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:

Chapter 1.0: Introduction

1.1 Chemometrics

Chemometrics is a dynamic and inherently multidisciplinary subject. [1-3] Chemometrics is built around the analysis of chemical systems, using tools originating predominantly from the field of applied mathematics, these tools are used to extract interesting correlations from data acquired from the chemical system being studied. The extracted correlations, allow for the prediction of properties of new samples from the same system. Spectroscopic techniques represent a very rapid means of accumulating large data sets, and for this reason, they are the focus of many research projects. [4-10]

1.2 Infrared Spectroscopy

Infrared spectroscopy is one of the more common spectroscopic branches within chemometrics, [11-13] as with any spectroscopic technique, the information it provides is related to the transitions between different energy states. For infrared spectroscopy the transitions are between vibrational and rotational energy levels. The spectrum of infrared radiation ranges in totality from 14000 – 10 cm^{-1} (0.8 – 1000 μm).[14] Infrared radiation is further divided into three regions, classified according to their proximity to the visible region. These regions are the near-infrared (NIR), mid-infrared (MIR) and far-infrared (FIR), they encompass the spectral regions of 14000 – 4000 cm^{-1}, 4000 – 400 cm^{-1} and 400 – 10 cm^{-1}, respectively. In order for a molecule to absorb infrared radiation, the incident radiation must induce a change in the dipole moment of the molecule. The dipole

moment, $\boldsymbol{\mu}$ is related simply to the magnitude of the separated charges, q and the distance between them, $\boldsymbol{r}$ as follows. [15]

$$\boldsymbol{\mu} = q\boldsymbol{r}$$

Physically the dipole moment changes manifest as geometry deformations of some form. The infrared radiation induces these changes based on the electric field component it possesses. The oscillation of the electric field component induces molecular oscillations only when its frequency aligns with the natural oscillation frequency of a given molecular vibration.

1.3 Attenuated Total Reflectance Fourier transform infrared spectroscopy

In general, Fourier Transform Infrared (FTIR) spectroscopy is a very efficient and rapid means of obtaining an infrared spectrum. However, more traditional transmittance-based FTIR techniques were limited in terms of applications by sample and substrate requirements, namely, the substrate must transmit infrared radiation. For quantitative studies to give transmittances within a useful working range, sample thickness is on the order of a few microns. With reflectance based techniques, the strict requirement on sample thickness is avoided and infrared analysis is accessible to a wider range of sample types.[16-19] Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR) spectroscopy is a

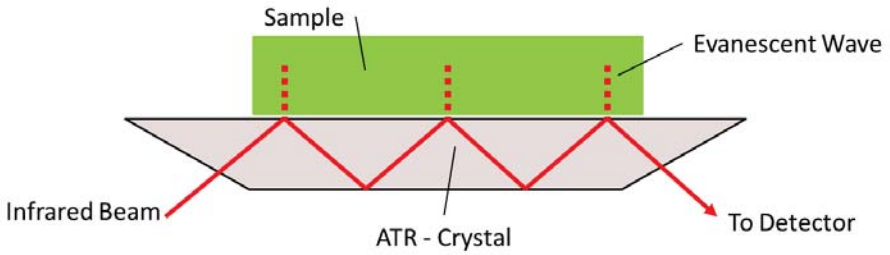

Figure 1.1: Graphical representation of ATR-FTIR sample analysis. Adapted from [16].

reflectance based MIR technique. ATR-FTIR requires the incident infrared radiation to be totally internally reflected within a crystal (which also acts as the sampling substrate), this requirement is due to the fact that ATR-FTIR probes the sample using an evanescent wave established at the sample – crystal interface(see Figure 1.1).[20] The evanescent wave occurs as a partial transmittance of the electromagnetic radiation, a consequence of solving Maxwell's equations including the relevant boundary conditions. The evanescent wave propagates along the interface, but decays exponentially in the direction normal to the interface (non-crystal side). The choice of crystal affects the penetration (or decay length) of the evanescent wave. Typical ATR – FTIR crystal types include diamond, germanium or zinc selenide. Diamond provides very long crystal life time, however, it is the most expensive option. As well as being expensive, diamond also introduces a *dead* region at around 2000 cm^{-1}.

In recent times ATR-FTIR has been preferred to transmittance-based FTIR across a wide range of applications. [21-27] This preference follows from the many benefits [28-33] ATR-FTIR provides such as non-destructive analysis, ease of application, rapid analysis and relatively low associated cost. Whilst ATR-FTIR possesses many benefits, there are still some inherent difficulties associated with the interpretation of the results. ATR-FTIR examines the MIR region which contains information regarding the fundamental vibration energy level transitions. [34, 35] The MIR region is relatively interpretable in the analysis of *pure simple* compounds, [36] however, rapidly becomes uninterpretable when more complex[37] samples are analysed. The complexity associated with these spectra is one of the main motivations for the employment of statistical methods.[38]

1.4 Samples

1.4.1 Forage Feed

Forage feeds is a term describing a range of plant species used as feed most commonly in production animals. Forage feed quality relates directly to the animal condition and therefore production. Forage quality may be evaluated through the identification of various chemical components such as neutral detergent fibre (NDF), protein and gross energy. As further motivation for studying these chemical components of forage feeds consider NDF, where high NDF values represent a greater contribution to the filling of the animal's gut. Hence, reduced dry matter intake which is undesirable for production animals, as the animal is consuming predominantly organic matter it cannot digest. However, it is a fine balancing act as low NDF values can result in a greater risk of sub-acute ruminal acidosis which will also cause reduced dry matter intake. [163] It is suggested the ideal range of forage NDF being 30 – 32 % [163] of the total ration although this varies with the introduction of non-forage NDF. [164]

The forage feed samples in this research may be divided further into an initial sample set and an expanded set. The motivation for the division is due to the fact that the expanded set includes additional plant samples, yet the variables present in these samples are not well accounted for. The initial sample set has far less ambiguity in terms of the species of plants present and the location and conditions of their growth. The initial sample set includes species that may be classified as either pasture or herb mixes. The pasture species include ryegrass and white clover. The herb mix species included chicory, plantain, white clover and red clover. The expanded sample set contained these species as well as

lucerne and unlabelled species. The chemical components identified for further analysis in the initial sample set were: metabolisable energy, neutral detergent fibre (NDF), acid detergent fibre (ADF), hemicellulose, dry matter (DM), digestible organic matter content (DOMD) and protein. For the expanded sample set fewer chemical component results were available only: NDF, hemicellulose, ADF, metabolisable energy and DM. These chemical components are currently determined either by wet chemical analysis or, on occasion, NIR analysis. Wet chemical methods are currently the standard in terms of accurate analysis but possess drawbacks in terms of time and cost. NIR analysis offers benefits over wet chemical methods, in terms of rapid analysis and time constraints, however, ATR-FTIR is suggested to offer further distinct benefits over NIR in terms of in field use, repeatability, and instrument lifetime.

1.4.2 Faecal

For the faecal sample set the known chemical components were: hemicellulose, dry matter, ash, nitrogen, gross energy, organic matter, NDF, ADF and lignin. The motivation for the analysis of these chemical components via ATR-FTIR is the development of a model to predict animal growth. Such a model would take into consideration variation chemical components before and after digestion, differences between these values provides a description of nutrient absorption by the animal. Successful prediction of chemical components from faecal samples would allow for an infield determination of the animal's current growth rate.

1.4.3 Chicken Feed

In the chicken feed sample set the known chemical components were: crude fibre, starch, fat, nitrogen, ash, dry matter, calcium, phosphorus and gross energy. The modelling of chicken feed alongside future chicken faecal models, would provide a means of assessing nutrient absorption in chickens. The chicken feed samples are made up of varying amounts of ground plant matter and meal. The feed portions were: maize, wheat, sorgum, soybean meal, canola meal, meat and bone meal, wheat bran, wheat dried distillers grain with solubles (DDGS), full fat soybean and palm kernel meal.

Table 1.1: Table of chemical components.

Chemical component	Description
Neutral Detergent Fibre	Neutral Detergent Fibre (NDF) represents the total fibre in the sample. NDF contains the plant cell wall components: cellulose, lignin hemicellulose, crude protein and ash.
Hemicellulose	Hemicellulose is one of the plant cell wall polysaccharides. Its chemical structure is that of a branched polymer chain of pentose and hexose sugars.
Acid Detergent Fibre	Acid Detergent Fibre (ADF) contains the least digestible portion of NDF. ADF contains the plant cell wall components: cellulose, lignin, crude protein and ash.
Lignin	Lignin is a bacterial barrier encasing the cellulose – hemicellulose structure of plant tissues. Lignin is essentially indigestible by all livestock. Lignin's chemical structure is that of a complex organic polymer. Lignins are cross-linked by phenolic polymers.[52]
Gross Energy	Gross energy is the total energy of the sample released as heat during combustion.

Table 1.1: Table of chemical components continued.

Chemical component	Description
Dry Matter	Dry matter is simply the amount of feed weight remaining post water removal. Dry matter is usually presented as a percentage of the initial sample weight.
Ash	Ash contains the total inorganic matter, essentially accounting for the mineral content of the feed.
Organic Matter	The difference between dry matter and ash content.
Nitrogen	The amount of nitrogen present in the sample. Measures nitrogen due to protein and non-protein sources. Evaluated typically through the Kjeldahl method.
In vitro Digestible Organic Matter in Dry matter(DOMD)	Determined using fluid from the rumen of research animals. DOMD describes the proportion of the dry matter the animal can digest.
Protein	Calculated from nitrogen content. With forages typically nitrogen content is representative of 16% of the protein content.
Crude Fibre	Accounts for the majority of fibrous components in the sample.
Fat	Triglycerides and fatty acids in the sample. Extracted from the dry matter using ether.

Table 1.1: Table of chemical components continued.

Chemical Component	Description
Calcium	Calcium content is captured in the ash content. Calcium may be explicitly determined in elemental analysis.
Phosphorus	Phosphorus content is captured in the ash content. Phosphorus may be explicitly determined in elemental analysis.
Metabolisable Energy	Metabolisable energy is the energy utilised by the animals. It may be calculated using the digestibility of the relevant sample.
Starch	A polymeric carbohydrate, provides glucose stores for the plant

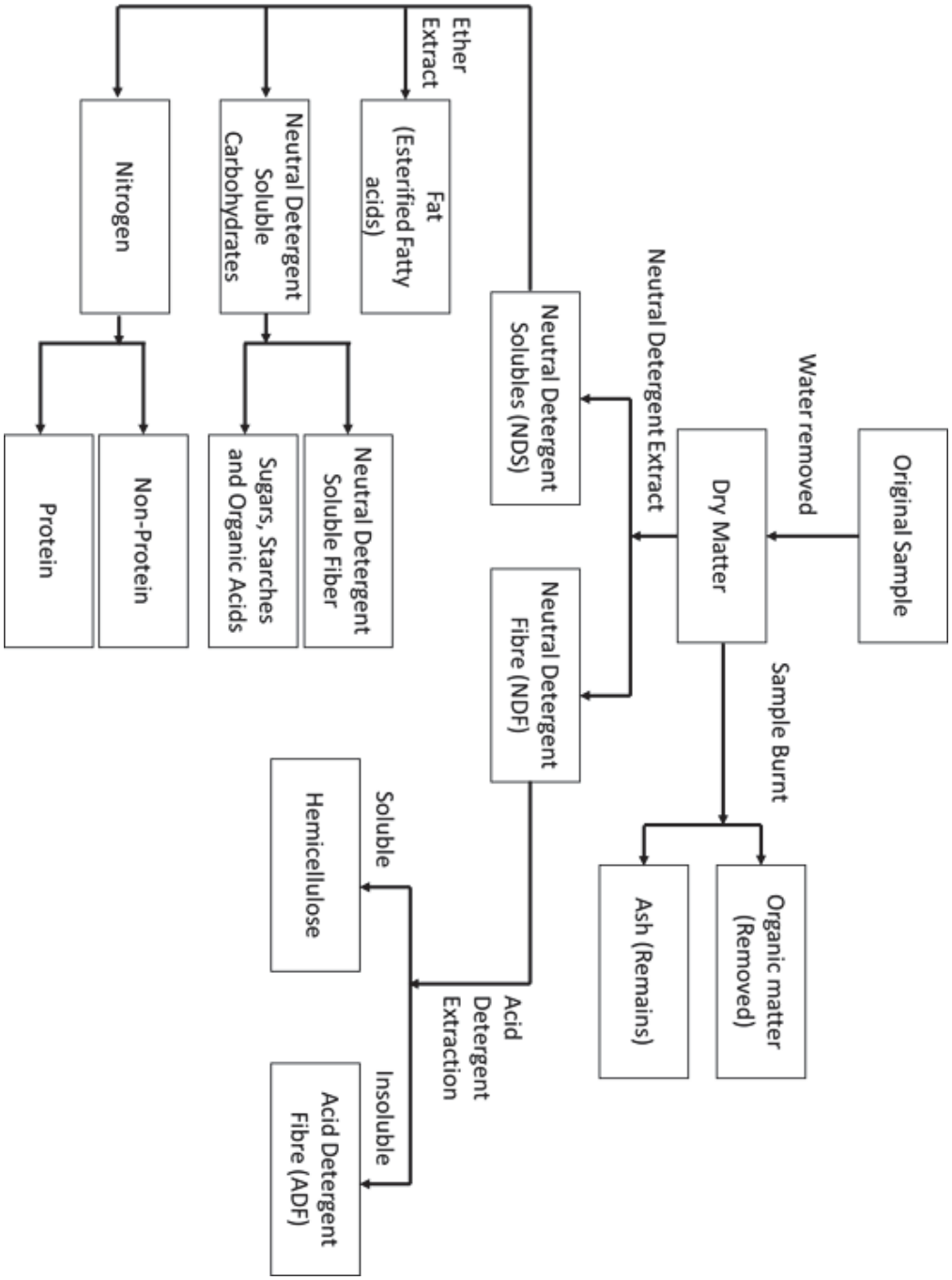

Figure 1.2: Relationship between chemical components.

1.4.4 Hyperaccumulators

Hyperaccumulators represent a particular evolutionary response adopted by a plant species due to a high concentration of heavy metals in the soil [39,40] or aquatic [41] environment. As the name suggests hyperaccumulators acquire internal concentrations of heavy metals, which exceed those in the surrounding soil environment. Depending on the particular heavy metal, the threshold concentration required for hyperaccumulator status varies and Table 1.2 illustrates this.

Table 1.2: Heavy metals and their concentrations within the plant tissue that warrant hyperaccumulator status

Metal	Au	Ag	Cd	Se	Ta	Cu	Co	Cr	Ni
mg/kg	> 1	> 1	> 100	> 100	> 100	> 1000	> 1000	> 1000	> 1000
Species Count	-	-	1	20	-	34	30	-	320

Metal	Pb	U	As	Mn	Zn
mg/kg	> 100	> 1000	> 1000	> 10000	> 10000
Species Count	14	-	4	10	11

Different applications of this interesting classification of plants can be seen throughout the literature, ranging from the removal of toxic heavy metals known as *phytoremediation,* [40] to the acquisition of precious heavy metals known as *phytomining.* [42] This research project will focus on nickel analysis in plant species such as *Alyssum* Sp and *Berkheya Coddii*. *Berkheya Coddii* is of particular interest because of having a relatively large biomass in comparison to other hyperaccumulators, whilst possessing an accumulated metal content under the right conditions of roughly one percent.[43] *Berkheya Coddii* uptakes Ni^{2+} through the same pathways as Ca^{2+} and Mg^{2+} as the presence of the latter two

ions hinder the uptake of the Ni^{2+} ion.[44] In the case of *Alyssum*, nickel has been found chelated by histidine in the xylem for transportation.[45] The nickel eventually becomes sequestrated in the leaf vacuole[46] in a complex form chelated by citrate and malate. Nickel hyperaccumulation is common with roughly 320 known plant species. [47] Table 1.2 illustrates the variation in the number of known species at this stage. [47]

Abou Auda *et al.*[48] in the early 2000s identified the mechanisms by which *Alyssum Murale* tolerate nickel and magnesium accumulation. Shortly afterwards Fujimori *et al.* [49] demonstrated an interesting 24-fold increase in relative expression of the AtSDC gene of *Arabidopsis thaliana,* when Ni was introduced. The AtSDC gene is responsible for the generation of serine decarboxylase (SDC), which suggests that SDC is involved in the cellular response to the presence of metal ions in some fashion. Research has also emerged illustrating the increased expression of genes coding for the production of metallothioneins in *Azolla Filiculoides* [50] and *Hebeloma Cylindrosporum*.[51] The gene up regulation follows from the addition of heavy metals, and it appears the metallothioneins capture the metal ions via complexation.

A range of interesting organic compounds have been shown to be able to be processed (e.g. by pyrolysis) from plant matter. Unfortunately, the products of interest are produced with little selectivity. However, there is reason to believe nickel presence in plant matter alters the selectivity ensuring the desired synthetic products. The ability for ATR-FTIR to predict nickel content may provide a means of infield plant analysis for future agricultural industry.

1.5 Mathematical Pre – treatment

Prior to statistical analysis of spectroscopic data, it is necessary to perform mathematical pre-treatments (PTs) to remove variations of intensity due to non-chemical variations. Non – chemical signal variations are introduced by a range of causes, such as partial size, chemical interference and methods of acquisition, [32,53] these issues manifest as noise or baseline deviations. [30] However, there are a range of noise suppression and baseline removal techniques that manage to, for the most part, remove these issues. [54,55]

1.5.1 Scatter Correction Pre – treatments

Scatter correction pre-treatments account for signal artefacts generated by flaws or limitations inherent to the instrument, such as low signal intensities or scattering. These pre-treatments typically include a combination of centring and scaling, the two most prevalent methods are known as Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV). [56]

1.5.1a Multiplicative Scatter Correction

Multiplicative scatter correction was introduced first by Martens *et al.* in 1983. [57] The process begins by regressing a reference spectrum onto the spectrum being corrected. Typically, this reference spectrum is chosen to be the average spectrum. If $\boldsymbol{x}_{i,original}$ represents a column vector containing p elements each representing the reflectance at a particular wavenumber. The linear relationship between this spectrum and a second reference spectrum may be outlined as follows.

$$\boldsymbol{x_{i,original}} = a_i - b_i \boldsymbol{x_{i,reference}} \quad (1)$$

Where a_i and b_i are coefficients accounting for offsets and slope deviations.

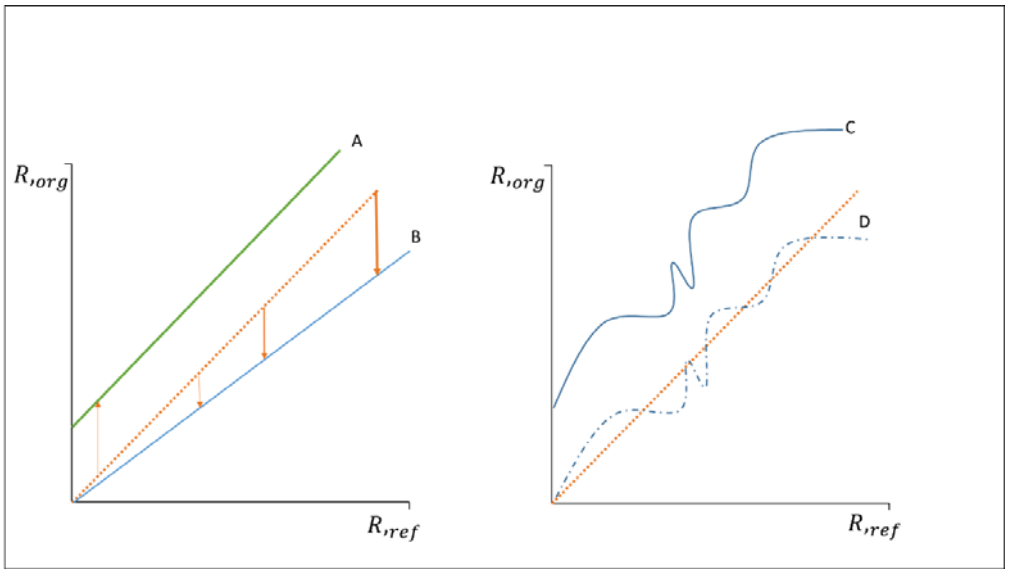

Figure 1.3: Example of Multiplicative Scatter Correction.

Figure 1.3 contains a visual representation of some example transformations. The axes in each of the plots represent the reflectance values at a given wavenumber. The line A is representative of a constant offset from the reference spectrum. Line B represents a linear slope deviation from the reference spectrum. However, in reality the intensity values of the original spectrum are not identical to the reference spectrum aside from slope or offsets. Line C represents a somewhat more realistic scenario where the intensities differ beyond the simple transformations discussed. Finally line D demonstrates the corrected form of line C, still possessing the undulations of line C which are attributed to useful information. The equation for the corrected spectra can be seen below.

$$\boldsymbol{x_{i,corr}} = \frac{\boldsymbol{x_{i,org}} - a_i}{b_i} \quad (2)$$

MSC has been used in the context of spectral pre-treating since as early as 1985, [58] and it is still commonly used across applications in NIR and MIR chemometrics. [59-61]

1.5.1b Standard Normal Variate

Another very common method of scatter correction, standard normal variate, SNV formulates an expression for the corrected spectrum that is very similar to the MSC. However, SNV has the benefit of not requiring a reference spectrum.

$$\boldsymbol{x}_{\boldsymbol{i,corr}}^{\boldsymbol{T}} = \frac{\boldsymbol{x}_{\boldsymbol{i,org}}^{\boldsymbol{T}} - \mu_{x_{org}}}{\sigma_{x_{org}}} \tag{3}$$

Where $\sigma_{x_{org}}$ and $\mu_{x_{org}}$ are the standard deviation and mean of the original spectrum, respectively. SNV is applied predominantly in the same situations as MSC, as such SNV is also very prevalent amongst IR literature. [62-64]

1.5.2 Savitsky – Golay Filtering

The Savitsky – Golay filter [65] is a combination of a spectral derivative and polynomial convolution smoothing. The smoothing step is included to reduce the exaggeration of high frequency terms. The smoothing step provides new data points based on a summation of the previous data points, within some user-defined window. The summation includes coefficients, the values of which vary in the same fashion as a polynomial (demonstrated in Figure 1.4) the order of which is, again, user-defined, the benefit of smoothing data in this fashion is that peak structures are maintained. Other smoothing options such as Norris-Williams [166] provide no preference to any data point, and suppress peak structures as a result.

$$x_i' = \frac{1}{n}\sum_{i}^{j} c_{n,i}x_i = \frac{1}{n}\boldsymbol{c}\boldsymbol{x}^{\boldsymbol{T}} \tag{4}$$

In equation 4, x_i' represents the smoothed data point c represents the corresponding weighting coefficient, and n is the number of points in the summation window. A linear algebra version is also included in equation 4,

$$C^T = [C_{2,1} \quad C_{2,2} \quad C_{2,3}]$$
$$C^T = [0.3 \quad 1 \quad 0.3]$$

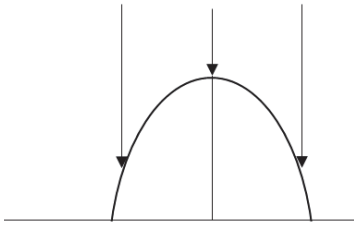

Figure 1.4: Simplified example of three polynomial weights.

where $\boldsymbol{c}$ is a column vector containing polynomial weights, $\boldsymbol{x}^{\boldsymbol{T}}$ is a portion of the

original data to be smoothed. The vectors $\boldsymbol{c}$ and $\boldsymbol{x^T}$ contain $j - i$ elements, where i is the first data point of the smoothing window and j the last. The Savitsky – Golay filter has been seen to be useful in combination with either of the scatter correction methods [66, 67] mentioned previously.

1.5.3 Asymmetric Least Squares

The asymmetric least squares (AsLS) algorithm identifies the baseline by using the spectrum itself as a starting point. The AsLS algorithm builds upon the work of Whittaker.[68] Consider a spectrum to be represented by a column vector, $\boldsymbol{y}$ with the dimensions $k \times 1$, for ATR-FTIR this vector contains reflectance values and, k represents the total number of reflectance values identified, thus, each element may be labelled with the appropriate wavenumber. Asymmetric least squares approximates the baseline by finding a second vector $\boldsymbol{z}$, of equal dimensions, this vector will satisfy the conditions of being similar to $\boldsymbol{y}$ and being smooth. Both the similarity and smoothness of $\boldsymbol{z}$ can be adjusted in the process to generate a satisfactory baseline. These features may be represented in the following equation. [69]

$$F = \sum_i (y_i - z_i)^2 + \sum_i (z_i - z_{i-1})^2$$

In matrix form.

$$\boldsymbol{f} = \boldsymbol{W}|\boldsymbol{y} - \boldsymbol{z}|^2 + \lambda|\boldsymbol{D}\boldsymbol{z}|^2 \qquad (5)$$

Where $|\boldsymbol{y} - \boldsymbol{z}|^2$ is a vector representing the squared differences between the vectors $\boldsymbol{y}$ and $\boldsymbol{z}$, this vector contributes to the term representing the similarity of $\boldsymbol{y}$ and $\boldsymbol{z}$. The other portion of the similarity term is $\boldsymbol{W}$, a matrix containing weights along the diagonal.

$$\boldsymbol{W} = \begin{bmatrix} w_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & w_i \end{bmatrix}$$

The values of the diagonal entries of $\boldsymbol{W}$ are determined based on the following relationship.

$$w_i = \begin{cases} p; & y_i > z_i \\ 1-p; & y_i < z_i \end{cases}$$

The weighting matrix, $\boldsymbol{W}$ contains a parameter p which the user can vary, small values of p generate baselines below the vector $\boldsymbol{y}$ and vice versa. Small values of p are desirable in the context of Raman spectroscopy and hence it has seen relatively frequent use in this field.[70-73] However, for ATR-FTIR reflectance spectra, large values of p are required.The second term of $\boldsymbol{f}$ describes the smoothness, by generating a vector which contains in its elements values relating to the difference of some element z_i with some preceding element z_{i-1}. The differences are generated using the following matrix, $\boldsymbol{D}$.

$$\boldsymbol{D} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

$\boldsymbol{D}$ will, in reality, have k columns and $k-1$ rows. In equation 5, λ is a parameter that represents the significance of the smoothness term, if λ is large the vector $\boldsymbol{z}$ will have more freedom to undulate. For some additional detail on the derivation of , see Appendix 1. For a brief graphical representation of the identification of $\boldsymbol{z}$ see Figure 1.5.

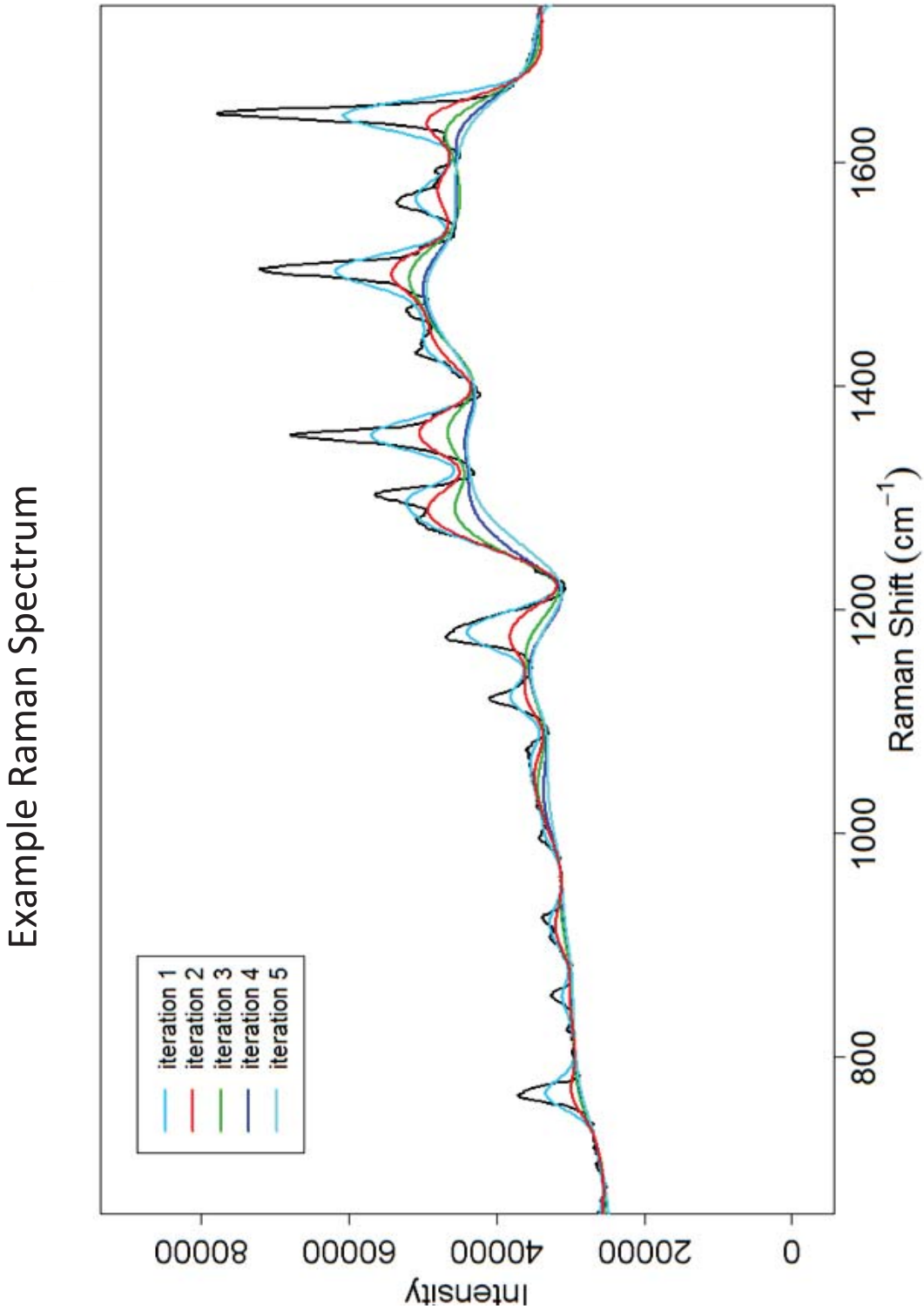

Figure 1.5: Example of the variation in the identified baseline after 5 iterations ($\lambda = 10^4, p = 0.005$). Note the baseline can be manipulated to sit further below the spectrum and undulate less.

1.5.4 Discrete Wavelet Transform

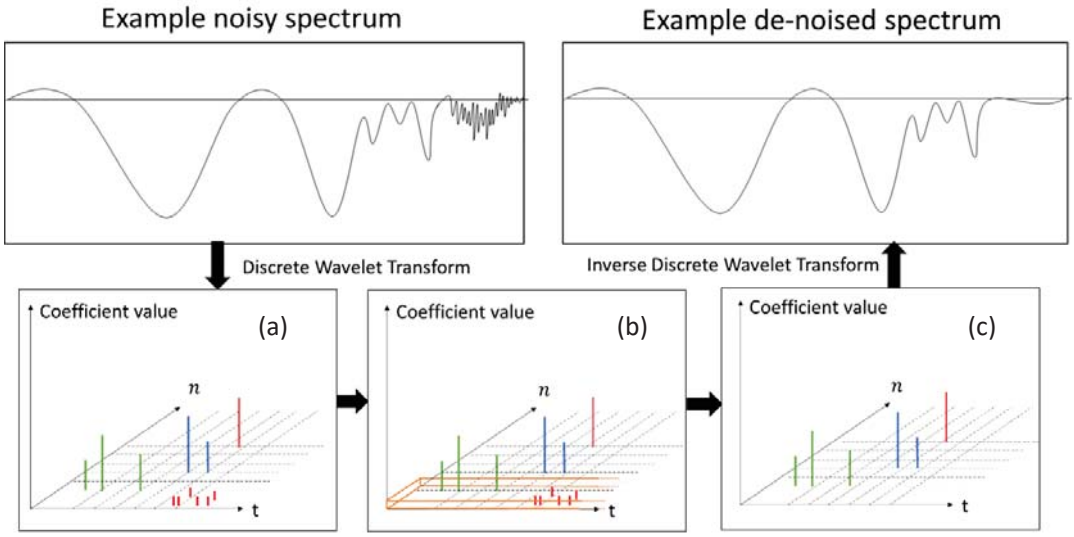

Figure 1.6: An overview of noise removal using the hard threshold approach for DWT. (a) Representation of the decomposed function in the scale (n), translation (t) and coefficient value space. (b) The orange region represents the noise domain within which components of the original function possess noise-like characteristics. (c) The decomposed function now with noise-like terms removed.

The Discrete Wavelet Transform (DWT) decomposes a given spectrum into combinations of scaling and wavelet functions, of varying scale and translation. In the context of spectroscopy, the translations are shifts in wavenumber, whilst the scale is inverse wavenumber (wavelength). A typical wavelet transformation of an $N \times 1$ vector, $\boldsymbol{f}$ may be represented in the language of linear algebra as below.[165]

$$\boldsymbol{\alpha} = \boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{f} \qquad (6)$$

$$\boldsymbol{W}\boldsymbol{W}^{\boldsymbol{T}} = \mathbf{I}$$

Where, $\boldsymbol{\alpha}$ contains the wavelet coefficients and $\boldsymbol{W}$ is an $N \times N$ matrix containing the orthonormal basis vectors (constructed in the same fashion as the example in Appendix 2). The Discrete wavelet form of the general wavelet transform is seen in Equation 7.

$$\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{d} \end{bmatrix} = \begin{bmatrix} \boldsymbol{S} \\ \boldsymbol{D} \end{bmatrix} \boldsymbol{f} \tag{7}$$

Where $\boldsymbol{a}$ and $\boldsymbol{d}$ contain the approximation and detail coefficients. The matrices $\boldsymbol{S}$ and $\boldsymbol{D}$ contain orthogonal discrete basis functions (scaling and wavelet, respectively), the rows of the matrices represent shifts of these basis functions. The process of decomposition is iterative using the algorithm of Mallat, [74] identifying relative coefficients as follows, where n is the level of decomposition.

$$\boldsymbol{a^n = S a^{n-1}} \tag{8}$$

$$\boldsymbol{S S^T = \mathrm{I}}$$

$$\boldsymbol{d^n = D a^{n-1}} \tag{9}$$

$$\boldsymbol{D D^T = \mathrm{I}}$$

The function may now be represented as follows.

$$\begin{aligned} \boldsymbol{f} &= \boldsymbol{a^0} \\ &= \boldsymbol{D_1^T d^1 + S_1^T a^1} \\ &= \boldsymbol{D_1^T d^1 + S_1^T \left[D_2^T d^2 + S_2^T a^2 \right]} \\ &= \boldsymbol{D_1^T d^1 + S_1^T \left[D_2^T d^2 + S_2^T \left[D_3^T d^3 + S_3^T a^3 \right] \right]} \end{aligned} \tag{10}$$

The benefit of this method of decomposition over other forms such as the Fourier transform (FT), is clear when it is applied to data with localised features, in such a scenario DWT provides fewer terms than the FT. There exist many different *familie*s of wavelets, differing in their geometry, as such wavelets need to be selected such that they allow the decomposition and accurate reconstruction of the signal. In general wavelet selection is fairly subjective, however,

the literature reflects that the Daubechies[75,76] family performs well [77] for infrared spectra. Noise suppression can be achieved through several means.[78] The approach utilised in this project is known as thresholding, specifically hard thresholding. Hard thresholding, as used in this research is equivalent to identifying expected features of noise contributions, and removing wavelets possessing these features (see Figure 1.6).[78] The two features outlined as indicative of noise, are low coefficient values and high scaling (low n), as illustrated in Figure 1.6. The coefficient threshold is important especially if the original spectrum is expected to possess high *frequency* terms of significance (as found in derivative plots). For a brief example of identifying scaling and wavelet functions, and on the decomposition of a function by discrete wavelets, see Appendix 2. [79]

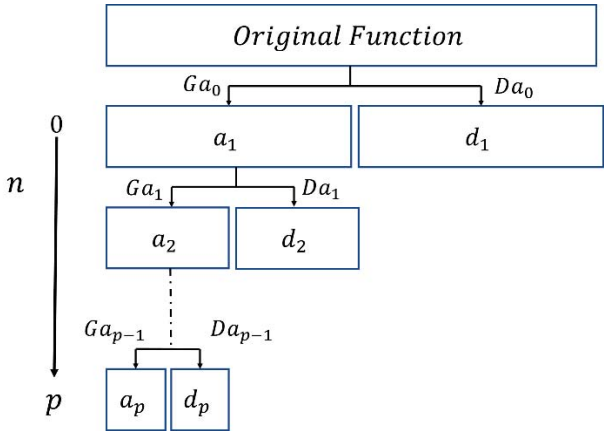

Figure 1.7: Graphical representation of the DWT decomposition process.

1.6 Multivariate Regression

Multivariate regression is concerned with identifying relevant information within a large set of variates, in order to predict response values. The methods of multivariate regression employed in this research are linear regression methods. They all build upon the basic principles of ordinary least squares regression. [80] Ordinary least squares formulates linear combinations of the predictor variates in the following fashion.

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta}_{\boldsymbol{ols}} + \boldsymbol{e} \quad (11)$$

Where $\hat{\boldsymbol{y}}$ is an $n \times 1$ column vector containing response values, $\boldsymbol{X}$ is an $n \times p$ matrix where each row is an observation and each column is a variable. $\boldsymbol{\beta}_{\boldsymbol{ols}}$ is a $p \times 1$ column vector of regression coefficients and $\boldsymbol{e}$ is an $n \times 1$ column vector of residual terms. In the context of spectroscopy, each row of $\boldsymbol{X}$ represents a spectrum obtained for a given sample. The equivalent row of $\hat{\boldsymbol{y}}$ contains some chemical component of interest, for instance, protein content, ideally, it should be possible to relate the entire spectrum of each sample to each component. Practically, however, spectroscopic data possesses several undesirable features that make regression analysis by OLS very difficult. The performance of OLS decreases drastically if neighbouring columns of the $\boldsymbol{X}$ matrix vary together (multi-collinear), this issue results in ill – conditioning [81] of the $\boldsymbol{X}$ matrix, a discussion of how this arises and the particular effects that result may be viewed hapter 2.2.1. The high dimensionality of the $\boldsymbol{X}$ matrix is also an issue, this refers to the fact $p \gg n$, this results in a greater number of unknowns (coefficients) than equations, as such no unique solutions. There are two prominent approaches to dealing with any issues introduced through high dimensionality, these are known as feature selection and feature extraction.

1.7 Principal Component Analysis

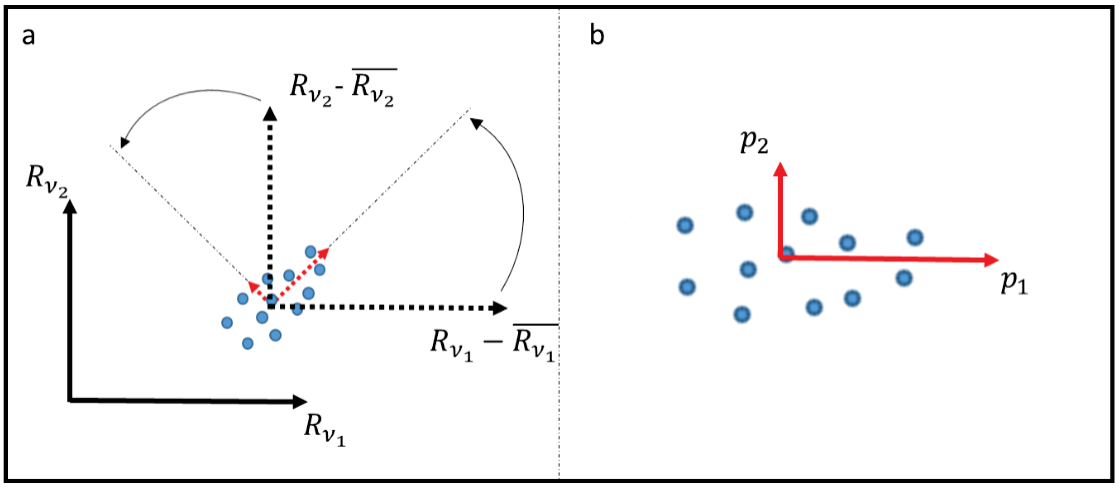

Figure 1.8: (a) illustrates the initial variable space (relative to the project) in a reduced form. The dashed axes are the mean subtracted axes, the red dashed axes represent the principal components. (b) represents the original data after the passive transformation to principal component space.

Principal component analysis[82] (PCA) is not a regression method, but rather a means of identifying what variables are significant within the $\boldsymbol{X}$ matrix. Using a geometrical interpretation, if we consider each variable as a "direction" in a p −dimensional space, then the principal components align with the direction of greatest variance within the initial $p-$ dimensional space. The principal components are formed as linear combinations of the initial p variable vectors. These components provide great insight into trends already present within the $\boldsymbol{X}$ matrix, such as clustering of observations. The components also provide insight into what variables of the original dataset are varying the most between observations. Principal component analysis results are often presented as scores and loadings plots. The scores plots use the principal components as the axes and display observations relative to these new axes, equivalent to a passive coordinate transformation. The loading plots reflect how well-aligned each original variable vector is with a given principal component, as such the loading values are equivalent to inner products.

1.8 Regression Methods

As has been mentioned, issues associated with high dimensional data matrices, may be overcome through feature selection or construction. The regression methods utilised in this project all fall into one of the two categories. Before discussing these methods, the method for identifying the regression vector $\boldsymbol{\beta}_{ols}$ in ordinary least squares will be described. The primary criterion for determining $\boldsymbol{\beta}_{ols}$ is minimisation of the squared error function ($\boldsymbol{ee^T}$):

$$\boldsymbol{ee^T} = (\boldsymbol{y} - \boldsymbol{X\beta})^{\boldsymbol{T}}(\boldsymbol{y} - \boldsymbol{X\beta}) \tag{12}$$

The vector, $\boldsymbol{\beta}_{ols}$ that minimises $\boldsymbol{ee^T}$ is given by the following equation**,**

$$\boldsymbol{\beta}_{ols} = (\boldsymbol{X^T X})^{\boldsymbol{-1}}\boldsymbol{X^T y} \tag{13}$$

The first three regression methods used in this thesis provide additional terms to the squared error function, which act to reduce any negative results due to multi-collinearity. These issues are supressed at the expense of an increased squared error associated with the identified regression vector. The latter two methods discussed will construct features, combining them to form a new matrix $\boldsymbol{X_C}$ possessing fewer multi-collinearity related issues.

1.8.1 Ridge Regression

Ridge regression was first described in 1970[83] by Hoerl *et al.*, it introduces a parameter, λ, that shrinks the elements of the regression coefficient vector.[84] The parameter is varied until a model with suitable low mean squared error is produced. Equations 14 and 15 below present the squared error function with Ridge penalty and Ridge regression coefficient vector, respectively.

$$\boldsymbol{ee^T}(Ridge) = (\boldsymbol{y} - \boldsymbol{X\beta})^T(\boldsymbol{y} - \boldsymbol{X\beta}) + \lambda_1|\boldsymbol{\beta}|^2 \tag{14}$$

$$\boldsymbol{\beta_R} = (\boldsymbol{X^TX} + \lambda_1\mathbf{I})^{-1}\boldsymbol{X^Ty} \tag{15}$$

Shrinkage occurs more significantly as λ_1 is increased, due to the growth of the inverse term in the equation for $\boldsymbol{\beta_R}$.

1.8.2 LASSO Regression

The Least Absolute Shrinkage and Selection Operator [85] (LASSO), is a fairly recent method providing a subtle but significant alteration to the Ridge regression method. The penalty imposed by the LASSO method varies only slightly to that of the Ridge method. The regression coefficient vector produced by the LASSO method will likely contain elements of zero value, this result provides the most distinct difference between the Ridge and LASSO methods. The Ridge method will shrink the contribution of variables of lesser relevance, but, LASSO will remove them entirely. Equations 16 and 17 present the squared error function with LASSO penalty and LASSO regression coefficient vector, respectively.

$$\boldsymbol{ee^T}(LASSO) = (\boldsymbol{y} - \boldsymbol{X\beta})^T(\boldsymbol{y} - \boldsymbol{X\beta}) + \lambda_2|\boldsymbol{\beta}| \tag{16}$$

$$\boldsymbol{\beta_L} = (\boldsymbol{X^T X})^{-1}(\boldsymbol{X^T y} - \frac{\lambda_2 \boldsymbol{W}}{2}) \quad (17)$$

$$\boldsymbol{diag(W)} = \begin{cases} 1 & \beta_i > 0 \\ -1 & \beta_i < 0 \end{cases}$$

Now the modification to the typical $\boldsymbol{\beta_{ols}}$ equation appears as a negative term in the parentheses, the effect of which is to set elements of $\boldsymbol{\beta_L}$ to zero as they shrink below a limit governed by λ_2. LASSO's built – in variable selection has led to it being utilised in a variety of studies involving high dimensional data. [86-89]

1.8.3 Elastic Net Regression

The Elastic Net method [90] combines the approaches of Ridge and LASSO in order to generate a regression coefficient vector possessing the desirable traits of both the LASSO and Ridge equivalents. Equations 18 and 19 below demonstrate the Elastic Net penalty and the resulting Elastic Net regression coefficient vector, respectively.

$$\boldsymbol{ee^T} = (\boldsymbol{y} - \boldsymbol{X\beta})^{\boldsymbol{T}}(\boldsymbol{y} - \boldsymbol{X\beta}) + \lambda[(1-\alpha)|\boldsymbol{\beta}| + \alpha|\boldsymbol{\beta}|^2] \quad (18)$$

$$\alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2}$$

$$\boldsymbol{\beta_{elnet}} = (\boldsymbol{X^T X} + \lambda\alpha\boldsymbol{I})^{-1}(\boldsymbol{X^T y} - \frac{\lambda(1-\alpha)\boldsymbol{w}}{2}) \quad (19)$$

From the equation for $\boldsymbol{\beta_{elnet}}$, it is immediately clear the shrinkage and selection characteristic are both present. The Elastic Net also has seen application across of wide range of studies, despite its relative recent development. [91-99]

1.8.4 Principal Component Regression

This method is the first of two feature construction methods employed in this research. Principal component regression constructs a new matrix $\boldsymbol{X_{PCR}}$ for regression analysis, this has the dimensions of $n \times c$ where c is the number of principal components augmented. These principal components are the same as previously discussed in the PCA section. These components are orthogonal to one another and as such the constructed matrix is devoid of issues associated with column linear dependence (i.e. multi – collinearity). Another important result of this method is $c < p$ thus this method also achieves dimension reduction.

$$\boldsymbol{p_i} = \boldsymbol{X\gamma_i} \tag{20}$$

$$\boldsymbol{p_i}\ being\ n \times 1$$

$$\boldsymbol{X_{PCR}} = [\boldsymbol{p_1} \quad \boldsymbol{p_2} \quad \cdots \quad \boldsymbol{p_c}]$$

$$\hat{\boldsymbol{y}} = \boldsymbol{X_{PCR}\beta} + \boldsymbol{e} \tag{21}$$

$$\boldsymbol{\beta_{PCR}} = \left(\boldsymbol{X_{PCR}^T X_{PCR}}\right)^{-1} \boldsymbol{X_{PCR}^T y} \tag{22}$$

Where $\boldsymbol{p_i}$ is the i^{th} principal component, and $\boldsymbol{\gamma_i}$ is the i^{th} eigenvector of $\boldsymbol{X^T X}$. As is clear from the equations above, PCR proceeds in the usual fashion to derive $\boldsymbol{\beta_{PCR}}$. Whilst PCR has existed in some capacity for a relatively long time, it still sees application across a range of disciplines. [100-103]

1.8.5 Partial Least Squares

Partial Least squares was developed in the mid 1970's to early 1980's, through the joint efforts of father and son Herman and Svante Wold,[104-106] two prominent figures in the field of chemometrics. Partial least squares defines several vectors, the vectors may be grouped in the following fashion. The vectors $\boldsymbol{w}$ and $\boldsymbol{c}$ are referred to as weight vectors, they assign weights to the variables (columns) of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. Weighting is allocated based on the relevance of the column to either the $\boldsymbol{t}$ or $\boldsymbol{u}$ vector. The vectors, $\boldsymbol{t}$ and $\boldsymbol{u}$ are referred to as scores vectors, representing how significant an observation is to the corresponding weight vectors. Finally $\boldsymbol{p}$ and $\boldsymbol{q}$ are referred to as loading vectors, the phrase carries a similar meaning as it does in the context of principal components analysis, the loadings reflect variable significance to the scores vectors.

$$\boldsymbol{u} = \boldsymbol{y}$$

$$\boldsymbol{w} = \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{u} \tag{22}$$

$$\boldsymbol{w} = \frac{\boldsymbol{w}}{|\boldsymbol{w}|}, \qquad \boldsymbol{normalisation\ of\ w} \tag{23}$$

$$\boldsymbol{t} = \boldsymbol{X}\boldsymbol{w} \tag{24}$$

$$\boldsymbol{c} = \frac{\boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{t}}{\boldsymbol{t}^{\boldsymbol{T}}\boldsymbol{t}} \tag{25}$$

$$\boldsymbol{c} = \frac{\boldsymbol{c}}{|\boldsymbol{c}|}, \qquad \boldsymbol{normalisation\ of\ c} \tag{26}$$

$$\boldsymbol{u} = \boldsymbol{Y}\boldsymbol{c} \tag{27}$$

Equations 22 – 27 are carried out iteratively in the order presented, until the vector, $\boldsymbol{u}$ stabilises, at which point the algorithm identifies two further vectors, $\boldsymbol{p}$ and $\boldsymbol{q}$.

$$\boldsymbol{p} = \frac{\boldsymbol{X^T t}}{\boldsymbol{t^T t}} \tag{28}$$

$$\boldsymbol{q} = \frac{\boldsymbol{Y^T u}}{\boldsymbol{u^T u}} \tag{29}$$

Some of these vectors may be combined to form a matrix, which may then be subtracted from the original $\boldsymbol{X}$ and $\boldsymbol{Y}$ matrices.

$$\boldsymbol{X_i} = \boldsymbol{X_{i-1}} - \boldsymbol{tp^T} \tag{30}$$

$$\boldsymbol{Y_i} = \boldsymbol{Y_{i-1}} - \boldsymbol{tc^T} \tag{31}$$

At which point the process begins again, inserting these matrices back into the previous. The matrices $\boldsymbol{X_i}$ and $\boldsymbol{Y_i}$ are reduced in dimension and represent so-called deflation matrices, *i* refers to the iteration. The matrices subtracted at each deflation may be combined to form matrices of all the constructed information. This constructed matrix may be used for further regression analysis as was seen for PCR. PLS has seen incorporation in a wide variety of research projects, [107-113] including an assortment of FTIR based studies. [114-120]

1.9 Chemometrics: Infrared Spectroscopy of Forage Feeds

Mention of infrared-based chemometric analysis of lamb faecal samples, chicken feed and hyperaccumulator plants is limited. However, amongst the literature the use of infrared spectroscopy based techniques in the prediction of agriculturally relevant chemical compositions is quite common. [121-127] In this field NIR techniques are the most widespread. In terms of forage quality predictions, NIR has been prevalent since 1976. [128] In 2016 a communication paper was published by Berzaghi *et al*. [129] summarising the predictive performance of a portable NIR instrument for farm forage analysis, they developed prediction models from a sample set of more than 300 samples. Berzaghi *et al*. identified with moderate accuracy NDF and DM in new samples (R^2: 0.82, 0.86). It is worth noting that the instrument outlined as "portable" in this research was not ergonomic and its dimensions seem quite cumbersome. [130]

NDF determination via NIR predictions [131] was also carried out in the work of Monrroy *et al.*, where they also predicted ADF, cellulose and CP. In their research they utilised a smaller sample set than Berzaghi, however, they looked exclusively at the species *Brachiaria* spp. Their external validation results produced a variety of model qualities; NDF and ADF were modelled rather well (R^2: 0.86, 0.90), whilst cellulose and crude protein were poorly accounted for (R^2: 0.73, 0.53).

Asekova *et al.* investigated the potential of NIR – models to predict NDF, ADF, CP and crude fat in soybeans. Their findings were based on models developed using over 300 samples, and employing a modified version of PLS. They tested the developed models on ~ 90 external validation samples and recorded large R^2 for CP and crude fat (0.91,0.93). [167] However, ADF and NDF were predicted

with less accuracy than has been seen in the previously mentioned literature, now with R^2 of 0.77 and 0.75. Despite the abundant use of NIR in the literature, in terms of real world applications its most valuable form is with hand held instruments. There is reason to believe hand-held NIR devices, such as diffuse reflectance based devices, are very susceptible to a variety of environmental factors. [132] Such sensitivity brings into question the ability of NIR –models to be transferred to hand held devices for *in situ* analysis. [132]

The work of *Pierna et al.* [133] claims to demonstrate the simple transfer of a large NIR calibration database from one instrument to another. The transfer process re-calculated the models on a combined sample set incorporating a portion of the new samples. The results are somewhat debatable based on the tremendous difference of sample sizes. The research of Santos *et al.* demonstrated explicitly that for classification and quantification of milk product adulteration, hand-held MIR (ATR-FTIR) out performs hand-held NIR.[134] Some authors have also suggested models derived from the MIR region possess greater predictive ability, although their suggestions are in a slightly different context.[135,136]

ATR-FTIR has been used as a technique for the analysis and differentiation of natural fibres, [137] with cellulosic components. Cellulosic fibres are also found in a variety of the chemical components investigated in the research of this thesis, suggesting ATR – FTIR may perform well in this context.

1.10 Project aims

Outlined in the literature are numerous predictions of chemical compositions by infrared spectroscopy, the agricultural domain seeing common employment of NIR. This project aims to explore the potential links between ATR-FTIR spectra, and a variety of chemical components relevant to agricultural based science. The project will employ several different methods in order to establish whether or not the link exists. The project aims to compare the ability of different combinations of mathematical pre-treatments and regression methods, comparisons will be made based on relevant statistical metrics. The project aims to provide a technique capable of introducing a new form of chemical composition analysis of biological samples from agricultural and environmental systems. The use of ATR-FTIR for chemical composition analysis is suggested to be an additional tool rather than one intended to supersede current methods.

Chapter 2.0: Methodology and Technical Background

2.1: Methodology

2.1.1 Initial Forage Feed Samples

A total of 140 herbage samples were used for analysis, consisting of 84 ryegrass - white clover samples and 56 herb-clover mix samples including chicory, plantain , white clover and red clover. The samples were collected in 2012, 2013, 2015 and 2016 on Massey University farms close to Palmerston North in New Zealand (40°38'S and 175°37' E). The samples were frozen (-20 °C) and then freeze dried before chemical analysis.

2.1.2 Expanded Forage Feed Samples

A total of 188 samples were used for analysis. The samples were collected in 2012, 2013, 2015 and 2016 on Massey University farms close to Palmerston North in New Zealand (40°38'S and 175°37' E). The samples were frozen (-20 °C) and then freeze dried before chemical analysis. Samples contained a mix of herb, pasture, lucerne and unknown sample species.

2.1.3 Chicken Feed Samples

A total of 56 feed samples were used for analysis. The feed samples were provided courtesy of Dr Fifi Zaefarian. They comprised 56 mixtures of 10 feed ingredients, including maize, wheat, sorghum, soybean meal, canola meal, meat and bone meal, wheat bran, wheat DDGS, full fat soybean and palm kernel meal. All 10 ingredients have been included in each feed mixture with either of three different inclusion levels (2, 42 or 82%).

2.1.4 Faecal Samples

A total of 54 lamb faecal samples were used for analysis. The faecal samples were collected in 2015, and provided for this research project by Antoinette Danso, a PhD student under the supervision of Professor Patrick Morel of the Insitute of Veterinary, Animal and Biological Sciences, Massey University.

2.1.5 Hyperaccumulators

A total of 40 hyperaccumulator plant samples were used for analysis. The samples consisted of ground leaf matter from *Alyssum sp* and *Bherkeya coddii* species, kindly provided by Associate Professor Chris Anderson.

2.1.6 Wet Chemical Analysis

Aside from nickel content in the hyperaccumulators, all chemical compositions were determined at the Massey University Nutrition Laboratory. The samples were analysed for N using the Dumas total combustion method (AOAC method 968.06),[138] and dry matter (DM) using a convection oven at 105 ^{O}C (AOAC methods 930.15 and 925.10).[138] Neutral detergent fibre (NDF), Lignin and acid detergent fibre (ADF) were determined using the Tecator Fibretec System following the method described by Robertson and Van Soest[139] (1981) (AOAC method 2002.04).[138] The in vitro digestible organic matter content (DOMD) and metabolisable energy (ME = 0.16 × DOMD) were determined by the method of Roughan and Holland.[140] Ash was determined in a furnace at 550 °C (AOAC method 942.05) [138] and OM (AOAC method 942.05).[138] Gross energy was determined using a bomb calorimeter (Leco, AC 350, Leco Corporation, St. Joseph, MI). Fat was determined by the Soxtec extraction method (AOAC method 991.36). [138] Nickel content of hyperaccumulator samples was determined using a microwave plasma-atomic emission spectrometer

(MP-AES). The chemical compositions not mentioned were carried out using standard methods at the Massey University Nutrition Laboratory, all of the determined compositions were of the expected range (see Appendix 8). [173-177]

2.1.7 ATR-FTIR Spectra Collection

Infrared spectra were recorded in reflectance mode using a Thermofisher Scientific Nicolet 5700 equipped with a germanium crystal Smart iTR™ Attenuated total reflectance sampling accessory and OMNIC™ series software. Dried samples were ground prior to analysis in order to distribute the organic components as homogeneously as possible. Reflectance data was recorded over the wavenumber range of $675 - 4000$ cm^{-1} with 64 scans per spectrum and a spectral resolution of 4 cm^{-1}. The acquisition time for a single spectrum was 66 seconds. Background spectra were collected with no samples present on the crystal, and under the same experimental conditions. Fresh background spectra were acquired periodically for data collection sessions longer than one hour. Beam alignment was checked before collection. For every sample three spectra were collected, an average was evaluated from these in order to further account for inhomogeneity. Repeatability was determined by collecting spectral information of one particular sample over a day, at regular intervals. Comparing these results to those of other samples demonstrated consistent clustering based on the sample, rather than the session of data collection (demonstrated using PCA, see Appendix 7). The files were exported as comma separated value (csv) files and imported into the *R* environment for analysis (all spectra are stored online, and available at request).

2.1.8 Statistical Analysis

From a total of 969 initial spectra, averaging generated 323 spectra, with each averaged spectrum representing a different sample (from within one of the four distinct sample types).

Pre-treatment of the averaged spectra was required to eliminate random effects due to sample presentation on the ATR crystal.[53] Six pre-treatments (PT) were investigated, the first (PT1) was a scatter correction technique, Multiplicative Scatter Correction (MSC). The second (PT2) consisted of another scatter correction technique, Standard Normal Variate (SNV) scaling. [128] PT3 applied a Savitzky-Golay (SG) derivative[65] after SNV to remove slowly varying components unrelated to changes in chemical composition. This research utilised a first – order derivative followed by smoothing using a polynomial weighting. Here, the smoothing polynomial was third-order and the window of smoothing was 11 units. A Discrete Wavelet Transform (DWT)[141] filter was utilised for PT4 and applied post-SNV. The mother wavelet for the transform was a Daubechies-8 wavelet, and the noise was reduced by suppression of non-signal high frequency terms using a wavelet-coefficient threshold approach. [142] PT5 was a product of PT3 followed by DWT-filtering with a Daubechies-12 mother wavelet. PT6 was AsLS applied post-SNV, with a weighting of 0.9999 and λ of 10^6.

A preliminary analysis of the both the whole dataset, and each different sample set, was carried out using Principal Component Analysis [82] (PCA). For the forage sample sets, the samples were divided into calibration and external validation sets (insufficient samples to allow for this in the other sample sets), with ~75% of the sample set used for model formulation and the remainder used as an external validation data set. A modification was made in the algorithms, allowing one initial model calibration to

identify and remove prediction outliers (with errors of prediction further than 2.5 standard deviations away from the average value), after which the model was recalibrated on the reduced calibration subset. During model formulation, a k-fold cross-validation[143] approach was incorporated to refine and identify the best model, with k = 10.

The optimum number of components (PCR and PLS) or variables (Ridge, LASSO or Elastic Net) in the regression model was determined based on the values of root-mean-squared error of cross-validation (RMSECV). The optimum number of components (or variables) was the smallest value that gave a RMSECV within one standard error of the lowest RMSECV value. [144] RMSECV was calculated using the following equation.[145]

$$\mathbf{RMSECV} = \sqrt{\frac{\sum_i \left(y_{est,i} - y_{known,i}\right)^2}{n}}$$

Where n is the number of values, and with the square root the numerator contains the total squared difference between estimated and known values. The selected model's performance was determined for the calibration, cross-validation and external validation datasets. A robust PCR or PLS model will possess a lower number of components, n_c,[146,147] as this reflects a lower risk of overfitting.[148]

The following metrics were used for the evaluation of prediction quality, root-mean-square error of prediction (RMSEP), standard error of prediction (SEP_c),[149] the coefficient of determination, R^2 and relative performance deviation (RPD_c).

Note that for each metric the acronym and equation vary slightly based on the context, therefore for clarity the equations used in this work are given below. The equations for SEP, bias[149] and RPD are:

$$\mathrm{SEP} = \sqrt{\frac{\sum_i \left(y_{est,i} - y_{known,i} - \mathrm{bias}\right)^2}{n-1}}$$

$$\mathrm{bias} = \frac{\sum_i (y_{est,i} - y_{known,i})}{n}$$

$$\mathrm{RPD} = \frac{\sigma}{\mathrm{SEP_c}}$$

Where σ is the standard deviation.

2.2.0 Technical Background: Regression Methods

This section is designed to provide an additional level of detail relating to the regression methods employed in this research. The information discussed in this section allows for further intuition regarding the motivation for the methods, and the geometric relationships between the mathematical objects of each method. The motivation will be developed by highlighting the short comings of the original *ordinary least squares* method, each method employs a slightly different approach to dealing with these short comings. The different approaches of each method will be discussed and explained, some additional information has been included in the appendices and is referenced when appropriate. The level of detail this section provides is beyond the level required to interpret the upcoming results and discussion sections, however, as mentioned it will enhance the reader's intuition and familiarity with the concepts.

2.2.1 Introduction

Regression is the process of relating two variable spaces. A linear regression equation appears as follows

$$y_n = b_0 + \sum_{i=1}^{n} b_i x_i + \sum_{i=0}^{n} \epsilon_i$$

Or the equivalent matrix form.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} \tag{1}$$

Where in this equation and for the remainder of the chapter the following notation will be used:

- bold capital letter implies a matrix
- bold lowercase letter implies a column vector
- bold lowercase letter with superscript **T** implies a row vector or in general the transpose of an object
- regular lowercase letter implies a scalar quantity

Hence the matrix equation provided illustrates the column vector of responses, $\boldsymbol{y}$ being equal to the product of a matrix $\boldsymbol{X}$ with a column vector of regression coefficients $\boldsymbol{\beta}$, and an additional disturbances term denoted as $\boldsymbol{\epsilon}$.

The equation above shows a single response variable, for some of this chapter the general case of $\boldsymbol{X}$ being of $n \times p$ dimensions and $\boldsymbol{Y}$ of $n \times m$ dimensions will also be used. For this research n is the number of observations (samples in $\boldsymbol{X}$, corresponding chemical composition in $\boldsymbol{Y}$), p is the number of predictor variables (reflectance at a wavenumber) and m is the number of response variables.

Equation 1 includes the disturbance term ϵ, where ϵ captures the difference between the measured value and the true value. The term ϵ is, by its very nature, unobservable. Section 2 will begin with a description of the ordinary least squares (OLS) model, the description will lead to a motivation for more advanced regression methods. Two classes of regression methods utilised in this research are.

(i) Feature extraction methods

(ii) Feature construction methods.

In this work, the feature extraction methods employed are Ridge, LASSO and Elastic Net regression. The feature construction methods are principal components regression and partial least squares regression.

2.2.2 Ordinary Least Squares regression

The OLS model contains a *residual* term $\boldsymbol{e}$ in place of the disturbances term. The residual is observable and represents the difference between an observed and estimated value. The ordinary least squares model is.

$$\boldsymbol{y} = \boldsymbol{X\beta_{ols}} + \boldsymbol{e} \tag{2}$$

Equation 2 can be easily expanded to a multiple response version with the following modifications.

$$\boldsymbol{y} \rightarrow \boldsymbol{Y}$$

$$\boldsymbol{\beta_{ols}} \rightarrow \boldsymbol{B_{ols}}$$

$$\boldsymbol{e} \rightarrow \mathbf{E}$$

The matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ are mean centred. Having at this stage identified a general relationship between two variable spaces, the next step is to find an expression for $\boldsymbol{\beta_{ols}}$ under the condition of

$$\frac{\partial \boldsymbol{e^T e}}{\partial \boldsymbol{\beta^T}} = 0$$

where $\boldsymbol{e^T e} = \left|\boldsymbol{e^2}\right|$ or the squared *error*. The motivation for minimising $\boldsymbol{e^T e}$ is it ensures $|\boldsymbol{e}| \cong 0$. The procedure for identifing $\boldsymbol{\beta_{ols}}$ in this fashion is as follows

$$\begin{aligned}\boldsymbol{e^T e} &= (\boldsymbol{y} - \boldsymbol{X\beta_{ols}})^T(\boldsymbol{y} - \boldsymbol{X\beta_{ols}}) \\ &= \left(\boldsymbol{y^T} - \boldsymbol{\beta_{ols}}^T\boldsymbol{X^T}\right)(\boldsymbol{y} - \boldsymbol{X\beta_{ols}}) \\ &= \boldsymbol{y^T y} - \boldsymbol{y^T X\beta_{ols}} - \boldsymbol{\beta_{ols}}^T\boldsymbol{X^T y} + \boldsymbol{\beta_{ols}}^T\boldsymbol{X^T X\beta_{ols}}\end{aligned}$$

Where $$\boldsymbol{y^T X \beta_{ols}} = g$$

$$(\boldsymbol{y^T X \beta_{ols}})^T = \boldsymbol{\beta_{ols}}^T \boldsymbol{X^T y} = g^T = g$$

$$g \text{ being some scalar.}$$

Therefore

$$\boldsymbol{e^T e} = \boldsymbol{y^T y} - 2\boldsymbol{\beta_{ols}}^T \boldsymbol{X^T y} + \boldsymbol{\beta_{ols}}^T \boldsymbol{X^T X \beta_{ols}}$$

$$\frac{\partial \boldsymbol{e^T e}}{\partial \boldsymbol{\beta_{ols}^T}} = -2\boldsymbol{X^T y} + \mathbf{2}\boldsymbol{X^T X \beta_{ols}} = 0 \qquad (3)$$

Rearranging

$$\boldsymbol{X^T y} = (\boldsymbol{X^T X})\boldsymbol{\beta_{ols}}$$

$$\boldsymbol{\beta_{ols}} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T y} \qquad (4)$$

The subscript OLS refers to *ordinary least squares.* [150] The previous assumption that the condition of

$$\frac{\partial e^T e}{\partial \boldsymbol{\beta_{ols}}} = 0$$

would provide a solution for *minimised* squared error is true because the second derivative will be positive. Looking at the linear regression equation with the substitution of the relation for $\boldsymbol{\beta_{ols}}$ the equation becomes

$$\boldsymbol{\hat{y}_{OLS}} = \boldsymbol{X\beta_{ols}} = (\boldsymbol{X}(\boldsymbol{X^T X})^{-1}\boldsymbol{X^T})\boldsymbol{y} = \boldsymbol{\Phi_{ols}}\, \boldsymbol{y} \qquad (5)$$

The *Gauss- Markov theorem* [151] follows from the Markov – Gauss assumptions (appendix 3). The Gauss- Markov theorem asserts that if these are to hold, then the

ordinary least squares estimator $\boldsymbol{\beta_{ols}}$ is the best linear, unbiased and efficient estimator, however, this is not the case if violations to the Markov – Gauss assumptions (Appendix 3) occur. The second Markov – Gauss assumption states the $\boldsymbol{X}$ matrix must be of column full rank, this assumption suggests that every column of the predictor dataset will have no relation to its adjacent columns. Experimental spectroscopic data clearly violates the second Markov – Gauss assumption, which is the reason that models different to OLS will be required to accurately generate regression models in the context of spectroscopy. In order to obtain clear mathematical motivation for the introduction of some of the upcoming regression techniques, some analysis of the properties of the OLS regression vector will be explored. The first point of interest is the unbiased nature of the estimator, which is to say there will be no differences between $\boldsymbol{E}[\boldsymbol{\beta_{ols}}]$ and $\boldsymbol{E}[\boldsymbol{\beta}]$. Note as $\boldsymbol{\beta}$ is non-stochastic $\boldsymbol{E}[\boldsymbol{\beta}] = \boldsymbol{\beta}$.

$$\boldsymbol{\beta_{ols}} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T y} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T}(\boldsymbol{X\beta} + \boldsymbol{\epsilon}) = \boldsymbol{\beta} + (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T \epsilon}$$

Taking the expectation value

$$\boldsymbol{E}(\boldsymbol{\beta_{ols}}) = \boldsymbol{E}(\boldsymbol{\beta}) + \boldsymbol{E}((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T \epsilon}) \tag{6}$$

Because $\boldsymbol{X}$ and $\boldsymbol{\beta}$ are non-stochastic and using the third Markov – Gauss assumption (Appendix 3) states $\boldsymbol{E}(\boldsymbol{\epsilon}) = 0$.

$$\boldsymbol{E}(\boldsymbol{\beta_{ols}}) = \boldsymbol{E}(\boldsymbol{\beta}) + \boldsymbol{E}((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T \epsilon}) = \boldsymbol{\beta} + (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T E}(\boldsymbol{\epsilon}) = \boldsymbol{\beta}$$

As such

$$\boldsymbol{E}(\boldsymbol{\beta_{ols}}) = \boldsymbol{\beta}$$

This reflects that the average regression estimator vector, equates to the ideal regression estimator vector. This result will be used to gain insight into how the ideal, $\boldsymbol{\beta}$ and estimated vectors, $\boldsymbol{\beta_{ols}}$ differ.

The next derivation is regarding the variance – covariance matrix of the OLS estimator.

$$Cov(\boldsymbol{\beta_{ols}}) = \boldsymbol{E}\left[\left(\boldsymbol{\beta_{ols}} - \boldsymbol{E}(\boldsymbol{\beta_{ols}})\right)\left(\boldsymbol{\beta_{ols}} - \boldsymbol{E}(\boldsymbol{\beta_{ols}})\right)^{T}\right] \tag{7}$$

From the previous result we see

$$\boldsymbol{\beta_{ols}} - \boldsymbol{E}(\boldsymbol{\beta_{ols}}) = \boldsymbol{\beta_{ols}} - \boldsymbol{\beta} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T}\boldsymbol{\epsilon}$$

So

$$\begin{aligned} Cov(\boldsymbol{\beta_{ols}}) &= \boldsymbol{E}\left[\boldsymbol{\beta_{ols}}\boldsymbol{\beta_{ols}^T}\right] \\ &= \boldsymbol{E}[(\boldsymbol{X^T X})^{-1}\boldsymbol{X^T}\boldsymbol{\epsilon}((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T}\boldsymbol{\epsilon})^{T}] \\ &= \boldsymbol{E}[(\boldsymbol{X^T X})^{-1}\boldsymbol{X^T}\boldsymbol{\epsilon}\boldsymbol{\epsilon^T}\boldsymbol{X}(\boldsymbol{X^T X})^{-1}] \end{aligned}$$

Due to the fact $\boldsymbol{X}$ is non-stochastic.

$$Cov(\boldsymbol{\beta_{ols}}) = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T}\boldsymbol{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon^T})\boldsymbol{X}(\boldsymbol{X^T X})^{-1} \tag{8}$$

Assuming the disturbances, ϵ_i and ϵ_j when $i \neq j$ are independent, but they are homoscedastic.

$$var(\epsilon_i) = var(\epsilon_j)$$

We then have.

$$\boldsymbol{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\boldsymbol{T}}) = \begin{bmatrix} \epsilon_1\epsilon_1 & 0 & \dots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \epsilon_i\epsilon_i \end{bmatrix}$$

$$\epsilon_1\epsilon_1 = \epsilon_i\epsilon_i = variance(\boldsymbol{\epsilon}) = \sigma^2$$

$$Cov(\epsilon_i, \epsilon_j) = 0\ , i \neq j$$

Which is the fourth Markov – Gauss assumption.

$$\boldsymbol{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\boldsymbol{T}}] = \sigma^2\mathbf{I}$$

We then see

$$Cov\ (\boldsymbol{\beta_{ols}}) = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T}(\sigma^2\mathbf{I})\boldsymbol{X}(\boldsymbol{X^T X})^{-1}$$

$$= (\sigma^2\mathbf{I})(\boldsymbol{X^T X})^{-1}\boldsymbol{X^T X}(\boldsymbol{X^T X})^{-1} = \sigma^2(\boldsymbol{X^T X})^{-1}$$

The diagonal entries in $Cov(\boldsymbol{\beta_{ols}})$ are the expected values of the differences between $\boldsymbol{\beta_{ols}}$ and $\boldsymbol{\beta}$ because.

$$\boldsymbol{E}[\boldsymbol{\beta_{ols}}]\ = \boldsymbol{\beta}$$

Defining a vector $\boldsymbol{\beta_\Delta}$ whose components are the diagonal entries of $Cov(\boldsymbol{\beta_{ols}})$ we can then derive an expected value for the magnitude of this vector.

$$\boldsymbol{E}\left[\boldsymbol{\beta_\Delta^T}\boldsymbol{\beta_\Delta}\right] = \sigma^2\boldsymbol{Tr}(\boldsymbol{X^T X})^{-1} \qquad (9)$$

To provide more insight as to the significance of this relationship we utilise an Eigen-decomposition of $\boldsymbol{X^T X}$, achieved as follows.

$$(\boldsymbol{X^T X})\boldsymbol{V} = (\boldsymbol{\lambda}\mathbf{I})\mathbf{V} = \mathbf{V}(\boldsymbol{\lambda}\mathbf{I}) = \boldsymbol{V\Lambda}$$

$$\boldsymbol{X^T X} = \boldsymbol{V\Lambda V^T}$$

Where $\boldsymbol{V}$ is an orthogonal matrix. Using this expression of $\boldsymbol{X^T X}$ and cyclic permeation of the trace, it can be shown that the distance between an estimated regression coefficient vectors, $\boldsymbol{\beta_{ols}}$ and the true vector, $\boldsymbol{\beta}$ depends on the nature of the eigenvalues of $\boldsymbol{X^T X}$.

$$\boldsymbol{Tr}(\boldsymbol{X^T X}) = \boldsymbol{Tr}(\boldsymbol{V\Lambda V^T}) = \sum_i (\boldsymbol{V\Lambda V^T}) = \sum_i \sum_j (\boldsymbol{V})_{ij} \, (\boldsymbol{\Lambda V^T})_{ji}$$

$$= \sum_j \sum_i (\boldsymbol{\Lambda V^T})_{ji} (\boldsymbol{V})_{ij} = \boldsymbol{Tr}(\boldsymbol{\Lambda V^T V}) = \boldsymbol{Tr}(\boldsymbol{\Lambda})$$

Therefore

$$\boldsymbol{Tr}(\boldsymbol{X^T X})^{-1} = \boldsymbol{Tr}(\boldsymbol{\Lambda})^{-1} = \sum_i \frac{1}{\lambda_i} \tag{10}$$

Now the length of the vector defining the difference between estimated and true regression coefficient vectors can be expressed as.

$$\boldsymbol{E}\left[\boldsymbol{\beta_\Delta^T \beta_\Delta}\right] = \sigma^2 \sum_i \frac{1}{\lambda_i} \tag{11}$$

And $\boldsymbol{E}\left[\boldsymbol{\beta_{ols}^T \beta_{ols}}\right]$ may be expressed as.

$$\boldsymbol{E}\left[\boldsymbol{\beta_{ols}^T \beta_{ols}}\right] = \boldsymbol{E}[\boldsymbol{\beta^T \beta}] + \boldsymbol{E}\left[\boldsymbol{\beta_\Delta^T \beta_\Delta}\right] = \boldsymbol{\beta^T \beta} + \sigma^2 \sum_i \frac{1}{\lambda_i} \tag{12}$$

It is a known result [151] that as a matrix deviates away from the unit matrix, towards an ill-conditioned matrix, the spread of eigenvalues deviates away from a single eigenvalue of 1 to a spread of many large and many very small eigenvalues. An ill-conditioned matrix is equivalent to a matrix whose columns are not perfectly linearly-independent.[151] The lack of linear independence of the columns of $\boldsymbol{X^T X}$ can be introduced by linearly dependent columns in $\boldsymbol{X}$. In ATR-FTIR spectroscopy (or spectroscopy in general) this is inevitable, due to the fact neighbouring wavenumber reflectance values are covariant. Small eigenvalues generate large $\boldsymbol{\beta_\Delta}$ in terms of length, which reflects poor estimates of the parameters of the system and is undesirable. Several approaches exist to combat this result. The present research will consider two approaches, the first aims to constrain the growth of the vector $\boldsymbol{\beta_{ols}}$, whilst the second reduces the linearly-dependence of the columns of $\boldsymbol{X^T X}$.

2.2.3 Ridge Regression

Ridge regression[83,84] utilises the first approach to dealing with the undesirable inflation of the $\boldsymbol{\beta_{ols}}$ due to the ill condition of $\boldsymbol{X^T X}$. The derivation of the OLS regression coefficient vector follows from identifying a vector $\boldsymbol{\beta_{ols}}$ which minimises the squared error of the OLS relationship. Ridge regression identifies a vector which minimises the following relationship.

$$\boldsymbol{e^T e} + \lambda \boldsymbol{\beta^T \beta} \tag{13}$$

Where λ is introduced as a parameter to control the length of $\boldsymbol{\beta}$, as such the second term is a penalty based on the Euclidean or l_2-*norm* of $\boldsymbol{\beta}$. The derivation of $\boldsymbol{\beta_{ridge}}$ is as follows, where $\boldsymbol{R}$ represents the Ridge squared error.

$$\boldsymbol{R} = (\boldsymbol{y} - \boldsymbol{X\beta})^{\boldsymbol{T}}(\boldsymbol{y} - \boldsymbol{X\beta}) + \lambda\boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{\beta} \tag{14}$$

$$\frac{\partial \boldsymbol{R}}{\partial \boldsymbol{\beta}^{\boldsymbol{T}}} = \frac{\partial}{\partial \boldsymbol{\beta}^{\boldsymbol{T}}}[\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{y} - \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{X\beta} - \boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y} + \boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X\beta} + \lambda\boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{\beta}]$$

$$= -2\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y} + 2\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X\beta} + 2\lambda\boldsymbol{\beta} = \boldsymbol{0}$$

$$\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y} = \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X\beta} + \lambda\boldsymbol{\beta} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} + \lambda\mathbf{I})\boldsymbol{\beta}$$

$$\boldsymbol{\beta}_{\boldsymbol{ridge}} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} + \lambda\mathbf{I})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y} \tag{15}$$

Whilst this estimator will not experience the same length deformation as the OLS estimator, in the scenario that $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}$ is ill conditioned. The Ridge estimator achieves this at the cost of some bias. The larger λ is, the smaller the Euclidean length of the resulting estimator vector. A simplified interpretation of the process of Ridge regression can be seen in Figure 2.2.1.

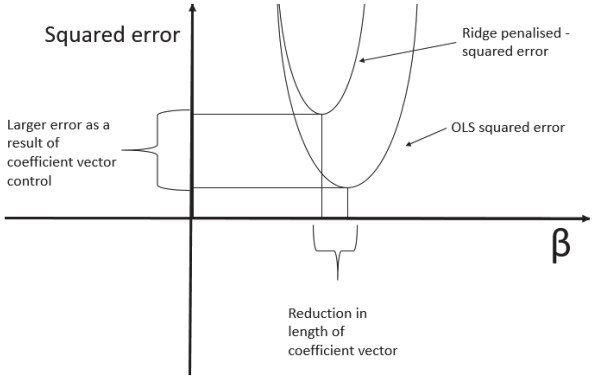

Figure 2.2.1: Simplified example of how the Ridge penalty would influence the squared error function.

In the figure, the OLS error function is depicted as a simple parabola. Figure 2.2.1 shows how the penalty term affects a very simple squared error function. Adding the squared term to the function shifts its minimum value closer to the origin along the $\boldsymbol{\beta}$ axis than the original function, as a consequence of controlling the $\boldsymbol{\beta}$ length, the

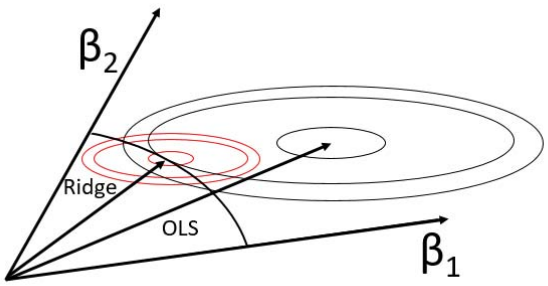

Figure 2.2.2: Illustration of how the Ridge penalty effects elements of $\boldsymbol{\beta}$.

squared error has increased. Figure 2.2.2 shows the projections of the squared error parabolas onto a two dimensional space that represents two elements of the $\boldsymbol{\beta}$ vector. The arc that connects $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ represents another approach to formulating this problem. Rather than finding the minimum value of the Ridge parabola (or the central point of the ridge contours), an equivalent approach is minimisation of squared error subject to the Euclidean length of the $\boldsymbol{\beta}$ vector being restricted to lie within the arc illustrated. This second approach identifies a vector that lies on the edge of the arc and touches the inner most contour of the OLS squared error function. Both approaches yield the same solution, however, they go about it in slightly different ways. It is worth noting at this stage that a down side of Ridge regression is that it is very unlikely to produce *sparse* solutions. A sparse solution is one possessing zero contribution from many of the components. This can be interpreted visually in Figure 2.2.2, because of the smooth nature of the constraint region, there is a far greater likelihood that the Ridge vector will occur at a location not exactly overlapping one of the element vectors or *axes*. Mention has been made of the l_2 – norm, generally an l_n norm is defined by the following equation.

$$|x|_n = \sqrt[n]{\sum |x_i|^n} \qquad (16)$$

$$n \text{ is an integer} > 0$$

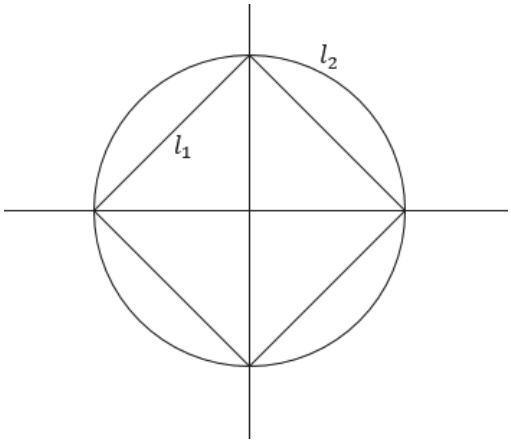

Figure 2.2.3: Difference between l_1 and l_2 norm geometries in two dimensions.

Whilst the variation of n in equation 16 will still generate the same absolute value of x, the various norms differ in their mathematical properties. The difference between the geometry of the l_2 and l_1 norms may be seen in Figure 2.2.3. The key difference between the two is that the l_1-*norm* possesses corners whilst the l_2-*norm* does not, it will become apparent shortly as to why this is important.

2.2.4 Least Absolute Shrinkage and Selection Operator

The second method to incorporate regression vector length control is known as *least absolute shrinkage and selection operator* [85] (LASSO). LASSO controls the length of the regression vector in a fashion very similar to Ridge with the key difference being that the length is accounted for using the l_1-*norm* rather than the l_2-*norm.* This is a subtle but important difference, as will be illustrated. It has been mentioned that Ridge does not successfully induce sparse solutions, as a consequence all elements of the estimator vector are retained in the penalised solution, but they are all scaled to fit. The scaling affects large magnitude elements more than the smaller counter parts, the result of which is a smoothing out of the contributions of the variables and reduced interpretability of the final solution. Alongside reduced interpretability it may also outright reduce the performance of the solution. The LASSO squared error function,

$\boldsymbol{L}$ may be seen below followed by the derivation of the regression coefficient vector $\boldsymbol{\beta_{lasso}}$.

$$\boldsymbol{L} = (\boldsymbol{Y} - \boldsymbol{X\beta})^{\boldsymbol{T}}(\boldsymbol{Y} - \boldsymbol{X\beta}) + \lambda|\boldsymbol{\beta}| \tag{17}$$

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{\beta^T}} = -\boldsymbol{2X^T y} + \boldsymbol{2X^T X\beta} + \frac{\partial(\lambda|\boldsymbol{\beta}|)}{\partial \boldsymbol{\beta^T}}$$

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{\beta^T}} = -\boldsymbol{2X^T y} + \boldsymbol{2X^T X\beta} + \lambda \boldsymbol{w} = \boldsymbol{0}$$

$$\boldsymbol{w} = \begin{Bmatrix} -1 \; if \; b_i < 0 \\ 1 \; if \; b_i > 0 \end{Bmatrix}$$

$$\boldsymbol{\beta_{lasso}} = (\boldsymbol{X^T X})^{\boldsymbol{-1}} (\boldsymbol{X^T y} - \frac{\lambda \boldsymbol{w}}{2}) \tag{18}$$

As λ increases the LASSO estimator vector shrinks, and if

$$|b_i| < \frac{\lambda}{2}$$

the coefficient will shrink to zero, larger coefficients will be more resistant and will be retained longer. It was mentioned earlier that the geometry of the l_1-*norm* is such that it possesses corners. These corners increase the likelihood of a sparse solution, as a contour of the squared error function has a greater chance of touching a corner than a straight edge of the region. Visually this may be seen in the simple example of the Figure 2.2.4.

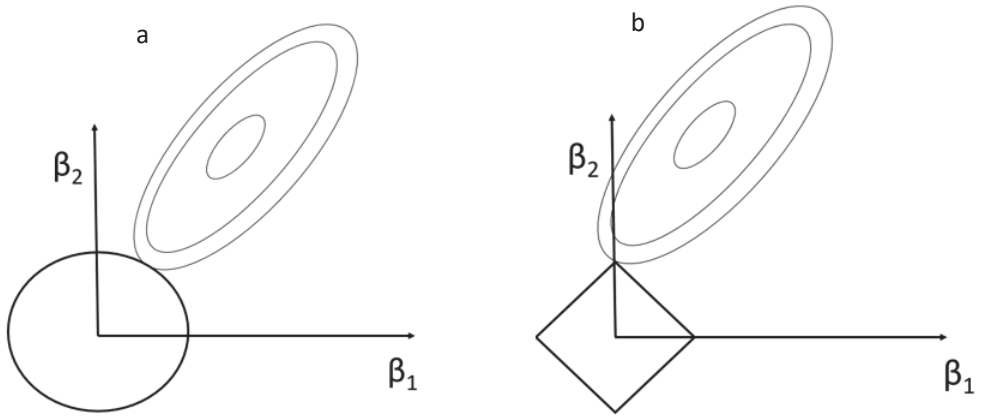

Figure 2.2.4: Influence of norm geometry (a: l_2-norm, b: l_1-norm) upon solution selection, showing greater likelihood of a sparse solution for l_1-norm.

The sparsity of the LASSO estimator is a desirable trait as it allows for the identification and retention of important features of the data set, whilst removing those of little contribution. The results of the LASSO regression method then provide greater insight into the relevant information of the initial data set. However, the LASSO algorithm selects single variables [90] within a highly correlated region of variables, and does so indiscriminately. The indiscriminate nature of the variable selection in LASSO is undesirable as it reduces the insight one can gain upon inspecting the selected variables, and may lead to poor selection within a correlated group.

2.2.5 Elastic Net

Somewhat recently a combination of Ridge regression and LASSO has been suggested, combining the two in a linear fashion. This technique is known as *Elastic Net*[90] and combines LASSO and Ridge introducing a tuning parameter, α which ranges from 0 (Ridge) to 1 (LASSO). Elastic Net boasts the retention of desirable features of both models, incorporating shrinkage, selection and smoother variable selection process than the discrete – indiscriminate process LASSO utilises. The Elastic Net squared error, $\boldsymbol{EN}$ and estimator vector $\boldsymbol{\beta_{elnet}}$ may be seen below.

$$\boldsymbol{EN} = (\boldsymbol{Y} - \boldsymbol{X\beta})^{\boldsymbol{T}}(\boldsymbol{Y} - \boldsymbol{X\beta}) + [(1-\alpha)|\boldsymbol{\beta}| + \alpha\boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{\beta}] \tag{19}$$

$$\boldsymbol{\beta_{elnet}} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} + \alpha\boldsymbol{I})^{-\boldsymbol{1}}\,(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y} - \frac{(1-\alpha)\boldsymbol{w}}{2}) \tag{20}$$

Whilst Elastic Net possesses several desirable properties, there is no guarantee that it will perform the best of the three methods mentioned.

2.2.6 Geometry Background [152]

This section will introduce several concepts which will allow for a geometric interpretation of the feature construction regression methods that will be used in this work, that is principal component analysis and partial least squares regression.

2.2.6.1 Oblique Projections

In general, projections are named by the geometry by which the projection occurs. Such projections are given names such as *oblique* or *orthogonal* to provide this geometrical insight. The PLS projection matrix will be shown to not be symmetric and as such its projections are oblique. The aim of this section is to briefly discuss *oblique projections.* For an oblique projection the line connecting the initial vector (**a**) to its projection (**b**) is not at a right angle (as with orthogonal projections) to the projected vector (dashed line connecting vectors below). As such contained within the projection there needs to be a reference to along which axis the projection occurs.

Figure 2.2.5 illustrates, on the left hand side, a two dimensional example, of the projection of $\boldsymbol{a}$ onto $\boldsymbol{c}$ along $\boldsymbol{e}$**.** Similarly, $\boldsymbol{d}$ is the oblique projection of $\boldsymbol{a}$ onto $\boldsymbol{e}$ along

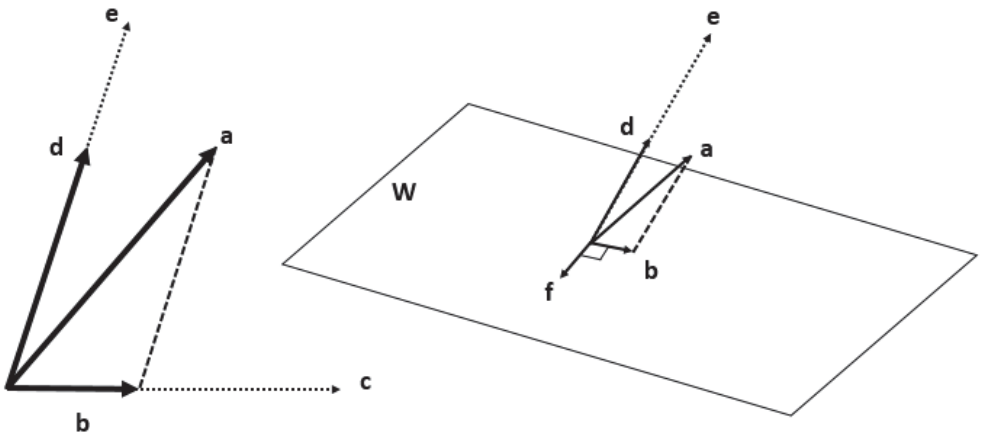

Figure 2.2.5 Examples of oblique projections. Adapted from [152.

$\boldsymbol{c}$. The right hand side depicts the extension into three dimensions, vector $\boldsymbol{b}$ from the two-dimensional example lies in the plane $\boldsymbol{W}$. Note that $\boldsymbol{a}$ cannot project onto $\boldsymbol{f}$, this is a consequence of $\boldsymbol{f}$ not lying in the plane defined by $\boldsymbol{b}$ and $\boldsymbol{d}$, another way of stating this is to say $\boldsymbol{f}$ is linearly independent from vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{d}$. The direction of the projection and the projection subspace should span the entire space, which would be satisfied in the projection of $\boldsymbol{a}$ onto $\boldsymbol{e}$ along the plane $\boldsymbol{W}$. The projection direction may not be contained in the projection subspace, using the current example this would be equivalent to a projection of **a** along **W** onto **b.** In general terms, for a space ξ^n with subspaces γ^m and ρ^p the projection onto γ^m along ρ^p is only possible if the dimensions of the subspaces sum to the dimensions of the space, or rather that the subspaces span the space entirely. They should also be disjoint such that the only shared vector between the spaces is the null vector. Consider the case of a projection onto γ^m along $\rho^{p\perp}$ with $\rho^{p\perp}$ and γ^m disjoint and $m + p = n$. If some matrices $\boldsymbol{A}\ (\boldsymbol{n} \times \boldsymbol{p})$and $\boldsymbol{B}\ (\boldsymbol{n} \times \boldsymbol{p})$ span γ^m and ρ^p respectively then the equation of the projector can be expressed as.[153]

$$\text{H}_{A.B^{\perp}} = A(B^T A)^{-1} B^T \tag{21}$$

This can be verified by first noting that $\text{H}_{A.B^{\perp}}$ is idempotent and hence is a projector. Consequently, any vector in the range of **A** is left unchanged. The columns of $\boldsymbol{B}^{\perp}$ represent the null space of $\boldsymbol{B^T}$, any vector in the subspace $\boldsymbol{B}^{\perp}$ will generate a matrix product $\boldsymbol{B^T B}^{\perp}$, the resulting matrix will have elements that represent products of the row vectors of $\boldsymbol{B^T}$, with the Null vectors of $\boldsymbol{B^T}$, and as such the elements will not exist and the vectors will be annihilated. This result indicates the oblique projection is along $\boldsymbol{B}^{\perp}$. If $\boldsymbol{A}$ and $\boldsymbol{B}$ span the same subspace, γ^m the projector equation becomes.

$$\mathrm{H}_{A.A^{\perp}} = A(A^T A)^{-1} A^T = \mathrm{H}_A \tag{22}$$

Which is equivalent to an orthogonal projection onto the subspace γ^{m}.

2.2.6.2 Geometry of Ellipsoids

If $\boldsymbol{A}$ $(p \times p)$ is a symmetric matrix of rank p, then the set of p vectors $\boldsymbol{z}$ satisfying.

$$\boldsymbol{z^T A z} = constant = c$$

Which is an expression stating that the vector formed by $\boldsymbol{Az}$ alters the inner product $\boldsymbol{z^T z}$ in a constant fashion. The result of this process is the generation of an ellipsoidal hypersurface in p −dimensional space. Ellipses and ellipsoids ($p \geq 3$) defined by the above equation will be denoted by, $\breve{\mathrm{A}}_c^-$.

$$\breve{\mathrm{A}}_c^- \equiv \{\boldsymbol{z} | \boldsymbol{z}^T \boldsymbol{A z} = c\} \tag{23}$$

The shape of the ellipsoid is determined by the relative lengths of its principal axes and it is easy to show these are related to the eigenvalues of $\boldsymbol{A}$ as follows.

$$\boldsymbol{Av} = \lambda \boldsymbol{v} = \boldsymbol{v} \lambda$$

$$\boldsymbol{AV} = \boldsymbol{V \Lambda}$$

$\boldsymbol{V}$ is a matrix of the eigenvectors of $\boldsymbol{A}$ and is orthogonal, $\boldsymbol{VV^T} = \mathbf{I}$

Which means

$$\boldsymbol{AVV^T} = \boldsymbol{V \Lambda V^T} = \boldsymbol{A}$$

The ellipse equation then becomes.

$$(\boldsymbol{z^T V})\boldsymbol{\Lambda}(\boldsymbol{V^T z}) = \boldsymbol{z'^T \Lambda z'} = c$$

As an example if $\boldsymbol{\Lambda}$ was a 2×2 matrix and $c = 1$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

$$\boldsymbol{z'^T \Lambda z'} = c$$

$$= [z_1' \quad z'_2] \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & \lambda_2 \end{bmatrix} \begin{bmatrix} z_1' \\ z_2' \end{bmatrix}$$

$$= [z_1' \quad z'_2] \begin{bmatrix} \lambda_1 z_1' \\ \lambda_2 z_2' \end{bmatrix}$$

$$= z_1' \lambda_1 z_1' + z'_2 \lambda_2 z_2'$$

$$= 1$$

If we take z_1' to be x and z_2' to be y.

$$\boldsymbol{z'^T \Lambda z'} = \lambda_1 x^2 + \lambda_2 y^2 = 1$$

Which is the equation of an ellipse with principal axes equal in length to

$\frac{1}{\sqrt{\lambda_1}}$ and $\frac{1}{\sqrt{\lambda_2}}$.

which shows that in general for an $i \times i$ diagonalisable matrix, $\boldsymbol{A}$ the principal axes will have lengths $\left(\frac{c}{\lambda_i}\right)^{\frac{1}{2}}$.

For any exponent m, the relationship is

$$\boldsymbol{A^m = V \Lambda^m V^T}$$

$$\check{\mathrm{A}}_{\boldsymbol{c}}^{-\boldsymbol{sign(m)}} \equiv \{\, \boldsymbol{z} | (\boldsymbol{z^T V})\, \boldsymbol{\Lambda^m} (\, \boldsymbol{V^T z}) \; = c \,\} \tag{24}$$

Equation 24 illustrates nicely the fact that the eigenvectors do not change upon the consideration of different m. For example $m = -1$ describes the inverse of $\boldsymbol{A}$, which generates an ellipsoid, $\check{\mathrm{A}}_{\boldsymbol{c}}$ with equivalent eigenvectors, however, now the eigenvalues are inverted in the diagonal of $\boldsymbol{\Lambda^{-1}}$**.** For a simple two dimensional example, with $c = 1$ this amounts to a rotation and scaling of the ellipse and can be seen in Figure 2.2.6.

There is a preference to work with $m = -1$ because the length of the principal axes of the resulting ellipse are now proportional to the size of the eigenvalues of $\boldsymbol{A}$. However, as has been shown above this is not the case for $m = 1$. The motivation for

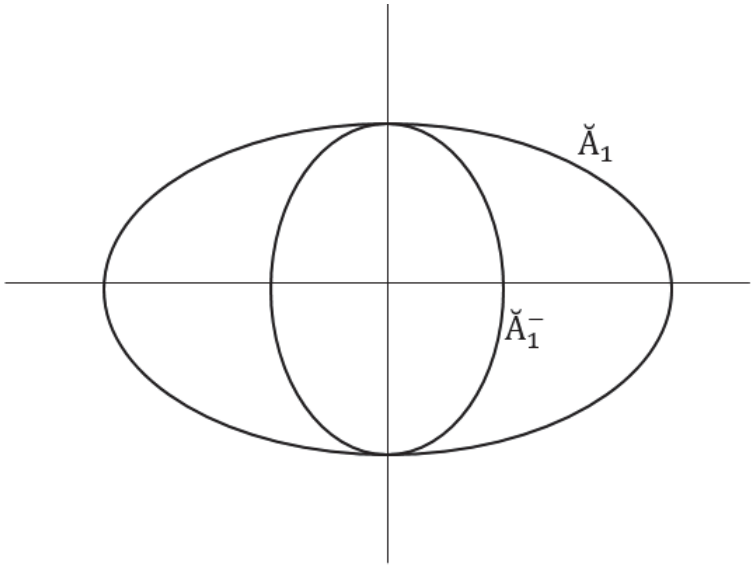

Figure 2.2.6: Example of the effects of eigenvalue inversion. Adapted from [152].

discussing ellipse geometry, is that it will provide a means for geometrical intuition in the upcoming discussion of various feature construction methods.

2.2.6.3 Planes of Tangency

The proceeding discussion will involve the ellipsoid $\check{A}_{\boldsymbol{c}}$. We will build towards identifying the nature of tangent rotations beginning first with a discussion of the algebraic operations of 25 and 26.

$$\boldsymbol{Aa} = \boldsymbol{b} \tag{25}$$

$$\boldsymbol{a} = \boldsymbol{A^{-1}b} \tag{26}$$

A being a $p \times p$ matrix, $\boldsymbol{b}$ and $\boldsymbol{a}$ are both p vectors. We begin with the assumption $\boldsymbol{b}$ lies on the surface of $\check{A}_{\boldsymbol{c}}$, such that we can write.

$$\boldsymbol{b^T A^{-1} b} = c \tag{27}$$

Consider now a selection of vectors equivalent to $\boldsymbol{b} + \boldsymbol{\delta}$ as $\boldsymbol{\delta} \to \boldsymbol{0}$ the vectors will describe an infinitesimally small element around the set of vectors $\boldsymbol{b}$. This region may be regarded as a surface element around $\boldsymbol{b}$ on $\check{A}_{\boldsymbol{c}}$. As such these vectors satisfy

$$(\boldsymbol{b} + \boldsymbol{\delta})^T \boldsymbol{A^{-1}} (\boldsymbol{b} + \boldsymbol{\delta}) = c$$

$$= (\boldsymbol{\delta^T} + \boldsymbol{b^T})(\boldsymbol{A^{-1}b} + \boldsymbol{A^{-1}\delta})$$

$$= \boldsymbol{\delta^T A^{-1} b} + \boldsymbol{\delta^T A^{-1} \delta} + \boldsymbol{b^T A^{-1} b} + \boldsymbol{b^T A^{-1} \delta}$$

$$= \boldsymbol{b^T A^{-1} b}$$

$$= c$$

Discarding second order term in $\boldsymbol{\delta}$ and simplifying

$$\boldsymbol{\delta^T A^{-1} b + b^T A^{-1} \delta = 2 * \delta^T A^{-1} b = 0}$$

$$\boldsymbol{\delta^T A^{-1} b = 0} \tag{28}$$

Remembering that the above is the inner product of $\boldsymbol{A^{-1}b}$ and $\boldsymbol{\delta}$ it can be quickly recognised they must be orthogonal to one another. So the set of vectors $\boldsymbol{\delta}$ as $\boldsymbol{\delta} \to 0$ constitute a surface element around the origin of which the direction of the normal is $\boldsymbol{A^{-1}b}$. Because $\boldsymbol{A^{-1}b}$ is orthogonal to $\boldsymbol{\delta}$ the addition of $\boldsymbol{b}$ simply translates the element to the surface of the ellipsoid without altering the direction, as the ellipsoid has been shown to be unaltered. The implication of this is that the hyperplane tangent to the ellipsoid at $\boldsymbol{b}$ has $\boldsymbol{a = A^{-1}b}$ as it's normal. We can now interpret

$$\boldsymbol{b = Aa}$$

as follows, given $\boldsymbol{A}$ $(p \times p)$ and $\boldsymbol{a}$ $(p \times 1)$, the point of tangency between the ellipsoid $Ă_{\boldsymbol{c}}$ and the $(p - 1)$ dimensional hyperplane that is orthogonal to $\boldsymbol{a}$ gives a vector that is a scalar multiple of $\boldsymbol{Aa}$ (where $\boldsymbol{Aa = b}$). A transformation of this type is known as a tangent rotation of the vector $\boldsymbol{a}$ on the ellipsoid $Ă_{\boldsymbol{c}}$.

As a simple example consider the scenario depicted in figure 2.2.7. Where $p = 2$, the vector $\boldsymbol{a}$ is shown alongside a family of perpendicular lines, these lines span the space of $\boldsymbol{\delta}$. The member of the orthogonal vector family that is tangent to the ellipse $Ă_{\boldsymbol{c}}$ will be equal to $k\boldsymbol{Aa}$ at the point of tangency. Equivalently it will be a scalar multiple of the vector $\boldsymbol{b}$ of which spans the surface of the ellipse. This approach can be extended to higher dimensions but quickly becomes difficult to visualise. *Reverse Tangent Rotations* can be carried out by constructing a line tangent to the ellipse at the point where $k\boldsymbol{b}$ intersects the ellipse. Then the vector that is orthogonal to this tangent line

will be a scalar multiple of the vector previously identified as being tangent to $\boldsymbol{\delta}$ namely $\boldsymbol{A^{-1}b}$.

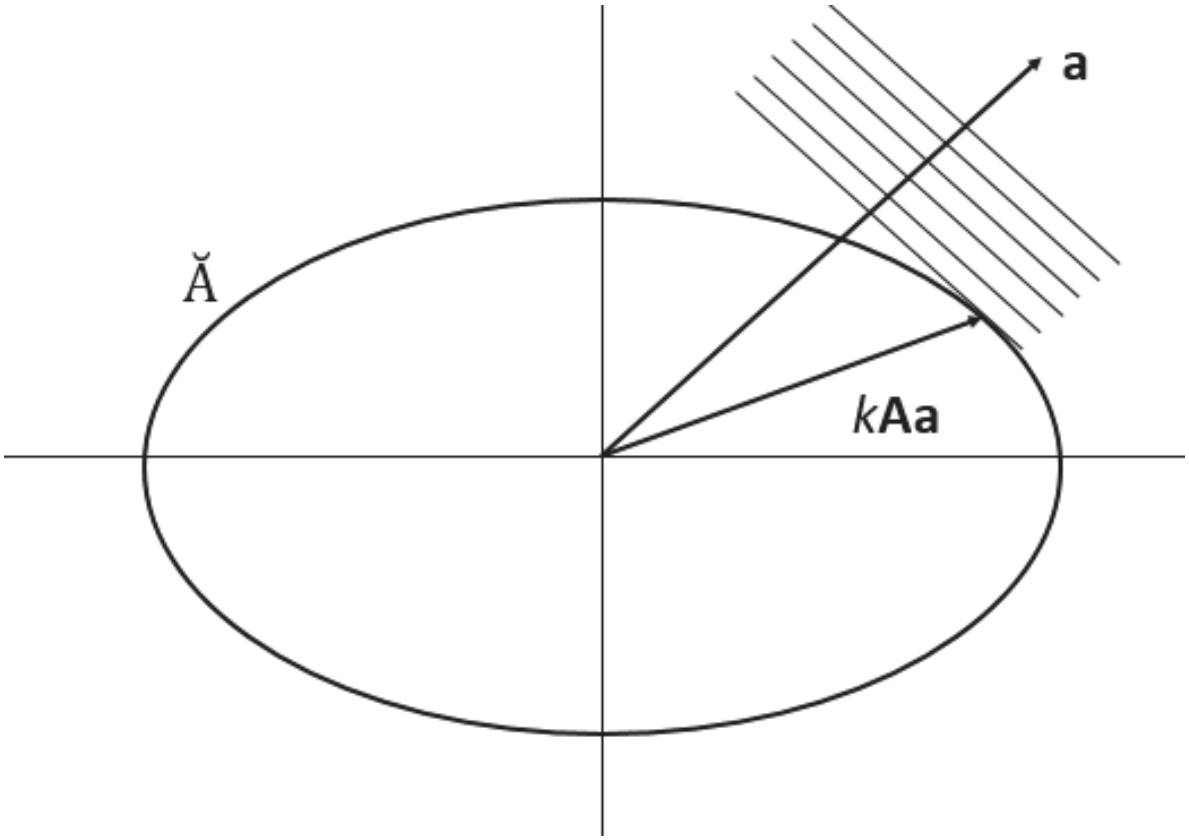

Figure 2.2.7: The tangent rotation of a vector $\boldsymbol{a}$. Adapted from [152].

2.2.6.4 The Power Method [154]

The power method is an approach to finding eigenvalues and eigenvectors for a symmetric matrix, it is based upon an iterative tangent rotation of a vector. The process itself starts with a vector $\boldsymbol{z^0}$, which will then be tangentially rotated onto the surface of the above ellipse. The process appears as follows

$$\boldsymbol{z^1} = k_1\boldsymbol{A z^0}$$

$$\boldsymbol{z^2} = k_2\boldsymbol{A^2 z^0} = k_2'\boldsymbol{A z^1}$$

$$\boldsymbol{z^3} = k_3\boldsymbol{A^3 z^0} = k_3'\boldsymbol{A z^2}$$

$$\vdots$$

$$\boldsymbol{z^i} = k_i\boldsymbol{A^i z^0} = k_i'\boldsymbol{A z^{i-1}}$$

In the context of this chapter it's sufficient to select scalars k and k' such that

$$\boldsymbol{z^{(i)T} A^{-1} z^{(i)}} = 1$$

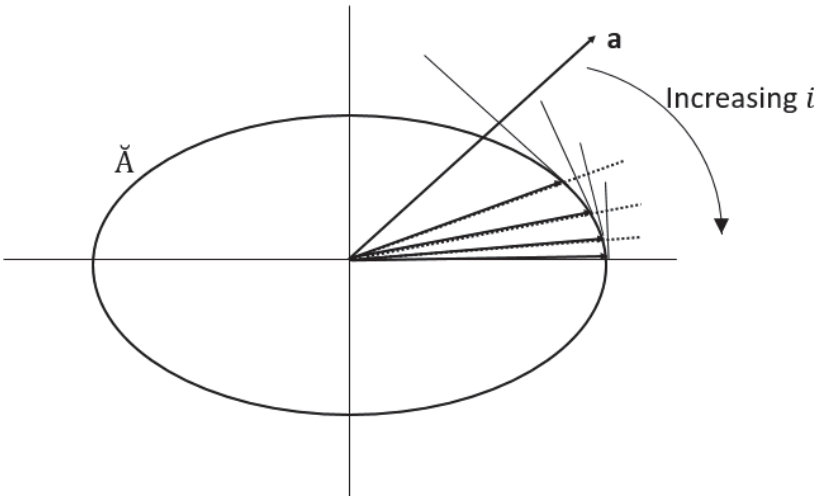

Figure 2.2.8: The iterative power method illustrated with a selection of vectors calculated as tangent rotations of the previous. Adapted from [152].

The sequence will converge on the first Eigen pair $(\lambda_1, \boldsymbol{v_1})$ and Figure 2.2.8 illustrates why. In order to identify the second eigenvector, we need to change the direction of vector generation, this is achieved using the reverse tangent rotation that has been discussed previously. To find subsequent eigenvectors and eigenvalues it is necessary to deflate the space as follows

$$A^i = A - \sum_{j=1}^{i} \lambda_j \boldsymbol{v_j} \boldsymbol{v_j^T}, i = 1,2, \dots, p-1$$

Geometrically deflation, after the i eigenvectors are obtained, can be visualized as *slicing* a $(p-(i+1))$ −dimensional ellipsoid along the $(p-i)$ −dimensional hyperplane orthogonal to the i^{th} eigenvector. Deflation is used in PCR and PLS to remove the constructed features from the original matrix.

2.2.7 Principal Component Regression

Principal Component Regression [155] (PCR) is the first technique (of two) we will look at to resolve the ill condition issues (discussed at the end of section 2.2.1), by restructuring the $\boldsymbol{X^T X}$ matrix. Principal component regression generates principal components by identifying the eigenvectors of the $\boldsymbol{X^T X}$ matrix (variance covariance matrix) and pre-multiplying these eigenvectors by $\boldsymbol{X}$. A principal component is effectively the projection of the original data set onto the eigenvectors of the covariance-variance matrix. As such these principal components capture the variations of the original data set. This technique reduces the ill condition of the original $\boldsymbol{X^T X}$ matrix due to the fact each principal component is a projection onto a different and orthogonal eigenvector.

$$\boldsymbol{PC = X\gamma_i = p_i} \tag{29}$$

The principal components are used as the predictor variables. The dimensions of the principal component matrix, $\boldsymbol{\mathcal{P}}$ will be $n \times k$ where k is the number of eigenvectors $\boldsymbol{\gamma}$ which are extracted for projection. The eigenvectors are selected based on the corresponding eigenvalues, selecting them based on eigenvalue size, as this reflects the degree of *variance* along the direction of the eigenvector. Because eigenvectors are chosen in this fashion the result is $k < p$ and the eigenvalues are fewer in number with the smallest $\lambda_k > \lambda_p$ and typically $\lambda_k \geq 1$. Constructing the predictor matrix in this fashion avoids small eigenvalues and hence dramatic estimator vector growth. The predicted response values can be expressed based on their orthogonal projections onto the principal components analogous to the OLS result.

$$\hat{\boldsymbol{y}} = \boldsymbol{\mathcal{P}}(\boldsymbol{\mathcal{P}}^{\boldsymbol{T}}\boldsymbol{\mathcal{P}})^{-\mathbf{1}}\boldsymbol{\mathcal{P}}^{\boldsymbol{T}}\boldsymbol{y} \tag{30}$$

$$\boldsymbol{\mathcal{P}}^{\boldsymbol{T}}\boldsymbol{\mathcal{P}} = \boldsymbol{\Gamma}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^{\boldsymbol{T}}(\boldsymbol{\lambda_k}\boldsymbol{\Gamma}) = \boldsymbol{\lambda}\boldsymbol{\Gamma}^{\boldsymbol{T}}\boldsymbol{\Gamma} = \boldsymbol{\lambda}$$

$$\boldsymbol{\lambda_k} = Diag\{\lambda_1, \lambda_2, \dots, \lambda_k\}$$

$$\lambda_1 > \lambda_2 > \lambda_k$$

$$\therefore \hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\Gamma}(\boldsymbol{\lambda_k})^{-\mathbf{1}}\boldsymbol{\Gamma}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}$$

$$\boldsymbol{\Gamma} = [\gamma_1, \gamma_2, \dots \gamma_k\ ]$$

$$\boldsymbol{\beta_{pcr}} = \boldsymbol{\Gamma}(\boldsymbol{\lambda_k})^{-\mathbf{1}}\boldsymbol{\Gamma}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y} \tag{31}$$

Inserting $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}$ into equation 31 we can derive a relation between $\boldsymbol{\beta_{pcr}}$ and $\boldsymbol{\beta_{ols}}$ as follows.

$$\boldsymbol{\beta_{pcr}} = \boldsymbol{\Gamma}(\boldsymbol{\lambda_k})^{-\mathbf{1}}\boldsymbol{\Gamma}^{\boldsymbol{T}}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\ )\ \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}$$

$$= \boldsymbol{\Gamma}(\boldsymbol{\lambda_k})^{-\mathbf{1}}\boldsymbol{\Gamma}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{\beta_{ols}}$$

$$= \boldsymbol{\Gamma}(\boldsymbol{\lambda_k})^{-\mathbf{1}}\boldsymbol{\Gamma}^{\boldsymbol{T}}(\boldsymbol{\Gamma}\boldsymbol{\lambda_k}\boldsymbol{\Gamma}^{\mathbf{T}})\boldsymbol{\beta_{ols}}$$

$$= \boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\boldsymbol{T}}\boldsymbol{\beta_{ols}}$$

Since $\boldsymbol{\Gamma}^{\boldsymbol{T}}\ \boldsymbol{\Gamma}\ =\ \mathbf{I}$ we can write the final equality as

$$\boldsymbol{\beta_{pcr}} = \boldsymbol{\Gamma}(\boldsymbol{\Gamma}^{\boldsymbol{T}}\boldsymbol{\Gamma})^{-\mathbf{1}}\boldsymbol{\Gamma}^{\boldsymbol{T}}\boldsymbol{\beta_{ols}} \tag{32}$$

Which is nothing more than an orthogonal projection of the OLS result onto the subspace spanned by the principal components.

This method results in

$$cov(\boldsymbol{\beta_{pcr}}) = \sigma^2(\boldsymbol{\mathcal{P}^T\mathcal{P}})^{-1} \tag{33}$$

And

$$\boldsymbol{E(\beta_{\Delta}^T\beta_{\Delta})} = \sigma^2\boldsymbol{Tr}(\boldsymbol{\mathcal{P}^T\mathcal{P}}) = \sigma^2\sum_{i=1}^{k}\frac{1}{\lambda_i} \tag{34}$$

This vector is far more stable now as a result of the following inequality.

$$\sum_{i=1}^{k}\frac{1}{\lambda_i} < \sum_{i=1}^{p}\frac{1}{\lambda_i} \tag{35}$$

Hence the expected size of the difference vector is smaller.

2.2.8 Principal Component Regression Geometry

The geometry of PCR[152] is fairly straightforward, as we now know the technique works by projecting observation vectors from $\boldsymbol{X}$ onto the eigenvectors of $\boldsymbol{X^T X}$. Each principal component may also be expressed in terms of the eigenvectors of $\boldsymbol{XX^T}$ as follows. The singular value decomposition of **X** is

$$\boldsymbol{X = U\Sigma V^T}$$

Thus each principal component may be expanded, remembering that the right singular vector, $\boldsymbol{v}$ is equivalent to the $\boldsymbol{X^T X}$ eigenvector, $\boldsymbol{\gamma}$.

$$\boldsymbol{X\gamma = Xv = u\sigma v^T v = u\sigma}$$

The left singular vector of $\boldsymbol{X}$, $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{XX^T}$. The square – symmetric matrix $\boldsymbol{XX^T}$ may be used to generate an ellipsoid, $\mathbb{D}$ governed by the following equation.

$$\boldsymbol{z^T XX^T z = z^T Dz = c}$$

Figure 2.2.9 illustrates the decomposition of variable vectors $\boldsymbol{v}$, onto the unit norm eigenvectors $\boldsymbol{u}$ in the process of constructing the principal components $\boldsymbol{P}$. It has already been identified that the PCR response prediction vector may be found by projecting orthogonally the OLS response prediction vector onto the hyperplane constructed with the *principal components*, this may be seen in Figure 2.2.10.

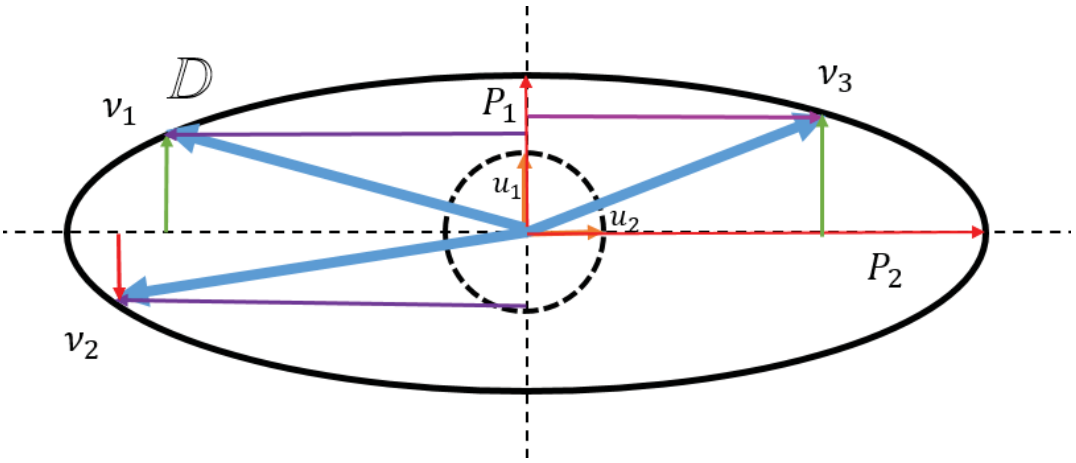

Figure 2.2.9: Geometric relation between variable vectors v, principal components P and the ellipse $\mathbb{D}$ in two dimensions. Adapted from [152].

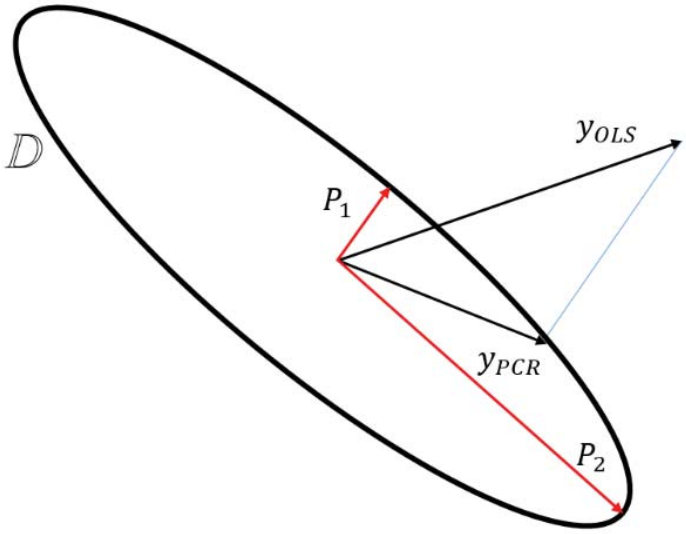

Figure 2.2.10: Orthogonal projection of OLS response estimations when two principal components are utilised. Adapted from [152].

2.2.9 Partial Least Squares

Partial least squares[104,156] implicitly links $\boldsymbol{Y}$ and $\boldsymbol{X}$ by identifying a set of new variables which capture the important variance and covariance between $\boldsymbol{X}$ and $\boldsymbol{Y}$. The discussion will continue for the time being, considering a matrix of responses, $\boldsymbol{Y}$. The initial step of the process involves extracting the first column of $\boldsymbol{Y}$, which will be denoted as $\boldsymbol{u}$, the elements of which contain responses for every n observation of $\boldsymbol{X}$ (or every row). The next stage of the process is to identify another column vector, $\boldsymbol{w}$ which will shed light on how to weight the variables of $\boldsymbol{X}$.

$$\boldsymbol{w} = \frac{\boldsymbol{X}^T\boldsymbol{u}}{\boldsymbol{u}^T\boldsymbol{u}} \tag{36}$$

The elements of $\boldsymbol{w}$ are large if the sums $\sum_i^n \boldsymbol{X}_{\boldsymbol{pi}}^{\boldsymbol{T}}\boldsymbol{u}_{\boldsymbol{i}}$ representing them are large. Each element of $\boldsymbol{w}$ is equal to the inner product of a column vector of $\boldsymbol{X}$ and the vector $\boldsymbol{u}$, where the inner product is then scaled based on the length of $\boldsymbol{u}$. After obtaining the weight vector, it is scaled to unit norm. This weights vector is used to generate a weighted version, of the observations in $\boldsymbol{X}$ as follows.

$$\boldsymbol{t} = \frac{\boldsymbol{X}\boldsymbol{w}}{\boldsymbol{w}^T\boldsymbol{w}} = \boldsymbol{X}\boldsymbol{w}; \tag{37}$$

$$\text{as } \boldsymbol{w}^T\boldsymbol{w} = 1 \text{ due to unit norm scaling}$$

Where $\boldsymbol{t}$ is a column vector containing elements which represent a weighted form of the information contained within each observation of $\boldsymbol{X}$ **.**

Since **w** is a unit norm vector, each element of $\boldsymbol{t}$ may be thought of as the *w*-component of the corresponding observation or row of $\boldsymbol{X}$. The *w*-components capture the most important information in $\boldsymbol{X}$. Similarly for $\boldsymbol{Y}$, we define a column vector $\boldsymbol{c}$ such that.

$$\boldsymbol{c} = \frac{\boldsymbol{Y}^T\boldsymbol{t}}{\boldsymbol{t}^T\boldsymbol{t}} \tag{38}$$

The vector $\boldsymbol{c}$ is also scaled to be unit norm. Now, $\boldsymbol{u}$ is recalculated using $\boldsymbol{c}$.

$$\boldsymbol{u} = \frac{\boldsymbol{Y}\boldsymbol{c}}{\boldsymbol{c}^T\boldsymbol{c}} = \boldsymbol{Y}\boldsymbol{c} \tag{39}$$

This equation has an equivalent interpretation as that for $\boldsymbol{t}$, in that it extracts the *c*-component of the observations in $\boldsymbol{Y}$. At this stage of the process we have identified a new vector $\boldsymbol{u}$ which can be used to recalculate the vectors that have been defined thus far, this iterative process repeats until the following convergence condition is achieved.

$$\boldsymbol{u}_{n+1} \cong \boldsymbol{u}_n$$

The next portion of the algorithm deals with identifying the appropriate loadings for both the $\boldsymbol{X}$ and $\boldsymbol{Y}$ data sets. The vectors $\boldsymbol{t}$ and $\boldsymbol{u}$ are referred to as *scores* vectors, we now define the *loadings* vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ for $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively. The loading vectors will provide a scaled summary of how well each column of $\boldsymbol{X}$ or $\boldsymbol{Y}$ aligns with $\boldsymbol{t}$ or $\boldsymbol{u}$.

$$\boldsymbol{p} = \frac{\boldsymbol{X}^T\boldsymbol{t}}{\boldsymbol{t}^T\boldsymbol{t}} \tag{40}$$

$$\boldsymbol{q} = \frac{\boldsymbol{Y}^T\boldsymbol{u}}{\boldsymbol{u}^T\boldsymbol{u}} \tag{41}$$

We can now calculate residual matrices by removing a matrix constructed from the relevant scores and loadings vectors.

$$\boldsymbol{X_r = X - tp^T} \tag{42}$$

$$\boldsymbol{Y_r = Y - tc^T} \tag{43}$$

The relations above for formulating residual matrices are commonly referred to as deflations (akin to those defined in 2.2.6.4). The inclusion of the $\boldsymbol{X}$ scores vector, $\boldsymbol{t}$ in the deflation process for $\boldsymbol{Y}$ is because $\boldsymbol{Y}$ is being decomposed based on its relation to $\boldsymbol{X}$. The result of this style of deflation is that the most relevant information is removed at each cycle of the algorithm. The process then begins once more, calculating the relevant vectors using these residual matrices. The iterations stop when some predetermined criteria is met. From the relationships outlined it is possible to derive eigenvalue equations (see appendix 4). The eigenvalue equations at convergence (vectors between iterations vary by a negligible amount) are.

$$\boldsymbol{YY^T XX^T\, u} = a\boldsymbol{u} \tag{44}$$

$$\boldsymbol{Y^T XX^T Y\, c} = a\boldsymbol{c} \tag{45}$$

$$\boldsymbol{XX^T YY^T t} = a\boldsymbol{t} \tag{46}$$

$$\boldsymbol{X^T YY^T Xw} = a\boldsymbol{w} \tag{47}$$

Where $a = (\boldsymbol{c_n^T c_n})(\boldsymbol{t_n^T t_n})(\boldsymbol{w_n^T w_n})(\boldsymbol{u_n^T u_n})$

This iterative approach is analogous to the power method, due to this relationship we observe that a is in fact the maximum eigenvalue for the above eigenvalue equations. The vectors $\boldsymbol{u}, \boldsymbol{c}, \boldsymbol{t}$ and $\boldsymbol{w}$ are the eigenvectors for the corresponding matrix products on the left hand side. Thus the iterative process discussed so far is nothing more than

the identification of the above eigenvectors at each stage of deflation, and then using them to further deflate the matrices. It turns out that in fact for computations it is sufficient to identify the eigenvector $\boldsymbol{w}$ and the other vectors can be calculated from it. The approach of identifying the eigenvector $\boldsymbol{w}$ as the starting point for the process, is used in an algorithm known as the PLSR *Kernel*[105,157-159] algorithm.

The matrix $\boldsymbol{X^T Y Y^T X}$ may be thought of as a weighted covariance-variance matrix for $\boldsymbol{X}$. The weighting is implemented by the insertion of $\boldsymbol{YY^T}$ between $\boldsymbol{X^T X}$. Where $\boldsymbol{YY^T}$ is a matrix containing information about the size of the row vectors of $\boldsymbol{Y}$ on the diagonal, and information about the alignment of different row vectors in the remaining elements. So the vector $\boldsymbol{w}$ may be thought of as an eigenvector of a covariance – variance matrix capturing covariance between $\boldsymbol{X}$ and $\boldsymbol{Y}$ matrices. The scores vectors $\boldsymbol{t}$ are the weighted form of the principal components identified in PCR.

$$\boldsymbol{Y} = \begin{Bmatrix} \boldsymbol{y_1} \\ \vdots \\ \boldsymbol{y_n} \end{Bmatrix}$$

$$\boldsymbol{y_1 \rightarrow y_n \; being \; the \; observation \; or \; row \; vectors \; of \; Y}$$

$$\boldsymbol{YY^T} = \begin{bmatrix} |\boldsymbol{y_1}|^2 & \cdots & \boldsymbol{y_1 y_n^T} \\ \vdots & \ddots & \vdots \\ \boldsymbol{y_n y_1^T} & \cdots & |\boldsymbol{y_n}|^2 \end{bmatrix}$$

2.2.9.1 Residual Matrices

The residual matrices and the $\boldsymbol{X_{i+1}^T X_{i+1}}$ residual product can be written in the following way. For the full set of residual matrix products, see Appendix 6, products of these matrices are involved in the convergence eigenvalue equations mentioned previously.

$$\boldsymbol{X_{i+1} = X_i - t_i p_i^T}$$

$$\boldsymbol{Y_{i+1} = Y_i - t_i c_i^T}$$

$$\boldsymbol{X_{i+1}^T X_{i+1} = \left(X_i - t_i p_i^T\right)^T \left(X_i - t_i p_i^T\right)}$$

$$\boldsymbol{= X_i^T X_i - X_i^T t_i p_i^T - p_i t_i^T X_i + p_i \left(t_i^T t_i\right) p_i^T}$$

where

$$\boldsymbol{t_i^T X_i = \left(t_i^T t_i\right) p_i^T;}$$

therefore

$$\boldsymbol{X_{i+1}^T X_{i+1} = X_i^T X_i - X_i^T t_i p_i^T - p_i \left(t_i^T t_i\right) p_i^T + p_i \left(t_i^T t_i\right) p_i^T}$$

$$\boldsymbol{= X_i^T X_i - X_i^T t_i p_i^T}$$

$$\boldsymbol{= X_i^T X_i - \left(t_i^T X_i\right)^T p_i^T = X_i^T X_i - p_i \left(t_i^T t_i\right) p_i^T}$$

$$\boldsymbol{= X_i^T X_i - p_i p_i^T \left(t_i^T t_i\right)} \qquad (48)$$

Appendix 6 shows that the product of the deflation of both $\boldsymbol{X}$ and $\boldsymbol{Y}$ residuals is equivalent to the product of the deflation of just one. At this point in time we have identified the structure of the new predictor covariance – variance matrix, which we

may denote as $\boldsymbol{X_C^T X_C}$ the $\boldsymbol{C}$ alludes to the fact this information has been constructed from the original full predictor covariance variance matrix. The construction has already been expressed above and can be rewritten as follows.

$$\boldsymbol{X^T X = X_C^T X_C + X_R^T X_R} \tag{49}$$

The subscript R will refer to the residual matrix. The original discussion regarding these regression methods began by identifying the issues present in the OLS estimate as a consequence of the ill-conditioned nature of $\boldsymbol{X^T X}$. The fact that $\boldsymbol{X^T X}$ was ill conditioned lead to it possessing many large and more importantly many very small eigenvalues,[151] which we identified as leading to undesirable growth of the vector $\boldsymbol{\beta_\Delta}$. Using equation 49 it can be shown the undesirable growth is reduced.

$$\boldsymbol{X^T X = X_C^T X_C + X_R^T X_R}$$

$$\boldsymbol{tr(X^T X) = tr\left(X_C^T X_C + X_R^T X_R\right) = tr(X_C^T X_C) + tr(X_R^T X_R)}$$

$$\boldsymbol{tr(X^T X) - tr\left(X_R^T X_R\right) = tr(X_C^T X_C)}$$

$$\boldsymbol{tr(\Gamma \Sigma \Gamma^T) - tr\left(\Gamma_R \Sigma_R \Gamma_R^T\right) = tr(\Gamma_C \Sigma_C \Gamma_C^T)}$$

$$\sum_i \left(\sum_j (\boldsymbol{\Gamma^T})_{ij} (\boldsymbol{\Gamma \Sigma})_{ji} \right) - \sum_i \left(\sum_j \left(\boldsymbol{\Gamma_R^T}\right)_{ij} (\boldsymbol{\Gamma_R \Sigma_R})_{ji} \right) = \sum_i \left(\sum_j \left(\boldsymbol{\Gamma_C^T}\right)_{ij} (\boldsymbol{\Gamma_C \Sigma_C})_{ji} \right)$$

$$\boldsymbol{tr(\Sigma - \Sigma_R) = tr(\Sigma_C)} \tag{50}$$

$$\text{if } \lambda_{Ri} > 0$$

$$\sum_{i=1}^{p} \frac{1}{\lambda_i} > \sum_{i=1}^{p} \frac{1}{\lambda_{C,i}}$$

$$|\boldsymbol{\beta_{OLS}} - \boldsymbol{\beta}| > |\boldsymbol{\beta_{pls}} - \boldsymbol{\beta}| \tag{51}$$

Which demonstrates the PLS regression vector is closer to the true parameter estimator vector than the OLS counterpart.

2.2.9.2 Properties of the Partial Least Squares Algorithm

An interesting feature of the PLS algorithm is that whilst it began as purely an algorithm, several mathematical and geometrical interpretations have since been identified. We have mentioned briefly before the weighted covariance interpretation for finding $\boldsymbol{w}$. We will now build towards another interpretation beginning first by considering a rotation denoted as a right multiplication of matrix, $\boldsymbol{X}$ by a rotation matrix $\boldsymbol{O_x}$ whereby $\boldsymbol{O_x O_x^T} = \mathbf{I}$ so the rotation is orthogonal.

Define

$$\boldsymbol{S} = \boldsymbol{X O_x} = \{\boldsymbol{s_1}, \boldsymbol{s_2}, \dots, \boldsymbol{s_k}\} \tag{52}$$

With

$$\sum_i s_i^2 = Tr(S^T S) = Tr(O_x^T X^T X O_x) \tag{53}$$

$$= \sum_i (O_x^T X^T X O_x)_{ii} = \sum_i \sum_j (O_x^T)_{ij} (X^T X O_x)_{ji} = \sum_j \sum_i (X^T X O_x)_{ji} (O_x^T)_{ij}$$

$$= Tr(X^T X O_x O_x^T)$$

$$= Tr(X^T X)$$

Now let's define an analogous rotation for $\boldsymbol{Y}$.

$$J = Y O_y \tag{54}$$

Matrices $\boldsymbol{S}$ and $\boldsymbol{J}$ contain columns of vectors $\boldsymbol{s_i}$ and $\boldsymbol{j_i}$. Remember because $\boldsymbol{Y}$ and $\boldsymbol{X}$ are centred matrices, the matrix products $\boldsymbol{X^T X}$ and $\boldsymbol{Y^T Y}$ are nothing more than covariance – variance matrices for the variables of each space. The traces of which represent the total variation for the space. Thinking of the vectors $\boldsymbol{s_i}$ and $\boldsymbol{j_i}$ we could ask, for given rotations how close can we get to $\boldsymbol{s_i = j_i}$. The relationship for the magnitude of the total difference between the vectors is.

$$\sum_{i=1}^{m} |s_i - j_i|^2 + \sum_{m+1}^{k} |s_i|^2 = |\Delta_{sj}|^2 \tag{55}$$

The latter summation reflects the fact that the number of $\boldsymbol{s_i}$ vectors exceeds the number of $\boldsymbol{j_i}$ based on the differences in the dimensions of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

$$\left|\boldsymbol{\Delta_{sj}}\right|^2 = \sum_{i=1}^{m}|\boldsymbol{s_i}|^2 - 2\sum_{i=1}^{m}|\boldsymbol{s_i j_i}| + \sum_{i=1}^{m}|\boldsymbol{j_i}|^2 + \sum_{m+1}^{k}|\boldsymbol{s_i}|^2$$

$$= \sum_{i=1}^{k}|\boldsymbol{s_i}|^2 - 2\sum_{i=1}^{m}|\boldsymbol{s_i j_i}| + \sum_{i=1}^{m}|\boldsymbol{j_i}|^2$$

$$= \boldsymbol{Tr}(\boldsymbol{S^T S}) - \boldsymbol{2Tr}(\boldsymbol{S^T J}) + \boldsymbol{Tr}(\boldsymbol{J^T J})$$

$$= \boldsymbol{Tr}(\boldsymbol{X^T X}) - \boldsymbol{2Tr}\left(\boldsymbol{O_x^T X^T Y O_y}\right) + \boldsymbol{Tr}(\boldsymbol{Y^T Y});$$

$$\left|\boldsymbol{\Delta_{sj}}\right|^2 = \boldsymbol{tr}(\boldsymbol{X^T X}) + \boldsymbol{tr}(\boldsymbol{Y^T Y}) - \boldsymbol{2tr}\left(\boldsymbol{O_x^T X^T Y O_y}\right) \tag{56}$$

Equation 56 defines the magnitude of a vector which is equal to the difference in lengths of two vectors $\boldsymbol{s_i}$ and $\boldsymbol{j_i}$. It can be seen that if one aims to reduce the difference between the two vectors it is necessary to increase the magnitude of the final term $-\boldsymbol{2tr}\left(\boldsymbol{O_x^T X^T Y O_y}\right)$, the question then becomes what rotation matrices $\boldsymbol{O_y}$ and $\boldsymbol{O_x^T}$ would maximise this term. Well if $\boldsymbol{O_x^T X^T Y O_y}$ is a diagonal matrix then the above term is maximised.

$$\boldsymbol{O_x^T X^T Y O_y} = \boldsymbol{D}; \tag{57}$$

$\boldsymbol{D}$ being some matrix with only diagonal entries, ordered by size.

$$\boldsymbol{X^T Y O_y = O_x D}$$

$$\boldsymbol{X^T Y = O_x D O_y^T} \tag{58}$$

So to find $\boldsymbol{O_x}$ and $\boldsymbol{O_y^T}$ we proceed as follows;

$$\boldsymbol{X^T Y Y^T X = O_x D O_y^T O_y D O_x^T = O_x D^2 O_x^T} \tag{59}$$

$$\boldsymbol{Y^T X X^T Y = O_y D O_x^T O_x D O_y^T = O_y D^2 O_y^T} \tag{60}$$

So we see $\boldsymbol{O_x}$ is a matrix containing the eigenvectors of $\boldsymbol{X^T Y Y^T X}$ which we know to be $\boldsymbol{w}$ thus $\boldsymbol{O_x}$ contains as its columns the vectors $\boldsymbol{w}$. $\boldsymbol{O_y}$ is a matrix which contains the eigenvectors of $\boldsymbol{Y^T X X^T Y}$ which we know to be $\boldsymbol{c}$, and as such the columns of $\boldsymbol{O_y}$ are the vectors $\boldsymbol{c}$. It now becomes clear that the rotation matrices are nothing more than matrices containing the weight vectors $\boldsymbol{w}$ and $\boldsymbol{c}$, so then the matrices $\boldsymbol{S}$ and $\boldsymbol{J}$ contain as their columns the vectors $\boldsymbol{t}$ and $\boldsymbol{u}$ respectively. So under the requirement of minimised vector differences for $\boldsymbol{s}$ and $\boldsymbol{j}$, the result follows that the rotation matrices become $\boldsymbol{W}$ and $\boldsymbol{C}$ and the rotated matrices become $\boldsymbol{T}$ and $\boldsymbol{U}$. So next we look to identify the covariance of these two vectors, because ideally this should be as large as possible. The covariance of two components is defined in general as.

$$cov(\boldsymbol{f}, \boldsymbol{g}) = \frac{\boldsymbol{f^T g}}{\boldsymbol{N}} \tag{61}$$

So for $\boldsymbol{t}$ and $\boldsymbol{u}$ we have;

$$[Cov(\boldsymbol{t},\boldsymbol{u})]^2 = [Cov(\boldsymbol{Xw},\boldsymbol{Yc})]^2 = [\boldsymbol{w^T X^T Y c}]^2;$$

$$D_{11} > D_{22} > D_{33} > \cdots > D_{ii}$$

$$\max[\boldsymbol{w^T X^T Y c}]^2 = D_{11} \tag{62}$$

Hence the vectors have maximal covariance when $\boldsymbol{w} = \boldsymbol{w_1}$ and $\boldsymbol{c} = \boldsymbol{c_1}$. It turns out that the second eigenvalue diagonal entry of $\boldsymbol{D}$ is actually less than the first diagonal entry of the matrix for the second iteration, so PLS identifies $\boldsymbol{t}$ and $\boldsymbol{u}$ as a single pair for each iteration.

2.2.9.3 Partial Least Squares for Linear Regression

As before the regression equation can be written adopting $\boldsymbol{X} = \boldsymbol{X_C}$, as before $\boldsymbol{X_C}$ will be the information the PLS algorithm has constructed**.**

$$\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{E} \tag{63}$$

$$\boldsymbol{X} = \boldsymbol{TP^T} = \sum_i \boldsymbol{t_i p_i^T}$$

Let $\boldsymbol{T} = \boldsymbol{XW} = \boldsymbol{TP^T W}$

$$\boldsymbol{XW} = \boldsymbol{T}(\boldsymbol{P^T W}) \tag{64}$$

$$\boldsymbol{T} = \boldsymbol{XW}(\boldsymbol{P^T W})^{-1} = \boldsymbol{XR} \tag{65}$$

$$\boldsymbol{W}(\boldsymbol{P^T W})^{-1} = \boldsymbol{R} \tag{66}$$

$$\boldsymbol{P^T R} = \mathbf{I}$$

The weighting matrices $\boldsymbol{R}$ and $\boldsymbol{W}$ contain weight vectors which span the same space.

Equation 63 becomes

$$\boldsymbol{Y = TP^T B + E = XRP^T B + E} \tag{67}$$

The decomposition of $\boldsymbol{Y}$ using scores and weights can be seen below,

$$\boldsymbol{Y = UC^T + E} \tag{68}$$

On occasion reference is made to a regression vector which links vectors $\boldsymbol{t}$ and $\boldsymbol{u}$, in fact these vectors are inherently related as may be shown below.

$$\boldsymbol{u^T t = \frac{c^T (Y^T t)}{c^T c} = \frac{c^T (c(\, t^T t))}{c^T c} = t^T t}$$

$$\boldsymbol{u \cong t}$$

This is the result of convergence under the outlined constraints, and $\boldsymbol{E \cong 0}$. The mixed relationship between $\boldsymbol{Y}$ and $\boldsymbol{X}$ is illustrated by taking advantage of vector inner product result.

$$\boldsymbol{Y = TC^T + E = TP^T B + E}$$

$$\boldsymbol{C^T = P^T B}$$

2.2.9.4 Derivation of PLS Regression Coefficient Vector

The regression model of $\boldsymbol{Y}$ and $\boldsymbol{T}$ is $\boldsymbol{Y = TB + E}$; and $\boldsymbol{y = T\beta + e}$.

$$\boldsymbol{e = y - T\beta}$$

$$\boldsymbol{e^T e = (y - T\beta)^T (y - T\beta)}$$

$$\boldsymbol{= (y^T - \beta^T T^T)(y - T\beta)}$$

$$\boldsymbol{= y^T y - y^T Tb - \beta^T T^T y + \beta^T T^T T\beta}$$

$\boldsymbol{y^T T\beta}$ is a scalar and $\boldsymbol{(y^T T\beta)^T = \beta^T T^T y}$

$$\boldsymbol{e^T e = y^T y - 2\,\beta T^T y + \beta^T T^T T\beta};$$

Minimised when $\boldsymbol{\frac{\partial e^T e}{\partial \beta} = 0}$ and $\boldsymbol{\frac{\partial^2 e^T e}{\partial \beta^2} > 0}$

$$\boldsymbol{\frac{\partial e^T e}{\partial \beta} = -2T^T y + 2T^T T\beta = 0}$$

$$\boldsymbol{\beta = (T^T T)^{-1} T^T y}$$

so

$$\hat{y}_{\boldsymbol{pls}} \boldsymbol{= XR(T^T T)^{-1} T^T y};$$

since

$$\boldsymbol{\beta_{ols} = (X^T X)^{-1} X^T y}, \qquad \boldsymbol{X^T y = (X^T X)\beta_{ols}}$$

$$\hat{y}_{\boldsymbol{pls}} \boldsymbol{= XR(T^T T)^{-1} (XR)^T y = XR(T^T T)^{-1} R^T X^T y}$$

$$\boldsymbol{= XR(T^T T)^{-1} R^T (X^T X)\beta_{ols} = XR(T^T T)^{-1} (XR)^T X\beta_{ols}}$$

$$\boldsymbol{= XR(T^T T)^{-1} T^T X\beta_{ols}}$$

$$\boldsymbol{p} = \frac{\boldsymbol{X^T t}}{\boldsymbol{t^T t}} \; so \; \boldsymbol{P} = \boldsymbol{X^T T}(\boldsymbol{T^T T})^{-1} \text{ and}$$

$$\boldsymbol{P^T} = (\boldsymbol{T^T T})^{-1}\boldsymbol{T^T X}$$

$$\hat{y}_{\boldsymbol{pls}} = \boldsymbol{XR}(\boldsymbol{T^T T})^{-1}\boldsymbol{T^T X}\boldsymbol{\beta}_{\boldsymbol{ols}} = \boldsymbol{XRP^T}\boldsymbol{\beta}_{\mathbf{ols}} = \boldsymbol{X}\boldsymbol{\beta}_{\boldsymbol{pls}}$$

$$\boldsymbol{\beta}_{\boldsymbol{pls}} = \boldsymbol{RP^T}\boldsymbol{\beta}_{\boldsymbol{ols}} \tag{69}$$

The above result illustrates the connection between the ordinary least squares coefficients and the partial least squares coefficients. Another interesting observation can be made as follows.

$$\hat{y}_{\boldsymbol{pls}} = \boldsymbol{T}(\boldsymbol{T^T T})^{-1}\boldsymbol{T^T X}(\boldsymbol{X^T X})^{-1}\boldsymbol{X^T y} = \boldsymbol{\Phi}_{\boldsymbol{pls}}\boldsymbol{y} \tag{70}$$

$$\boldsymbol{\Phi}_{\boldsymbol{pls}} = \boldsymbol{T}(\boldsymbol{T^T T})^{-1}\boldsymbol{T^T X}(\boldsymbol{X^T X})^{-1}\boldsymbol{X^T} \tag{71}$$

Where $\boldsymbol{\Phi}_{\boldsymbol{pls}}$ is the PLS projection matrix projecting $\boldsymbol{y}$ onto the column space of $\boldsymbol{X}$ then the column space of $\boldsymbol{T}$ to generate $\boldsymbol{y}$ estimates, $\hat{y}$. The Projection matrix $\boldsymbol{RP^T}$ is idempotent but not symmetric, therefore it generates oblique projections.

2.2.10 Geometry of the PLS Algorithm [152]

The projection matrix for the regression coefficient matrices, found in the PLS algorithm can be represented as follows.

$$\boldsymbol{R}\boldsymbol{P}^{\boldsymbol{T}}\boldsymbol{\beta}_{\boldsymbol{ols}} = \boldsymbol{\beta}_{\boldsymbol{pls}}$$

$$\boldsymbol{R} = \boldsymbol{W}(\boldsymbol{P}^{\boldsymbol{T}}\boldsymbol{W})^{-\mathbf{1}}$$

$$\boldsymbol{W}(\boldsymbol{P}^{\boldsymbol{T}}\boldsymbol{W})^{-\mathbf{1}}\boldsymbol{P}^{\boldsymbol{T}} = \mathrm{H}_{\boldsymbol{W}.\boldsymbol{P}^{\boldsymbol{T}}\perp}$$

The column vectors of $\boldsymbol{R}$ and $\boldsymbol{W}$ span the same space and as such projections onto this space are equivalent, that is

$$\mathrm{H}_{\boldsymbol{R}.\boldsymbol{P}^{\boldsymbol{T}}\perp} = \boldsymbol{R}(\boldsymbol{P}^{\boldsymbol{T}}\boldsymbol{R})^{-\mathbf{1}}\boldsymbol{P}^{\boldsymbol{T}} = \boldsymbol{R}(\mathbf{I})^{-\mathbf{1}}\boldsymbol{P}^{\boldsymbol{T}} = \boldsymbol{R}\boldsymbol{P}^{\boldsymbol{T}} = \boldsymbol{W}(\boldsymbol{P}^{\boldsymbol{T}}\boldsymbol{W})^{-\mathbf{1}}\boldsymbol{P}^{\boldsymbol{T}} = \mathrm{H}_{\boldsymbol{W}.\boldsymbol{P}^{\boldsymbol{T}}\perp}$$

These projections occur along the space perpendicular to $\boldsymbol{P}^{\boldsymbol{T}}$ onto the space spanned by $\boldsymbol{R}$ and $\boldsymbol{W}$.

2.2.10.1 Single Response, Two Predictor Variables

For the situation of a single response variable and two predictor variables the problem is simplified greatly, that is because the only criterion of interest is the maximisation of the squared covariance of $\boldsymbol{t}_{\boldsymbol{i}}$ with $\boldsymbol{y}$ itself. As such the weights can be determined explicitly.

2.2.10.2 The Object Space

In the object space it is sufficient to work only with $\hat{y}_{OLS}$, because

$$\boldsymbol{t_i^T \hat{y}_{OLS} = r_i^T X^T \hat{y}_{OLS} = r_i^T X^T X (X^T X)^- X^T y = r_i^T X^T \, y}$$

The weights vectors $\boldsymbol{r_i}$ satisfy $\boldsymbol{r_i^T r_i} = 1$, which in this instance is simply the equation of a circle of radius one in two-dimensional *variable* space. The negative symbol represents the use of the Moore-Penrose pseudoinverse or generalised inverse. [18] The image of this circle in n −dimensional object space is an ellipse.

$$\boldsymbol{(XX^T)^- = X^{T-} X^- = D^-}$$

$$\boldsymbol{r_i^T r_i = r_i^T X^T X^{T-} X^- X r_i = r_i^T X^T (XX^T)^- X r_i = t_i^T D^- t_i} = 1$$

Because $\boldsymbol{X^T}$ is the conjugate transpose of $\boldsymbol{X}$, $\boldsymbol{(XX^T)}$ can be expanded as $\boldsymbol{X^{T-} X^-}$.

$$\boldsymbol{t_i^T D^- t_i} = 1 \tag{72}$$

Is an equation representing the ellipse $\mathbb{D}_1$ of which the vectors $\boldsymbol{t_1}$ and $\boldsymbol{t_2}$ lie on, orthogonal to one another, as proven in appendix 5.

This ellipse occupies a subspace of the n-dimensional object space of which the row vectors of $\boldsymbol{X}$, $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ occupy. The principal axes of $\mathbb{D}_1$ are the eigenvectors of $\boldsymbol{D^-}$ as can be visualised in Figure 2.2.11.

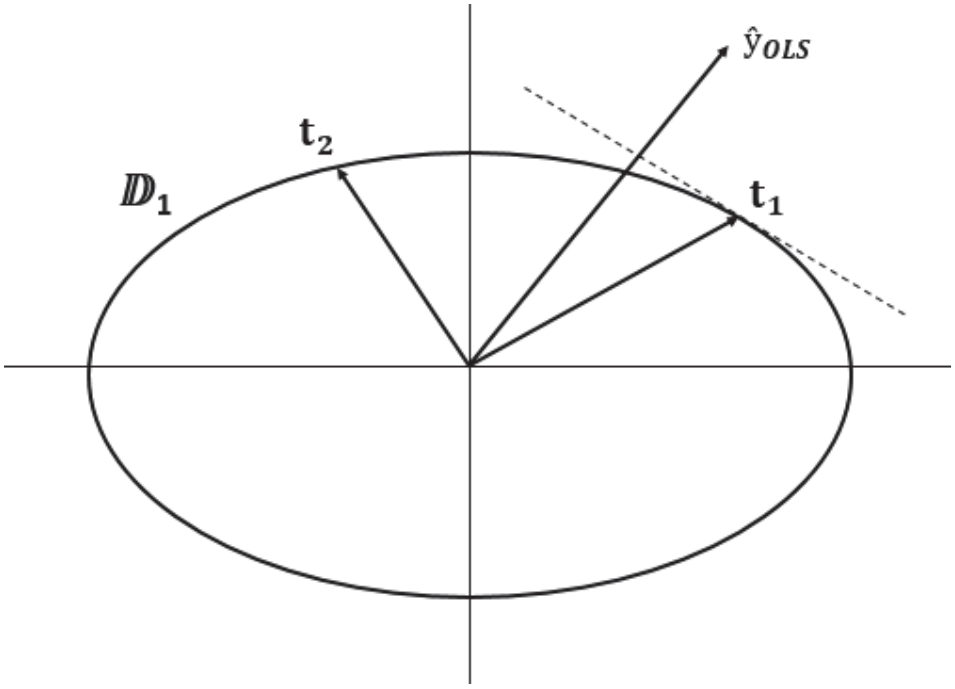

Figure 2.2.11 Two dimensional demonstration of OLS response vector and t vectors from the NIPALS algorithm

In Figure 2.2.11 it can be seen that the OLS vector lies in this object space, and the objective of PLS is to determine the vectors $\boldsymbol{t_1}$ and $\boldsymbol{t_2}$. First, the optimal weight vector $\boldsymbol{r_1}$ that maximises $\left(\hat{y}_{OLS}^{T}\boldsymbol{X r_1}\right)^2$ which is the squared inner product of $\boldsymbol{t_1}$ with $\hat{y}_{OLS}$.

$$\boldsymbol{t_1} = \boldsymbol{X r_1} = \boldsymbol{X}(k\boldsymbol{X}^{T}\hat{y}_{OLS}) = k\boldsymbol{D}\hat{y}_{OLS} \tag{73}$$

Which is the familiar tangent rotation of $\hat{y}_{OLS}$. After identifying $\boldsymbol{t_1}, \boldsymbol{t_2}$ is found as the vector directly orthogonal and lying on the ellipse.

2.2.10.3 The Variable space

Within the variable space the response vector $\boldsymbol{y}$ is contained in the direction of $\boldsymbol{\beta_{ols}}$. The first PLS weight vector $\boldsymbol{r_1} = \boldsymbol{w_1}$ can easily be constructed by a tangent rotation of $\boldsymbol{\beta_{OLS}}$ onto $\boldsymbol{X^T X}$ where we will define an ellipse $\mathbb{S}$

$$\boldsymbol{S} = (\boldsymbol{X^T X})^{-}$$

To see this remember

$$\boldsymbol{r_1} = k\boldsymbol{X^T}\hat{y}_{\boldsymbol{OLS}} = k\boldsymbol{X^T X}\boldsymbol{\beta_{OLS}} = k\boldsymbol{S^{-}}\boldsymbol{\beta_{ols}}$$

If the weight vectors are further restricted to unit norm, they will occupy a circle with a radius of 1. With $\boldsymbol{r_1}$ we can identify the corresponding loading vector $\boldsymbol{p_1}$ using the relationship

$$\boldsymbol{p_1} = \frac{\boldsymbol{X^T t_1}}{\boldsymbol{t_1^T t_1}} = \left(\frac{1}{\boldsymbol{t_1^T t_1}}\right)\boldsymbol{X^T X r_1} = k\boldsymbol{S^{-}}\boldsymbol{r_1}$$

Which once again is a tangent rotation. The length can be identified utilising the fact that

$$\boldsymbol{p_1^T r_1} = \frac{\boldsymbol{t_1^T X r_1}}{\boldsymbol{t_1^T t_1}} = \frac{\boldsymbol{t_1^T t_1}}{\boldsymbol{t_1^T t_1}} = 1$$

As such the length of $\boldsymbol{p_1}$ can be found by dropping a perpendicular line from the point of $\boldsymbol{p_1}$ on to $\boldsymbol{r_1}$. After finding $\boldsymbol{w_1}$ the algorithm defines $\boldsymbol{w_2}$ as the unit norm vector orthogonal to $\boldsymbol{w_1}$.

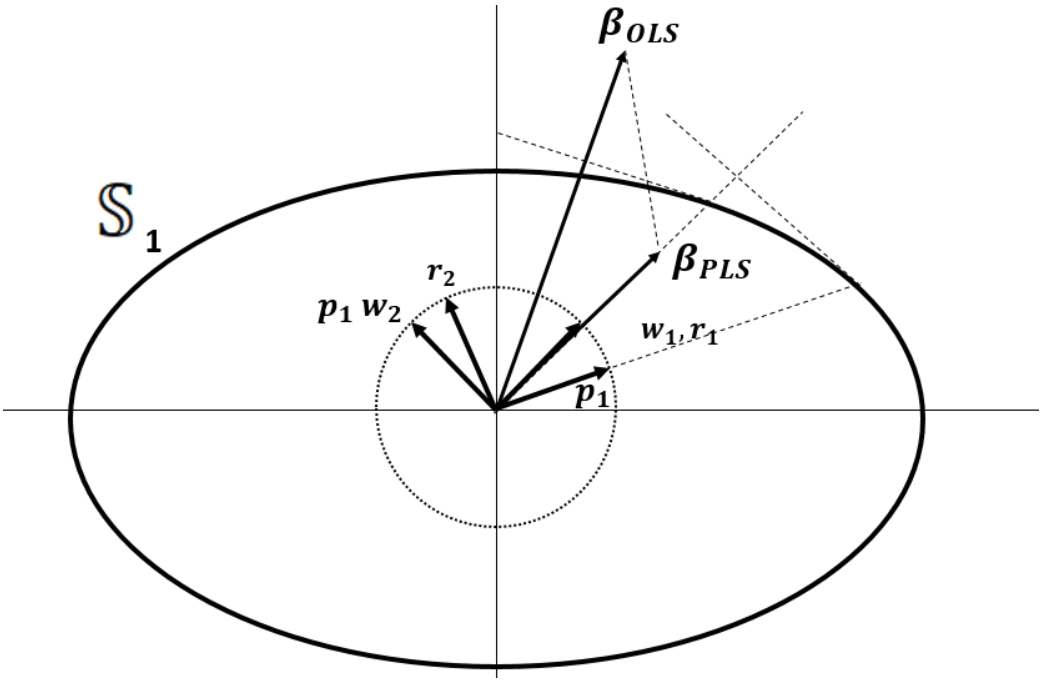

Figure 2.2.12: A selection of vectors occupying the variable space, generated from the NIPALS algorithm. Adapted from [152].

The construction of the PLS model will be finished once the $\boldsymbol{\beta_{pls}}$ regression vector is created, this vector can be found using the oblique projection matrix identified earlier as.

$$\boldsymbol{W(P^TW)^{-1}P^T = RP^T}$$

$\boldsymbol{\beta_{pls}}$ in the above image represents the regression coefficient vector obtained after one iteration of its determination as an oblique projection of $\boldsymbol{\beta_{OLS}}$ onto $\boldsymbol{r_1}$ along a direction perpendicular to $\boldsymbol{p_1}$ i.e $\boldsymbol{r_2}$. The $\boldsymbol{\beta_{pls}}$vector becomes the $\boldsymbol{\beta_{OLS}}$ vector if the iterative process is fully completed, which is equivalent to a full deflation of the **X** matrix. However, practically this may not be necessary as successive residual matrices contain less useful information than their predecessors as such often the process may stop with

$\boldsymbol{X_r \neq 0}$ which results in $\boldsymbol{\beta_{OLS} \neq \beta_{pls}}$

Chapter 3.0: Spectral Pre-treatment and Principal Components Analysis Results.

Prior to regression analysis, several different mathematical pretreatments were employed, these pretreatments ensured variations between spectra were due to only chemical variations. Figures 3.1 and 3.2 contain visual representations of the raw spectra as well as plots of spectra after three of the more visually distinct pre-treatments. SNV and SG methods produce noticeable changes but the methods that introduce noise suppression introduce subtler changes. These steps generate changes apparent in statistical analysis but unnoticeable upon simple inspection.

Examining the raw spectra, it is clear that the baselines of each spectrum differ. This difference in baseline values is one of the more important issues to resolve, as it induces differences in intensity values that are certainly not of chemical origin. The three remaining figures all display significantly reduced baseline drift. Noise reduction steps were also employed during this research, however, their effects are often not discernable upon visual inspection. The low impact of noise removal is suggested to be partially due to the high signal-to-noise ratio of the raw data, a result of the new germanium crystal used for this research. It is worth noting that germanium ATR crystals do eventually wear over time leading to significant degradation in signal. As mentioned the data collected in this work was obtained with a new germanium ATR crystal which showed no signs of degradation over the duration of the project (hours of use > 150).

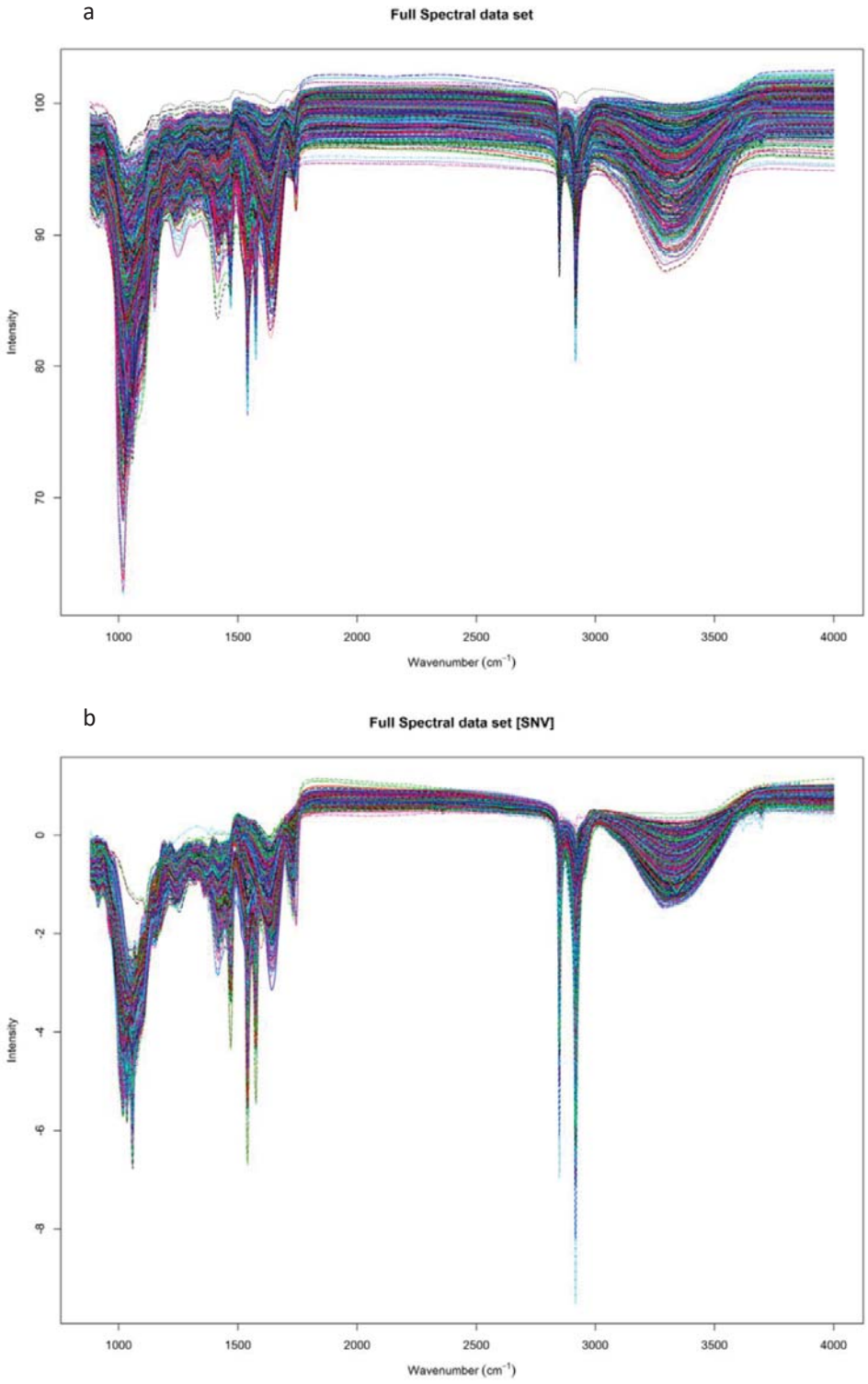

Figure 3.1: (a) The full raw infrared reflectance data set. (b) The full data set, post – SNV.

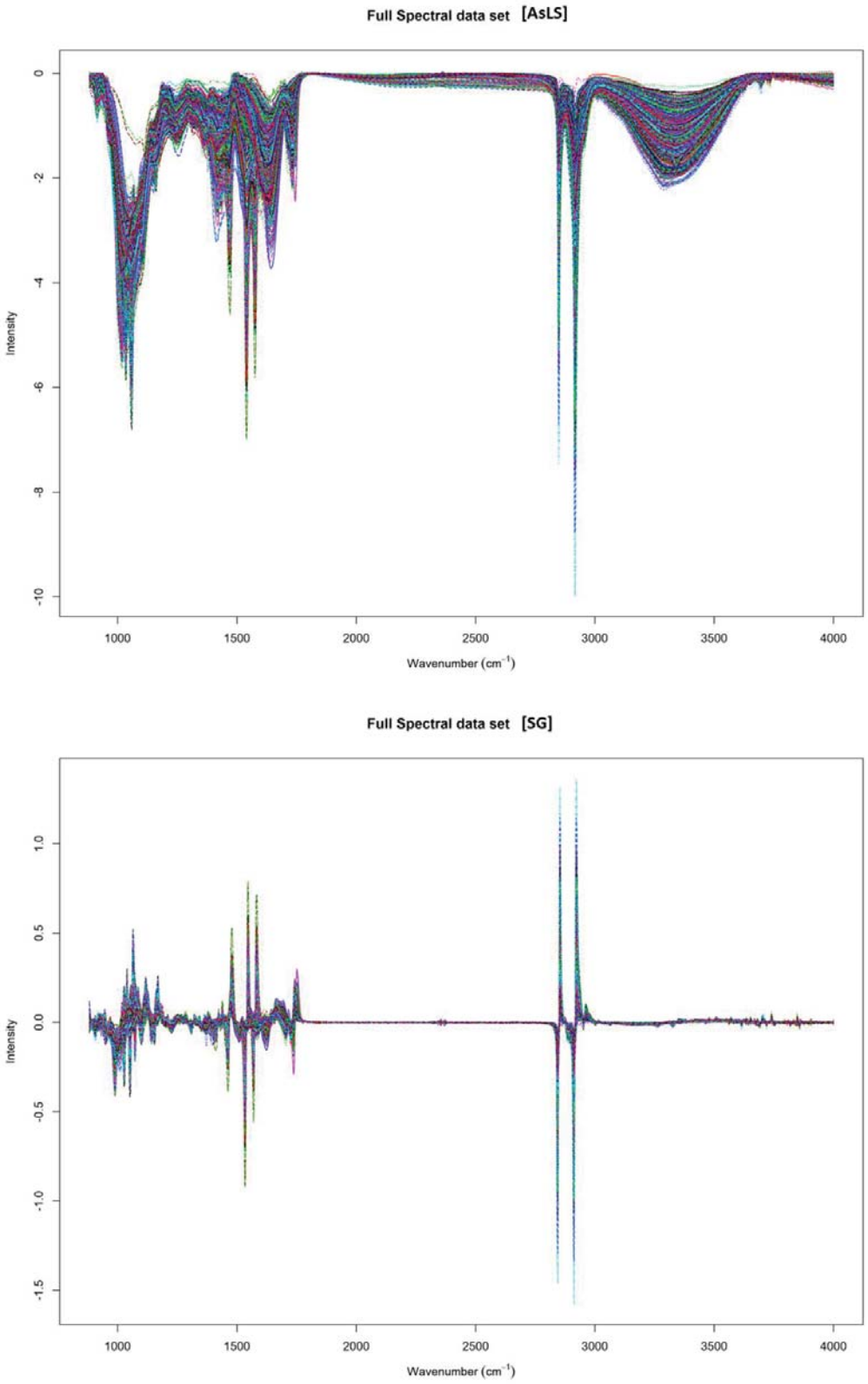

Figure 3.2: (a) Full data set post – PT4. (b) Full Data set post PT3.

Principal component analysis is the first form of analysis carried out in this project. The intention of the technique is to provide insight into trends present in the dataset prior to further analysis. These trends are identified by expressing the results in terms of the data set's principal components, rather than the unwieldy initial variables. In this instance each variable corresponds to the reflectance at a given wavenumber, thus each spectrum may be considered a point in an observation space with 1618 dimensions (not orthogonal). Each principal component is a linear combination of the prior variables, such that the principal components capture the directions of the greatest variance in the dataset. The method of obtaining the principal components also ensures they are orthogonal, *i.e.* linearly independent.

In Figures 3.3 – 3.5 are the results of principal components analysis on the full data set. The figures contain plots of the spectra relative to the first three principal components. The plots follow from PT2, PT6 and PT3, respectively. These three pre-treatments offer the most distinctive differences at this stage of the analysis. Displayed also in the figures are the *variable vectors* (labelled by wavenumber) possessing the largest loadings. The loadings refer to how aligned a variable vector is with a principal component. From the figure it is clear that the PT3 model provides three wavenumber variables which align well with the first three principal components. This figure also demonstrates rather distinct clustering, suggesting that these wavenumber regions contain relevant information to general structural variations between the sample types.

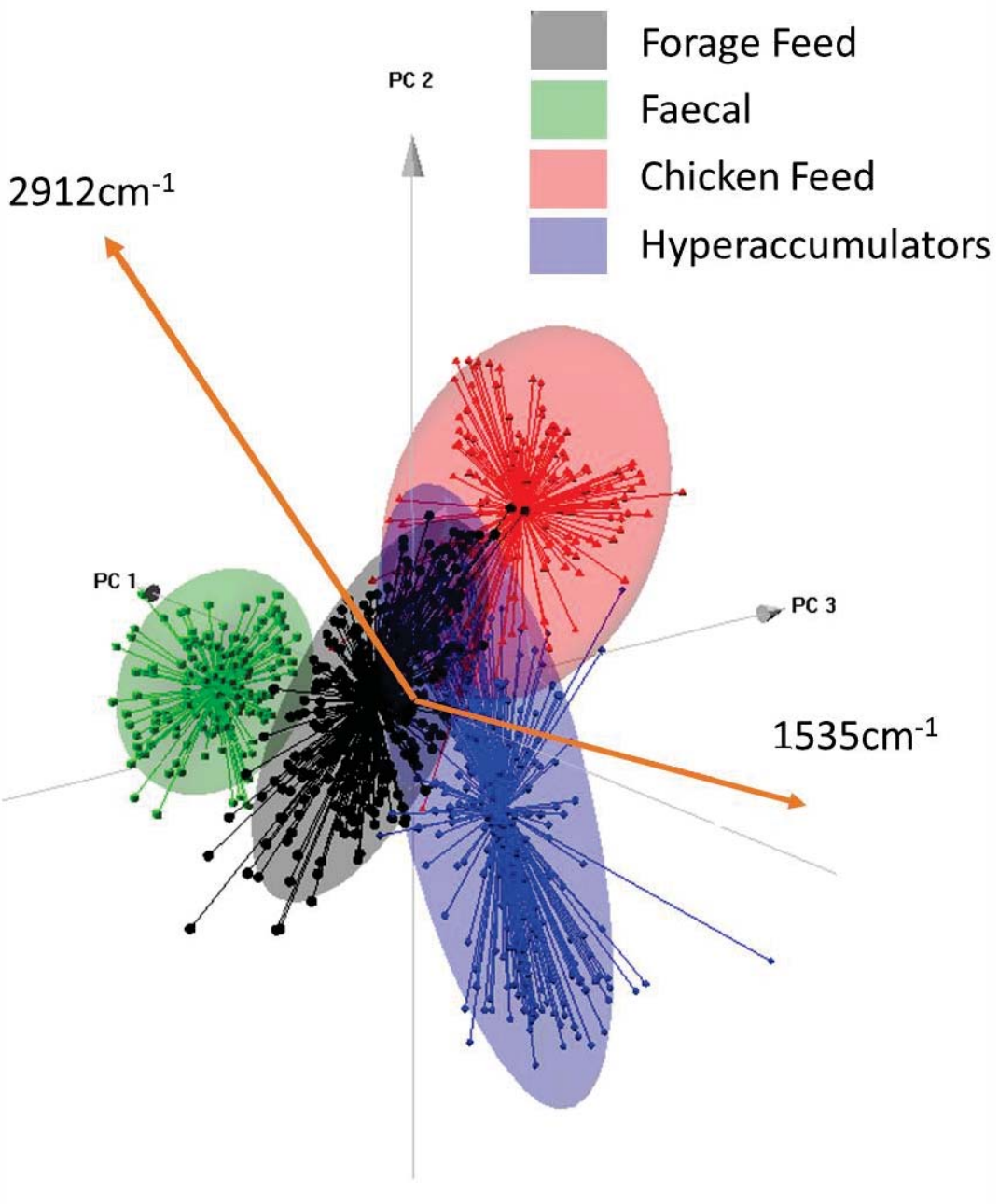

Figure 3.3: PT2 PCA plot for the first three principal components.
Orange arrows reflect variable vectors with the largest loadings to a particular PC.

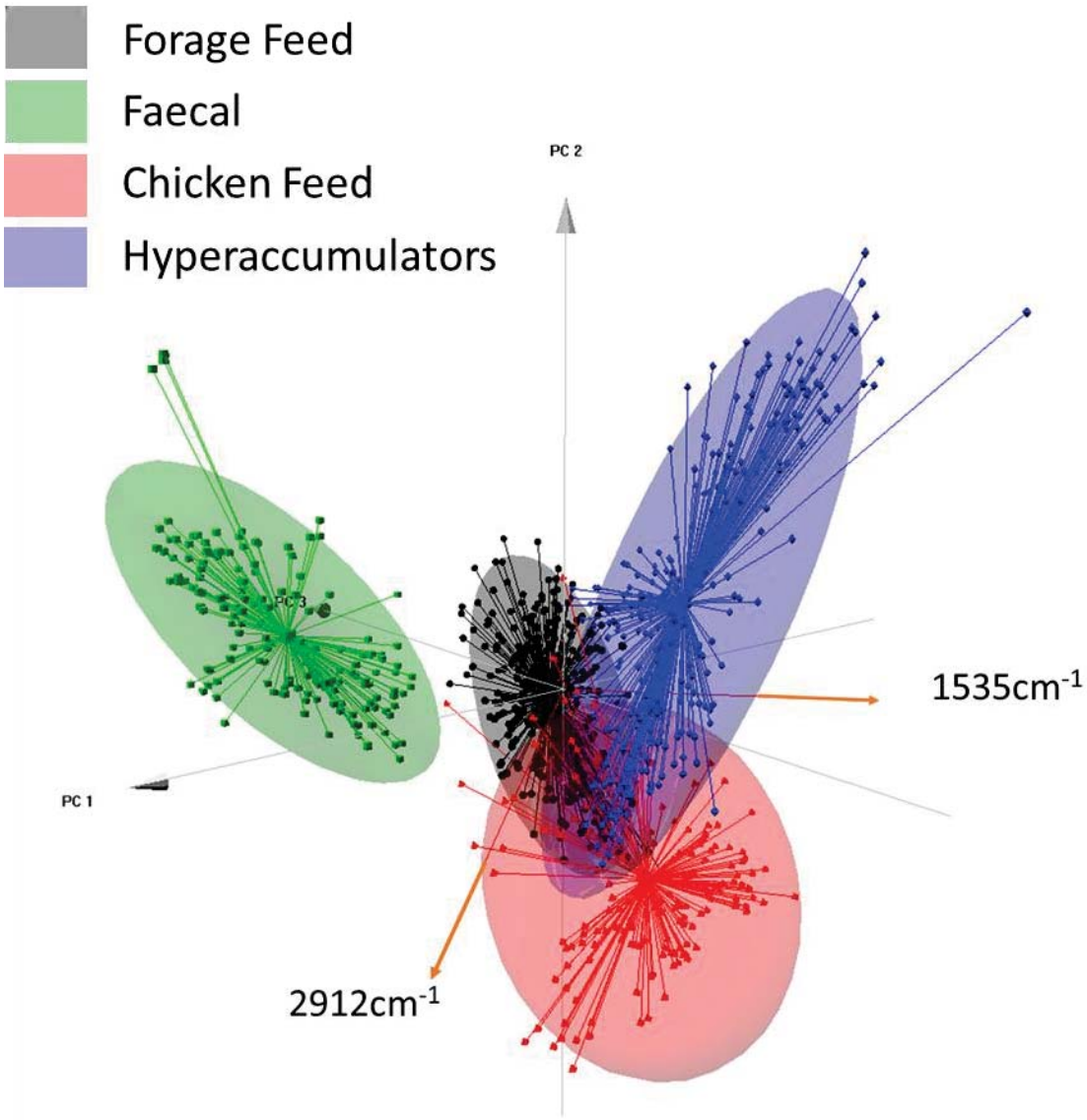

Figure 3.4: PT6 PCA plot for the first three principal components.
Orange arrows reflect variable vectors with the largest loadings to a particular PC.

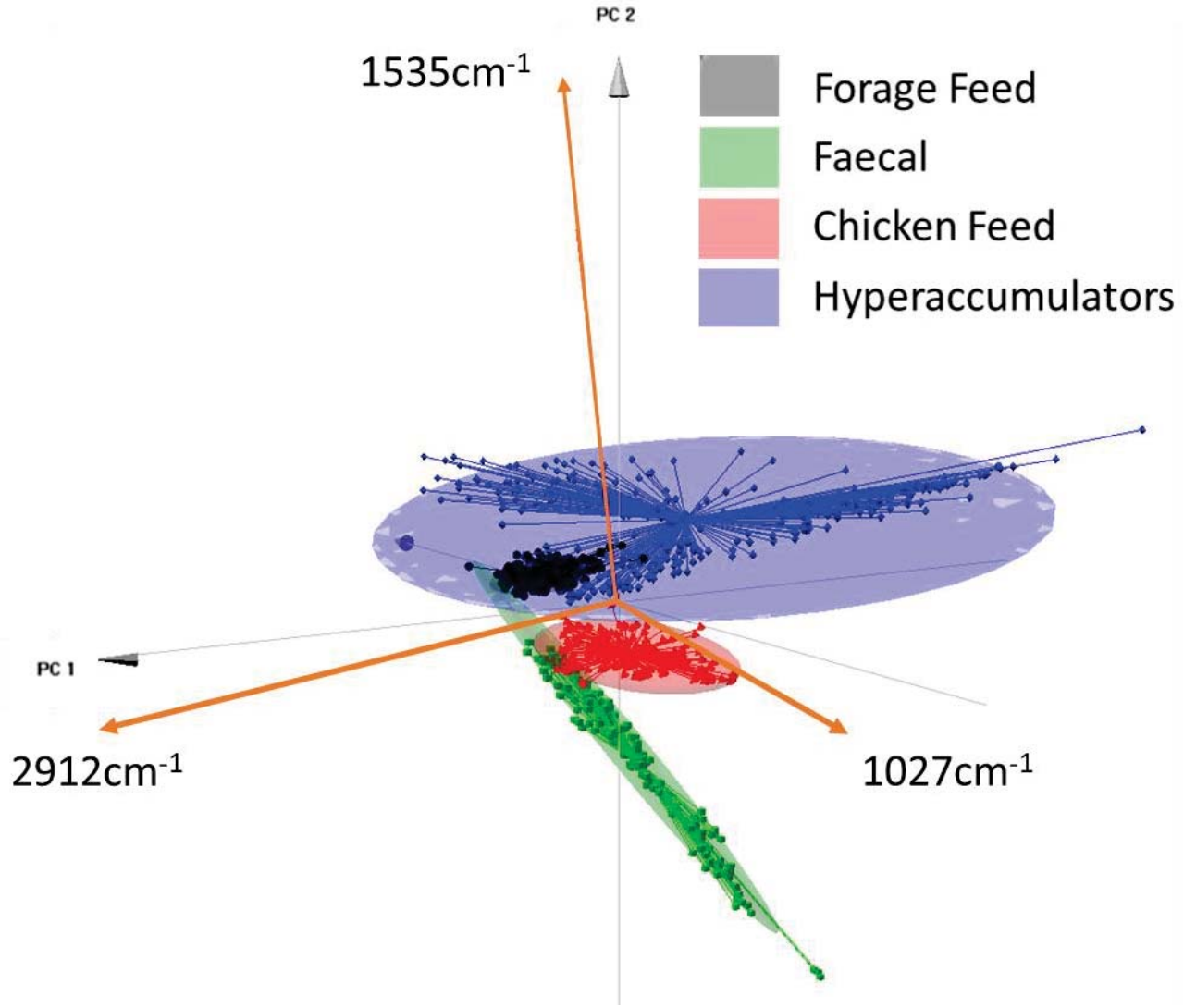

Figure 3.5: PT3 PCA plot for the first three principal components.
Orange arrows reflect variable vectors with the largest loadings to a particular PC.

The other plots generally identify the same wavenumbers as more significant, but not to the same extent. The faecal sample set cluster is the most separated in the plots, followed by the chicken feed samples and lastly the forage and hyperaccumlators. This pattern is sensible given that the chicken feed, forage and hyperaccumulator samples all contain plant matter prior to digestion.

Figure 3.6 contains two dimensional plots from the PT3 PCA results. The PC2 and PC3 plot demonstrates separation based on sample type as either plant matter, chicken feed or faecal matter. Upon further inspection of figure 3.6 (b), it is apparent that within the region stated as containing plant matter further division is possible. Splitting the hyperaccumulator samples into leaves or ground plant matter, demonstrates that the ground forage feed plant matter and hyperaccumulators cluster well, while the leaf samples show tremendous variation in PC3. PC3 possesses the most significant loading from the 1027 cm^{-1} region. The 1027 cm^{-1} region contains contributions from the O – H and C – OH stretch associated in the literature as due to cell wall

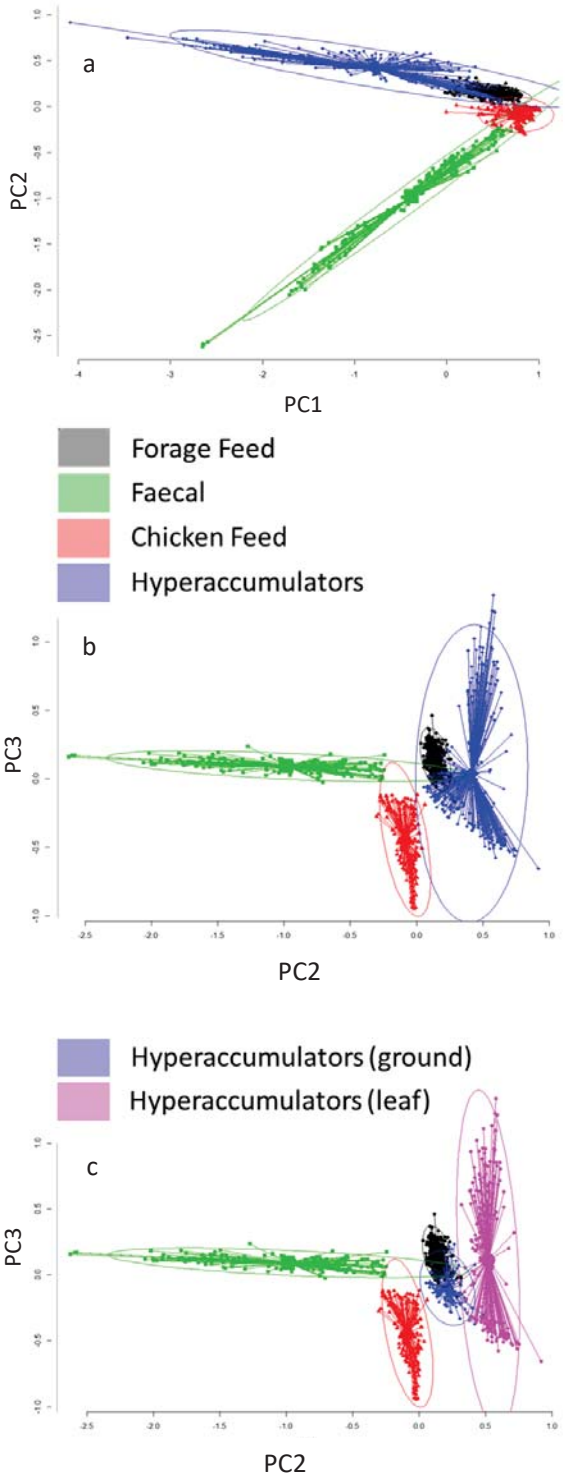

Figure 3.6: (a) PC1 vs PC2 for the PT3 spectra. (b) PC2 vs PC3 for the PT3 spectra. (c) PC2vs PC3 for the PT3, now with separate coloring for hyperaccumulator sample type.

polysaccharides.[162] The distinct variation in peak intensity for the leaf hyperaccumulator samples is suggested to be due to the variation of plant structures the infrared radiation was interacting with during analysis.

By contrast, the faecal samples contain a tremendous amount of variation relative to PC 2, which possesses a large loading from the 1535 cm^{-1} region. This region is associated with the amide II band, [160,162] suggesting variations in protein content are most distinct in the initial faecal spectral data set.

Performing PCA on each sample set provided interesting results only for the forage feed samples. Within this sample set, clustering was only observed in the PC1 verse PC2 scores plots, for PT3 and PT5 (see Figure 3.7 for illustration).

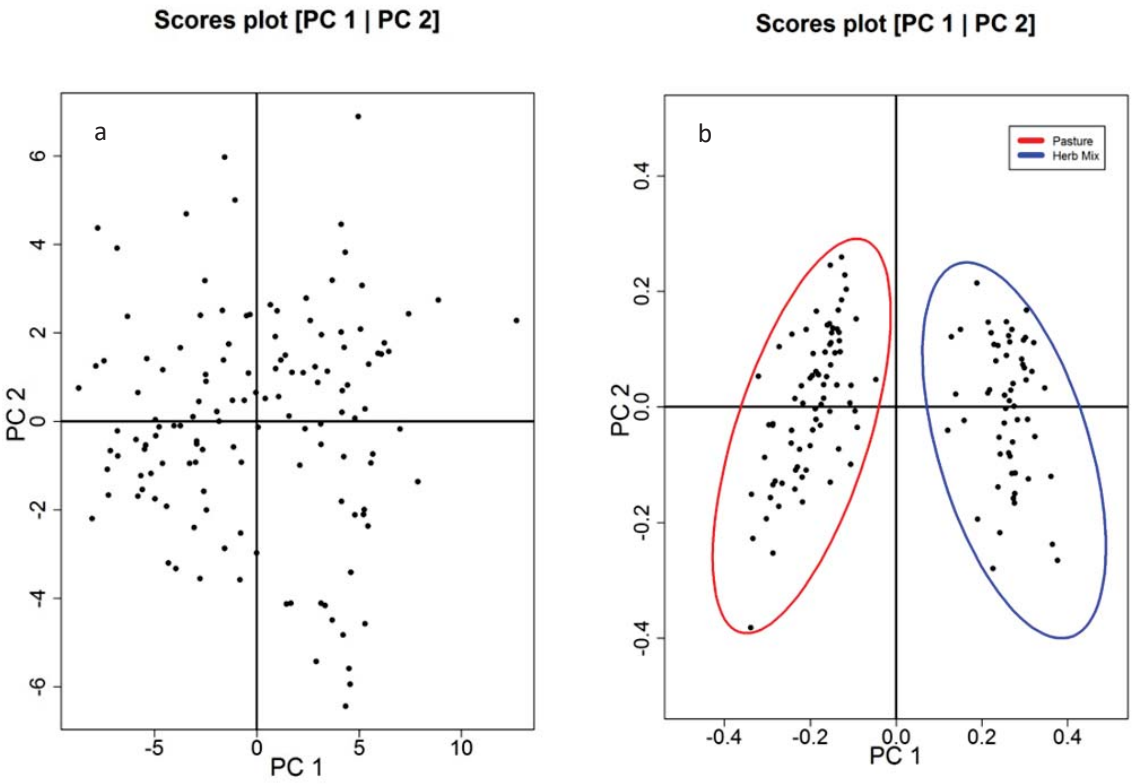

Figure 3.7: (a) PC1 vs PC2 for the PT2 spectra. (b) PC1 vs PC2 for the PT3 spectra.

The remaining pre-treatments failed to induce this effect in the principal component analysis. The key difference between these particular pre-treatments

was the use of the Savitsky-Golay derivative filter, this result follows from the previous example of derivative plots producing better clustering in the principal component space. The separation is induced predominantly by the scores values for the first component.

Figure 3.8 contains the loadings for the first two components of the PT2 and PT3 principal components. Comparing the loadings of the first components of each, the PT3 component has greater loadings from the 1030 cm^{-1} region as well as the 2800 – 3000 cm^{-1} region. The 3200 – 3500 cm^{-1} region is suppressed in the PT3 plot, demonstrating the suppression of water contribution. The 1030 cm^{-1} region still possesses significant loading in the PT2 first component plot, however, there is more significance distributed to the 1600 cm^{-1} and 3200 – 3500 cm^{-1}. Because the first component of the PT3 generates clustering based on the type of forage feed (herb mix or pasture), the regions of 1030 cm^{-1} and 2800 – 3000 cm^{-1} must contain information relative to forage types. The region of 1030 cm^{-1} was discussed previously as containing information relating to cell wall polysaccharides, whilst the 2800 – 3000 cm^{-1} region is the aliphatic C – H stretch region.[162] Therefore, both regions contain information that may have been expected to vary based on species' classifications.

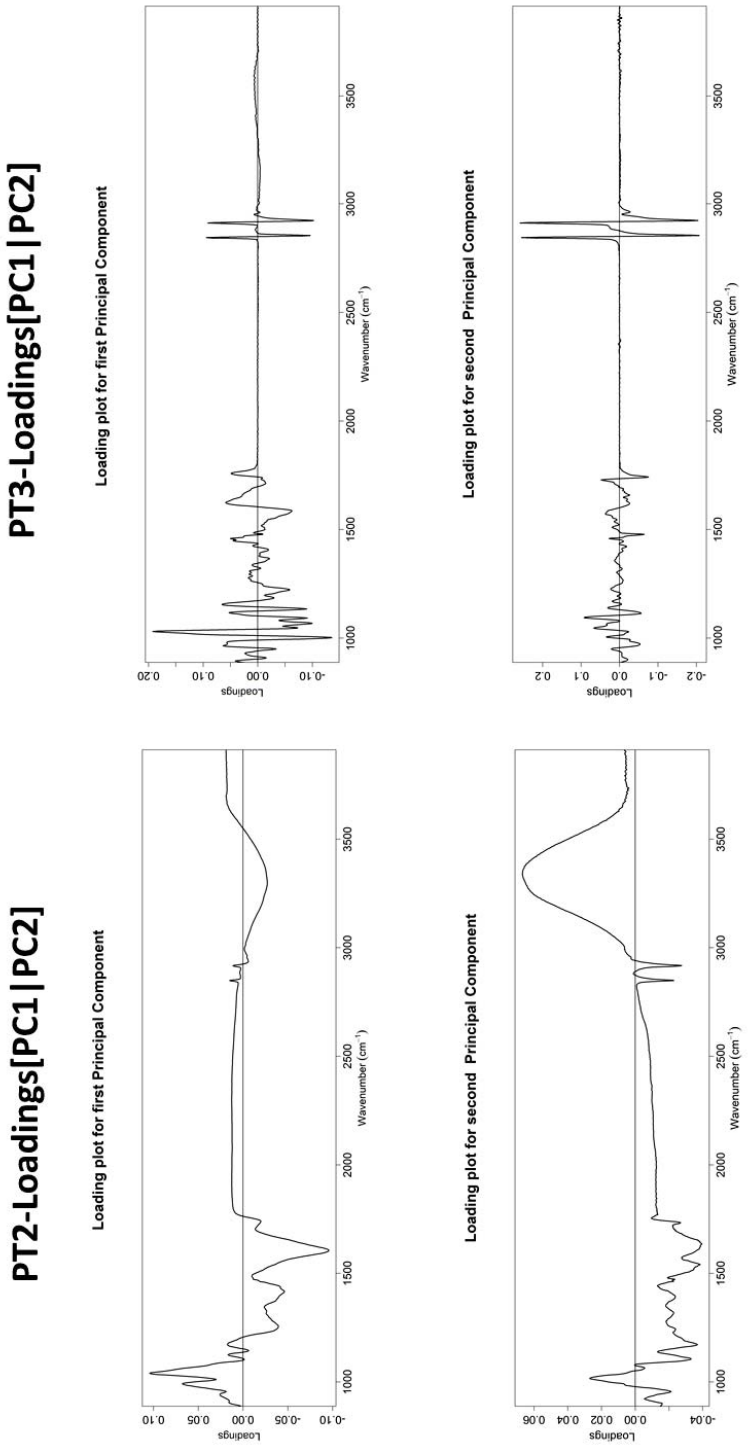

Figure 3.8: (a) Loadings of PC 1 and PC 2 for PT2. (b) Loadings of PC 1 and PC 2 for PT3.

Chapter 4.0: Results and Discussion I
Feature Extraction Regression Methods

This section will present the results and subsequent discussion for the Ridge, LASSO and Elastic Net regression methods. These methods are grouped together as they may all be considered feature extraction techniques. The discussion will present information relating to model performance in cross – validation or external validation context, the latter is only available for the forage based sample sets. For all of this chapter and the following, calibration data is available but not discussed. The motivation for excluding calibration data, is as follows, if the calibration results are high quality and the model is reliable, then the cross – validation (or external validation) data should also be high quality. If the calibration results are high quality and the cross – validation (or external – validation) results are poor, then the model is over – fitting the calibration data. A model's true worth is identified in its ability to predict new observations, for these reasons calibration data is omitted. Literature thresholds for high quality models have been noted as $RPD > 2$ and $R^2 > 0.8$. [171,172] The present results and discussion section is divided into the three logical divisions based on the methods, the methods are arranged in chronological order based on their development.

4.1 Ridge Regression

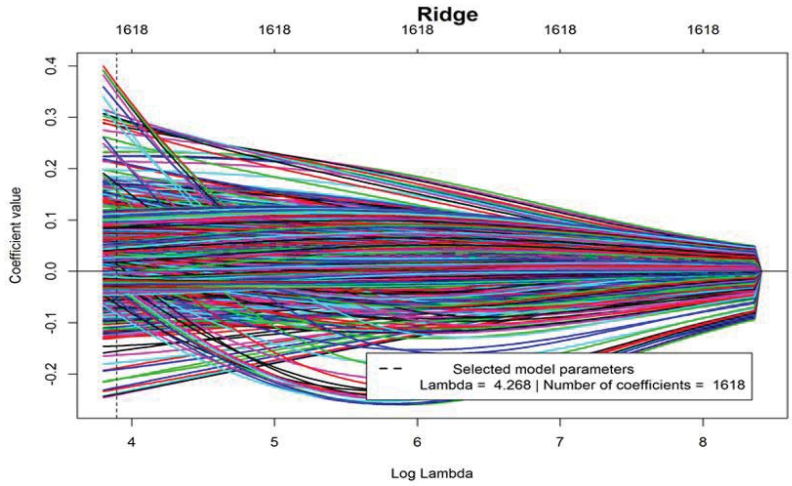

Figure 4.1.1 : Coefficent trace resulting from the Ridge method

Because the field of chemometrics is so closely tied to analytical chemistry it is important to maintain an ability to provide a chemical interpretation of the results at hand. Ridge Regression provides perhaps the least interpretable results of the methods this research employs, a direct consequence of the non – sparse nature of the solutions it generates. However, the results are not completely uninterpretable, the insight is just far less obvious. Despite the condition that all coefficients are non – zero (or all zero in the extreme), model parameter selection still needs to occur, whilst the number of coefficients is constant, their position in terms of significance to the model alters as the λ parameter is tuned. Model parameter selection is, as usual, achieved by taking the parameter that produces a RMSECV one standard error above the minimum value. An example plot of the process may be seen in Figure 4.1.1. The figure shows the paths of coefficient values, each path corresponding to a variable (wavenumber reflectance). It is clear in this image as λ increases, the degree of importance of a variable can be seen to change, this change manifests as pathways crossing. Groups of paths indicate high correlation between the reflectance values of neighbouring wavenumbers within the same band.

Table 4.1.1: Ridge regression results for the initial forage feed samples

Initial forage feed results												
		Metabolisable Energy			NDF			Dry Matter				
	PT	R^2	RMSEP(MJ/kg)	RPD(P)	R^2	RMSEP (%)	RPD(P)	R^2	RMSEP (%)	RPD(P)		
	1	0.73	0.40	1.68	0.84	4.36	2.48	0.35	1.31	1.26		
	2	0.73	0.40	1.59	0.83	4.44	2.36	0.33	1.34	1.20		
	3	0.78	0.35	1.91	0.90	3.40	3.16	0.34	1.33	1.17		
	4	0.77	0.41	1.57	0.90	4.31	2.42	0.44	1.34	1.21		
	5	0.78	0.35	1.89	0.90	3.40	3.13	0.34	1.33	1.18		
	6	0.73	0.39	1.69	0.83	4.43	2.40	0.27	1.41	1.12		
		ADF			Hemicellulose			DOMD			Protein	
PT	R^2	RMSEP (%)	RPD(P)	R^2	RMSEP (%)	RPD(P)	R^2	RMSEP (%)	RPD(P)	R^2	RMSEP (%)	RPD(P)
1	0.66	3.36	1.71	0.92	2.06	3.33	0.73	2.47	1.68	0.54	2.46	1.66
2	0.66	3.36	1.70	0.91	2.17	2.89	0.73	2.51	1.59	0.56	2.38	1.60
3	0.75	2.87	2.00	0.95	1.61	4.03	0.78	2.18	1.91	0.68	1.95	2.01
4	0.75	3.33	1.71	0.95	2.14	2.91	0.78	2.51	1.59	0.68	2.39	1.59
5	0.76	2.83	2.02	0.95	1.62	3.93	0.78	2.21	1.89	0.69	1.94	2.00
6	0.61	3.46	1.65	0.91	2.08	3.20	0.58	2.96	1.34	0.55	2.55	1.45

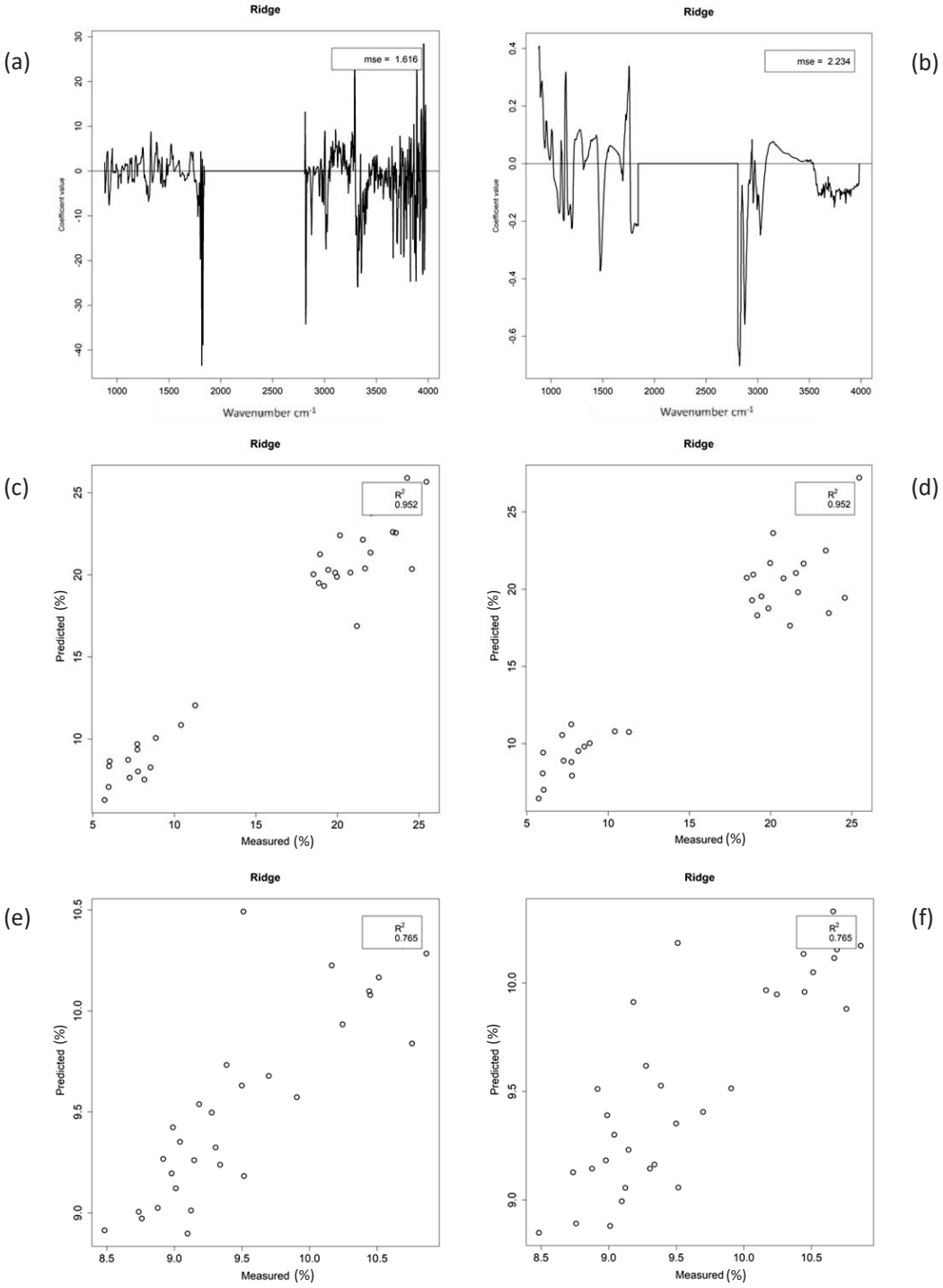

Figure 4.1.2: (a) Hemicellulose PT3 model coefficient plot. (b) Hemicellulose PT4 model coefficient plot. (c) Hemicellulose predicted vs measured values for the PT3 model. (d) Hemicellulose predicted vs measured values for the Pt4 model. Metabolisable energy predicted vs measured plots for PT3 (e) and PT4 (f).

This discussion begins by examining the performance of Ridge regression in the prediction of the chemical compositions of interest for the initial forage feed samples. As will be seen to be a trend for the forage feed samples, hemicellulose and neutral detergent fibre were the best predicted with large R^2 values ($0.95\ [PT3]$,$0.90[PT3]$) and large RPD (4.03,3.16). Hemicellulose and NDF were predicted equally well by Ridge regression models utilising either PT3, PT4 or PT5. What is of particular interest is the contrast in the coefficients each different PT – model has selected. For example, the hemicellulose PT3 model epitomises a lack of interpretability as seen in Figure 4.1.2, the noise is likely the result of the small reflectance values post – SNV and SG. Looking at the RPD it is clear PT3 provides higher quality predictions, specifically in terms of errors and bias. PT4 and PT5 differ only by the presence of a derivative – smoothing step in the pre – processing, examining the NDF and hemicellulose RPD values between these two PT's demonstrates the importance this step plays in removing bias. The PT4 model is relatively more interpretable than PT3 (and PT5), as seen in Figure 4.1.2, for example in terms of magnitude more significance is placed on the C – H type stretching band[162] just below 3000 cm^{-1} followed by bands from the fingerprint region. While chemical interpretability is desirable from a confidence stand point, it is not entirely correct to dismiss a model solely based on poor interpretability, in this instance the predictions of PT3 are more reliable than PT4, despite PT3's poor interpretability. This predictive ability is demonstrated subtly in figure 4.1.2. ME was modelled well by PT4 in terms of R^2 (0.77), yet the same model possesses the second lowest RPD (1.57). The small RPD is a consequence of larger bias and prediction error in the model.

Comparing the prediction plots of PT3 and PT4 (Figure 4.1.2) it is apparent that the predictions are less accurate for the latter, these two models once again have quite contrasting coefficient plots. The PT3 model contains a somewhat distinct contribution occurring around 3200 – 3500 cm^{-1}, but interpretation beyond this is difficult. The PT4 plot again allows for simpler chemical interpretation, and it is immediately clear that the C – H stretch[162] (aliphatic) is significant to the model. There is also a prominent contribution occurring at around 1550 – 1650 cm^{-1}, a region containing C = C and C = O stretches and aromatic stretches, respectively.[162] DOMD is modelled well with an R^2 of 0.78, for PT3, PT4 and PT5 models. The PT4 model possessed the smallest RPD of these three models, which supports the previous suggestion of derivative – smoothing reducing bias. The coefficient plots of both DOMD PT4 and ME PT4 are virtually identical, this result is expected as DOMD and ME are nothing more than scaled forms of one another. ADF was predicted with an R^2 of 0.75 by the PT3 model, the relatively low R^2 is surprising as hemicellulose and NDF were both modelled quite well, and ADF may be calculated

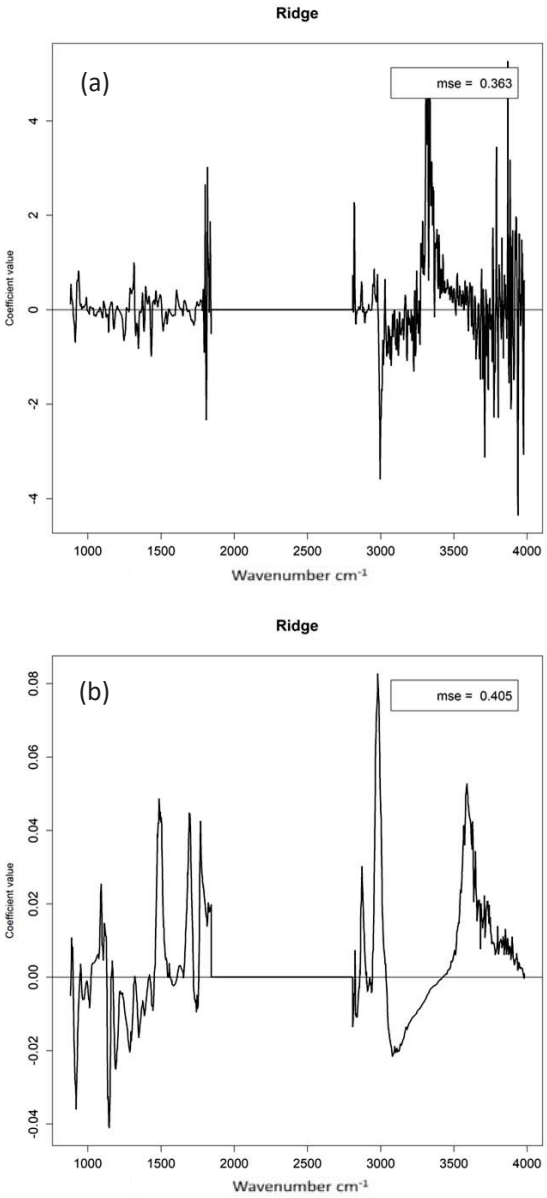

Figure 4.1.3: Coefficient plots of ME for the PT3 (a) and PT4 (b) models.

as the difference between NDF and hemicellulose. However, as NDF contains both ADF and hemicellulose, perhaps the presence of hemicellulose accounts for the well modelled nature of NDF. Protein and DM content represent the worst predicted chemical compositions of the initial forage feed samples with R^2 values of 0.68 (PT4) and 0.44(PT4). The RPD values of the protein models, are low but comparable to ME models, of which have much larger R^2 values. The comparable RPD values imply that although variation in protein content is somewhat poorly accounted for, the bias and errors are relatively low, the same is not seen for DM where low R^2 values are matched by low RPD values. DM describes how much of the plant matter is not water, the difficulties in predicting this chemical composition are suggested to originate, in part, from the fact the samples analysed were freeze dried and as such possessed rather low water content when analysed. Not only was water removed for the most part, but the initial variation in water content may have been due to reasons not related to variations in the chemical structure of the forage sample but rather conditions it was exposed to prior to collection. Overall it appears as though PT3 has produced the best models in terms of R^2 and RPD values, in the context of this sample set. At the other end of the performance spectrum is PT6 which seems to routinely perform poorly, it is possible this is due to improper parameters, allowing some useful information to be distorted.

Table 4.1.2: Ridge regression results for the expanded forage feed samples

Expanded forage feed results									
	NDF			Hemicellulose					
PT	R^2	RMSEP (%)	RPD(P)	R^2	RMSEP (%)	RPD(P)			
1	0.63	6.81	1.60	0.73	3.61	1.90			
2	0.62	6.92	1.46	0.73	3.70	1.78			
3	0.67	6.69	1.46	0.75	3.20	1.79			
4	0.72	6.83	1.49	0.75	3.53	1.74			
5	0.67	6.60	1.49	0.76	3.15	1.87			
6	0.57	7.23	1.43	0.69	3.83	1.79			
	ADF			Metabolisable Energy			Dry Matter		
PT	R^2	RMSEP (%)	RPD(P)	R^2	RMSEP (MJ/kg)	RPD(P)	R^2	RMSEP (%)	RPD(P)
1	0.42	4.25	1.30	0.46	0.48	1.41	0.10	1.55	1.02
2	0.42	4.28	1.17	0.47	0.47	1.34	0.09	1.56	0.89
3	0.50	4.06	1.18	0.52	0.47	1.28	0.30	1.37	1.01
4	0.43	4.24	1.19	0.25	0.48	1.31	0.30	1.56	0.89
5	0.51	4.06	1.18	0.53	0.46	1.33	0.29	1.39	0.99
6	0.39	4.37	1.14	0.47	0.47	1.33	0.13	1.55	0.90

When the initial forage feed sample set was expanded to include a greater variety of species, the general trend of prediction quality was down. Hemicellulose possessed the largest overall R^2 values but relative to the previous sample set was still low $(R^2 : 0.76, PT4)$. NDF was the second best predicted chemical composition of the expanded sample set, this result supports the previous observation of NDF and hemicellulose being the best predicted chemical compositions. The general decline of predictions may result from the introduction of new species or sample growth conditions, development of future models in this context with single species may be more successful.

Table 4.1.3: Ridge regression results for the chicken feed samples

Chicken feed results									
		Crude Fibre			Starch			Fat	
PT	R^2	RMSECV (%)	RPD	R^2	RMSECV (%)	RPD	R^2	RMSECV (%)	RPD
1	0.72	1.29	1.99	0.90	4.58	3.15	0.56	2.35	1.51
2	0.72	1.29	1.99	0.89	4.91	2.94	0.55	2.38	1.49
3	0.84	1.04	2.47	0.90	4.66	3.09	0.52	2.47	1.43
4	0.84	1.11	2.31	0.89	4.99	2.89	0.47	2.62	1.35
5	0.84	1.04	2.49	0.88	5.08	2.84	0.46	2.63	1.34
6	0.73	1.28	2.02	0.89	4.79	3.02	0.31	2.94	1.20
		Nitrogen			Ash			Dry Matter	
PT	R^2	RMSECV (%)	RPD	R^2	RMSECV (%)	RPD	R^2	RMSECV (%)	RPD
1	0.87	0.49	2.77	0.93	1.23	4.13	0.69	0.83	1.80
2	0.88	0.45	2.99	0.92	1.27	3.98	0.70	0.82	1.80
3	0.87	0.51	2.68	0.94	1.15	4.42	0.73	0.76	1.95
4	0.88	0.50	2.74	0.94	1.11	4.57	0.73	0.76	1.95
5	0.88	0.50	2.74	0.94	1.11	4.59	0.75	0.74	2.01
6	0.88	0.48	2.83	0.93	1.16	4.39	0.66	0.84	1.78
		Calcium			Phosphorus			Gross Energy	
PT	R^2	RMSECV (mg/g)	RPD	R^2	RMSECV (mg/g)	RPD	R^2	RMSECV (kJ/g)	RPD
1	0.93	4.72	4.20	0.92	2.01	4.00	0.64	0.61	1.56
2	0.92	4.92	4.04	0.93	1.88	4.31	0.58	0.64	1.48
3	0.94	4.25	4.69	0.94	1.77	4.56	0.75	0.49	1.96
4	0.95	4.04	4.92	0.94	1.71	4.72	0.76	0.48	1.99
5	0.95	3.99	4.99	0.95	1.67	4.83	0.75	0.49	1.95
6	0.93	4.43	4.50	0.93	1.84	4.38	0.64	0.60	1.60

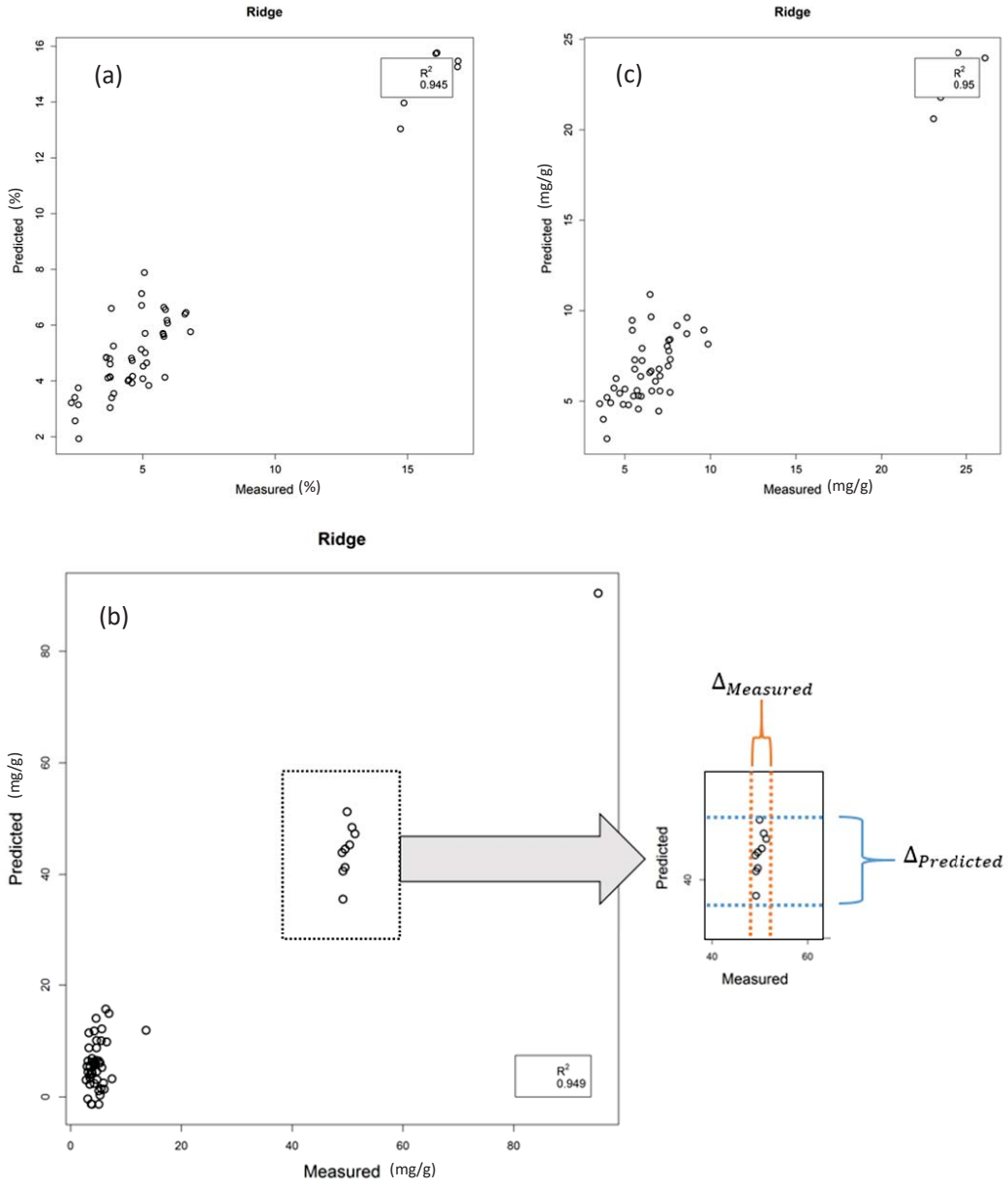

Figure 4.1.4: Predicted vs measured plots of Ash (a), calcium (b) and phosphorus (c). (b) also includes an illustration of the high variability of predictions.

Examining the R^2 values for the various chemical component models relating to the chicken feed samples, it is obvious that ash possesses one of the largest R^2 values $(0.94, PT3)$. However, in this instance it is only upon examining the corresponding graph (figure 4.1.4) of measured and predicted values that it becomes clear these predictions may in fact not be as good as the R^2 or even RPD implies. The issue is that there are distinct clusters of values, each cluster could be considered as a *pseudo* – point on a plot whereby the R^2 corresponds to that of a line passing close to both pseudo – points. Such a scenario is referred to as having high leverage, inducing *homoscedastic* clusters (equal variance about the centroid of the cluster). This leverage effect in this example is not as large as will be seen in later results. The prediction of calcium content in the chicken feed represents a second, and more distinct example of leverage related issues. Whilst the R^2 value is large ($R^2: 0.95, PT4$) it is undeniable that if one was to isolate any one of the clusters there would be little useful prediction information available, as figure 4.1.4 indicates. The figure explicitly demonstrates that the variation in the predicted axes is far larger than that of the measured. Phosphorus content was the third and final chemical composition to demonstrate large R^2 $(0.95, PT5)$, as a result of leveraging as seen in figure 4.1.4 .The impact of the leveraging is not quite as significant as seen in the calcium example. Leveraging issues demonstrate the benefit of visual inspection of the prediction plots. Examining the results of PT3, PT4 and PT5 for calcium and phosphorus they demonstrate higher RPD for PT4 relative to the other two. This increase in RPD is the opposite trend to what was observed in the previous two sample sets, and may be another subtler consequence of the issues outlined for these chemical compositions.

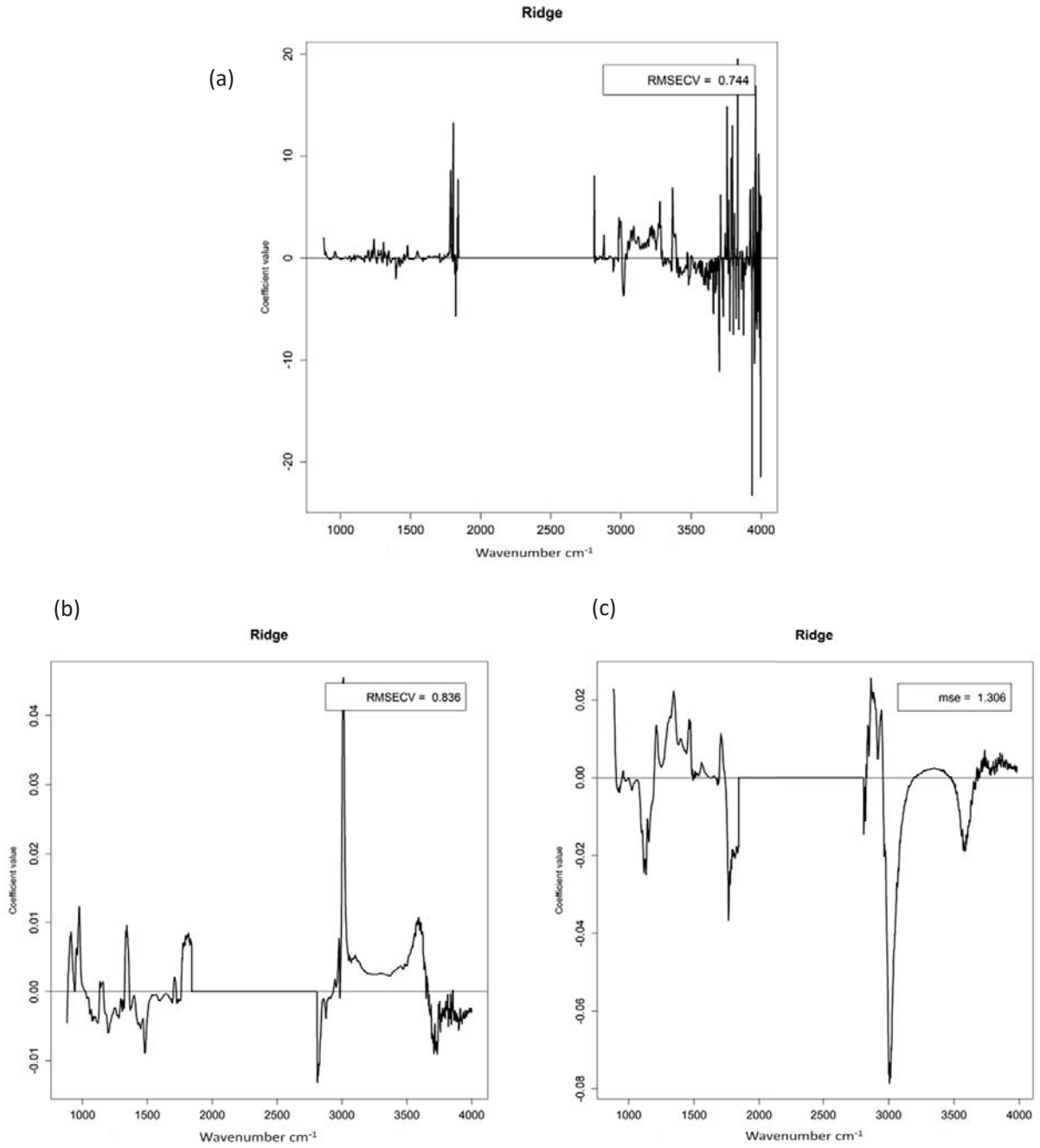

Figure 4.1.5: (a) coefficient plot of the PT5 DM model. Coefficient plots of the DM PT1 models for the chicken feed samples (b) and the initial forage feed (c)

Crude fibre was predicted rather well (R^2 : 0.84, $PT3$) and devoid of any of the issues encountered with the previous two models. DM was predicted better than in any previous attempt (R^2: 0.75, $PT5$). To gain some insight as to why DM is predicted more accurately here compared to the forage samples, the first avenue of exploration would be the coefficient plots. Unfortunately, PT5 is a derivative based model and interpretation of the contributions is difficult as seen in figure 4.1.5. However, at the price of a decrease in model quality we can utilise the coefficient plots of a more interpretable model, for example the PT1 model (R^2: 0.69,). Comparing the coefficient plot of PT1 from the grass model to the chicken feed model they appear relatively similar in terms of the largest absolute value regions. Beyond the location of strong contributions at ~3000 cm^{-1} and 3500 cm^{-1} to both models, it is difficult to single out any particular absorption contributions. The geometry of the chicken feed DM model aliphatic C – H stretch contribution appears sharper than the somewhat broad peak in the grass coefficient plot. The fact that the plots are so similar makes it difficult to assert differences in performance are originating from the differences in the model, it is

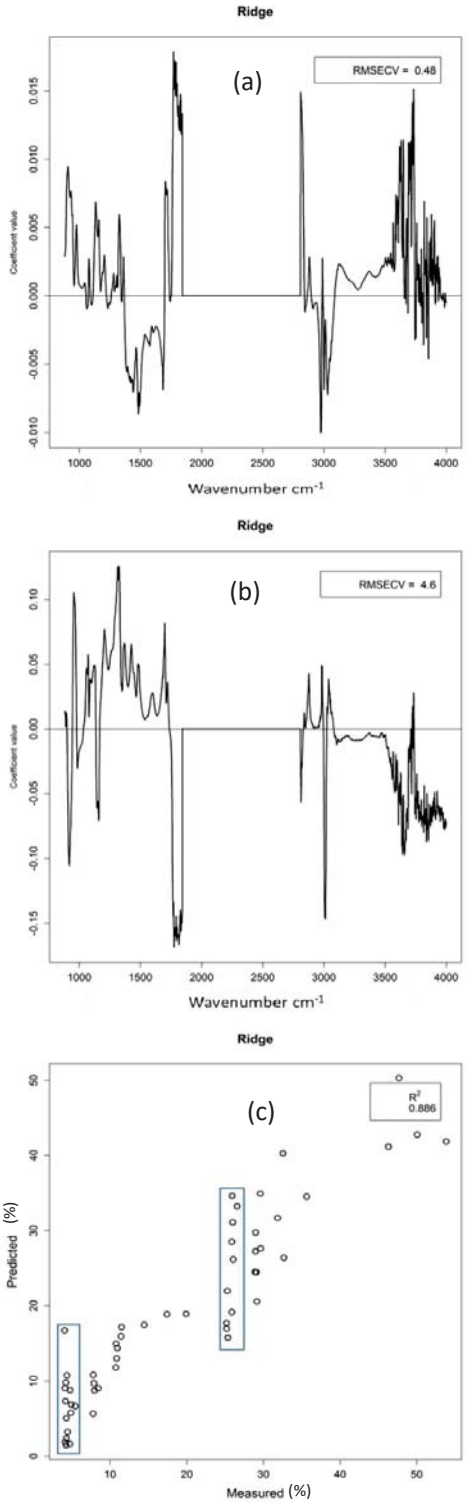

Figure 4.1.6: The coefficient plots for nitrogen (a) and starch (b) PT1 – models. (c) Illustration of collimation in predicted vs measure plot for starch.

possible that some of the apparent superior performance arises from the fact that the chicken feed data is cross – validation data whilst the grass is external validation.

Nitrogen was modelled well by all models ($R^2 \cong 0.88$), examining the coefficient plot of PT1 in figure 4.1.6 it can be seen than there is more similarity between the magnitude of coefficients around the amide I and II regions, and the C – H aliphatic stretch. This is a logical result as it is expected that the importance of the amide regions would increase when predicting compositions of which should absorb in the amide I and II bands. Starch was modelled well (R^2: 0.90, RPD: 3.15 $PT1$), the coefficient plot (figure 4.1.6) has the most prominent contribution once again from the aliphatic C – H stretching[162] at 3000 cm^{-1}, subsequent significant contributions seem to originate from the fingerprint region. The prediction plot for starch (figure 4.1.6) possesses columns of data with narrow variation in the measured value and large prediction variation, these regions are clearly a weakness of this particular model. Gross energy is predicted quite well by PT3, PT4 and PT5 all of which possess $R^2 \cong 0.75$ and RPD$\cong$ 1.96. The Gross Energy results demonstrate the need for combinations of pre – processing methods beyond purely scatter correction, this can be observed by examining the increase in prediction quality when comparing PT1 or PT2 to any of PT3, PT4 and PT5. Fat stands out in this sample set as the worst predicted chemical composition (R^2: 0.44, $PT2$)

Table 4.1.4: Ridge regression results for the faecal samples

Faecal results									
	Hemicellulose			Dry Matter			Ash		
PT	R^2	RMSECV (%)	RPD	R^2	RMSECV (%)	RPD	R^2	RMSECV (%)	RPD
1	0.81	2.92	2.23	0.44	0.46	1.00	0.88	2.32	2.81
2	0.78	3.17	2.05	0.44	0.46	0.99	0.84	2.67	2.31
3	0.86	2.67	2.54	0.48	0.45	1.07	0.88	2.45	2.78
4	0.85	2.74	2.07	0.52	0.43	1.07	0.88	2.41	2.81
5	0.85	2.71	2.56	0.51	0.44	1.07	0.88	2.42	2.79
6	0.79	3.09	2.11	0.47	0.44	1.03	0.88	2.33	2.76
	Nitrogen			Gross Energy			Organic Matter		
PT	R^2	RMSECV (%)	RPD	R^2	RMSECV (%)	RPD	R^2	RMSECV (%)	RPD
1	0.84	0.40	2.37	0.94	0.50	3.98	0.88	2.20	2.85
2	0.84	0.41	2.34	0.92	0.52	3.88	0.83	2.62	2.38
3	0.87	0.37	2.56	0.93	0.52	3.87	0.87	2.31	2.71
4	0.87	0.36	2.66	0.93	0.51	3.88	0.88	2.26	2.77
5	0.87	0.37	2.58	0.93	0.52	3.85	0.88	2.27	2.75
6	0.83	0.41	2.33	0.92	0.52	3.82	0.88	2.22	2.82
	NDF			ADF			Lignin		
PT	R^2	RMSECV (%)	RPD	R^2	RMSECV (%)	RPD	R^2	RMSECV (%)	RPD
1	0.84	4.71	2.45	0.82	2.21	2.36	0.80	0.47	2.34
2	0.82	4.97	2.32	0.81	2.30	2.27	0.78	0.50	2.21
3	0.88	4.38	2.64	0.83	2.25	2.33	0.80	0.49	2.23
4	0.87	4.48	2.58	0.82	2.30	2.27	0.79	0.51	2.14
5	0.88	4.42	2.61	0.83	2.26	2.31	0.80	0.50	2.22
6	0.80	5.30	2.18	0.79	2.46	2.13	0.80	0.48	2.27

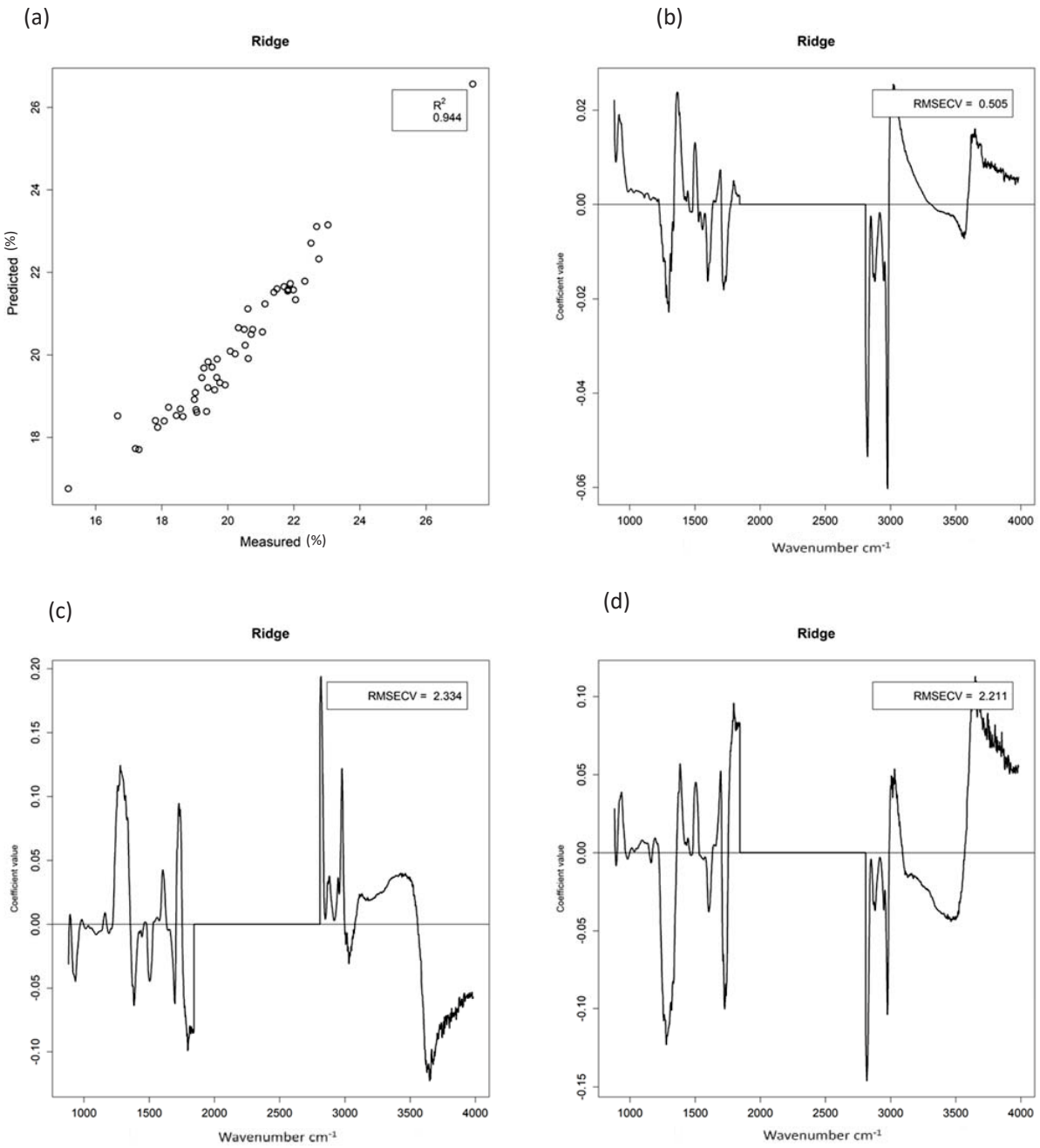

Figure 4.1.7: (a) the predicted vs measured plot for gross – energy. (b) coefficient plot for gross – energy PT1 model. Coefficient plots for Ash (c) and OM (d) PT1 – models.

Examining the Ridge regression results for the faecal samples, GE represents the best predicted chemical composition. GE was modelled with the largest R^2 (0.94) and RPD (3.98) in the PT1 based model, although all PT – models are high quality. The prediction plot in figure 4.1.7, illustrates no presence of leveraging in this chemical component and the coefficient plot illustrates the importance of C – H stretching to the model. The PT1 based models for ash and OM exhibited the largest R^2 $(0.88, 0.88)$, their coefficient plots are interesting as they are nearly identical to one another aside from a change in sign. This highlights a fundamental negative relationship between the two, as OM increases ash decreases which follows from the fact ash is proportional to inorganic matter content. The faecal sample set is the first to contain predictions of near equivalent quality for both NDF and ADF $(R^2: 0.87, 0.82, PT4)$. Since NDF includes ADF it is reasonable to have expected to see this trend occur among the other sample sets, however, that has not been the case. As to why this has not been the case it has previously been speculated that the portion of NDF not contained in

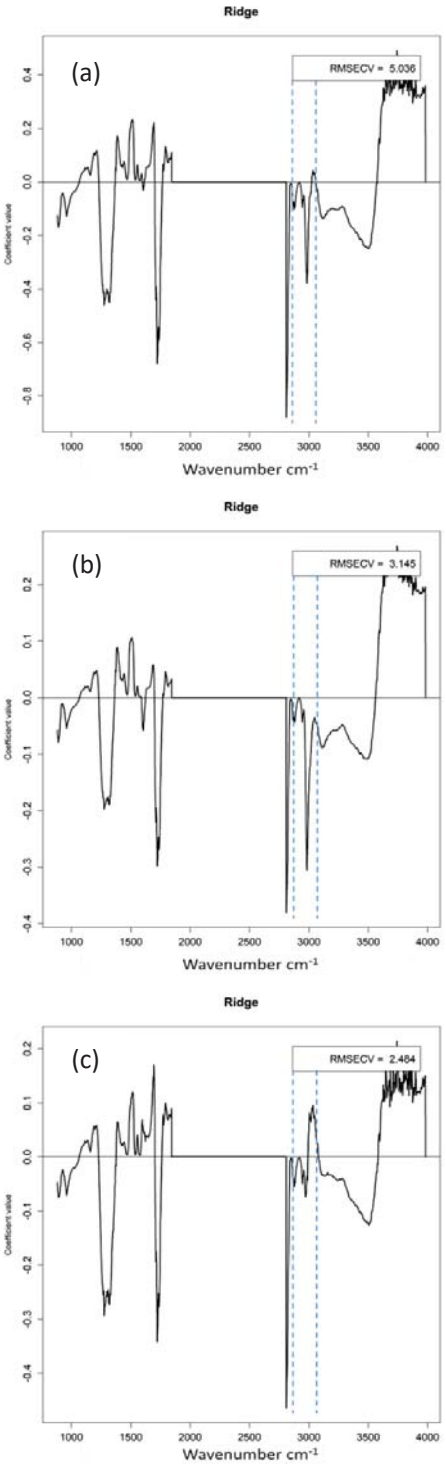

Figure 4.1.8: Coefficient plots of NDF (a), hemicellulose (b) and ADF (c).

ADF was accounting for NDF's high quality modelling. Upon first examination of the figure 4.1.8, which displays the coefficient plots of NDF, hemicellulose and ADF for PT2, they all appear very similar. The dashed blue lines represent the approximate region where most of the changes between plots occurs. The changes in this region occur predominantly at ~3000 cm^{-1}, the aliphatic C – H stretch region, implying this region provides useful information for the model to differentiate between hemicellulose, ADF and NDF. Nitrogen was predicted most accurately using a PT4 based model. Again the coefficient plot of a nitrogen content model demonstrated the logical shift of importance towards the region of amide I and II absorption bands, as seen in figure 4.1.9. Lignin was predicted to quite a high standard by all PT's, all of which produced very similar R^2 and RPD values.

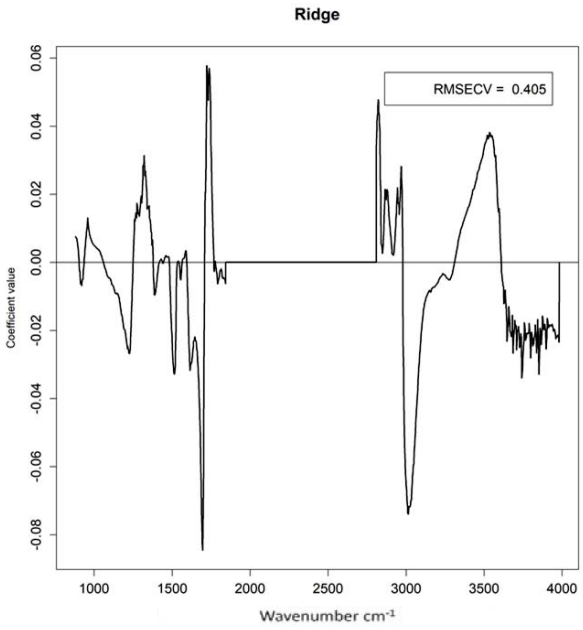

Figure 4.1.9: Coefficient plot for the PT4 – model of nitrogen.

Ridge was unable to successfully predict nickel content in hyperaccumulators using infrared reflectance information, with any pre – treatment based model. Of the different models, PT2 stands out in terms of R^2 as being substantially lower than the others.

Table 4.1.5: Ridge regression results for the hyperaccumulator samples

	Hyperaccumulators results		
PT	R^2	RMSECV (%)	RPD
1	0.30	0.31	1.21
2	0.08	0.36	0.92
3	0.34	0.30	1.24
4	0.34	0.30	1.24
5	0.35	0.29	1.25
6	0.35	0.29	1.25

4.2 Least Absolute Shrinkage and Selection Operator (LASSO)

LASSO regression will generate models possessing far fewer variable contributions than the Ridge method, which allows for easier chemical interpretation. The coefficient value traces also become more interesting as pathways will rapidly tend to zero if they are non – essential to the prediction. One of the short comings of this technique is that the selection of coefficients is carried out indiscriminately, in practice this results in the selection of a coefficient from a highly correlated cluster without preference. For infrared spectroscopy this may manifest as selection of one wavenumber within a band, or the selection of one wavenumber within two highly correlated bands. The risk of either scenario is the misplacement of importance on a

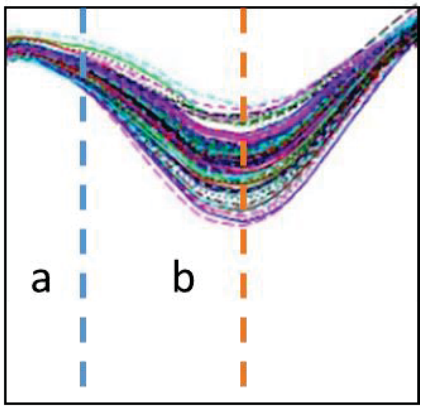

Figure 4.2.1: Example of difference between variance at wavenumber a vs b.

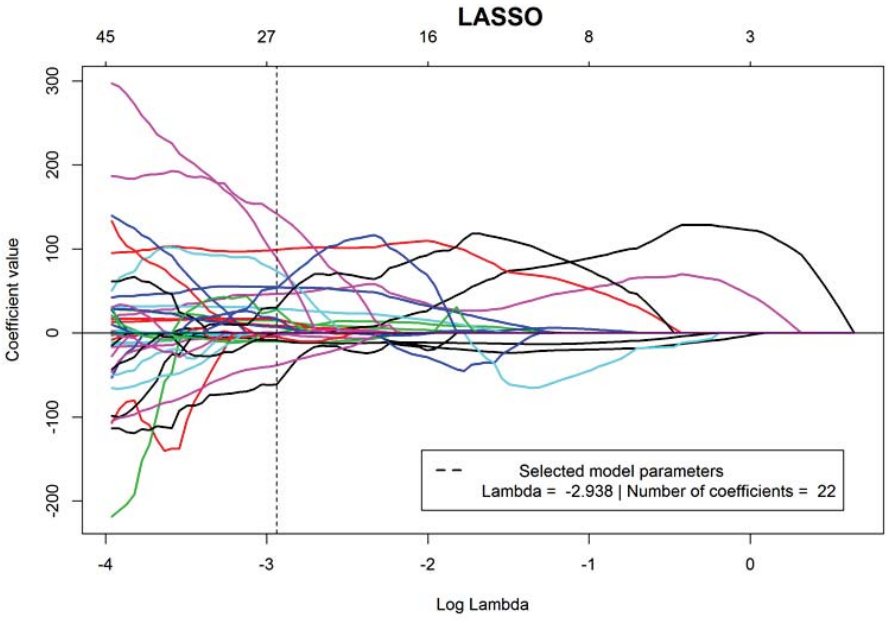

Figure 4.2.2: Coefficient trace resulting from using the LASSO algorithm.

variable which is less sensitive to chemical change. An example of the consequence of indiscriminate variable selection from within a band may be seen in Figure 4.2.1, where it is clear if wavenumber a was selected it would possess less variation between samples than wavenumber b. In the trace plot of coefficients (figure 4.2.2), the selection process is seen to force values to zero if they are unimportant to the model. The importance of a variable refers to whether or not its inclusion assists in reducing the error of prediction for the model. Likewise, any line that resists trending towards zero is associated with a wavenumber of which is deemed important to the model.

Table 4.2.1: LASSO regression results for the initial forage feed samples

Initial forage feed results												
	Metabolisable Energy				NDF				Dry Matter			
PT	*n*	R^2	RMSEP (MJ/kg)	RPD(P)	*n*	R^2	RMSEP (%)	RPD(P)	*n*	R^2	RMSEP (%)	RPD(P)
1	21	0.76	0.36	1.86	30	0.88	3.76	2.88	34	0.26	1.57	1.05
2	26	0.75	0.36	1.79	30	0.89	3.65	2.69	36	0.34	1.45	1.09
3	25	0.67	0.41	1.63	19	0.85	4.12	2.57	28	0.37	1.31	1.21
4	14	0.84	0.38	1.71	19	0.85	3.54	2.84	33	0.05	1.49	1.06
5	12	0.62	0.45	1.49	17	0.87	3.98	2.64	31	0.30	1.37	1.14
6	19	0.70	0.39	1.70	31	0.85	4.33	2.37	34	0.03	3.65	0.44

	ADF				Hemicellulose				DOMD					Protein			
PT	*n*	R^2	RMSEP (%)	RPD(P)	*n*	R^2	RMSEP (%)	RPD(P)	*n*	R^2	RMSEP (%)	RPD(P)		*n*	R^2	RMSEP (%)	RPD(P)
1	25	0.66	3.24	1.77	29	0.92	2.10	3.27	21	0.76	2.23	1.86	1	19	0.60	2.40	1.70
2	30	0.67	3.19	1.76	18	0.93	1.85	3.52	26	0.75	2.26	1.79	2	22	0.62	2.34	1.74
3	34	0.67	3.19	1.79	23	0.95	1.57	4.21	24	0.67	2.55	1.63	3	21	0.70	1.93	1.97
4	31	0.67	3.14	1.75	18	0.95	2.07	3.04	14	0.67	2.38	1.71	4	21	0.70	2.33	1.75
5	40	0.67	3.16	1.81	37	0.93	1.86	3.45	12	0.62	2.78	1.49	5	13	0.67	2.03	1.84
6	43	0.63	3.41	1.66	28	0.91	2.13	3.00	18	0.70	2.44	1.70	6	33	0.49	3.18	1.24

Table 4.2.1 shows the LASSO results for the initial forage feed samples. In the intial forage feed sample set analysis, NDF and hemicellulose were predicted most accurately (R^2: 0.89, 0.95) with PT2 and PT4 models respectively. ADF was once again predicted with less accuracy (R^2: 0.66 $PT1$) than NDF or ADF, as seen for the equivalent Ridge results. Using the PT1 models for each of the three, ADF contains the fewest coefficients (14) whilst NDF contains the most (27), figure 4.2.3 displays the variables associated with these coefficients. The hemicellulose plot (figure 4.2.3) has three distinct peaks outside of the fingerprint region. The two peaks around 2700 – 2900 cm^{-1} are clearly of C – H stretching origin, and the peaks at ~3500 cm^{-1} originate from the O – H band.[161] The C – H and O – H stretching peaks are present to some extent in all models. However, the O – H peak is most prominent in the NDF and ADF models. Annotated in the figures are several peaks of which possess relatively larger coefficient values. The peaks at ~1750 cm^{-1} are significant in both the

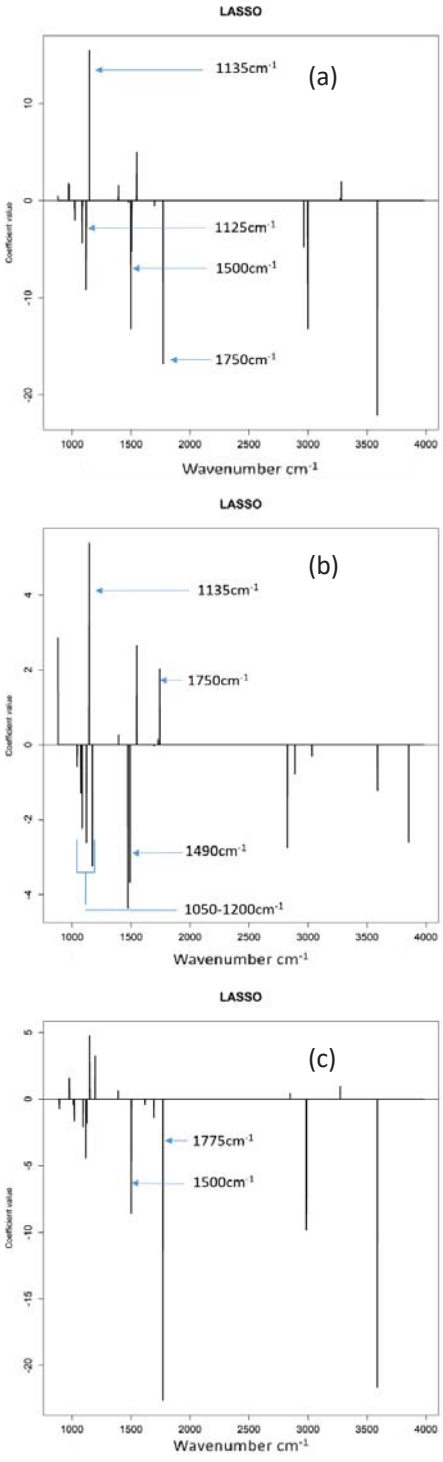

Figure 4.2.3: (a) NDF coefficient plot for PT1. (b) Hemicellulose coefficient plot for PT1. (c) ADF coefficient plot for PT1

NDF and ADF models, as well as the peak at 1500 cm^{-1}, which appears also in the hemicellulose model. The peak at 1750 cm^{-1} may originate from the C = O stretch of a saturated ester.[162] Amongst the literature referenced as due to several potential sources[162] one of which is lignin, since lignin is a constituent of ADF, this result is sensible. The peak at ~1500 cm^{-1} is present in all models, cited in the literature as due to C = C aromatic stretching, as such this is a sensible inclusion across all models. The hemicellulose model shares peaks with NDF at 1135 cm^{-1} and a cluster between 1050 – 1200 cm^{-1}. The peak at 1135 cm^{-1} is likely due to C – O – C asymmetric stretching[162] associated with β – 1.4 glucan, which is a component of hemicellulose. Most of the shared peaks within the 1050 – 1200 cm^{-1} range (excluding the 1135 cm^{-1}) lie around 1075 – 1100 cm^{-1}. This region contains contributions from xyloglucan and cutin, the former of which is explicitly related to hemicellulose content.

The expanded forage feed sample set (table 4.2.2) again shows a general decline of prediction accuracy across all the chemical components. The PT4 performs well in both NDF and hemicellulose predictions relative to other PT – models. For the sake of comparison, the NDF and hemicellulose PT1 model coefficient plots for the present sample set (bottom) and previous forage feed sample set (top) are displayed in figure 4.2.4. The O – H and aliphatic C – H stretching regions are present in the NDF and hemicellulose models of the expanded sample set. The NDF model lacks either of the 1135 cm^{-1} or 1750 cm^{-1} contributions present in the equivalent model of the previous sample set. The 1500cm^{-1} peak and peaks between 1050 – 1150 cm^{-1} are more exaggerated in the NDF model of the expanded forage feed sample set. The hemicellulose model contains more similarities with its previous sample set model. The

significant peaks of the previous model are still present, however, reduced in size relative to the aliphatic C – H and O – H stretching regions. Hemicellulose was modelled with greater accuracy than NDF, which may be the reason why its coefficient plot aligns better to that of the previous sample set.

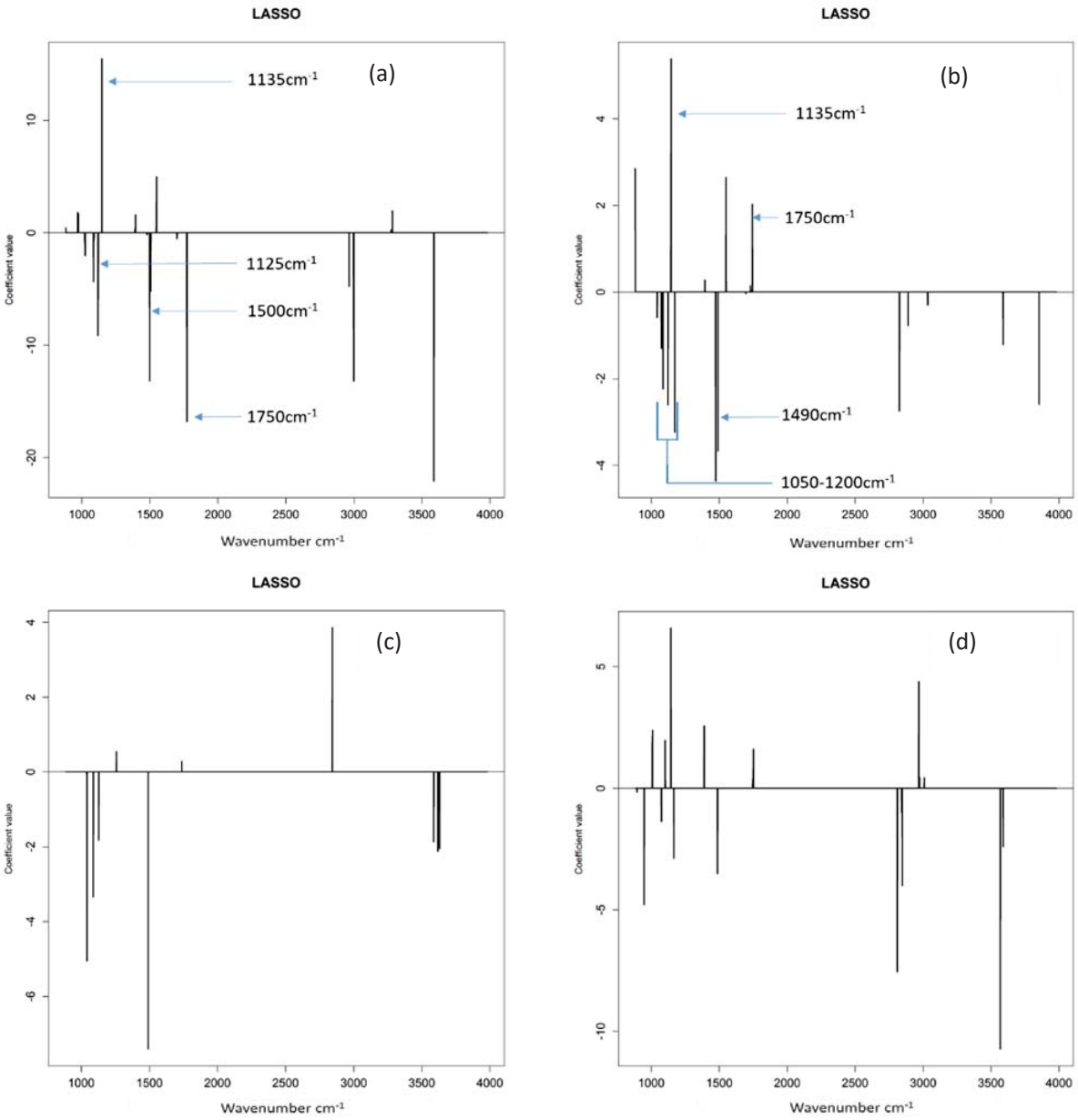

Figure 4.2.4: (a) and (b) are the coefficient plots of NDF and Hemicellulose in the initial forage sample set. (c) and (d) NDF and Hemicellulose models for the expanded sample set. All plots are for PT1 models.

Table 4.2.2: LASSO regression results for the expanded forage feed samples

Expanded forage feed results												
	NDF				Hemicellulose							
PT	n	R^2	RMSEP (%)	RPD(P)	n	R^2	RMSEP (%)	RPD(P)				
1	10	0.64	6.84	1.60	22	0.70	3.45	1.87				
2	11	0.64	6.79	1.49	21	0.72	3.36	1.89				
3	7	0.66	7.35	1.31	50	0.76	3.15	1.90				
4	11	0.66	6.78	1.48	22	0.76	3.30	1.92				
5	8	0.64	7.45	1.29	21	0.76	3.12	1.91				
6	16	0.53	7.60	1.37	39	0.72	3.43	1.70				
	ADF				Metabolisable Energy				Dry Matter			
PT	n	R^2	RMSEP (%)	RPD(P)	n	R^2	RMSEP (MJ/kg)	RPD(P)	n	R^2	RMSEP (%)	RPD(P)
1	17	0.49	3.92	1.41	14	0.51	0.45	1.48	23	0.18	1.54	1.03
2	14	0.41	4.22	1.21	15	0.55	0.43	1.50	6	0.11	1.53	0.92
3	24	0.46	4.18	1.14	15	0.53	0.47	1.29	30	0.19	1.46	0.97
4	16	0.33	4.08	1.27	16	0.31	0.45	1.44	30	0.19	1.55	0.94
5	24	0.40	4.33	1.13	13	0.51	0.48	1.25	47	0.21	1.45	1.01
6	8	0.39	4.55	1.06	20	0.50	0.46	1.40	4	0.13	1.53	0.92

Table 4.2.3: LASSO regression results for the chicken feed samples

Chicken feed results												
	Crude Fibre				Starch				Fat			
PT	*n*	R^2	RMSECV (%)	RPD	*n*	R^2	RMSECV (%)	RPD	*n*	R^2	RMSECV (%)	RPD
1	9	0.81	1.06	2.43	9	0.93	3.80	3.79	13	0.62	2.19	1.61
2	6	0.77	1.16	2.23	9	0.93	3.92	3.68	18	0.59	2.27	1.55
3	15	0.87	0.90	2.86	17	0.91	4.28	3.37	15	0.49	2.53	1.40
4	18	0.82	0.87	2.49	15	0.92	4.11	3.51	7	0.42	2.77	1.28
5	16	0.85	1.00	2.57	15	0.91	4.29	3.36	9	0.45	2.66	1.33
6	8	0.78	1.15	2.23	10	0.94	3.66	3.94	13	0.46	2.58	1.37
	Nitrogen				Ash				Dry Matter			
PT	*n*	R^2	RMSECV (%)	RPD	*n*	R^2	RMSECV (%)	RPD	*n*	R^2	RMSECV (%)	RPD
1	9	0.90	0.43	3.20	19	0.96	0.86	5.91	8	0.74	0.77	1.92
2	5	0.91	0.44	3.10	22	0.96	0.87	5.86	5	0.73	0.78	1.90
3	12	0.91	0.43	3.16	11	0.97	0.74	6.86	10	0.73	0.79	1.89
4	19	0.91	0.41	3.31	18	0.98	0.65	7.81	8	0.71	0.83	1.78
5	11	0.90	0.45	2.99	10	0.98	0.68	7.47	10	0.73	0.77	1.94
6	4	0.91	0.45	3.01	30	0.97	0.76	6.72	10	0.74	0.75	1.98
	Calcium				Phosphorus				Gross Energy			
PT	*n*	R^2	RMSECV (mg/g)	RPD	*n*	R^2	RMSECV (mg/g)	RPD	*n*	R^2	RMSECV (kJ/g)	RPD
1	20	0.97	3.18	6.26	15	0.96	1.37	5.90	6	0.72	0.52	1.84
2	18	0.97	3.22	6.20	17	0.97	1.09	7.42	7	0.68	0.55	1.73
3	21	0.99	2.06	9.67	13	0.97	1.11	7.28	18	0.71	0.53	1.78
4	14	0.98	2.39	8.34	12	0.98	1.06	7.62	13	0.75	0.50	1.92
5	12	0.98	2.31	8.65	9	0.98	1.06	7.64	11	0.74	0.51	1.87
6	22	0.97	2.70	7.38	22	0.97	1.13	7.16	8	0.76	0.47	2.02

The chemical compositions of ash, calcium and phosphorus were predicted with exceptionally large R^2 values for the chicken feed sample set. Previously it has been noted that these three chemical compositions are at risk of artificially large R^2 values, based on the distinct clustering of the response values. LASSO has, however, managed to perform better than Ridge regression in terms of reducing the size of the predicted value variation for ash and phosphorus. Ash and phosphorus are much more reliably described by the R^2 value associated with them, than calcium is, this is supported by the fact calcium has relatively large RMSECV. The starch model also generated a large R^2 value (0.93, PT1). Starch was modelled more accurately using LASSO when compared with Ridge, with increases in both RPD and R^2. Unfortunately, as can be seen in figure 4.2.5 this model still has distinct regions where measured variation is far lower than the variation in prediction. Nitrogen was modelled well with a large R^2 (0.90) and moderate RPD, and like the starch models, demonstrated improved prediction compared with the Ridge results. The PT2 model utilised only 5 variables in building this model, which is a strikingly low number. The coefficient plot of these variables may be seen in figure 4.2.5. It is interesting to note the lack of any variable input from outside the fingerprint region. Previously tentative connections have been made with Ridge nitrogen models and amide I and II absorption bands, for this particularly sparse model, the connection is more distinct. The majority of the variables this model uses lie between the approximate region of 1400 – 1700 cm^{-1}, with most of the variables closer to the 1400 cm^{-1} side and hence amide II band[162] (C – N stretch and N – H bend).

Crude fibre was most accurately predicted in the PT3 model, as it possessed the largest R^2 (0.87) and RPD (2.86). The coefficient plot for the crude fibre PT3 model (figure 4.2.5) reveals significant contributions with the region of 3700 – 3900 cm^{-1} as well as a prominent peak at 1335 cm^{-1}. The peak at 1335 cm^{-1} is likely due to C – H bending, associated with cellulose.[162] The region of 3700 – 3900 cm^{-1} contains some peaks which perhaps may be associated with the tail end of the O – H stretch region.[160]

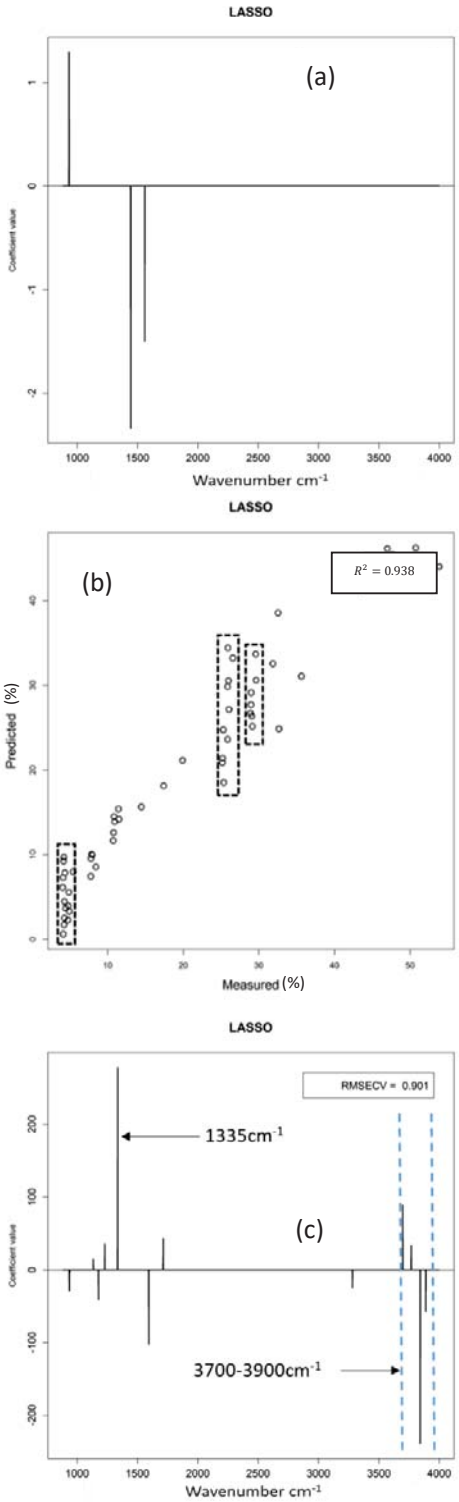

Figure 4.2.5: (a) PT2 Nitrogen model coefficient plot. (b) Predicted vs measured value plot, for starch (PT1). (c) Coefficient plot of the crude fibre model (PT3).

Table 4.2.4: LASSO regression results for the faecal samples

Faecal results												
			Hemicellulose				Dry Matter				Ash	
PT	n	R^2	RMSECV (%)	RPD(CV)	n	R^2	RMSECV (%)	RPD(CV)	n	R^2	RMSECV (%)	RPD(CV)
1	12	0.85	2.52	2.59	17	0.65	0.36	1.69	15	0.95	1.50	4.33
2	9	0.83	2.74	2.38	14	0.65	0.35	1.70	26	0.90	2.01	3.23
3	12	0.84	2.97	1.25	19	0.71	0.33	1.82	14	0.91	2.02	3.20
4	20	0.86	2.48	1.16	12	0.64	0.36	1.68	19	0.91	1.93	3.36
5	20	0.85	2.60	1.18	10	0.49	0.44	1.38	20	0.91	1.90	3.42
6	9	0.80	3.01	2.16	12	0.65	0.36	1.68	27	0.88	2.23	2.92
			Nitrogen				Gross Energy				Organic Matter	
PT	n	R^2	RMSECV (%)	RPD(CV)	n	R^2	RMSECV (%)	RPD(CV)	n	R^2	RMSECV (%)	RPD(CV)
1	12	0.83	0.39	2.43	14	0.96	0.40	4.94	15	0.93	1.64	3.83
2	15	0.82	0.41	2.35	12	0.95	0.38	5.21	18	0.91	1.85	3.39
3	13	0.83	0.41	2.32	14	0.93	0.48	4.15	18	0.88	2.19	2.86
4	14	0.86	0.37	2.54	17	0.93	0.46	4.35	19	0.90	1.98	3.16
5	14	0.83	0.41	2.30	20	0.93	0.48	4.20	19	0.86	2.37	2.65
6	19	0.83	0.41	2.32	16	0.96	0.38	5.30	10	0.87	2.31	2.71
			NDF				ADF				Lignin	
PT	n	R^2	RMSECV (%)	RPD(CV)	n	R^2	RMSECV (%)	RPD(CV)	n	R^2	RMSECV (%)	RPD(CV)
1	13	0.88	4.10	2.82	11	0.84	2.12	2.47	12	0.82	0.45	2.44
2	13	0.87	4.17	2.76	14	0.84	2.13	2.46	8	0.82	0.46	2.38
3	21	0.83	4.74	2.44	14	0.80	2.44	2.14	12	0.79	0.53	2.09
4	19	0.85	4.45	2.60	12	0.82	2.31	2.26	20	0.74	0.55	2.01
5	26	0.84	4.64	2.49	14	0.81	2.37	2.21	8	0.79	0.54	2.03
6	13	0.82	4.96	2.33	8	0.73	2.74	1.90	6	0.74	0.58	1.90

In the faecal sample set, as was apparent in the Ridge results, gross energy is the most accurately predicted composition. Gross energy was predicted with slightly larger R^2 than the Ridge analysis generated, and larger RPD, with all models possessing RPD > 4. Organic matter and ash also demonstrated marginal increases in R^2 values, in comparison to the Ridge results. This perhaps indicates that LASSO has successfully removed variable contributions to the model, which were obscuring the link between organic matter and MIR reflectance values. Interestingly the Ridge results for both OM and ash demonstrated PT2 to possess the lowest R^2 and RPD. Whilst the LASSO results position PT2 as one of the better pre – treatments, the PT1 based model provided the highest quality predictions of ash and organic matter with the highest R^2 and RPD values. Hemicellulose, ADF and NDF were all predicted with respectable accuracy by most PT's with PT1 possessing the largest values of R^2 and RPD. Previously it has been

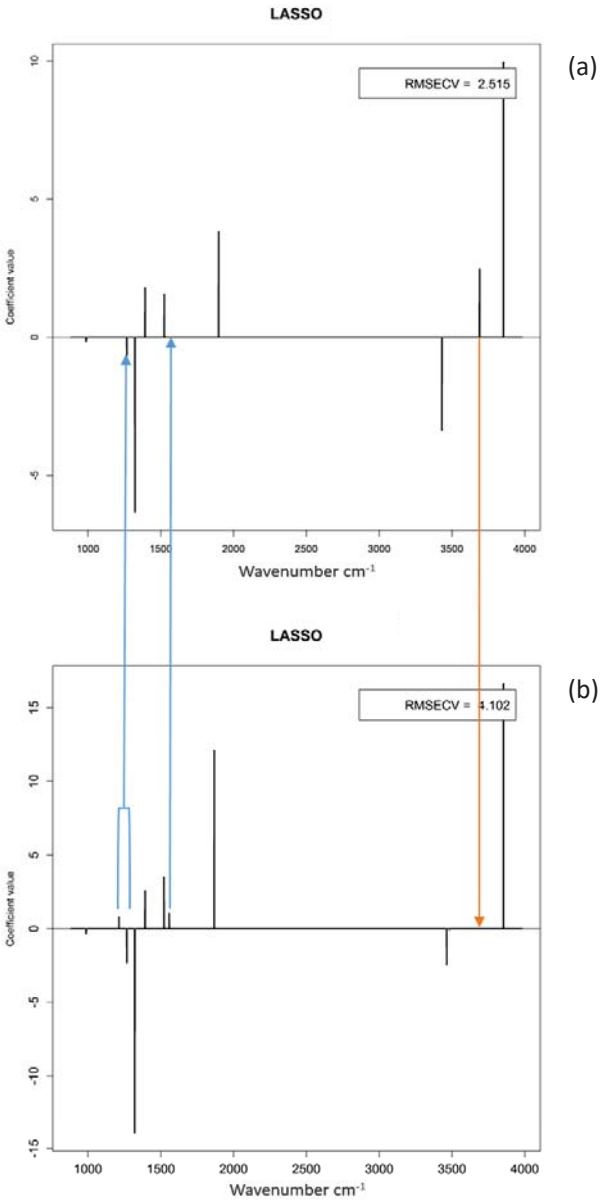

Figure 4.2.6: (a) NDF coefficient plot for PT1. (b) Hemicellulose coefficient plot for PT1. Blue and orange arrows highlight the differences between the two plots.

mentioned that it may be the hemicellulose fraction of NDF that is accounting for the well modelled nature of NDF relative to ADF. Whilst in this instance ADF is quite well predicted, examining the coefficient plots in figure 4.2.7 of hemicellulose and NDF shows they are very similar. They differ only subtly as indicted by the arrows in the figure. The similarity between these plots is evidence to support the suggestion hemicellulose plays a key role in the modelling of NDF.

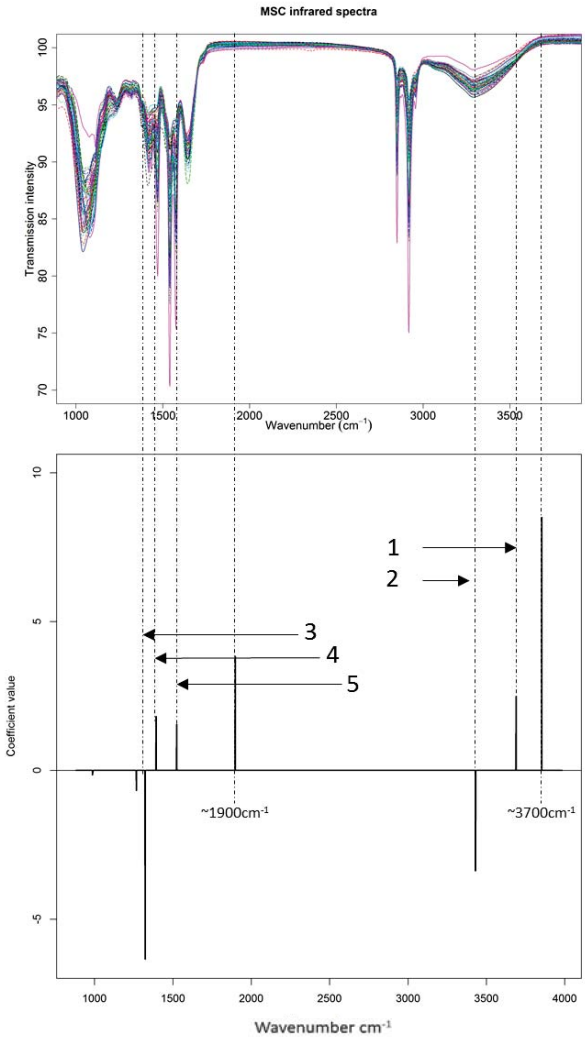

Figure 4.2.7: (a) PT1 spectra. (b) NDF coefficient plots for PT1, with significant regions number 1 – 5.

Overlaying one of these plots with the PT1 processed spectra allows for the interpretation of the chemical significance of these coefficients. The peaks at roughly 1900 cm^{-1} and 3700 cm^{-1} appear to originate from spectral regions of no chemical significance. Their inclusion is only to ensure the linear combination of reflectance values sum to the response value, however, this may impact the models sensitivity to chemical changes. Peaks 1 and 2 seem to be from the O – H stretch band. Peak 3 occupies the wavenumber range of 1350 – 1400 cm^{-1} this region contains the CH_2 bend attributed in the literature to cellulose and

hemicellulose.[162] Peak 4 lies between 1400 – 1500 cm^{-1} a region containing O – H bending (cell wall polysaccharide), CH_2 asymmetric bending and the amide III band.[162] Peak 5 lies between 1560 – 1625 cm^{-1} which may be a result of the amide II band or the aromatic C = O stretch.[160,162] The majority of these regions are therefore reasonable for inclusion in either NDF or hemicellulose models. Nitrogen was predicted accurately by all PT – models, with PT4 providing the highest quality prediction by a small margin. The coefficient plot for the PT4 model may be seen in 4.2.8. There are three significant peaks indicated at 1365 cm^{-1}, 3060 cm^{-1} and 3500 cm^{-1}, the two latter peaks are due to aliphatic C – H stretching and O – H stretching respectively.[162] The peak at 1365 cm^{-1} is suggested to be due to the C – H symmetric bending of CH_2 and CH_3 attributed in the literature to protein content.[162]

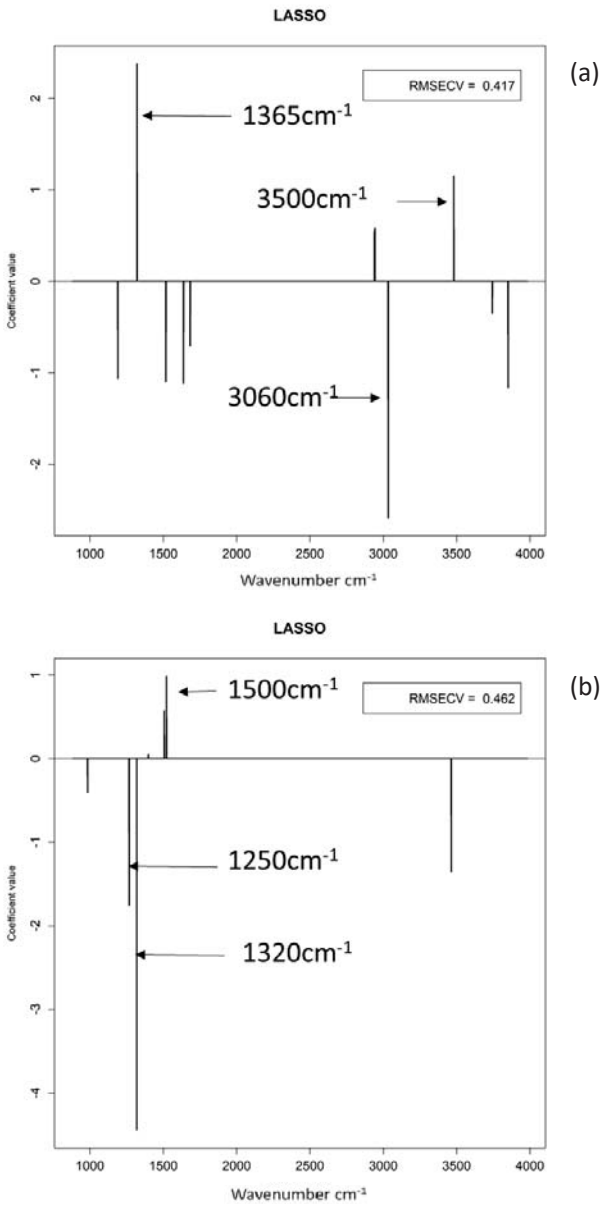

Figure 4.2.8: (a) Nitrogen coefficient plot for the PT4 model. (b) Lignin coefficient plot for the PT2 model.

The PT2 model provided modest predictions for lignin, the coefficient plot for this model may be seen in figure 4.2.8, there are three more distinct peaks

identified at 1250 cm^{-1}, 1320 cm^{-1} and 1500 cm^{-1}. All of these regions are demonstrated in the literature to contain information regarding lignin presence.[162] The peak at 1250 cm^{-1} is associated with the C = O stretch of lignin, whilst the peak at 1320 cm^{-1} is assigned to the presence of the syringyl units of lignin in the literature.[161,162] The peak at 1500 cm^{-1} corresponds to the region associated with the C = C aromatic stretching of lignin ring structures.[162]

Unfortunately, the modelling of nickel content in hyperaccumulators was no more successful using the LASSO algorithm than observed using the Ridge as seen in Table 4.2.5.

Table 4.2.5: LASSO regression results for the hyperaccumulator samples

Hyperaccumulators results			
n	R^2	RMSECV (%)	RPD
16	0.30	0.32	1.16
6	0.15	0.34	0.91
23	0.26	0.32	1.15
13	0.20	0.33	1.13
12	0.23	0.32	1.15
12	0.23	0.32	1.15

4.3 Elastic Net

The Elastic Net method is the third and final feature extraction regression method. Elastic Net models possess fewer variables than Ridge models, but typically more than the LASSO models. The Elastic Net method does not indiscriminately select variables from within highly correlated regions, because of the nature of the variable selection Elastic Net may be touted (in this thesis) as possessing the most interpretable solutions thus far. As well as possessing the already encountered tuning parameter λ the Elastic Net algorithm also utilises a parameter $\alpha \in [\,0,1]$. The α parameter tunes the Ridge ($\alpha = 0$) and LASSO ($\alpha = 1.0$) characters. The α parameter in this research was varied by increments of 0.1 over its range, selected based on the lowest RMSECV of the associated model.

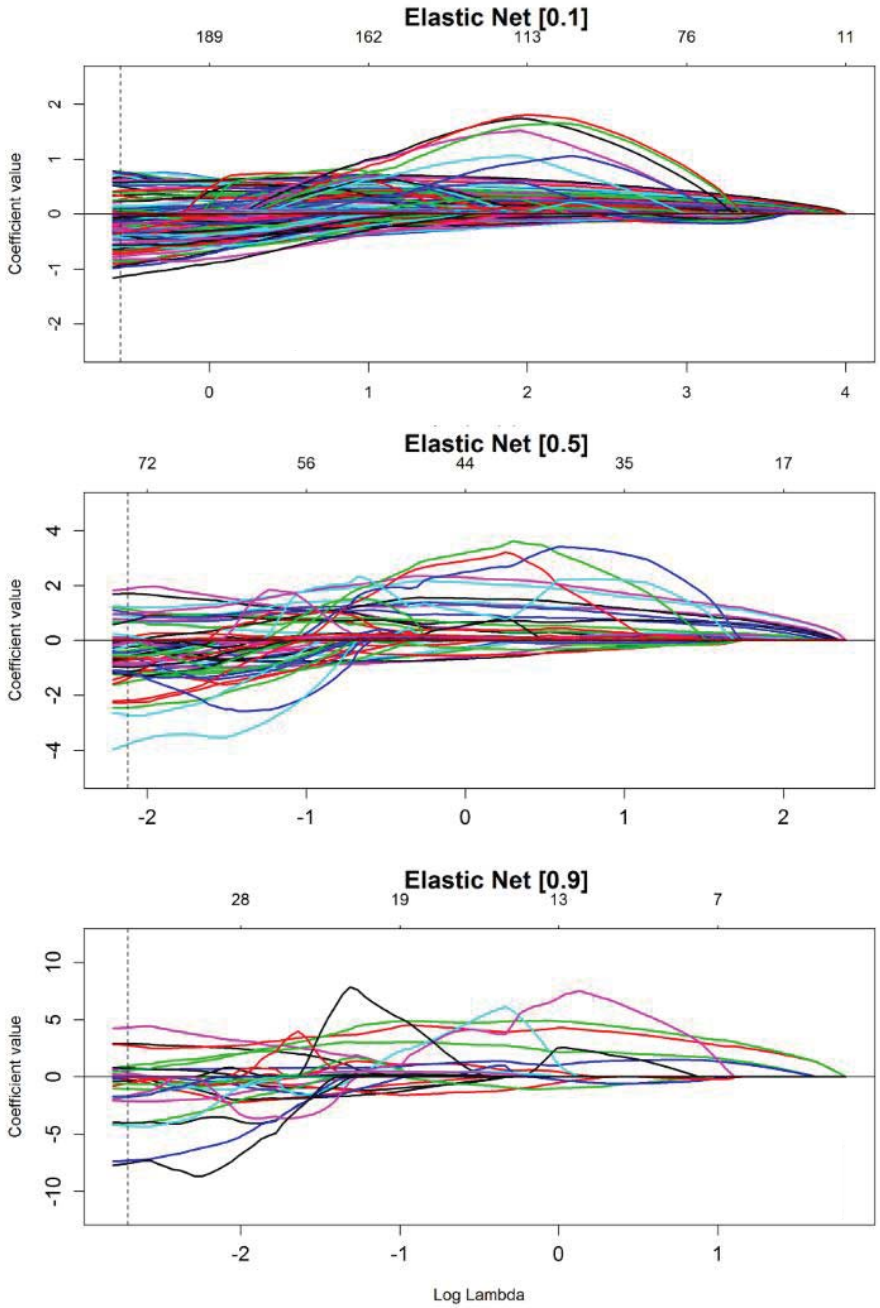

Figure 4.3.1: Three coefficient trace plots generated using the Elastic Net method. The values of α are provided in the title in square brackets.

Table 4.3.1:Elastic Net regression results for the initial forage feed samples

Initial Forage Feed Results										
	Metabolisable Energy					NDF				
PT	*n*	α	R^2	RMSEP (MJ/kg)	RPD(P)	*n*	α	R^2	RMSEP (%)	RPD(P)
1	36	0.90	0.78	0.40	1.89	172	0.10	0.90	4.49	2.41
2	151	0.20	0.76	0.41	1.76	177	0.10	0.90	4.55	2.70
3	105	0.10	0.74	0.36	1.81	115	0.10	0.88	3.29	2.67
4	218	0.10	0.74	0.41	1.74	109	0.30	0.90	4.43	2.76
5	117	0.10	0.71	0.37	1.75	112	0.10	0.89	3.28	2.74
6	164	0.10	0.72	0.39	1.76	83	0.40	0.85	4.49	2.40
	ADF					Hemicellulose				
PT	*n*	α	R^2	RMSEP (%)	RPD(P)	*n*	α	R^2	RMSEP (%)	RPD(P)
1	87	0.60	0.69	3.39	1.67	25	0.90	0.93	2.20	3.15
2	68	0.70	0.70	3.39	1.78	42	0.60	0.93	2.28	3.28
3	144	0.10	0.75	2.87	2.07	33	0.90	0.94	1.62	3.57
4	82	0.60	0.75	3.33	1.79	66	0.40	0.94	2.23	3.29
5	142	0.10	0.76	2.83	2.10	88	0.20	0.94	1.61	3.35
6	55	0.90	0.63	3.46	1.65	133	0.10	0.92	2.08	3.32

	Dry Matter					DOMD					Protein				
PT	*n*	α	R^2	RMSEP (%)	RPD(P)	*n*	α	R^2	RMSEP (%)	RPD(P)	*n*	α	R^2	RMSEP (%)	RPD(P)
1	229	0.10	0.29	1.31	1.03	35	0.90	0.78	2.51	1.89	176	0.10	0.62	2.43	1.69
2	74	0.70	0.35	1.32	1.10	151	0.20	0.76	2.53	1.76	156	0.10	0.63	2.39	1.80
3	40	0.80	0.26	1.35	1.04	105	0.10	0.74	2.27	1.81	110	0.10	0.73	1.95	2.09
4	237	0.10	0.26	1.32	1.09	217	0.10	0.74	2.53	1.74	109	0.20	0.73	2.39	1.78
5	42	0.50	0.28	1.33	1.06	117	0.10	0.71	2.29	1.75	61	0.30	0.73	1.94	2.07
6	52	0.90	0.26	1.41	1.06	164	0.10	0.72	2.42	1.76	165	0.10	0.54	2.52	1.41

Amongst the initial forage feed Elastic Net results, hemicellulose was predicted with the greatest degree of accuracy. For hemicellulose three models demonstrated very large R^2 values of 0.94, of these the PT3 model possessed the largest RPD (3.57). These three models also differ in terms of the α parameter selected. Figure 4.3.2 displays the coefficient plots of the PT3, PT4 and PT5 hemicellulose models. The coefficient plots of the PT3 and PT5 models have several values associated with wavenumbers close to the 4000 cm^{-1} region, those above 3800 cm^{-1} are very likely to be not of chemical significance. Ignoring this region, all three models are seen to possess contributions from the region around 3500 cm^{-1} which may be allocated to O – H stretching. [162] The PT3 model is the only model of the three to contain no contributions from the C – H stretching region at 2800 – 3000 cm^{-1}. [162] The fingerprint region of the PT5 model contains no relatively large

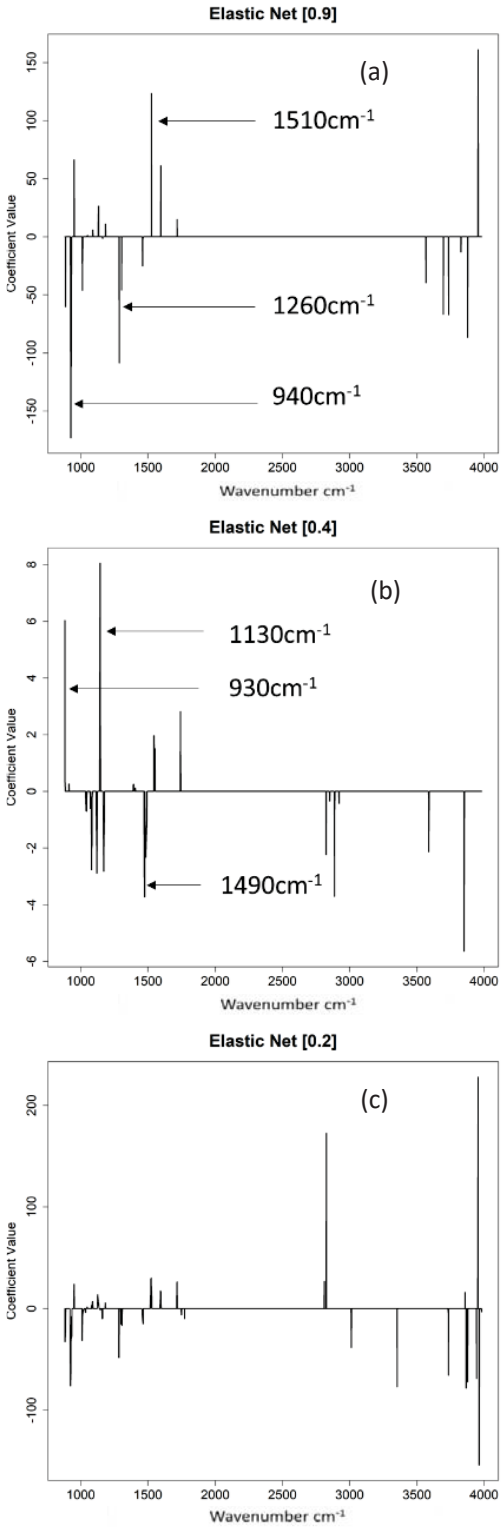

Figure 4.3.2: The coefficient plots for hemicellulose PT3 (a), PT4 (b) and PT5 (c) models.

peaks. For the PT5 model most of the significant contributions in terms of coefficient values lie above 2000 cm^{-1}. For both the PT3 and PT4 models, the fingerprint region contains some particularly distinct contributions. The peaks are indicated in Figure 4.3.2, the peaks occurring at ~935 cm^{-1} and ~1500 cm^{-1} are present in both models. The third distinct peak occurs in the PT3 and PT4 models at 1260 cm^{-1} and 1130 cm^{-1}, respectively. The peaks at ~935 cm^{-1} may be a result of C – O – H or C – O – C deformations associated with the presence of xyloglucan (hemicellulose). The peak at ~1500 cm^{-1} may be assigned to phenolic ring modes, which have been observed in literature as being associated with arabinoxylan phenolic cross – linking. [160-162] The peaks at 1260 cm^{-1} and 1130 cm^{-1} may in fact, originate from the same absorption band, the separation of them may be an unfortunate consequence of the differing degree of variable selection.

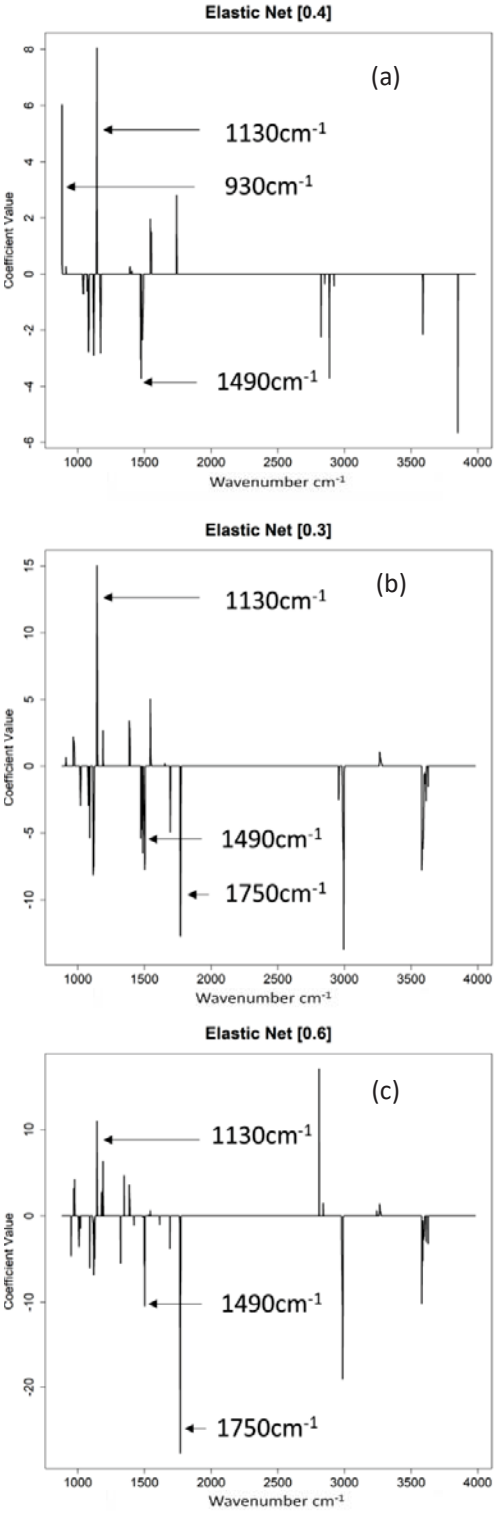

Figure 4.3.3: Displays the coefficient plots of the PT4 model for hemicellulose (a), NDF (b) and ADF (c).

The approximate region of 1130 – 1260 cm^{-1} contains contributions from several different sources, including C – O stretching of xylan and C = O stretching of hemicellulose.[162] It is clear from these assignments that the peaks identified are closely associated with hemicellulose structure. As such, their larger significance is expected. The PT4 model performs relatively well in the prediction of NDF and ADF as well. Due to the relationship between NDF, hemicellulose and ADF, and the fact that the α values are all rather similar, inspection of the coefficient plots is of particular interest. Figure 4.3.3 displays the coefficient plots of these PT4 models, indicating peaks of interest. It follows nicely that the NDF coefficient plot indicates the presence of peaks significant to either hemicellulose or ADF models. Relative to the hemicellulose model, the NDF model demonstrates the presence of a new peak at 1750 cm^{-1}, alongside the growth of the C – H stretch peak at ~3000 cm^{-1}. The peak at 1750 cm^{-1} is again present in the ADF model, where it is demonstrated as being even more distinct. The chemical origin of this peak is outlined in the literature as the C = O stretch of a saturated ester, which may be due to several plant components including lignin.[162] The other chemical components were modelled with moderate accuracy with the exception of dry matter.

The expansion of the initial forage sample set (Table 4.3.2) once again led to the general decline in model quality, the prediction of hemicellulose was least affected. As before, near identical model quality was provided by PT3, PT4 and PT5 models in terms of both R^2 (~0.77) and RPD (~1.92). Inspecting one of the sparser solutions (PT4, n = 30) as seen at the bottom of figure 4.3.4, we see the occurrence of the same peaks identified as significant to hemicellulose prediction (for the initial forage feed samples). The PT4 model, however, places more significance on the C – H stretch and O – H stretching regions than the equivalent model in the initial forage feed setting. This result may be due to the fact there are fewer coefficients included in this model.

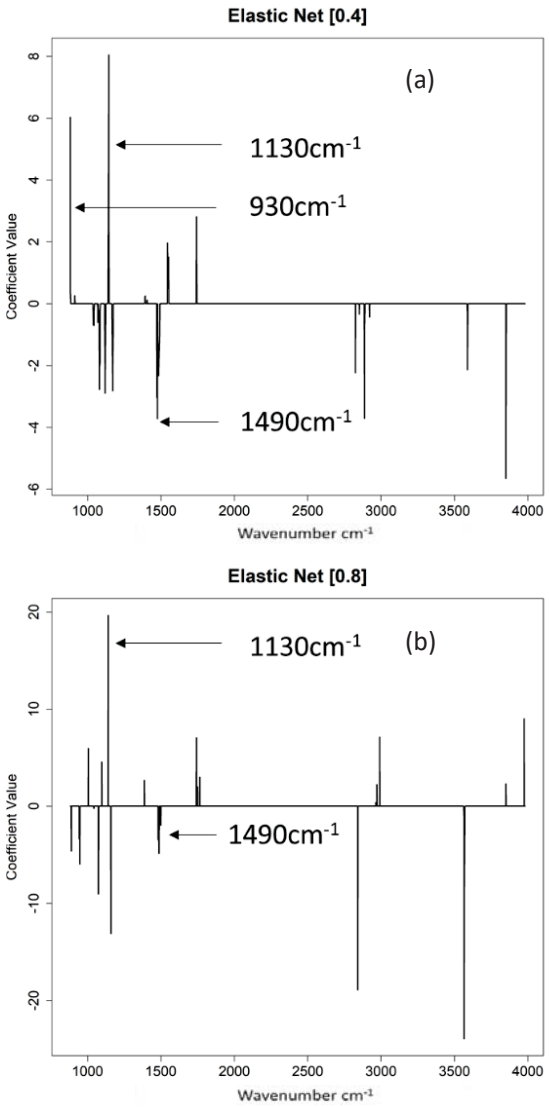

Figure 4.3.4: (a) hemicellulose PT4 model for initial forage sample set. (b) hemicellulose PT4 model for the expanded forage sample set.

Table 4.3.2: Elastic Net regression results for the expanded forage feed samples.

Expanded forage feed Results										
	NDF					Hemicellulose				
PT	n	α	R^2	RMSEP (%)	RPD(P)	n	α	R^2	RMSEP (%)	RPD(P)
1	74	0.30	0.67	6.84	1.71	55	0.50	0.71	4.00	1.92
2	65	0.40	0.67	6.96	1.60	177	0.10	0.72	3.58	1.91
3	24	0.60	0.67	6.59	1.44	148	0.10	0.77	3.20	1.92
4	83	0.30	0.67	6.82	1.58	30	0.80	0.77	3.53	1.92
5	40	0.30	0.67	6.48	1.41	23	0.90	0.76	3.15	1.93
6	103	0.20	0.53	7.22	1.34	126	0.20	0.72	3.77	1.76

	ADF					Metabolisable Energy					Dry Matter				
PT	n	α	R^2	RMSEP (%)	RPD(P)	n	α	R^2	RMSEP (MJ/kg)	RPD(P)	n	α	R^2	RMSEP (%)	RPD(P)
1	69	0.50	0.50	4.25	1.44	39	0.60	0.54	0.48	1.54	23	0.60	0.12	1.55	1.04
2	61	0.60	0.50	4.30	1.35	37	0.70	0.55	0.47	1.51	16	0.70	0.11	1.55	0.92
3	113	0.10	0.56	3.96	1.23	41	0.50	0.53	0.45	1.35	93	0.10	0.15	1.42	0.91
4	74	0.50	0.56	4.24	1.34	19	0.90	0.53	0.48	1.50	31	0.60	0.15	1.55	0.92
5	22	0.90	0.52	4.02	1.16	26	0.70	0.53	0.44	1.28	20	0.80	0.18	1.45	0.95
6	171	0.10	0.47	4.41	1.25	39	0.40	0.48	0.47	1.37	74	0.20	0.15	1.53	0.98

Table 4.2.3: Elastic Net regression results for the Chicken feed samples.

Chicken feed results															
	Crude Fibre					Starch					Fat				
PT	n	α	R^2	RMSECV (%)	RPD	n	α	R^2	RMSECV (%)	RPD	n	α	R^2	RMSECV (%)	RPD
1	22	0.70	0.84	1.28	2.67	32	0.70	0.94	4.60	4.05	85	0.20	0.68	2.34	1.76
2	25	0.70	0.82	1.28	2.50	62	0.20	0.93	4.72	3.67	93	0.20	0.67	2.36	1.72
3	27	0.60	0.90	0.96	3.35	65	0.20	0.93	4.31	3.73	103	0.10	0.63	2.50	1.62
4	28	0.70	0.82	1.28	2.49	68	0.20	0.93	4.71	3.66	111	0.20	0.67	2.37	1.73
5	20	0.80	0.90	0.97	3.31	32	0.60	0.93	4.74	3.73	93	0.10	0.57	2.63	1.52
6	54	0.40	0.85	1.27	2.71	68	0.20	0.94	4.68	3.91	17	0.90	0.57	2.91	1.52
	Nitrogen					Ash					Dry Matter				
PT	n	α	R^2	RMSECV (%)	RPD	n	α	R^2	RMSECV (%)	RPD	n	α	R^2	RMSECV (%)	RPD
1	66	0.20	0.90	0.48	3.21	50	0.80	0.97	1.23	6.30	27	0.80	0.77	0.84	2.02
2	38	0.50	0.91	0.44	3.45	55	0.80	0.97	1.27	6.35	24	0.80	0.76	0.84	2.02
3	34	0.50	0.93	0.49	3.83	98	0.10	0.98	1.08	8.46	48	0.20	0.78	0.76	2.04
4	34	0.50	0.91	0.44	3.43	42	0.80	0.97	1.27	6.27	23	0.80	0.76	0.84	2.00
5	37	0.50	0.93	0.47	3.57	18	0.90	0.98	1.05	8.70	55	0.20	0.78	0.74	2.14
6	39	0.50	0.91	0.47	3.29	50	0.80	0.97	1.18	7.02	82	0.20	0.75	0.85	2.07
	Calcium					Phosphorus					Gross Energy				
PT	n	α	R^2	RMSECV (mg/g)	RPD	n	α	R^2	RMSECV (mg/g)	RPD	n	α	R^2	RMSECV (kJ/g)	RPD
1	175	0.10	0.97	4.78	6.31	40	0.80	0.97	2.02	6.21	27	0.60	0.75	0.61	1.96
2	177	0.10	0.97	4.92	6.26	43	0.80	0.97	1.89	7.16	41	0.40	0.71	0.64	1.84
3	23	0.20	0.99	3.90	10.14	93	0.10	0.98	1.66	8.57	108	0.10	0.81	0.50	2.27
4	89	0.10	0.97	4.92	6.34	36	0.80	0.97	1.89	7.08	68	0.20	0.72	0.64	1.85
5	94	0.20	0.99	3.76	10.82	89	0.10	0.98	1.59	8.86	115	0.10	0.81	0.50	2.24
6	43	0.80	0.97	4.53	7.47	45	0.80	0.97	1.87	7.05	51	0.30	0.76	0.60	2.01

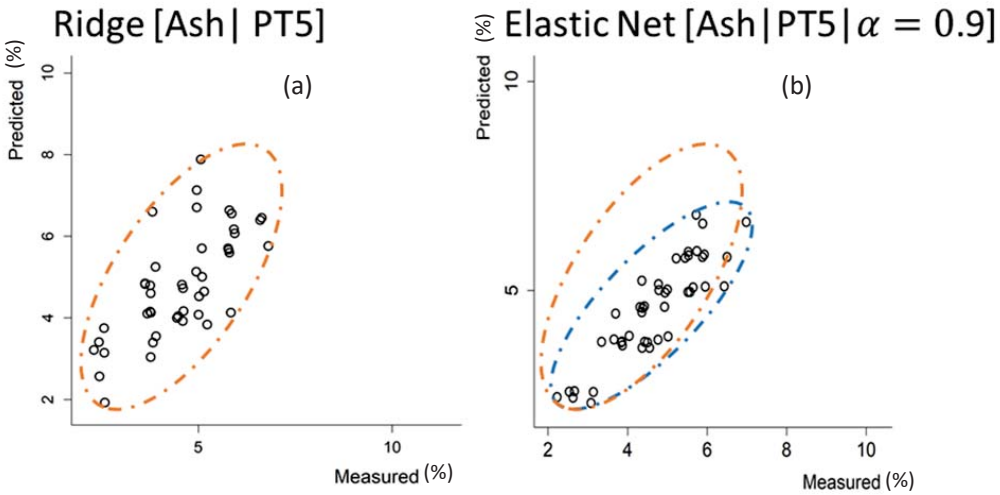

Figure 4.3.5: Comparison of Ash predicted vs measured plots for the PT5 model using Ridge (a) and Elastic Net (b).

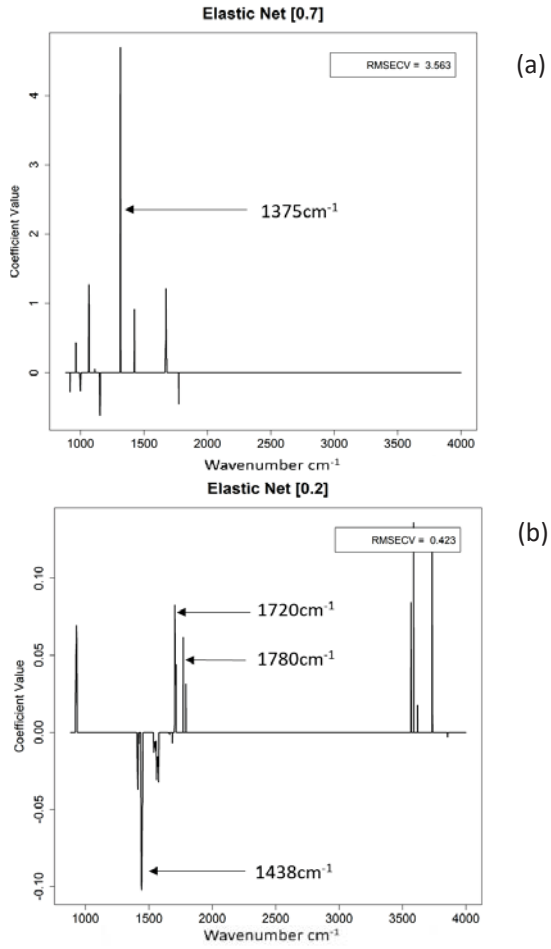

Figure 4.3.6: PT1 models of Starch (a) and Nitrogen (b).

The Elastic Net procedure reduces leveraging issues, the best of any method thus far. The phosphorus and ash predictions have improved from the Ridge and LASSO counter parts. Figure 4.3.6 displays the lower cluster of the ash prediction plots, for both Ridge and Elastic Net. The illustrated orange and blue oval represent the variance ellipses, the reduction in size is demonstrated by overlaying the Ridge ellipse on the Elastic Net plot. What is of interest is that Elastic Net has superior performance to LASSO in this sense, and it has done so by incorporating Ridge penalty character into the Elastic Net penalty. Starch and nitrogen were both modelled accurately using Elastic Net based models, both also demonstrated a tendency to have the majority of PT models selecting the same value of α (0.2 and 0.5 respectively). Figure 4.3.6 contains the coefficients of both the starch and nitrogen PT1 models. The starch model contains contributions only within the fingerprint region. Within this region there is one distinctive peak, at 1375 cm^{-1} it is likely this is due to the presence of CH_2 bending, of cellulose or hemicellulose.[162] Given the fact cellulose and starch are so similar in structure, the importance of this region is clear. The nitrogen model has three more prominent peaks between 1250 – 2000 cm^{-1}. The peak at 1438 cm^{-1} likely originates from the absorption due to C – H symmetric bending of CH_2 and CH_3 from cell wall polysaccharides, lipids and most relevantly, proteins.[162] The two prominent peaks at 1720 cm^{-1} and 1780 cm^{-1} are suggested to be portions of the saturated ester C = O stretch band, present in a variety of plant components.[162] There is a third cluster of peaks at ~1560 – 1590 cm^{-1} this region represents the incorporation of the amide II band (C = N and N – H) into the model.[160] Aside from the three peaks indicated in the figure and those at ~1560 – 1590 cm^{-1} there are significant contributions around 3500 – 3600 cm^{-1} which is the tail end of the O – H and N – H stretching

region.[162] The coefficient plots for the nitrogen and starch PT1 models, indicate the allocation of importance to appropriate wavenumber regions given the nature of the chemical composition being modelled. Crude fibre and gross energy were both predicted with the greatest accuracy in the derivative based pre – treatment models, the accuracy of prediction is demonstrated by the larger R^2 and RPD values observed. The accurate modelling of crude fibre and gross energy with PT3 and PT5 has been observed in both Ridge and LASSO results.

Table 4.3.4: Elastic Net regression results for the faecal samples

Faecal Results															
	Hemicellulose					Dry Matter					Ash				
PT	*n*	*α*	R^2	RMSECV (%)	RPD	*n*	*α*	R^2	RMSECV (%)	RPD	*n*	*α*	R^2	RMSECV (%)	RPD
1	25	0.90	0.88	2.88	2.90	141	0.20	0.65	0.44	1.70	37	0.70	0.95	2.33	4.34
2	24	0.90	0.87	3.14	2.76	146	0.20	0.65	0.45	1.71	59	0.40	0.93	2.74	3.87
3	68	0.20	0.88	2.56	2.84	30	0.50	0.65	0.46	1.66	103	0.10	0.93	2.15	3.84
4	18	0.90	0.87	3.15	2.75	157	0.20	0.66	0.45	1.72	61	0.40	0.93	2.75	3.87
5	73	0.20	0.89	2.54	2.90	56	0.20	0.52	0.45	1.43	33	0.70	0.93	2.13	3.87
6	103	0.10	0.83	3.08	2.38	48	0.40	0.65	0.41	1.68	23	0.80	0.91	2.09	3.37
	Nitrogen					Gross Energy					Organic Matter				
PT	*n*	*α*	R^2	RMSECV (%)	RPD	*n*	*α*	R^2	RMSECV (%)	RPD	*n*	*α*	R^2	RMSECV (%)	RPD
1	48	0.60	0.88	0.40	2.81	27	0.90	0.97	0.51	5.55	40	0.70	0.94	2.21	3.98
2	31	0.80	0.88	0.41	2.78	116	0.20	0.96	0.51	5.31	37	0.70	0.93	2.57	3.69
3	25	0.90	0.89	0.35	2.91	39	0.60	0.94	0.48	4.82	103	0.10	0.90	2.20	3.23
4	31	0.80	0.87	0.41	2.70	119	0.20	0.96	0.51	5.29	42	0.70	0.93	2.59	3.68
5	80	0.20	0.88	0.35	2.75	37	0.80	0.96	0.48	5.79	86	0.20	0.91	2.12	3.50
6	124	0.20	0.85	0.40	2.51	85	0.40	0.96	0.49	5.42	23	0.80	0.91	2.12	3.23
	NDF					ADF					Lignin				
PT	*n*	*α*	R^2	RMSECV (%)	RPD	*n*	*α*	R^2	RMSECV (%)	RPD	*n*	*α*	R^2	RMSECV (%)	RPD
1	24	0.80	0.89	4.73	2.99	30	0.80	0.87	2.35	2.75	26	0.80	0.84	0.48	2.60
2	32	0.70	0.88	5.03	2.89	31	0.80	0.86	2.47	2.67	50	0.40	0.83	0.50	2.49
3	82	0.20	0.88	4.41	2.88	72	0.20	0.84	2.27	2.45	19	0.90	0.80	0.46	2.22
4	18	0.90	0.88	5.04	2.83	27	0.80	0.84	2.48	2.55	51	0.40	0.82	0.50	2.43
5	81	0.20	0.89	4.41	3.00	77	0.20	0.86	2.27	2.59	51	0.30	0.82	0.45	2.20
6	102	0.10	0.83	5.24	2.41	103	0.10	0.75	2.51	1.99	48	0.50	0.79	0.48	2.29

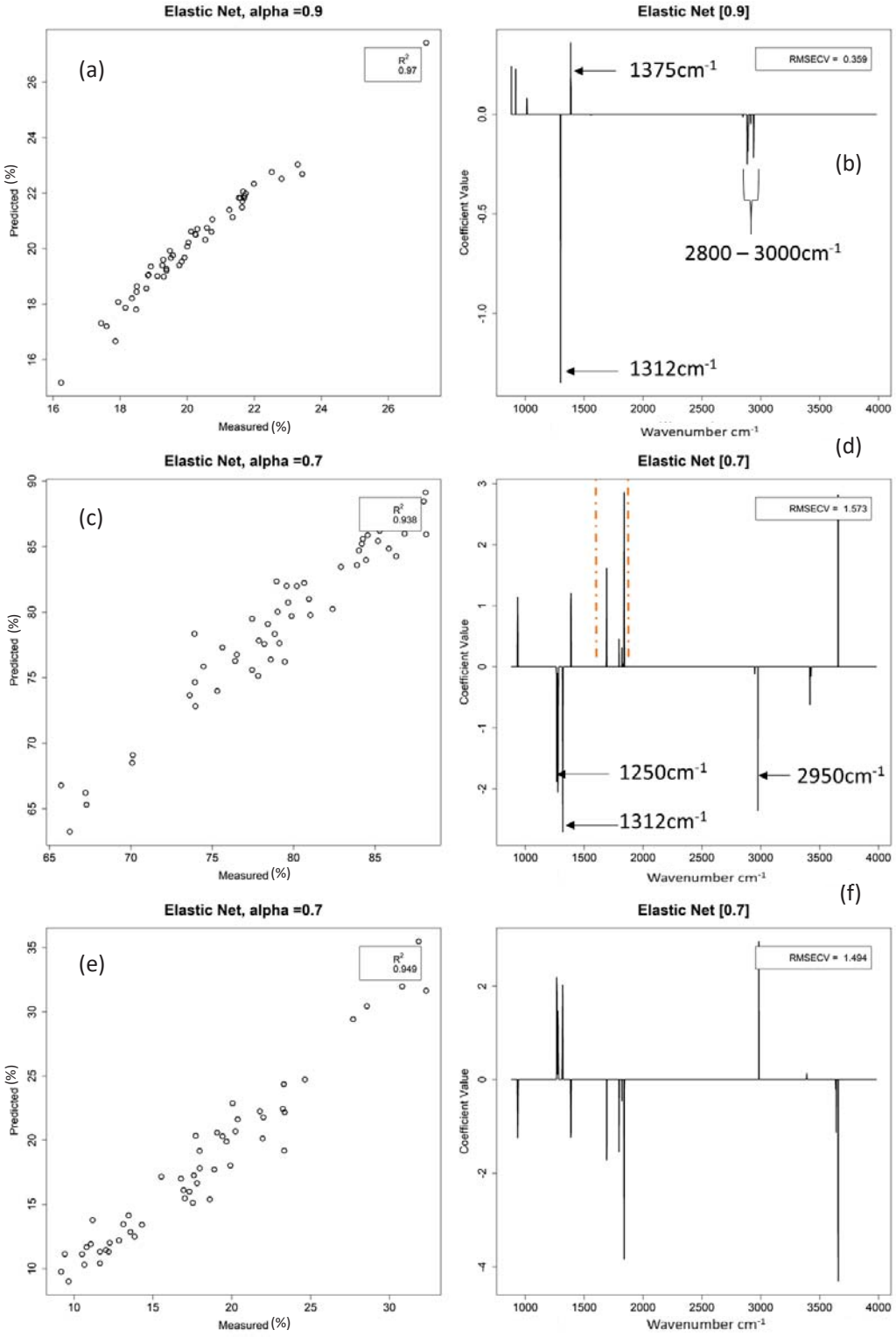

Figure 4.3.7: Left column contains predicted vs measured value plots for gross energy (a), organic matter (c) and ash (e). Right column contains coefficient plots of gross energy (b), organic matter (d) and ash (f).

Table 4.3.4 contains the Elastic Net results for the prediction of chemical compositions for the faecal samples. Figure 4.3.7 contains the prediction plots and coefficient plots for the PT1 models of gross energy, organic matter and ash. These chemical compositions are predicted with the highest accuracy among all chemical compositions in the faecal sample set. Inspection of the coefficient plots reveals that the organic matter and ash models identify effectively the same spectral regions of significance. The models also select the same values of α. The relationship between these models is appropriate given the mutually exclusive relationship between ash (inorganic content) and organic matter.

The coefficient plots for gross energy and organic matter both are labelled with peaks of interest. Assignment will be carried out in reference to gross energy and organic matter, however, clearly assignments to the latter are also present in the ash plot. The organic matter plot contains a significant, but unlabelled peak at ~3600 cm^{-1} which may be from the tail end of the O – H stretch band, and a cluster of peaks between the vertical dashed lines, 1625 – 1875 cm^{-1}.[162] Both the gross energy and organic matter models, have selected variables within 2800 – 3000 cm^{-1} containing wavenumbers associated with C – H stretch information. The peak at 1312 cm^{-1} is also present in both models, it is suggested to be a portion of the amide III band or also assigned in the literature to absorption due to the syringyl units of lignin.[161,162] The 1250 cm^{-1} peak of the organic matter plot is suggested to originate from the C = O stretch present in a variety of plant matter components.[162] The final peak of interest for the gross energy coefficient plot, lies at 1375 cm^{-1} a region observed in the literature to contain CH_2 bend information, originating from cellulose or hemicellulose.[162] The region of 1625 – 1875 cm^{-1} was mentioned previously, as it contains a cluster

of peaks in the coefficient plot of organic matter, this region captures information on several bond vibrations including C = O aromatic stretch, C = C stretch, C = O stretch (amide I) and the C = O stretch of saturated esters.[162] All of these bonds are found within the molecular structures of several plant components such as lignin, phospholipid, pectin and cutin esters, and phenolic compounds.[162] The bonds within this region (1625 – 1875 cm^{-1}) are clearly relevant to organic matter content, so the presence of significant peaks here is appropriate. Lignin was predicted with quite reasonable accuracy by the PT1 model (R^2: 0.84, RPD: 2.60). Amongst the literature lignin is noted as possessing

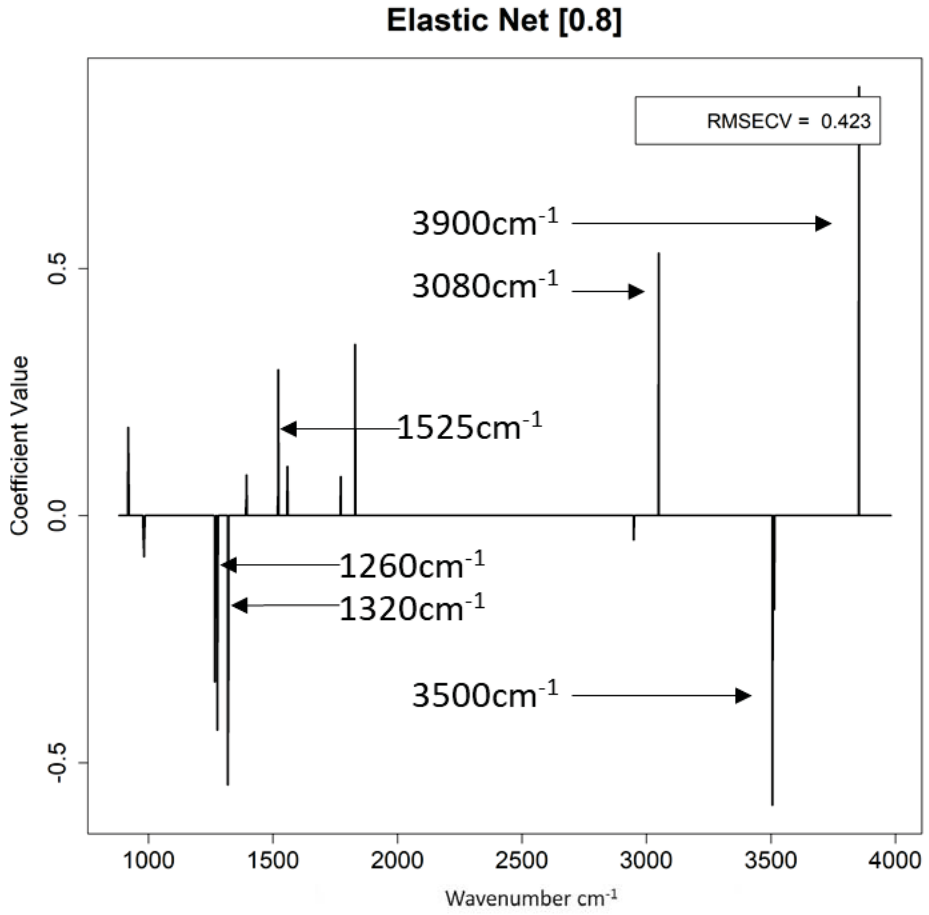

Figure 4.3.8: Coefficient plot for PT1 lignin model.

several mid – infrared absorption regions, at 1270 cm^{-1}, 1330 cm^{-1} and 1500 – 1600 cm^{-1}.[162] In figure 4.3.8 the coefficients utilised by the PT1 model of lignin content may be seen. The figure shows other significant variables. Those above 3000 cm^{-1} are attributed to the C – H stretch (3080 cm^{-1}), and the O – H stretch (3500 cm^{-1}, 3900 cm^{-1}), although the peak at 3900 cm^{-1} is perhaps unlikely to be a variable selection from within the O – H stretch band.[160,162] The three other peaks indicated align well to those identified in the literature as due to lignin presence. The peaks at ~1270 cm^{-1} and ~1330 cm^{-1} are described in the literature to originate from the syringyl and guaiacyl units of lignin, whilst the peak at 1525 cm^{-1} is likely due to the aromatic backbone of lignin.[161,162] The hemicellulose, NDF and ADF predictions are all fairly equivalent in the Elastic Net results. The PT1 model coefficients are displayed in figure 4.3.9 with peaks of interest numbered. The numbered regions outlined are of interest because of their relationship to the ADF and hemicellulose models. Peaks at region 4 are present amongst all

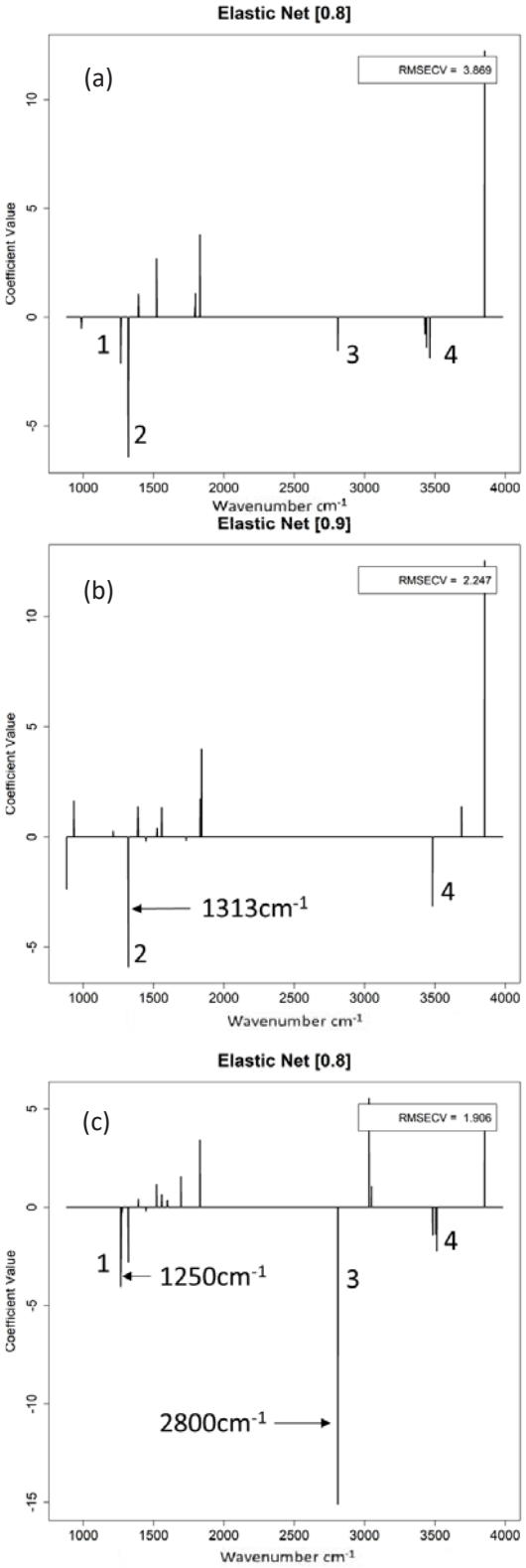

Figure 4.3.9: The PT1 models of NDF (a), hemicellulose (b) and ADF (c).

models. This result is expected as O – H stretching absorptions would be present for each composition. Peaks 1, 2 and 3 are either completely absent or significantly reduced in either ADF or hemicellulose, of these peaks, peak 2 is the only one present in the hemicellulose model. Peak 2 occurs at 1313 cm^{-1} very close to the literature values of 1316 – 1317 cm^{-1} associated with the presence of xyloglucan, a hemicellulose. Peaks 1 and 3 of the ADF model are larger than in the NDF model, these peaks occur at 1250 cm^{-1} and 2800 cm^{-1}, respectively. The peak at 1250 cm^{-1} may correspond to absorptions of the guaiacyl units of lignin, listed in the literature as occurring at 1270 cm^{-1}.[161] 2800 cm^{-1} is the familiar C-H stretching region, evidently this region offers more insight into ADF content than hemicellulose content, as suggested by the complete lack of contribution from this region to the hemicellulose model.

Unfortunately, nickel content in the hyperaccumulators was poorly modelled. The PT1 model produced the best quality predictions, however, they are still quite poor.

Table 4.3.5: Elastic Net regression results for the hyperaccumulator samples

Hyperaccumulators				
n	α	R^2	RMSECV (%)	RPD
100	0.60	0.50	0.31	1.39
46	0.10	0.15	0.36	0.91
31	0.70	0.27	0.30	1.15
71	0.20	0.27	0.30	1.18
18	0.70	0.27	0.29	1.18
18	0.70	0.27	0.29	1.18

Chapter 5.0: Results and Discussion II
Feature Construction Regression Methods

This section will present the results and discussion for PCR and PLS, grouped in this chapter as they are both feature construction methods. The same logic is applied regarding the exclusion of calibration results as described in section 4.0. Likewise, this chapter contains two subsections based on whether the results relate to PCR or PLS, these subsections once again are organised chronologically based on the development of the method.

5.1 Principal Component Regression

This technique generates principal components as linear combinations of the original variables. The components are extracted using the eigenvectors of the covariance – variance matrix of the initial predictor variable space, the same components used in PCA. The number of components is selected based on how the inclusion of the component affects the RMSECV of the resulting model. Plotting RMSECV as a function of the number of components in the model, the number chosen is the first to produce an RMSECV within one standard error of the minimum value generated. This selection method is suggested to provide a model which has the lowest number of components and still produce relatively low error, hence is less at risk of overfitting. Figure 5.1.1 displays the RMSECV plot on the left and the RMSEP on the right, RMSECV is most widely used in this project to select for components as RMSEP is not always available. The use of RMSECV should, however, provide insight into the number of components that would also perform well in the prediction of *truly* new observations (RMSEP). Figure 5.1.1 illustrates that 3 components is the optimal choice based on RMSECV. Figure 5.1.1 illustrates the RMSEP supports the use of three components, as the RMSEP for the three component model is very close to the minimum RMSEP. Figure 5.1.2 displays the loading plots of the first two components of some PCR model. Loading plots describe the importance of each variable to the relevant component.

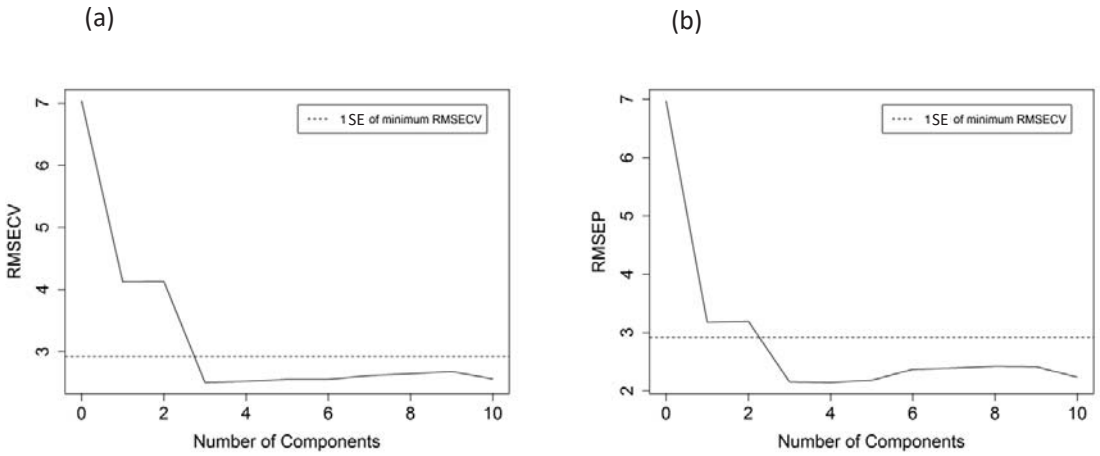

Figure 5.1.1: (a) Plot of RMSECV vs number of components. (b) Plot of RMSEP vs number of components for the same model as (a).

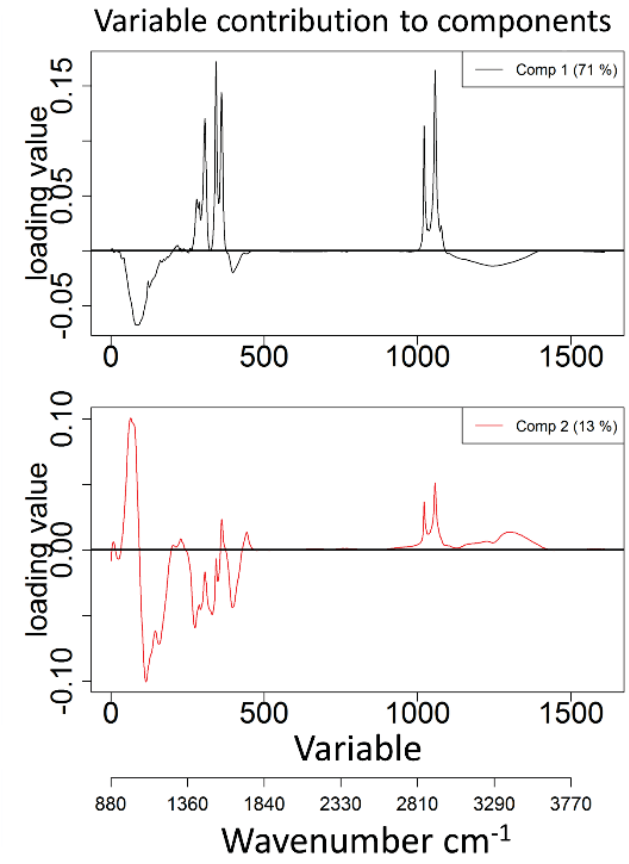

Figure 5.1.2: Example loading value plots for the first two components of a model. The legend contains the component number and explained variance, the latter is in parentheses.

Table 5.1.1: Principal components regression results for the initial forage feed samples

Initial forage feed results												
	Metabolisable Energy				NDF				Dry Matter			
PT	*n*	R^2	RMSEP (MJ/kg)	RPD(P)	*n*	R^2	RMSEP (%)	RPD(P)	*n*	R^2	RMSEP (%)	RPD(P)
1	3	0.56	0.47	1.50	3	0.67	6.30	1.67	8	0.41	1.27	1.25
2	4	0.62	0.44	1.60	3	0.67	6.40	1.64	8	0.40	1.28	1.24
3	8	0.70	0.40	1.70	10	0.92	3.25	2.65	10	0.37	1.29	1.21
4	3	0.56	0.47	1.50	3	0.67	6.41	1.64	8	0.40	1.28	1.24
5	7	0.71	0.40	1.61	5	0.78	5.02	2.11	2	0.21	1.43	1.12
6	4	0.55	1.29	0.29	5	0.69	25.25	0.21	8	0.41	1.48	0.87

	ADF				Hemi				DOMD				Protein			
PT	*n*	R^2	RMSEP (%)	RPD(P)	*n*	R^2	RMSEP (%)	RPD(P)	*n*	R^2	RMSEP (%)	RPD(P)	*n*	R^2	RMSEP (%)	RPD(P)
1	5	0.43	4.20	1.29	3	0.91	2.21	3.15	3	0.56	2.96	1.50	5	0.46	3.03	1.03
2	7	0.40	4.28	1.29	3	0.91	2.15	3.23	4	0.62	2.77	1.60	5	0.48	3.10	0.98
3	10	0.77	2.75	1.79	10	0.95	1.75	3.34	8	0.70	2.47	1.70	7	0.74	1.79	1.91
4	7	0.40	4.28	1.29	3	0.91	2.15	3.22	3	0.56	2.96	1.50	5	0.47	3.05	1
5	7	0.70	3.05	1.79	1	0.92	2.07	3.37	7	0.71	2.49	1.61	4	0.59	2.26	1.42
6	9	0.30	11.63	0.26	3	0.89	9.52	0.37	4	0.55	8.09	0.29	5	0.45	8.78	0.20

In the legend the brackets contain the variance accounted for by the component (predictor variance). The loadings are a representation of how important each variable of the original data set is to a particular component. Specifically, the value is the inner product of each variable vector with the corresponding left singular vector of the $\boldsymbol{X}$ matrix. Table 5.1.1 displays the PCR results for the initial forage feed samples, hemicellulose was again modelled very accurately. There is one distinct outlier amongst the hemicellulose results which is the PT6 model, where the RPD is far smaller than any other hemicellulose model. The PT3 model performed best within hemicellulose prediction, and likewise in the prediction of both NDF and ADF. The chemical compositions of NDF and ADF were predicted with large differences in model quality. The PT3 model selected a large number of components, ($n = 10$) in both NDF and ADF prediction, in stark contrast to the selections by other models especially in NDF prediction where no other model used greater than 5 components.

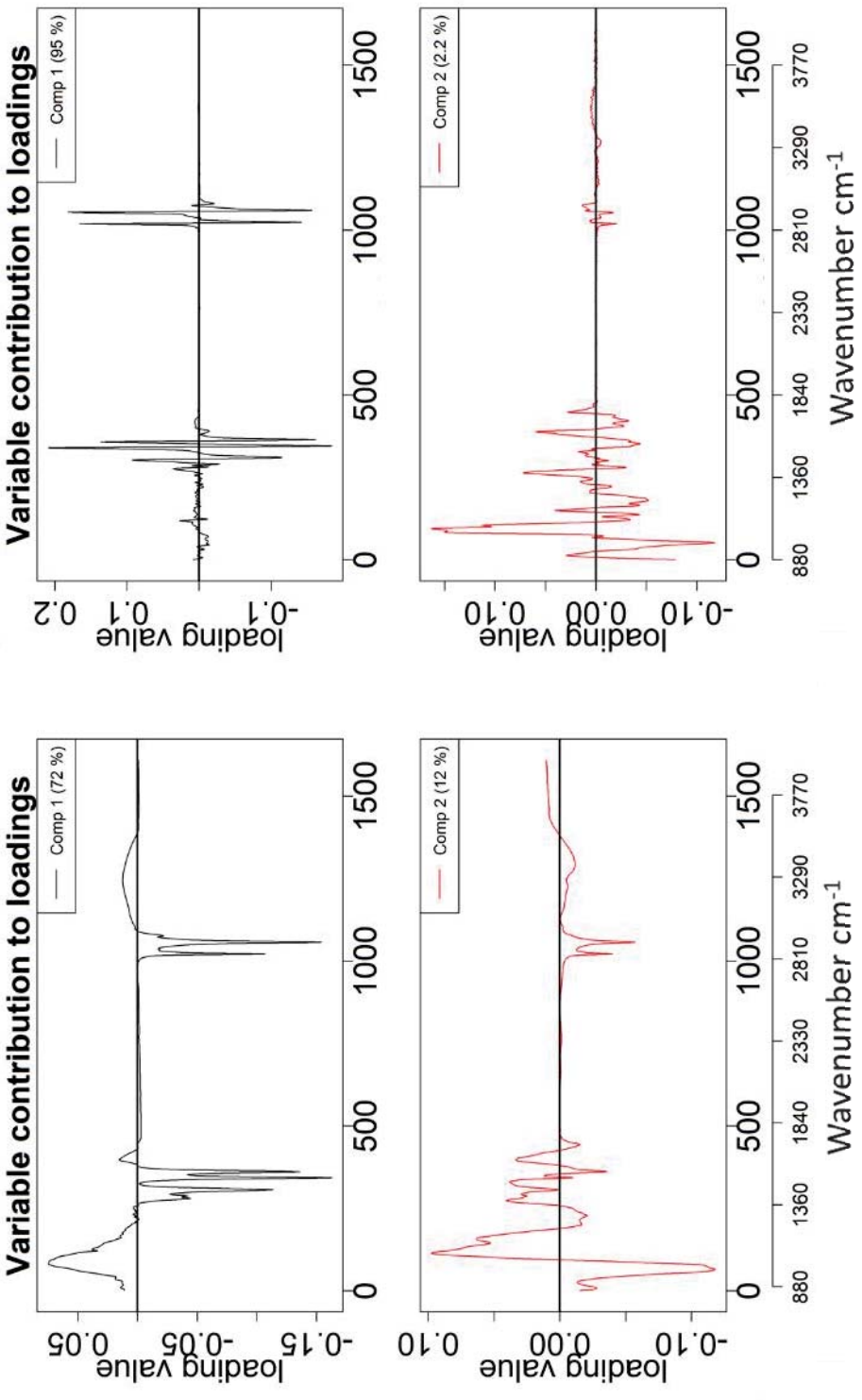

Figure 5.1.3: Loadings of the first two components of the PT2 (a) and PT3 (b) models for NDF.

Both PT3 and PT5 incorporate a Savitsky–Golay filter in their pre–processing steps. The present results suggest PCR gives superior performance in the analysis of 1st order derivative spectra relative to the other pre–processing utilised in this project. In fact, this trend extends to the other chemical compositions within the forage feed sample set, PT2 and PT3 differ only in this SG – filter step, yet differ quite substantially in terms of prediction quality. Figure 5.1.3 displays the first two components for each model in the prediction of NDF. The first component of the PT2 model possesses two contributions which are absent in the PT3 model, these are located at 880 – 1360 cm^{-1} and 3000 – 3500 cm^{-1}. The fingerprint regions, aside from geometry differences as a consequence of the derivative process, appear somewhat similar in the second components. The most obvious differences of the PT2 and PT3 loadings (figure 5.1.3) occur in the reduction of aliphatic C – H stretching in the PT3 model, as indicated in the reduced magnitude of the loadings of the two bands at ~3000 cm^{-1}. The broad O – H stretch makes an appearance again in the second component of the PT2 model, but appears absent in the PT3 model.[162] The accounted variances in the first two components are quite different, with the PT2 model possessing 72% and the PT3 95%. This results suggests that very little variance is associated with the remaining components of the PT3 model. Because we have access in this instance to *true* prediction performance metrics, the relatively large number of components with small variance is not tremendously concerning. The metabolisable energy, DOMD and protein content were all predicted modestly by PT3 with R^2 ~0.7 and RPD 1.70, 170 and 1.97, respectively.

Table 5.1.2: Principal components regression results for the expanded forage feed samples

Expanded forage feed results								
			NDF				Hemicellulose	
PT	n	R^2	RMSEP (%)	RPD(P)	n	R^2	RMSEP (%)	RPD(P)
1	3	0.51	7.84	1.27	3	0.63	4.25	1.59
2	3	0.52	7.81	1.27	3	0.63	4.22	1.60
3	10	0.71	6.09	1.58	5	0.77	3.44	1.84
4	3	0.52	7.81	1.27	3	0.63	4.21	1.60
5	1	0.51	7.90	1.24	1	0.70	3.86	1.70
6	1	0.49	19.30	0.30	3	0.70	12.96	0.27

			ADF				Metabolisable Energy				Dry Matter	
PT	n	R^2	RMSEP (%)	RPD(P)	n	R^2	RMSEP (MJ/kg)	RPD(P)	n	R^2	RMSEP (%)	RPD(P)
1	5	0.29	4.78	0.96	1	0.41	0.50	1.14	2	0.09	1.61	0.82
2	5	0.29	4.79	0.96	1	0.42	0.50	1.14	2	0.09	1.61	0.82
3	10	0.55	3.78	1.24	9	0.62	0.41	1.37	5	0.11	1.52	0.91
4	5	0.29	4.79	0.96	1	0.42	0.50	1.12	2	0.09	1.60	0.82
5	6	0.40	4.38	1.08	6	0.52	0.46	1.22	6	0.06	1.64	0.83
6	5	0.24	20.63	0.13	6	0.32	1.51	0.22	8	0.24	7.18	0.11

Table 5.1.2 contains the results of the PCR analysis on the expanded forage feed sample set. The PT3 models, and to a lesser extend the PT5 models continue the trend of superior performance relative to non – derivative containing PT's. Again PCR failed to generate any high quality predictions in the expanded sample set, which has been the case for previously discussed methods as well.

Table 5.1.3: Principal components regression results for the chicken feed samples

Chicken feed results												
PT	*n*	R^2	Crude Fibre RMSECV (%)	RPD(CV)	*n*	R^2	Starch RMSECV (%)	RPD(CV)	*n*	R^2	Fat RMSECV (%)	RPD(CV)
1	3	0.66	1.52	1.71	2	0.83	5.99	2.40	3	0.43	2.66	1.31
2	3	0.69	1.44	1.78	2	0.83	5.93	2.43	1	0.36	2.51	1.25
3	5	0.67	1.50	1.74	1	0.88	5.04	2.88	6	0.52	2.50	1.43
4	6	0.84	1.04	2.46	1	0.89	4.78	3.02	6	0.56	2.36	1.48
5	4	0.64	1.58	1.66	1	0.88	5.03	2.88	6	0.55	2.37	1.46
6	4	0.19	1.58	0.97	1	0.91	5.03	1.18	6	0.51	2.37	0.85
PT	*n*	R^2	Nitrogen RMSECV (%)	RPD(CV)	*n*	R^2	Ash RMSECV (%)	RPD(CV)	*n*	R^2	Dry Matter RMSECV (%)	RPD(CV)
1	3	0.88	0.46	2.90	3	0.94	1.31	3.84	2	0.65	0.88	1.69
2	1	0.71	0.69	1.86	3	0.93	1.39	3.64	2	0.65	0.88	1.69
3	4	0.89	0.46	2.97	4	0.91	1.28	3.32	2	0.57	0.93	1.52
4	3	0.75	0.67	2.01	4	0.91	1.28	3.32	2	0.57	0.93	1.51
5	4	0.89	0.46	2.97	4	0.91	1.28	3.31	2	0.57	0.93	1.52
6	4	0.90	0.46	1.00	4	0.91	1.28	3.31	2	0.57	0.93	1.52
PT	*n*	R^2	Calcium RMSECV (mg/g)	RPD(CV)	*n*	R^2	Phosphorus RMSECV (mg/g)	RPD(CV)	*n*	R^2	Gross Energy RMSECV (kJ/g)	RPD(CV)
1	3	0.94	4.81	4.09	3	0.92	2.27	3.53	2	0.57	0.63	1.53
2	3	0.94	4.76	4.13	3	0.92	2.23	3.59	3	0.69	0.47	1.78
3	4	0.91	4.82	3.38	4	0.91	1.97	3.39	6	0.74	0.50	1.89
4	4	0.91	4.83	3.37	4	0.91	1.98	3.38	6	0.75	0.49	1.94
5	4	0.91	4.83	3.37	4	0.91	1.98	3.38	6	0.74	0.49	1.93
6	4	0.91	4.83	3.37	4	0.91	1.98	3.38	6	0.31	0.49	1.08

Table 5.1.3 contains the chicken feed results, the large R^2 and RPD values of PT1 in the predictions of ash, calcium and phosphorus are unfortunately once again misleading. Visual inspection of figure 5.1.4 reveals the clustering and presence of leveraging that has caused issues previously. Unlike in some prior methods, PCR has been unable to successfully predict values within each cluster, as demonstrated by the spherical or column geometry of the clusters. The calcium content is the clearest example of this lack of prediction ability. Again the variance of prediction values and measured values within the clusters is very different. The PT3 and PT5 models predicted nitrogen content equally accurately, and more accurately than any other PT – model in terms of R^2 and RPD. In figure 5.1.5 the first two component loadings are displayed for both the PT2 model and the PT3, the vertical dashed green lines of represent the more significant contributions to both

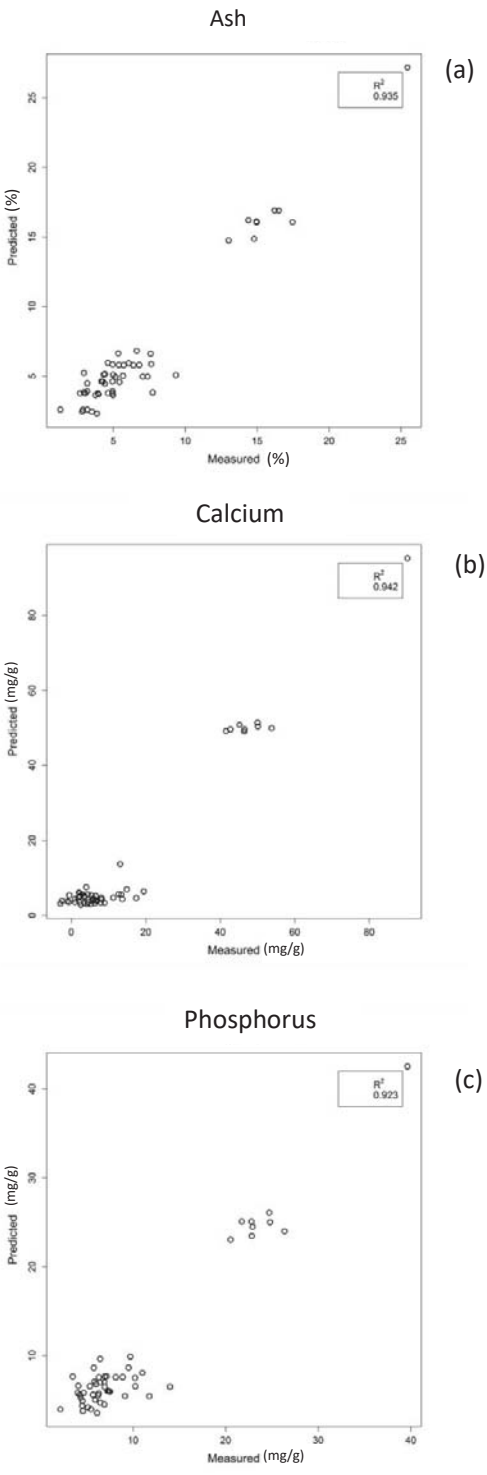

Figure 5.1.4: Predicted vs measured plots of ash (a) , calcium (b) and phosphorus (c).

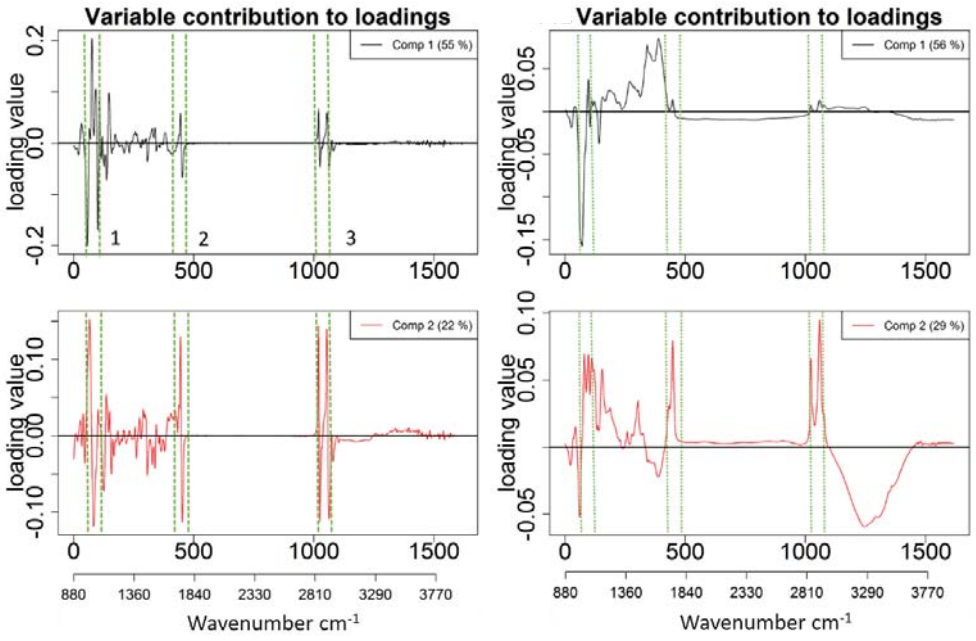

Figure 5.1.5: Left column contains loadings of the nitrogen PT3 models. Right column contains loadings of the nitrogen PT2 model.

components one and two of the PT3 model. These regions are numbered in the top left graph as 1, 2 and 3 they cover the wavenumbers ranges of 1000 – 1120 cm^{-1}, 1700 – 1820 cm^{-1} and 2830 – 2950 cm^{-1}, respectively. Region 1 is likely originating from cutin C – O – C symmetric stretching.[162] Region 2 is likely the C = O stretch of a saturated ester, which may originate from a variety of plant components.[162] Region 3 is considered to contain CH_3 symmetric stretching and CH_2 symmetric and asymmetric stretching,[162] originating from protein and lipid content predominantly. PT2 demonstrates a large reduction in regions 2 and 3 for the first component, with perhaps an increase at the amide I band slightly below region 2. The second component demonstrates more parity to the PT3 model within regions 1, 2 and 3, however, the N-H stretching region at 3200 cm^{-1} possess significant contribution in this model. It may be interpreted that the increase of the presence of the amide I band in the first component of the PT4 model, accompanied by the reduction of regions 2 and 3 may account for some of the observed prediction quality reduction. Starch and crude fibre were predicted most accurately by the PT4

model, interestingly the starch PT3 model utilized only one component. The use of only one component in the starch PT3 model allows for great insight into which spectral regions are important to the model. Figure 5.1.6 displays the loadings of the component in the PT4 starch model, once again the dashed green lines indicate prominent contributions to the model. Region 1 spans 1000 – 1060 cm^{-1} and likely absorptions in this region originate from the stretching of O – H and C – OH in cell wall polysaccharides.[162] Region 2 lies between 1420 – 1540 cm^{-1} this region contains a range of absorptions including C = C aromatic stretching, amide III, C – H asymmetric bending of CH_2 and CH_3 as well as O – H bending.[162] The plant components containing these groups are lignin, protein, lipid and cell wall polysaccharides, [162] aside from the amide III band, it is logical to see these contributions as important in a model predicting starch content.

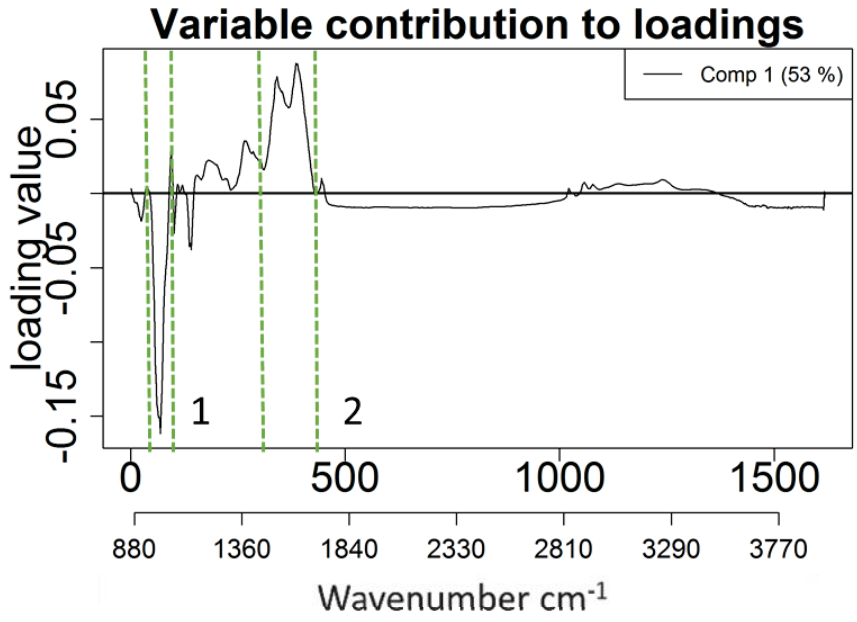

Figure 5.1.6: Loadings of the first component for the starch PT4 model.

Table 5.1.4: Principal components regression results for the faecal samples

Faecal results												
			Hemicellulose				Dry Matter				Ash	
PT	*n*	R^2	RMSECV (%)	RPD(CV)	*n*	R^2	RMSECV (%)	RPD(CV)	*n*	R^2	RMSECV (%)	RPD(CV)
1	5	0.73	3.41	1.91	8	0.48	0.44	1.35	5	0.66	3.73	1.70
2	5	0.72	3.43	1.90	10	0.49	0.45	1.34	5	0.66	3.68	1.70
3	3	0.72	3.38	1.87	10	0.35	0.49	1.21	6	0.77	2.88	2.01
4	3	0.72	3.37	1.87	9	0.23	0.54	1.11	4	0.67	3.44	1.72
5	3	0.72	3.37	1.88	9	0.22	0.54	1.10	4	0.67	3.46	1.72
6	4	0.69	3.61	1.79	9	0.61	0.37	1.57	6	0.82	2.71	2.33
			Nitrogen				Gross Energy				Organic Matter	
PT	*n*	R^2	RMSECV (%)	RPD(CV)	*n*	R^2	RMSECV (%)	RPD(CV)	*n*	R^2	RMSECV (%)	RPD(CV)
1	2	0.68	0.54	1.75	3	0.80	0.89	2.22	5	0.68	3.45	1.77
2	2	0.67	0.55	1.72	4	0.88	0.70	2.79	6	0.83	2.52	2.41
3	2	0.79	0.43	2.18	3	0.81	0.75	2.28	4	0.67	3.31	1.73
4	2	0.79	0.43	2.20	3	0.81	0.75	2.27	4	0.71	3.26	1.86
5	2	0.80	0.43	2.20	3	0.81	0.75	2.27	4	0.75	3.17	1.97
6	2	0.66	0.56	1.70	4	0.86	0.74	2.63	6	0.85	2.34	2.60
			NDF				ADF				Lignin	
PT	*n*	R^2	RMSECV (%)	RPD(CV)	*n*	R^2	RMSECV (%)	RPD(CV)	*n*	R^2	RMSECV (%)	RPD(CV)
1	4	0.66	6.79	1.68	2	0.67	3.04	1.72	2	0.66	0.62	1.69
2	4	0.67	6.66	1.72	2	0.68	2.99	1.75	2	0.67	0.61	1.73
3	3	0.73	5.92	1.91	2	0.73	2.69	1.92	2	0.72	0.56	1.88
4	3	0.72	5.99	1.88	2	0.72	2.72	1.90	2	0.71	0.57	1.86
5	3	0.72	5.98	1.89	2	0.73	2.72	1.90	2	0.71	0.57	1.86
6	2	0.64	6.95	1.66	2	0.68	2.97	1.76	2	0.67	0.61	1.75

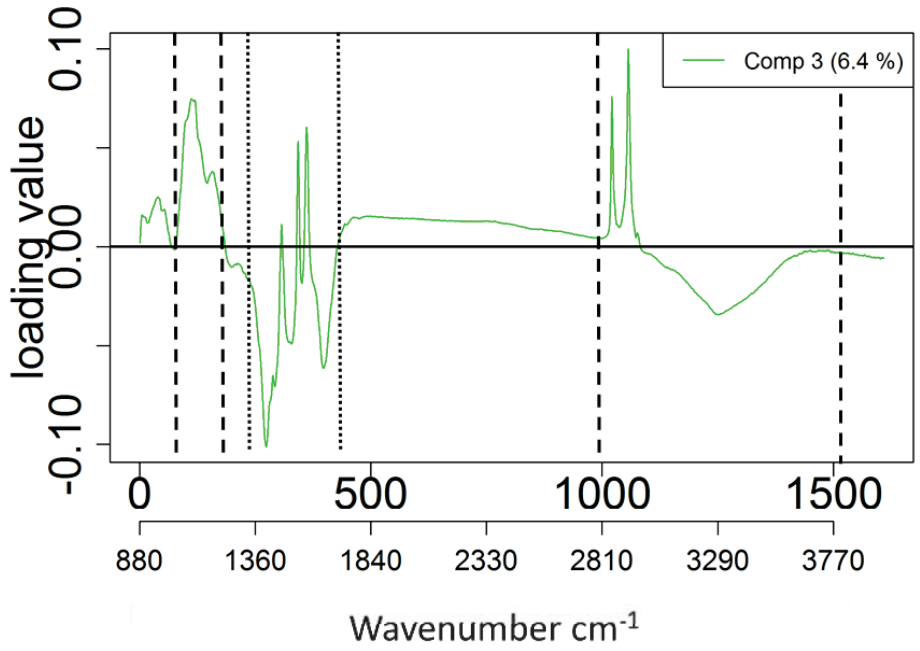

Figure 5.1.7: Loadings of the third component of the hemicellulose PT3 model.

Table 5.1.4 contains the results of PCR prediction of chemical compositions within the faecal sample set, these results are noticeably worse than those of the feature extraction based methods. The highest quality predictions of this sample set occur for the PT2 models of gross energy and organic matter. The PT3 model achieves near equivalent prediction quality amongst NDF, ADF and hemicellulose predictions. The loadings of the first two components are identical in all three PT3 models, however, hemicellulose and NDF incorporate a third component. The use of the third component in the NDF and hemicellulose models suggests this component contains contributions that are significant to both NDF and hemicellulose. This component contains three regions of distinct variable contribution, these regions occupy the wavenumber ranges of 1018 – 1194 cm^{-1}, 1355 – 1739 cm^{-1} and 2805 – 3760 cm^{-1}. The third region contains three peaks, the two on the 2800 cm^{-1} side of this region are C – H stretches whilst the third may be a combination of O – H and N – H stretching.[162] Its bandshape is closer to that of

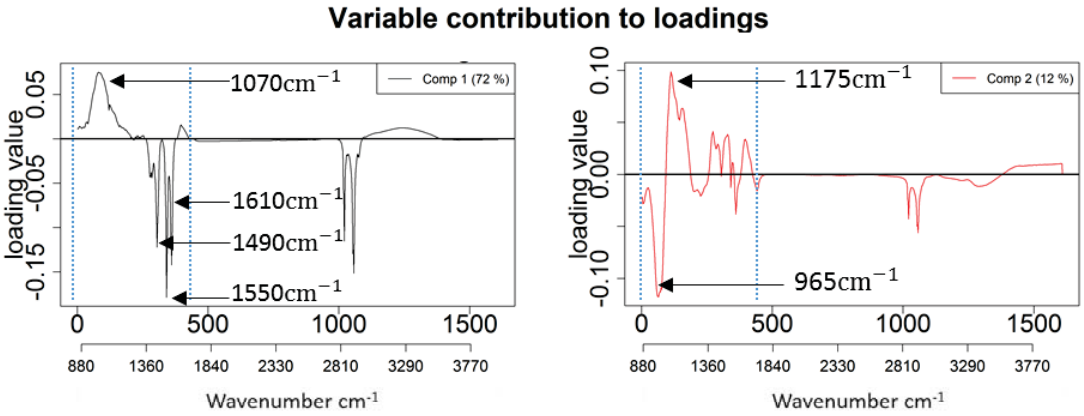

Figure 5.1.8: Loadings of components one and two for the nitrogen PT4 model.

a primary amide band at this wavenumber location. The two other regions deemed significant contain an assortment of bond contributions. The first region contains two clear peaks located at 1060 cm^{-1} the second at 1148 cm^{-1}. The prior of these two may be due to arabinan O – H and C – OH stretching, the latter a result of C – O – C asymmetric stretching of β – 1.4 glucan.[162] Both peaks therefore may originate from cellulose or hemicellulose structures, their inclusion in models of hemicellulose suggests the latter. The region of 1355 – 1739 cm^{-1} contains five peaks located at 1448 cm^{-1}, 1530 cm^{-1}, 1570 cm^{-1}, 1580 cm^{-1} and 1670 cm^{-1}. The first is likely due to cell wall polysaccharide C – H asymmetric bending, the following three peaks are in the approximate location of the amide II band (C = N and N – H) as such likely originating from plant proteins. [162] The final peak in this region is suggested to be due to C = O stretching (amide I band) suggested in the literature to be potentially due to cellulose. [162] The peaks that have been assigned from this component support the suggestion that the component may contain information of particular relevance to hemicellulose. All nitrogen models selected the optimal number of components to be two. The PT4 model performed amongst the best in terms of R^2 and the best in terms of RPD, the loading of components one and two may be seen in figure 5.1.8. The two components both possess

contributions from C – H aliphatic stretching, however, the majority of the contributions originate between 860 – 1700 cm^{-1} (dashed blue lines). Within this region of the first component, there are four significant peaks, the first of these occurs at 1070 cm^{-1} most likely due to C – O ring stretching from rhamnogalactorunan or b-galactan. [162] The three other peaks span the spectral region where absorptions due to lignin (C = C and C = O aromatic stretches) and protein (C = N and N – H stretches). [162] The most significant peak within this region is the peak at 1550 cm^{-1}, which is very likely due to the amide II band and aforementioned C = N and N – H stretches. [162] The two peaks of the second component are likely portions of the C – O ring stretching band due to rhamnogalctorunan or b – galactan. [162] Inspection of the lignin and ash models illustrates PT3 to provide relatively accurate models in both instances. The lignin PT3 model possesses sign-inverted loadings when compared to the loadings of nitrogen, indicating inverted correlation of the peaks to lignin content. Nickel content within hyperaccumulator plants was predicted most accurately by the PT1 model, however, the model quality is still clearly poor.

Table 5.1.5: Principal components regression results for the hyperaccumulator samples

Hyperaccumulators			
n	R^2	RMSECV (%)	RPD(CV)
10	0.41	0.28	1.28
5	0.06	0.37	0.98
6	0.27	0.31	1.17
6	0.27	0.31	1.17
6	0.27	0.31	1.17
6	0.34	0.30	1.23

5.2 Partial Least Squares

Partial least squares is the second of two methods based around the construction of *features* from the original predictor variables. Partial least squares is an iterative method, identifying the most relevant information to the response variables from within the predictor variables. Each iteration generates several vectors, two of these, namely the scores and loadings vectors, are combined to form a component. The iterative process begins again identifying subsequent components. These components are summed to give a constructed matrix, on which ordinary least squares may be performed. The number of components utilised is selected in the same fashion as the PCR components. The nature of the PLS process allows for a new level of insight upon inspection of the loading plots. The PCR loadings describe the proportion of variance ascribed to each variable within the predictor matrix. The PLS loadings describe how important a variable is to the response value, as such the connection between IR reflectance values and the nature of the chemical composition is more explicit. Figure 5.2.2 compares the first two components of a PT3 model, for the PLS and PCR methods. In figure 5.2.1 there are three approximate regions where distinct change occurs (aside from sign changes). Region 1 reduces whilst regions 2 and 3 grow, this result indicates that even though regions 2 and 3 account for less of the variance, they are in fact more significant to the variance of the responses than region 1.

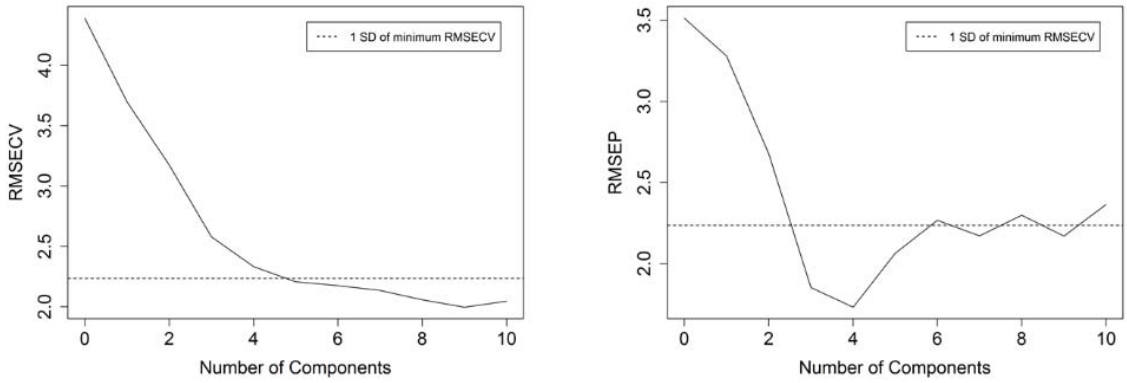

Figure 5.2.1: (a) RMSECV vs number of components for a PT2 – model. RMSEP vs Number of components for the same PT2 model.

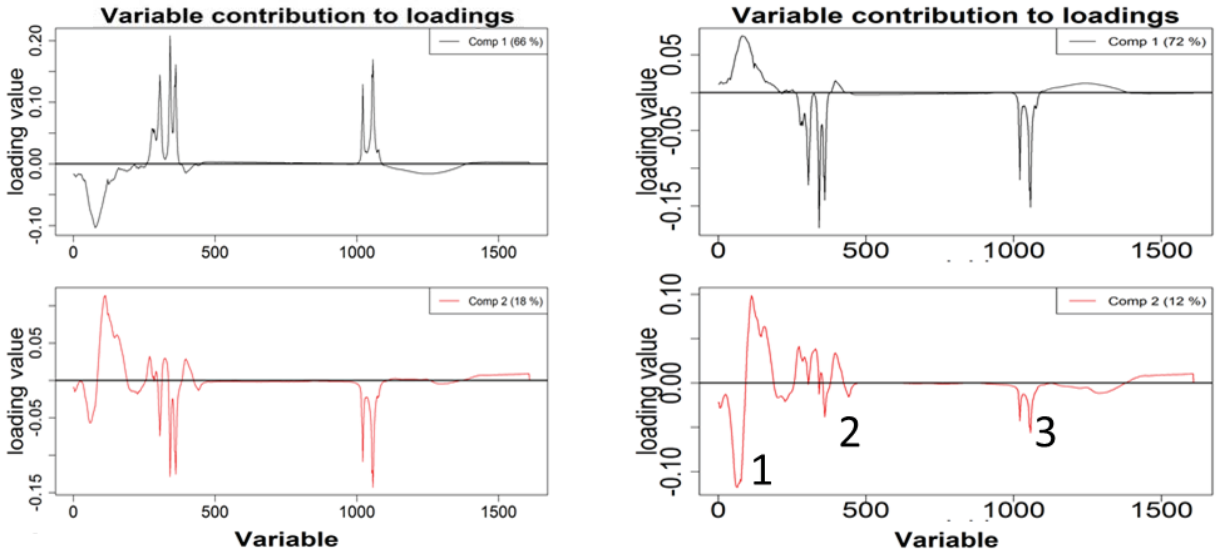

Figure 5.2.2: Loading values for the first two components of a PLS PT3 – model (left column). Loading values for the equivalent PCR PT3 model (right column).

Table 5.2.1: Partial least squares regression results for the initial forage feed samples

Initial forage feed results												
	Metabolisable Energy				NDF				Dry Matter			
PT	*n*	R^2	RMSEP (MJ/kg)	RPD(P)	*n*	R^2	RMSEP (%)	RPD(P)	*n*	R^2	RMSEP (%)	RPD(P)
1	10	0.82	0.31	2.16	8	0.90	3.84	2.17	3	0.44	1.24	1.29
2	10	0.81	0.32	2.03	9	0.88	3.89	2.29	3	0.43	1.24	1.29
3	5	0.76	0.36	1.90	5	0.94	3.06	2.59	5	0.44	1.30	1.12
4	10	0.81	0.32	2.03	9	0.88	3.89	2.29	3	0.43	1.24	1.29
5	5	0.76	0.36	1.89	5	0.94	3.07	2.57	5	0.45	1.30	1.12
6	7	0.82	0.32	1.89	8	0.90	3.69	2.27	3	0.41	1.25	1.26

	ADF				Hemicellulose				DOMD				Protein			
PT	*n*	R^2	RMSEP (%)	RPD(P)	*n*	R^2	RMSEP (%)	RPD(P)	*n*	R^2	RMSEP (%)	RPD(P)	*n*	R^2	RMSEP (%)	RPD(P)
1	8	0.72	3.02	1.68	2	0.91	2.09	3.33	10	0.82	1.92	2.16	7	0.69	2.02	1.68
2	8	0.72	3.06	1.64	2	0.92	2.05	3.39	10	0.81	2.01	2.03	7	0.70	2.02	1.67
3	5	0.81	2.51	1.98	4	0.95	1.82	3.12	5	0.76	2.22	1.90	4	0.76	1.73	1.97
4	8	0.72	3.06	1.63	2	0.92	2.05	3.39	10	0.81	2.02	2.03	7	0.70	2.02	1.67
5	5	0.81	2.51	1.97	4	0.95	1.82	3.11	5	0.76	2.23	1.89	4	0.76	1.73	1.97
6	8	0.72	3.00	1.74	3	0.93	1.94	3.60	7	0.82	2.02	1.89	6	0.55	2.80	1.22

Table 5.2.1 displays the results of PLS analysis for the initial forage feed samples. Hemicellulose and NDF have been predicted once again with high accuracy. The PT3 model demonstrates high quality predictions for hemicellulose and NDF and relatively high for ADF, in terms of both R^2 and RPD. Not only did the derivative based models perform well in terms of R^2 and RPD they also possessed the fewest components in the modelling of all chemical compositions aside from hemicellulose and dry matter. Hemicellulose requires very few components in its modelling relative to other chemical compositions. A visual comparison of the PT1 models for hemicellulose generated by both PCR and PLS can be seen in figure 5.2.3. The process of generating loadings for PLS can easily generate vastly different loadings from PCR, as mentioned previously, this is because PLS places importance on predictor variances that align well with the response variances. In the present example PLS identifies very similar components to the components identified by PCR, the main changes are the sign inversion of PCR components one and two, and the exchange of the order of PCR components two and three. This result illustrates hemicellulose content is well correlated to variances already present in the predictor matrix, it also demonstrates that whilst PCR component three accounts for around 7% of the variance, it is more relevant to the variance in hemicellulose content than PCR component two. Previously it has been common to see PT6 perform significantly worse than the other pre – treatments, in table 5.2.1 this is seen to no longer be the case, as aside from the prediction of protein content, PT6 is on par with the others. This result illustrates the ability of PLS to extract relevant information.

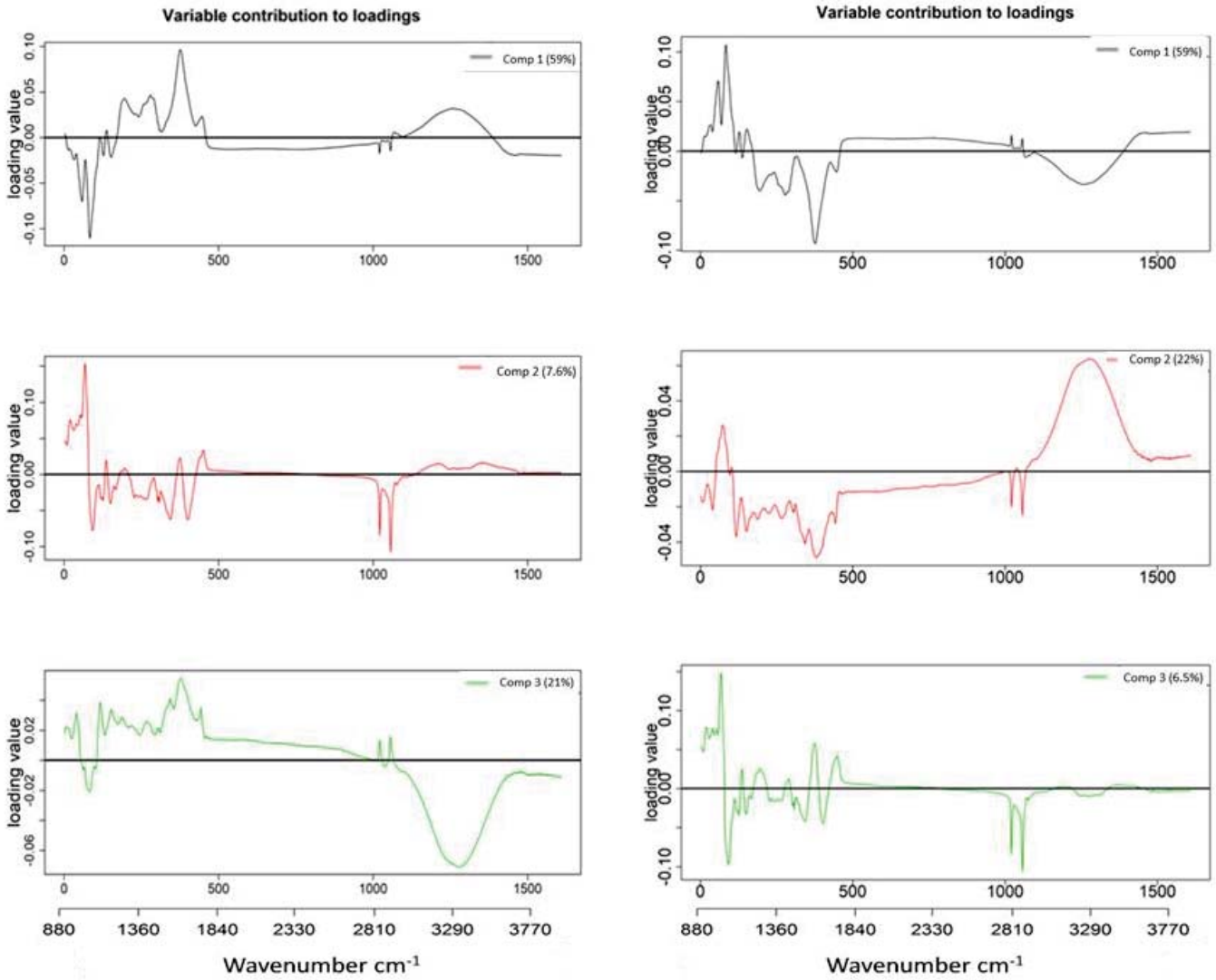

Figure 5.2.3: Loading values for PT1 hemicellulose model, displaying the first three components of the PLS model (left column) and equivalent PCR model (right column).

Table 5.2.2: Partial least squares regression results for the expanded forage feed samples

Expanded forage feed results												
	NDF				Hemicellulose							
PT	n	R^2	RMSEP (%)	RPD(P)	n	R^2	RMSEP (%)	RPD(P)				
1	9	0.68	6.20	1.71	2	0.73	3.58	1.89				
2	9	0.68	6.20	1.73	2	0.73	3.60	1.85				
3	4	0.73	5.88	1.68	3	0.75	3.50	1.91				
4	9	0.68	6.21	1.73	2	0.73	3.59	1.85				
5	4	0.73	5.87	1.68	3	0.75	3.50	1.91				
6	8	0.64	6.65	1.57	4	0.74	3.57	1.90				
	ADF				Metabolisable Energy				Dry Matter			
PT	n	R^2	RMSEP (%)	RPD(P)	n	R^2	RMSEP (MJ/kg)	RPD(P)	n	R^2	RMSEP (%)	RPD(P)
1	9	0.58	3.56	1.40	5	0.56	0.43	1.44	1	0.16	1.49	0.91
2	9	0.58	3.55	1.42	5	0.54	0.44	1.41	1	0.16	1.50	0.91
3	5	0.61	3.53	1.33	5	0.64	0.39	1.53	6	0.31	1.37	1.04
4	9	0.58	3.55	1.42	5	0.54	0.44	1.41	1	0.16	1.50	0.91
5	5	0.61	3.53	1.33	5	0.64	0.39	1.54	6	0.31	1.37	1.04
6	10	0.58	3.58	1.40	8	0.58	0.43	1.42	1	0.15	1.50	0.91

The expanded forage feed sample set was once again poorly predicted (Table 5.2.2). For hemicellulose and NDF the PT3 model generated relatively good quality results (R^2: 0.75,0.73 RPD: 1.91, 1.68). The number of components required for the modelling of hemicellulose is quite low, by comparison NDF requires several more. A comparison between the first two components of the PCR and PLS PT1 models may be seen in figure 5.2.4. The PLS model utilises only these two components whilst the PCR model requires all three. The PLS model places less significance on the O – H stretching region and more on the fingerprint region, contrary to the PCR model.

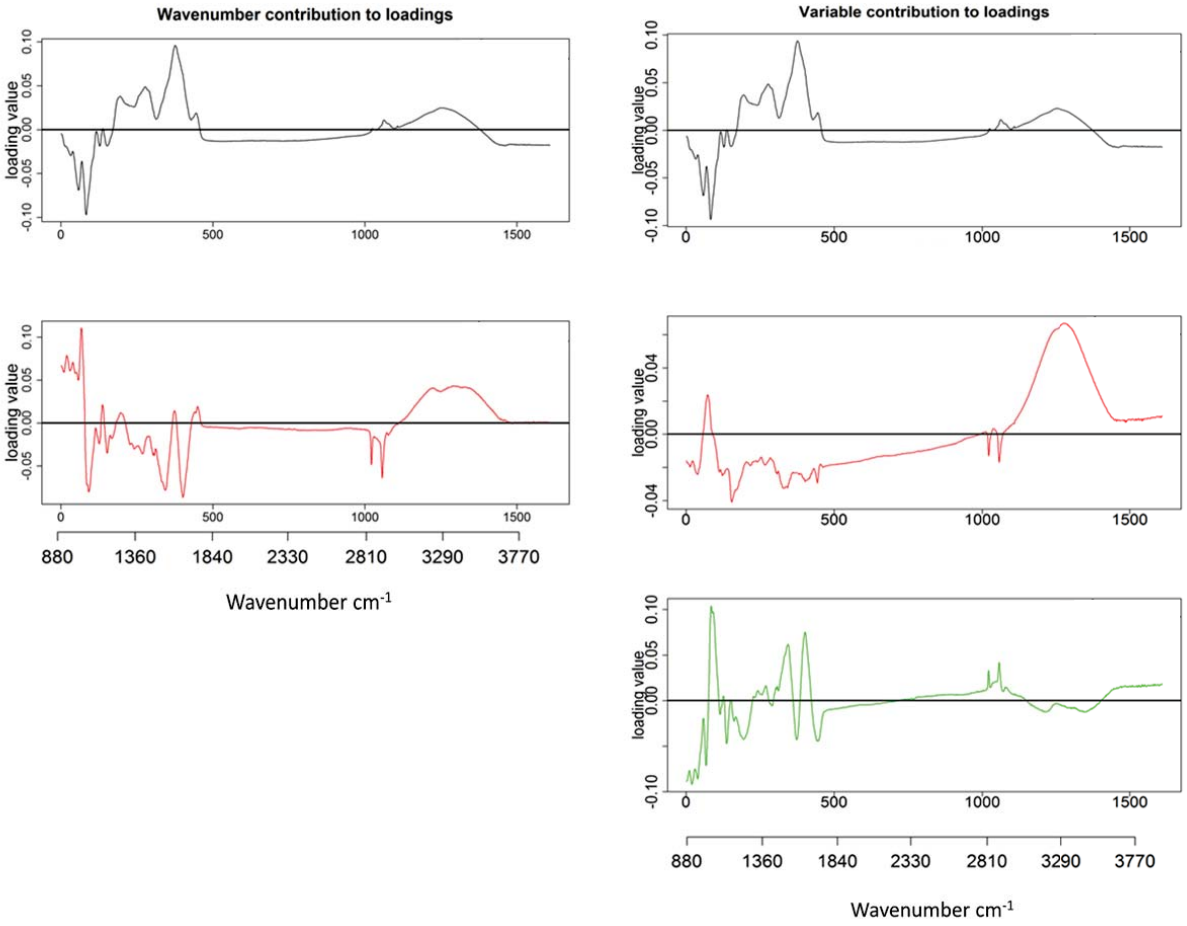

Figure 5.2.4: loadings of the two components in the PLS hemicellulose PT1 model (left column). Loadings of three components of the equivalent PCR PT1 model (right column).

Table 5.2.3: Partial least squares regression results for the chicken feed samples

Chicken feed results												
		Crude Fibre				Starch				Fat		
PT	n	R^2	RMSECV (%)	RPD	n	R^2	RMSECV (%)	RPD	n	R^2	RMSECV (%)	RPD
1	6	0.88	0.84	2.86	3	0.91	4.43	3.24	9	0.72	1.88	1.87
2	6	0.87	0.87	2.75	3	0.92	4.20	3.44	9	0.69	1.98	1.78
3	5	0.88	0.87	2.80	1	0.90	4.68	3.06	5	0.57	2.33	1.51
4	6	0.92	0.74	3.48	1	0.91	4.36	3.31	5	0.51	2.50	1.39
5	5	0.87	0.87	2.77	1	0.91	4.36	3.31	5	0.51	2.49	1.40
6	5	0.87	0.87	2.77	1	0.91	4.36	3.31	5	0.51	2.49	1.40
		Nitrogen				Ash				Dry Matter		
PT	n	R^2	RMSECV (%)	RPD	n	R^2	RMSECV (%)	RPD	n	R^2	RMSECV (%)	RPD
1	3	0.89	0.44	3.05	4	0.95	0.98	4.47	4	0.74	0.76	1.94
2	3	0.88	0.45	2.84	4	0.95	0.98	4.44	4	0.73	0.77	1.92
3	5	0.91	0.42	3.27	5	0.98	0.66	6.46	4	0.72	0.75	1.89
4	5	0.90	0.44	3.10	5	0.98	0.66	6.46	4	0.72	0.77	1.86
5	5	0.90	0.44	3.10	5	0.98	0.66	6.45	4	0.72	0.75	1.89
6	5	0.90	0.44	3.10	5	0.98	0.66	6.45	4	0.72	0.75	1.89
		Calcium				Phosphorus				Gross Energy		
PT	n	R^2	RMSECV (mg/g)	RPD	n	R^2	RMSECV (mg/g)	RPD	n	R^2	RMSECV (kJ/g)	RPD
1	5	0.96	3.59	4.71	4	0.95	1.62	4.26	7	0.71	0.52	1.85
2	5	0.95	3.67	4.61	4	0.95	1.52	4.39	8	0.69	0.53	1.80
3	5	0.98	2.29	7.10	5	0.98	1.03	6.52	7	0.76	0.48	1.99
4	5	0.98	2.32	7.02	5	0.98	1.03	6.51	7	0.77	0.47	2.02
5	5	0.98	2.31	7.05	5	0.98	1.03	6.51	7	0.76	0.48	2.01
6	5	0.98	2.31	7.05	5	0.98	1.03	6.51	7	0.76	0.48	2.01

There are three more distinct peaks within the fingerprint region of the second component of the PLS model. These occur at 1000 cm^{-1} ,1480 cm^{-1} and 1600 cm^{-1}. The first of these peaks is likely to have originated from an O – H and C – OH stretch of arabinans.[162] The second and third peaks may be attributed to the amide III and C = O aromatic stretch. Hence, all three peaks are clearly related to the structure of hemicellulose and as such their preferential inclusion is not surprising. [162] The derivative based pre – treatments provided the most accurate models in general, although often only by a small margin.

Upon inspection of the PLS results for the chicken feed samples, seen in Table 5.2.3, the ash, calcium and phosphorus results stand out immediately as very well modelled. Unfortunately, again the distribution of the response values of these chemical compositions has introduced extremely high leverage. Examining figure 5.2.5 the issues are clear, two distinct clusters may be seen within which prediction and measured values align poorly. Starch, crude fibre, and nitrogen content, are all modelled with considerable accuracy with the majority of PT's generating $R^2 \cong 0.9$ and RPD values > 2.7. Previously the PCR results for these components reflected greater variation in model quality dependent on the PT utilised. The PLS results demonstrate far greater consistency between PT models, again a consequence of PLS's superior extraction of relevant information from the predictor matrix.

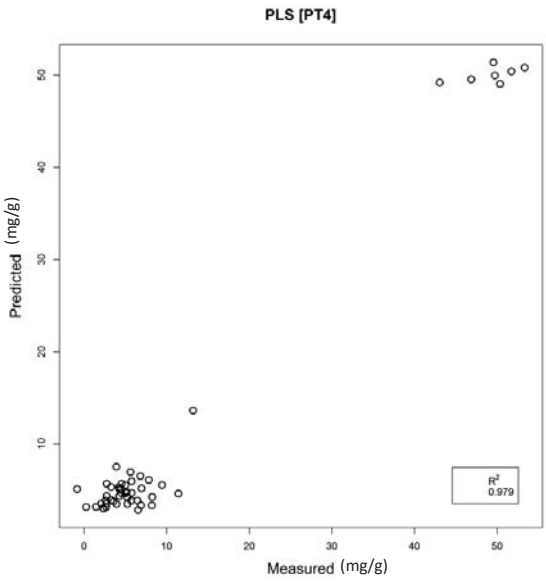

Figure 5.2.5: Predicted vs measured value plot for the calcium PT4 model.

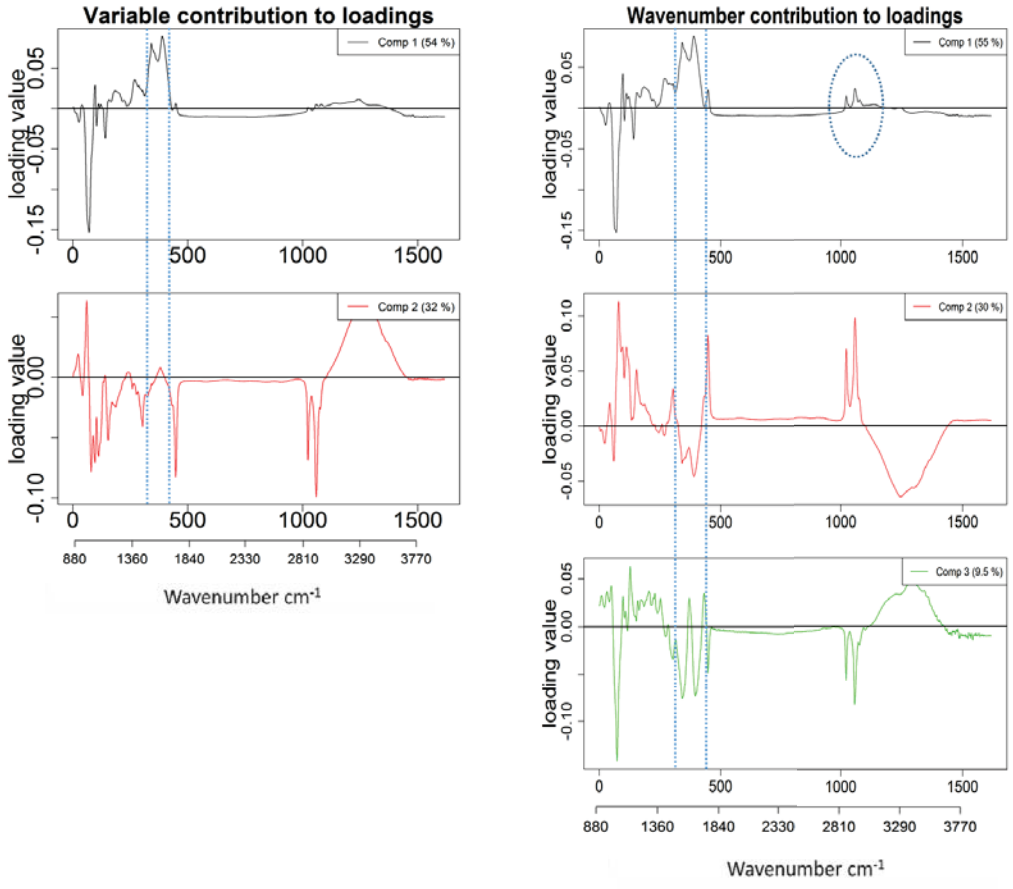

Figure 5.2.6: The loadings of the first two components of the PLS PT1 model for starch (left). Loadings of the first three components of the equivalent PCR model (right).

The PT1 models of starch and nitrogen both possess a similar or identical number of components when compared with their PCR counterparts, of these two chemical compositions the starch model demonstrates a fair improvement in model quality in the PLS results. Starch consists of a large number of glucose molecules assembled into a polysaccharide structure, as such it is expected that the superior PLS model will have placed more significance on bands common to polysaccharide molecular structures. Figure 5.2.6 contains the two components the PCR model utilises on the left hand side, and the three components utilised by the PLS model on the right. Indicated in Figure 5.2.6 are two regions of interest, one defined by the dashed blue circle the other by two sets of two vertical dashed blue lines. The former is included to indicate the only discernible difference between the two first components of each model, and that is the growth of the significance of the C – H stretch region.[160] The vertical lines enclose the spectral region of 1480 – 1720 cm^{-1}, this region represents the second most notable difference between the loading plots of each. Examining the PCR components there is a clear reduction in the loadings for this region when comparing components one and two.

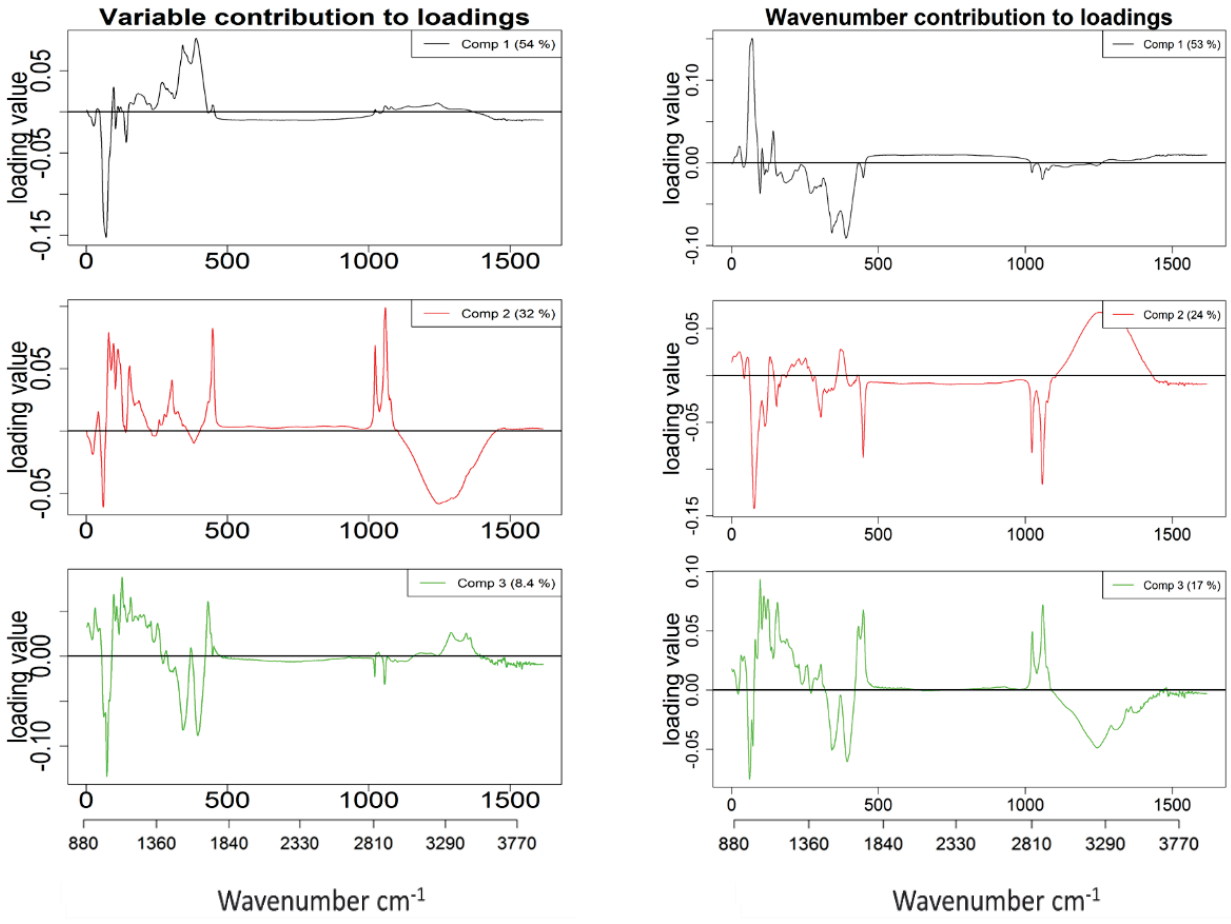

Figure 5.2.7: Loadings of the first three components of the PLS (left) and PCR (right) PT1 – models for nitrogen in chicken feed

This trend is not present in the PLS figure, where it is demonstrated that this region maintains significance across all three components. The PLS model has identified this region as significant to the starch content of the sample, and examination of the absorptions occurring in this region will shed light as to whether this is reasonable or not. The region of 1480 – 1720 cm^{-1} contains C = C aromatic stretching, the amide II band (C = N and N – H stretch), C = O aromatic stretch, and the amide I band (C = O stretch).[162] Aside from the amide II band, these absorptions are for the most part very reasonable to be of importance to a starch model. In the modelling of nitrogen content, the PT1 models for both PCR and PLS utilise three components, which allows for useful comparisons between the two. Figure 5.2.7 contains the loadings of the three components of each model, PCR on the right and PLS the left. The most distinct differences occur in the loadings of the third component, where greater loading is placed on the C – H (aliphatic) stretching and what appears to be the N – H stretching band (~3290 cm^{-1}).[160,162] Of these two changes, the dependence on the N – H stretch is expected based on the chemical composition being predicted. Fat, dry matter and gross energy are also predicted with improved accuracy using the PLS method, relative to the PCR results.

Table 5.2.4: Partial least squares regression results for the faecal samples

Faecal results												
PT	*n*	R^2	Hemicellulose RMSECV (%)	RPD(CV)	*n*	R^2	Dry Matter RMSECV (%)	RPD(CV)	*n*	R^2	Ash RMSECV (%)	RPD(CV)
1	5	0.86	2.50	2.61	7	0.60	0.40	1.50	8	0.93	1.70	3.78
2	7	0.86	2.46	2.65	8	0.58	0.41	1.46	8	0.95	1.48	4.26
3	3	0.84	2.56	2.53	6	0.67	0.35	1.71	5	0.93	1.66	3.79
4	3	0.85	2.52	2.56	6	0.66	0.35	1.68	5	0.93	1.66	3.80
5	3	0.85	2.52	2.57	6	0.66	0.35	1.69	5	0.93	1.66	3.81
6	5	0.81	2.83	2.30	7	0.60	0.39	1.53	7	0.95	1.45	4.33
PT	*n*	R^2	Nitrogen RMSECV (%)	RPD(CV)	*n*	R^2	Gross Energy RMSECV (%)	RPD(CV)	*n*	R^2	Organic Matter RMSECV (%)	RPD(CV)
1	4	0.84	0.39	2.45	4	0.94	0.50	3.92	8	0.92	1.76	3.52
2	5	0.84	0.38	2.48	5	0.94	0.44	3.90	9	0.93	1.63	3.73
3	3	0.86	0.36	2.68	4	0.94	0.41	4.14	5	0.92	1.70	3.54
4	3	0.87	0.34	2.77	4	0.94	0.41	4.12	4	0.90	1.90	3.20
5	3	0.86	0.36	2.63	4	0.94	0.41	4.12	4	0.90	1.90	3.20
6	5	0.86	0.36	2.68	5	0.93	0.45	3.75	6	0.91	1.85	3.29
PT	*n*	R^2	NDF RMSECV (%)	RPD(CV)	*n*	R^2	ADF RMSECV (%)	RPD(CV)	*n*	R^2	Lignin RMSECV (%)	RPD(CV)
1	5	0.88	4.09	2.83	4	0.77	2.49	2.08	5	0.84	0.42	2.50
2	5	0.86	4.40	2.61	4	0.76	2.54	2.05	4	0.78	0.49	2.13
3	3	0.86	4.33	2.64	3	0.82	2.21	2.34	3	0.83	0.44	2.41
4	3	0.86	4.26	2.68	3	0.82	2.19	2.37	3	0.80	0.47	2.25
5	3	0.86	4.26	2.69	3	0.82	2.19	2.37	3	0.83	0.44	2.43
6	5	0.82	4.95	2.33	2	0.70	2.86	1.81	5	0.79	0.49	2.17

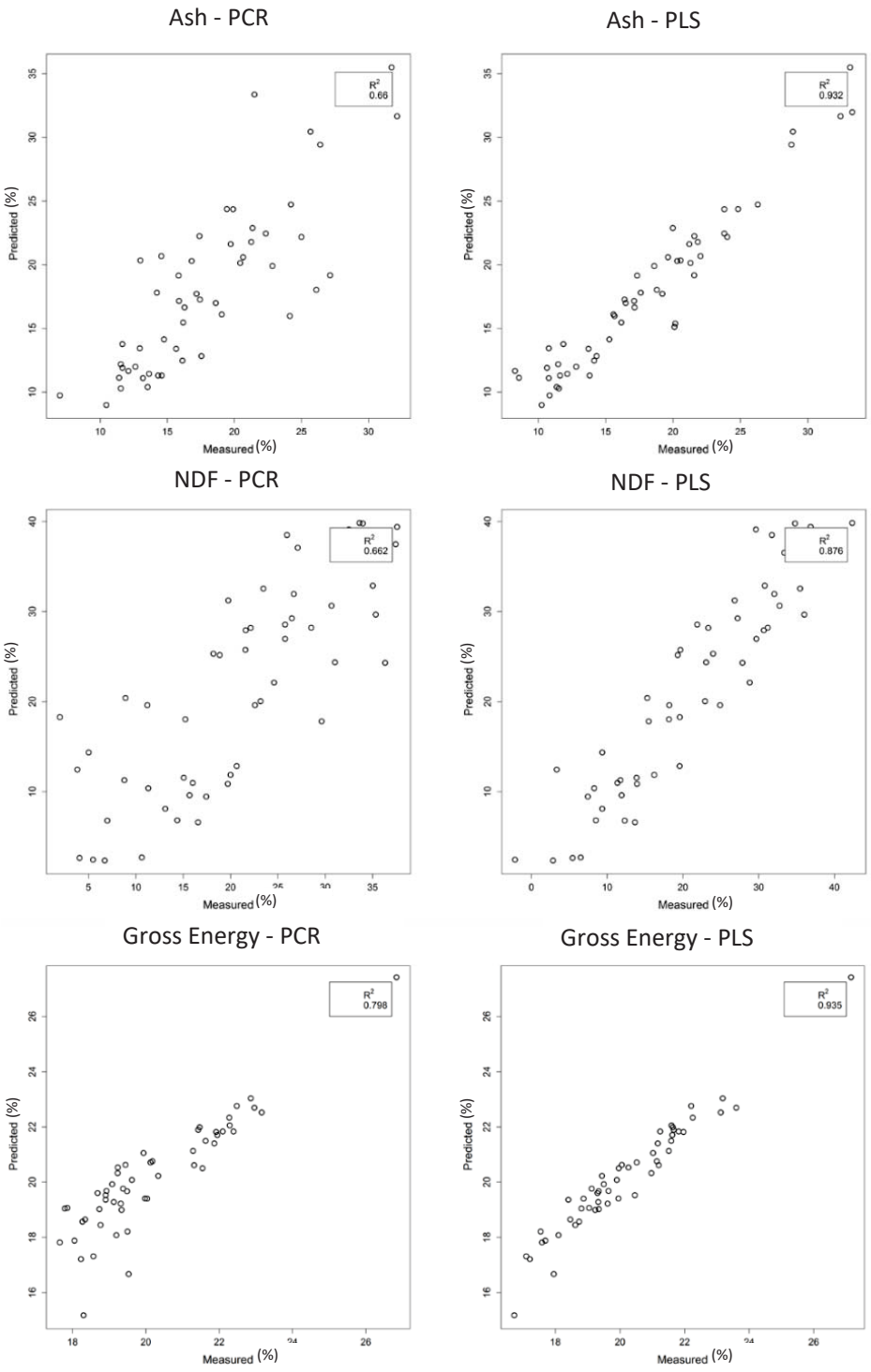

Figure 5.2.8: NDF, gross energy and ash PCR (left column) and PLS (right column) results.

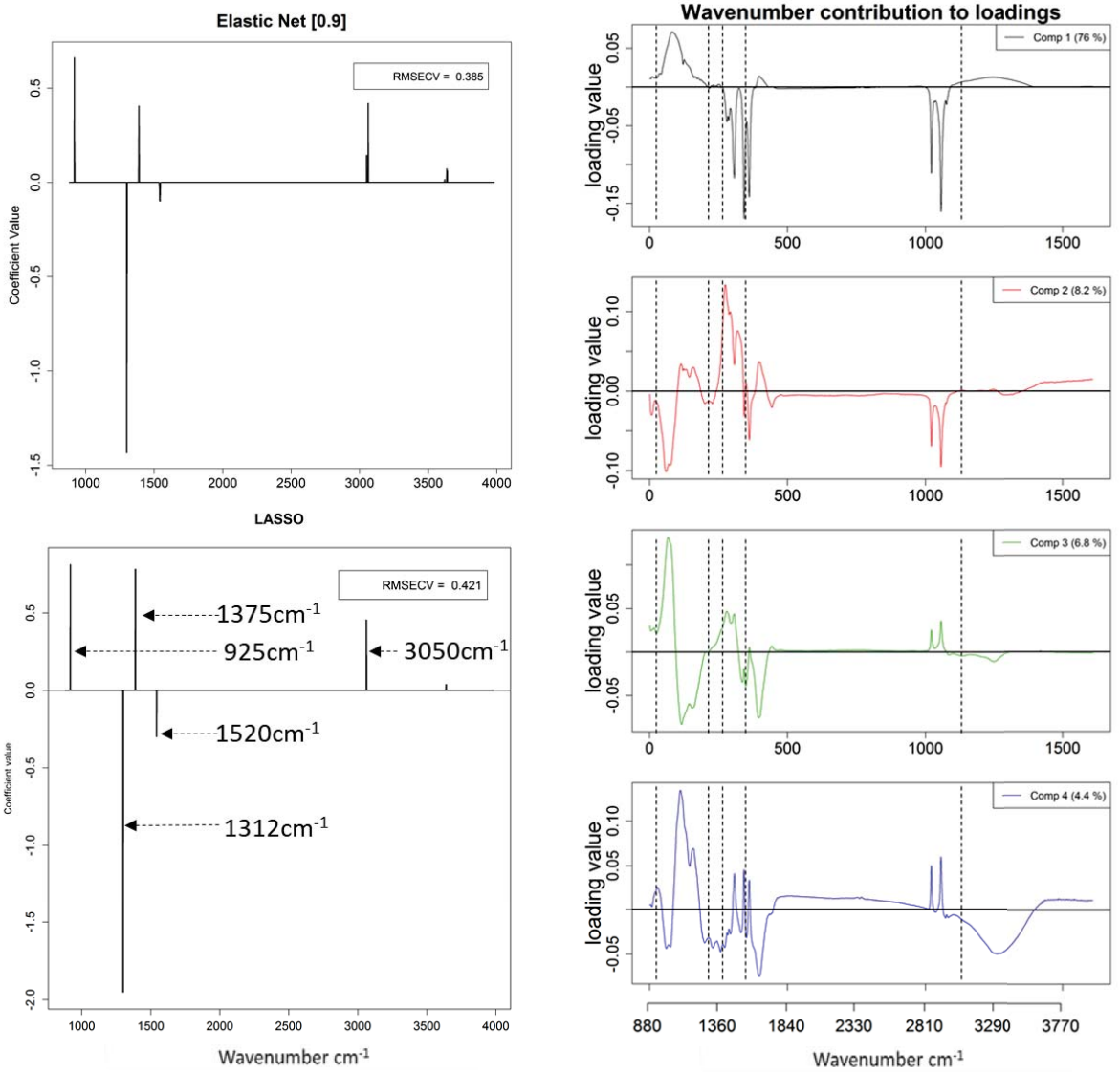

Figure 5.2.9: Top left is the coefficient plot of the gross energy PT1 – model identified using the Elastic Net method. Bottom left is the LASSO gross energy PT1 – model coefficient plot. Right column contains the loadings of the first four PLS components for the gross energy PT1 – model.

In the faecal sample set (Table 5.2.4) some tremendous improvements may be seen relative to the performance of the PCR method. This indicates that for the chemical compositions of this sample set, the information is less explicitly contained in the variances within the spectra. In figure 5.2.8 there are the specific comparisons of prediction plots of NDF, gross energy and ash with the left hand column the PCR results and the right hand column the PLS results. In order to find comparable prediction evaluation metrics, it is necessary to examine the results of the feature extraction methods. The feature extraction methods also possess an ability to sort through the predictor matrix to identify the variables most necessary for successful modelling. Figure 5.2.9 contains information relating to the PT1 models of gross energy content as constructed using LASSO, Elastic Net or PLS. The figure displays the coefficients selected by Elastic Net and LASSO on the left hand side. The Ridge method also generated high quality predictions but due to the sparsity of the Elastic Net and LASSO their coefficients are more readily identified. Interestingly there appears to be some discrepancy between the wavenumber intensities deemed significant by the Elastic Net and LASSO methods and the loading values of the regions in the PLS loading plots. The 925 cm^{-1} peak appears of no substantial importance to any of the four components in the PLS model. The peaks at 1312 cm^{-1} and 1375 cm^{-1} are of some significance to the fourth component. The peak at 1520 cm^{-1} is distinctly present in the loadings plot of the first component likely either a portion of the amide II band or a C = C aromatic stretch.[162]

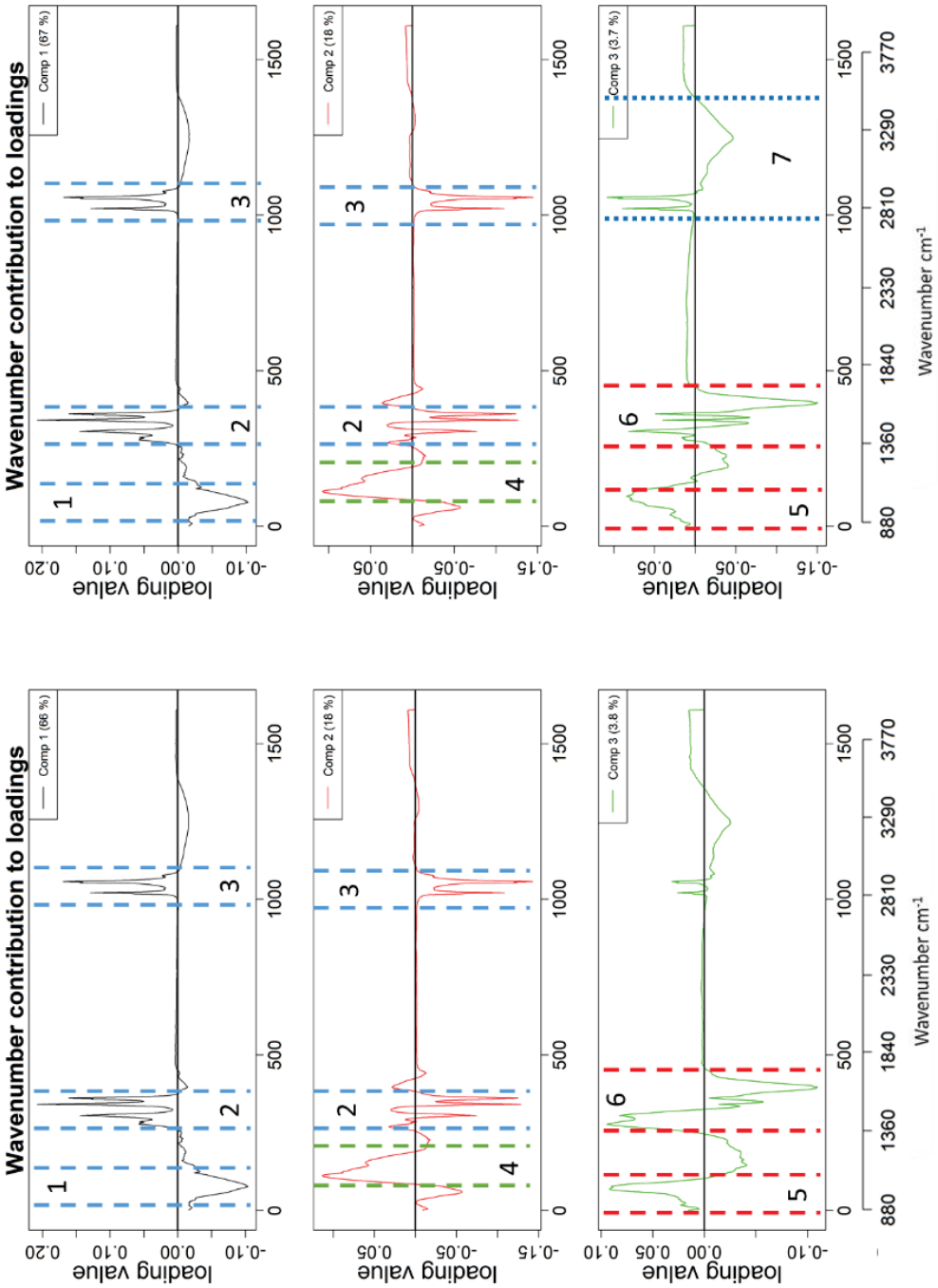

Figure 5.2.10: Left column contains loadings of the first three components of the PLS PT4 hemicellulose model. Right column contains loadings of the first three components of the PLS PT4 ADF model.

The figure 5.2.10 contains loading plots for both the hemicellulose and ADF PT4 models, as well as numbered regions bounded by dashed vertical lines. These regions are highlighted as they represent regions of more significant loading values, equivalent regions are numbered the same. Hemicellulose and ADF contain very similar loading values for the majority of regions, specifically those associated with components one and two. For the third component, values start to differ more noticeably, there are two clear differences occurring in region 6 and 7. These are the reduction of the peak near the lower wavenumber side of region six and the enlargement of all contributions in region 7 for the ADF model. Region 6 covers 1350 – 1710 cm^{-1} the peak of interest occupies 1350 – 1410 cm^{-1}. This region contains CH_2 bending absorptions at ∼1370 cm^{-1} due to cellulose and hemicellulose, and CH_2 symmetric bending at ∼1380 cm^{-1} due to cell wall polysaccharides.[162] The second region to observe change was region 7 which represents the wavenumber range of 2750 – 3470 cm^{-1}, the C – H stretching portion of this region is not particularly insightful. However, it is likely the N – H stretching occurring at 3200 – 3500 cm^{-1} is providing useful information for ADF content, as ADF contains insoluble forms of nitrogen.

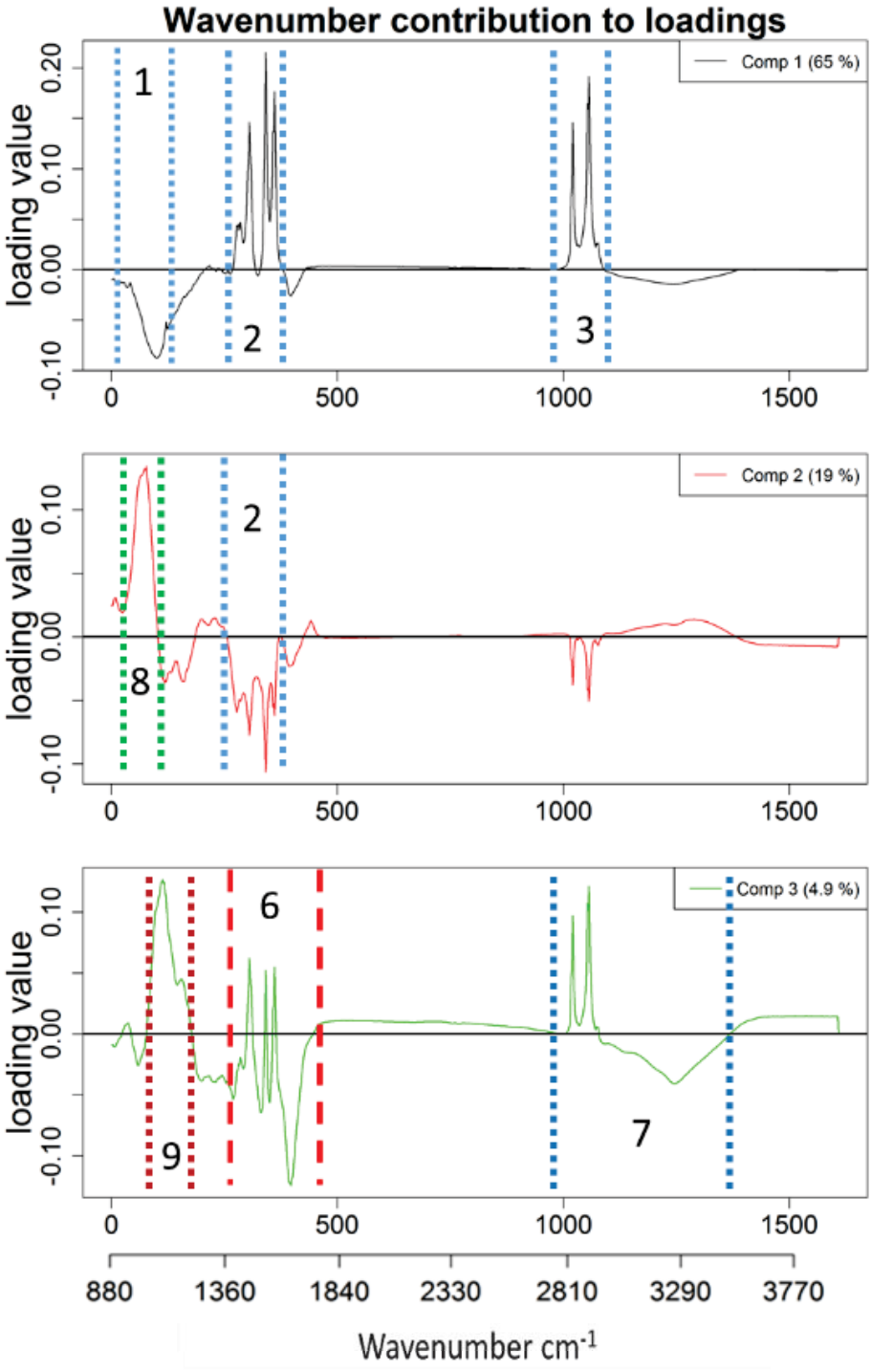

Figure 5.2.11: Loadings of the first three components of the PLS PT4 nitrogen model.

The PT4 model also produced high quality predictions of nitrogen content with relatively few components utilised. The low number of components again allows for great insight into the important spectral regions identified in the modelling. Figure 5.2.11 displays the first three loadings, utilising the same labelling scheme as before. Interestingly the loading values of each component for this model, aside from component 1, differ more substantially than what was observed between hemicellulose and ADF. This observation is logical, due to the fact hemicellulose and ADF are more inherently related to each other, than either is to nitrogen content. In the second component of the nitrogen content model, the loading values of region two appear to have reduce slightly. However, the peak within region 8 is larger than in the models of hemicellulose and ADF. Region 8 covers the wavenumber range of 890 – 1070 cm^{-1} this region contains O – H and C – OH stretches as well as C – O ring stretches, [162] so it appears to contain no explicit information regarding nitrogen content. The contributions from region 9 are likely to have been C – O ring stretching, again not explicitly related to nitrogen content.[162] Regions 6 and 7 appear to contain loading values that match very closely to those of regions 6 and 7 in the ADF model. Previously it was suggested that growth around 3200 – 3500 cm^{-1} in region 7 for the ADF model, may have been capturing variation in the insoluble nitrogen portion of ADF. The occurrence of this region in the nitrogen model appears to support that claim. The fact that region 6 also appears very similar in both the nitrogen and ADF models suggests this region may also be utilised to model the nitrogen portion of ADF.

Table 5.2.5: Partial least squares regression results for the hyperaccumulator samples

Hyperaccumulators			
n	R^2	RMSECV	RPD(CV)
7	0.48	0.26	1.37
6	0.32	0.30	1.14
5	0.54	0.25	1.46
5	0.54	0.25	1.46
5	0.54	0.25	1.46
5	0.54	0.25	1.46

Unfortunately, as table 5.2.5 demonstrates, the hyperaccumulators were poorly modelled. It is perhaps worth noting the far more even distribution of R^2 and RPD values relative to any previous regression method. The metrics are generally better than have been produced in any other method. Calibration results demonstrated R^2 values as large as 0.7, yet without the support of cross – validation results this means little.

Chapter 6.0: Conclusions and Future work

6.1 Conclusions

Overall the impact of pre-treatment selection had the least impact in the context of PLS models. Amongst the remaining models, AsLS seemed to perform poorly. The poor performance of AsLS may be due to the subjective nature of the parameter selection process. As selection is manual this leads to potential for incorrect baseline evaluation in some spectra, which may have negatively impacted the results. Overall the most accurately modelled chemical component was faecal gross energy, using the Elastic Net PT1 model (R^2: 0.97, RPD: 5.55). Across all models, a wide variety of chemical components were modelled with a fair degree of accuracy (R^2 >0.80, RPD ≥2.5). Ridge and PCR seemed to generate the lowest quality results of all five regression methods. The LASSO, Elastic Net and PLS methods all performed very well. Amongst LASSO, Elastic Net and PLS the use of SNV and Savitsky – Golay filtering (PT3) seems to most frequently produce the best results.

This thesis has demonstrated that MIR reflectance methods along with appropriate data pre-treatment gives performance as good as, or better than, existing techniques such as NIR (in the context of forages[168,169]), specifically in terms of interpretability and model robustness. It is suggested these are the most important advantages of MIR over NIR reflectance for predicting sample chemical components. MIR reflectance spectroscopy combines the low cost and efficiency of NIR spectroscopic methods with better prediction of sample chemical components due to the rich spectral features of the MIR region. Existing handheld MIR spectrometers offer the potential for *in situ* sample analysis using hand-held MIR devices with appropriate data pre-treatments such as PT3.

6.2 Future work

It would be a useful venture to expand the chicken feed, faecal and hyperaccumulator sample sets in order to allow for the generation of an external validation subset. The external validation data would provide a more reliable representation of the prediction of unknown samples. There is reason to believe there may be other interesting response values to attempt to model in the hyperaccumulator samples. This avenue may prove more fruitful than the prediction of nickel content. The use of a diamond crystal rather than the germanium crystal may also provide an opportunity for the formulation of more accurate models. This is suggested due to the lower refractive index of diamond allowing for greater penetration of the evanescent wave into the sample. [170] Exploring the performance of ATR – FTIR based predictions of fresh pasture samples or samples prepared using *in situ* methods would be useful. The expanded forage feed samples possessed relatively poor predictions, as such, further investigation of factors that might influence model transferability such as soil types and growing conditions is required. The analysis of samples using a hand – held ATR-FTIR instrument would be useful as a means of truly identifying the potential of this technique for *in situ analysis*.

The regression models utilised in this research project may be classed as linear regression models. As such it is suggested that the investigation of other methods such as the *Genetic Algorithm* and *Support Vector Machine*, may offer in some instances superior performance. Likewise, the inclusion of variable selection in the PLS analysis may bolster its prediction performance in the external validation setting. It may be possible to implement a built – in variable selection feature using loading vector l_1 - *norm* constraints. A large variety of variable selection methods currently exist in the context of PLS, and they warrant investigation.

Chapter 7.0: Appendices

Appendix 1: Asymmetric Least Squares Process

$$\boldsymbol{f} = \boldsymbol{W}|\boldsymbol{y} - \boldsymbol{z}|^2 + \lambda|\boldsymbol{D}\boldsymbol{z}|^2 \quad (1)$$

The first term of $\boldsymbol{f}$ is a product of a squared difference vector and a weight matrix, which is a weighted squared difference vector, $\boldsymbol{v}$. The second term of $\boldsymbol{f}$ describes the smoothness, it does so by generating a vector which contains in its elements values relating to the difference of some element z_i with some preceding element z_{i-1}. This is achieved using the following matrix, $\boldsymbol{D}$.

$$\boldsymbol{D} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \quad (2)$$

$\boldsymbol{D}$ will, in reality, have k columns and $k - 1$ rows, λ is a parameter that represents the significance of the smoothness term. If λ is large the vector $\boldsymbol{z}$ will have more freedom to undulate. So the second term will generate a second vector $\boldsymbol{d}$ which describes the squared difference between elements in $\boldsymbol{z}$ as multiplied by λ.

So $\boldsymbol{f}$ is now,

$$\boldsymbol{f} = \boldsymbol{v} + \boldsymbol{d} \quad (3)$$

One key observation is that both terms depend on the vector $\boldsymbol{z}$, and that we would like to identify a vector $\boldsymbol{z}$ which is close to $\mathbf{y}$ as possible (small $\boldsymbol{v}$) but also as smooth as possible (small $\boldsymbol{d}$). Hence, the determination of $\boldsymbol{z}$ is nothing more than a question of solving the first order derivative

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{z}^{\boldsymbol{T}}} = 0$$

for $\boldsymbol{z}$.

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{z}^{\boldsymbol{T}}} = \frac{\partial}{\partial \boldsymbol{z}^{\boldsymbol{T}}}\left(\boldsymbol{W}|\boldsymbol{y}-\boldsymbol{z}|^2 + \lambda|\boldsymbol{D}\boldsymbol{z}|^2\right)$$

$$= \frac{\partial}{\partial \boldsymbol{z}^{\boldsymbol{T}}}\left(\boldsymbol{W}\left(\boldsymbol{y}^{\boldsymbol{T}}-\boldsymbol{z}^{\boldsymbol{T}}\right)(\boldsymbol{y}-\boldsymbol{z}) + \lambda \boldsymbol{D}^{\boldsymbol{T}}\boldsymbol{z}^{\boldsymbol{T}}\boldsymbol{D}\boldsymbol{z}\right)$$

$$= \boldsymbol{W}\frac{\partial}{\partial \boldsymbol{z}^{\boldsymbol{T}}}\left(\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{y} - \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{z} - \boldsymbol{z}^{\boldsymbol{T}}\boldsymbol{y} + \boldsymbol{z}^{\boldsymbol{T}}\boldsymbol{z}\right) + \frac{\partial}{\partial \boldsymbol{z}^{\boldsymbol{T}}}\lambda \boldsymbol{D}^{\boldsymbol{T}}\boldsymbol{z}^{\boldsymbol{T}}\boldsymbol{D}\boldsymbol{z}$$

$$= -2\boldsymbol{W}\boldsymbol{y} + 2\boldsymbol{W}\boldsymbol{z} + 2\lambda \boldsymbol{D}^{*}\boldsymbol{D}\boldsymbol{z}$$

$$= \boldsymbol{0}$$

Yields,

$$\boldsymbol{W}\boldsymbol{y} = (\boldsymbol{W} + \lambda \boldsymbol{D}^{\boldsymbol{T}}\boldsymbol{D})\boldsymbol{z} \qquad (4)$$

A vector $\boldsymbol{z}$ satisfying the above equation will be a vector that minimises the balance between similarity to $\boldsymbol{y}$ and smoothness.

$$\boldsymbol{z} = (\boldsymbol{W} + \lambda \boldsymbol{D}^{*}\boldsymbol{D})^{-\mathbf{1}}\boldsymbol{W}\boldsymbol{y} \qquad (5)$$

The process begins with an estimate of $\boldsymbol{z}$, typically an evenly weighted function with some restriction on the smoothness.

$$\boldsymbol{z} = (\boldsymbol{I} + \lambda \boldsymbol{D}^{*}\boldsymbol{D})^{-\mathbf{1}}\boldsymbol{y} \qquad (6)$$

The corresponding weights are then calculated, then using these weights and equation (7.5) we can identify a new $\boldsymbol{z}$**.** This process repeats until the calculated weights match those of the preceding iteration, this typically occurs in under 10 iterations.

Appendix 2: Discrete Wavelet Transform – Brief Example

A brief example of generating basis wavelet and scaling functions using the Gram – Schmidt method will be developed now in the context of Haar[165] wavelets. The process begins first with the scaling function, $\boldsymbol{\phi_n}$ defined as follows.

$$\boldsymbol{\phi_p} = [1 \quad 1 \quad 1 \quad 1]^T$$

Where 2^n denotes the window size, in this instance $n = p$ where $2^p = N$. The $n = p/2$ may be seen below, it is clear n defines the extent of scaling.

$$\boldsymbol{\phi_{p/2}} = [1 \quad 1 \quad 0 \quad 0]^T$$

The Gram-Schmidt process sets the first basis vector as follows.

$$\boldsymbol{\phi'_0} = \frac{\boldsymbol{\phi'_p}}{\left|\boldsymbol{\phi'_p}\right|}$$

The second is identified as

$$\boldsymbol{\phi'_1} = \boldsymbol{\phi_{p/2}} - \frac{\boldsymbol{\phi^T_{p/2}\phi'_0}}{\boldsymbol{\phi'^T_0\phi'_0}}\boldsymbol{\phi'_0}$$

$$\boldsymbol{\phi'_1} = \frac{\boldsymbol{\phi'_1}}{|\boldsymbol{\phi'_1}|}$$

This results in the following orthonormal vectors. Combining to form the relevant $\boldsymbol{W}$ yields,

$$\boldsymbol{W} = \begin{bmatrix}\boldsymbol{\phi'_0}\\ \boldsymbol{\phi'_1}\end{bmatrix} = \begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\frac{1}{2}$$

The first row of $\boldsymbol{W}$ above contains positive terms only whilst the second contains alternating signs. In fact this trend extends to any general $N \times N$ matrix $\boldsymbol{W}$ where the first row will always contain all positive terms. In the discrete wavelet transform this

row is referred to as the *scaling* function. The other row is referred to as the detail function or mother wavelet. The DWT decomposes a function, $\boldsymbol{f}$ by projecting it onto the orthogonal basis as it is shifted along $\frac{N}{2^n}$ times. The function is decomposed iteratively, using the algorithm of Mallat. There is a total of n decompositions calculated. Each, n is referred to as a decomposition *level*. The scaling function for a given decomposition level, accounts for low *frequency* trends in the function. Whilst the detail vector accounts for higher *frequency* trends. The process of decomposition is outlined below in equation form. [79]

$$\boldsymbol{a^n = Sa^{n-1}}$$

$$\boldsymbol{SS^T = \mathrm{I}}$$

$$\boldsymbol{d^n = Da^{n-1}}$$

$$\boldsymbol{DD^T = \mathrm{I}}$$

Where $\boldsymbol{a}$ and $\boldsymbol{d}$ are the scaling and detail coefficients respectively. The matrices $\boldsymbol{S}$ and $\boldsymbol{D}$ contain the scaling and detail functions, these matrices are $\frac{N}{2^n} \times N$. Each row of $\boldsymbol{S}$ or $\boldsymbol{D}$ corresponds to a shift of the scaling of detail function by one interval relative to the above row. The general wavelet transform equation can then be rewritten as follows.

$$\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{d} \end{bmatrix} = \begin{bmatrix} \boldsymbol{S} \\ \boldsymbol{D} \end{bmatrix} \boldsymbol{f}$$

Scaling and Wavelet Shift Matrices Example

Considering $N = 8$ the $\boldsymbol{S}$ and $\boldsymbol{D}$ matrices of the Haar system would be.

$$\boldsymbol{S_1} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\boldsymbol{D_1} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

$$\boldsymbol{S_2} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\boldsymbol{D_2} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

$$\boldsymbol{S_3} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \end{bmatrix}$$

$$\boldsymbol{D_3} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \end{bmatrix}$$

Appendix 3: Properties of OLS residuals and Markov – Gauss assumptions.

There is a selection of interesting properties and assumptions regarding the residuals and disturbances of the OLS model and true model respectively. These properties will now be outlined. Firstly, the expression derived during the minimisation of the squared residual term may be expanded as follows

$$(\boldsymbol{X^T X})\boldsymbol{\beta_{ols}} = \boldsymbol{X^T y} = \boldsymbol{X^T}(\boldsymbol{X\beta_{ols}} + \boldsymbol{e})$$

$$(\boldsymbol{X^T X})\boldsymbol{\beta_{ols}} = (\boldsymbol{X^T X})\boldsymbol{\beta_{ols}} + \boldsymbol{X^T e} = (\boldsymbol{X^T X})\boldsymbol{\beta_{ols}}$$

Which implies

$$\boldsymbol{X^T e} = 0$$

This result provides several insights regarding the residuals

1. The inner product of the variables of $\boldsymbol{X}$ and the residuals is 0. This implies the variables themselves are uncorrelated to the residuals.

2. If there is a constant term, $\boldsymbol{\beta_0}$ the first column of $\boldsymbol{X}$ will be all 1's. This implies $\sum \boldsymbol{e_i} = 0$.

3. If the sum of the residuals is zero, then the mean must be zero also.

Since $\boldsymbol{e_{avg}} = \frac{\sum \boldsymbol{e_i}}{n} = \frac{0}{n} = 0$

4. If the mean residual is zero then the regression hyperplane will pass perfectly through the means of $\boldsymbol{X}$ and $\boldsymbol{y}$, as

$$\boldsymbol{y_{avg}} = \boldsymbol{X_{avg}\beta_{ols}} + \boldsymbol{e_{avg}} = \boldsymbol{X_{avg}\beta_{ols}}$$

5. The predicted values of $\boldsymbol{y}$ are equal to $\boldsymbol{X\beta_{ols}}$ and as such,

$$\boldsymbol{y^T e} = \left(\boldsymbol{\beta_{ols}}^{\boldsymbol{T}} \boldsymbol{X^T}\right)\boldsymbol{e} = 0$$

Which implies the predicted values are uncorrelated to the residuals.

Markov-Gauss assumptions

These assumptions pertain predominantly to the nature of the *disturbances* of the true model.

1. $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$

   The first of the Markov - Gauss assumptions, and simply put it is the assumption there is in fact a linear relationship between response and predictor values.

2. $\boldsymbol{X}$ is a $n \times k$ matrix of *full column rank*

3. $\boldsymbol{E}[\boldsymbol{\epsilon}\,|\boldsymbol{X}\,] = \mathbf{0}$

   This assumption asserts that the average effect of the disturbances will be zero for observations of $\boldsymbol{X}$, or equally that the observations of the predictor variables provide no insight regarding the disturbances. This assumption leads to the fact that the average response value will be modelled with no disturbance.

4. $\boldsymbol{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'|\boldsymbol{X}] = \sigma^2\mathbf{I}$

   This result illustrates that disturbances have equal variance at any predictor variable, and are entirely uncorrelated to one another therefore only have entries on the diagonal.

5. $\boldsymbol{X}$ may be fixed or random, so long as it's generation does not relate to $\boldsymbol{\epsilon}$

Appendix 4: Derivation of the Eigenvalue Equations for PLS

The numerical properties of these vectors become clearer when the relations are expanded relative to the vectors of the preceding iteration, denoted as $n-1$. We see:

$$\boldsymbol{u_n} = \frac{\boldsymbol{Yc_n}}{\boldsymbol{c_n^T c_n}}$$

$$= \frac{\boldsymbol{Y}}{\boldsymbol{c_n^T c}}\left(\frac{\boldsymbol{Y^T t_n}}{\boldsymbol{t_n^T t_n}}\right) = \frac{\boldsymbol{YY^T}}{(\boldsymbol{c_n^T c_n})(\boldsymbol{t_n^T t_n})}\left(\frac{\boldsymbol{Xw_n}}{\boldsymbol{w_n^T w_n}}\right) = \frac{\boldsymbol{YY^T X}}{(\boldsymbol{c_n^T c_n})(\boldsymbol{t_n^T t_n})(\boldsymbol{w_n^T w_n})}\left(\frac{\boldsymbol{X^T u_{n-1}}}{\boldsymbol{u_{n-1}^T u_{n-1}}}\right)$$

$$\boldsymbol{c_n} = \frac{\boldsymbol{Y^T t_n}}{\boldsymbol{t_n^T t_n}}$$

$$= \frac{\boldsymbol{Y^T}}{\boldsymbol{t_n^T t_n}}\left(\frac{\boldsymbol{Xw_n}}{\boldsymbol{w_n^T w_n}}\right) = \frac{\boldsymbol{Y^T}}{\boldsymbol{t_n^T t_n w_n^T w_n}}\left(\frac{(\boldsymbol{XX^T u})}{\boldsymbol{u_n^T u_n}}\right) = \frac{\boldsymbol{Y^T XX^T}}{(\boldsymbol{t_n^T t_n})(\boldsymbol{u_n^T u_n})(\boldsymbol{w_n^T w_n})}\left(\frac{\boldsymbol{Yc_{n-1}}}{(\boldsymbol{c_{n-1}^T c_{n-1}})}\right)$$

$$\boldsymbol{t_n} = \frac{\boldsymbol{Xw_n}}{\boldsymbol{w_n^T w_n}}$$

$$= \frac{\boldsymbol{X}}{\boldsymbol{w_n^T w_n}}\left(\frac{\boldsymbol{X^T u_n}}{\boldsymbol{u_n^T u_n}}\right) = \frac{\boldsymbol{XX^T}}{(\boldsymbol{u_n^T u_n})(\boldsymbol{w_n^T w_n})}\left(\frac{\boldsymbol{Yc_n}}{\boldsymbol{c_n^T c_n}}\right) = \frac{\boldsymbol{XX^T Y}}{(\boldsymbol{u_n^T u_n})(\boldsymbol{c_n^T c_n})(\boldsymbol{w_n^T w_n})}\left(\frac{\boldsymbol{Y^T t_{n-1}}}{\boldsymbol{t_{n-1}^T t_{n-1}}}\right)$$

$$\boldsymbol{w_n} = \frac{\boldsymbol{X^T u}}{\boldsymbol{u_n^T u_n}}$$

$$= \frac{\boldsymbol{X^T}}{\boldsymbol{u_n^T u_n}}\left(\frac{\boldsymbol{Yc_n}}{\boldsymbol{c_n^T c_n}}\right) = \frac{\boldsymbol{X^T Y}}{(\boldsymbol{u_n^T u_n})(\boldsymbol{c_n^T c_n})}\left(\frac{\boldsymbol{Y^T t_n}}{\boldsymbol{t_n^T t_n}}\right) = \frac{\boldsymbol{X^T YY^T}}{(\boldsymbol{u_n^T u_n})(\boldsymbol{c_n^T c_n})(\boldsymbol{t_n^T t_n})}\left(\frac{\boldsymbol{Xw_{n-1}}}{\boldsymbol{w_{n-1}^T w_{n-1}}}\right)$$

The factor of $\frac{\boldsymbol{1}}{\boldsymbol{w_{n-1}^T w_{n-1}}}$ occurs as result of the required normalisation for the previous iteration. The above relationships become eigenvalue equations at the point of convergence, where for some vector $\boldsymbol{v}$

$$\boldsymbol{v_n = v_{n-1}}$$

The numerators are various matrices and the denominators are scalars all of which are equal via the associativity of scalars. We then acquire the following eigenvalue equations at convergence;

$$\boldsymbol{YY^T XX^T\, u} = a\boldsymbol{u}$$

$$\boldsymbol{Y^T XX^T Y\, c} = a\boldsymbol{c}$$

$$\boldsymbol{XX^T YY^T t} = a\boldsymbol{t}$$

$$\boldsymbol{X^T YY^T Xw} = a\boldsymbol{w}$$

Where $a = (\boldsymbol{c_n^T c_n})(\boldsymbol{t_n^T t_n})(\boldsymbol{w_n^T w_n})(\boldsymbol{u_n^T u_n})$

Appendix 5: Geometric Properties of the PLS Vectors [11]

The residual matrix $\boldsymbol{X_i}$ is related to the previous residual matrix $\boldsymbol{X_{i-1}}$ as follows;

$$\boldsymbol{X_i = X_{i-1} - t_{i-1} p_{i-1}^T}$$

Which can be expanded as;

$$\boldsymbol{X_i = X_{i-1} - t_{i-1} \left(\frac{X_{i-1}^T t_{i-1}}{t_{i-1}^T t_{i-1}} \right)^T}$$

$$\boldsymbol{= X_{i-1} - \frac{t_{i-1} t_{i-1}^T X_{i-1}}{t_{i-1}^T t_{i-1}}}$$

$$\boldsymbol{= \left[\mathbf{I} - \frac{\mathbf{t_{i-1}} t_{i-1}^T}{t_{i-1}^T t_{i-1}} \right] X_{i-1}}$$

$$\boldsymbol{= \left[\mathbf{I} - \frac{\mathbf{t_{i-1}} t_{i-1}^T}{t_{i-1}^T t_{i-1}} \right] \left[X_{i-2} - \frac{t_{i-2} t_{i-2}^T X_{i-2}}{t_{i-2}^T t_{i-2}} \right]}$$

The result generalises to $\boldsymbol{X_i = Z\, X_{i-n}}$ where $\boldsymbol{Z}$ is some matrix.

Property 1

The vectors $\boldsymbol{w_i}$ and $\boldsymbol{w_j}$ are orthogonal when i and j are different iterative steps.

$$\boldsymbol{w_i^T w_j = 0}, \qquad \boldsymbol{i \neq j}$$

Proof

If $i < j$

We have

$$\boldsymbol{X_j = Z\left[X_i - \frac{t_i t_i^T X_i}{t_i^T t_i}\right]}$$

Let's first show that $\boldsymbol{X_j\, w_i}$ = 0, because if this holds then we can see that via rearranging the previous eigenvalue equation for $\boldsymbol{w}$ and substituting into $\boldsymbol{w_i^T\, w_j}$ we will arrive at a product of $\boldsymbol{X_j}$ and $\boldsymbol{w_i}$ which if equal to zero will imply $\boldsymbol{w_i^T w_j = 0}$ and hence weight vectors of different iterations are orthogonal.

Using the above relation for $\boldsymbol{X_j}$

$$\boldsymbol{X_j w_i = Z\left[X_i - \frac{t_i t_i^T X_i}{t_i^T t_i}\right] w_i}$$

$$\boldsymbol{= Z\left[X_i w_i - \frac{t_i t_i^T X_i w_i}{t_i^T t_i}\right];\;\; t_i = X_i w_i}$$

Hence

$$\boldsymbol{X_j w_i = Z\left[t_i - \frac{t_i\left(t_i^T t_i\right)}{t_i^T t_i}\right] = Z[t_i - t_i] = 0}$$

As has already been alluded to, the following result can be found. Using the convergence relation and taking the transpose we find;

$$\boldsymbol{X_j^T Y_j Y_j^T X_j w_j} = a_j \boldsymbol{w_j}$$

$$\boldsymbol{w_j^T X_j^T Y_j Y_j^T X_j} \ = a_j \boldsymbol{w_j^T}$$

$$\frac{\boldsymbol{w_j^T X_j^T Y_j Y_j^T X_j}}{a_j} \ = \boldsymbol{w_j^T}$$

We can show that the inner product of $\boldsymbol{w_j}$ and $\boldsymbol{w_i}$ will be zero as;

$$\boldsymbol{w_j^T w_i} = \frac{\boldsymbol{w_j^T X_j^T Y_j Y_j^T X_j w_i}}{a_j} = 0; as \ \boldsymbol{X_j w_i} = 0$$

This proves the orthogonality of weight vectors from different iterations.

Property 2

$$\boldsymbol{t_i^T t_j} = 0 \text{ for } \boldsymbol{i \neq j}$$

Proof

If $\boldsymbol{i < j}$

$$\boldsymbol{X_j = X_{j-1} - \frac{t_{j-1}\, t_{j-1}^T X_{j-1}}{t_{j-1}^T t_{j-1}} = X_{j-1} - \frac{(X_{j-1} w_{j-1}) t_{j-1}^T X_{j-1}}{t_{j-1}^T t_{j-1}}}$$

$$\boldsymbol{= X_{j-1}\left[\mathbf{I} - \frac{w_{j-1} t_{j-1}^T X_{j-1}}{t_{j-1}^T t_{j-1}}\right]}$$

$$\boldsymbol{= X_{j-2}\left[\mathbf{I} - \frac{w_{j-2} t_{j-2}^T X_{j-2}}{t_{j-2}^T t_{j-2}}\right]\left[\mathbf{I} - \frac{w_{j-1} t_{j-1}^T X_{j-1}}{t_{j-1}^T t_{j-1}}\right]}$$

$$\boldsymbol{= X_{j-n}\prod_{l=0}^{n+1}\left[\mathbf{I} - \frac{w_{j-(n-l)} t_{j-(n-l)}^T X_{j-(n-l)}}{t_{j-(n-l)}^T t_{j-(n-l)}}\right]}$$

Which becomes after $\boldsymbol{n_i = j - (i+1)}$ steps.

$$\boldsymbol{X_{i+1}\prod_{l=0}^{n_i+1}\left[\mathbf{I} - \frac{w_{j-(n_i-l)} t_{j-(n_i-l)}^T X_{j-(n_i-l)}}{t_{j-(n_i-l)}^T t_{j-(n_i-l)}}\right] = X_{i+1} H}$$

$$\boldsymbol{t_i^T t_j = t_i^T X_j w_j;}$$

$$\boldsymbol{t_i^T X_j = t_i^T X_{i+1} H = t_i^T\left[X_i - \frac{t_i t_i^T X_i}{t_i^T t_i}\right] H}$$

$$\boldsymbol{= \left[t_i^T X_i - \frac{t_i^T t_i t_i^T X_i}{t_i^T t_i}\right] H = 0 \; for \; i < j}$$

$$\boldsymbol{which \; proves \; t_i^T t_j = 0, \qquad i \neq j}$$

Property 3

$\boldsymbol{w_i^T p_j} = 0, i < j$

Proof

$$\boldsymbol{p_j} = \frac{\boldsymbol{X_j^T t_j}}{\boldsymbol{t_j^T t_j}}$$

$$\boldsymbol{p_j} = \frac{[\boldsymbol{Z X_{i+1}}]^{\boldsymbol{T}} \boldsymbol{t_j}}{\boldsymbol{t_j^T t_j}} = \frac{\boldsymbol{X_{i+1}^T Z^T t_j}}{\boldsymbol{t_j^T t_j}}$$

$$\boldsymbol{w_i^T p_j} = \frac{\boldsymbol{w_i^T X_{i+1}^T Z^T t_j}}{\boldsymbol{t_j^T t_j}} = \frac{(\boldsymbol{X_{i+1} w_i})^{\boldsymbol{T}} \boldsymbol{Z^T t_j}}{\boldsymbol{t_j^T t_j}}$$

$$\boldsymbol{i < i + 1 < j}$$

$$\boldsymbol{X_{i+1} w_i} = \left[\boldsymbol{X_i} - \frac{\boldsymbol{t_i t_i^T X_i}}{\boldsymbol{t_i^T t_i}}\right] \boldsymbol{w_i} = \left[\boldsymbol{X_i w_i} - \frac{\boldsymbol{t_i t_i^T X_i w_i}}{\boldsymbol{t_i^T t_i}}\right] = \left[\boldsymbol{t_i} - \frac{\boldsymbol{t_i}(\boldsymbol{t_i^T t_i})}{\boldsymbol{t_i^T t_i}}\right] = 0$$

Therefore

$$\boldsymbol{w_i^T p_j} = \frac{[\boldsymbol{0}]^{\boldsymbol{T}} \boldsymbol{Z^T t_j}}{\boldsymbol{t_i^T t_i}} = 0$$

Appendix 6: Residual Matrices Derivations.

$$X_{i+1}^T X_{i+1} = \left(X_i - t_i p_i^T\right)^T \left(X_i - t_i p_i^T\right)$$
$$= X_i^T X_i - X_i^T t_i p_i^T - p_i t_i^T X_i + p_i \left(t_i^T t_i\right) p_i^T$$
$$t_i^T X_i = \left(t_i^T t_i\right) p_i^T;$$
$$X_{i+1}^T X_{i+1} = X_i^T X_i - X_i^T t_i p_i^T - p_i \left(t_i^T t_i\right) p_i^T + p_i \left(t_i^T t_i\right) p_i^T$$
$$= X_i^T X_i - X_i^T t_i p_i^T$$
$$= X_i^T X_i - \left(t_i^T X_i\right)^T p_i^T = X_i^T X_i - p_i \left(t_i^T t_i\right) p_i^T = X_i^T X_i - p_i p_i^T \left(t_i^T t_i\right)$$

$$Y_{i+1}^T Y_{i+1} = \left(Y_i - t_i c_i^T\right)^T \left(Y_i - t_i c_i^T\right)$$
$$= Y_i^T Y_i - Y_i^T t_i c_i^T - c_i t_i^T Y_i + c_i t_i^T t_i c_i^T$$
$$Y_i^T t_i = \left(t_i^T t_i\right) c_i;$$
$$Y_i^T Y_i = Y_i^T Y_i - \left(c_i \left(t_i^T t_i\right) c_i^T\right) - c_i \left(c_i \left(t_i^T t_i\right)\right)^T + c_i t_i^T t_i c_i^T$$
$$= Y_i^T Y_i - c_i c_i^T \left(t_i^T t_i\right)$$

$$Y_{i+1}^T X_{i+1} = \left(Y_i - t_i c_i^T\right)^T \left(X_i - t_i p_i^T\right)$$
$$= Y_i X_i - Y_i^T t_i p_i^T - c_i t_i^T X_i + c_i t_i^T t_i p_i^T$$
$$p_i^T = \frac{t_i^T X_i}{t_i^T t_i}; t_i^T X_i = \left(t_i^T t_i\right) p_i^T$$
$$Y_{i+1}^T X_{i+1} = Y_i X_i - Y_i^T t_i p_i^T - c_i \left(t_i^T t_i\right) p_i^T + c_i t_i^T t_i p_i^T$$
$$= Y_i^T X_i - Y_i^T t_i p_i^T = Y_i^T \left(X_i - t_i p_i^T\right) = Y_i^T X_{i+1}$$

$$X_{i+1}^T Y_{i+1} = \left(X_i - t_i p_i^T\right)^T \left(Y_i - t_i c_i^T\right)$$
$$= X_i^T Y_i - X_i^T t_i c_i^T - p_i t_i^T Y_i + p_i t_i^T t_i c_i^T$$
$$= X_i^T Y_i - X_i^T t_i c_i^T - p_i \left(c_i^T \left(t_i^T t_i\right)\right) + p_i t_i^T t_i c_i^T$$
$$= X_i^T Y_i - X_i^T t_i c_i^T$$
$$= X_i^T \left(Y_i - t_i c_i^T\right) = X_i^T Y_{i+1}$$

Appendix 7: Demonstration of repeatability of ATR-FTIR

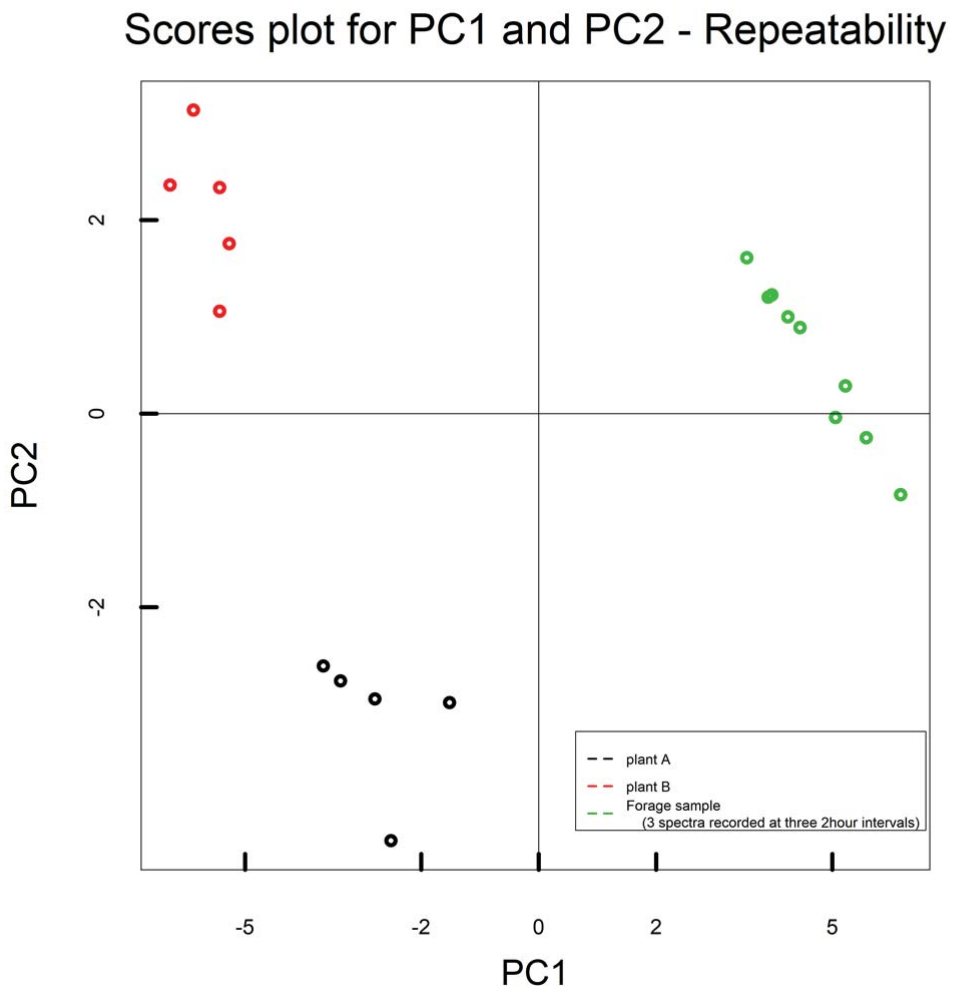

Figure 8.1: Demonstration of close clustering of spectral measurements made over a day (green dots). Distinctly different from other species displayed.

Appendix 8: Tables of Wet Chemistry Results

Table 8.1: Forage feed wet chemistry results.

	As Fed Basis						
sample	DM %	Protein %	Hemicellulose %	NDF %	ADF %	In vitro DOMD%	ME MJ/kg
4A-7-2013	93.98	13.62	20.54	47.92	27.37	55.09	8.81
5A-8-2013	94.13	12.71	24.05	47.87	23.81	56.44	9.03
7A-1-2013	93.90	9.29	25.45	52.96	27.50	53.02	8.48
8A-2-2013	94.08	7.73	25.59	53.27	27.68	54.25	8.68
10A-11-2013	93.42	16.17	8.78	27.18	18.40	65.55	10.48
11A-12-2013	93.62	18.22	9.65	31.56	21.91	63.88	10.22
12A-5-2013	92.98	14.83	7.73	25.25	17.52	65.30	10.44
14A-6-2013	93.24	15.04	5.97	24.99	19.01	59.44	9.51
20A-9-2013	92.76	16.06	13.20	38.20	24.99	60.86	9.73
21A-10-2013	93.50	19.80	12.26	37.76	25.49	62.59	10.01
26A-3-2013	92.52	16.50	11.87	36.47	24.59	60.29	9.64
27A-4-2013	93.37	17.78	10.40	32.83	22.43	67.90	10.86
TN12-139-1-12-Sep	91.78	23.21	18.93	39.25	20.31	58.36	9.33
TN12-139-1A-10-Nov	93.61	14.86	19.44	47.21	27.77	56.18	8.99
TN12-139-1A-11-Jun	92.57	24.04	17.69	41.20	23.50	57.90	9.26
TN12-139-2-12-Sep	92.13	22.28	19.90	39.77	19.86	58.57	9.37
TN12-139-2A-11-Jun	92.73	25.27	17.46	38.61	21.15	58.47	9.35
TN12-139-4A-25-Jul	91.79	22.61	19.00	39.90	20.90	58.24	9.31
TN12-139-6A-25-Jul	92.27	25.37	19.25	39.89	20.64	56.44	9.03
TN12-139-10-12-Sep	92.05	17.86	8.74	25.65	16.90	63.71	10.19
TN12-139-13-12-Sep	92.12	15.09	7.78	23.83	16.05	64.64	10.34
TN12-139-20-12-Sep	91.41	21.41	12.86	26.25	13.38	62.17	9.94
TN12-139-22-12-Sep	91.50	16.79	9.59	26.25	16.66	62.26	9.96
TN12-139-1-12-Sep			0.00				
TN12-139-1A-10-Nov	92.47	13.96	23.58	44.32	20.73	57.39	9.18
TN12-139-1A-11-Jun	93.44	13.56	18.36	42.46	24.10	58.94	9.43
TN12-139-2-12-Sep	92.84	19.51	7.77	21.89	14.11	65.70	10.51
TN12-139-2A-11-Jun	92.75	17.43	7.86	23.74	15.88	63.95	10.23
TN12-139-4A-25-Jul	91.51	17.90	6.96	23.40	16.44	63.29	10.12
TN12-139-6A-25-Jul	93.34	13.75	9.46	29.84	20.37	62.91	10.06
TN12-139-10-12-Sep	92.27	17.01	12.07	33.71	21.64	59.65	9.54
TN12-139-13-12-Sep	92.58	19.61	8.61	26.07	17.45	63.99	10.23
TN12-139-20-12-Sep	92.05	13.82	12.52	29.47	16.95	63.47	10.15
TN15-98-01	96.04		20.16	54.16	34.00	51.73	8.27
TN15-98-02	96.13		20.04	52.08	32.04	55.52	8.88
TN15-98-03	93.68		15.35	36.79	21.44	61.11	9.77
TN15-98-04	95.64		17.26	39.73	22.46	62.06	9.93
TN15-98-05	95.86		19.43	40.38	20.94	61.35	9.81

Table 8.1: Forage feed wet chemistry results (continued).

sample	DM %	Protein %	Hemicellulose %	NDF %	ADF %	In vitro DOMD %	ME MJ/kg
TN15-98-06	95.34		20.20	46.68	26.47	58.05	9.28
TN15-98-07	95.43		21.79	40.58	18.79	62.16	9.94
TN15-98-08	96.51		20.16	50.10	29.94	56.31	9.01
TN15-98-09	95.04		21.55	45.15	23.59	57.83	9.25
TN15-98-10	94.59		23.50	48.93	25.42	55.57	8.89
TN15-98-11	94.59		19.78	39.02	19.23	60.30	9.64
TN15-99-01	94.51		17.82	47.48	29.65	53.87	8.61
TN15-99-02	95.16		16.31	48.53	32.21	55.69	8.91
TN15-99-03	94.05		21.19	41.00	19.80	58.66	9.38
TN15-99-04	94.04		18.59	37.76	19.16	59.21	9.47
TN15-99-05	92.92		4.13	15.82	11.69	67.03	10.72
TN15-99-06	94.09		3.82	15.51	11.68	68.53	10.96
TN15-99-07	92.93		4.11	15.73	11.62	67.43	10.78
TN15-99-08	92.72		3.87	17.46	13.59	66.07	10.57
TN15-99-09	94.46		5.00	19.25	14.25	67.39	10.78
TN15-99-10	94.75		7.73	22.69	14.96	66.63	10.66
TN15-99-11	94.29		6.92	21.47	14.54	66.46	10.63
TN15-99-12	94.33		8.54	23.42	14.88	65.85	10.53
TN15-99-13	95.76		18.62	53.92	35.29	52.40	8.38
TN15-99-14	95.64		18.86	49.22	30.36	56.12	8.98
TN15-99-15	95.08		20.58	43.73	23.14	58.46	9.35
TN15-99-16	94.44		19.93	40.54	20.60	60.89	9.74
TN15-99-17	96.49		8.16	26.00	17.83	66.68	10.66
TN15-99-18_2	96.94		9.64	30.36	20.72	64.27	10.28
TN15-99-19_2	96.59		6.00	24.19	18.19	66.80	10.68
TN15-99-20	96.72		7.09	25.14	18.04	67.15	10.74
TN15-99-21	95.66		5.19	28.16	22.97	64.86	10.37
TN15-99-22	95.54		6.98	27.42	20.44	65.16	10.42
TN15-99-23	96.13		11.50	45.35	33.85	57.82	9.25
TN15-99-24	95.31		7.16	28.68	21.51	63.53	10.16
TN15-99-25	96.00		23.62	49.24	25.61	56.44	9.03
TN15-99-26	96.32		21.69	49.20	27.51	57.16	9.14
TN15-99-27	94.85		19.97	43.21	23.24	59.36	9.49
TN15-99-28	95.45		19.86	50.08	30.21	55.48	8.87
TN15-104-01	96.31		17.67	39.70	22.03	62.22	9.95
TN15-104-02	96.46		17.39	40.77	23.38	61.86	9.89
TN15-104-03	96.65		19.18	42.55	23.37	61.91	9.90
TN15-104-04	96.54		17.07	40.16	23.09	61.98	9.91
TN15-104-05	96.79		19.31	46.30	26.98	59.04	9.44

Table 8.1: Forage feed wet chemistry results (continued).

sample	DM %	Protein %	Hemicellulose %	NDF %	ADF %	In vitro DOMD %	ME MJ/kg
TN15-104-06	93.98		18.53	41.61	23.08	60.61	9.69
TN15-104-07	94.48		20.77	42.65	21.88	59.33	9.49
TN15-104-08	94.63		22.07	45.06	22.98	58.15	9.30
TN15-104-09	93.96		21.47	46.77	25.30	56.54	9.04
TN15-104-10	96.43		22.01	44.33	22.32	60.01	9.60
TN15-246-01	97.26		20.33	41.12	20.79	65.94	10.55
TN15-246-02	97.36		24.27	54.55	30.27	55.75	8.92
TN15-246-03	97.21		22.72	46.16	23.44	60.98	9.75
TN15-246-04	96.55		15.25	37.76	22.50	63.18	10.10
TN15-246-05	96.98		17.67	38.74	21.06	66.51	10.64
TN15-246-06	97.40		22.03	46.94	24.90	59.47	9.51
TN15-96-01	94.35		24.74	47.92	23.17	56.89	9.10
TN15-96-02	94.92		23.51	49.50	25.99	56.73	9.07
TN15-96-03	94.89		21.56	52.67	31.11	54.60	8.73
TN15-96-04	94.57		22.91	49.27	26.36	55.37	8.85
TN15-96-05	93.75		6.03	22.41	16.38	64.69	10.35
TN15-96-06	94.13		6.32	24.12	17.80	64.67	10.34
TN15-96-07	95.04		-6.64	23.12	29.77	65.70	10.51
TN15-96-08	94.86		6.45	22.15	15.69	65.58	10.49
TN15-96-09	95.75		23.39	54.71	31.31	54.74	8.75
TN15-96-10	96.17		24.27	51.20	26.92	56.86	9.09
TN15-96-11	95.65		27.12	54.36	27.24	55.30	8.84
TN15-96-12	96.07		27.76	55.10	27.34	56.00	8.96
TN15-96-13	96.10		7.96	31.15	23.19	63.28	10.12
TN15-96-14	96.05		8.62	31.45	22.82	63.16	10.10
TN15-96-15	96.58		7.14	28.15	21.01	63.50	10.16
TN15-96-16	95.24		7.67	28.62	20.94	63.21	10.11
TN15-96-17	97.72		26.00	57.32	31.32	55.82	8.93
TN15-96-18	98.13		27.79	61.75	33.95	54.16	8.66
TN15-96-19	98.18		28.20	59.07	30.87	55.70	8.91
TN15-96-20	98.08		26.63	59.72	33.09	55.82	8.93
TN15-96-21	95.74		9.08	31.56	22.48	62.83	10.05
TN15-96-22	95.38		5.72	27.01	21.28	65.26	10.44
TN15-96-23	95.76		7.66	35.09	27.42	61.15	9.78
TN15-96-24	97.62		7.71	36.05	28.34	62.42	9.98
TN15-96-25	92.47		19.28	36.23	16.95	60.42	9.66
TN15-96-26	92.70		18.74	36.30	17.56	61.15	9.78
TN16-565-11	92.90	18.70	19.89	43.95	24.06	56.12	8.98
TN16-565-12	94.40	18.88	20.79	44.08	23.29	56.51	9.04

Table 8.1: Forage feed wet chemistry results (continued).

sample	DM %	Protein %	Hemicellulose %	NDF %	ADF %	In vitro DOMD %	ME MJ/kg
TN16-565-13	95.90	13.23	11.28	47.46	36.18	55.74	8.91
TN16-565-14	96.40	12.88	16.82	54.01	37.19	54.64	8.74
TN16-565-15	95.60	14.23	10.93	41.68	30.75	59.14	9.46
TN16-565-16	93.60	14.29	8.71	42.31	33.60	57.19	9.15
TN16-565-17	93.80	13.77	24.48	54.57	30.08	51.68	8.27
TN16-565-21	94.70	14.66	11.34	49.98	38.63	53.51	8.56
TN16-565-22	93.50	20.69	6.46	33.05	26.58	61.41	9.82
TN16-565-23	94.20	21.47	10.86	43.89	33.02	55.42	8.86
TN16-565-02	92.40	18.46	21.95	47.85	25.89	54.70	8.75
TN16-565-24	93.90	16.60	8.87	42.45	33.58	57.97	9.27
TN16-566-01	94.70	19.66	22.50	48.69	26.19	59.64	9.54
TN16-566-02	93.40	16.48	21.58	46.24	24.65	58.32	9.33
TN16-566-03	94.90	17.20	22.52	49.77	27.24	57.99	9.27
TN16-566-04	95.10	16.10	24.57	53.54	28.97	57.01	9.12
TN16-566-05	93.50	15.90	7.26	33.72	26.46	67.24	10.75
TN16-566-06	93.10	16.26	6.95	35.16	28.21	66.17	10.58
TN16-566-07	93.50	14.89	7.36	36.55	29.18	64.70	10.35
TN16-566-08	89.60	19.35	6.03	32.27	26.23	64.03	10.24
TN16-565-09	93.70	15.68	21.62	49.53	27.91	54.63	8.74
TN16-565-10	94.30	20.07	19.77	44.51	24.73	56.46	9.03
TN16-565-06	92.05	15.92	9.25	32.35	23.10	61.28	9.80
TN16-565-07	92.31	14.48	10.57	35.98	25.40	59.54	9.52
TN16-565-08	92.42	14.40	8.76	30.20	21.44	62.84	10.05
TN16-565-01	91.76	18.35	20.35	43.68	23.32	55.71	8.91
TN16-565-03	91.67	18.70	19.88	43.95	24.06	56.12	8.98
TN16-565-18	93.73	18.36	20.35	48.11	27.75	54.83	8.77
TN16-565-19	94.68	17.33	20.54	46.69	26.14	55.58	8.89

Table 8.2: Table of wet chemistry results for the additional samples present in the expanded forage feed set.

	As Fed Basis				
Sample	DM %	Hemicellulose %	ME %	ADF %	NDF%
TN15-623-01	95.79	16.48	9.63	28.37	44.85
TN15-623-02	95.52	18.73	9.81	24.17	42.90
TN15-623-03	95.07	19.02	9.84	22.61	41.64
TN15-623-04	95.45	16.64	9.75	21.68	38.33
TN15-623-09	96.35	21.12	9.79	25.35	46.47
TN15-623-10	96.16	21.82	9.39	29.83	51.66
TN15-623-11	96.22	20.57	9.71	25.85	46.43
TN15-623-12	96.21	21.95	9.68	25.46	47.41
TN15-624-01	95.16	21.14	10.02	24.81	45.96
TN15-624-02	95.23	18.72	10.31	22.99	41.71
TN15-624-03	94.18	23.14	9.78	26.25	49.39
TN15-624-04	95.14	18.98	10.43	22.43	41.41
TN15-624-05	93.45	8.64	10.78	18.95	27.59
TN15-624-06	93.67	8.72	10.80	18.36	27.08
TN15-624-07	94.37	9.28	10.69	18.49	27.78
TN15-624-08	94.77	7.92	10.68	19.11	27.03
TN15-624-09	93.81	6.82	10.87	17.39	24.21
TN15-624-10	93.65	9.09	10.66	18.76	27.85
TN15-624-11	94.00	9.66	10.73	17.74	27.40
TN15-624-12	94.72	7.86	10.85	17.54	25.41
TN15-625-02	94.61	21.94	10.15	23.75	45.70
TN15-625-03	94.66	22.46	9.98	23.93	46.40
TN15-625-05	94.24	9.29	10.61	20.72	30.01
TN15-625-06	94.43	9.17	10.67	19.73	28.90
TN15-625-07	94.32	12.03	10.69	18.87	30.90
TN15-625-08	94.42	9.95	10.55	20.14	30.09
TN15-626-02	92.02	19.26	10.2	23.90	43.17
TN15-626-05	91.46	25.29	9.69	31.62	56.92
TN15-626-08	93.54	21.62	9.36	30.29	51.91
TN15-626-11	94.08	24.56	9.95	28.38	52.95

Table 8.2: Table of wet chemistry results for the additional samples present in the expanded forage feed set (continued).

Sample	DM %	Hemicellulose %	ME %	ADF %	NDF %
TN15-626-14	94.20	22.79	9.06	31.08	53.88
TN15-626-17	92.56	15.35	10.49	22.94	38.29
TN15-626-20	94.09	15.98	10.01	24.30	40.28
TN15-626-23	94.02	21.61	9.74	26.78	48.40
TN15-626-25	95.56	26.28	8.83	32.51	58.79
TN15-626-26	94.33	28.10	8.9	30.85	58.96
TN15-626-27	94.72	27.43	8.8	35.51	62.95
TN15-626-28	93.26	27.72	8.98	30.84	58.57
TN15-626-29	95.70	28.87	8.77	33.45	62.33
TN15-626-30	95.22	24.84	8.84	31.50	56.35
TN15-626-31	92.30	18.45	10.04	20.58	39.04
TN15-626-32	91.68	21.24	9.75	26.20	47.44
TN15-626-33	90.80	21.71	9.25	26.52	48.24
TN15-626-34	95.01	28.53	8.75	33.21	61.74
TN15-626-35	94.18	20.58	9.63	22.88	43.47
TN15-626-36	93.52	25.37	9.15	28.80	54.18
TN16-565-01	91.76	22.18	9.71	25.41	47.60
TN16-565-03	91.67	21.69	9.8	26.24	47.94
TN16-565-04	92.35	22.52	9.79	25.21	47.73
TN16-565-05	93.28	16.51	9.98	28.39	44.91
TN16-565-06	92.05	10.05	10.65	25.09	35.14
TN16-565-07	92.31	11.45	10.32	27.51	38.97
TN16-565-18	93.73	21.71	9.36	29.60	51.32
TN16-565-19	94.68	21.69	9.39	27.61	49.31
TN16-565-20	93.69	18.48	9.56	29.60	48.08

Table 8.3: Wet chemistry results for the lamb faeces samples.

	As Fed Basis								
sample name	Hemicellulose %	Dry matter %	Ash %	Nitrogen %	GE %	Organic matter %	NDF %	ADF %	Lignin %
TN15-44-128	1.35	98.18	31.97	4.46	19.59	66.21	2.41	1.05	0.53
TN15-44-130	3.64	98.35	24.36	4.14	21.83	73.99	6.81	3.17	1.05
TN15-44-131	6.83	98.65	33.35	4.87	15.17	65.29	17.80	10.96	4.32
TN15-44-133	4.40	98.61	20.29	4.51	22.33	78.32	9.59	5.19	1.82
TN15-44-134	4.92	98.37	24.71	4.16	20.50	73.65	10.96	6.04	1.65
TN15-44-137	1.11	98.73	35.48	4.82	16.49	63.24	2.34	1.22	0.93
TN15-44-140	7.29	98.02	22.17	5.32	18.21	75.85	14.35	7.06	1.59
TN15-44-141	5.22	98.94	30.44	4.51	16.66	68.49	11.52	6.30	1.78
TN15-44-145	4.29	98.49	20.67	4.03	22.51	77.81	9.43	5.13	1.59
TN15-44-149	1.38	98.43	31.64	5.98	17.31	66.79	2.62	1.23	0.80
TN15-44-152	1.21	98.52	29.42	5.37	19.39	69.10	2.66	1.45	0.74
TN15-44-153	2.80	98.00	22.87	4.09	21.70	75.12	6.59	3.78	1.03
TN15-44-154	5.40	98.51	22.24	4.04	20.60	76.26	11.85	6.45	1.44
TN15-44-155	4.83	98.54	21.77	3.84	22.05	76.76	10.85	6.02	1.42
TN15-44-797a-9kg	5.19	96.54	20.33	7.93	19.52	76.21	6.76	1.57	0.76
TN15-44-797b-16kg	15.16	97.19	21.60	4.32	17.20	75.58	24.30	9.14	2.47
TN15-44-798a-9kg	21.16	97.04	11.11	4.28	19.03	85.92	37.48	16.31	3.78
TN15-44-798b-16kg	23.43	98.19	9.74	4.06	20.32	88.45	39.83	16.40	4.36
TN15-44-799a-9kg	15.33	96.84	13.40	4.23	20.07	83.44	28.20	12.86	3.03
TN15-44-799b-16kg	18.14	97.15	11.30	3.37	22.76	85.85	32.56	14.41	3.27
TN15-44-800a-9kg	7.04	96.77	19.14	5.39	20.71	77.63	12.83	5.78	1.43
TN15-44-800b-16kg	9.45	96.75	16.99	4.02	21.89	79.76	18.03	8.57	1.99
TN15-44-802a-9kg	12.83	97.38	17.15	5.25	19.60	80.22	22.11	9.28	3.03
TN15-44-802b-16kg	21.94	97.97	10.29	3.54	20.52	87.68	39.40	17.45	3.95

Table 8.3: Wet chemistry results for the lamb faeces samples (continued).

sample name	Hemicellulose %	Dry matter %	Ash %	Nitrogen %	GE %	Organic matter %	NDF %	ADF %	Lignin %
TN15-44-803a-9kg	7.79	97.51	19.17	4.49	21.83	78.33	12.43	4.64	1.51
TN15-44-803b-16kg	14.70	97.68	15.46	3.40	21.81	82.22	25.73	11.02	2.86
TN15-44-804a-9kg	11.13	97.08	16.09	4.48	19.66	80.98	19.59	8.46	2.33
TN15-44-804b-16kg	16.49	97.40	12.18	4.28	20.75	85.21	26.96	10.47	2.84
TN15-44-805a-16kg	7.75	97.09	22.44	7.04	18.64	74.65	11.26	3.51	0.99
TN15-44-805b-16kg	17.33	97.42	13.44	3.91	20.62	83.98	29.67	12.33	3.41
TN15-44-806a-9kg	20.45	97.41	11.99	4.32	18.56	85.41	36.54	16.08	3.39
TN15-44-806b-16kg	16.42	97.12	15.11	4.24	18.07	82.00	31.94	15.52	3.05
TN15-44-807a-9kg	14.93	97.72	15.37	4.47	19.36	82.34	28.18	13.24	3.46
TN15-44-807b-16kg	22.36	98.12	8.99	3.35	21.04	89.12	39.78	17.42	4.15
TN15-44-808a-16kg	14.11	96.96	20.58	4.17	21.39	76.38	19.58	5.47	1.51
TN15-44-808b-16kg	11.25	97.63	11.65	2.79	27.41	85.97	18.26	7.01	1.90
TN15-44-809a-9kg	12.76	97.46	19.89	5.03	17.87	77.57	25.31	12.54	3.23
TN15-44-809b-16kg	19.85	97.95	10.39	3.86	19.92	87.56	36.94	17.08	3.65
TN15-44-810a-9kg	14.16	96.80	17.71	4.69	18.44	79.09	24.35	10.18	2.33
TN15-44-810b-16kg	15.27	97.62	11.43	3.25	22.69	86.18	28.56	13.29	2.69
TN15-44-811a-9kg	13.22	97.35	16.63	5.72	19.06	80.72	25.15	11.93	2.87
TN15-44-811b-16kg	14.97	98.04	12.47	3.41	21.98	85.57	29.24	14.27	3.15
TN15-44-816a-9kg	5.65	97.17	24.34	6.75	19.01	72.83	8.09	2.43	0.54
TN15-44-816b-16kg	18.88	97.82	17.79	3.84	19.40	80.02	32.88	14.00	3.41
TN15-44-817a-9kg	16.01	97.53	12.83	3.92	19.76	84.70	30.64	14.63	3.29
TN15-44-817b-16kg	15.05	96.75	11.90	3.39	23.03	84.85	27.91	12.86	2.84
TN15-44-818a-9kg	12.12	97.52	18.02	4.52	20.22	79.49	20.38	8.25	2.02

Table 8.3: Wet chemistry results for the lamb faeces samples (continued).

sample name	Hemicellulose %	Dry matter %	Ash %	Nitrogen %	GE %	Organic matter %	NDF %	ADF %	Lignin %
TN15-44-818b-16kg	21.67	98.20	11.29	3.35	21.49	86.90	38.51	16.84	3.77
TN15-44-819a-9kg	16.77	97.97	15.97	4.69	17.81	81.99	31.23	14.46	3.07
TN15-44-819b-16kg	20.91	97.71	14.13	3.83	18.99	83.58	39.12	18.20	3.90
TN15-44-820a-9kg	6.95	97.43	20.13	5.34	21.12	77.30	10.36	3.41	1.44
TN15-44-820b-9kg	21.44	97.90	11.10	3.63	19.68	86.79	38.28	16.83	3.59
TN15-44-825a-9kg	12.26	96.95	17.25	5.00	19.21	79.69	20.04	7.78	1.91
TN15-44-825b-16kg	19.58	98.02	13.76	4.06	19.27	84.26	37.09	17.50	3.97

Table 8.4: Wet chemistry results for the Chicken Feed samples.

	As Fed Basis								
	DM%	Ash %	N %	Fat %	Starch %	CF %	Ca mg/g	P mg/g	GE kJ/g
AA	86.15	6.62	6.14	5.8	4.87	6.87	6.5	8.55	17.55
AB	85.27	6.6	6.76	5.17	5.44	3.97	5.52	7.48	17.42
AC	87.59	3.82	2.54	6.39	25.21	8.2	4.6	5.43	17.37
AD	86.29	3.76	3.93	13.19	25.37	3.51	3.74	4.91	18.71
AE	86.85	3.9	2.71	6.48	28.89	5.24	3.33	6.01	17.26
AF	85.82	3.76	2.48	6.11	31.86	6.54	3.33	6.01	16.86
AG	89.22	14.87	4.44	9.11	25.19	2.89	49.04	23.46	16.4
AH	85.63	4.61	3.84	5.87	25.3	6.57	5.16	7	17.15
AI	85.5	4.58	4.46	5.24	25.87	3.67	4.17	5.93	17.01
AJ	85.36	2.57	2.16	5.3	46.3	3.38	2.83	4.37	16.61
AK	87.21	3.83	2.58	5.98	25.91	8.13	4.74	5.22	17.23
AL	86.74	3.78	3.97	12.77	26.06	3.45	3.89	4.7	18.57
AM	87.06	3.91	2.75	6.06	29.59	5.17	3.48	5.8	17.12
AN	86.52	3.77	2.52	5.69	32.55	6.47	3.47	5.8	16.71
AO	88.55	14.88	4.48	8.69	25.88	2.83	49.19	23.25	16.25
AP	87.34	4.62	3.88	5.45	25.99	6.51	5.3	6.79	17
AQ	86.5	4.6	4.5	4.82	26.56	3.61	4.32	5.72	16.87
AR	86.28	2.58	2.2	4.88	46.99	3.31	2.97	4.17	16.47
AS	86.19	2.59	2.24	4.46	47.69	3.25	3.12	3.96	16.32
AT	87.67	3.69	2.39	6.39	28.98	8.04	4.76	5.01	17.23
AU	86.86	3.63	3.78	13.19	29.14	3.36	3.9	4.49	18.58
AV	87.35	3.77	2.57	6.47	32.66	5.09	3.49	5.59	17.13
AW	86.5	3.63	2.33	6.1	35.63	6.38	3.49	5.59	16.72
AX	90.1	14.74	4.29	9.1	28.96	2.74	49.2	23.04	16.26
AY	85.68	4.48	3.69	5.86	29.07	6.42	5.32	6.58	17.01
AZ	85.01	4.45	4.31	5.23	29.64	3.52	4.33	5.51	16.88
BA	84.78	2.44	2.01	5.3	50.07	3.22	2.99	3.96	16.48
BB	84.7	2.45	2.05	4.88	50.76	3.16	3.13	3.75	16.33
BC	84.62	2.31	1.87	5.29	53.84	3.07	3.15	3.54	16.34
BD	85.65	6.81	3.99	8.27	19.91	5.93	13.61	9.87	17.45
U	86.79	6.64	5.52	6.43	4.29	9.77	7.49	9.62	17.68
V	88.17	5.84	4.84	6.33	4.78	8.49	5.95	6.98	17.77
W	86.36	5.78	6.23	13.12	4.94	3.81	5.09	6.46	19.12
X	87.51	5.92	5.01	6.41	8.46	5.53	4.68	7.56	17.66
Y	85.61	5.77	4.78	6.04	11.43	6.83	4.68	7.57	17.26
Z	87.93	16.89	6.74	9.04	4.76	3.19	50.39	25.01	16.8
A	89.94	5.07	2.92	7.49	4.12	13.02	6.37	6.48	18.13
B	87.38	5.02	4.31	14.28	4.28	8.33	5.52	5.96	19.47
C	86.61	4.96	5.69	21.08	4.44	3.65	4.66	5.44	20.81
D	88.16	5.15	3.09	7.57	7.8	10.06	5.11	7.06	18.02
E	86.02	5.1	4.48	14.36	7.96	5.38	4.25	6.54	19.36

Table 8.4: Wet chemistry results for the Chicken Feed samples (continued).

ID	DM%	Ash %	N %	Fat %	Starch %	CF %	Ca mg/g	P mg/g	GE kJ/g
F	86.42	5.23	3.27	7.65	11.49	7.1	3.84	7.64	17.91
G	87.09	5.01	2.86	7.2	10.77	11.36	5.1	7.06	17.62
H	84.78	4.96	4.25	13.99	10.93	6.67	4.25	6.54	18.96
I	86.64	5.09	3.03	7.28	14.45	8.4	3.84	7.64	17.51
J	85.64	4.95	2.8	6.91	17.42	9.7	3.83	7.65	17.1
K	90.25	16.12	4.82	10.2	4.1	7.71	50.82	24.51	17.16
L	88.8	16.07	6.21	16.99	4.25	3.03	49.96	23.99	18.5
M	89.18	16.2	4.99	10.28	7.78	4.76	49.55	25.09	17.05
N	88.71	16.06	4.76	9.91	10.74	6.05	49.55	25.09	16.64
O	90.58	27.17	6.72	12.91	4.07	2.41	95.26	42.54	16.18
P	87.62	5.86	4.22	6.96	4.21	11.39	6.93	8.05	17.9
Q	86.51	5.8	5.61	13.75	4.36	6.71	6.07	7.53	19.25
R	87.15	5.94	4.39	7.04	7.89	8.44	5.66	8.63	17.8
S	86.65	5.8	4.16	6.67	10.85	9.73	5.66	8.64	17.39
T	90.01	16.91	6.12	9.67	4.18	6.09	51.37	26.08	16.93

Table 8.5: Nickel content of hyperaccumulator plant samples.

sample name	Ni %	sample name	Ni %
song 2 -1	1.52	BC-2-2-1	1.62
am1-1	1.60	BC-2-2a-1	1.62
song 2- 2	1.92	BC-2-3-1	1.13
am1 -2	1.74	BC-2-3a-1	1.13
song 3 -2	0.67	BC-2-4-1	1.42
am1 - 3	1.43	BC-2-4a-1	1.42
song 2 - 3	1.85	BC-2-5-1	1.28
song 3- 3	0.74	BC-2-5a-1	1.28
am1- 4	1.76	BC-2-6-1	1.44
song 3 - 4	0.80	BC-2-6a-1	1.44
song 2 -5	1.57	BC-2-7-1	1.40
song 3 - 5	0.61	BC-2-7a-1	1.40
song 2- 6	1.60	BC-2-8-1	1.04
am1 -7	2.09	BC-2-8a-1	1.04
song 2 -7	1.10	BC-2-9-1	1.84
song 2 - 8	1.31	BC-2-9a-1	1.84
am1-8	1.29	BC-3-10-1	1.60
am1 - 9	1.84	BC-3-10a-1	1.60
am1- 10	1.93		
am1 -5	1.49		
song 3 -1	0.57		
BC-2-1-1	1.50		
BC-2-1a-1	1.50		

Chapter 8.0: References

[1] Aroca-Santos, R., Cancilla, J. C., Perez-Perez, A., Moral, A., & Torrecilla, J. S. (**2016**). Quantifying binary and ternary mixtures of monovarietal extra virgin olive oils with UV-vis absorption and chemometrics. *Sensors and Actuators B-Chemical, 234*, 115-121.

[2] Huang, Y. P., Wu, Z. W., Su, R. H., Ruan, G. H., Du, F. Y., & Li, G. K. (**2016**). Current application of chemometrics in traditional Chinese herbal medicine research. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 1026*, 27-35.

[3] Monakhova, Y. B., Mushtakova, S. P., Kolesnikova, S. S., & Astakhov, S. A. (**2010**). Chemometrics-assisted spectrophotometric method for simultaneous determination of vitamins in complex mixtures. *Analytical and Bioanalytical Chemistry, 397*(3), 1297-1306.

[4] Holland, J. K., Kemsley, E. K., & Wilson, R. H. (**1998**). Use of Fourier transform infrared spectroscopy and partial least squares regression for the detection of adulteration of strawberry purees. *Journal of the Science of Food and Agriculture, 76*(2), 263-269.

[5] Hu, T., Jin, W. Y., & Cheng, C. G. (**2011**). Classification of five kinds of moss plants with the use of Fourier transform infrared spectroscopy and chemometrics. *Spectroscopy-Biomedical Applications, 25*(6), 271-285.

[6] Kumar, R., Kumar, V., & Sharma, V. (**2017**). Fourier transform infrared spectroscopy and chemometrics for the characterization and discrimination of writing/photocopier paper types: Application in forensic document examinations. *Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy, 170*, 19-28.

[7] Rohman, A., Wibowo, D., Sudjadi, Lukitaningsih, E., & Rosman, A. S. (**2015**). Use of Fourier Transform Infrared Spectroscopy in Combination with Partial Least Square for Authentication of Black Seed Oil. *International Journal of Food Properties, 18*(4), 775-784.

[8] Theophilou, G., Lima, K. M. G., Martin-Hirsch, P. L., Stringfellow, H. F., & Martin, F. L. (**2016**). ATR-FTIR spectroscopy coupled with chemometric analysis discriminates normal, borderline and malignant ovarian tissue: classifying subtypes of human cancer. *Analyst, 141*(2), 585-594.

[9] Travo, A., Paya, C., Deleris, G., Colin, J., Mortemousque, B., & Forfar, I. (**2014**). Potential of FTIR spectroscopy for analysis of tears for diagnosis purposes. *Analytical and Bioanalytical Chemistry, 406*(9-10), 2367-2376.

[10] Ulewicz-Magulska, B., & Wesolowski, M. (**2013**). A Chemometric Approach to Distribution of Selenium in Medicinal Plants Cultivated in Poland. *Journal of Medicinal Food, 16*(5), 460-466.

[11] Rohman, A., Nugroho, A., Lukitaningsih, E., & Sudjadi. (**2014**). Application of Vibrational Spectroscopy in Combination with Chemometrics Techniques for Authentication of Herbal Medicine. *Applied Spectroscopy Reviews, 49*(8), 603-613.

[12] Rohman, A., & Man, Y. B. C. (**2012**). Application of Fourier Transform Infrared Spectroscopy for Authentication of Functional Food Oils. *Applied Spectroscopy Reviews, 47*(1), 1-13.

[13] Mostaco-Guidolin, L. B., & Bachmann, L. (**2011**). Application of FTIR Spectroscopy for Identification of Blood and Leukemia Biomarkers: A Review over the Past 15 Years. *Applied Spectroscopy Reviews, 46*(5), 388-404.

[14] Barbara, Stuart H. Infrared Spectroscopy: Fundamentals and Applications. John Wiley & Sons, Ltd. 5 JUL **2005**, accessed 2017.

[15] Aroca, R. (**2006**) Theory of Molecular Vibrations. The Origin of Infrared and Raman Spectra, in Surface-Enhanced Vibrational Spectroscopy, John Wiley & Sons, Ltd, Chichester, UK.

[16] chemists.princeton.edu/bernasek/atr-ftir - Accessed 16/10/17

[16] Fale, P. L. V., & Chan, K. L. A. (**2017**). Preventing damage of germanium optical material in attenuated total reflection-Fourier transform infrared (ATR-FTIR) studies of living cells. *Vibrational Spectroscopy, 91*, 59-67.

[17] Sharma, V., & Kumar, R. (**2017**). Fourier transform infrared spectroscopy and high performance thin layer chromatography for characterization and multivariate discrimination of blue ballpoint pen ink for forensic applications. *Vibrational Spectroscopy, 92*, 96-104.

[18] Sharm, S., & Uttam, K. N. (**2017**). Rapid analyses of stress of copper oxide nanoparticles on wheat plants at an early stage by laser induced fluorescence and attenuated total reflectance Fourier transform infrared spectroscopy. *Vibrational Spectroscopy, 92*, 135-150.

[19] Haputhanthri, R., Ojha, R., Izgorodina, E. I., Guo, S. X., Deacon, G. B., McNaughton, D., & Wood, B. R. (**2017**). A spectroscopic investigation into the binding of novel platinum(IV) and platium(II) anticancer drugs with DNA. *Vibrational Spectroscopy, 92*, 82-95.

[20] Ramer, G. and Lendl, B. **2013**. Attenuated Total Reflection Fourier Transform Infrared Spectroscopy. Encyclopedia of Analytical Chemistry Published Online: 15 MAR 2013 (Wiley Online Library)

[21] R. Anthony, H.H. Mantsch, Infrared spectroscopy in clinical and diagnostic analysis, in: R.A. Meyers (Ed.), Encyclopedia of Analytical Chemistry, John Wiley& Sons Ltd, Chichester, **2011**, pp. 1–19.

[22] Trevisan, J., Angelov, P. P., Carmichael, P. L., Scott, A. D., & Martin, F. L. (**2012**). Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. *Analyst, 137*(14), 3202-3215.

[23] Gautam, R., Vanga, S., Ariese, F., & Umapathy, S. (**2015**). Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *Epj Techniques and Instrumentation, 2*.

[24] Ozgenc, O., Durmaz, S., Boyaci, I. H., & Eksi-Kocak, H. (**2017**). Determination of chemical changes in heat-treated wood using ATR-FTIR and FT Raman spectrometry. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy, 171*, 395-400.

[25] Peets, P., Leito, I., Pelt, J., & Vahur, S. (**2017**). Identification and classification of textile fibres using ATR-FT-IR spectroscopy with chemometric methods. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy, 173*, 175-181.

[26] Parhizkar, E., Ghazali, M., Ahmadi, F., & Sakhteman, A. (**2017**). PLS-LS-SVM based modeling of ATR-IR as a robust method in detection and qualification of alprazolam. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy, 173*, 87-92.

[27] Lima, C. A., Goulart, V. P., Correa, L., & Zezell, D. M. (**2016**). Using Fourier transform infrared spectroscopy to evaluate biological effects induced by photodynamic therapy. *Lasers in Surgery and Medicine, 48*(5), 538-545.

[28] B.H. Stuart, Infrared Spectroscopy, Fundamentals and Applications (Analytical Techniques in the Sciences), John Wiley & Sons, Ltd, Chichester, UK, **2004**.

[29] M.R. Derrick, D. Stulik, J.M. Landry, Infrared Spectroscopy in Conversation Science, Scientific Tools for Conservation, The Getty Conservation Institute, LosAngeles, **1999**.

[30] B.C. Smith, Fundamentals of Fourier Transform Infrared Spectroscopy, 2nd ed.,CRC Press, Boca Raton, **2011**.

[31] J. Fahrenfort, Attenuated total reflection: a new principle for the prediction of useful infrared reflection spectra of organic compounds, Spectrochim. Acta 17(**1961**) 698–709.

[32] J.M. Chalmers, H.G.M. Edwards, M.D. Hargreaves, Vibrational Spectroscopy sampling techniques (Eds.), in: J.M. Chalmers, H.G.M. Edwards, M.D. Hargreaves (Eds.), Infrared and Raman spectroscopy in Forensic Science, John Wiley & Sons, Chichester, **2012**, pp. 45–82.

[33] F.M. Mirabella, Internal Reflection Spectroscopy: Theory and Applications, Marcel Dekker, New York, **1993**.

[34] U.P. Fringeli, ATR and reflectance IR spectroscopy, applications (Ed.), in: J.C. Lindon, G.E. Tranter, D.W. Koppenaal (Eds.), Encyclopedia of Spectroscopy and Spectrometry, 3rd ed., Elsevier, The Netherlands, **2016**, pp. 115–129.

[35] Schädle, T. and Mizaikoff, B, (**2016**) Mid-Infrared Waveguides: A Perspective, Appl. Spectrosc.70, 1625-1638

[36] Szady-Chelmieniecka, A., Kolodziej, B., Morawiak, M., Kamienski, B., & Schilf, W. (**2018**). Spectroscopic studies of the intramolecular hydrogen bonding in o-hydroxy Schiff bases, derived from diaminomaleonitrile, and their deprotonation reaction products. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy, 189*, 330-341

[37] Upadhyay, N., Jaiswal, P., & Jha, S. N. (**2018**). Application of attenuated total reflectance Fourier Transform Infrared spectroscopy (ATR-FTIR) in MIR range coupled with chemometrics for detection of pig body fat in pure ghee (heat clarified milk fat). *Journal of Molecular Structure, 1153*, 275-281.

[38] Materazzi, S., Gregori, A., Ripani, L., Apriceno, A., & Risoluti, R. (**2017**). Cocaine profiling: Implementation of a predictive model by ATR-FTIR coupled with chemometrics in forensic chemistry. *Talanta, 166*, 328-335.

[39] Brooks, R. R., Chambers, M. F., Nicks, L. J., & Robinson, B. H. (**1998**). Phytomining. *Trends in Plant Science, 3*(9), 359-362.

[40] Macek, T., Mackova, M., Kucerova, P., Chroma, L., Burkhard, J., & Demnerova, K. (**2002**). Phytoremediation. *Biotechnology for the Environment: Soil Remediation, Vol 3b, 3B*, 115-137.

[41] Hadad, H. R., Maine, M. A., & Bonetto, C. A. (**2006**). Macrophyte growth in a pilot-scale constructed wetland for industrial wastewater treatment. *Chemosphere, 63*(10), 1744-1753.

[42] Anderson, C. W. N., Brooks, R. R., Chiarucci, A., LaCoste, C. J., Leblanc, M., Robinson, B. H., . . . Stewart, R. B. (**1999**). Phytomining for nickel, thallium and gold. *Journal of Geochemical Exploration, 67*(1-3), 407-415.

[43] Robinson, B. H., Brooks, R. R., & Clothier, B. E. (**1999**). Soil amendments affecting nickel and cobalt uptake by *Berkheya coddii*: Potential use for phytomining and phytoremediation. *Annals of Botany, 84*(6), 689-694.

[44] Robinson, B. H., Lombi, E., Zhao, F. J., & McGrath, S. P. (**2003**). Uptake and distribution of nickel and other metals in the hyperaccumulator Berkheya coddii. *New Phytologist, 158*(2), 279-285.

[45] Chaney, R. L., Malik, M., Li, Y. M., Brown, S. L., Brewer, E. P., Angle, J. S., & Baker, A. J. M. (**1997**). Phytoremediation of soil metals. *Current Opinion in Biotechnology, 8*(3), 279-284.

[46] Milner, M. J., & Kochian, L. V. (**2008**). Investigating heavy-metal hyperaccumulation using *Thlaspi caerulescens* as a model system. *Annals of Botany, 102*(1), 3-13.

[47] Koptsik, G. N. (**2014**). Problems and prospects concerning the phytoremediation of heavy metal polluted soils: A review. *Eurasian Soil Science, 47*(9), 923-939.

[48] Abou Auda, M. M., Symeonidis, L., Hatzistavrou, E., & Yupsanis, T. (**2002**). Nucleolytic activities and appearance of a new DNAse in relation to nickel and manganese accumulation in *Alyssum murale*. *Journal of Plant Physiology*, *159*(10), 1087-1095.

[49] Fujimori, K., & Ohta, D. (**2003**). Heavy metal induction of *Arabidopsis* serine decarboxylase gene expression. *Bioscience Biotechnology and Biochemistry, 67*(4), 896-898.

[50] Schor-Fumbarov, T., Goldsbrough, P. B., Adam, Z., & Tel-Or, E. (**2005**). Characterization and expression of a metallothionein gene in the aquatic fern *Azolla filiculoides* under heavy metal stress. *Planta, 223*(1), 69-76.

[51] Bellion, M., Courbot, M., Jacob, C., Guinet, F., Blaudez, D., & Chalot, M. (**2007**). Metal induction of a *Paxillus involutus* metallothionein and its heterologous expression in *Hebeloma cylindrosporum*. *New Phytologist, 174*(1), 151-158.

[52] Lebo, Stuart E. Jr.; Gargulak, Jerry D.; McNally, Timothy J. (**2001**). "Lignin". Kirk-Othmer Encyclopedia of Chemical Technology. Kirk Othmer Encyclopedia of Chemical Technology. John Wiley & Sons,

[53] Lee, L. C., Liong, C. Y., & Jemain, A. A. (**2017**). A contemporary review on Data Preprocessing (DP) practice strategy in ATR-FTIR spectrum. *Chemometrics and Intelligent Laboratory Systems, 163*, 64-75.

[54] Geladi, P, Dabakk E, An Overview of Chemometrics Application In Near Infrared Spectrometry, *J. Infrared Spec*. 3 (**1995**) 119–132.

[55] Geladi, P, Chemometrics in spectroscopy. Part 1. Classical chemometrics, *Spectrochim. Acta. B* 58 (**2003**) 767–782.

[56] A.S. Rinna, L. Norgaard, F. van den Berg, J. Thygesen, R. Bro, S.B. Engelsen, Data pre-processing (Ed.), in: D.-W. Sun (Ed.), Infrared spectroscopy for Food Quality Analysis and Control, Academic Press, Burlington, **2009**, pp. 29–48.

[57] H. Martens, S.A. Jensen, P. Geladi, Multivariate linearity transformations for near infrared reflectance spectroscopy, in: O.H.J. Christie (Editor), Proc. Nordic Symp. Applied Statistics, Stokkland Forlag, Stavanger, Norway, **1983**, pp. 205–234.

[58] Iversen, A. J., & Palm, T. (**1985**). Multiplicative Scatter Correction of Visible Reflectance Spectra In Color Determination Of Meat Surfaces. *Applied Spectroscopy, 39*(4), 641-646

[59] Pyzowski, J., Lenartowicz, M., Sobanska, A. W., & Brzezinska, E. (**2017**). Fast and Convenient NIR Spectroscopy Procedure for Determination of Metformin Hydrochloride in Tablets. *Journal of Applied Spectroscopy, 84*(4), 710-715.

[60] Sila, A. M., Shepherd, K. D., & Pokhariyal, G. P. (**2016**). Evaluating the utility of mid-infrared spectral subspaces for predicting soil properties. *Chemometrics and Intelligent Laboratory Systems, 153*, 92-105.

[61] Luna, A. S., da Silva, A. P., Ferre, J., & Boque, R. (**2013**). Classification of edible oils and modeling of their physico-chemical properties by chemometric methods using mid-IR spectroscopy. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy, 100*, 109-114

[62] Shen, Y. X., Zhao, Y. L., Zhang, J., Jin, H., & Wang, Y. Z. (**2016**). Study on the Discrimination of FTIR Spectroscopy of Gentiana Rigescens with Different Harvest Time. *Spectroscopy and Spectral Analysis, 36*(5), 1358-1362.

[63] Jolayemi, O. S., Tokatli, F., Buratti, S., & Alamprese, C. (**2017**). Discriminative capacities of infrared spectroscopy and e-nose on Turkish olive oils. *European Food Research and Technology, 243*(11), 2035-2042.

[64] Wang, C. K., Zhang, T. L., & Pan, X. Z. (**2017**). Potential of visible and near-infrared reflectance spectroscopy for the determination of rare earth elements in soil. *Geoderma, 306*, 120-126.

[65] Savitzky, A., Golay, M. J. E. (**1964**). Smoothing + Differentiation of Data By Simplified Least Squares Procedures. *Analytical Chemistry, 36*(8), 1627

[66] Lara, M. A., Diezma, B., Lleo, L., Roger, J. M., Garrido, Y., Gil, M. I., & Ruiz-Altisent, M. (**2014**). Hyperspectral images for the detection of salinity effect on lettuce leaves. *Vii Congreso Iberico De Agroingenieria Y Ciencias Horticolas: Innovar Y Producir Para El Futuro. Innovating and Producing for the Future*, 1644-1649.

[67] Luo, X. F., Yu, X., Wu, X. M., Cheng, Y. Y., & Qu, H. B. (**2008**). Rapid determination of Paeoniae Radix using near infrared spectroscopy. *Microchemical Journal, 90*(1), 8-12.

[68] Whittaker, E. (**1922**). On a New Method of Graduation. Proceedings of the Edinburgh Mathematical Society, 41, 63-75.

[69] Eilers P. H. and Boelens H. F., "Baseline correction with asymmetric least squares smoothing," Leiden University Medical Centre Report (**2005**).

[70] Jentzsch, P. V., Ciobota, V., Salinas, W., Kampe, B., Aponte, P. M., Rosch, P., . . . Ramos, L. A. (**2016**). Distinction of Ecuadorian varieties of fermented cocoa beans using Raman spectroscopy. *Food Chemistry, 211*, 274-280.

[71] Kim, H. M., Park, H. S., Cho, Y., Jin, S. M., Lee, K. T., Jung, Y. M., & Suh, Y. D. (**2014**). Noninvasive deep Raman detection with 2D correlation analysis. *Journal of Molecular Structure, 1069*, 223-228.

[72] Mabbott, S., Correa, E., Cowcher, D. P., Allwood, J. W., & Goodacre, R. (**2013**). Optimization of Parameters for the Quantitative Surface-Enhanced Raman Scattering Detection of Mephedrone Using a Fractional Factorial Design and a Portable Raman Spectrometer. *Analytical Chemistry, 85*(2), 923-931.

[73] Vardaki, M. Z., Matousek, P., & Stone, N. (**2016**). Characterisation of signal enhancements achieved when utilizing a photon diode in deep Raman spectroscopy of tissue. *Biomedical Optics Express, 7*(6), 2130-2141.

[74] Mallat, S. G. (**1989**). A Theory for Multiresolution Signal Decomposition - The Wavelet Representation. *Ieee Transactions on Pattern Analysis and Machine Intelligence, 11*(7), 674-693.

[75] Daubechies, I. (**1990**). The Wavelet Transform, Time-Frequency Localization and Signal Analysis. *Ieee Transactions on Information Theory, 36*(5), 961-1005.

[76] Daubechies, I (**1992**) Ten Lectures on Wavelets. CBMS-NSF Regional Conference Series in Applied Mathematics, Philadelphia

[77] Jiang, H., Liu, G. H., Mei, C. L., & Chen, Q. S. (**2013**). Qualitative and quantitative analysis in solid-state fermentation of protein feed by FT-NIR spectroscopy integrated with multivariate data analysis. *Analytical Methods, 5*(7), 1872-1880.

[78] Cai, C. S., & Harrington, P. D. (**1998**). Different discrete wavelet transforms applied to denoising analytical data. *Journal of Chemical Information and Computer Sciences, 38*(6), 1161-1170.

[79] Walczak, B., & Massart, D. L. (**1997**). Noise suppression and signal compression using the wavelet packet transform. *Chemometrics and Intelligent Laboratory Systems, 36*(2), 81-94.

[80] Ordinary least-squares regression. (**1999**). In Hutcheson, G. D. The multivariate social scientist (pp. 56-113). : SAGE Publications Ltd

[81] Alin, A. (2010), Multicollinearity. WIREs Comp Stat, 2: 370–374.

[82] Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. *Chemometr. Intell. Lab Syst.* 2, 37 – 52

[83] Hoerl, A. E., & Kennard, R. W. (**1970**). Ridge Regression - Applications to Nonorthogonal Problems. Technometrics, 12(1), 69.

[84] Hoerl, A. E., & Kennard, R. W. (**2000**). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 42(1), 80-86.

[85] Tibshirani, R. (**1996**). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B-Methodological, 58*(1), 267-288.

[86] Ferrand-Calmels, M., Palhiere, I., Brochard, M., Leray, O., Astruc, J. M., Aurel, M. R., . . . Larroque, H. (**2014**). Prediction of fatty acid profiles in cow, ewe, and goat milk by mid-infrared spectrometry. *Journal of Dairy Science, 97*(1), 17-35.

[87] Trueman, B. F., MacIsaac, S. A., Stoddart, A. K., & Gagnon, G. A. (**2016**). Prediction of disinfection by-product formation in drinking water via fluorescence spectroscopy. *Environmental Science-Water Research & Technology, 2*(2), 383-389.

[88] Segal, M. R., Dahlquist, K. D., & Conklin, B. R. (**2003**). Regression approaches for microarray data analysis. *Journal of Computational Biology, 10*(6), 961-980.

[89] Gauthier, P. A., Scullion, W., & Berry, A. (**2017**). Sound quality prediction based on systematic metric selection and shrinkage: Comparison of stepwise, lasso, and elastic-net algorithms and clustering preprocessing. *Journal of Sound and Vibration, 400*, 134-153.

[90] Zou, H., & Hastie, T. (**2005**). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B-Statistical Methodology, 67, 301-320.

[91] Hong, D., & Zhang, F. (**2010**). Weighted Elastic Net Model for Mass Spectrometry Imaging Processing. *Mathematical Modelling of Natural Phenomena, 5*(3), 115-133.

[92] Zemmour, C., Bertucci, F., Finetti, P., Chetrit, B., Birnbaum, D., Filleron, T., & Boher, J. M. (**2015**). Prediction of Early Breast Cancer Metastasis from DNA Microarray Data Using High-Dimensional Cox Regression Models. *Cancer Informatics, 14*, 129-138.

[93] Lara, M. A., Diezma, B., Lleo, L., Roger, J. M., Garrido, Y., Gil, M. I., & Ruiz-Altisent, M. (**2014**). Hyperspectral images for the detection of salinity effect on lettuce leaves. *Vii Congreso Iberico De Agroingenieria Y Ciencias Horticolas: Innovar Y Producir Para El Futuro. Innovating and Producing for the Future*, 1644-1649.

[94] Das, J., Gayvert, K. M., Bunea, F., Wegkamp, M. H., & Yu, H. Y. (**2015**). ENCAPP: elastic-net-based prognosis prediction and biomarker discovery for human cancers. *Bmc Genomics, 16*, 13.

[95] Algamal, Z. Y., & Lee, M. H. (**2015**). Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification. *Computers in Biology and Medicine, 67*, 136-145.

[96] Piaskowski, J. L., Brown, D., & Campbell, K. G. (**2016**). Near-Infrared Calibration of Soluble Stem Carbohydrates for Predicting Drought Tolerance in Spring Wheat. *Agronomy Journal, 108*(1), 285-293.

[97] Valitalo, P. A. J., Griffioen, K., Rizk, M. L., Visser, S. A. G., Danhof, M., Rao, G., . . . van Hasselt, J. G. C. (**2016**). Structure-Based Prediction of Anti-infective Drug Concentrations in the Human Lung Epithelial Lining Fluid. *Pharmaceutical Research, 33*(4), 856-867.

[98] Sharif, B., Makowski, D., Plauborg, F., & Olesen, J. E. (**2017**). Comparison of regression techniques to predict response of oilseed rape yield to variation in climatic conditions in Denmark. *European Journal of Agronomy, 82*, 11-20.

[99] Pinheiro, C. T., Rendall, R., Quina, M. J., Reis, M. S., & Gando-Ferreira, L. M. (**2017**). Assessment and Prediction of Lubricant Oil Properties Using Infrared Spectroscopy and Advanced Predictive Analytics. *Energy & Fuels, 31*(1), 179-187.

[100] Devangad, P., Unnikrishnan, V. K., Tamboli, M. M., Shameem, K. M. M., Nayak, R., Choudhari, K. S., & Santhosh, C. (**2016**). Quantification of Mn in glass matrices using laser induced breakdown spectroscopy (LIBS) combined with chemometric approaches. *Analytical Methods, 8*(39), 7177-7184.

[101] Yu, H. W., Liu, H. Z., Wang, N., Yang, Y., Shi, A. M., Liu, L., . . . Wang, Q. (**2016**). Rapid and visual measurement of fat content in peanuts by using the hyperspectral imaging technique with chemometrics. *Analytical Methods, 8*(41), 7482-7492.

[102] Altieri, G., Genovese, F., Tauriello, A., & Di Renzo, G. C. (**2017**). Models to improve the non-destructive analysis of persimmon fruit properties by VIS/NIR spectrometry. *Journal of the Science of Food and Agriculture, 97*(15), 5302-5310.

[103] Mahesh, S., Jayas, D. S., Paliwal, J., & White, N. D. G. (**2015**). Comparison of Partial Least Squares Regression (PLSR) and Principal Components Regression (PCR) Methods for Protein and Hardness Predictions using the Near-Infrared (NIR) Hyperspectral Images of Bulk Samples of Canadian Wheat. *Food and Bioprocess Technology, 8*(1), 31-40.

[104] Wold, S., Sjostrom, M., & Eriksson, L. (**2001**). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems, 58*(2), 109-130.

[105] Lindgren, F., Geladi, P., & Wold, S. (**1993**). The Kernel Algorithm for Pls. *Journal of Chemometrics, 7*(1), 45-59.

[106] Wold, H. Soft modelling, The basic design and some extensions,in: K.-G. Joreskog, H. Wold Eds. , Systems Under Indirect Observation, vols. I and II, North-Holland, Amsterdam,1982

[107] Yin, M., Tang, S. F., & Tong, M. M. (**2016**). Identification of edible oils using terahertz spectroscopy combined with genetic algorithm and partial least squares discriminant analysis. *Analytical Methods, 8*(13), 2794-2798.

[108] Li, B. Y., Casamayou-Boucau, Y., Calvet, A., & Ryder, A. G. (**2017**). Chemometric approaches to low-content quantification (LCQ) in solid-state mixtures using Raman mapping spectroscopy. *Analytical Methods, 9*(44), 6293-6301.

[109] Cheng, J. H., Jin, H. L., Xu, Z. Y., & Zheng, F. P. (**2017**). NIR hyperspectral imaging with multivariate analysis for measurement of oil and protein contents in peanut varieties. *Analytical Methods, 9*(43), 6148-6154.

[110] Terra, L. A., Filgueiras, P. R., Alves, J. C. L., & Poppi, R. J. (**2017**). Quantification of the contents in biojet fuel blends using near infrared spectroscopy and multivariate calibration. *Analytical Methods, 9*(31), 4616-4621.

[111] Lu, Y. Z., Du, C. W., Yu, C. B., & Zhou, J. M. (**2015**). Determination of Nitrogen in Rapeseed by Fourier Transform Infrared Photoacoustic Spectroscopy and Independent Component Analysis. *Analytical Letters, 48*(7), 1150-1162.

[112] De Morais, C. D. M., & de Lima, K. M. G. (**2015**). Determination and analytical validation of creatinine content in serum using image analysis by multivariate transfer calibration procedures. *Analytical Methods, 7*(16), 6904-6910.

[113] Lin, S. S., & Zhang, X. M. (**2016**). A rapid and novel method for predicting nicotine alkaloids in tobacco through electronic nose and partial least-squares regression analysis. *Analytical Methods, 8*(7), 1609-1617

[114] Wu, Z. Z., Long, J., Xu, E., Wu, C. S., Wang, F., Xu, X. M., . . . Jiao, A. Q. (2015). Application of FT-NIR spectroscopy and FT-IR spectroscopy to Chinese rice wine for rapid determination of fermentation process parameters. *Analytical Methods, 7*(6), 2726-2737.

[115] Bezerra, R. M. R., Neves, A. C. O., Pimenta, A. S., & Lim, K. M. G. (2015**).** Estimation of Brazilian charcoal properties using attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectrometry coupled with multivariate analysis. *Analytical Methods, 7*(13), 5695-5701.

[116] Wubshet, S. G., Mage, I., Bocker, U., Lindberg, D., Knutsen, S. H., Rieder, A., . . . Afseth, N. K. (**2017**). FTIR as a rapid tool for monitoring molecular weight distribution during enzymatic protein hydrolysis of food processing by-products. *Analytical Methods, 9*(29), 4247-4254.

[117] Ellis, D. I., Ellis, J., Muhamadali, H., Xu, Y., Horn, A. B., & Goodacre, R. (**2016**). Rapid, high-throughput, and quantitative determination of orange juice adulteration by Fourier-transform infrared spectroscopy. *Analytical Methods, 8*(28), 5581-5586

[118] Nurrulhidayah, A. F., Man, Y. B. C., Amin, I., Salleh, R. A., Farawahidah, M. Y., Shuhaimi, M., & Khatib, A. (2015**).** FTIR-ATR Spectroscopy Based Metabolite Fingerprinting as A Direct Determination of Butter Adulterated with Lard. *International Journal of Food Properties, 18*(2), 372-379.

[119] Da Silva, D. J., & Wiebeck, H. (**2017**). Using PLS, iPLS and siPLS linear regressions to determine the composition of LDPE/HDPE blends: A comparison between confocal Raman and ATR-FTIR spectroscopies. *Vibrational Spectroscopy, 92*, 259-266.

[120] Zhang, H., Ma, J. K., Miao, Y. L., Tuchiya, T., & Chen, J. Y. (**2015**). Analysis of Carbonyl Value of Frying Oil by Fourier Transform Infrared Spectroscopy. *Journal of Oleo Science, 64*(4), 375-380

[121] Fernandez, B., Andres, S., Prieto, N., Mantecon, A. R., & Giraldez, F. J. (**2008**). Prediction of chemical composition of sugar beet pulp by near infrared reflectance spectroscopy. *Journal of near Infrared Spectroscopy, 16*(2), 105-110

[122] Restaino, E. A., Fernandez, E. G., La Manna, A., & Cozzolino, D. (**2009**). Prediction of the nutritive value of pasture silage by near infrared spectroscopy (nirs). *Chilean Journal of Agricultural Research, 69*(4), 560-566.

[123] Ye, H, Xia, A, Zhang, X, Chen Z, Zhou, X. (**2012)** 'Development and Application of Portable Quality Analyzer Based on Long Wave Near Infrared Spectroscopy'. *Chin. J. Sci. Ins.*. 33(1): 85-90.

[124] Ji, H. Y., Wen, M., & Ieee. (**2006**). Design of portable LED-based NIR integrity wheat component intelligent measuring apparatus. *Wcica 2006: Sixth World Congress on Intelligent Control and Automation, Vols 1-12, Conference Proceedings*, 4944

[125] Correia, R. M., Tosato, F., Domingos, E., Rodrigues, R. R. T., Aquino, L. F. M., Filgueiras, P.,Romao, W. (**2018**). Portable near infrared spectroscopy applied to quality control of Brazilian coffee. *Talanta, 176*, 59-68

[126] Ji, H. Y., Wen, M., & Hao, B. (**2006**). Building artificial neural networks model on portable NIR integrity wheat component measuring apparatus. *Spectroscopy and Spectral Analysis, 26*(1), 57-59.

[127] Wen, M., & Ji, H. Y. (**2004**). Development of portable LED-based NIR integrity wheat component measuring apparatus. *Spectroscopy and Spectral Analysis, 24*(10), 1276-1279.

[128] Norris, K. H., Barnes, R. F., Moore, J. E., & Shenk, J. S. (**1976**). Predicting forage quality by infrared reflectance spectroscopy. *Journal of Animal Science, 43*(4), 889-897.

[129] Berzaghi, P., Serva, L., Piombino, M., Mirisola, M., & Benozzo, F. (**2005**). Prediction performances of portable near infrared instruments for at farm forage analysis. *Italian Journal of Animal Science, 4*, 145-147.

[130] Compact Sensor unit, Corona 45 NIR 1.7, Specification 10/2001. Spectral Sensors by Carl Zeiss. Accessed 15/10/17. https://tp.dresden-concept.de/userdata/6ba91cd980e5fd859e4d8551cc60cad3.pdf

[131] Monrroy, M., Gutierrez, D., Miranda, M., Hernandez, K., & Garcia, J. R. (**2017**). Determination of Brachiaria spp. Forage quality by near-infrared spectroscopy and partial least squares regression. *Journal of the Chilean Chemical Society, 62*(2), 3472-3477.

[132] Dos Santos, C. A. T., Lopo, M., Pascoa, R., & Lopes, J. A. (**2013**). A Review on the Applications of Portable Near-Infrared Spectrometers in the Agro-Food Industry. *Applied Spectroscopy, 67*(11), 1215-1233.

[133] Pierna, J. A. F., Vermeulen, P., Lecler, B., Baeten, V., & Dardenne, P. (**2010**). Calibration Transfer from Dispersive Instruments to Handheld Spectrometers. *Applied Spectroscopy, 64*(6), 644-648.

[134] Santos, P. M., Pereira, E. R., & Rodriguez-Saona, L. E. (**2013**). Application of Hand-Held and Portable Infrared Spectrometers in Bovine Milk Analysis. *Journal of Agricultural and Food Chemistry, 61*(6), 1205-1211. doi:10.1021/jf303814g

[135] Vohland, M., Ludwig, M., Thiele-Bruhn, S., & Ludwig, B. (**2014**). Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. Geoderma, 223, 88-96.

[136] Soriano-Disla, J. M., Janik, L. J., Allen, D. J., & McLaughlin, M. J. (**2017**). Evaluation of the performance of portable visible-infrared instruments for the prediction of soil properties. *Biosystems Engineering, 161*, 24-36.

[137] Garside, P., & Wyeth, P. (**2003**). Identification of cellulosic fibres by FTIR spectroscopy - Thread and single fibre analysis by attenuated total reflectance. *Studies in Conservation, 48*(4), 269-275.

[138] Association of Official Analytical Chemists (AOAC), (**1990)**. Official Methods of Analysis, 15th ed. AOAC International, Arlington, VA

[139] Robertson, J. B., Van Soest, P. J., (**1981)**. The Detergent System of Analysis And Its Application To Human Foods. In: W. P. T. James and O. Theander (eds.) The analysis of dietary fiber in food. Marcel Dekker, New York.

[140] Roughan, P. G., Holland, R. (**1977)**. Predicting In Vivo Digestibilities Of Herbages By Exhaustive Enzymic-Hydrolysis Of Cell-Walls. *J. Sci. Food Agric.* 28, 1057 – 1064.

[141] Shao, L. M., Lin, X. Q., Shao, X. G., (**2002)**. A Wavelet Transform And Its Application To Spectroscopic Analysis. Appl. Spectrosc. Rev. 37, 429 – 450\

[142] Donoho, D. L., Johnstone, I. M., (**1995)**. Adapting to unknown smoothness via wavelet shrinkage. J. Am. Stat. Assoc. 90, 1200 - 1224.

[143] Stone, M., (**1974)**. Cross-validatory choice and assessment of statistical predictions. J. R. Stat. Soc. Series B Stat. Methodol. 36, 111 - 147.

[144] Hastie, T., Tibshirani, R., Friedman, J., (**2009)**. The elements of statistical learning: data mining, inference and prediction, 2nd edition, Springer.

[145] Chen, Q. S., Zhao, J. W., Liu, M. H., Cai, J. R., Liu, J. H., (**2008)**. Determination of total polyphenols content in green tea using FT-NIR spectroscopy and different PLS algorithms. J. Pharm Biomed. Anal. 46, 568 – 573

[146] De Bei, R., Cozzolino, D., Sullivan, W., Cynkar, W., Fuentes, S., Dambergs, R., Pech, J., Tyerman, S., (**2011)**. Non-destructive measurement of grapevine water potential using near infrared spectroscopy. Aust. J. Grape Wine Res., 17, 62-71

[147] Feret, J. B., Francois, C., Gitelson, A., Asner, G. P., Barry, K. M., Panigada, C., Richardson, A. D., Jacquemoud, S., (**2011)**. Optimizing spectral indices and chemometric analysis of leaf chemical properties using radiative transfer modeling. Remote Sens. Environ. 115, 2742 – 2750

[148] Gowen, A. A., Downey, G., Esquerre, C., O'Donnell, C. P., (**2011)**. Preventing over-fitting in PLS calibration models of near-infrared (NIR) spectroscopy data using regression coefficients. J. Chemom. 25, 375 – 381.

[149] Kramer, R., (**1998)**. Chemometric techniques for quantitative analysis, Marcel Dekker, New York.

[150] Aldrich, J. (**1998**). "Doing Least Squares: Perspectives from Gauss and Yule". International Statistical Review. 66 (1): 61–81.

[151] Watkins, DS. (**2002)** Fundamentals of Matrix Computations New York: John, Wiley and Sons, Inc., 7.

[152] Phatak, A., & DeJong, S. (**1997**). The Geometry of Partial Least Squares. *Journal of Chemometrics*, 11(4).

[153] K. Takeuchi, H. Yanai and B. Mukherjee, The Foundations of Multivariate Analysis, Wiley Eastern, New Delhi (**1982**).

[154] Anderson, T. W., Introduction to Multivariate Statistical Analysis. New York: John, Wiley and Sons, Inc., (**1958)**.

[155] Massy, W. F. (**1965**). Principal Components Regression in Exploratory Statistical Research. *Journal of the American Statistical Association*, 60(309), 234-256.

[156] Höskuldsson, A. (**1988**). PLS regression methods Journal of Chemometrics 2 pp. 211–228.

[157] Dayal, B. S., & MacGregor, J. F. (**1997**). Improved PLS algorithms. Journal of Chemometrics, 11(1), 73-85.

[158] Dejong, S., & Terbraak, C. J. F. (**1994**). Comments on The PLS Kernel Algorithm. Journal of Chemometrics, 8(2), 169-174.

[159] Rannar, S., Lindgren, F., Geladi, P., & Wold, S. (**1994**). A PLS Kernel Algorithm for Data Sets with Many Variables and Fewer Objects .1. Theory and Algorithm. Journal of Chemometrics, 8(2), 111-125.

[160] Kumar, S., Lahlali, R., Liu, X., & Karunakaran, C. (**2016**). Infrared spectroscopy combined with imaging: A new developing analytical tool in health and plant science. *Applied Spectroscopy Reviews, 51*(6), 466-483.

[161] Largo-Gosens, A., Hernandez-Altamirano, M., Garcia-Calvo, L., Alonso-Simon, A., Alvarez, J., & Acebes, J. L. (**2014**). Fourier transform mid infrared spectroscopy applications for monitoring the structural plasticity of plant cell walls. *Frontiers in Plant Science, 5*, 15.

[162] Turker-Kaya, S., & Huck, C. W. (**2017**). A Review of Mid-Infrared and Near-Infrared Imaging: Principles, Concepts and Applications in Plant Tissue Analysis. *Molecules, 22*(1), 20.

[163] Zebeli, Q., Mansmann, D., Steingass, H., & Ametaj, B. N. (**2010**). Balancing diets for physically effective fibre and ruminally degradable starch: A key to lower the risk of sub-acute rumen acidosis and improve productivity of dairy cattle. *Livestock Science, 127*(1),

[164] Bradford, B. J., & Mullins, C. R. (**2012**). Invited review: Strategies for promoting productivity and health of dairy cattle by feeding nonforage fiber sources. *Journal of Dairy Science, 95*(9), 4735-4746.

[165] Haar, Alfréd (**1910**), "Zur Theorie der orthogonalen Funktionensysteme [On the theory of orthogonal function systems]", Mathematische Annalen, 69 (3): 331–371

[166] K.H. Norris, Extracting information from spectrophotometric curves - Predicting chemical composition from visible and near infrared spectraFood, (**1983**) Research and Data Analysis–Proc. IUFOST Symposium, Applied Science Publishers, London, UK, pp. 95–113.

[167] Asekova, S., Han, S. I., Choi, H. J., Park, S. J., Shin, D. H., Kwon, C. H., . . . Lee, J. D. (**2016**). Determination of forage quality by near-infrared reflectance spectroscopy in soybean. *Turkish Journal of Agriculture and Forestry, 40*(1), 45-52.

[168] Cozzolino, D., Fassio, A., Fernandez, E., Restaino, E., La Manna, A., (**2006)**. Measurement of chemical composition in wet whole maize silage by visible and near infrared reflectance spectroscopy. *Anim. Feed Sci. Technol. 129, 329 – 336.*

[169] Xiccato, G., Trocino, A., De Boever, J. L., Maertens, L., Carabano, R., Pascual, J. J., Perez, J. M., Gidenne, T., Falcao-E-Cunha, L., (**2003)**. Prediction of chemical composition, nutritive value and ingredient composition of European compound feeds for rabbits by near infrared reflectance spectroscopy (NIRS). *Anim. Feed Sci. Technol. 104, 153 - 168.*

[170] Boulet-Audet, M., Buffeteau, T., Boudreault, S., Daugey, N., & Pezolet, M. (**2010**). Quantitative Determination of Band Distortions in Diamond Attenuated Total Reflectance Infrared Spectra. *Journal of Physical Chemistry B, 114*(24), 8255-8261.

[171] Chang, C. W., Laird, D. A., Mausbach, M. J., Hurburgh, C. R., (**2001**). Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.* 65, 480-490.

[172] Zornoza, R., Guerrero, C., Mataix-Solera, J., Scow, K. M., Arcenegui, V., & Mataix-Beneyto, J. (**2008**). Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biology & Biochemistry, 40*(7), 1923-1930.

[173] Bakken, A. K., Randby, A. T., & Uden, P. (**2011**). Changes in fibre content and degradability during preservation of grass-clover crops. *Animal Feed Science and Technology, 168*(1-2), 122-130

[174] Boominathan, R., Saha-Chaudhury, N. M., Sahajwalla, V., & Doran, P. M. (**2004**). Production of nickel bio-ore from hyperaccumulator plant biomass: Applications in phytomining. *Biotechnology and Bioengineering, 86*(3), 243-250.

[175] Cheng, L., Judson, H. G., Bryant, R. H., Mowat, H., Guinot, L., Hague, H., . . . Edwards, G. R. (**2017**). The effects of feeding cut plantain and perennial ryegrass-white clover pasture on dairy heifer feed and water intake, apparent nutrient digestibility and nitrogen excretion in urine. *Animal Feed Science and Technology, 229*, 43-46

[176] Khan, M. M. H., & Chaudhry, A. S. (**2011**). A Comparative Study of Low And High Quality Forages For Chemical Composition And In Vitro Degradability. *Journal of Animal and Plant Sciences, 21*(4), 715-723.

[177] Parrini, S., Acciaioli, A., Crovetti, A., & Bozzi, R. (**2018**). Use of FT-NIRS for determination of chemical components and nutritional value of natural pasture. *Italian Journal of Animal Science, 17*(1), 87-91.