix

The result of the modelling would be the assignment of the pixels/objects to the different classes, namely the different fungal isolates. Both object and pixel wise PLS-DA models were calculated on identical calibration sets based on two of the three triplicates, to compare and find the optimum method to distinguish the fungi. Full cross-validation was only performed on the object wise PLS-DA models, as the pixel wise models contained a much larger number of data points.

The last of the three triplicate images was used to validate the models. Using the classification images, the predictions for the different classes, i.e. the fungal isolates, were totalled to determine the incorrect and correct classifications for both the pixel-wise and object-wise method. The “no class” classification class, which refers to pixels or objects that were not been assigned to any of the existing classes, was regarded as false negatives for this study.

The classification accuracy was calculated per equation 3.4, together with the false positive error and false negative error shown in equation 3.5 and 3.6 respectively. The accuracy illustrates how effective the overall model is (Sokolova *et al.*, 2006). The false positive error refers to the algorithm classifying the other isolates in the model as a selected

isolate, whereas the false negative error describes the misclassification of a selected isolate to another isolate in the model.

Sensitivity, also known as the recall or true positive rate, refers to the probability that a given fungal isolate in a model would be correctly classified, as per equation 3.7 (Sokolova *et al.*, 2006). It describes the effectiveness of the model on a class basis. The specificity, also known as the true negative rate, describes the probability that the remaining fungal isolates in the model would be correctly classified (equation 3.8). The sensitivity and specificity is thus used to assess the classification algorithm on a single class. The precision refers to the predictive power of the model by estimating the predictive value per class, as per equation 3.9. F1 score (equation 3.10) refers to the weighted mean between precision and recall (Sokolova *et al.*, 2006). The misclassification rate was also calculated for each of the models (equation 3.11).

$$\text{Classification accuracy (\%)} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total}} \times 100\% \quad (\text{eq. 3.4})$$

$$\text{False positive error (\%)} = \frac{\text{False Positives}}{\text{Total}} \times 100\% \quad (\text{eq. 3.5})$$

$$\text{False negative error (\%)} = \frac{\text{False Negatives}}{\text{Total}} \times 100\% \quad (\text{eq. 3.6})$$

$$\text{Sensitivity or Recall (\%)} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \times 100\% \quad (\text{eq. 3.7})$$

$$\text{Specificity (\%)} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \times 100\% \quad (\text{eq. 3.8})$$

$$\text{Precision (\%)} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \times 100\% \quad (\text{eq. 3.9})$$

$$\text{F1 Score (\%)} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (\text{eq. 3.10})$$

$$\text{Misclassification rate (\%)} = \frac{\text{False Negatives} + \text{False Positives}}{\text{Total}} \times 100\% \quad (\text{eq. 3.11})$$

Where:

True Positives = Isolate 1 correctly classified as isolate 1

True Negatives = Isolates 2, 3 and 4 correctly classified as isolates 2, 3 and 4, respectively.

False Positives = Isolates 2, 3 and 4 incorrectly classified as isolate 1

False Negatives = Isolate 1 incorrectly classified as isolates 2, 3 and 4

Total = Sum of all the fungal isolates in the given model

3.4. RESULTS AND DISCUSSION

3.4.1. Spectral Analysis

The mean spectra of each pathogen isolate, computed by the software in the range of 1103 nm to 2464 nm, was plotted on one graph as shown in Figure 3.5.

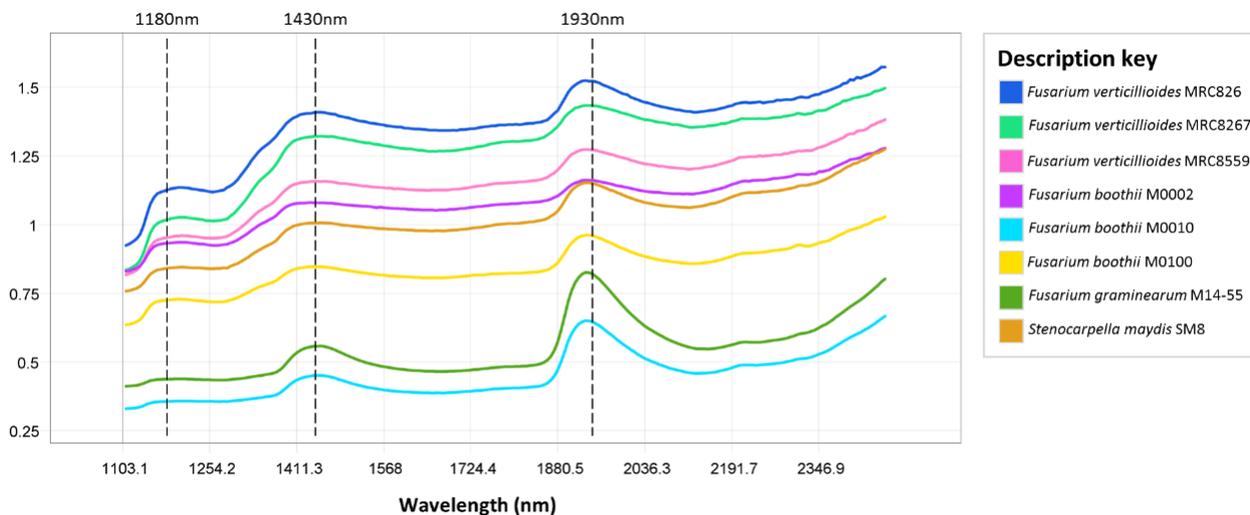


Figure 3.5 Mean spectra of the pathogen isolates, depicting *Fusarium verticillioides* MRC826 (dark blue), *Fusarium verticillioides* MRC8267 (teal), *Fusarium verticillioides* MRC8559 (pink), *F. boothii* M0010 (purple), *F. boothii* M0010 (light blue), *F. boothii* M0010 (yellow), *F. graminearum* s.s. M14-55 (green) and *Stenocarpella maydis* SM8 (orange), used in the four different mosaics.

The mean spectra of the eight pathogen isolates have similar absorption bands, although they vary in intensity. The prominent bands are seen at 1180, 1430 and 1930 nm. The band at 1180 nm is related to the C-H second overtone from carbohydrates (Delwiche & Hareland, 2004) and is also related to carbonyl compounds found in *F. verticillioides* (Berardo *et al.*, 2005). The 1430 nm band indicates the presence of starch and sucrose (O-H stretch first overtone) as stated by Berardo *et al.* (2005). Although Osborne, Fearn, & Hindle, (1993) stated that 1430 nm indicates the presence of protein (N-H stretch first overtone), their study referred to food and beverage analysis and not pathogen contamination. It can thus be said that the 1430 nm band can be related to starch, sucrose and protein. Lastly, the 1930 nm band represents the O-H stretch and O-H deformation combination and O-H bend second overtone related to moisture, as well as N-H stretch second overtone related to the CONH group (Osborne *et al.*, 1993; Berardo *et al.*, 2005; Manley, 2014). Although the spectra of the pathogen isolates vary in intensity, it cannot be

attributed to the chemical composition as the raw spectra may contain effects of light scattering and agar.

The separation between the mean spectra of the eight isolates was further investigated by subtracting the mean spectra from one another to identify where the bands are that account for the difference in the spectra. The difference spectra of each isolate with the remaining isolates is shown in Addendum A (Figure A1 – A8). From this, it was seen that there was little difference between the *F. boothii* M0100 and *S. maydis* SM8 isolates due to a semi-straight line (Figure A6 and A8). The separation between the remainder of the isolates is a result of the 1930 nm band caused by the O-H stretch and O-H deformation combination, as well as the O-H bend second overtone related to moisture as stated by Osborne *et al.* (1993). This is possibly due to the moisture accumulation on the lid surface within the Petri dish that varied.

3.4.2. Multivariate Data Analysis

3.4.2.1. Principal Component Analysis

PCA was used to explore the data through examining the differences between selected groups of pathogen isolates clustered together in a mosaic. The results of the mosaics, calculated for both pixel and object wise analysis are shown in Figure 3.6 to 3.15.

The scores and loadings plots were used to explore the data and get an indication of the relationship between the pathogen isolates. The score plot of PC2 vs PC3 clearly identified the dead pixel lines as they were distinctly separated from the rest of the image pixels. This is the result of the detector having lines that do not capture any data during imaging. These lines were removed in all the mosaics and the PC's were recalculated before proceeding with the interpretation of the data. The score images were also used to explore the separation between the isolates. The score images (Figure 3.6c, 3.7c, 3.9c and 3.10c) illustrated the similarity between the isolates, where positive values are red (warm colours) and negative values are blue (cool colours).

The score plot of PC1 vs. PC2 for mosaic one showed good separation with three distinct clusters (Figure 3.6a), which when shaded per the pathogen isolates illustrated that each cluster corresponded with an isolate (Figure 3.6d). PC1 accounted for 80.2% of the variance in the data (Figure 3.6a), and PC2 for 14.7% (Figure 3.6a, b), whereas PC3 (Figure 3.6b) only accounted for approximately 1% of the variance in the data. This shows that most of the separation is explained in the first two components. The variance seen in the first PC can be attributed to the difference in the biological composition of the three pathogen

isolates since the three species produce different toxins. This separation is due to a difference in carbohydrates, proteins and moisture as shown in the loadings plot (Figure 3.8a).

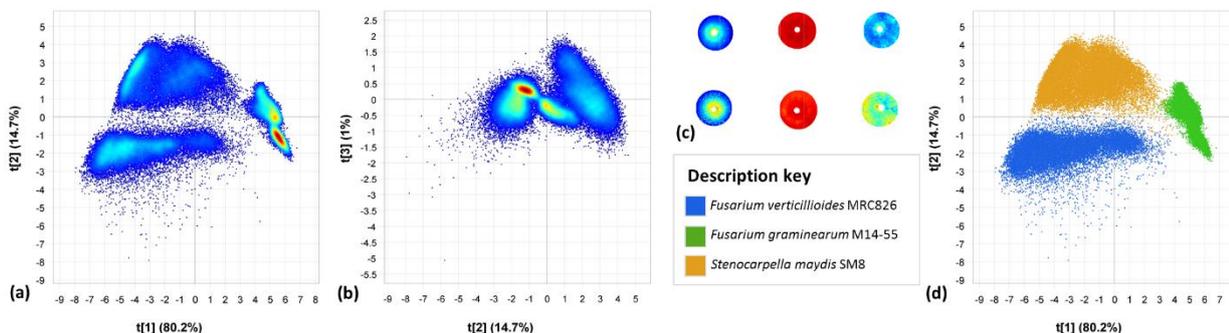


Figure 3.6 Pixel wise PCA of mosaic one illustrating good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (80.2%) vs. PC2 (14.7%) and (b) PCA score plot of PC2 (14.7%) vs. PC3 (1%) shaded per the density of the points. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. graminearum* s.s. M14-55 (middle) and *S. maydis* SM8 (right). (d) PCA score plot of PC1 vs. PC2 illustrating the separation between the *F. verticillioides* MRC826 (blue), *F. graminearum* s.s. M14-55 (green) and *S. maydis* SM8 (orange) isolates.

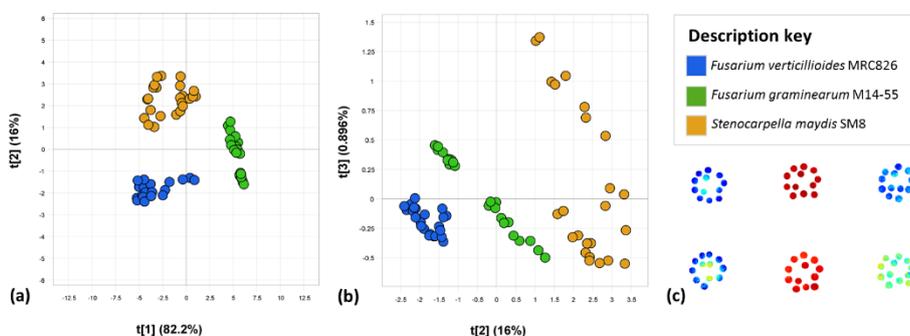


Figure 3.7 Object wise PCA of mosaic one illustrating good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (82.2%) vs. PC2 (16%) and (b) PCA score plot of PC2 (16%) vs. PC3 (0.896%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. graminearum* s.s.M14-55 (middle) and *S. maydis* SM8 (right).

When comparing the pixel wise PCA score plots to the plots of the object wise PCA, the object wise PCA plots had noticeably better separation between the isolates. The variance explained by PC1 (82.2%) was improved by 2% and PC2 (16%) by 1.3% (Figure 3.7a). Although this illustrates a small increase, a smaller number of data points was used

in the calculation, thus the separation of the isolates was improved. The score plot shown in Figure 3.7b of PC2 vs. PC3 also illustrated three distinct clusters in the direction of PC2 (16%), which when compared to the pixel wise approach Figure 3.6b, was a vast improvement of the separation between the isolates. The third PC (0.896%) showed separation between the duplicates used in the study, thus illustrating the internal variance of the samples. This however did not affect the overall separation between the isolates.

The score images of both the object and pixel wise analysis confirms what was seen in the score plots (Figure 3.6d and Figure 3.7a), where a clear separation in PC1 (Figure 3.6c and Figure 3.7c) is observed. The heat map illustrates the similarity of isolates through similar colours based on the score values, thus *Fusarium graminearum* s.s. that is red and has a positive score value is clearly distinguishable from *F. verticillioides* and *S. maydis* with their blue-yellow colour and negative score values (Figure 3.8a).

The loadings plots, which are used in combination with the score plots, explain the origin of the variance between the isolates. This is based on the spectral points used during the calculations and thus explains the variance as wavelength bands that contribute to the separation. The loadings were the same for the object and pixel wise data as the same images were used for both analyses. The positive bands in a loading plot correlate to the positive side of the corresponding PC's score plot and is an indication of the intensity of the spectral response (Esbensen *et al.*, 2002b). The same is exhibited by the negative bands correlating with the negative side of the score plot.

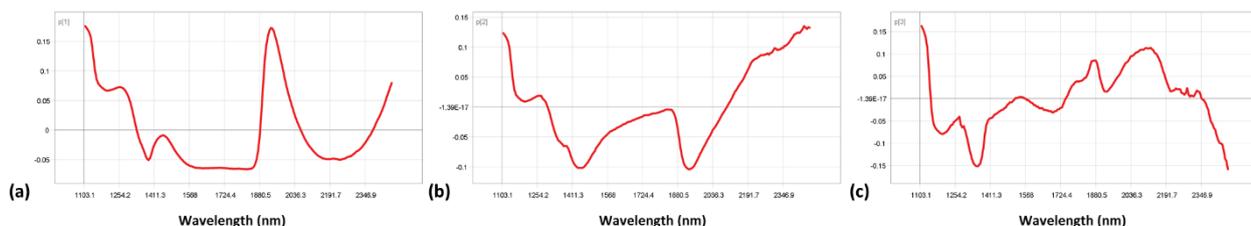


Figure 3.8 PCA loadings plots of mosaic one for both the pixel and object wise analysis is given for (a) PC1 with bands at 1260, 1400 and 1930 nm, (b) PC2 with bands at 1430 and 1930 nm and (c) PC3 with a band at 1200 nm.

The loading plot of PC1 (Figure 3.8a) showed prominent positive bands at 1260 and 1930 nm and a negative band at 1400 nm. When looking at the positive loading bands of PC1, in conjunction with the score plot of PC1 (80.2%) vs. PC2 (14.7%) (Figure 3.6d and 3.7a), the separation of *F. graminearum* s.s. from the other isolates in mosaic one is mostly based on the spectral band at 1930 nm, as this has the highest intensity followed by 1260

nm. This is related to the CONH group [N-H stretch second overtone] corresponding to the protein composition of the isolates, although the moisture [O-H stretch, O-H deformation combination and O-H bend second overtone] could also affect the separation (Osborne *et al.*, 1993; Berardo *et al.*, 2005; Manley, 2014). It is unclear whether the moisture bands observed is a result of the constituents of the pathogen isolates, or a direct result of the moisture accumulation on the inside of the petri dish. The moisture accumulates on the inside of the petri dish lid due to the pathogens respiring in the presence of the nutrient rich agar, thus resulting in the formation of both carbon dioxide and water. Another factor is also the temperature fluctuations experienced by the isolates as they travelled to where they were scanned with the NIR hyperspectral imaging system.

The loading plot of the second PC (Figure 3.8b) showed that the separation of *F. verticillioides* and *S. maydis* was based on the two negative bands at 1430 nm, which is associated with starch and sucrose, and 1930 nm, related to protein (Berardo *et al.* 2005). The negative band at 1200 nm seen in PC3 (Figure 3.8c), and to a lesser extent also in the first two PC's, was due to the presence of carbohydrates [C-H stretch second overtone] and carbonyl compounds (Osborne *et al.*, 1993; Delwiche & Hareland, 2004; Berardo *et al.*, 2005). Delwiche and Hareland (2004) also found that at 1200 nm, it was possible to discriminate between two samples, or isolates, based on the presence of the DON mycotoxin produced by the *F. graminearum* species complex.

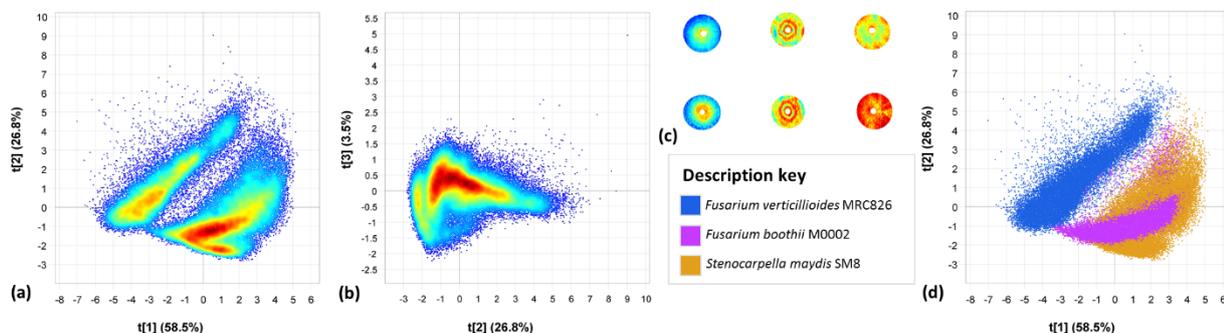


Figure 3.9 Pixel wise PCA of mosaic two illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (58.5%) vs. PC2 (26.8%) and (b) PCA score plot of PC2 (26.8%) vs. PC3 (3.5%) shaded per the density of the points. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. boothii* M0002 (middle) and *S. maydis* SM8 (right) is also given. PCA score plot of PC1 vs. PC2 (d) colour coded per pathogen isolate as seen in the accompanying legend is also given.

The pixel and object wise PCA results of the second mosaic is shown in Figure 3.9 and 3.10, respectively. PC1 captured 58.5% of the variance and a further 26.8% by PC2. PC3 accounted for 3.5% of the separation, although there was a lack of distinct clustering on the PC3 axes (Figure 3.9b). The clustering observed in the score plot of PC1 vs PC2 (Figure 3.9a) indicates that there are two isolates that are not separate from one another. This is confirmed with the score plot, colour coded according to each pathogen isolate, where *F. boothii* (purple) and *S. maydis* (orange) overlap almost entirely. *F. verticillioides*, on the other hand, can be distinguished from the other two isolates.

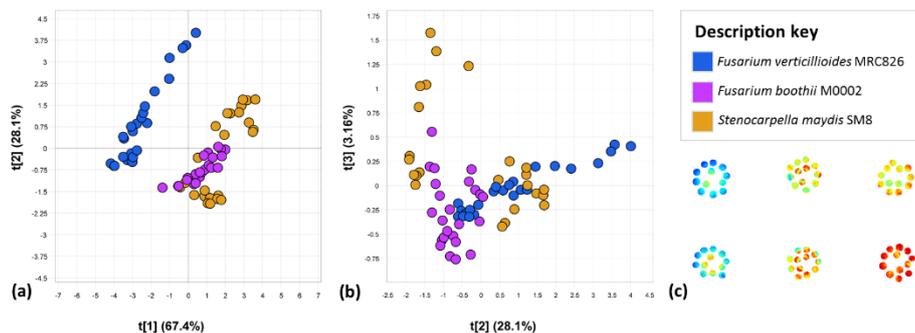


Figure 3.10 Object wise PCA of mosaic two illustrates separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (67.4%) vs. PC2 (28.1%) and (b) PCA score plot of PC2 (28.1%) vs. PC3 (3.16%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. boothii* M14-55 (middle) and *S. maydis* SM8 (right).

The lack of separation between *F. boothii* and *S. maydis* is also seen in the object wise PCA score plot of PC1 (67.4%) vs. PC2 (28.1%). This is seen by the two clusters on the score plots (Figure 3.10a, b), specifically the *S. maydis* isolate (orange). The score image (Figure 3.10c) confirms this as the top (right column) duplicate is shaded yellow-orange, like the *F. boothii* isolates (middle column) and the bottom (right column) duplicate is red. This was also noticed during the exploration of mosaic one's PCA score image, where the bottom duplicate was coloured lighter indicating a different score value to that of the top duplicate. The variance between the duplicates of the isolate could be attributed to the density of the fungal growth on the agar surface.

The *S. maydis* digital image B (Figure 3.2) has less dense growth on the surface, thus showing more of the agar which was subsequently captured by the imaging system and influenced the data. The variation in the growth density could be attributed to the composition of the agar [potato extract, glucose and agar] having slight variation due to manual mixing and pouring. The score plot of PC1 vs PC2 showed separation of the *F.*

verticillioides isolate in the direction of both PC1 and PC2, which is again confirmed by the distinctive colour difference (isolates are shaded blue) in the score image (Figure 3.10c) when compared to the other two isolates.

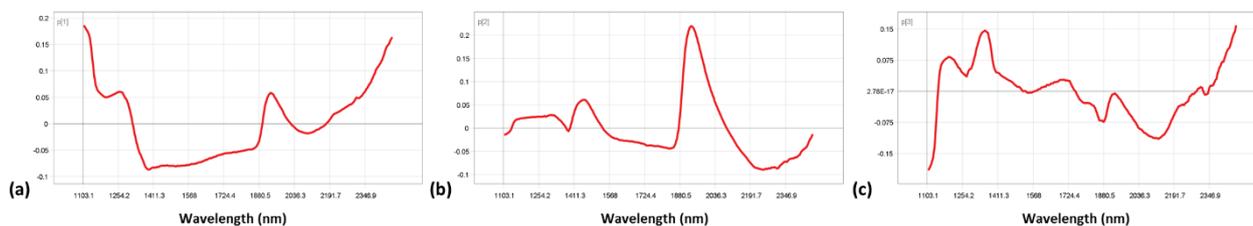


Figure 3.11 PCA loadings plots of mosaic two for both the pixel and object wise analysis is given for (a) PC1 with bands at 1260, 1400 and 1930 nm, (b) PC2 with bands at 1450 and 1930 nm and (c) PC3 with a band at 1200 and 1880 nm.

The loading plot of PC1 (Figure 3.11a) shows prominent positive bands at 1260 and 1930 nm and a negative band at 1400 nm. The negative band corresponds with the negative side of the score plot (Figure 3.10a), thus it is responsible for the separation of *F. verticillioides* from the other isolates. This separation can be attributed to the ROH structure [O-H stretch first overtone] which is associated with starch (Osborne *et al.*, 1993; Williams *et al.*, 2012a). This correlates with what was seen in the loadings plots of mosaic one (Figure 3.8b), where the second PC distinguished *F. verticillioides* from the other isolates based on the starch and protein content as stated by Berardo *et al.* (2005). Although the loading plots of PC2 and PC3 had prominent bands at 1260 and 1930 nm, and 1200 and 1880 nm, respectively, there was a lack of separation between the isolates shown in Figure 3.9d and Figure 3.10b. These bands are associated with the DON mycotoxin (1200 nm), sucrose and starch (1450 nm), protein and moisture (1930 nm), also seen in the previous models' results (Berardo *et al.*, 2005).

As the first two mosaics proved that object wise analysis was the superior method, it was decided to only use object wise analysis for the last two mosaics. Mosaic three consisted of three *F. verticillioides* isolates, namely MRC826, MRC8267 and MRC8559, whereas mosaic 4 consisted of *F. boothii* M0002, M0010 and M0100.

The PCA score plots (Figure 3.12a, b) showed that 83.2% of the variance was captured by PC1, 10.8% by PC2 and 4.72% by PC3. The three isolates were separated into three clusters in the direction of PC1. The score plot of PC2 vs. PC3 also further illustrated clustering in the vertical axis. The score image of PC1 vs PC2 does not have a distinct colour difference between the MRC826 and MRC8267 isolates when a heat map was

applied (Figure 3.12c), which was seen in the score plot of PC1 vs. PC2 as these two isolates overlapped slightly.

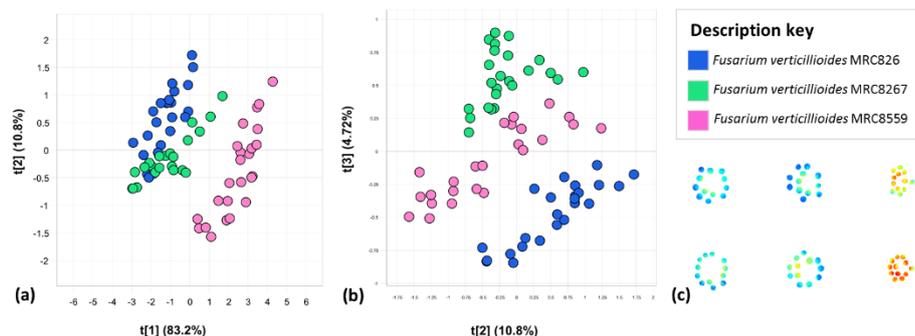


Figure 3.12 Object wise PCA of mosaic three illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (83.2%) vs. PC2 (10.8%) and (b) PCA score plot of PC2 (10.8%) vs. PC3 (4.72%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. verticillioides* MRC8267 (middle) and *F. verticillioides* MRC8559 (right).

As with the previous two models, the separation observed in the score plots were explained by the corresponding loadings plot, where the MRC8559 isolate was separated due to a band at 1930 nm on PC1 (Fig.3.13a) relating to the protein content (Berardo *et al.*, 2005). The lack of a high intensity band at 1200nm is due to none of the isolates producing the DON mycotoxin found to be produced by the *F. graminearum* species complex (Delwiche & Hareland, 2004). The driver for separation between the *F. verticillioides* isolates was thus the protein and starch differences found at 1430 and 1930 nm shown in Figure 3.13a (Osborne *et al.*, 1993; Berardo *et al.*, 2005).

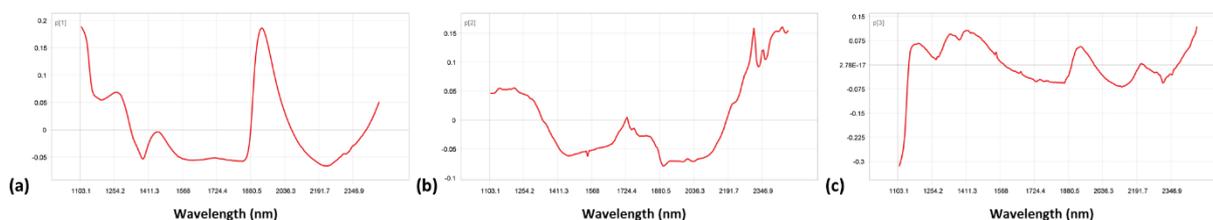


Figure 3.13 PCA loadings plots of mosaic three is given for (a) PC1 with bands at 1260, 1400 and 1930 nm, (b) PC2 with bands at 1430 and 1880 nm and (c) PC3 with a few low intensity bands at 1380, 1400 and 1930 nm.

The PCA results of the fourth mosaic also suggested that separation between isolates within the *F. boothii* species were possible. The score plot of PC1 vs. PC2 illustrated that

with the 92.4% variance captured in PC1 (Figure 3.14a), the three isolates could be distinguished from one another as they clustered together. The variance captured by PC2 (5.39%) and PC3 (1.43%) in Figure 3.14b, did not contribute to the separation between the isolates, however PC2 could be showing variance between the duplicates of the isolates.

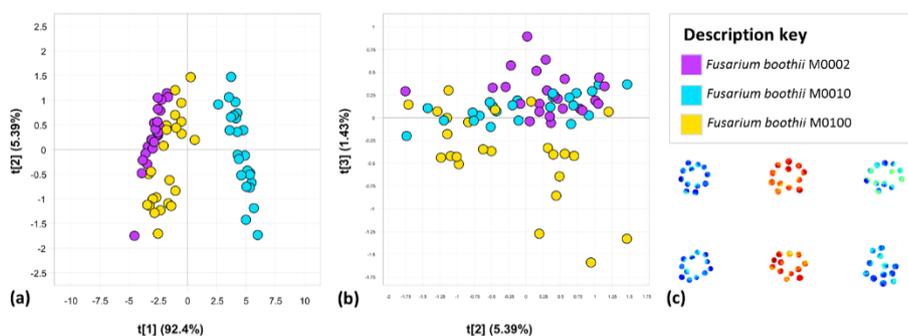


Figure 3.14 Object wise PCA of mosaic four illustrates the separation between *F. boothii* isolates. Scores illustrated by (a) PCA score plot of PC1 (92.4%) vs. PC2 (5.39%) and (b) PCA score plot of PC2 (5.39%) vs. PC3 (1.43%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. boothii* M0002 (left), *F. boothii* M0010 (middle) and *F. boothii* M0100 (right).

Visually the digital images of isolates M0010 and M0100 are similar (Figure 3.2), although this is not the case when looking at the score image (Figure 3.14c) as *F. boothii* M0010 (middle) has a positive (red) colour when the heat map is applied and *F. boothii* M0100 (right) has a negative (blue) colour. This illustrates that although samples may appear visually similar, chemically they are different and can in fact be distinguished from one another.

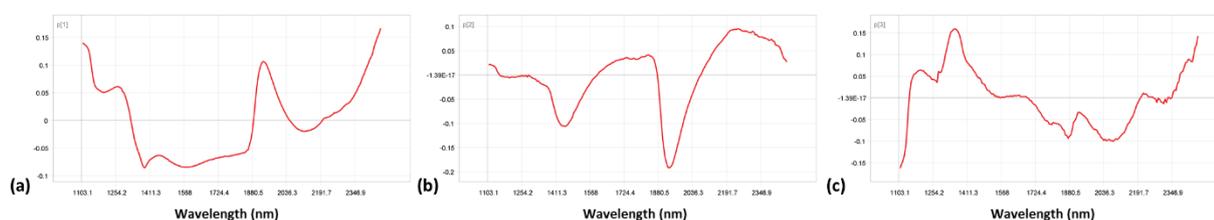


Figure 3.15 PCA loadings plots of mosaic four is given for (a) PC1 with bands at 1260, 1400 and 1930 nm, (b) PC2 with bands at 1430 and 1880 nm and (c) PC3 with a band at 1400 nm.

The loadings plot of PC1 (Figure 3.15a) has bands at 1260, 1400 and 1930 nm. This corresponds with the previous mosaics, thus these bands pertain to the chemical

characteristics of these pathogens in general. The 1260 nm band is associated with carbohydrates and the 1930 nm band with the protein [N-H stretch second overtone] present in the sample (Delwiche & Hareland, 2004; Berardo *et al.*, 2005). These two bands account for the separation between the three isolates in the first PC, which captured the most variance.

3.4.2.2. Partial Least Squares Discriminant Analysis

To determine the extent of the separation possible between the pathogens, both pixel wise and object wise PLS-DA models were calculated for each of the three mosaics. Both the object and pixel wise models were calculated on identical calibration sets, consisting of the triplicates B and C per isolate, and validated with triplicate A (Figure 3.2).

The performance of the models was evaluated by determining the classification accuracy, sensitivity, specificity and misclassification rate. The coefficient of determination, or R^2 value, was calculated for both the pixel and object wise models to evaluate how the classification model fit to the calibration data (Table 3.2). The cross-validated coefficient of determination, or Q^2 value, was also calculated for the object wise models based on the full cross-validation of the dataset. The full cross-validation was only performed on the object wise PLS-DA models, as the pixel wise models consisted of an abundant number of pixels.

Table 3.2 The performance measures used to assess the overall pixel and object wise PLS-DA models of the pathogen isolates.

Analysis	Mosaic	R^2	Q^2	Classification accuracy (%)	Misclassification rate (%)
Pixel wise	1	0.9242	-	99.72	0.49
	2	0.6682	-	82.05	34.82
Object wise	1	0.9702	0.9590	100.00	0.00
	2	0.8680	0.8410	86.11	16.13
	3	0.8943	0.8783	100.00	0.00
	4	0.8562	0.8081	88.89	25.00

The PLS-DA model of mosaic one, consisting of *F. verticillioides*, *F. graminearum* s.s. and *Stenocarpella maydis*, gave the best classification results. The pixel wise PLS-DA model of mosaic one achieved an overall classification accuracy of 99.7%, with the object wise model improving the classification accuracy to 100% (Table 3.2).

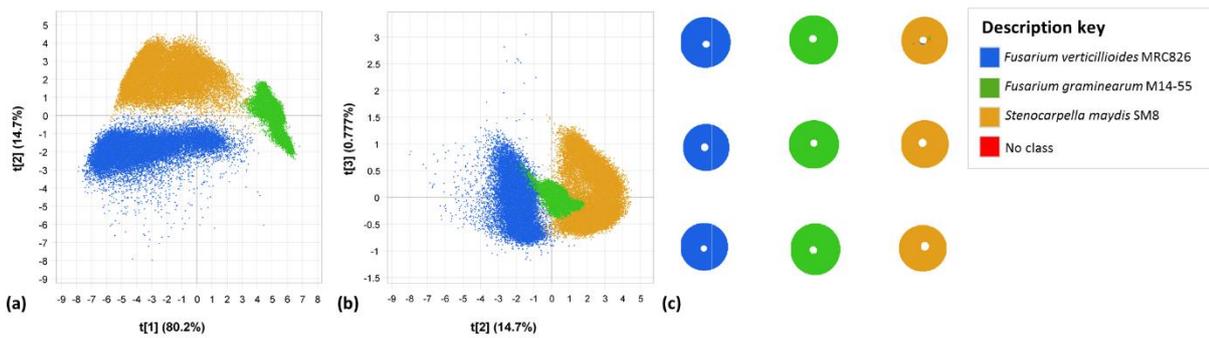


Figure 3.16 Pixel wise PLS-DA of mosaic one illustrating the (a) score plot of PLS factor 1 (80.2%) vs. factor 2 (14.7%) colour coded per pathogen isolate, the (b) score plot of factor 2 (14.7%) vs factor 3 (0.777%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. graminearum* s.s. M14-55 (green) and *S. maydis* SM8 (orange), illustrating the predicted pixels, with the pathogen categories provided in the legend, where the first row was used as validation and the second and third as calibration.

The pixel wise classification image of mosaic one (Figure 3.16c) shows how well the model classified the isolates (top row). Only a few misclassified pixels were observed in *S. maydis* isolate (top right) of Figure 3.16c. This accounted for the 0.49% misclassification rate of the overall model (Table 3.2). The sensitivity was above 99.6% for all three of the isolates and the specificity above 99.7% for the pixel wise model of mosaic one (Table 3.3). This indicates that the model has a high probability of correctly classifying all the isolates.

The object wise PLS-DA model of mosaic one showed improvement in the variance captured by the first two factors as seen in the score plot of factor 1 vs factor 2 (Figure 3.17a, b). The first factor accounted for 82.2% of the variance and the second factor a further 16%. The classification accuracy, together with the sensitivity, specificity and F1 score were 100% (Table 3.3), which means the object wise model classified the isolates without fault. This was confirmed when looking at the score image shaded according to the isolates, where the top row is the classified isolates. A high R^2 value of 0.9702 and a Q^2 value of 0.9590 was also determined, indicating the ease of distinguishing between the three isolates (Table 3.2).

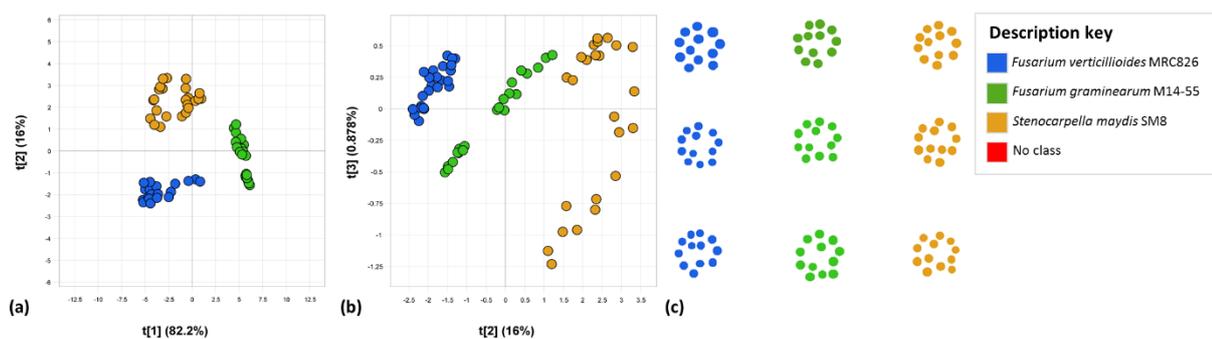


Figure 3.17 Object wise PLS-DA of mosaic one illustrating the (a) score plot of PLS factor 1 (82.2%) vs factor 2 (16%) colour coded per pathogen isolate, the (b) score plot of factor 2 (16%) vs factor 3 (0.878%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. graminearum* s.s. M14-55 (green) and *S. maydis* SM8 (orange), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used as validation and the second and third as calibration.

The second mosaic, where *F. graminearum* s.s. was replaced with *F. boothii* M0002 due to the two isolates producing the same mycotoxin (DON), the object wise PLS-DA model (Figure 3.19) again achieved better results than that of the pixel wise model (Figure 3.18). This was seen in the improvement of the R^2 value with 20%, which accounts for how easily the model fits to the data (Table 3.2). A Q^2 value of 0.8410 was obtained for the object wise model, which is lower than what was observed in mosaic one, but still indicates that distinguishing between the three isolates can be done with ease (Table 3.2). The overall accuracy of the mosaic two models was lower than that of mosaic one, although both the pixel and object wise models of mosaic two had high classification accuracies of 82.05% and 86.11%, respectively (Table 3.2). This suggests that the three major ear rot pathogens can easily and accurately be distinguished if either of the *F. graminearum* species complex isolates, *F. graminearum* s.s. or *F. boothii*, are included.

The score plot of factor 1 vs factor 2 accounted for 50.3% and 30% of the variance, respectively (Figure 3.18a). This allowed for the separation the *F. verticillioides* on the positive side of factor 1. The score image of factor 1 vs factor 2 (Figure 3.18c) illustrated that the classification of the *F. verticillioides* pixels (top left) was almost all classified correctly. This was confirmed by the classification accuracy (98.96%), sensitivity (97.58%) and specificity (99.94%) in Table 3.4. Even though these results are very accurate, the object wise model achieved 100% accuracy, sensitivity and specificity (Table 3.4) when classifying *F. verticillioides* (Figure 2.19c).

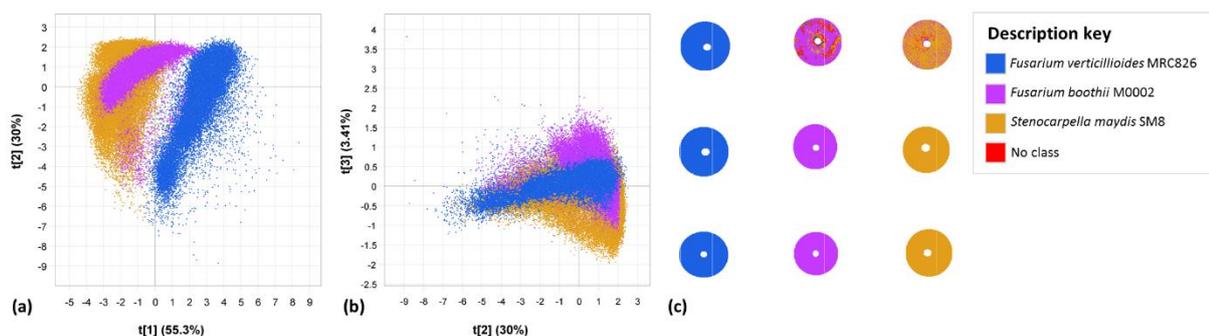


Figure 3.18 Pixel wise PLS-DA of mosaic two illustrating the (a) score plot of PLS factor 1 (55.3%) vs factor 2 (30%) colour coded per pathogen isolate, the (b) score plot of factor 2 (30%) vs factor 3 (3.41%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. boothii* M0002 (purple) and *S. maydis* SM8 (orange), illustrating the predicted pixels with the pathogen categories provided in the legend, where the first row was used as validation and the second and third as calibration.

The classification accuracy, sensitivity and specificity (Table 3.4) for the *F. boothii* and *S. maydis* isolates were lower than that of *F. verticillioides*. The object wise model greatly improved the classification of *S. maydis* from the pixel wise model, with the classification accuracy increasing from 88.10% (pixel wise) to 100% (object wise) shown in Table 3.4. The 13.89% misclassification rate of *S. maydis* in Table 3.4 was seen in the predicted object wise score image in Figure 3.19c (top middle), where 5 out of the 12 objects were misclassified as “no class”. The same area of the isolate was misclassified with the pixel wise analysis (Figure 3.18c) as the top part of the image contained a cluster of “red” misclassified pixels. This was attributed to the presence of agar due to the growth formation of the given pathogen isolate. *F. boothii* M0002 grows in a ring formation, resulting in alternating rings of thick and thin growth, as seen in the digital images of the isolate (Figure 3.2). The object wise analysis takes the average spectra of a selected cluster of pixels, so it could be that there were more pixels representing the agar present when the spectra of the cluster of pixels was converted into an object with a single spectrum. The specificity of the *F. boothii* isolate in mosaic two was 100% (Table 3.4), which indicates that the remaining isolates in the model were all accurately classified.

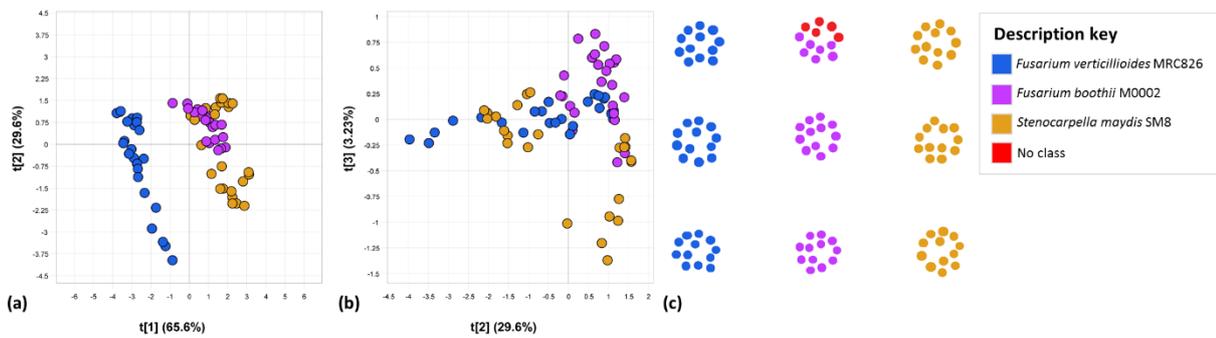


Figure 3.19 Object wise PLS-DA of mosaic two illustrating the (a) score plot of PLS factor 1 (65.6%) vs factor 2 (29.6%) colour coded per pathogen isolate, the (b) score plot of factor 2 (29.6%) vs factor 3 (3.23%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. boothii* M0002 (purple) and *S. maydis* SM8 (orange), illustrating the predicted objects with the pathogen categories provided in the legend, where the first row was used as validation and the second and third as calibration.

The PLS-DA results of mosaic three confirmed what was seen in the PCA score plots. The separation between the three *F. verticillioides* isolates was achieved with 100% overall accuracy on Day 5 of growth (Table 3.2). The overall model also had a high R^2 value of 0.8943. As the overall model achieved perfect classification, each isolate was also classified with 100% accuracy, together with 100% sensitivity and specificity. It is thus possible to distinguish between isolates from the same species, i.e. *F. verticillioides*, after five days of growth.

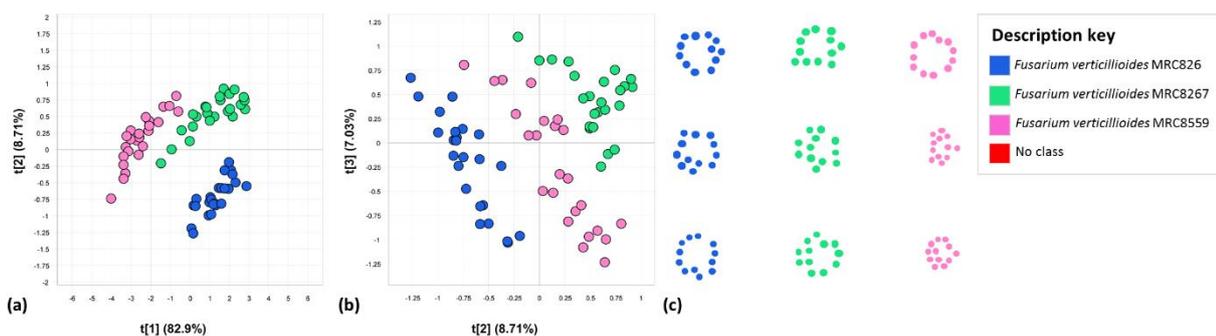


Figure 3.20 Object wise PLS-DA of mosaic three illustrating the (a) score plot of PLS factor 1 (82.9%) vs factor 2 (8.71%) colour coded per pathogen isolate, the (b) score plot of factor 2 (8.71%) vs factor 3 (7.03%) colour coded per pathogen isolate and the (c) classification image of *Fusarium verticillioides* MRC826 (blue), *F. verticillioides* MRC8267 (teal) and *F. verticillioides* MRC8559 (pink), illustrating the predicted objects with the pathogen categories provided in the legend, where the first row was used as validation and the second and third as calibration.

Lastly it was decided to investigate the possibility of distinguishing between the three *F. boothii* isolates that are known to cause DER in maize. The PLS-DA model of mosaic four showed promising results for the differentiation of the *F. boothii* isolates as it had an overall accuracy of 88.89% (Table 3.2).

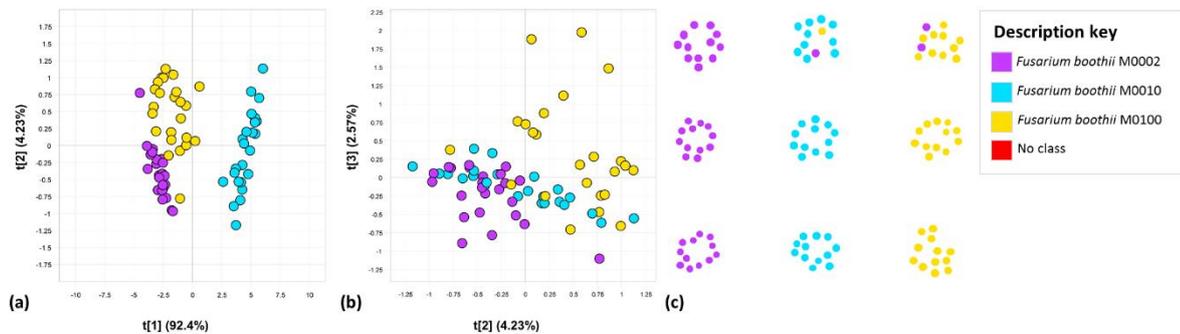


Figure 3.21 Object wise PLS-DA of mosaic four illustrating the (a) score plot of PLS factor 1 (92.4%) vs factor 2 (4.23%) colour coded per pathogen isolate, the (b) score plot of factor 2 (4.23%) vs factor 3 (2.57%) colour coded per pathogen isolate and the (c) classification image of *Fusarium boothii* M0002 (purple), *F. boothii* M0010 (light blue) and *F. boothii* M0100 (yellow), illustrating the predicted objects with the pathogen categories provided in the legend, where the first row was used as validation and the second and third as calibration.

The calibration (bottom two rows) and validation (top row) set seen in Figure 3.21c illustrates the accuracy with which the model classified the three *F. boothii* isolates. The M0010 isolate achieved the highest accuracy with 94.12%, followed by M0002 and M0100 each with 91.43% (Table 3.4). The classification accuracy of M0002 was lower due to the two false negative objects classified in the remaining two isolates (Figure 3.21c). The sensitivity and specificity for the M0010 and M0100 isolates were above 91%, suggesting a good effective model, even though the sensitivity for M0002 was only 80%, which is still considered high. It can be said that the *F. boothii* isolates can effectively be distinguished from one another with only a small percent of misclassification (Table 3.4).

Table 3.3 The performance measures used to assess the pixel wise PLS-DA models of the pathogen isolates used in the four mosaics.

Mosaic	Pathogen isolate	Classification accuracy (%)	Sensitivity (%)	Specificity (%)	F1 Score (%)	Misclassification rate (%)
1	<i>F. verticillioides</i> MRC826	99.88	99.73	99.97	99.84	0.12
	<i>F. graminearum</i> s.s. M14-55	99.89	99.68	100.00	99.84	0.11
	<i>S. maydis</i> 8	99.74	99.76	99.73	99.57	0.26
2	<i>F. verticillioides</i> MRC826	98.96	97.58	99.94	98.74	1.04
	<i>F. boothii</i> M0002	83.16	64.55	91.97	71.13	16.84
	<i>S. maydis</i> 8	88.10	82.54	91.23	83.35	11.90

Table 3.4 The performance measures used to assess the object wise PLS-DA models of the pathogen isolates used in the four mosaics.

Mosaic	Pathogen isolate	Classification accuracy (%)	Sensitivity (%)	Specificity (%)	F1 Score (%)	Misclassification rate (%)
1	<i>F. verticillioides</i> MRC826	100.00	100.00	100.00	100.00	0.00
	<i>F. graminearum</i> s.s. M14-55	100.00	100.00	100.00	100.00	0.00
	<i>S. maydis</i> 8	100.00	100.00	100.00	100.00	0.00
2	<i>F. verticillioides</i> MRC826	100.00	100.00	100.00	100.00	0.00
	<i>F. boothii</i> M0002	86.11	58.33	100.00	73.68	13.89
	<i>S. maydis</i> 8	100.00	100.00	100.00	100.00	0.00
3	<i>F. verticillioides</i> MRC826	100.00	100.00	100.00	100.00	0.00
	<i>F. verticillioides</i> MRC8267	100.00	100.00	100.00	100.00	0.00
	<i>F. verticillioides</i> MRC8559	100.00	100.00	100.00	100.00	0.00
4	<i>F. boothii</i> M0002	91.43	80.00	100.00	88.89	8.57
	<i>F. boothii</i> M0010	94.12	100.00	91.67	90.91	5.88
	<i>F. boothii</i> M0100	91.43	90.91	91.67	86.96	8.57

3.5. CONCLUSION

NIR hyperspectral imaging coupled with MIA could accurately distinguish between the major maize ear rot pathogens on day five of growth. *Fusarium verticillioides*, *F. boothii*, *F. graminearum* s.s. and *Stenocarpella maydis* isolates were used in varying combinations and could all be distinguished from one another with reasonably high accuracy. The study also found that pathogen identification is more accurate when using an object wise approach, as one colony would be representative of a sample and not each pixel as biological samples would differ from pixel to pixel. This was confirmed when the object wise analysis proved to be superior to that of pixel wise when considering the type of sample used in this study. Occasionally, a model had a lower classification accuracy, yet still showed good prediction results due to the size of the datasets. The three *F. boothii* isolates and three *F. verticillioides* isolates were also accurately distinguished from one another using an object wise approach. Throughout the study, the models containing the *F. boothii* isolates performed poorer in relation to that of the other models. This is attributed to the radial growth of the specific pathogen as its growths circles vary in density. Even though this was observed, the *F. boothii* isolates could still be effectively distinguished from one another with only a small percentage of misclassified points. The growth age of the pathogen isolates may influence the classification accuracy and needs to be investigated as younger cultures would allow for more rapid results when applying this method in a microbiology laboratory.

3.6. REFERENCES

- Amigo, J.M., Martí, I. & Gowen, A. (2013). Hyperspectral Imaging and Chemometrics: A Perfect Combination for the Analysis of Food Structure, Composition and Quality. In: *Data Handling in Science and Technology*, Pp. 343–370.
- Berardo, N., Pisacane, V., Battilani, P., Scandolara, A., Pietri, A. & Marocco, A. (2005). Rapid detection of kernel rots and mycotoxins in maize by near-infrared reflectance spectroscopy. *Journal of Agricultural and Food Chemistry*, **53**, 8128–8134.
- Beukes, I., Rose, L.J., Shephard, G.S., Flett, B.C. & Viljoen, A. (2017). Mycotoxigenic *Fusarium* species associated with grain crops in South Africa – A review. *South African Journal of Science*, **113**, 1–12.
- Boutigny, A., Ward, T.J., Coller, G.J. Van, Flett, B., Lamprecht, S.C., O'Donnell, K. & Viljoen, A. (2011). Analysis of the *Fusarium graminearum* species complex from wheat, barley

- and maize in South Africa provides evidence of species-specific differences in host preference. *Fungal Genetics and Biology*, **48**, 914–920.
- Boutigny, A.L., Beukes, I., Small, I., Zühlke, S., Spiteller, M., Rensburg, B.J. Van, Flett, B. & Viljoen, A. (2012). Quantitative detection of *Fusarium* pathogens and their mycotoxins in South African maize. *Plant Pathology*, **61**, 522–531.
- Brereton, R.G. & Lloyd, G.R. (2014). Partial least squares discriminant analysis: taking the magic away. *Journal of Chemometrics*, **28**, 213–225.
- Burger, J.E. (2006). *Hyperspectral NIR Image Analysis*. PhD Thesis. Swedish University of Agricultural Sciences, Umeå.
- Burger, J. & Gowen, A. (2011). Data handling in hyperspectral image analysis. *Chemometrics and Intelligent Laboratory Systems*, **108**, 13–22.
- Department of Agriculture, Forestry and Fisheries. (2017). Summer Crops: Third Production Forecast (2017) and Winter Cereals: Intentions to plant (2017). [Internet document] URL <http://www.daff.gov.za/daffweb3/Home/Crop-Estimates>. Accessed 24/07/2017.
- Delwiche, S.R. & Hareland, G.A. (2004). Detection of scab-damaged hard red spring wheat kernels by near-infrared reflectance. *Cereal Chemistry*, **81**, 643–649.
- Del Fiore, A., Reverberi, M., Ricelli, A., Pinzari, F., Serranti, S., Fabbri, A.A., Bonifazi, G. & Fanelli, C. (2010). Early detection of toxigenic fungi on maize by hyperspectral imaging analysis. *International Journal of Food Microbiology*, **144**, 64–71.
- EPPO. (2017). *Stenocarpella macrospora* and *Stenocarpella maydis* [Internet document] URL https://www.eppo.int/QUARANTINE/data_sheets/.../DIPDSP_ds.pdf. Accessed 24/07/2017.
- EIMasry, G.M. & Nakauchi, S. (2016). Image analysis operations applied to hyperspectral images for non-invasive sensing of food quality - A comprehensive review. *Biosystems Engineering*, **142**, 53–82.
- Esbensen, K.H., Guyot, D., Westad, F. & Houmøller, L.P. (2002). Principal Component Analysis (PCA) - Introduction. In: *Multivariate data analysis in practice: an introduction to multivariate data analysis and experimental design*. 5th edn. CAMO Process AS.
- Foca, G., Ferrari, C., Ulrici, A., Sciutto, G., Prati, S., Morandi, S., Brasca, M., Lavermicocca, P., Lanteri, S. & Oliveri, P. (2016). The potential of spectral and hyperspectral-imaging techniques for bacterial detection in food: A case study on lactic acid bacteria. *Talanta*, **153**, 111–119.
- Gelderblom, W.C.A., Seier, J. V., Snijman, P.W., Schalkwyk, D.J. Van, Shephard, G.S. & Marasas, W.F.O. (2001). Toxicity of culture material of *Fusarium verticillioides* strain MRC 826 to nonhuman primates. *Environmental Health Perspectives*, **109**, 267–276.

- Gokul, A. (2015). *Fusarium graminearum species complex (FGSC) composition in South African wheat and maize grown in rotation*. MSc Thesis. Stellenbosch University, Stellenbosch.
- Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Downey, G. & Frias, J.M. (2007). Hyperspectral imaging - an emerging process analytical tool for food quality and safety control. *Trends in Food Science and Technology*, **18**, 590–598.
- Kammies, T., Manley, M., Gouws, P.A. & Williams, P.J. (2016). Differentiation of foodborne bacteria using NIR hyperspectral imaging and multivariate data analysis. *Applied Microbiology and Biotechnology*, **100**, 9305–9320.
- Kucheryavskiy, S. (2013). A new approach for discrimination of objects on hyperspectral images. *Chemometrics and Intelligent Laboratory Systems*, **120**, 126–135.
- Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chemical Society Reviews*, **43**, 8200–8214.
- Marasas, W.F.O., Thiel, P.G., Rabie, C.J., Nelson, P.E. & Toussoun, T.A. (1986). Moniliformin Production in *Fusarium* Section *Liseola*. *Mycologia*, **78**, 242–247.
- Munkvold, G.P. (2003). Cultural and genetic approaches to managing mycotoxins in maize. *Annual Review of Phytopathology*, **41**, 99–116.
- Musumarra, G., Trovato-Salinaro, A., Scirè, S., Foti, A., Barresi, V., Fortuna, C.G., Strazzulla, G., Condorelli, D.F. (2007). Identification of genes involved in radiation-induced G1 arrest. *Journal of Chemometrics*, 398–405.
- Nirenberg, H.I. & O'Donnell, K. (1998). New *Fusarium* species and Combinations within the *Gibberella fujikuroi* species complex. *Mycologia*, **90**, 434–458.
- Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). *Practical NIR spectroscopy with applications in food and beverage analysis*. 2nd edn. Longman Scientific & Technical, Essex.
- Peraica, M., Radić, B., Lucić, A. & Pavlović, M. (1999). Toxic effects of mycotoxins in humans. *Bulletin of the World Health Organization*, **77**, 754–766.
- Prats-Montalbán, J.M., Juan, A. de & Ferrer, A. (2011). Multivariate image analysis: A review with applications. *Chemometrics and Intelligent Laboratory Systems*, **107**, 1–23.
- Rabie, C.J., Rensburg, S.J. Van, Kriek, N.P.J. & Lübben, A. (1977). Toxicity of *Diplodia maydis* to laboratory animals. *Applied and Environmental Microbiology*, **34**, 111–114.
- Rinnan, A., Van Den Berg, F. & Engelsens, S.B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, **28**.
- Rossouw, J.D., Van Rensburg, J.B.J. & Van Deventer, C.S. (2002). Breeding for resistance to ear rot of maize, caused by *Stenocarpella maydis* (Berk) Sutton. 1. Evaluation of selection criteria. *South African Journal of Plant and Soil*, **19**, 182–187.

- Schilling, A.G., Möller, E.M. & Geiger, H.H. (1996). Polymerase Chain Reaction-Based Assays for Species-Specific Detection of *Fusarium culmorum*, *F. graminearum*, and *F. avenaceum*. *Phytopathology*, **86**, 515–522.
- Sendin, K. (2017). *Characterisation of whole white maize kernels*. MSc Thesis. Stellenbosch University, Stellenbosch.
- Seo, Y., Park, B., Hinton, A., Yoon, J.S. & Lawrence, K.C. (2016). Identification of *Staphylococcus* species with hyperspectral microscope imaging and classification algorithms. *Journal of Food Measurement and Characterization*, **10**, 253–263.
- Shephard, G.S. (2008). Impact of mycotoxins on human health in developing countries. *Food Additives and Contaminants*, **25**, 146-151.
- Snyder, W.C. & Hansen, H.N. (1940). The Species Concept in *Fusarium*. *American Journal of Botany*, **27**, 64–67.
- Sokolova, M., Japkowicz, N. & Szpakowicz, S. (2006). Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In: *Advances in Artificial Intelligence: 19th Australian Joint Conference on Artificial Intelligence* (edited by A. Sattar & B. Kang). Pp. 1015–1021. Springer Berlin Heidelberg.
- Strausbaugh, C.A., Overturf, K. & Koehn, A.C. (2005). Pathogenicity and real-time PCR detection of *Fusarium* spp. in wheat and barley roots. *Canadian Journal of Plant Pathology*, **27**, 430–438.
- ThermoFischer Scientific. (2017). Overview of ELISA [Internet document]. URL <https://www.thermofisher.com/za/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-elisa.html>. Accessed 27/09/2017.
- Turner, N.W., Bramhmbhatt, H., Szabo-Vezse, M., Poma, A., Coker, R. & Piletsky, S.A. (2015). Analytical methods for determination of mycotoxins: An update (2009-2014). *Analytica Chimica Acta*, **901**, 12–33.
- Velusamy, V., Arshak, K., Korostynska, O., Oliwa, K. & Adley, C. (2010). An overview of foodborne pathogen detection: In the perspective of biosensors. *Biotechnology Advances*, **28**, 232–254.
- Williams, P.J. (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. Stellenbosch University, Stellenbosch.
- Williams, P.J., Geladi, P., Britz, T.J. & Manley, M. (2012a). Near-infrared (NIR) hyperspectral imaging and multivariate image analysis to study growth characteristics and

differences between species and strains of members of the genus *Fusarium*. *Analytical and Bioanalytical Chemistry*, **404**, 1759–1769.

Williams, P.J., Geladi, P., Britz, T.J. & Manley, M. (2012b). Investigation of fungal development in maize kernels using NIR hyperspectral imaging and multivariate data analysis. *Journal of Cereal Science*, **55**, 272–278.

Williams, P.J. & Kucheryavskiy, S. (2016). Classification of maize kernels using NIR hyperspectral imaging. *Food Chemistry*, **209**, 131–138.

Yao, H., Hruska, Z., Kincaid, R., Brown, R.L. & Cleveland, T.E. (2008). Differentiation of toxigenic fungi using hyperspectral imagery. *Sensing and Instrumentation for Food Quality and Safety*, **2**, 215–224.

CHAPTER 4

The influence of growth age on the accuracy of maize ear rot pathogen differentiation using NIR hyperspectral imaging

4.1. ABSTRACT

This study focussed on determining the influence of the growth age, on the accuracy of maize ear rot pathogen differentiation with NIR hyperspectral imaging. Eight maize ear rot pathogen isolates comprising of *Fusarium verticillioides* [MRC826, MRC 8267 and MRC8559], *F. graminearum* s.s. [M14-55], *F. boothii* [M0002, M0010 and M0100] and *Stenocarpella maydis* [SM8] were plated on potato dextrose agar, in triplicate, and incubated at 25 °C for 3, 5 and 7 days. The plates were imaged with a SisuChema short wave infrared camera, in the range 1000 to 2500 nm, after which the data was analysed using multivariate image analysis. Principal component analysis (PCA) was initially used for data exploration, after which object wise classification models were calculated using partial least squares discriminant analysis (PLS-DA). Standard normal variate and mean-centering were applied to all the images. Models were calibrated and optimised with full cross-validation, thereafter independent validation images were tested. The overall classification accuracy of the PLS-DA models showed that day 5 was the best day to distinguish all the pathogens effectively. Day 3 could potentially be used to distinguish the three major ear rot isolates, *F. verticillioides* MRC826, *F. graminearum* s.s. M14-55 and *S. maydis* SM8, with 100% accuracy. Classification of the three *F. verticillioides* isolates on day 3 also achieved a 97% accuracy. The F1 score, i.e. the weighted mean between the predictive power of a model and the effectiveness of the model on a class basis, also demonstrated that pathogen isolates were accurately classified on day 5. NIR hyperspectral imaging can thus accurately distinguish between the major maize ear rot pathogens on day 3 or 5 depending on the pathogens, thus proving that the growth age does influence the accuracy of the differentiation of pathogens.

4.2. INTRODUCTION

The fungal pathogens *Fusarium* spp. and *Stenocarpella* spp. greatly affect the maize industry as it results in the formation of maize ear rot. Maize is a staple food in South Africa and is thus considered one of the most important cereal grains of our diet. As the maize is prone to pathogenic infection by organisms such as *Fusarium* and *Stenocarpella* spp., detecting contamination is key in securing food safety.

Crop losses can, however, occur due to physical plant damage or disease, which can be attributed to weather changes and infestation. Two of these diseases are Fusarium ear rot (FER) and Gibberella ear rot (GER), which is caused by *Fusarium* spp. and is proven to affect the yield and quality of maize (Munkvold, 2003). FER is predominantly caused by *Fusarium verticillioides* (syn: *F. moniliforme*) and manifests as a light pink or white mould on maize ears and produces the fumonisin toxin (Munkvold, 2003; Boutigny *et al.*, 2012; Beukes *et al.*, 2017). The *Fusarium graminearum* species complex (FGSC) comprises 16 phylogenetically distinct species of which *F. graminearum* s.s. and *F. boothii* pathogens are commonly associated with GER of maize in South Africa (Boutigny *et al.*, 2011; Gokul, 2015). GER appears as pink or red mould that initiates at the tip of the maize ear and produces deoxynivalenol (DON) toxin (Munkvold, 2003; Beukes *et al.*, 2017; Peraica *et al.*, 1999). It is difficult to distinguish the *Fusarium* genus due its taxonomy (Snyder & Hansen, 1940; Williams, 2013), however, it is important to be able to easily distinguish between the species since maize is highly susceptible to these mycotoxin-producing pathogens. The third disease most prominently associated with South African maize grain is Diplodia ear rot (DER), caused by the *Stenocarpella maydis* plant pathogen. It presents as white mould between the kernels of the maize ear, turning the ear a greyish-brown colour and producing the diplodia (DIP) toxin (Rossouw *et al.*, 2002; EPPO, 2017). It is important to note that these mycotoxigenic fungi may contaminate grain without the presence of visual symptoms.

Currently, molecular and immunological detection methods such as the polymerase chain reaction (PCR) (Strausbaugh *et al.*, 2005) and enzyme-linked immunosorbent assay (ELISA) (ThermoFischer Scientific, 2017) have been used for the enumeration and identification of the pathogens. These conventional microbiological methods, both molecular and immunological, can be tedious and time-consuming as specialized growth media and further testing or staining are generally required (Turner *et al.*, 2015). There is thus a need for rapid screening methods that accurately distinguish pathogens on growth media as an initial step in the microbiology laboratory.

Near infrared (NIR) hyperspectral imaging offers the ability for rapid sample measurement that provides both spatial and spectral information for each image of a given sample. The NIR region lies between 780 and 2500 nm, thus these images comprise numerous wavelength bands (Yao *et al.*, 2008). Through this, the chemical and structural information of a sample can be used as a fingerprint to aid in the identification thereof (Burger, 2006). The images are data rich, which makes the analysis thereof challenging.

Multivariate image analysis (MIA) has been shown to aid this problem by using methods such as principal component analysis (PCA) to reduce the dimensionality of the data (Esbensen *et al.*, 2002b; Prats-Montalbán *et al.*, 2011). PCA is used as an initial data exploration step. As hyperspectral imaging produces three-dimensional (3D) datasets (hypercubes), $(\lambda; x; y)$, the data is first unfolded by multiplying the x and y coordinates, which produces two-dimensional (2D) dataset $(\lambda; x \cdot y)$ or matrix. PCA decomposes the data set into a smaller number of variables, referred to as principal components (PC), which are composed of scores and loadings. These explain the original variance the data in descending order of PC's. As stated by Manley (2014), the variance captured in PC1 does not always represent the properties responsible for the separation between the samples and may be misleading and therefore looking at the lower PC's is also needed to determine which properties are truly responsible for the separation.

Partial least squares discriminant analysis (PLS-DA) is used as a classification method to model the relationship between the spectral data (X) and the pre-determined responses (Y) that assign the data to a specific fungal pathogen (Næs *et al.*, 2002; Gowen *et al.*, 2007). PLS-DA thus aims to find a straight line that divides the space into two regions that results in the separation of two classes. This method, when applied to multi-class models, separates the classes in higher order spatial directions (Brereton and Lloyd, 2014). Both PCA and PLS-DA ensure rapid sample screening that has the potential of distinguishing pathogen isolates within minutes through the construction of calibrated prediction models. These two methods can both be done using either a pixel or object wise approach as stated by Williams and Kucheryavskiy (2016). The object wise approach was found to be superior to that of the pixel wise approach as it groups the large volume of pixels in an image into separate smaller clusters taking the mean of the spectra as a data point (Kucheryavskiy, 2013).

Yao *et al.* (2008) showed that the differentiation of toxigenic fungi is possible on growth media, using NIR hyperspectral imaging, however, they focussed on fungal isolates from vastly different fungal species, including *Penicillium* spp., *Fusarium* spp. and *Aspergillus* spp. A study by Williams *et al.* (2012a) found that *Fusarium* spp. could be distinguished

from one another on PDA media by using both PCA and PLS-DA analysis, but the study only focussed on FER in maize and not the other two maize ear rot diseases GER and DER.

The previous study, Chapter 3, illustrated that it was in fact possible to distinguish between the major ear rot pathogens on day 5 of growth. It was seen that not all the pathogens could be distinguished from one another with 100% accuracy and it was therefore decided to investigate whether the growth age of the isolates influences the models.

The study is thus aimed at determining the influence of the growth age on the classification accuracy of the maize ear rot pathogens on growth media using NIR hyperspectral imaging. The two main objectives of this study were to (a) distinguish between the major ear rot pathogen species and (b) distinguish between the different isolates within the same species at three different growth ages, specifically day 3, day 5 and day 7.

4.3. MATERIALS AND METHODS

4.3.1. Sample Preparation

The fungal pathogens most commonly associated with maize ear rot in South Africa, specifically *Fusarium* spp. and *Stenocarpella* spp. were obtained from the Department of Plant Pathology at Stellenbosch University, South Africa. Eight isolates were used in varying combinations in this study, which included *Fusarium verticillioides* [MRC826, MRC8267 and MRC8559], *F. graminearum* s.s. [M14-55], *F. boothii* [M0002, M0010 and M0100], as well as *Stenocarpella maydis* [SM8]. The *Fusarium* spp. were stored at -80 °C in 15% glycerol and the *S. maydis* at room temperature in water. The details of the origin of the isolates can be seen in Chapter 3, Table 3.1. The pathogen isolates were plated on PDA in glass Petri dishes by placing a piece of previously inoculated PDA in the centre of the agar filled plates. Each of the isolates were plated out in triplicate, and incubated at 25°C for three, five and seven days as seen in Addendum B, Figure B1. The digital images of the pathogen isolates were taken on day 3 (Addendum B, Figure B2), day 5 (Addendum B, Figure B3) and day 7 (Addendum B, Figure B4) for comparison during the modelling.

4.3.2. NIR Hyperspectral Imaging and Analysis

A SisuChema short-wave infrared (SWIR) pushbroom imaging system (Specim Spectral Imaging Ltd, Oulu, Finland) with Chemadaq version 3.62.183.19 controlling software was used. The specifications of the hyperspectral imaging system were the same as stated in Chapter 3. Once the hyperspectral images were captured they were analysed using hyperspectral imaging software, namely Prediktera Evince v.2.7.0 (Umeå, Sweden).

4.3.2.1. Mosaic Construction

The hyperspectral images were merged together to form mosaics before proceeding with further analysis. Each mosaic consisted of three pathogens and their respective triplicates to form a nine-image merged mosaic representing three clusters of pathogens. Two replicates of each pathogen (B and C) were used for the calibration set, with the remaining one (A) forming the validation set. Each mosaic was then analysed individually using an object wise approach.

The aim of the first three mosaics was to distinguish between the three major ear rot pathogens, specifically *F. verticillioides* MRC826, *F. graminearum* s.s. M14-55 and *S. maydis* SM8. Mosaic one, two and three were created with images captured on day 3, 5 and 7 respectively. The second cluster of three mosaics consisted of *F. verticillioides* MRC826, *F. boothii* M0002 and *S. maydis* SM8. The GER causing pathogen was replaced due to *F. boothii* being the most prominent contributor to GER in South Africa (Boutigny *et al.*, 2011) after *F. graminearum* s.s. and both isolates produces the DON mycotoxin. Mosaic four, five and six was thus formed with images of these isolates captured on day 3, 5 and 7 respectively. Mosaic seven (day 3), eight (day 5) and nine (day 7) was formed with the aim of distinguishing between isolates within the same species, thus using *F. boothii* M0002, M0010 and M0100. Lastly, mosaic ten, eleven and twelve consisted of *F. verticillioides* MRC826, MRC8267 and MRC8559 images captured on day 3, 5 and 7 respectively.

4.3.2.2. Image Correction and Cleaning

The image calibration and absorbance correction was done according to equation 4.1 (Sendin, 2017) in the Evince software package before any further processing. The wavelengths before 1103 nm and after 2464 nm were also removed as it was associated with noise.

$$I_{\lambda,n} = -\log_{10} \left[\left(\frac{S_{\lambda,n} - B_{\lambda,n}}{W_{\lambda,n} - B_{\lambda,n}} \right) \right] \quad (\text{eq. 4.1})$$

Where:

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

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

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

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

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

The data was mean-centred prior to the calculation of the principal components. The score images and score plots were used interactively, together with the digital images (Addendum B, Figure 4.18 to 4.20), to select regions of interest. The images were randomly subdivided into twelve circular areas to increase the sample size used during further analysis, after which the PCA was recalculated. Figure 4.1 broadly illustrates the stages of image correction and cleaning.

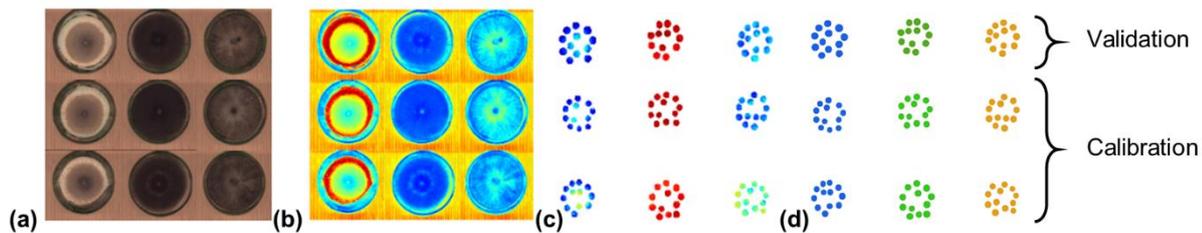


Figure 4.1 Step-wise approach used for the analysis of the various models. (a) RGB image of the mosaic, illustrating the triplicates of three pathogen isolates, (b) the uncleaned PC score image before the subdivision, (c) the PC score image of the subdivided isolates and (d) the PLS-DA score image illustrating the calibration and validation sets.

4.3.2.3. Spectral Analysis

The mean spectra of the pathogen isolates were calculated between 1103 nm and 2464 nm and plotted on three graphs, corresponding with their growth day, in order to compare all the isolates. The mean spectra were investigated to determine the chemical properties of the pathogen isolates.

4.3.2.4. Multivariate Image Analysis

Numerous pre-processing methods were assessed, after which it was concluded that only standard normal variate (SNV) transformation was needed apart from mean-centering applied to the data (Rinnan *et al.*, 2009).

In the previous chapter (chapter 3) the two approaches to analyse, namely pixel or object wise approach as stated by Williams and Kucheryavskiy (2016) were investigated. Based on those findings, only object wise analysis will be used in this study to classify maize ear rot pathogens.

Twelve PCA models were calculated, one for each mosaic, using object wise analysis as per equation 4.2. Each PCA model comprised of three PC's for consistency. The score and loading plots together with the score images were used to explore the data.

$$X = TP' + E \quad (\text{eq. 4.2})$$

Where:

X = Unfolded hypercube, or data matrix, obtained from the hyperspectral image

T = Scores matrix

P = Loadings matrix

E = Residual matrix

The PCA score images were used in conjunction with the loadings and score plots to illustrate the similarity, or lack thereof between the pathogen isolates resulting from their chemical characteristics. When applying a heat map to these images the similarity can be seen where similar colours represent closely related pathogens. The score plots visually illustrated the association between the pixels, where each point represented a single unit of measurement. The loading plots provided information related to the variables and illustrated which of these are responsible for the variance observed in the score plots the explanation

To distinguish the pathogen isolates from one another, PLS-DA was calculated for each mosaic using equation 4.3. PLS-DA classes the objects by using predetermined values and determining their orientation towards these values, thus modelling the relationship between the spectral data and the response variables (Amigo *et al.*, 2013). This results in a model that assigns each object to a pre-determined class, namely the pathogens isolates used.

$$y = X\beta + f \quad (\text{eq. 4.3})$$

Where:

y = Response matrix

X = Unfolded hypercube, or data matrix, obtained from the hyperspectral image

β = Beta coefficient

f = Residual or "error" matrix

Object wise PLS-DA models were calculated with four factors and full cross-validation was performed on each model. Two of the triplicate images for each isolate were used in the calibration of the model, with the last one forming the validation set as seen in Figure 4.1. After validating the models, classification images and prediction tables were obtained. The classification images and prediction tables were used to determine the amount of incorrect and correct classification predictions. The "no class" classification class,

which refers to pixels that have not been assigned to any of the existing classes, was regarded as false negatives for the purpose of this study.

Various parameters were calculated to assess and compare the twelve PLS-DA models based on how it distinguished the isolates from one another. The classification accuracy was calculated according to equation 4.5, and illustrates how effective the overall model is (Sokolova *et al.*, 2006). The sensitivity and specificity are two parameters that are used to assess the classification algorithm on a single class basis. The sensitivity, calculated per equation 4.6 refers to the probability that a given isolate in a model would be correctly classified, thus describing the effectiveness of the model on a class basis (Sokolova *et al.*, 2006). On the other hand, the specificity, as per equation 4.7, describes the probability that the remaining fungal isolates in the model would be correctly classified.

$$\text{Classification accuracy (\%)} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total}} \times 100\% \quad (\text{eq. 4.5})$$

$$\text{Sensitivity or Recall (\%)} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \times 100\% \quad (\text{eq. 4.6})$$

$$\text{Specificity (\%)} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \times 100\% \quad (\text{eq. 4.7})$$

$$\text{Precision (\%)} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \times 100\% \quad (\text{eq. 4.8})$$

$$\text{F1 Score (\%)} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (\text{eq. 4.9})$$

$$\text{Misclassification rate (\%)} = \frac{\text{False Negatives} + \text{False Positives}}{\text{Total}} \times 100\% \quad (\text{eq. 4.10})$$

Where:

True Positives = a specified isolate correctly classified as such

True Negatives = remaining isolates correctly classified as such

False Positives = remaining isolates incorrectly classified as a specified isolate

False Negatives = a specified isolate incorrectly classified as the remaining

Total = Sum of all the isolates in the given mosaic

The F1 score (equation 4.9) refers to the weighted mean between precision and recall (Sokolova *et al.*, 2006), where the precision is associated with the predictive power of the model seen in equation 4.8 and the recall is a synonym for sensitivity. Lastly, the misclassification rate was also calculated by equation 4.10 for each of the twelve mosaics.

4.4. RESULTS AND DISCUSSION

4.4.1. Spectral Analysis

The mean spectra of the eight pathogen isolates investigated in this study, were plotted on three graphs, each representing the three growth days, day 3 (Figure 4.2), day 5 (Figure 4.3) and day 7 (Figure 4.4).

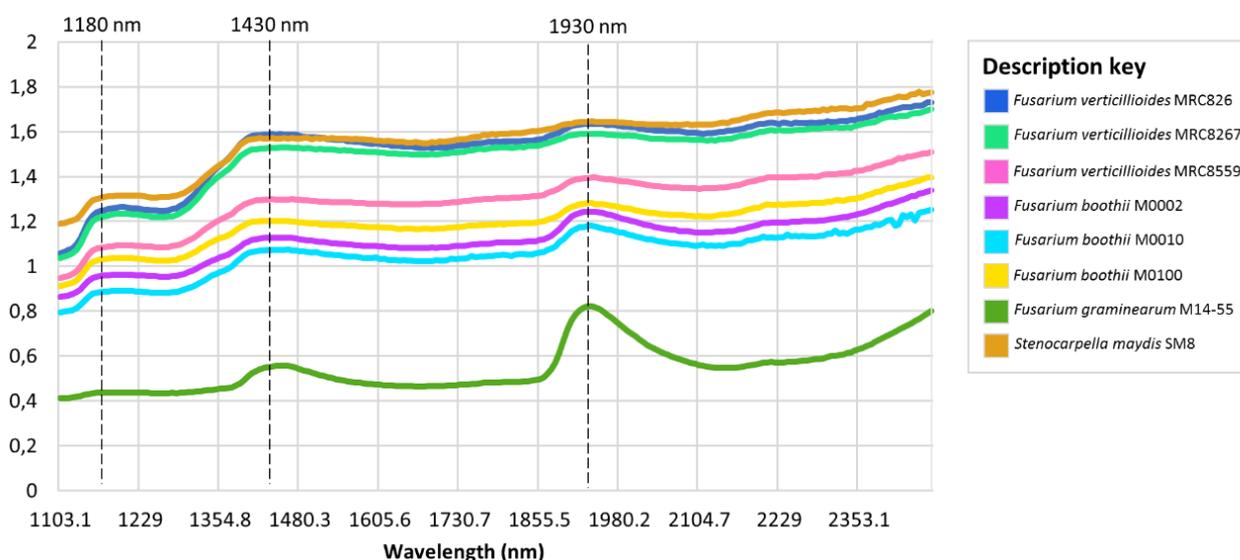


Figure 4.2 Mean spectra of the pathogen isolates on day 3 of growth, depicting *Fusarium verticillioides* MRC826 (blue), *F. verticillioides* MRC8267 (teal), *F. verticillioides* MRC8559 (pink), *F. boothii* M0010 (purple), *F. boothii* M0010 (light blue), *F. boothii* M0010 (yellow), *F. graminearum* s.s. M14-55 (green) and *Stenocarpella maydis* SM8 (orange), used in the different mosaics.

The mean spectra of the pathogen isolates on day 3 of growth illustrated the same absorption bands at 1180, 1430 and 1930 nm as seen in the day 5 mean spectra (Chapter 3, Figure 3.5). These three bands were the most prominent bands seen in all eight pathogens throughout the three growth days investigated.

The absorption band at 1180 nm is associated with the C-H second overtone, found in carbohydrates as reported by Delwiche and Hareland (2004), while Berardo *et al.* (2005) showed it was related to the carbonyl compounds found in *F. verticillioides* species. The band at 1430 nm is associated with the sucrose and starch content as stated by Berardo *et al.* (2005), which is related to the O-H stretch first overtone. On the other hand Osborne, Fearn, & Hindle, (1993) found that the 1430 nm band indicates the presence of protein [N-H stretch first overtone], their scope was food and beverage analysis and not pathogen

contamination as in the case of Berardo *et al.* (2005). The absorption band at 1930 nm band is associated with the O-H stretch and O-H deformation combination, as well as the O-H bend second overtone related to moisture and it is also related to the CONH group of protein [N-H stretch second overtone] (Osborne *et al.*, 1993; Berardo *et al.*, 2005; Manley, 2014). The separation of pathogen isolates is thus based on the biological composition thereof.

The intensity of the bands, specifically at 1930 nm, increased with the age of the colony (Figure 4.3), which is possibly a result of the moisture accumulation due to respiration.

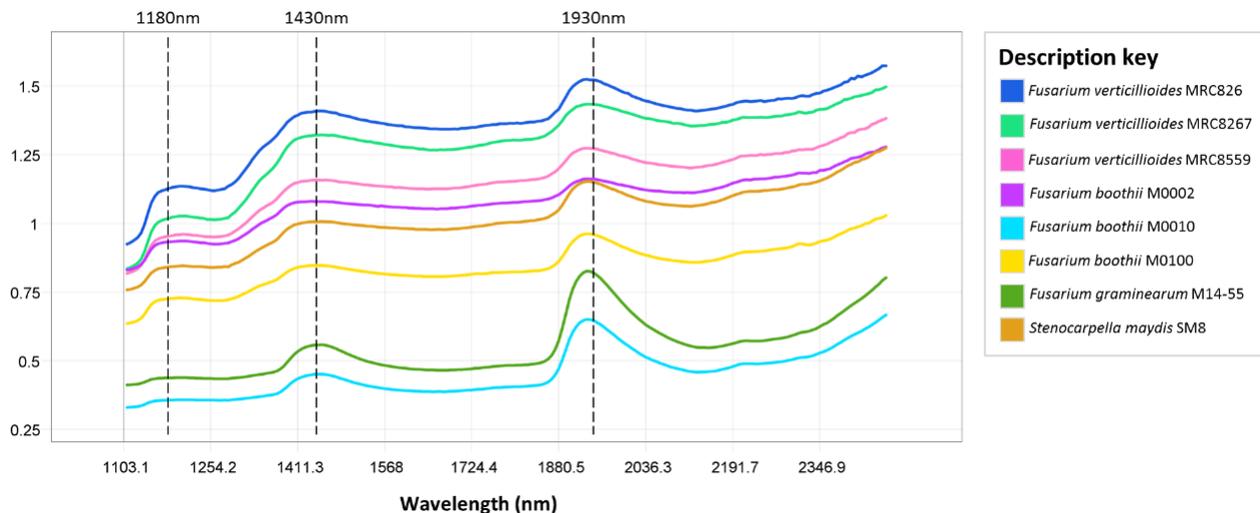


Figure 4.3 Mean spectra of the pathogen isolates on day 5 of growth, depicting *Fusarium verticillioides* MRC826 (blue), *F. verticillioides* MRC8267 (teal), *F. verticillioides* MRC8559 (pink), *F. boothii* M0002 (purple), *F. boothii* M0010 (light blue), *F. boothii* M0100 (yellow), *F. graminearum* s.s. M14-55 (green) and *Stenocarpella maydis* SM8 (orange), used in the different mosaics.

The mean spectra for each isolate on all three days were plotted on individual graphs in order to compare the effect of the age on the pathogen's spectra (Addendum B, Figure B5 to B12). It was noticed that for each isolate, the intensity of the band at 1930 nm increased as the age of the colony increased. This could be the result of the respiration of the pathogens, where the moisture associated with the O-H stretch and O-H deformation combination, as well as the O-H bend second overtone as stated by Osborne *et al.* (1993), is increased.

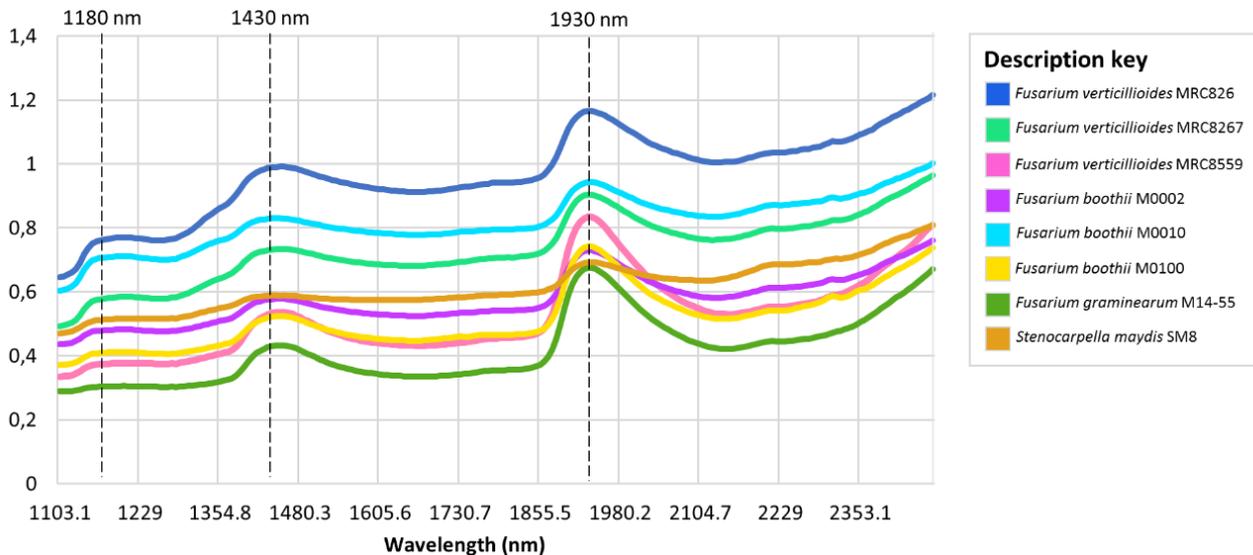


Figure 4.4 Mean spectra of the pathogen isolates on day 7 of growth, depicting *Fusarium verticillioides* MRC826 (blue), *F. verticillioides* MRC8267 (teal), *F. verticillioides* MRC8559 (pink), *F. boothii* M0010 (purple), *F. boothii* M0010 (light blue), *F. boothii* M0010 (yellow), *F. graminearum* s.s. M14-55 (green) and *Stenocarpella maydis* SM8 (orange), used in the different mosaics.

4.4.2. Multivariate Data Analysis

4.4.2.1. Principal Component Analysis

The differences between the pathogens in each mosaic were initially explored with PCA. The PCA scores and loading plots, together with the score images of each mosaic was obtained and is shown in Figure 4.5 (mosaic one) and Addendum B, Figure B13 to B23.

The scores plot of PC2 vs PC3 was used to identify the dead pixel lines, which resulted from a line of inactive pixels on the detector of the instrument. These were removed before continuing the analysis and interpretation of the data.

The score plot and score image of each mosaic was used to explore the differences between the isolates (Figure 4.5a, b and c). It illustrated isolate similarities after a heat map was applied, where positive score values are red (warm colours) and negative score values are blue (cool colours) corresponding to the loading plots positive and negative bands. The loading plots provide possible explanations for the variance observed in both the score plots and images based on the spectral points used during the PCA calculations. The variance is thus expressed as wavelength bands that contribute to pathogen separation. The positive bands in a loading plot correlate to the positive side of the corresponding PC's score plot and is an indication of the intensity of the spectral response (Esbensen *et al.*, 2002b). The same is seen with the negative bands correlating to the negative side of the score plot. The

PCA score plots (Figure 4.5a, b) that illustrate separation would thus be explained by the corresponding loading plots (Figure 4.5d, e), which would identify the corresponding wavelength band responsible for the variance.

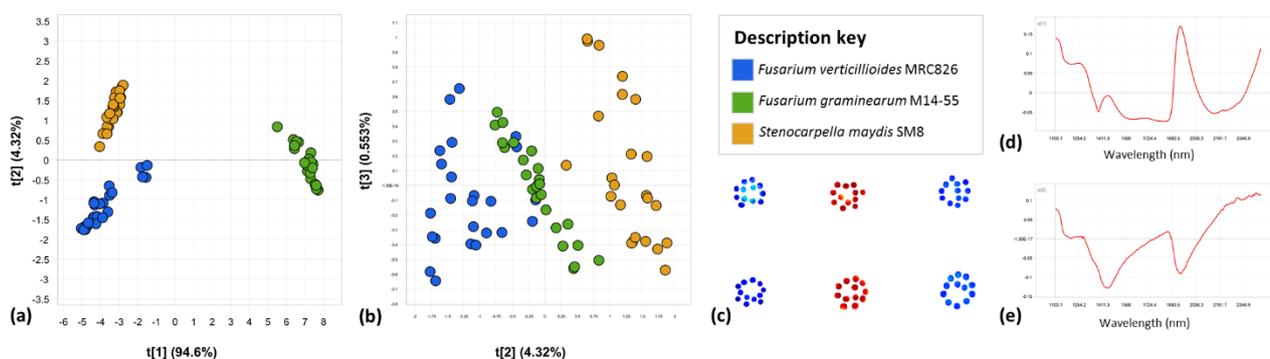


Figure 4.5 Object wise PCA of mosaic one (Day 3) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (94.6%) vs. PC2 (4.32%) and (b) PCA score plot of PC2 (4.32%) vs. PC3 (0.553%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. graminearum* s.s. M14-55 (middle) and *S. maydis* SM8 (right). The loading plots is given for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1930 nm.

The first three mosaics, comprising of *F. verticillioides* MRC826, *F. graminearum* s.s. M14-55 and *S. maydis* SM8, represented the three growth days, i.e. day 3 (Figure 4.5), day 5 (Addendum B, Figure B13) and day 7 (Addendum B, Figure B14). The score plot of mosaic one illustrating PC1 (94.6%) vs. PC2 (4.32%) showed very good separation between the three pathogen isolates (Figure 4.5a). The first PC accounted for most of the variance observed with the remaining variance (< 6%) covered by the subsequent PC's (Figure 4.5b). The large variance in the first PC could be attributed to the differences in the biological composition of the three pathogen isolates that results in a different type of ear rot formation on maize. The corresponding loading plots (Figure 4.5d, e) indicates the separation is due to a difference in carbohydrates, proteins and/or moisture.

The PC1 loading plot of mosaic one (Figure 4.5d) showed prominent bands at 1260 nm (positive), 1400 nm (negative) and 1930 nm (positive). This suggests that the separation between *F. graminearum* s.s. and the other two isolates, namely *F. verticillioides* and *S. maydis* is due to a difference in protein and moisture content. The 1260 nm band is associated with the C-H stretch second overtone found in CH₃ groups (Berardo *et al.*, 2005) and the 1930 nm band is related to the CONH group [N-H stretch second overtone]

corresponding to protein, as well as moisture [O-H stretch, O-H deformation combination and O-H bend second overtone] (Osborne *et al.*, 1993; Berardo *et al.*, 2005; Manley, 2014). The negative band at 1400 nm is also related to the moisture. The PC2 loading plot (Figure 4.5e) showed prominent negative peaks at 1430 and 1930 nm, thus suggesting the separation seen between *F. verticillioides* and *S. maydis* is also a result of variance in the starch and moisture content. The 1430 nm band is associated with starch and sucrose, and 1930 nm is related to protein as stated by Berardo *et al.* (2005).

The score images of all three mosaics (1, 2 and 3) confirms what was seen in the score plots (Figure 4.5a and Addendum B, Figure B13a and B14a), where clear differences in the first PC is seen when a heat map is applied. The applied heat map showed the similarity of the isolates through similar colours, where the red isolate *Fusarium graminearum* s.s. was clearly distinguishable from the other blue-yellow isolates, *F. verticillioides* and *S. maydis*. The two blue-yellow isolates have negative score values, which shows that their separation from *F. graminearum* s.s. is based on the 1400 nm band (Figure 4.5d). A difference in score values was also observed between *F. verticillioides* and *S. maydis*, as the latter tends to lean more to lighter blue-green shading. The same was seen in the score and loading plots of the day 5 (Addendum B, Figure B13) and day 7 (Addendum B, Figure B14) mosaics, i.e. mosaic two and three respectively.

The second set of mosaics consisted of *F. verticillioides* MRC826, *F. boothii* M0002 and *S. maydis* SM8, which represented the three maize ear rots, FER, GER and DER over the three growth days (Addendum B, Figure B15 to B17). The score images did not show prominent differences between isolates after a heat map was applied, although the score plots showed some separation. Mosaic four (day 3) and five (day 5) showed overlap between *F. boothii* and *S. maydis*, whereas mosaic six (day 7) clearly separated all three isolates in the score plot of PC1 vs PC2 (Addendum B, Figure B17a). The loading plots of PC1 and PC2 (Addendum B, Figure B17d, e) illustrated the same prominent bands as mosaic one to three, thus these bands are associated with maize ear rot pathogens overall.

The PCA results of the third set of mosaics, namely mosaic seven (day 3), eight (day 5) and nine (day 7) are shown in Addendum B, Figure B18 to B20. These results suggested that separation between isolates within the *F. boothii* species [M0002, M0010 and M0100] were possible as the colonies aged. The day 5 isolates of mosaic eight yielded the best separation (Addendum B, Figure B19) compared to that of cultures from day 3 and 7.

Lastly, mosaic ten, eleven and twelve consisted of three *F. verticillioides* isolates, namely MRC826, MRC8267 and MRC8559 on day 3, 5 and 7 respectively (Addendum B, Figure B21 to B23). As with the previous models, the separation observed in the score plots

are explained by the corresponding loadings plot, where the *F. verticillioides* MRC8559 isolate is separated due to a band at 1930 nm on PC1 (Addendum B, Figure B22a), which is related to the protein content as stated by Berardo *et al.* (2005). The driver of the separation between the *F. verticillioides* isolates is thus the protein and starch variation observed at the 1430 and 1930 nm bands as shown in Addendum B, Figure B22d (Osborne *et al.*, 1993; Berardo *et al.*, 2005).

It is unclear whether the moisture bands observed during the analysis of all twelve mosaics were a result of the constituents of the pathogen isolates, or a direct result of the moisture accumulation. The moisture accumulates on the inside of the petri dish due to pathogen respiration, followed by condensation resulting from external temperature fluctuations. The respiration is enhanced by the presence of the nutrient rich agar, which results in the formation of water and carbon dioxide. The condensation of the water produced during respiration was increased by the temperature fluctuations exerted on the pathogens due to transportation.

4.4.2.2. Partial Least Squares Discriminant Analysis

The classification results for the twelve mosaic models are shown in Figure 4.6 to 4.17 and Table 4.1 to 4.5. The PLS-DA score plots of PLS factor 1 vs. factor 2 (Figure 4.6a) and PLS factor 2 vs. factor 3 (Figure 4.6b) shaded according to the pathogen isolates were obtained for each mosaic model. The prediction image (Figure 4.6c) depicts the calibration set (row 2 and 3) and the validation set (row 1) together with an accompanying legend. The parameters used to assess the effectiveness of the overall classification models included the classification accuracy, misclassification rate, coefficient of determination (R^2) and the cross-validated coefficient of determination (Q^2) as shown in Table 4.1. The classification accuracy, sensitivity, specificity, F1 score and misclassification rate was also calculated for the individual isolates of each of the models (Table 4.2 to 4.5).

The PLS-DA model of mosaic one, two and three, consisting of *F. verticillioides* MRC826, *F. graminearum* s.s. M14-55 and *Stenocarpella maydis* SM8 isolates showed that all three models could predict the isolates with 100% accuracy and a R^2 above 0.95 (Table 4.1). The score plots of the three mosaics (Figure 4.6a, 4.7a and 4.8a) showed that the isolates were separated effectively with no overlap occurring. To distinguish the three major ear rot pathogens effectively, a growth age of three days (mosaic one) would yield accurate results and reduce the time needed for identification in the laboratory.

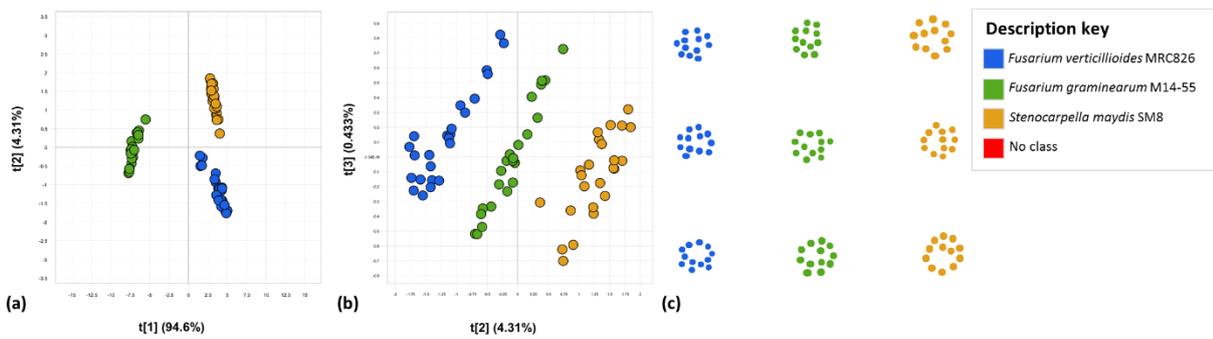


Figure 4.6 PLS-DA of mosaic one (Day 3) illustrating the (a) score plot of factor 1 (94.6%) vs factor 2 (4.31%) colour coded per pathogen isolate, the (b) score plot of factor 2 (4.31%) vs factor 3 (0.433%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. graminearum* s.s. M14-55 (green) and *S. maydis* SM8 (orange), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.

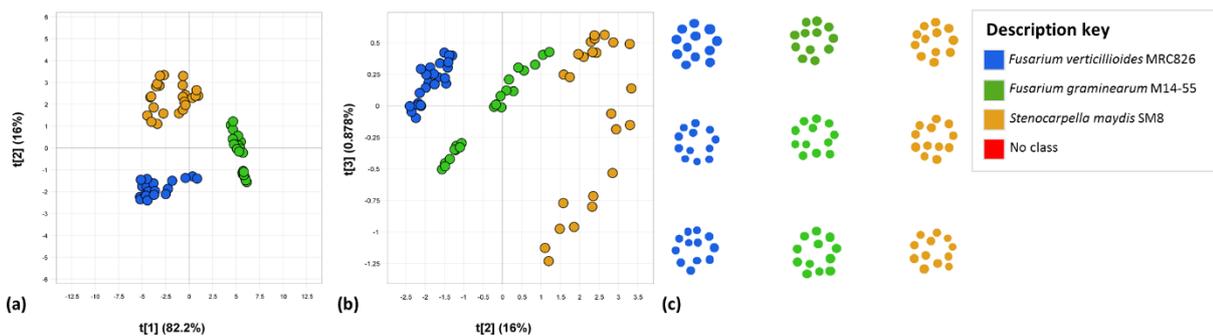


Figure 4.7 PLS-DA of mosaic two (Day 5) illustrating the (a) score plot of PLS factor 1 (82.2%) vs factor 2 (16%) colour coded per pathogen isolate, the (b) score plot of factor 2 (16%) vs factor 3 (0.878%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. graminearum* s.s. M14-55 (green) and *S. maydis* SM8 (orange), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.

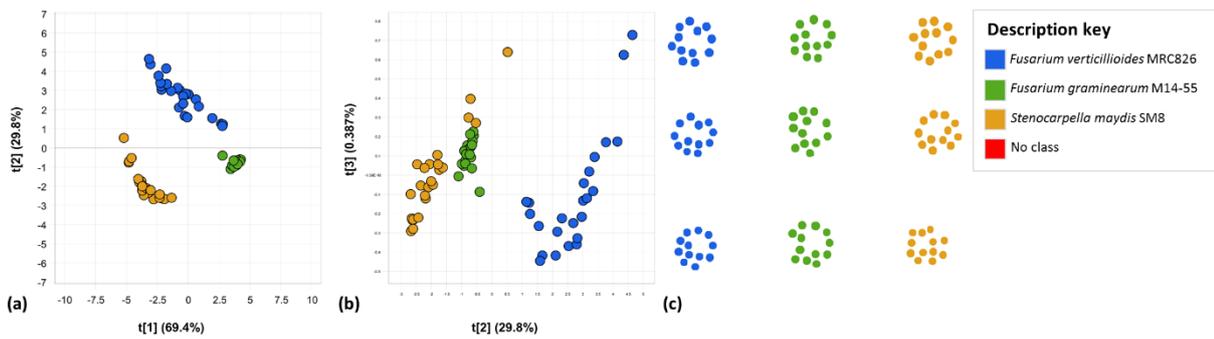


Figure 4.8 PLS-DA of mosaic three (Day 7) illustrating the (a) score plot of factor 1 (69.4%) vs factor 2 (29.8%) colour coded per pathogen isolate, the (b) score plot of factor 2 (29.8%) vs factor 3 (0.387%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. graminearum* s.s. M14-55 (green) and *S. maydis* SM8 (orange), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.

The models of mosaic four, five and six consisted of *F. verticillioides* MRC826, *F. boothii* M0002 and *S. maydis* SM8, which represented the three maize ear rots, FER, GER and DER over the three growth days (Figure 4.9 to 4.11). *F. graminearum* s.s. was replaced by the *F. boothii* isolate as both these isolates are the leading cause of GER and produce the same mycotoxins. The overall accuracy of mosaic four (75%), mosaic five (86%) and mosaic six (100%) suggested that a growth age of seven days would provide perfect classification (Table 4.1).

In practice, it is not necessary to have 100% accuracy as 86% would also be deemed an appropriate level to proceed with laboratory work based on the presence of possible mycotoxins. The decrease in the accuracy of mosaic four and five is mainly due to the misclassification of the objects assigned to the *F. boothii* class (Table 4.3). The sensitivity of *F. boothii* M0002 in mosaic five is 58%, yet the specificity is 100%, thus confirming that the model can accurately predict the remaining isolates although it struggles with *F. boothii* M0002 (Table 4.3). This phenomenon can be explained by investigating the digital images (Addendum B, Figure B2 and B3) of this specific isolate. It is clear that this isolate grows in circular patterns varying in growth density, thus hindering the separation between the isolate and agar.

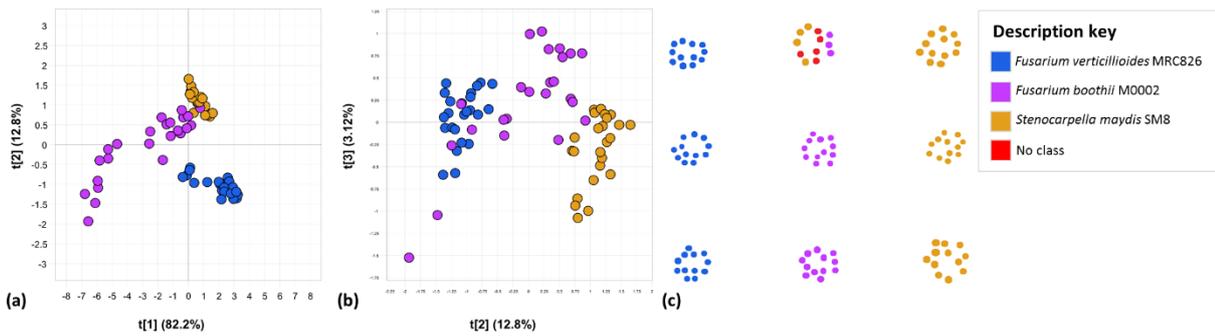


Figure 4.9 PLS-DA of mosaic four (Day 3) illustrating the (a) score plot of factor 1 (82.2%) vs factor 2 (12.8%) colour coded per pathogen isolate, the (b) score plot of factor 2 (12.8%) vs factor 3 (3.12%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. boothii* M0002 (purple) and *S. maydis* SM8 (orange), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.

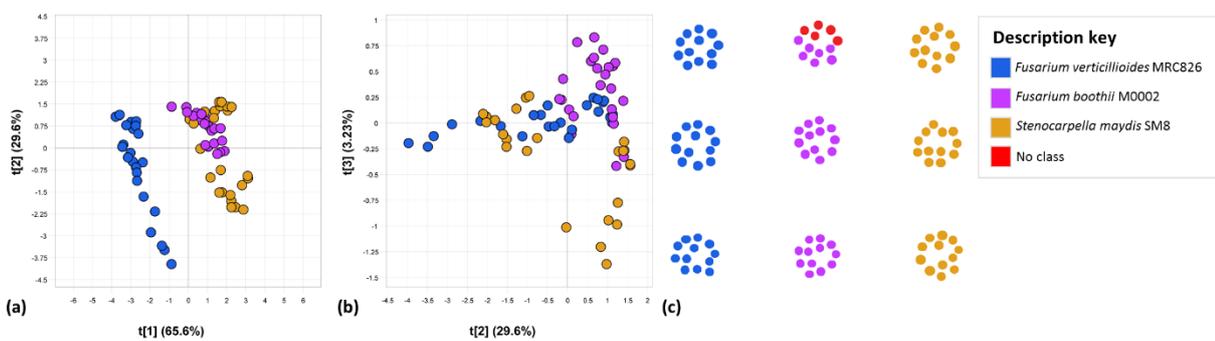


Figure 4.10 PLS-DA of mosaic five (Day 5) illustrating the (a) score plot of PLS factor 1 (65.6%) vs factor 2 (29.6%) colour coded per pathogen isolate, the (b) score plot of factor 2 (29.6%) vs factor 3 (3.23%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. boothii* M0002 (purple) and *S. maydis* SM8 (orange), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.

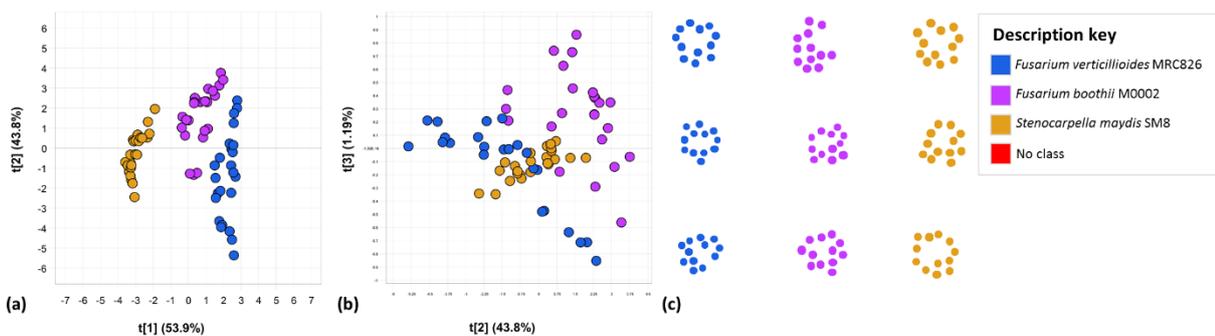


Figure 4.11 PLS-DA of mosaic six (Day 7) illustrating the (a) score plot of factor 1 (53.9%) vs factor 2 (43.8%) colour coded per pathogen isolate, the (b) score plot of factor 2 (43.8%)

vs factor 3 (1.19%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. boothii* M0002 (purple) and *S. maydis* SM8 (orange), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.

The PLS-DA models of the third set of mosaics, mosaic seven (day 3), eight (day 5) and nine (day 7) are shown in Figure 4.12 to 4.14. The results showed that the separation between the three *F. boothii* isolates [M0002, M0010 and M0100] was possible, with day 5 being the most accurate. The overall accuracy of mosaic seven (day 3) was only 22%, whereas mosaic eight and nine showed 89% and 69%, respectively (Table 4.1). When looking at the individual isolates in the three models (Table 4.4) it was observed that the classification accuracy of the isolates was low on day 3 (mosaic seven), which increased on day 5 (mosaic eight) and then decreased on day 7 (mosaic nine). The same trend was observed in the sensitivity, specificity and F1 score of the isolates. This could be due to the pathogen isolates producing the mycotoxins that, in effect, make them similar after seven days, thus reducing the accuracy. The isolates of *F. boothii* can thus be distinguished on day 5 of growth with 88% accuracy.

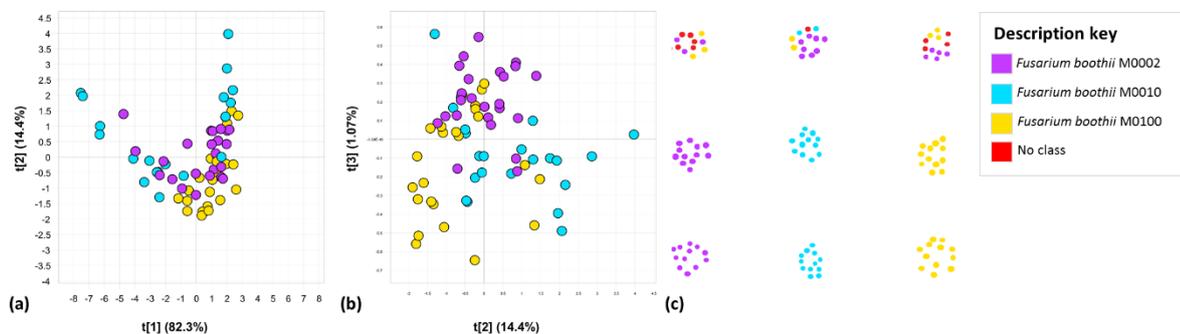


Figure 4.12 PLS-DA of mosaic seven (Day 3) illustrating the (a) score plot of factor 1 (82.3%) vs factor 2 (14.4%) colour coded per pathogen isolate, the (b) score plot of factor 2 (14.4%) vs factor 3 (1.07%) colour coded per pathogen isolate and the (c) classification image of *F. boothii* M0002 (purple), *F. boothii* M0010 (light blue) and *F. boothii* M0100 (yellow), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.

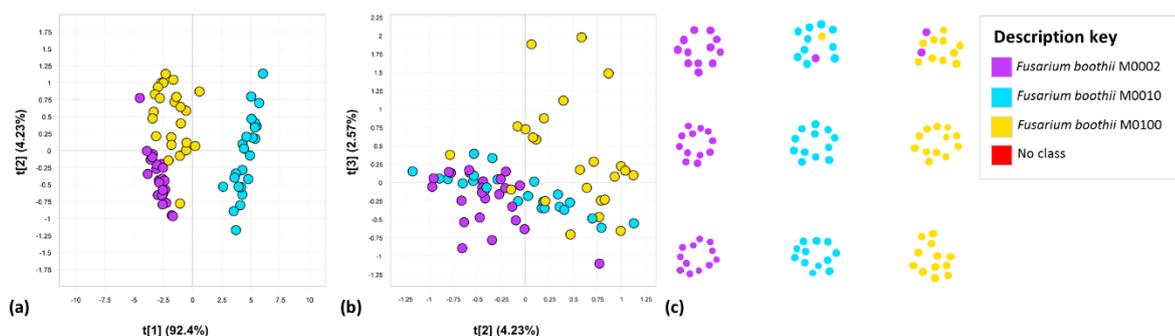


Figure 4.13 PLS-DA of mosaic eight (Day 5) illustrating the (a) score plot of factor 1 (92.4%) vs factor 2 (4.23%) colour coded per pathogen isolate, the (b) score plot of factor 2 (4.23%) vs factor 3 (2.57%) colour coded per pathogen isolate and the (c) classification image of *F. boothii* M0002 (purple), *F. boothii* M0010 (light blue) and *F. boothii* M0100 (yellow), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.

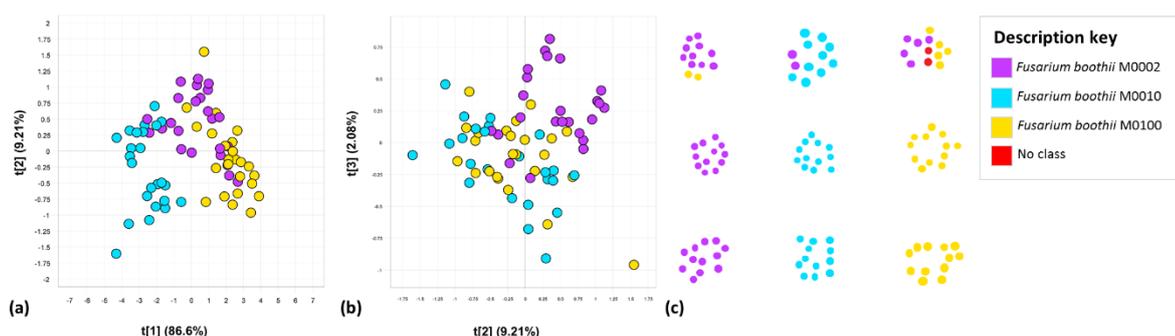


Figure 4.14 PLS-DA of mosaic nine (Day 7) illustrating the (a) score plot of factor 1 (86.6%) vs factor 2 (9.21%) colour coded per pathogen isolate, the (b) score plot of factor 2 (9.21%) vs factor 3 (2.08%) colour coded per pathogen isolate and the (c) classification image of *F. boothii* M0002 (purple), *F. boothii* M0010 (light blue) and *F. boothii* M0100 (yellow), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.

Lastly, the results of the PLS-DA models for the *F. verticillioides* isolates on day 3 (mosaic ten), day 5 (mosaic eleven) and day 7 (mosaic twelve) are depicted in Figure 4.15 to 4.17. The PLS-DA score plots of the three mosaics all showed good separation between the three isolates and was confirmed by the overall accuracies achieved by each model (Table 4.1). As with the previous three models, where isolates of *F. boothii* were distinguished from one another, it was seen that the classification accuracy was improved on day 5 from that seen in day 3, after which it decreased on day 7 again (Table 4.5). The overall accuracy of mosaic ten was already 97%, which improved to 100% with the addition

of two growth days. In practice, a sample would ideally be identified as soon as possible, thus it can be stated that the *F. verticillioides* isolates can be distinguished on day 3 of growth with 97% accuracy.

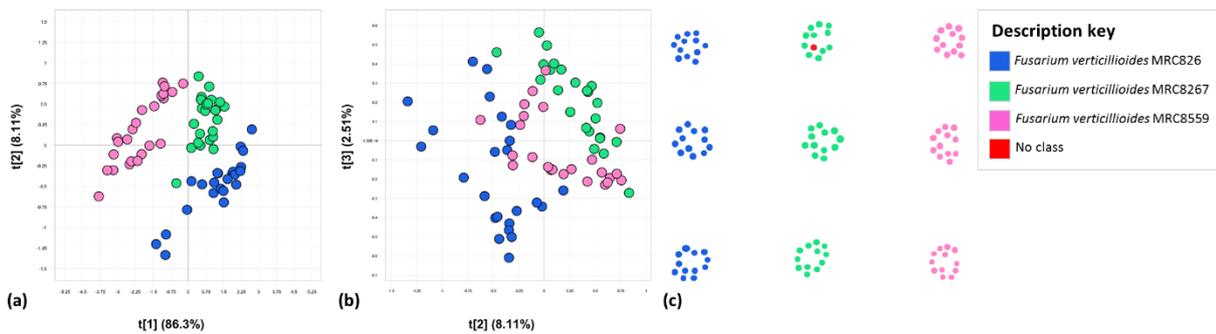


Figure 4.15 PLS-DA of mosaic ten (Day 3) illustrating the (a) score plot of factor 1 (86.3%) vs factor 2 (8.11%) colour coded per pathogen isolate, the (b) score plot of factor 2 (8.11%) vs factor 3 (2.51%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. verticillioides* MRC8267 (teal) and *F. verticillioides* MRC8559 (pink), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.

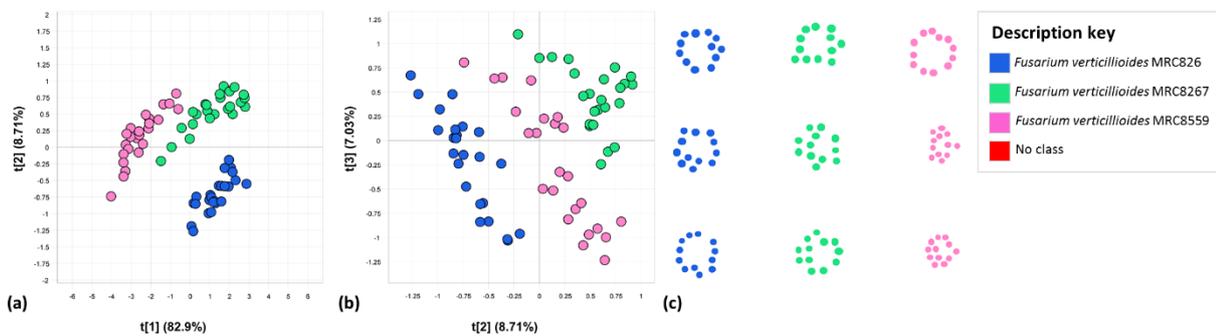


Figure 4.16 PLS-DA of mosaic eleven (Day 5) illustrating the (a) score plot of factor 1 (82.9%) vs factor 2 (8.71%) colour coded per pathogen isolate, the (b) score plot of factor 2 (8.71%) vs factor 3 (7.03%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. verticillioides* MRC8267 (teal) and *F. verticillioides* MRC8559 (pink), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.

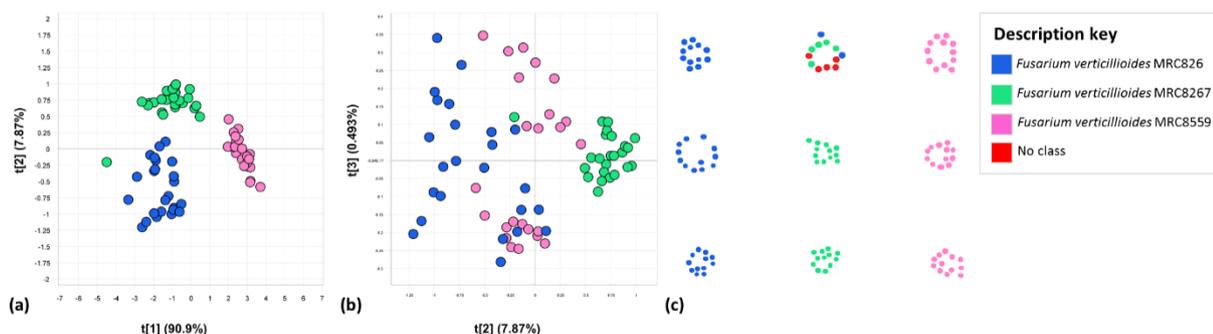


Figure 4.17 PLS-DA of mosaic twelve (Day 7) illustrating the (a) score plot of factor 1 (90.9%) vs factor 2 (7.87%) colour coded per pathogen isolate, the (b) score plot of factor 2 (7.87%) vs factor 3 (0.493%) colour coded per pathogen isolate and the (c) classification image of *F. verticillioides* MRC826 (blue), *F. verticillioides* MRC8267 (teal) and *F. verticillioides* MRC8559 (pink), illustrating the predicted objects, with the pathogen categories provided in the legend, where the first row was used for validation and the second and third for calibration.

Table 4.1 Performance measures used to assess the overall PLS-DA models of the pathogen isolates in the twelve different mosaics.

Mosaic	Day	R ²	Q ²	Classification accuracy (%)	Misclassification rate (%)
1	3	0,9852	0,9830	100,00	0,00
2	5	0,9702	0,9590	100,00	0,00
3	7	0,9592	0,9472	100,00	0,00
4	3	0,8249	0,7967	75,00	25,00
5	5	0,8680	0,8410	86,11	13,89
6	7	0,8619	0,8237	100,00	0,00
7	3	0,3979	0,2501	22,22	77,78
8	5	0,8562	0,8081	88,89	11,11
9	7	0,6798	0,6095	69,44	30,56
10	3	0,8315	0,7937	97,22	2,78
11	5	0,9418	0,9336	100,00	0,00
12	7	0,9139	0,8967	80,56	19,44

Table 4.2 Performance measures used to assess the PLS-DA models of the three pathogen isolates in mosaic 1, 2 and 3.

Mosaic	Day	Pathogen isolate	Classification accuracy (%)	Sensitivity (%)	Specificity (%)	F1 Score (%)	Misclassification rate (%)
1	3	<i>F. verticillioides</i> MRC826	100,00	100,00	100,00	100,00	0,00
		<i>F. graminearum</i> s.s. M14-55	100,00	100,00	100,00	100,00	0,00
		<i>S. maydis</i> 8	100,00	100,00	100,00	100,00	0,00
2	5	<i>F. verticillioides</i> MRC826	100,00	100,00	100,00	100,00	0,00
		<i>F. graminearum</i> s.s. M14-55	100,00	100,00	100,00	100,00	0,00
		<i>S. maydis</i> 8	100,00	100,00	100,00	100,00	0,00
3	7	<i>F. verticillioides</i> MRC826	100,00	100,00	100,00	100,00	0,00
		<i>F. graminearum</i> s.s. M14-55	100,00	100,00	100,00	100,00	0,00
		<i>S. maydis</i> 8	100,00	100,00	100,00	100,00	0,00

Table 4.3 Performance measures used to assess the PLS-DA models of the three pathogen isolates in mosaic 4, 5 and 6.

Mosaic	Day	Pathogen isolate	Classification accuracy (%)	Sensitivity (%)	Specificity (%)	F1 Score (%)	Misclassification rate (%)
4	3	<i>F. verticillioides</i> MRC826	100,00	100,00	100,00	100,00	0,00
		<i>F. boothii</i> M0002	75,00	37,50	85,71	40,00	25,00
		<i>S. maydis</i> 8	87,1	75,00	100,00	85,71	12,9
5	5	<i>F. verticillioides</i> MRC826	100,00	100,00	100,00	100,00	0,00
		<i>F. boothii</i> M0002	86,11	58,33	100,00	73,68	13,89
		<i>S. maydis</i> 8	100,00	100,00	100,00	100,00	0,00
6	7	<i>F. verticillioides</i> MRC826	100,00	100,00	100,00	100,00	0,00
		<i>F. boothii</i> M0002	100,00	100,00	100,00	100,00	0,00
		<i>S. maydis</i> 8	100,00	100,00	100,00	100,00	0,00

Table 4.4 Performance measures used to assess the PLS-DA models of the three *F. boothii* isolates in mosaic 7, 8 and 9.

Mosaic	Day	Pathogen isolate	Classification accuracy (%)	Sensitivity (%)	Specificity (%)	F1 Score (%)	Misclassification rate (%)
7	3	<i>F. boothii</i> M0002	27,59	14,29	62,50	22,22	72,41
		<i>F. boothii</i> M0010	44,44	66,67	40,00	28,57	55,56
		<i>F. boothii</i> M0100	36,36	25,00	50,00	30,00	63,64
8	5	<i>F. boothii</i> M0002	91,43	80,00	100,00	88,89	8,57
		<i>F. boothii</i> M0010	94,12	100,00	91,67	90,91	5,88
		<i>F. boothii</i> M0100	91,43	90,91	91,67	86,96	8,53
9	7	<i>F. boothii</i> M0002	73,53	58,82	88,24	68,97	26,47
		<i>F. boothii</i> M0010	92,59	100,00	88,24	90,91	7,41
		<i>F. boothii</i> M0100	73,53	55,56	80,00	52,63	26,47

Table 4.5 Performance measures used to assess the PLS-DA models of the *F. verticillioides* isolates in mosaic 10, 11 and 12.

Mosaic	Day	Pathogen isolate	Classification accuracy (%)	Sensitivity (%)	Specificity (%)	F1 Score (%)	Misclassification rate (%)
10	3	<i>F. verticillioides</i> MRC826	100,00	100,00	100,00	100,00	0,00
		<i>F. verticillioides</i> MRC8267	97,72	91,67	100,00	95,65	2,78
		<i>F. verticillioides</i> MRC8559	100,00	100,00	100,00	100,00	0,00
11	5	<i>F. verticillioides</i> MRC826	100,00	100,00	100,00	100,00	0,00
		<i>F. verticillioides</i> MRC8267	100,00	100,00	100,00	100,00	0,00
		<i>F. verticillioides</i> MRC8559	100,00	100,00	100,00	100,00	0,00
12	7	<i>F. verticillioides</i> MRC826	93,55	85,71	100,00	92,31	6,45
		<i>F. verticillioides</i> MRC8267	80,56	50,00	92,31	58,82	19,44
		<i>F. verticillioides</i> MRC8559	100,00	100,00	100,00	100,00	0,00

4.5. CONCLUSION

Day 3 could be used with confidence for the accurate classification of the major maize ear rot pathogens using NIR hyperspectral imaging. *Fusarium verticillioides*, *F. graminearum*, *F. boothii* and *Stenocarpella maydis* could effectively be distinguished from one another with more than 75% accuracy. It was also found that the three *F. verticillioides* isolates could be distinguished on day 3 of growth with 97% accuracy. The *F. boothii* isolates could not be distinguished on Day 3, however, an increase in the classification accuracy was observed on Day 5. A general trend was observed in the accuracy of the models that consisted of isolates from the same species, where day 5 showed higher accuracy than that of day 3 or day 7. This could be due to the pathogen isolates producing the mycotoxins that make them similar after seven days, thus reducing the accuracy. It can thus be deduced that the growth age does influence the accuracy of the models. Although in practice, it is not necessary to have 100% accuracy as only an appropriate level (>85%) is needed to proceed with subsequent laboratory work if this method was applied as a rapid screening tool. To further this work, focus should be placed on the classification of these pathogens on maize grain, as growth media is a step towards reaching a non-destructive pathogen detection method that can provide rapid information regarding the potential of mycotoxin contamination of maize.

4.6. REFERENCES

- Amigo, J.M., Martí, I. & Gowen, A. (2013). Hyperspectral Imaging and Chemometrics: A Perfect Combination for the Analysis of Food Structure, Composition and Quality. In: *Data Handling in Science and Technology*, Pp. 343–370.
- Berardo, N., Pisacane, V., Battilani, P., Scandolara, A., Pietri, A. & Marocco, A. (2005). Rapid detection of kernel rots and mycotoxins in maize by near-infrared reflectance spectroscopy. *Journal of Agricultural and Food Chemistry*, **53**, 8128–8134.
- Beukes, I., Rose, L.J., Shephard, G.S., Flett, B.C. & Viljoen, A. (2017). Mycotoxigenic *Fusarium* species associated with grain crops in South Africa – A review. *South African Journal of Science*, **113**, 1–12.
- Boutigny, A., Ward, T.J., Coller, G.J. Van, Flett, B., Lamprecht, S.C., O'Donnell, K. & Viljoen, A. (2011). Analysis of the *Fusarium graminearum* species complex from wheat, barley and maize in South Africa provides evidence of species-specific differences in host preference. *Fungal Genetics and Biology*, **48**, 914–920.

- Boutigny, A.L., Beukes, I., Small, I., Zühlke, S., Spiteller, M., Van Rensburg, B.J., Flett, B. & Viljoen, A. (2012). Quantitative detection of *Fusarium* pathogens and their mycotoxins in South African maize. *Plant Pathology*, **61**, 522–531.
- Brereton, R.G. & Lloyd, G.R. (2014). Partial least squares discriminant analysis: taking the magic away. *Journal of Chemometrics*, **28**, 213–225.
- Burger, J.E. (2006). *Hyperspectral NIR Image Analysis*. PhD Thesis. Swedish University of Agricultural Sciences, Umeå.
- Delwiche, S.R. & Hareland, G.A. (2004). Detection of scab-damaged hard red spring wheat kernels by near-infrared reflectance. *Cereal Chemistry*, **81**, 643–649.
- EPPO. (2017). *Stenocarpella macrospora* and *Stenocarpella maydis* [Internet document] URL https://www.eppo.int/QUARANTINE/data_sheets/.../DIPDSP_ds.pdf. Accessed 24/07/2017.
- Esbensen, K.H., Guyot, D., Westad, F. & Houmøller, L.P. (2002). Principal Component Analysis (PCA) - Introduction. In: *Multivariate data analysis in practice: an introduction to multivariate data analysis and experimental design*. 5th edn. CAMO Process AS.
- Gokul, A. (2015). *Fusarium graminearum species complex (FGSC) composition in South African wheat and maize grown in rotation*. MSc Thesis. Stellenbosch University, Stellenbosch.
- Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Downey, G. & Frias, J.M. (2007). Hyperspectral imaging - an emerging process analytical tool for food quality and safety control. *Trends in Food Science and Technology*, **18**, 590–598.
- Kucheryavskiy, S. (2013). A new approach for discrimination of objects on hyperspectral images. *Chemometrics and Intelligent Laboratory Systems*, **120**, 126–135.
- Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chemical Society Reviews*, **43**, 8200–8214.
- Munkvold, G.P. (2003). Cultural and genetic approaches to managing mycotoxins in maize. *Annual Review of Phytopathology*, **41**, 99–116.
- Næs, T., Isaksson, T., Fearn, T. & Davies, T. (2002). Data compression by PCR and PLS. In: *A user-friendly guide to multivariate calibration and classification*. Pp. 27-38. Chichester, U.K.: NIR Publications.
- Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). *Practical NIR spectroscopy with applications in food and beverage analysis*. 2nd edn. Longman Scientific & Technical, Essex.
- Peraica, M., Radić, B., Lucić, A. & Pavlović, M. (1999). Toxic effects of mycotoxins in humans. *Bulletin of the World Health Organization*, **77**, 754–766.

- Prats-Montalbán, J.M., De Juan, A. & Ferrer, A. (2011). Multivariate image analysis: A review with applications. *Chemometrics and Intelligent Laboratory Systems*, **107**, 1-23.
- Rinnan, A., Van Den Berg, F. & Engelsen, S.B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, **28**.
- Rossouw, J.D., Rensburg, J.B.J. van & Deventer, C.S. van. (2002). Breeding for resistance to ear rot of maize, caused by *Stenocarpella maydis* (Berk) Sutton. 1. Evaluation of selection criteria. *South African Journal of Plant and Soil*, **19**, 182–187.
- Sendin, K. (2017). *Characterisation of whole white maize kernels*. MSc Thesis. Stellenbosch University, Stellenbosch.
- Snyder, W.C. & Hansen, H.N. (1940). The Species Concept in *Fusarium*. *American Journal of Botany*, **27**, 64–67.
- Sokolova, M., Japkowicz, N. & Szpakowicz, S. (2006). Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In: *Advances in Artificial Intelligence: 19th Australian Joint Conference on Artificial Intelligence* (edited by A. Sattar & B. Kang). Pp. 1015–1021. Springer Berlin Heidelberg.
- Strausbaugh, C.A., Overturf, K. & Koehn, A.C. (2005). Pathogenicity and real-time PCR detection of *Fusarium* spp. in wheat and barley roots. *Canadian Journal of Plant Pathology*, **27**, 430–438.
- ThermoFischer Scientific. (2017). Overview of ELISA [Internet document] URL <https://www.thermofisher.com/za/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-elisa.html>. Accessed 27/09/2017.
- Turner, N.W., Bramhmbhatt, H., Szabo-Vezse, M., Poma, A., Coker, R. & Piletsky, S.A. (2015). Analytical methods for determination of mycotoxins: An update (2009-2014). *Analytica Chimica Acta*, **901**, 12–33.
- 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. Stellenbosch University, Stellenbosch.
- Williams, P.J., Geladi, P., Britz, T.J. & Manley, M. (2012). Near-infrared (NIR) hyperspectral imaging and multivariate image analysis to study growth characteristics and differences between species and strains of members of the genus *Fusarium*. *Analytical and Bioanalytical Chemistry*, **404**, 1759–1769.
- Williams, P.J. & Kucheryavskiy, S. (2016). Classification of maize kernels using NIR hyperspectral imaging. *Food Chemistry*, **209**, 131–138.

Yao, H., Hruska, Z., Kincaid, R., Brown, R.L. & Cleveland, T.E. (2008). Differentiation of toxigenic fungi using hyperspectral imagery. *Sensing and Instrumentation for Food Quality and Safety*, **2**, 215–224.

CHAPTER 5

General Discussion and Conclusion

Maize is one of the main dietary staple foods of Southern Africa. It is readily consumed by people and used in animal feed (Hellin *et al.*, 2012; Williams *et al.*, 2012). Unfortunately, it is also highly susceptible to both symptomatic and asymptomatic contamination by *Fusarium* spp. and *Stenocarpella* spp. These pathogens infect South African maize ears and result in three main maize ear rots, namely Fusarium ear rot (FER), Gibberella ear rot (GER) and Diplodia ear rot (DER). The presence of these pathogens allows the production of mycotoxins within the contaminated plants. *Fusarium verticillioides* mainly produces fumonisins, whereas the *Fusarium graminearum* species complex (FGSC) produces deoxynivalenol, among others. *Stenocarpella* spp. produces the toxin referred to as diplonin. These mycotoxins are a threat to the maize industry in South Africa as it can enter the food chain possibly resulting in cancer-promoting activity in both humans and animals. It was found that fumonisins lead to oesophageal cancer in humans and its carcinogenic properties cause liver cancer in rats (Yao *et al.*, 2008), as well as leukoencephalomalacia in horses and pulmonary edema syndrome in pigs (Gelderblom *et al.*, 2001).

With consumers placing a high demand on food safety, it was necessary to investigate a fast, efficient method that would determine the safety of maize grain. Near infrared (NIR) hyperspectral imaging addresses this by offering a non-destructive and rapid alternative scanning method. As NIR hyperspectral imaging is being used more frequently to distinguish pathogens and as an initial step towards the application in industry, studies have focussed on growth media (Yao *et al.*, 2008; Del Fiore *et al.*, 2010; Williams *et al.*, 2012b; Foca *et al.*, 2016; Kammies *et al.*, 2016; Seo *et al.*, 2016). This method was thus investigated as an alternative tool in laboratories for food analysis, due to current laboratory methods being too lengthy. Although the ideal application would be to apply this technology on maize grain, this study first evaluated the possibility of distinguishing between the ear rot pathogens on growth media.

In this study near infrared (NIR) hyperspectral imaging, together with statistical and multivariate image analysis (MIA), was used to distinguish between the major maize ear rot pathogens on growth media. Exploratory principal component analysis (PCA) was successful in reducing the dimensionality of the data and provided an essential initial step for hyperspectral image analysis. Although PCA showed that distinguishing between the pathogens was possible, it was necessary to apply classification algorithms in order to

predict unknown samples. One such algorithm is partial least squares discriminant analysis (PLS-DA), which can be used to predict the pathogens present in unknown samples.

MIA has generally been performed using a pixel wise approach, although the use of an object wise approach has increased (Kucheryavskiy, 2013; Williams & Kucheryavskiy, 2016; Sendin, 2017). During pixel wise analysis, the spectrum of each pixel present in a hyperspectral image is used as a unit of measurement for the classification of the data. Object wise analysis, on the other hand, uses a mean spectrum from a group of pixels to form an object which then acts as the smallest unit of measurement. Object wise analysis offers the ability to average out variation found within a given sample and allows for a sample to be seen as a “whole”. The two approaches were compared during this study and it was found that the object wise approach achieved a higher classification accuracy in the models tested. Using the object wise approach, it was possible to classify the pathogens with 100% accuracy, while none of the pixel wise models were error free.

This study initially found that the major maize ear rot pathogens, *Fusarium verticillioides*, *F. graminearum* s.s., *F. boothii* and *Stenocarpella maydis*, could be distinguished on day 5 of growth with reasonable accuracy. Furthermore, the three *F. verticillioides* isolates and three *F. boothii* isolates could also be distinguished from one another with good accuracy. This showed that not only could NIR hyperspectral imaging distinguish between different ear rot species, but also isolates from the same species. Compared to the study by Williams (2013) where poor separation between *F. verticillioides* isolates were obtained, this study proved that the *F. verticillioides* isolates could be distinguished with greater accuracy by using object wise analysis. Few studies have determined whether the age of a culture affects the accuracy of the classification and therefore a gap in the research was located.

The influence of the growth age on the classification accuracy of the maize ear rot pathogens was later investigated to ensure that the NIR hyperspectral imaging could be used as effectively as possible. From this it was found that after 3 days of growth (day 3) the major ear rot pathogens in the first and second combinations of *Fusarium verticillioides*, *F. graminearum* s.s. and *Stenocarpella maydis*, as well as *Fusarium verticillioides*, *F. boothii* and *Stenocarpella maydis*, could confidently and accurately be distinguished. Distinguishing the fungi on younger growth ages would allow for more rapid results when applying this method in a microbiology laboratory.

The three *F. verticillioides* isolates could also be distinguished from one another on day 3 with 97% accuracy, whereas the *F. boothii* isolates could only effectively be distinguished on day 5. It was observed that the accuracy of distinguishing between the

isolates from the same species was higher on day 5 compared to day 3 or day 7. This could be due to the pathogens isolates producing the same mycotoxins, thus chemically making their signature similar after seven days.

Throughout the study, the differentiation of *F. boothii* was deemed difficult with the poor accuracies obtained, as the pathogen exhibits circular growth varying in density. The less dense rings allowed the growth media to be more prominent, which influenced the spectra of the samples. This was not a problem for the remaining isolates used during the study. Even though the *F. boothii* isolates were difficult to distinguish, it was still possible to achieve separation.

Although the growth age does influence the accuracy of the models, in practice it is not necessary to have 100% accuracy. Only an appropriate level of accuracy (> 75%) should be required to proceed with subsequent laboratory work, as this method would be applied as a rapid screening tool, initially on growth media and later maize kernels. The suggested method could therefore be applied after 3 days to determine the pathogen species present in a sample. The method could thus be used as a rapid indication screening test in the microbiological laboratory, after which subsequent specialized testing methods could be applied to determine the exact species and isolates present in the sample.

NIR hyperspectral imaging is an excellent tool for the microbiology laboratory as it enables rapid screening and thus identification and differentiation of the pathogens. This work should be furthered by investigating more combinations of pathogens, as well as different growth media and growth ages. Furthermore, it should also be extended to the classification of maize ear rot pathogens on maize kernels, which would then allow for a non-destructive, pathogen detection method that would provide further information regarding potential mycotoxin contamination. The multivariate image analysis should also be expanded by investigating different analysis techniques such as soft independent modelling of class analogy (SIMCA) and support vector machines (SVM) to investigate the possibility that more accurate differentiation can be achieved.

NIR hyperspectral imaging offers a technique that can rapidly distinguish between the major ear rot pathogens and isolates of the same species accurately and reliably. This study provides the maize industry with an initial step towards applying the technology to maize grain, thus contributing to the food safety and grading of the maize grain.

5.1. REFERENCES

- Del Fiore, A., Reverberi, M., Ricelli, A., Pinzari, F., Serranti, S., Fabbri, A.A., Bonifazi, G. & Fanelli, C. (2010). Early detection of toxigenic fungi on maize by hyperspectral imaging analysis. *International Journal of Food Microbiology*, **144**, 64–71.
- Foca, G., Ferrari, C., Ulrici, A., Sciutto, G., Prati, S., Morandi, S., Brasca, M., Lavermicocca, P., Lanteri, S. & Oliveri, P. (2016). The potential of spectral and hyperspectral-imaging techniques for bacterial detection in food: A case study on lactic acid bacteria. *Talanta*, **153**, 111–119.
- Gelderblom, W.C.A., Seier, J.V., Snijman, P.W., Van Schalkwyk, D.J., Shephard, G.S. & Marasas, W.F.O. (2001). Toxicity of culture material of *Fusarium verticillioides* strain MRC 826 to nonhuman primates. *Environmental Health Perspectives*, **109**, 267–276.
- Hellin, J., Shiferaw, B., Cairns, J.E., Reynolds, M.P., Ortiz-monasterio, I., Banziger, M., Sonder, K. & La Rovere, R. (2012). Climate Change and Food Security in the Developing World: Potential of Maize and Wheat Research to Expand Options for Adaptation and Mitigation. *Journal of Development and Agricultural Economics*, **4**, 311–321.
- Kammies, T., Manley, M., Gouws, P.A. & Williams, P.J. (2016). Differentiation of foodborne bacteria using NIR hyperspectral imaging and multivariate data analysis. *Applied Microbiology and Biotechnology*, **100**, 9305–9320.
- Kucheryavskiy, S. (2013). A new approach for discrimination of objects on hyperspectral images. *Chemometrics and Intelligent Laboratory Systems*, **120**, 126–135.
- Sendin, K. (2017). *Characterisation of whole white maize kernels*. MSc Thesis. Stellenbosch University, Stellenbosch.
- Seo, Y., Park, B., Hinton, A., Yoon, J.S. & Lawrence, K.C. (2016). Identification of *Staphylococcus* species with hyperspectral microscope imaging and classification algorithms. *Journal of Food Measurement and Characterization*, **10**, 253–263.
- 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. Stellenbosch University, Stellenbosch.
- Williams, P.J., Geladi, P., Britz, T.J. & Manley, M. (2012a). Investigation of fungal development in maize kernels using NIR hyperspectral imaging and multivariate data analysis. *Journal of Cereal Science*, **55**, 272–278.

- Williams, P.J., Geladi, P., Britz, T.J. & Manley, M. (2012b). Growth characteristics of three *Fusarium* species evaluated by near-infrared hyperspectral imaging and multivariate image analysis. *Applied Microbiology and Biotechnology*, **96**, 803–813.
- Williams, P.J. & Kucheryavskiy, S. (2016). Classification of maize kernels using NIR hyperspectral imaging. *Food Chemistry*, **209**, 131–138.
- Yao, H., Hruska, Z., Kincaid, R., Brown, R.L. & Cleveland, T.E. (2008). Differentiation of toxigenic fungi using hyperspectral imagery. *Sensing and Instrumentation for Food Quality and Safety*, **2**, 215–224.

ADDENDUM A

Supplementary information pertaining to Chapter 3

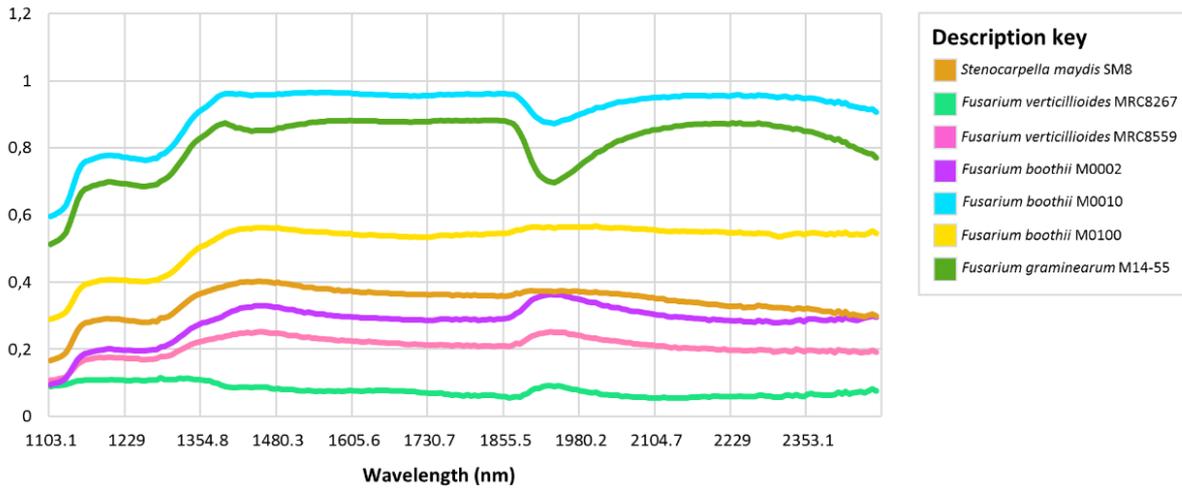


Figure A1 Difference spectra of *F. verticillioides* MRC826 and the remaining isolates.

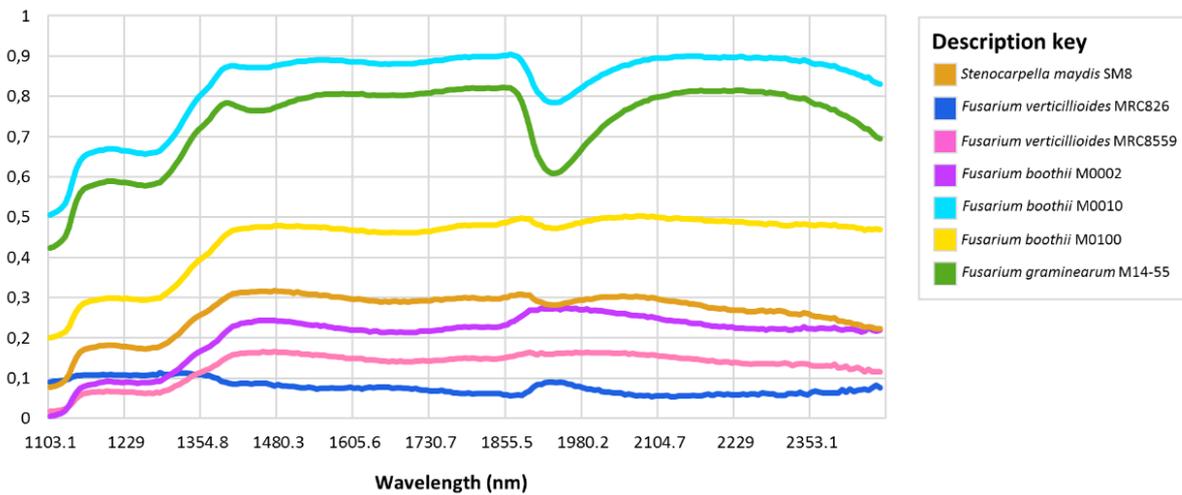


Figure A2 Difference spectra of *F. verticillioides* MRC8267 and remaining isolates.

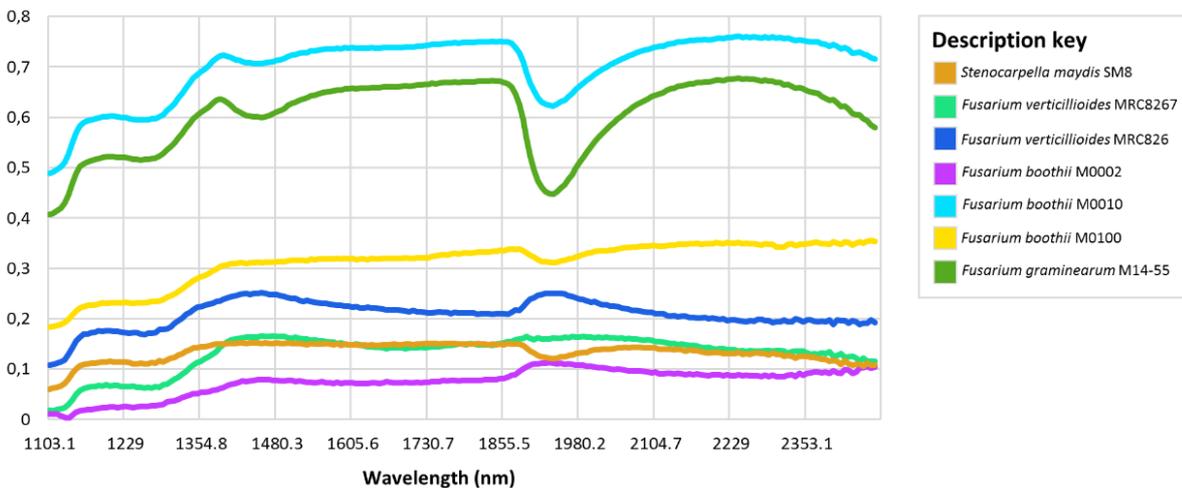


Figure A3 Difference spectra of *F. verticillioides* MRC8559 and the remaining isolates.

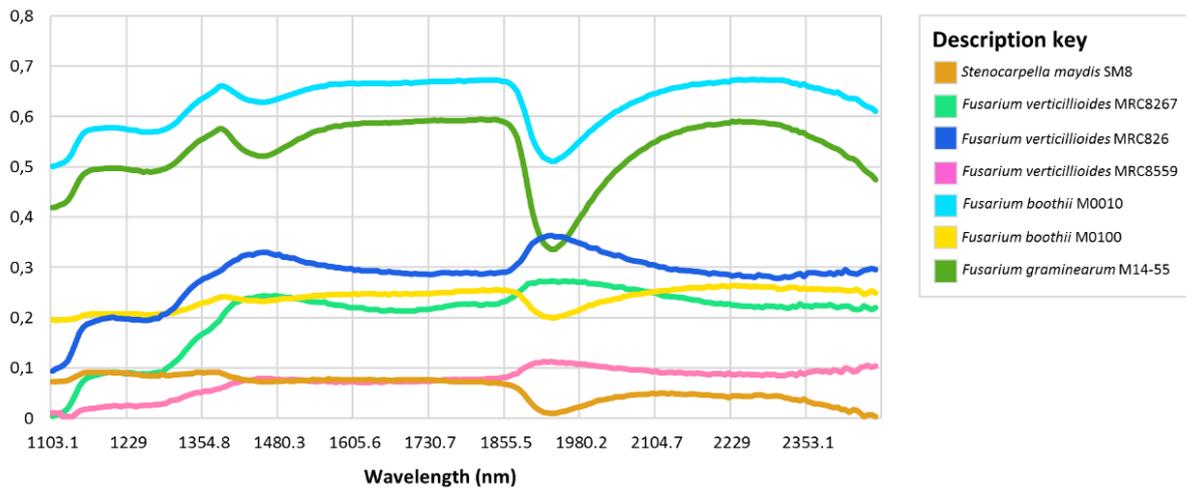


Figure A4 Difference spectra of *F. boothii* M0002 and the remaining isolates.

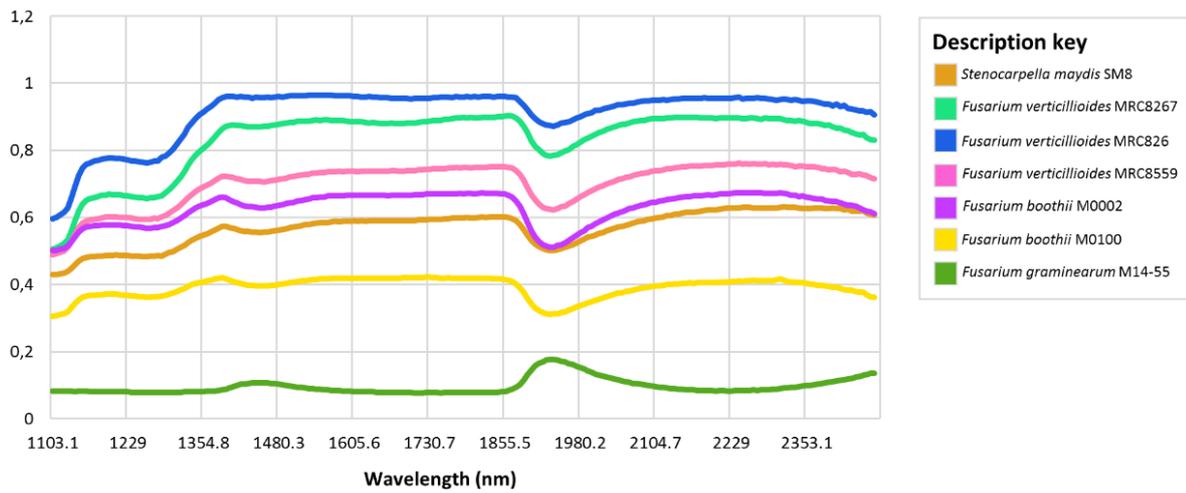


Figure A5 Difference spectra of *F. boothii* M0010 and the remaining isolates.

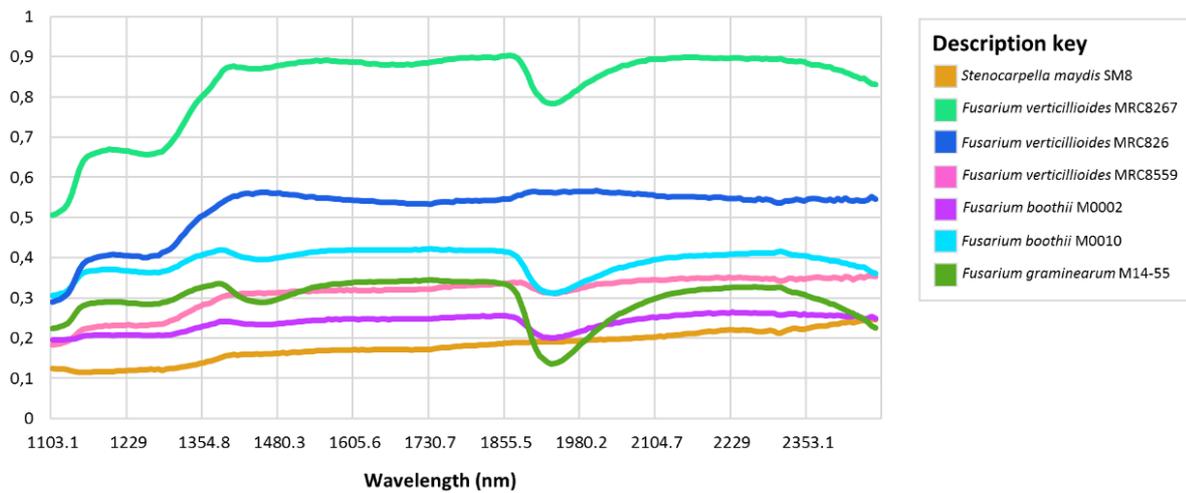


Figure A6 Difference spectra of *F. boothii* M0100 and the remaining isolates.

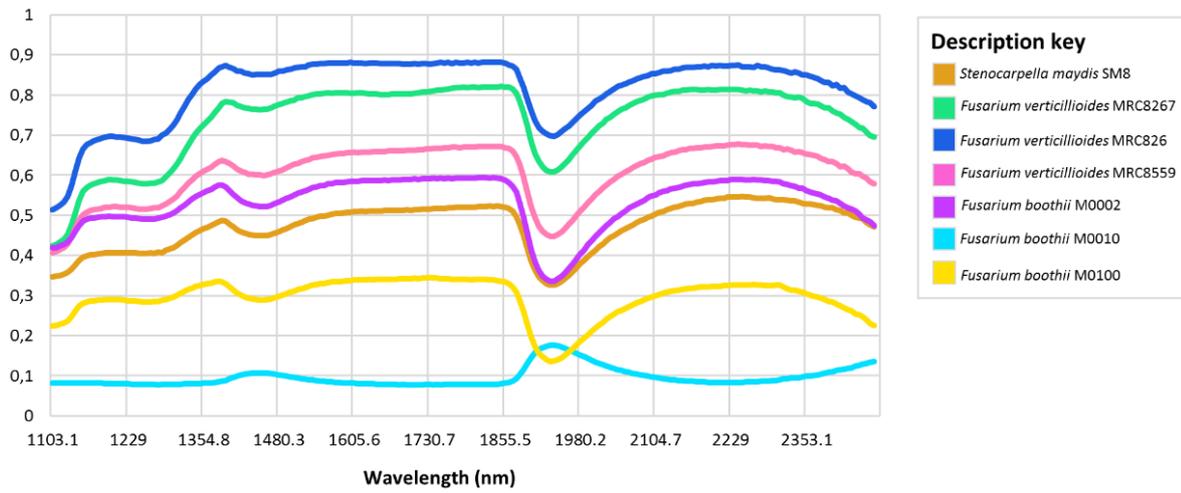


Figure A7 Difference spectra of *F. graminearum* M14-55 and the remaining isolates.

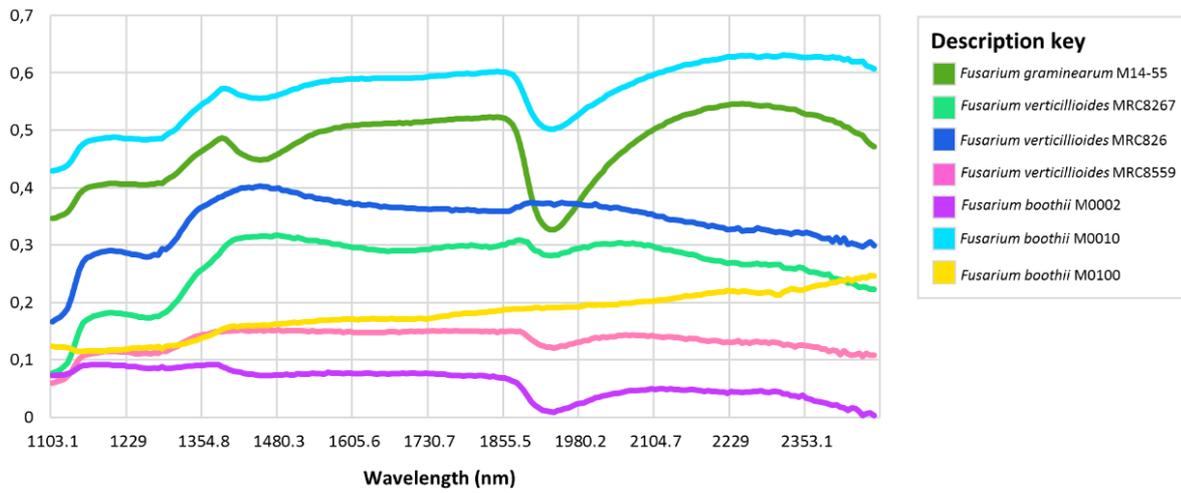


Figure A8 Difference spectra of *S. maydis* SM8 and the remaining isolates.

ADDENDUM B

Supplementary information pertaining to Chapter 4

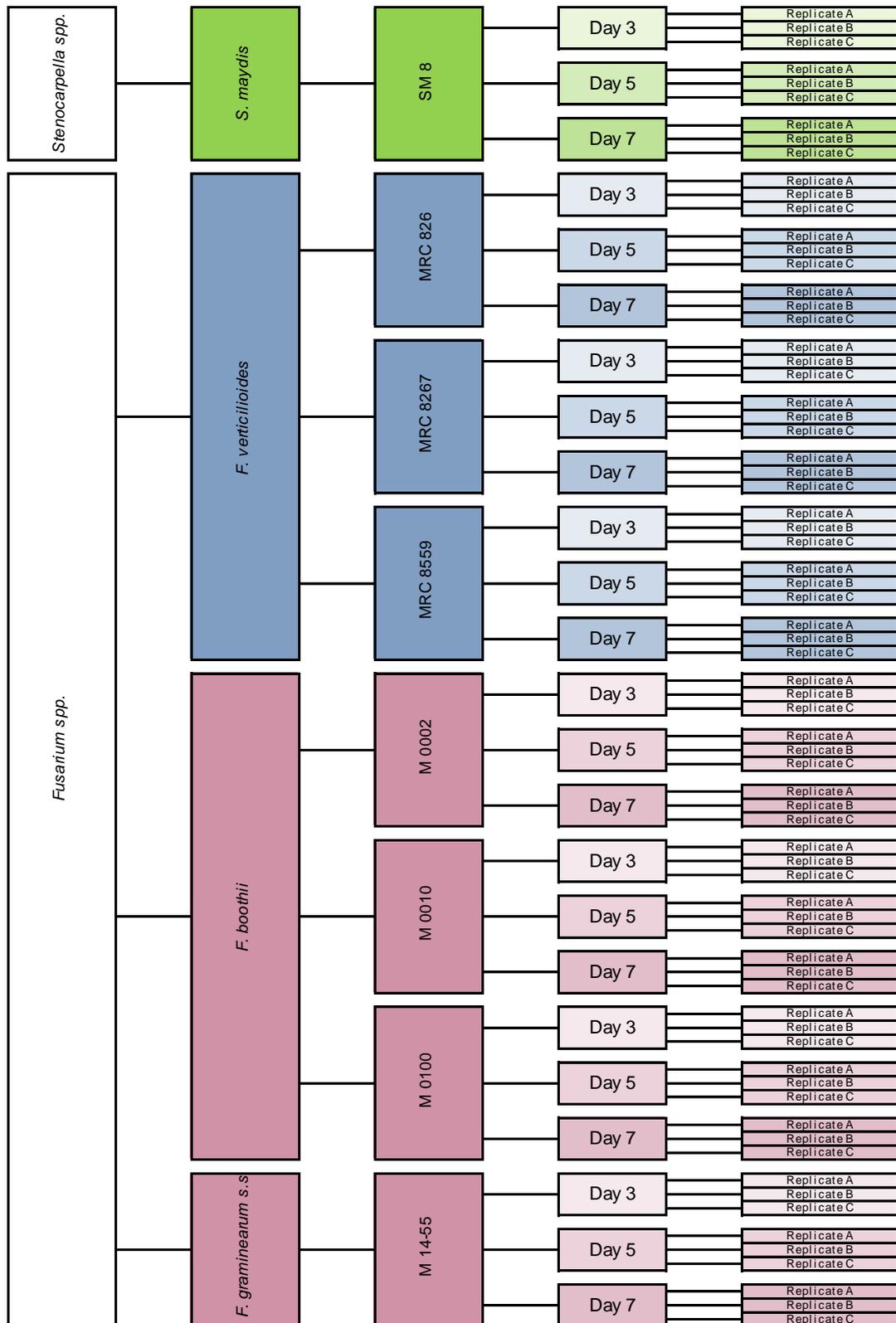


Figure B1 Pathogen diagram illustrating the isolates used in this study, together with the incubation times and number of replicates for each isolate, where maize ear rot types are distinguished by colour, i.e. FER (blue), GER (pink) and DER (green).

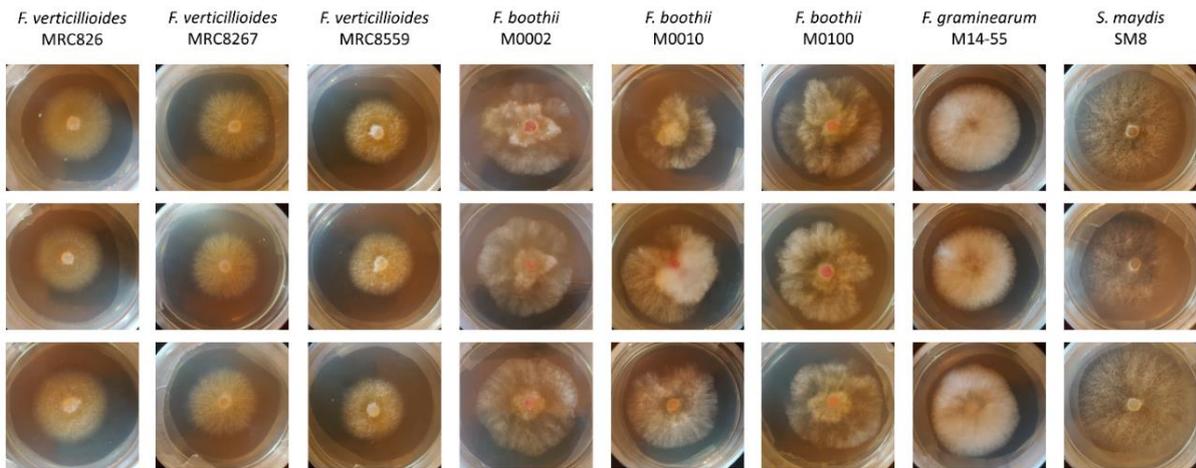


Figure B2 Digital images of the eight *F. verticillioides*, *F. graminearum* s.s., *F. boothii* and *S. maydis* isolates, on day 3 of growth, used in the study.

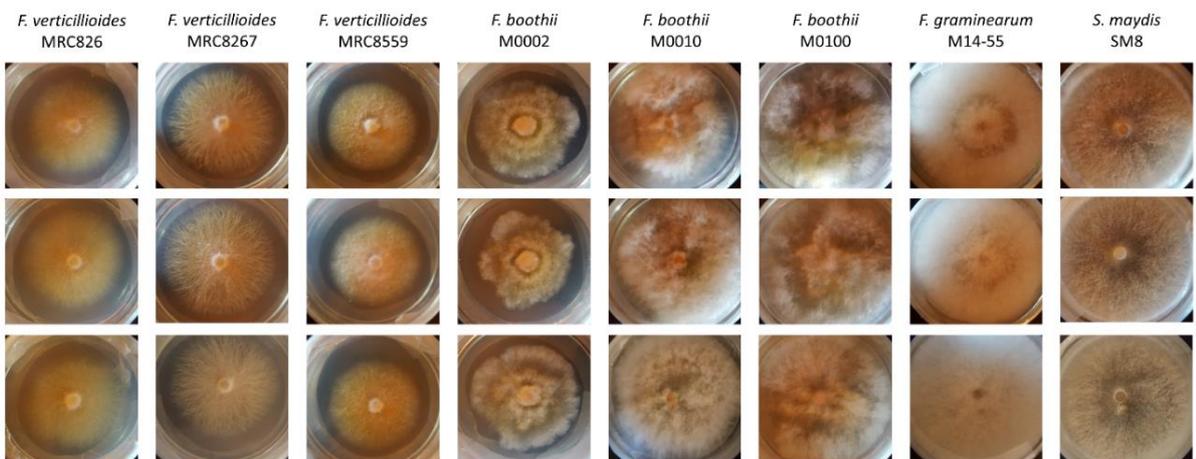


Figure B3 Digital images of the eight *F. verticillioides*, *F. graminearum* s.s., *F. boothii* and *S. maydis* isolates, on day 5 of growth, used in the study.

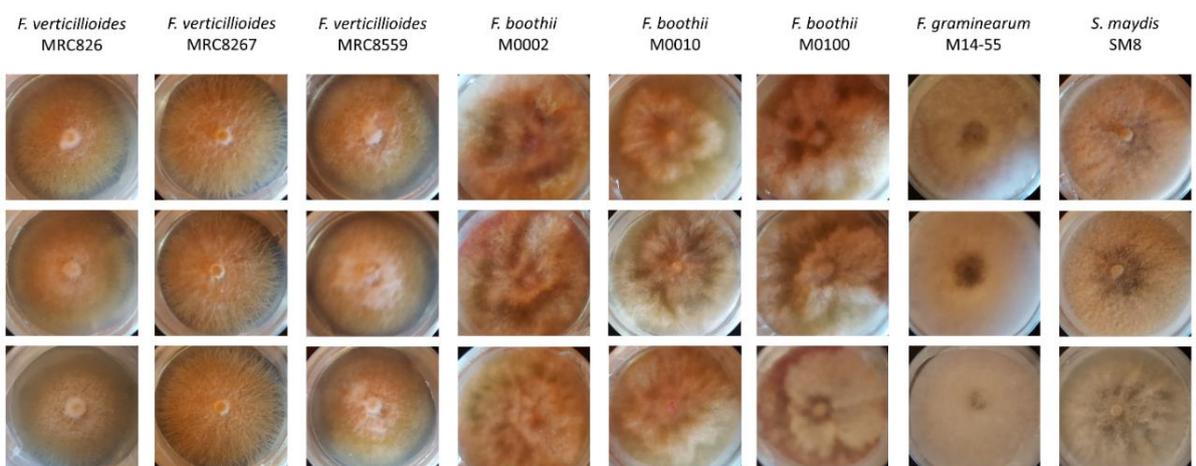


Figure B4 Digital images of the eight *F. verticillioides*, *F. graminearum* s.s., *F. boothii* and *S. maydis* isolates, on day 7 of growth, used in the study.

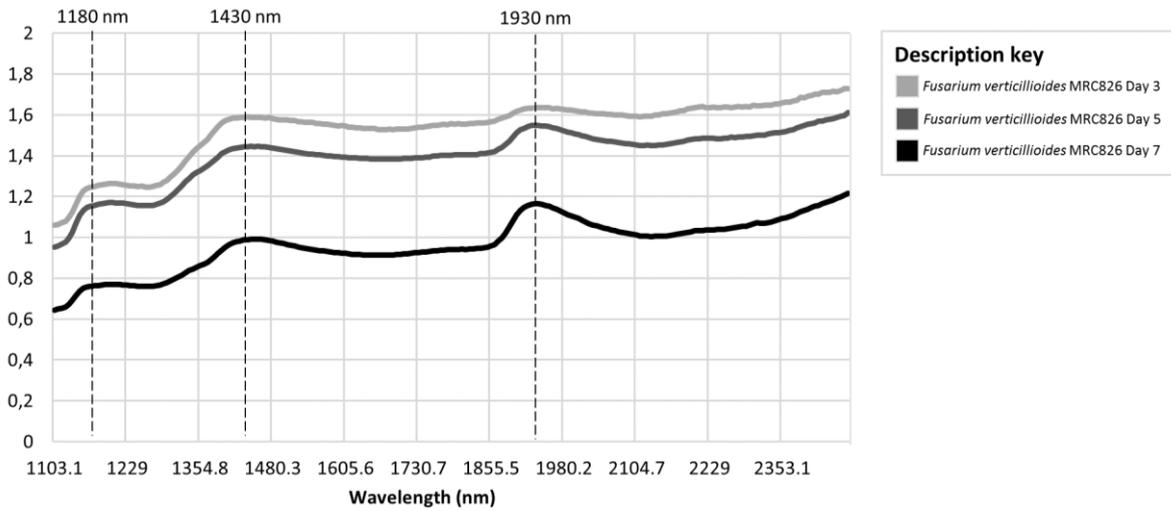


Figure B5 Mean spectra of *F. verticillioides* MRC826 on day 3, 5 and 7 of growth.

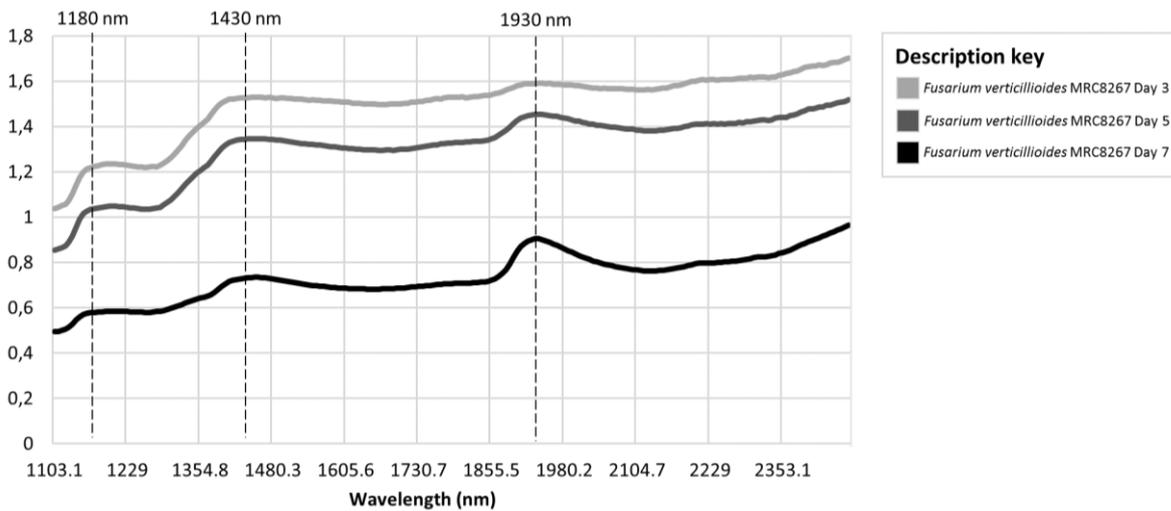


Figure B6 Mean spectra of *F. verticillioides* MRC8267 on day 3, 5 and 7 of growth.

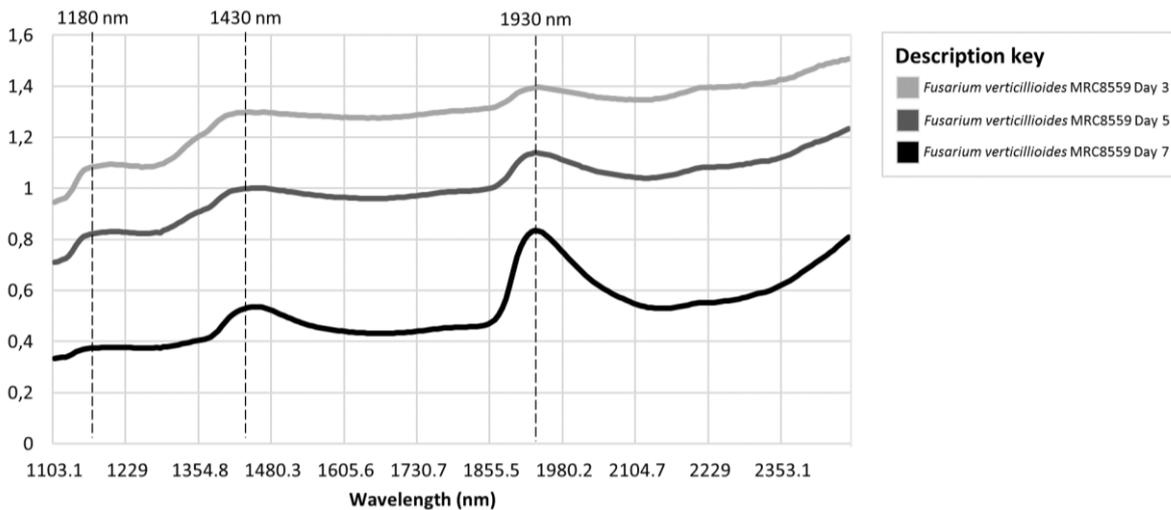


Figure B7 Mean spectra of *F. verticillioides* MRC8559 on day 3, 5 and 7 of growth.

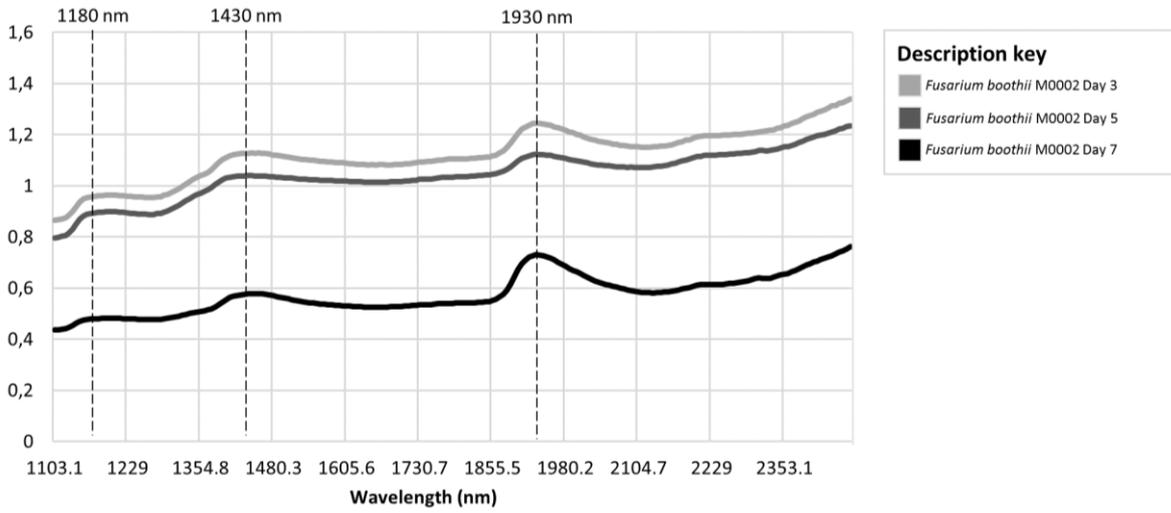


Figure B8 Mean spectra of *F. boothii* M0002 on day 3, 5 and 7 of growth.

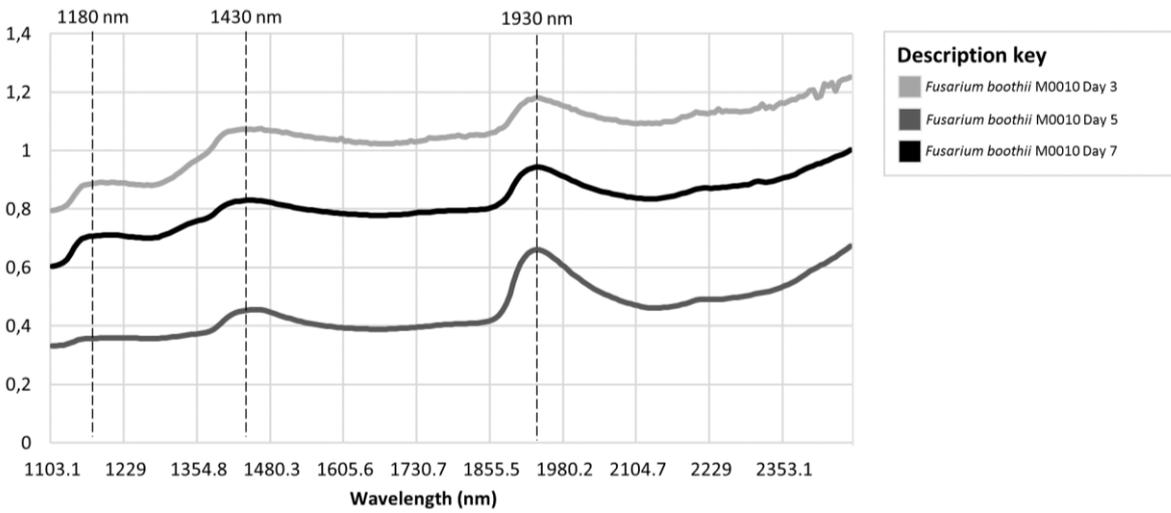


Figure B9 Mean spectra of *F. boothii* M0010 on day 3, 5 and 7 of growth.

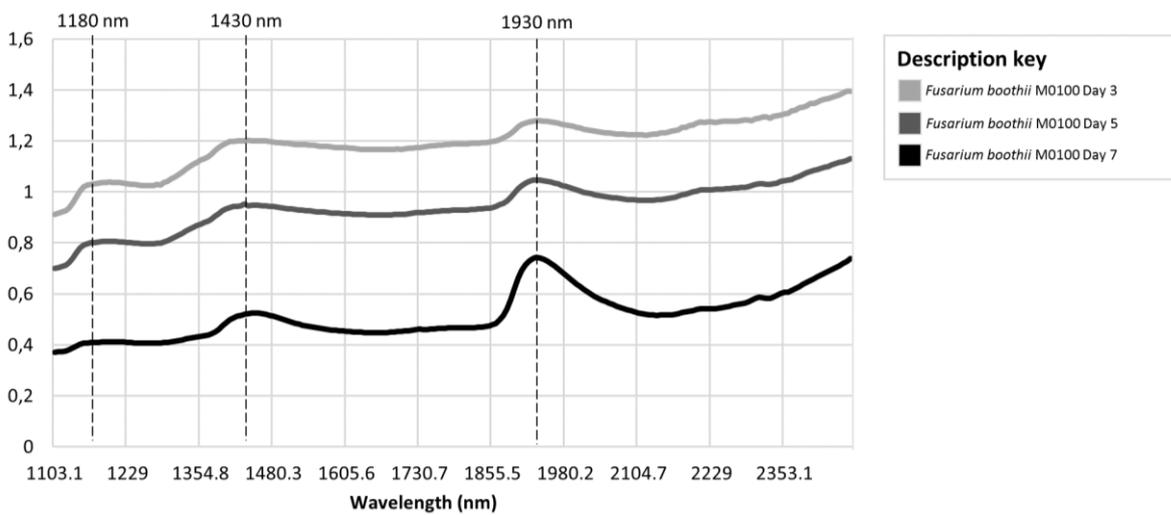


Figure B10 Mean spectra of *F. boothii* M0100 on day 3, 5 and 7 of growth.

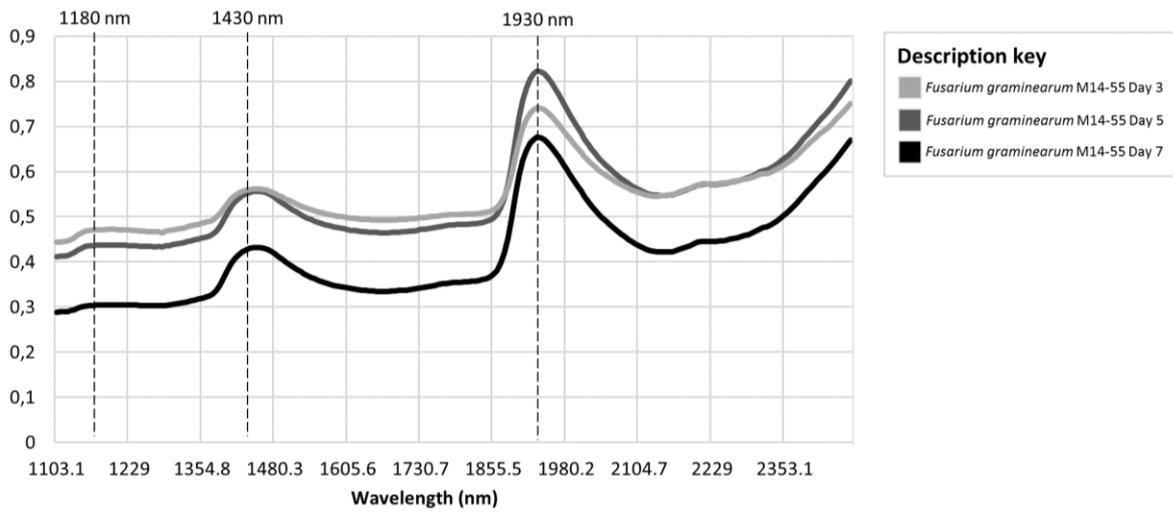


Figure B11 Mean spectra of *F. graminearum* s.s. M14-55 on day 3, 5 and 7 of growth.

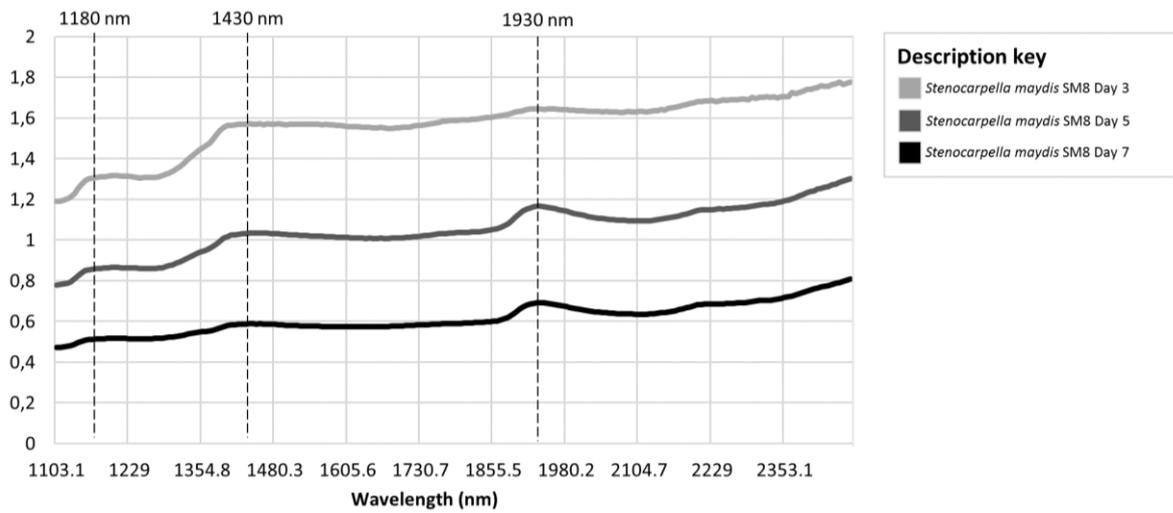


Figure B12 Mean spectra of *S. maydis* SM8 on day 3, 5 and 7 of growth.

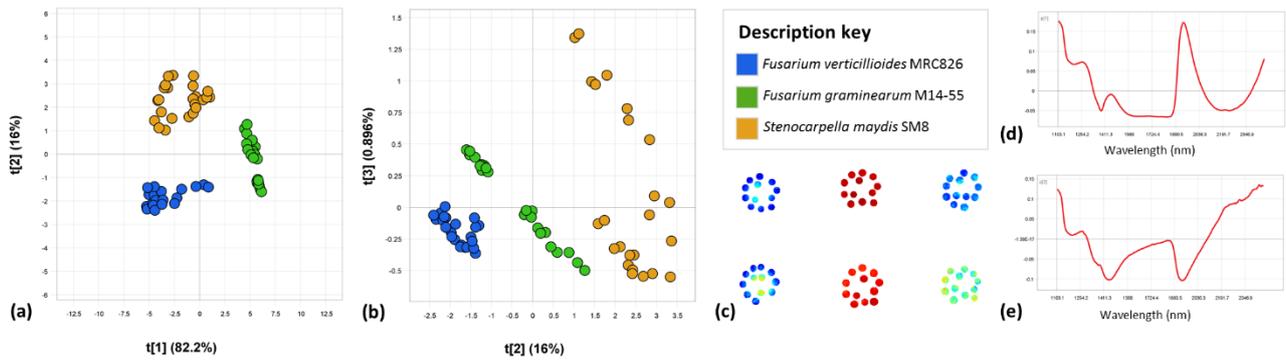


Figure B13 Object wise PCA of mosaic two (Day 5) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (82.2%) vs. PC2 (16%) and (b) PCA score plot of PC2 (16%) vs PC3 (0.896%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. graminearum* s.s. M14-55 (middle) and *S. maydis* SM8 (right). The loading plots for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1930 nm.

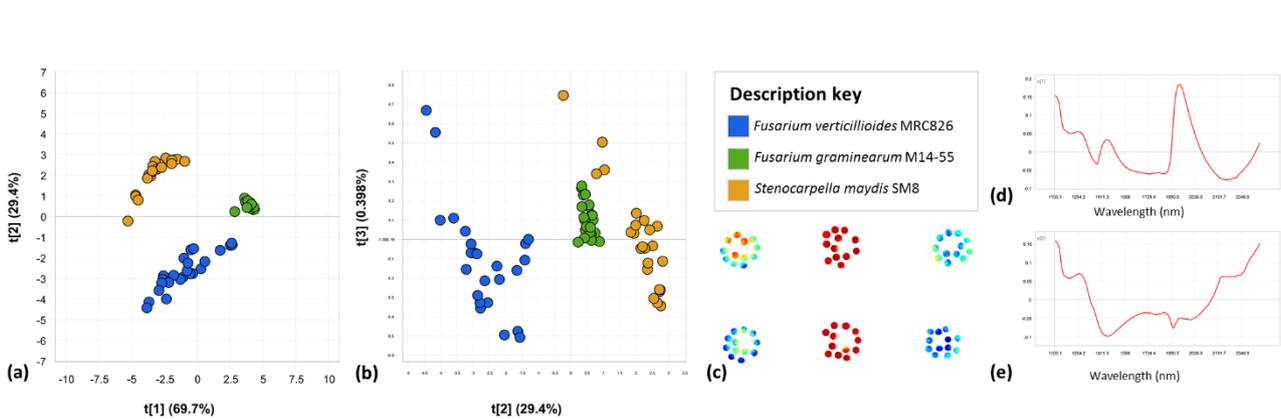


Figure B14 Object wise PCA of mosaic three (Day 7) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (69.7%) vs. PC2 (29.4%) and (b) PCA score plot of PC2 (29.4%) vs PC3 (0.398%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. graminearum* s.s. M14-55 (middle) and *S. maydis* SM8 (right). The loading plots for (d) PC1 with bands at 1260, 1400, 1430 and 1930 nm and (e) PC2 with bands at 1260, 1430 and 1880 nm.

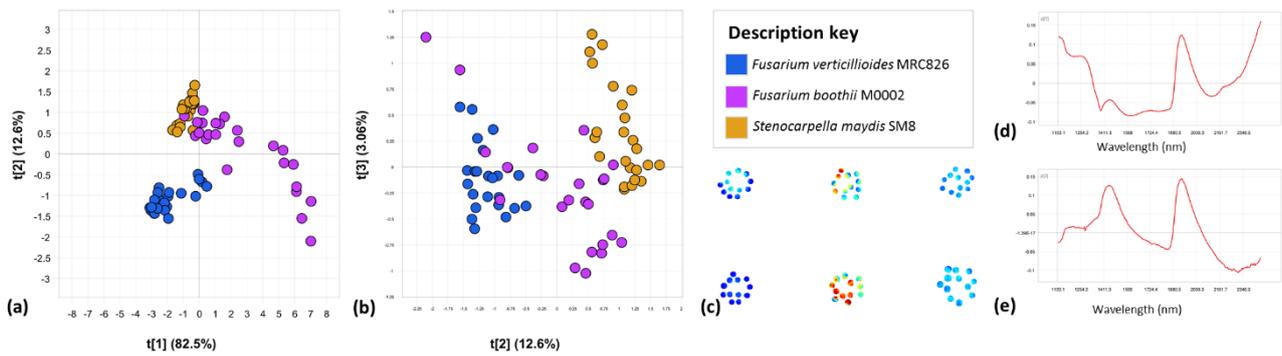


Figure B15 Object wise PCA of mosaic four (Day 3) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (82.5%) vs. PC2 (12.6%) and (b) PCA score plot of PC2 (12.6%) vs. PC3 (3.06%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. boothii* M0002 (middle) and *S. maydis* SM8 (right). The loading plots for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1930 nm.

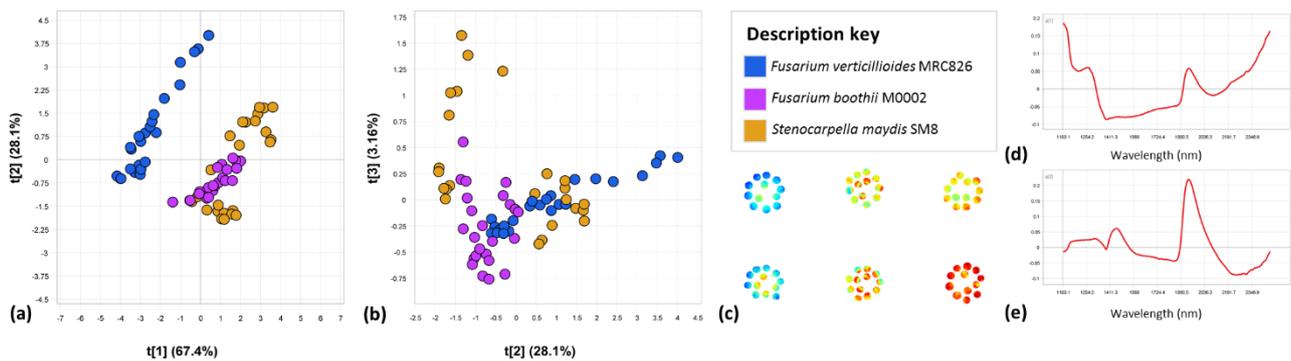


Figure B16 Object wise PCA of mosaic five (Day 5) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (67.4%) vs. PC2 (28.1%) and (b) PCA score plot of PC2 (28.1%) vs. PC3 (3.16%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. boothii* M0002 (middle) and *S. maydis* SM8 (right). The loading plots for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1930 nm.

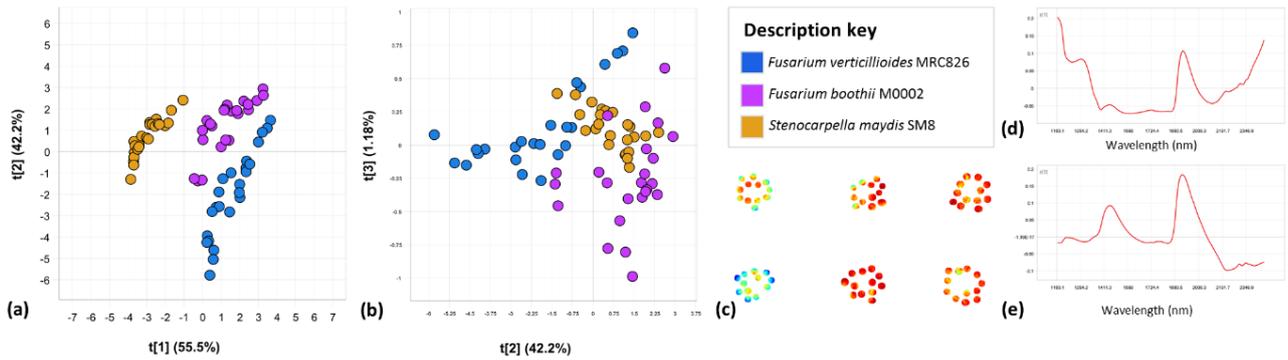


Figure B17 Object wise PCA of mosaic six (Day 7) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (55.5%) vs. PC2 (42.2%) and (b) PCA score plot of PC2 (42.2%) vs. PC3 (1.18%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. boothii* M0002 (middle) and *S. maydis* SM8 (right). The loading plots for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1930 nm.

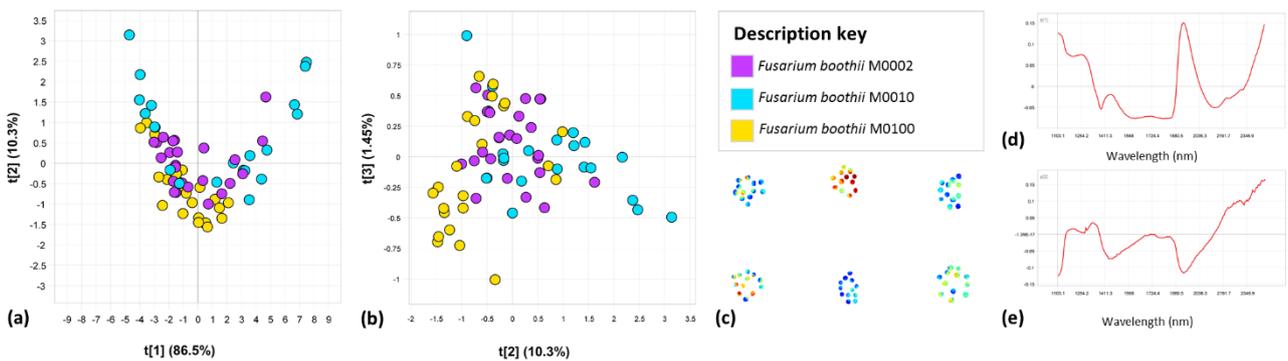


Figure B18 Object wise PCA of mosaic seven (Day 3) illustrates poor separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (86.5%) vs. PC2 (10.3%) and (b) PCA score plot of PC2 (10.3%) vs. PC3 (1.45%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. boothii* M0002 (left), *F. boothii* M0010 (middle) and *F. boothii* M0100 (right) is also given. The loading plots for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1930 nm.

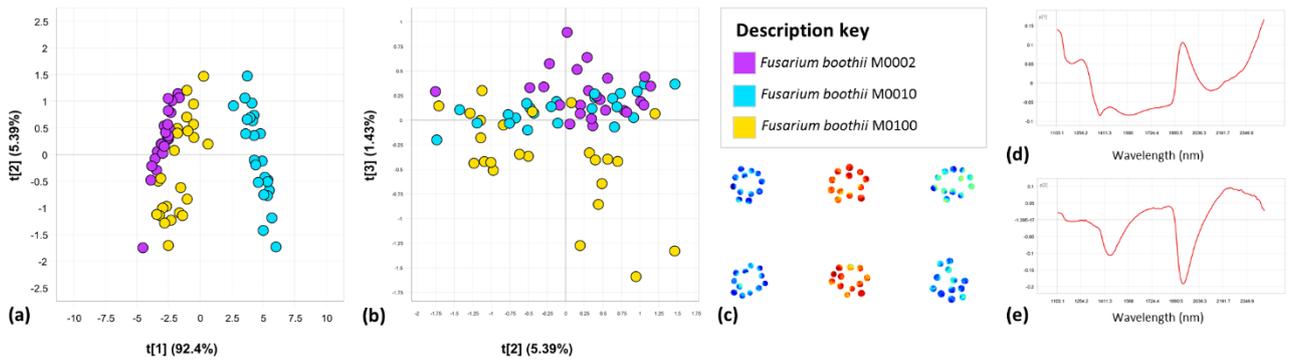


Figure B19 Object wise PCA of mosaic eight (Day 5) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (92.4%) vs. PC2 (5.39%) and (b) PCA score plot of PC2 (5.39%) vs. PC3 (1.43%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. boothii* M0002 (left), *F. boothii* M0010 (middle) and *F. boothii* M0100 (right). The loading plots for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1930 nm.

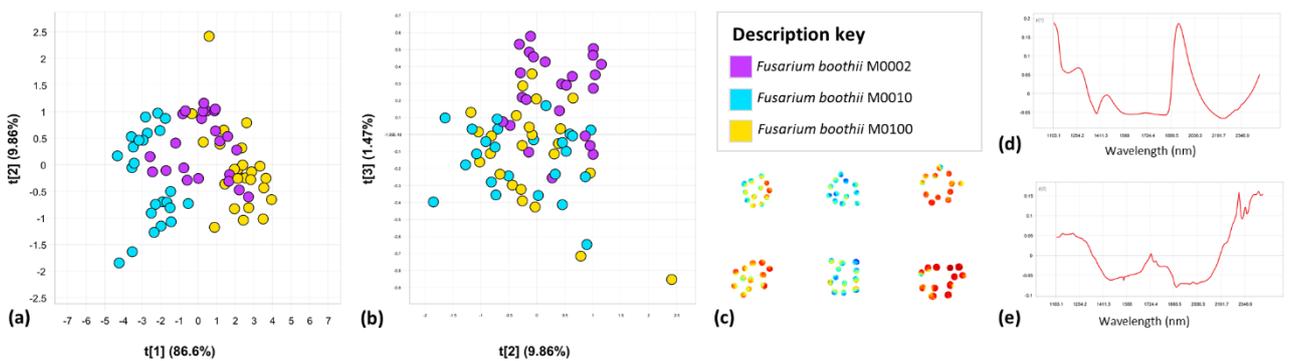


Figure B20 Object wise PCA of mosaic nine (Day 7) illustrates separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (86.6%) vs. PC2 (9.86%) and (b) PCA score plot of PC2 (9.86%) vs. PC3 (1.47%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. boothii* M0002 (left), *F. boothii* M0010 (middle) and *F. boothii* M0100 (right). The loading plots for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1900 nm.

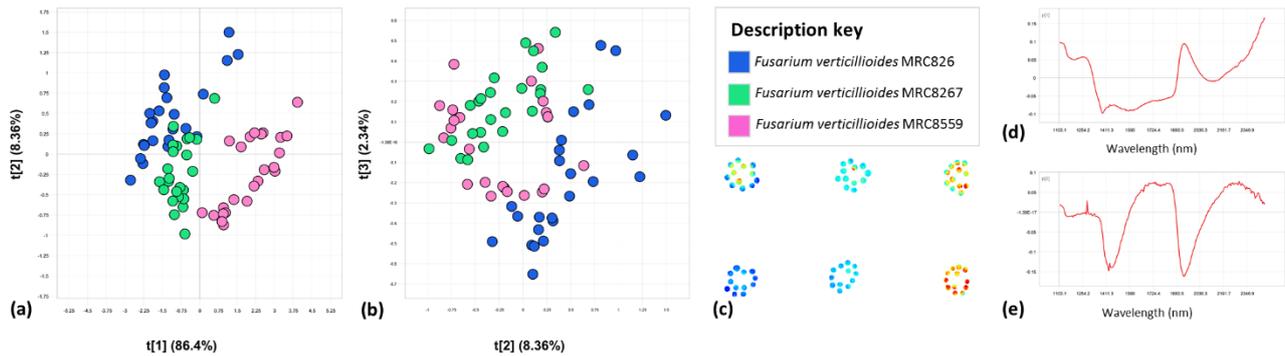


Figure B21 Object wise PCA of mosaic ten (Day 3) illustrates separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (86.4%) vs. PC2 (8.36%) and (b) PCA score plot of PC2 (8.36%) vs. PC3 (2.34%) colour coded per pathogen isolate as seen in the accompanying legend. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. verticillioides* MRC8267 (middle) and *F. verticillioides* MRC8559 (right). The loading plots for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1260, 1430 and 1930 nm.

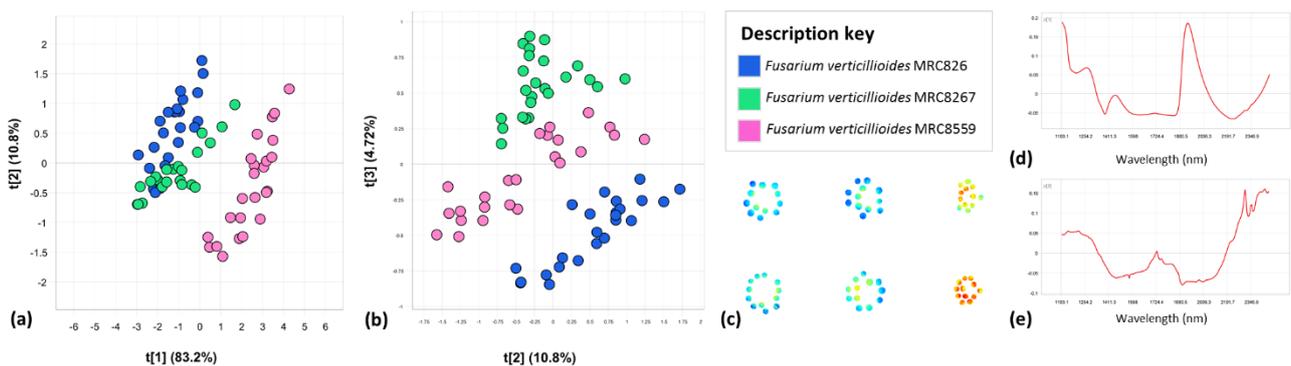


Figure B22 Object wise PCA of mosaic eleven (Day 5) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (83.2%) vs. PC2 (10.8%) and (b) PCA score plot of PC2 (10.8%) vs. PC3 (4.72%) colour coded per pathogen isolate as seen in the accompanying legend is also given. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. verticillioides* MRC8267 (middle) and *F. verticillioides* MRC8559 (right) is also given. The loading plots is given for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1430 and 1930 nm.

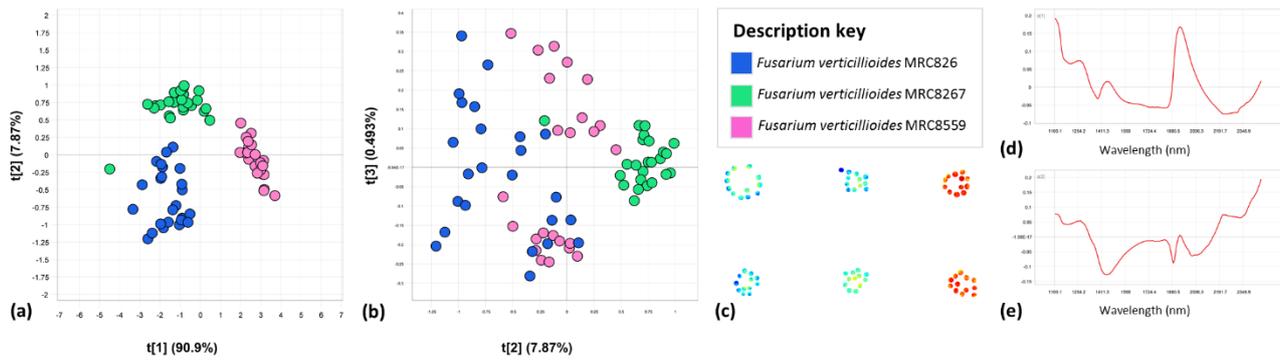


Figure B23 Object wise PCA of mosaic twelve (Day 7) illustrates good separation between isolates. Scores illustrated by (a) PCA score plot of PC1 (90.9%) vs. PC2 (7.87%) and (b) PCA score plot of PC2 (7.87%) vs. PC3 (0.493%) colour coded per pathogen isolate as seen in the accompanying legend is also given. (c) PCA score image (PC1) of *F. verticillioides* MRC826 (left), *F. verticillioides* MRC8267 (middle) and *F. verticillioides* MRC8559 (right) is also given. The loading plots is given for (d) PC1 with bands at 1260, 1400 and 1930 nm and (e) PC2 with bands at 1400, 1900 and 2100 nm.