

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

***ATR – FTIR Chemometrics for Biological
Samples***

*A thesis presented in partial fulfilment of the requirements
for the degree of*

Master of Science

in

Nanoscience

*at Massey University, Manawatū,
New Zealand.*

Josiah David Cleland

2018

Abstract:

This project has used the analytical infrared reflectance technique of *Attenuated Total Reflectance Fourier Transform Infrared* (ATR-FTIR) spectroscopy, for the prediction of chemical components in a range of biological samples. Data collection was carried out on 40 hyperaccumulator samples, 56 chicken feed samples, 54 lamb faecal samples and 188 forage feed samples. Predictions were made using several different regression methods including: Ridge, Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net, Principal Components (PCR) and Partial Least Squares (PLS). The best methods were identified as LASSO, Elastic Net and PLS. Several spectral data pre-treatments were explored, the best of which combined Standard Normal Variant scaling (SNV) with a first-order Savitzky – Golay (SG) spectral derivative and smoothing filter. Several of the resulting models illustrated high quality predictions ($R^2 > 0.8$, Relative Performance Deviation (RPD) ≥ 2). The SNV and SG pre-treatment almost completely reduces the contribution of strong water-based signals to the regression model, allowing the possibility of *in situ* prediction of forage feed composition with minimal sample preparation. ATR-FTIR spectrometers are available in a hand-held form, and the results of this research suggest that *in situ* forage quality analysis could be performed using mid – infrared (MIR) reflectance spectroscopy.

Preface:

A portion of this research thesis has been submitted to the Journal of Animal Feed Science and Technology for publishing. The article is titled “*Mid-Infrared Reflectance Spectroscopy as a tool for forage feed composition prediction*” and was authored by Josiah D. Cleland, Ellie Johnson, Patrick C. H. Morel, Paul R. Kenyon, Mark R. Waterland.

Acknowledgements:

First and foremost, I would like to acknowledge my wonderful wife, Jo and daughter Ayla, I wouldn't be where I am today without their love and support. I would like to also thank my parents, Darryl and Trish, for all their help and support throughout my studies.

A special thanks is in order for my supervisor, Associate Professor Mark Waterland, for his guidance and expert advice. Mark's enthusiasm for the theory involved in this project has led to many productive discussions, and contributed greatly to the successes of this research.

The project included a variety of sample types, an additional thanks is extended to the people who kindly provided them, namely Dr Fifi Zaefarian, Dr Felicity Jackson, Miss Antoinette Danso and Associate Professor Chris Anderson. I am also very thankful to Professor Patrick Morel, Professor Paul Kenyon and again Associate Professor Chris Anderson for the specific knowledge they were able to input into the project regarding the forage feed samples and hyperaccumulator samples, respectively.

I would like to also acknowledge Massey University for selecting me as a recipient of the *Massey University Masterate Scholarship*, the funds helped greatly during my studies.

Abbreviations:

| | |
|----------|---|
| PT | Pre – Treatment |
| SG | Savitsky – Golay filter |
| SNV | Standard normal variate scaling |
| MSC | Multiplicative Scatter Correction |
| DWT | Discrete Wavelet Transform |
| AsLS | Asymmetric Least Squares |
| PLS | Partial Least Squares |
| PCR | Principal Component Regression |
| PCA | Principal Component Analysis |
| RR | Ridge Regression |
| LASSO | Least Absolute Shrinkage and Selection Operator |
| EN | Elastic Net |
| NDF | Neutral Detergent Fibre |
| ADF | Acid Detergent Fibre |
| GE | Gross – Energy |
| ME | Metabolisable – Energy |
| OM | Organic Matter |
| DOMD | In vitro Digestible Organic Matter |
| PC | Principal Component |
| ATR FTIR | Attenuated Total Reflectance Fourier Transform Infrared |
| MIR | Mid – Infrared |
| NIR | Near – Infrared |
| FIR | Far – Infrared |
| MPAES | Microwave Plasma Atomic Emission Spectrometer |
| RMSEP | Root Mean Squared Error of Prediction |
| RMSECV | Root Mean Squared Error of Cross – Validation |
| SEP | Standard Error of Prediction |
| SECV | Standard Error of Cross – Validation |
| RPD | Relative Performance Deviation |

Table of Contents:

| | |
|---|-----------|
| 1 - Introduction | 1 |
| <i>1.1 Chemometrics</i> | <i>1</i> |
| <i>1.2 Infrared Spectroscopy.....</i> | <i>1</i> |
| <i>1.3 Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy</i> | <i>2</i> |
| <i>1.4 Samples</i> | <i>4</i> |
| <i>1.4.1 Forage Feed.....</i> | <i>4</i> |
| <i>1.4.2 Faecal</i> | <i>6</i> |
| <i>1.4.3 Chicken Feed</i> | <i>6</i> |
| <i>1.4.4 Hyperaccumulators.....</i> | <i>11</i> |
| <i>1.5 Mathematical Pre – Treatment</i> | <i>13</i> |
| <i>1.5.1 Scatter Correction Pre – treatments</i> | <i>13</i> |
| <i>1.5.1a Multiplicative Scatter Correction</i> | <i>13</i> |
| <i>1.5.1b Standard Normal Variate Scaling.....</i> | <i>15</i> |
| <i>1.5.2 Savitsky – Golay Filtering</i> | <i>16</i> |
| <i>1.5.3 Asymmetric Least Squares</i> | <i>17</i> |
| <i>1.5.4 Discrete Wavelet Transform</i> | <i>20</i> |
| <i>1.6 Multivariate Regression.....</i> | <i>22</i> |
| <i>1.7 Principal Component Analysis</i> | <i>24</i> |

| | |
|--|-----------|
| 1.8 Regression Methods..... | 25 |
| 1.8.1 Ridge Regression | 26 |
| 1.8.2 LASSO Regression..... | 26 |
| 1.8.3 Elastic Net Regression | 27 |
| 1.8.4 Principal Component Regression | 28 |
| 1.8.5 Partial Least Squares | 29 |
| 1.9 Chemometrics: Infrared Spectroscopy of Forage Feeds..... | 31 |
| 1.10 Project Aims | 33 |
| 2 - Methodology and Technical Background | 35 |
| 2.1 Methodology..... | 35 |
| 2.1.1 Initial Forage Feed Samples | 35 |
| 2.1.2 Expanded Forage Feed Samples | 35 |
| 2.1.3 Chicken Feed Samples..... | 35 |
| 2.1.4 Faecal Samples..... | 36 |
| 2.1.5 Hyperaccumulators | 36 |
| 2.1.6 Wet Chemical Analysis | 36 |
| 2.1.7 ATR – FTIR Spectra Collection..... | 37 |
| 2.1.8 Statistical Analysis..... | 38 |
| 2.2 Technical Background: Regression Methods | 41 |
| 2.2.1 Introduction..... | 42 |
| 2.2.2 Ordinary Least Squares..... | 44 |
| 2.2.3 Ridge Regression..... | 50 |

| | |
|--|------------|
| 2.2.4 Least Absolute Shrinkage and Selection Operator | 53 |
| 2.2.5 Elastic Net | 56 |
| 2.2.6 Geometry Background | 57 |
| 2.2.6.1 Oblique Projections | 57 |
| 2.2.6.2 Geometry of Ellipsoids..... | 59 |
| 2.2.6.3 Planes of Tangency | 62 |
| 2.2.6.4 The Power Method | 65 |
| 2.2.7 Principal Component Regression | 67 |
| 2.2.8 Principal Component Regression Geometry | 70 |
| 2.2.9 Partial Least Squares | 72 |
| 2.2.9.1 Residual Matrices | 76 |
| 2.2.9.2 Properties of the Partial Least Squares Algorithm | 78 |
| 2.2.9.3 Partial Least Squares for Linear Regression | 82 |
| 2.2.9.4 Derivation of PLS Regression Coefficient Vector | 84 |
| 2.2.10 Geometry of the PLS Algorithm | 86 |
| 2.2.10.1 Single Response, Two Predictor Variables..... | 86 |
| 2.2.10.2 The Object Space | 87 |
| 2.2.10.3 The Variable space | 89 |
| 3.0 - Spectral Pre-treatment and Principal Components Analysis Results..... | 91 |
| 4.0 - Results and Discussion I: Feature Extraction Regression Methods | 103 |
| 4.1 Ridge Regression | 104 |
| 4.2 LASSO Regression | 122 |
| 4.3 Elastic Net Regression | 138 |

| | |
|--|-----|
| 5.0 - Results and Discussion II: Feature construction Regression Methods | 157 |
| 5.1 <i>Principal Component Regression</i> | 158 |
| 5.2 <i>Partial Least Squares Regression</i> | 173 |
| 6.0 - Conclusions and Future Work | 195 |
| 6.1 <i>Conclusions</i> | 195 |
| 6.2 <i>Future Work</i> | 196 |
| 7.0 - Appendices | 197 |
| 7.1 <i>Appendix 1: Asymmetric Least Squares Process</i> | 197 |
| 7.2 <i>Appendix 2 Discrete Wavelet Transform – Brief Example</i> | 200 |
| 7.3 <i>Appendix 3: Properties of OLS residuals and Markov – Gauss assumptions.</i> | 203 |
| 7.4 <i>Appendix 4: Derivation of the Eigenvalue Equations for PLS</i> | 206 |
| 7.5 <i>Appendix 5: Geometric Properties of the PLS Vectors</i> | 208 |
| 7.6 <i>Appendix 6: Residual Matrices Derivations</i> | 213 |
| 7.7 <i>Appendix 7: Demonstration of repeatability of ATR-FTIR</i> | 214 |
| 7.8 <i>Appendix 8: Tables of Wet Chemistry Results</i> | 215 |
| 8.0 - References | 227 |