Investigating changes in the microstructure of calcified tissues using Raman microspectroscopy and chemometrics

A thesis presented in partial fulfilment of the

requirements for the degree of

Master of Science

in

Chemistry

at Massey University, Manawatū, New Zealand.

Eileen Ama Mansa Obben

2020

Abstract

Bone fracture is a growing health concern in medicine. Current clinical assessment methods for bone fracture risk are bone mineral density (BMD)-based, as bone quantity is the only aspect of bone strength that is most readily measured in clinical practice. On their own, these framework methods can not exactly predict the likelihood of individuals to fracture, since bone strength and health are influenced by both quantity and quality; it is a multi-factored probability. A sounder understanding of both facets of bone strength would expedite more accurate fracture risk assessment. Better prevention and treatment of orthopaedic diseases now rest on a greater understanding of bone quality and its underlying factors, since bone quantity has historically received more research attention. One route to confront this challenge in progressing comprehension of the underlying mechanisms of bone quality is to use animal models of human bone diseases like osteoporosis and osteoarthritis (OA). Given that any atypical chemical alterations to bone's main components are reflected in its microstructure, and therefore contribute to the development of various bone diseases, there is increasing interest in how molecular-level changes to bone affect overall bone quality.

Molecular vibrational spectroscopy is often used as a tool in disease diagnosis, as any disease-causing chemical alterations may be identified and monitored; it also holds the potential to enable prediction of any further complications. As Raman spectroscopy is not as water-sensitive as infrared is, it is highly beneficial for characterising biological specimens.

Bone tissues and other biological specimens are inherently intricate, as would be the chemical information collected from them; multivariate statistical analysis is required to aid in the simplification, extraction, and classification of these large volumes of chemical information collected. This cataloguing of the actual variation of bone tissue's chemical information would improve understanding of how damage affects the interplay between bone's various micro- and macrostructural aspects.

Principal component analysis (PCA) – one such dimensionality-reducing statistical technique – was conducted on Raman spectral data collected from two separate sets of equine bone specimens: fracture-prone third metatarsal (Mt3) and induced osteoarthritic (OA) carpal joint sections. The results from both aggregated data sets suggested that some localised microstructural differences were detectable – especially within parts of the subchondral bone. What was unclear, however, was the likely cause of these differences. These differences could

potentially be highlighting areas of hypermineralisation or some organic matrix degradation within fracture predilection sites or OA-induced sites that may well be indicative of early development of orthopaedic diseases like osteoporosis or OA. Some of the common questions the PCA results raised were the extent of similarity between individuals with respect to the organic matrix component, and the extent of heterogeneity between individuals with respect to the mineral component. In order for any potential predictions to be applicable, addressing the multi-level nature of the multivariate spectral data obtained would be the first step in preparing this type of work for further validation and classification. Widening the scope of data analysis might then help in clarifying the classification of the spectral data.

If not already available, condensed, fibre optic-style instrumentation might enable trialling of this technique in a practical, clinical setting. If it is practically feasible, instrumentation that even combines the two vibrational spectroscopic techniques in tandem with chemometrics to provide simultaneous groups of data from samples, could also be developed.

Acknowledgements

I want to thank my chief supervisor, Associate Professor Mark Waterland, for his instruction, patience, understanding, and guidance throughout this project. I would also like to thank my co-supervisor, Professor Geoff Jones, for his guidance with the statistical analysis. My thanks also go to Catherine Nicholson and Luca Panizzi at the School of Veterinary Sciences for their respective provision of the equine third metatarsal and carpal joint bone specimens, as well as offering feedback with enquiries regarding general aspects of equine anatomy. Thanks also go to fellow research group members Haidee Dykstra for her initial assistance with use of the Raman microscope – especially when I encountered technical difficulties – and Josiah Cleland for his guidance with the pre-processing of my spectral data.

I would like to thank my family – especially my parents – for their support and encouragement throughout this journey. Unfortunately, during this journey, my parents passed away within two years of each other, both from cancer; I dedicate this thesis in loving memory of them.

I also wish to convey my sincerest gratitude to hospice personnel, various co-members the of the Massey University African Students' Association and fellow members of my church parish for their continued support in these challenging times. I cannot name all of you, but you know who you are. If I have forgotten anyone else, my apologies – I would give an all-encompassing 'thank you' if I could!

Above all, thanks go to God Almighty, for allowing me to further my education, as well as the grace and strength to see this through to the end; may my perseverance in this endeavour be to Your glory, Lord.

Table of Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations

2D:	2-dimensional
3D:	3-dimensional
CH_2:	Alkyl
ANOVA:	Analysis of variance
ACC:	Articular calcified cartilage
AC:	Articular cartilage
ALS:	Asymmetric least squares
BSEM:	Backscattered electron microscopy
BMD:	Bone mineral density
CO_3^{2-}:	Carbonate
C=O:	Carbonyl
C_3:	Third carpal bone
C_4:	Fourth carpal bone
C_R:	Radial carpal bone
C_U:	Ulnar carpal bone
CCD:	Charge-coupled device
CLS:	Classical least-squares analysis
G660:	Collagenase-induced osteoarthritis, with 660 nm laser therapy treatment
G780:	Collagenase-induced osteoarthritis, with 780 nm laser therapy treatment
GCOL:	Collagenase-induced osteoarthritis, without treatment
GCON:	Collagenase-induced osteoarthritis control
CT:	Computed tomography
DA:	Discriminant analysis
DXA:	Dual-energy X-ray absorptiometry
FT-IR:	Fourier-Transform infrared
FT-Raman:	Fourier-Transform Raman
HCA:	Hierarchical cluster analysis
HAC:	Hyaline articular cartilage
OH:	Hydroxide

IR:	Infrared
ISCD:	International Society for Clinical Densitometry
L*n*:	Layer *n*
LDA:	Linear discriminant analysis
LLLT:	Low-level laser therapy
Mc3:	Third metacarpal bone
Mt3:	Third metatarsal bone
NIR:	Near-infrared
NA:	Numerical aperture
OA:	Osteoarthritis
OVX:	Ovariectomised
PLS-DA:	Partial least-squares-discriminant analysis
PLSR:	Partial least-squares regression
pQCT:	Peripheral quantitative computed tomography
Phe:	Phenylalanine
PO_4^{3-}:	Phosphate
PMMA:	Polymethylmethacrylate
PC(s):	Principal component(s)
PCA:	Principal component analysis
PCA-LDA:	Principal component analysis-linear discriminant analysis
PSB(s):	Proximal sesamoid bone(s)
ROC(s):	Receiver operating characteristic(s)
RPCA:	Relative principal component analysis
RMSECV:	Root mean-square-error-of-cross-validation
SFS:	School of Fundamental Sciences (formerly the Institute of Fundamental Sciences, IFS)
SVS:	School of Veterinary Sciences (formerly the Institute of Veterinary, Animal, and Biomedical Sciences, IVABS)
S*n*:	Section *n*
SHAM:	Sham-operated
SNV:	Single normal variate
SIMCA:	Soft-independent modelling of class analogy or analogies
SF FT-IR:	Specular reflectance Fourier-Transform infrared spectroscopy

SD(s):	Standard deviation(s)
S. aureus:	*Staphylococcus aureus*
SCB:	Subchondral bone
UV-Vis:	Ultraviolet-visible
VBG:	Volume Bragg Grating
WHO:	World Health Organisation

Chapter 1 – Introduction

1.1 Functions of the skeletal system

The mammalian skeletal system mainly imparts structural support and enables body movements, protects vital internal organs, and maintains mineral and lipid homeostasis (balance of physiological processes within an organism; it stores such necessary minerals as calcium, phosphorus, and magnesium) (1-4). It also acts as a centre for haematopoiesis (blood production and storage within the bone marrow of the medullary cavity of long bones) (1-4).

Bone is always responding to the loads and forces that are applied to it during physical activity, adapting itself accordingly. When these loads and forces exceed what bone can readily or easily handle to maintain its optimum architecture, bone tends to fracture.

Riggs (5) outlines three general pathways that can lead to fracture: abnormal loading of normal bone ('monotonic fracture'), normal loading of abnormal bone ('pathological fracture'), and intermittent, cyclic loading of fatigued bone ('fatigue fracture'). Physiological changes caused by disease, other pathological anomalies or prescriptive drugs disrupt the delicate balance between bone resorption and bone formation processes. These all distort the 'normal' state of bone's structure. Fracture susceptibility and manifestation are the weightiest clinical concerns about bone diseases (4) as they entail profound disability and morbidity – perhaps even mortality, in the most severe cases.

1.2 Structure and composition of bone

There are five classes of bone: long, short or cuboidal, sesamoid (small bones found within tendons), flat, and irregular. Long bones provide leverage; short or cuboidal bones offer stability, support, and shock absorption (2, 6, 7). Sesamoid bones shield tendons and joints from compressive and tensile forces; flat bones grant muscle attachment points and protection of internal organs; irregular bones also guard internal organs (2, 6, 7).

Bone has a complex hierarchical and heterogeneous structure. Macroscopically, mature bone tissue consists of a hard, dense exterior known as cortical (or 'compact') bone that surrounds highly porous trabecular (or 'cancellous' or 'spongy') bone. The perforation of cortical bone by holes and channels allows blood vessels and nerves into the medullary cavity where marrow resides; refer to Figure 1.1.

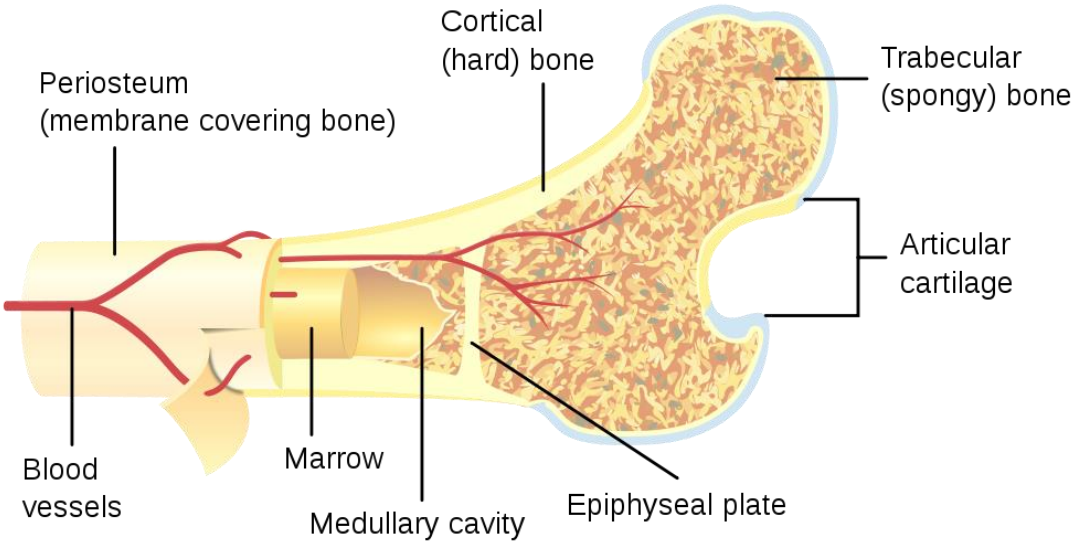

Figure 1.1 Cross-section of a long bone showing the trabecular and cortical bone encompassing marrow in the medullary cavity. From "File: Bone cross-section.svg", by Pbroks13, 2008 (https://commons.wikimedia.org/w/index.php?curid=5188772).
CC BY 3.0. (https://creativecommons.org/licenses/by/3.0/).

The surface where bones articulate is known as the articular surface. Underlying the articular surface is the articular cartilage (AC) which has two main zones (outermost to innermost): the hyaline articular cartilage (HAC) and the articular calcified cartilage (ACC; the ACC can simply be referred to as the calcified cartilage). The HAC is 'uncalcified' and can be further subdivided into three zones (again, outermost to innermost): superficial, middle, and deep. A tidemark or mineralising front separates the calcified and 'uncalcified' parts of the articular cartilage. Deeper are the subchondral bone (SCB) and cancellous bone (or 'trabecular bone'); these tissue components are more mineralised than that of the ACC.

Figure 1.2 illustrates the osteon (or 'Haversian system'), which is a repetitive microscopic unit consisting of concentric layers known as lamellar bone (or 'lamellae'). Osteons are found in cortical bone. Haversian canals, which run parallel to the long axis of a long bone, contain blood vessels and nerves that service the bone (that is, it is the nutrient canal of the bone). An osteon is where the bone has deposited and gradually mineralised. Like the rings of a tree, moving outwards from the centre of an osteon, the younger bone gives way to areas of older, more mineralised bone ('interstitial lamellae' or 'interstitial bone') and each osteon is separated from the others by a cement line. Volkmann's canals run perpendicular to the long axis of the bone, connecting each of the Haversian systems (this permits communication between them).

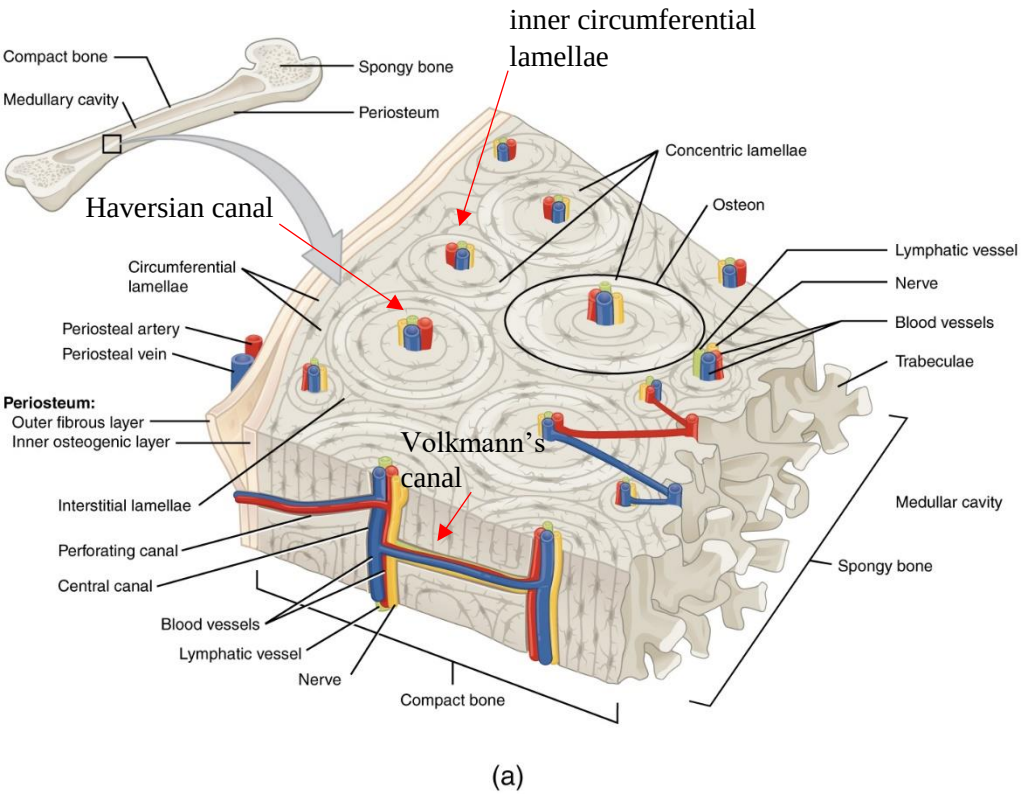

Figure 1.2 A cross-sectional view of the basic (microscopic) structural unit of bone: the osteon (adapted from "Bone structure" by OpenStax, 2013 (https://philschatz.com/anatomy-book/contents/m46281.html)). CC BY 4.0. (https://creativecommons.org/licenses/by/4.0/)

Bone is composed mainly of a mineral phase, an organic matrix, and water. The combination of mineral and organic components are what makes bone simultaneously strong, rigid, and flexible (4). Making up approximately 20 – 25 % of the composition by weight (1), the organic matrix is mostly made up of type I collagen, but also includes non-collagenous proteins and lipids (8). The collagenous component gives bone its flexibility and is a mineralisation template (9); this elasticity allows bone to withstand tensile loads and forces (10).

The mineral phase (making up approximately 70 % of the composition (1)) is based on apatite, a calcium phosphate-containing species. The mineral component confers compressive resistance to the matrix, as its crystals are bound to the collagen fibres (10).

1.3 Diagnosis of bone diseases and current methods of risk assessment

From a clinical perspective, bone mineral density (BMD) is perceived as a suitable surrogate for skeletal mineral content (11). Two main reasons for its prevalence in clinical settings are

ease of its *in situ* measurement and the tendency towards fracture outcomes from low BMD-related values (11). Current clinical assessment methods for bone fracture risk are BMD-based as bone quantity is the only aspect of bone strength that is most readily measured in clinical practice (11-13). For example, the BMD-based photon absorption techniques, single and dual X-ray absorptiometry (DXA), quantitatively assess the areal mineral content of the entire skeletal system and sites most vulnerable to fracture. Patients' DXA test results are reported back as so-called '*T*-scores' and '*Z*-scores'; these measurements are based on BMD criteria from the World Health Organisation (WHO) Fracture Risk Assessment Working Group and the International Society for Clinical Densitometry (ISCD) (14, 15).

The *T*-score – used for a patient who is aged 20 and above – is the number of standard deviations (SDs) that his or her bone density differs from that of an ethnicity- and sex-matched healthy young adult. That is, the *T*-score shows the level of deviation from ethnicity- and sex-matched peak bone mass and is therefore indicative of bone loss since early adulthood. The *Z*-score, by contrast, is reported for all age groups. The *Z*-score is the number of SDs that a patient's bone mass differs from that of age-, ethnicity-, height-, weight-, pubertal status-, and sex-matched bone mass (that is, deviation from appropriately matched reference individuals who have no fragility fractures) (14, 15). According to the WHO's criteria, a *T*-score at and above 1.0 SD suggests normal bone density. The threshold for diagnosis of osteopenia (low bone mass) is when a *T*-score falls between 1.0 and 2.5 SDs below the mean, and for osteoporosis is when a *T*-score is at or below 2.5 SDs. A patient with a history of at least one prior fracture caused by mild trauma and a *T*-score below 2.5 SDs is more likely to be diagnosed with severe or established osteoporosis (14, 15).

Where the *Z*-score is a more appropriate tool for osteoporosis or osteopenia – as, for instance, would especially be the case for paediatric patients – the criteria for diagnosis are two-fold: a significant fracture history and a *Z*-score below 2.0 SDs. According to the ISCD's guidelines, this signals that a patient's bone mass falls below the expected range for his or her age. A *Z*-score above 2.0 SDs signifies that a patient's bone mass falls within the expected range for his or her age (14, 15). It is due to this ease of categorisation of BMD-based values that DXA is considered the benchmark or gold standard diagnostic tool for diseases like osteoporosis.

On their own, these framework methods cannot unerringly predict the likelihood of individuals to develop a fracture, since bone strength and health are influenced by both quantity and quality; it is a multi-factored probability (16). An individual may have seemingly

normal bone density, yet inexplicably fracture. Another individual may have lower bone density but may not fracture (16). Most research attention has historically been given to bone quantity. Still, there has been greater recognition of the need to be additionally addressing bone quality and the factors that affect it (17).

Atypical alterations to either or both of bone's main constituents contribute to the development of various bone diseases. For instance, mutations in collagenous and non-collagenous proteins can make bone more brittle (giving rise to conditions such as osteogenesis imperfecta) (4). Osteomalacia is the end-result of inadequate mineralisation from dietary calcium deficiencies or phosphate deficiencies (4). Other methods are being sought to enable better grasp of bone's chemical and structural composition (that is, its microstructure). These have the potential to fill any gaps in knowledge left by BMD-based measurements. Studies of chemical composition can provide insight into how interactions and alterations to bone components may affect the overall bone quality (given that any perturbations to bone are reflected in its microstructure) (18). A sounder understanding of both facets of bone strength expedites more accurate fracture risk assessment, and more generally, better-targeted medical practices.

Bone fracture is a growing health concern in medicine. Annually, there are about 9 million osteoporotic fractures globally (13) because of low bone mass and tissue deterioration; these are more common in the hip, forearm, and spine (19). It is expected that as life expectancy increases (and, subsequently, a greater proportion of countries' populations move into higher age brackets) and as more individuals become increasingly sedentary, the rates of osteoporosis and other orthopaedic diseases will rise (20). Greater bone fragility means greater fracture susceptibility (12). Such alarming prospects highlight the obvious need for better prevention and treatment of orthopaedic diseases. One route to confront this challenge is to use animal models of bone disease to progress comprehension of underlying mechanisms.

1.4 The horse as a veterinary model for bone diseases in humans

The equine and human skeletal systems are somewhat similar physiologically. Both creatures undergo an initial phase of rapid musculoskeletal growth and development before attaining skeletal maturity and progressively ageing (21-23). Both are also capable of adaptation to

mechanical loading (21-23). As there are both human and equine athletes, some parallels can be drawn between ailments that can afflict both creatures (21-23). Progressive ageing may bring with it naturally occurring, age- and exercise-related musculoskeletal diseases (21-23). The veterinary field, therefore, is increasingly interested in how equine bone strength and resistance are altered at the molecular level. These alterations might show how equine models for bone diseases and other musculoskeletal disorders and injuries might also help in human medicine. Since many horses are often euthanised after significantly traumatic injuries, and many ethical considerations need to be taken into account regarding the use of any human tissue in research, equine models provide a good starting point. There are not yet clear-cut clinical methods of identifying individuals who are likely to develop fractures. The exact pathogenesis of fractures, likewise, remains at best rather speculative.

1.4.1 The equine skeletal system

Figure 1.3 shows a picture of the equine skeletal system.

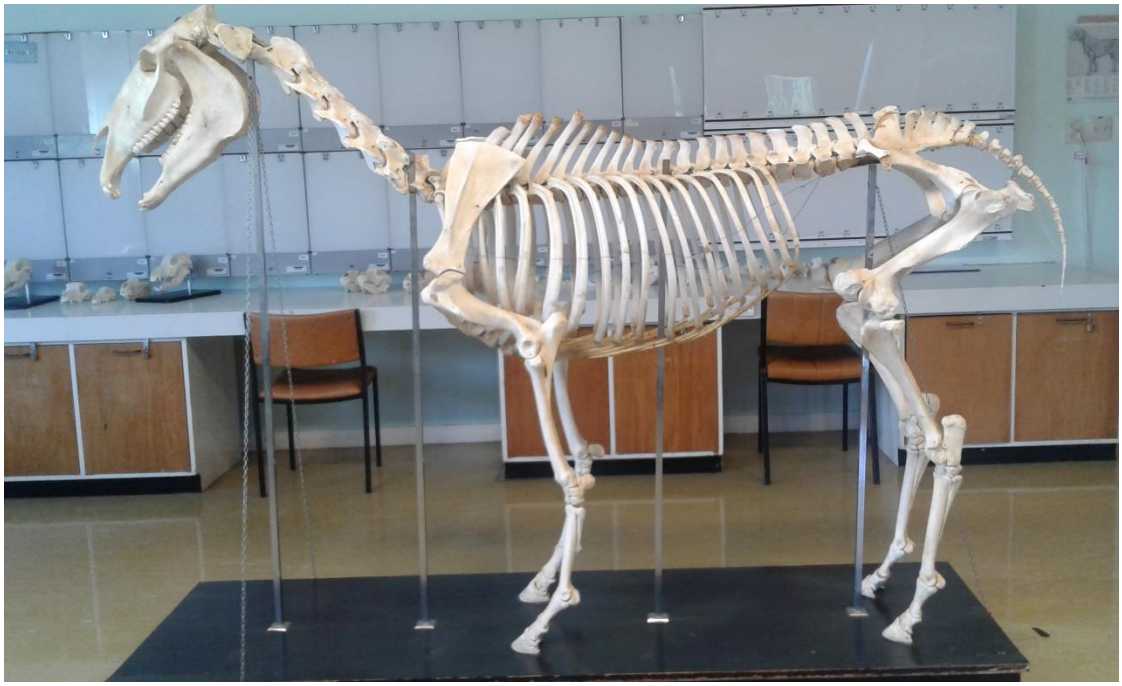

Figure 1.3 A picture of the equine skeletal system in one of the teaching laboratories of the School of Veterinary Sciences (SVS; formerly the Institute of Veterinary, Animal, and Biomedical Sciences, IVABS) at Massey University. Image by author.

1.4.2 Hind limb in the equine skeletal system

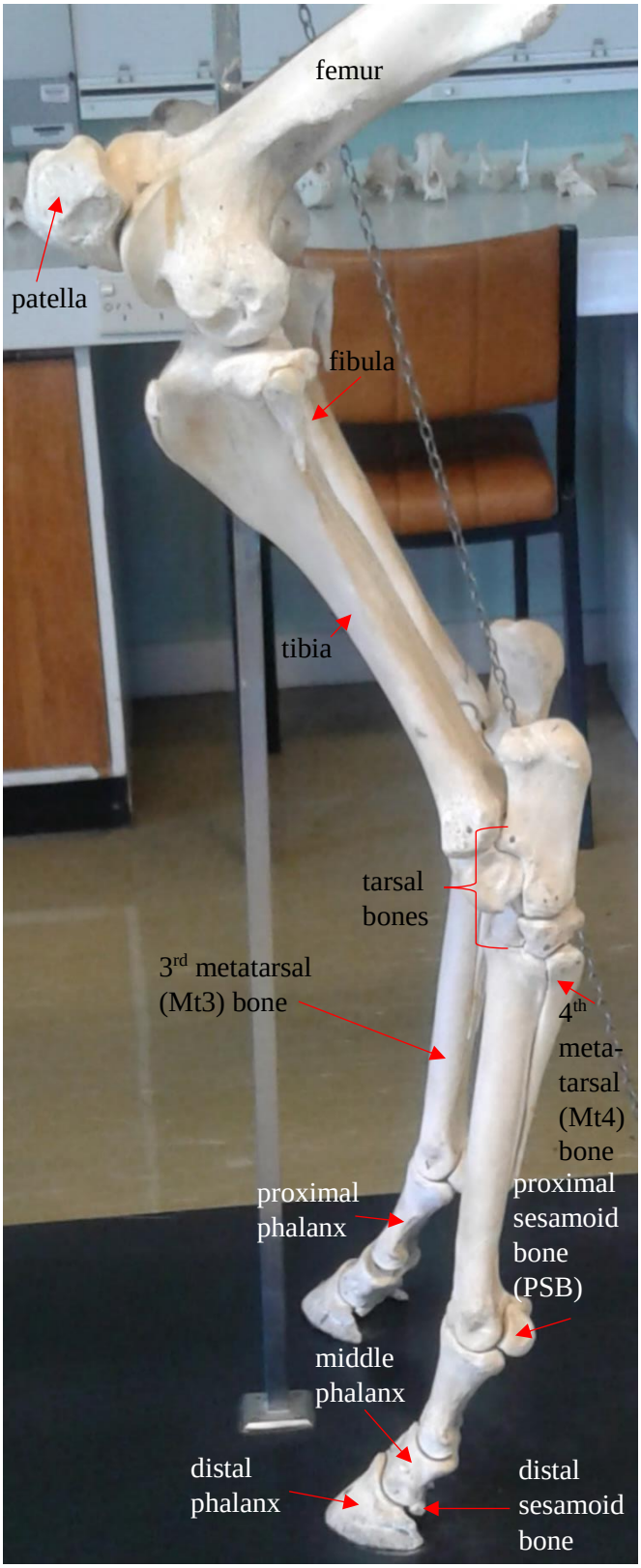

Figure 1.4 The equine hind limb; the third metatarsal (Mt3) bone was the most relevant with respect to the first part of the work conducted herein. Image by author.

The pelvis, femur, tibia and fibula, tarsus (or ‘hock’), three metatarsals (cannon and splint bones), three phalanges (long and short pastern bones, and the pedal bone), and three sesamoid bones (the ‘sesamoids’ and navicular bone) make up the hind limb of the horse; refer to Figure 1.4.

1.4.2.1 Third metatarsal (Mt3) bone

As indicated in Figure 1.5, the Mt3 bone (also known as the ‘cannon bone’) articulates proximally with the tarsus and distally with the phalanx.

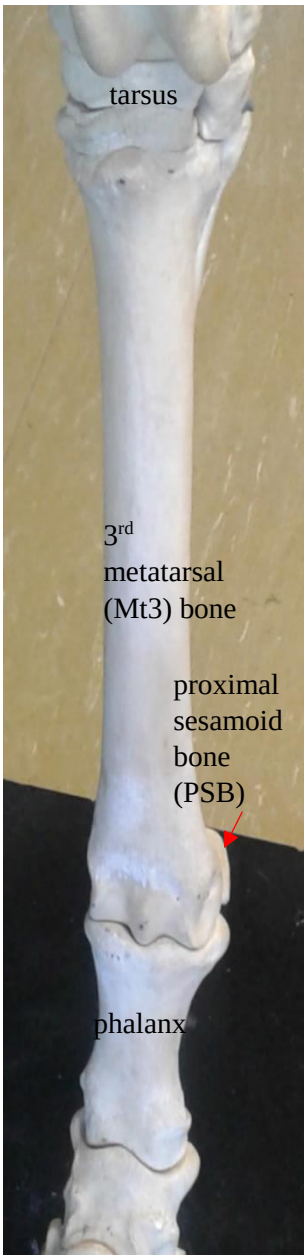

Figure 1.5 The equine third metatarsal (Mt3) bone and its articulations. Image by author.

The distal epiphysis consists of the lateral and medial condyles that are separated by a sagittal ridge. Immediately adjacent to the sagittal ridge are the lateral and medial parasagittal

grooves; these anatomical components are apparent in Figure 1.6. The Mt3 bone is one of the bones within the athletic equine skeletal system prone to fracturing; it was the distal part of the Mt3 that bone sections came from for the first part of the data analysis in this body of work.

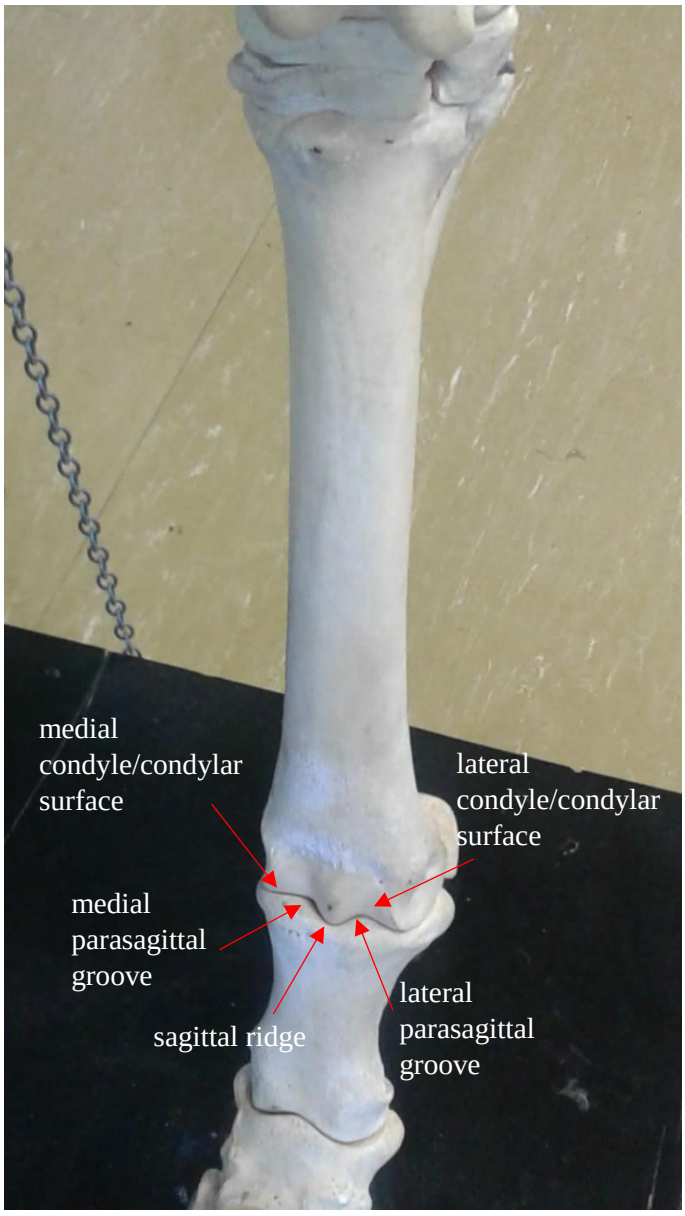

Figure 1.6 The areas of the distal epiphysis of the equine third metatarsal (Mt3) bone. Image by author.

1.4.3 Forelimb in the equine skeletal system

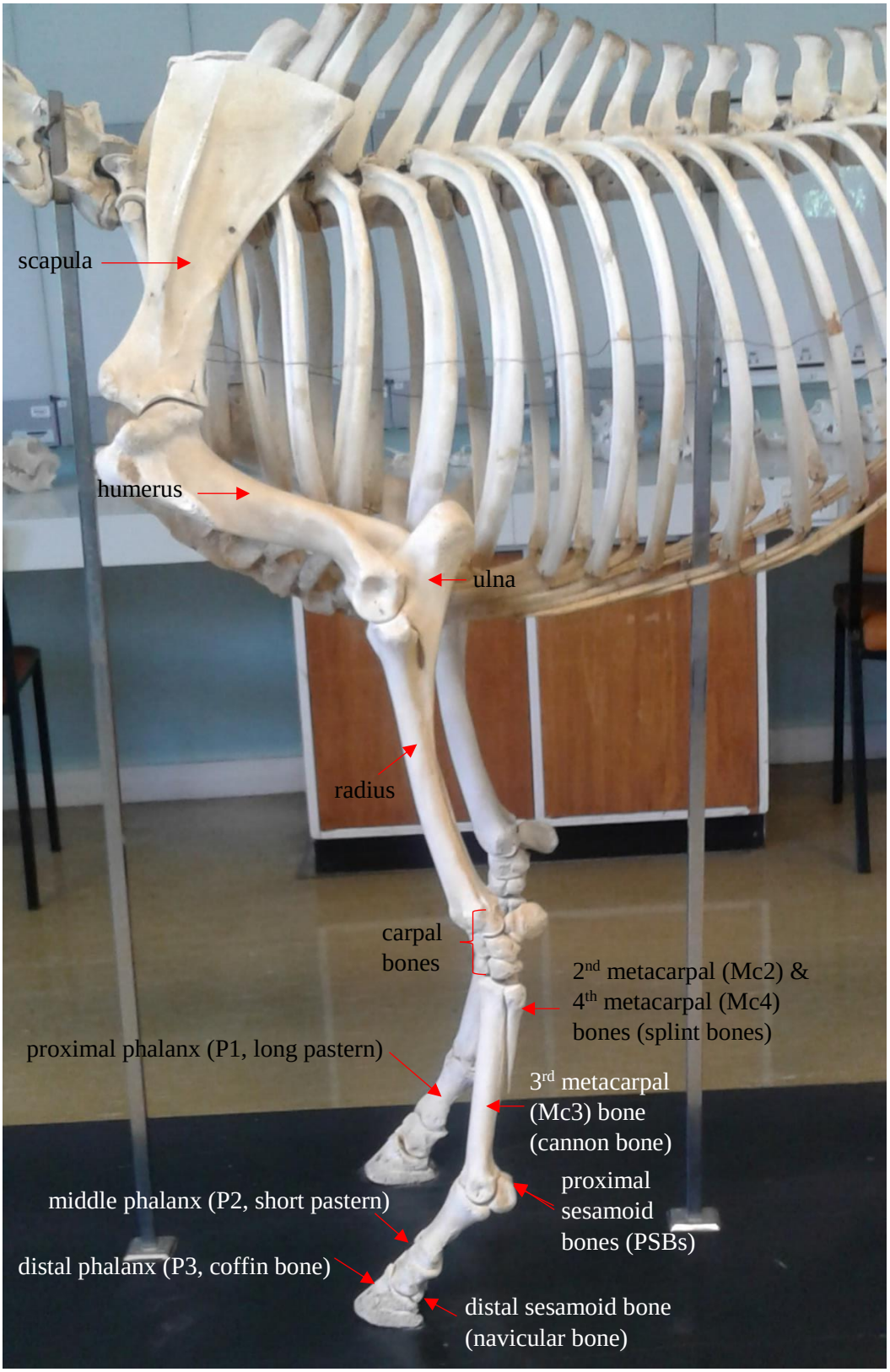

Figure 1.7 The equine forelimb; the bones of the carpal joint were those of interest with respect to the second part of the work conducted herein. Image by author.

As displayed in Figure 1.7, the forelimb of the horse is comprised of the scapula, humerus, radius and ulna, carpus (or 'knee'), three metacarpals (cannon and splint bones), three phalanges, and three sesamoid bones (again, sesamoids and navicular bone). The bones of the carpal joint were those of interest as regards the second half of the work conducted in this thesis.

1.4.3.1 Carpal joint

The bones in the carpus (or carpal joint) are examples of short bones. Seven bones make up the carpus, arranged in two rows: the radial carpal (C_R), intermediate carpal (C_I), ulnar carpal (C_U), and accessory carpal in the upper row; the second carpal (C_2), third carpal (C_3), and fourth carpal (C_4) are in the lower row. The radius is proximal to the C_R, C_I, C_U, and accessory carpal bones, the third and fourth metacarpal (Mc3 and Mc4, respectively) bones are proximal to the C_2, C_3, and C_4 bones. Figure 1.8 shows the distal part of the forelimb, and Figure 1.9 gives a close-up of the carpal joint.

It was the C_3, C_4, and C_R bones of the carpal joint from which bone sections were taken for the second part of the data analysis in this body of work.

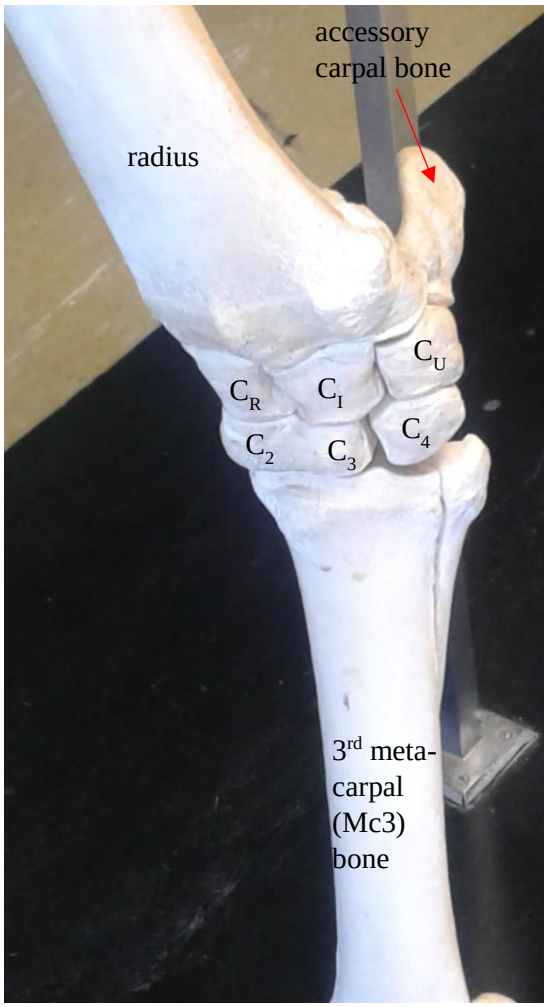

Figure 1.8 The distal equine carpal joint. Image by author.

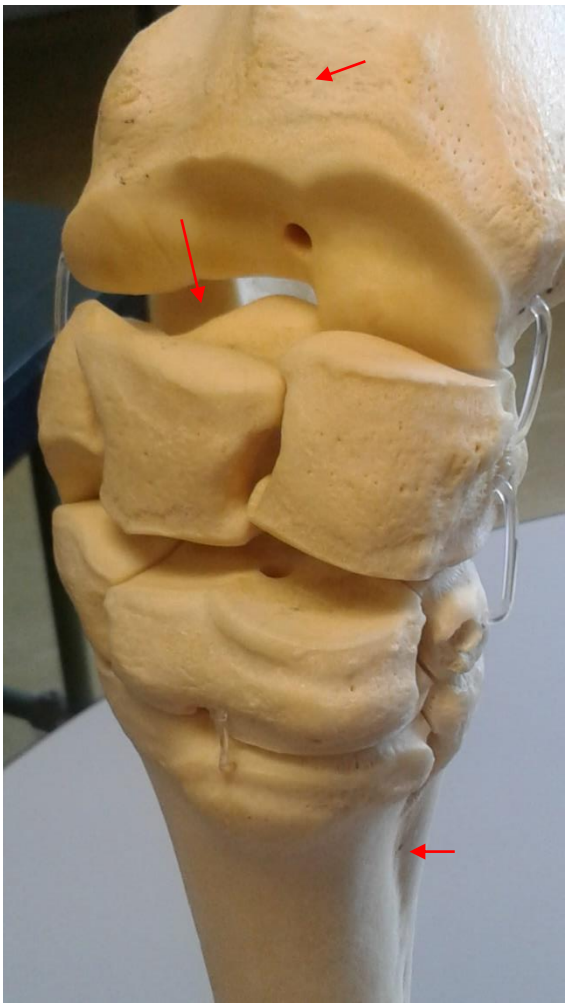

Figure 1.9 A close-up of the equine carpal joint. Image by author.

1.5 Molecular vibrational spectroscopy

Molecular vibrational spectroscopy probes the vibrational motion of bonds and groups of atoms in molecules and materials. Vibrational frequency mode is governed by atomic mass and bond strength, while intensity is determined by the magnitude of dipole change in the direction of vibrational motion. Generally, changes in chemical composition or structure will affect band position and intensity; this provides the basis for differentiating between samples of similar composition and structure. Figure 1.10 shows the infrared region of the electromagnetic spectrum, where molecular vibration occurs.

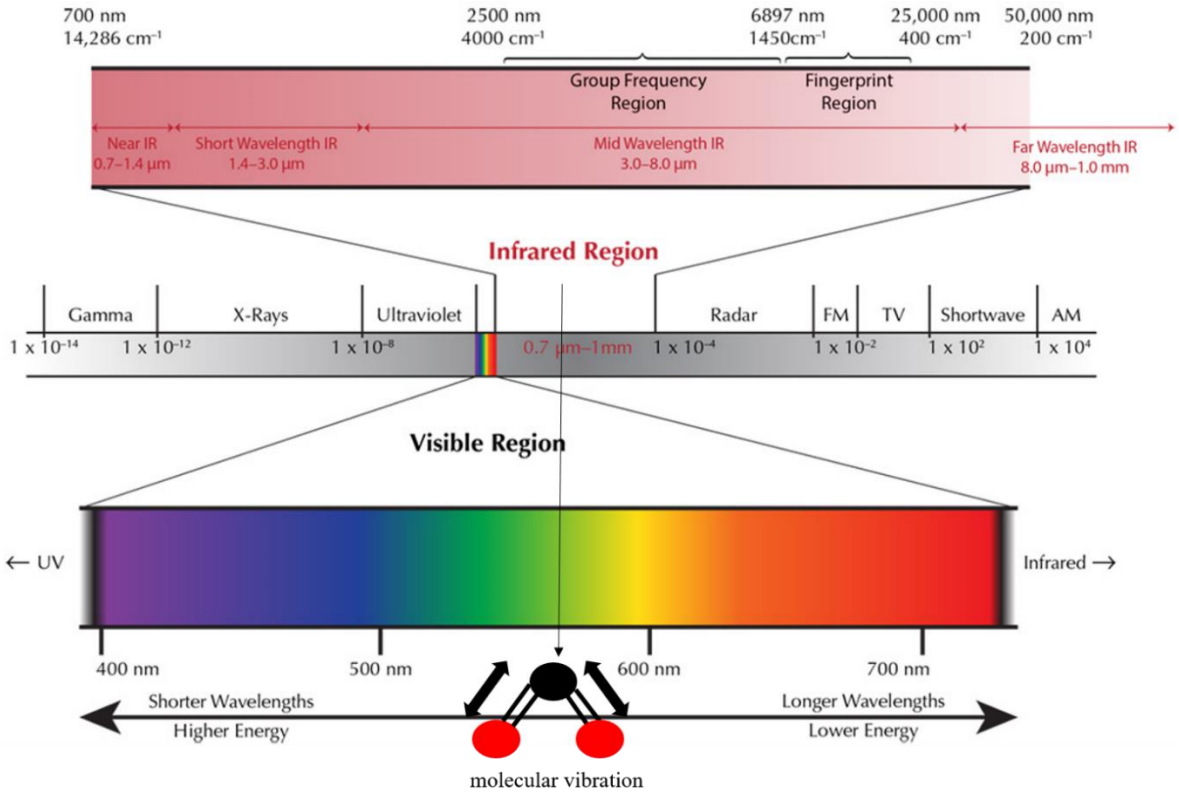

Figure 1.10 Schematic of the nature of molecular vibrational spectroscopy (adapted from "FTIR in Gem Testing" by Lotus Gemology, 2019 (http://www.lotusgemology.com/index.php/library/articles/294-ftir-in-gem-testing-ftir-intrigue-lotus-gemology)). Copyright 2019 by Lotus Gemology. Reprinted with written permission.

The application of molecular vibrational spectroscopy as a tool in disease diagnosis has become of increasing interest over the past two decades or so (24) – particularly as it allows for rapid structural characterisation and evaluation, and is a non-destructive technique (25, 26). With molecular vibrational spectroscopy, any chemical alterations due to a disease process may be identified and monitored, and perhaps even enable prediction of any further complications.

1.5.1 Infrared (IR) spectroscopy

The direct excitation of molecular vibrations from the absorption of infrared photons forms the basis of infrared (IR) spectroscopy. Excitation is the result of an exact match between the molecular vibrational frequency and the radiation frequency as it passes through a layer of the sample medium. Changes in molecular dipole moments (the net distribution of electrical charge across a molecule) during vibration is what IR spectroscopy ascertains. IR spectroscopy is usually carried out in the mid-IR region.

Samples of biological specimens are often aqueous. Infrared spectra of these types of samples show broad hydroxide (OH) bands; that is, water has a strong signal within the spectra. This

signal broadness is the main disadvantage of using IR spectroscopy as a characterisation technique for aqueous samples.

1.5.2 Raman spectroscopy

Raman spectroscopy arises from molecular vibrations caused by the difference in frequency between the incident and scattered photons; that is, it is a light-scattering technique. Only about one in every 10,000 photons, however, interacting with a molecule gives rise to the 'Raman effect', so this Raman effect is weak. There is a much higher probability of there being no loss in energy – and therefore no change in frequency since it is the same as that of the excitation source – as excited molecular vibrations return to the ground state upon re-emission of infrared radiation. This more probable event is known as 'Rayleigh' or 'elastic' scattering ('elastic' since energy from the incident photon is conserved).

Raman scattering is considered 'inelastic' since it results from frequency differences between the incident and scattered photons. Raman scattering manifests in two forms: 'Stokes' (when the frequency of the scattered radiation is lower than that from the excitation beam; a loss of energy), and 'anti-Stokes' (when the frequency of the scattered radiation is higher than that from the excitation beam; a gain of energy). Raman spectroscopy measures changes in molecular polarisation (the temporary distortion of a molecule's electron cloud).

The foremost advantage Raman spectroscopy holds over infrared is that it is not as water-sensitive; this is highly beneficial for samples of biological specimens.

Infrared (IR) and Raman both provide characteristic peaks or bands (so-called 'fingerprints') for constituent chemical species in a given sample. The most important ones in bone include those arising from phosphates and carbonates (mineral component) and amides (collagenous components). IR and Raman spectroscopies are complementary techniques (both involve molecular vibrations but have differing mechanistic origins; refer to Figure 1.11). Together, they provide a fuller picture of the molecular vibrations of samples under study; their use is dependent upon the purpose of the study matter.

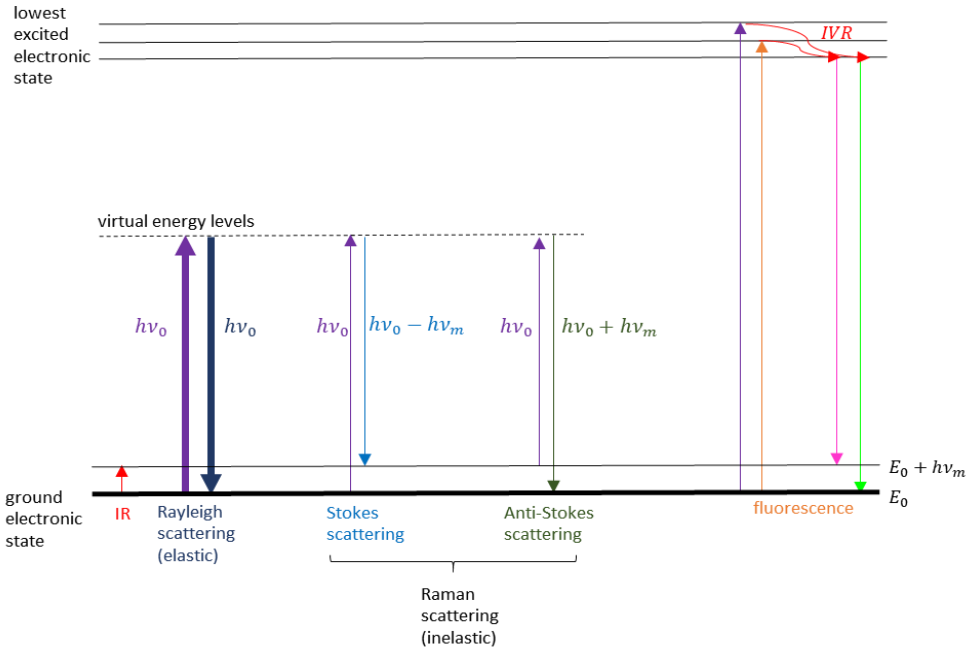

Figure 1.11 A Jablonski energy diagram depiction of the mechanistic origins of infrared (IR) and Raman spectroscopies, as well as the fluorescence that can overwhelm Raman spectra of biological specimens.
E_0: ground state energy; $h\nu_0$: energy of incident photon; $h\nu_m$: energy of molecular vibration; IVR: intramolecular vibrational relaxation. Diagram by author.

1.5.3 Use of near-infrared (NIR) lasers with biological specimens

Fluorescence (atomic or molecular excitation upon irradiation with photons, but followed by short-lived emission) interferes considerably in the collection of Raman spectral data from biological specimens, as it is so many orders of magnitude more intense than the inherently weak Raman scattering is. Near-infrared (NIR) lasers can be used to minimise fluorescence during the collection of Raman scattering from biological samples. NIR laser use is based on the premise that such lasers do not encounter as many strongly absorbing molecules and fluorescence in this part of the electromagnetic spectrum (26-28).

1.6 Multivariate statistical analysis: Chemometrics

Bone tissues and other biological samples are inherently intricate. Any chemical information presented from them will also be complicated and not necessarily straightforwardly interpretable (29, 30). Subtle variations (due, for instance, to pathological abnormalities) and overlaps in large volumes of spectral data are not readily detected manually (30). Unfortunately, any qualitative assignment of these changes is highly subjective and affected by extraneous and confounding factors (29, 30). This complexity necessitates the use of tools

to aid in simplification, extraction, and categorisation or classification of these data (26). Classification or categorisation of the actual variation of chemical information within bone tissue is critical to improving understanding of how damage affects the interplay between the various micro- and macrostructural aspects of bone (29, 31, 32).

Multivariate measurements are common in biology and analytical chemistry; chemometrics can be thought of as the use of statistics and mathematics for the analysis of such chemical data (33). Principal component analysis and linear discriminant analysis are just two examples of such quantitative analytical methods.

1.6.1 Principal component analysis (PCA)

The fundamental aim of principal component analysis (PCA) is to extract the underlying information in the multivariate data; this then requires exploration and quantitative interpretation of patterns within those data (33). Similarities and differences within the data can be highlighted (34). What PCA does is reduce a vast number of experimental variables into a much smaller number of variables, known as 'principal components' and these are used as independent variables in model construction.

These principal components (PCs) represent the direction of greatest variation observed in the original data, weighted by the amount of variation shown. Each PC has a characteristic value – a 'score' – that relates to a particular sample (34). These scores then comprise a new coordinate system used to build up a model that can eventually classify the spectra into different groups based on observed spectral differences caused by molecular structural variations.

1.6.2 Linear discriminant analysis (LDA)

The basic principle of linear discriminant analysis (LDA) is to find a linear combination of features to characterise or separate two or more groups of 'objects'. The resulting combination could be used either directly as a linear differentiate or as a dimensionality reduction in later classification. LDA models class differences to maximise the ratio of between-sample variation to within-sample variation.

What LDA has in common with PCA is that both dimensionality-reduction techniques try to find the linear combination of variables that best account for the variation within the original data. One distinguishing feature of LDA is that it explicitly attempts to model the difference

between the classes of data, based on *a priori* knowledge of the data. Another of its distinguishing features is its use of the class labels for model construction. Still, the risk in using LDA is that it always returns a classification, even if there are no separate classes within the data set.

1.6.3 Example of previous applications of PCA and LDA to equine bone tissue

Diagnosis of musculoskeletal diseases with the aid of molecular vibrational spectroscopy is recent. Most of the (readily available and accessible) literature on its applications to bone tissues centres on infrared (IR) spectroscopy (use of Raman has probably only gathered more considerable momentum within the last decade or so). Literature exists on the application of Raman spectroscopy to some human and other animal bone specimens – particularly murine (rats or mice) bone specimens. There appears to be a paucity of Raman applications to equine bone that are used in tandem with PCA (or even LDA) for classification purposes. Many IR and Raman studies of bone quality in bone specimens have tended to spotlight other aspects of compositional analysis, such as possible implications of changes in specific parameters. These parameters – which are characteristically univariate – tend to be mineral-to-matrix ratio (amount of mineralisation), carbonate-to-phosphate ratio (degree of carbonate accumulation within biological apatite, and mineral crystal maturation status), mineral maturity or crystallinity, and collagen quality (17, 35, 36).

Infrared probing of equine bone tissue is slightly more established than that of Raman but has focused on other subject matters. Proof-of-principle work conducted by Nicholson *et al*. in 2012 (37) combined a type of infrared spectroscopy with chemometrics – specifically, LDA – to try to gain further insight into underlying causes of fracture initiation based on chemical and microstructural information. Their work also explored the use of that chemometric method in analysing that type of IR spectral data.

The bone samples consisted of sections from the distal Mt3 bones of nine young thoroughbred racehorses (four newborn foals, four five-month-old foals, and a three-year-old horse) at sites where fractures are known to initiate (the medial and lateral parasagittal grooves). A third site – the medial condylar surface – acted as the ‘control’. The foals were clinically normal, but the three-year-old had known morphological abnormalities. The spectral collection itself was from the calcified cartilage and subchondral bone.

Among the authors' findings were that bone microstructure and its variations within each horse group led to a lack of fracture resistance (that is, the bone was unable to cope with applied loads and forces). These variations, however, raised the question of the likelihood of fracture in individual horses due to earlier disadvantageous microstructural changes. Even though the classification via LDA provided a clear separation of the horses both by age and disease status, there were no clear-cut conclusions from the study.

1.6.4 Other Raman, Raman-related, and Raman-chemometric applications to various bone tissue specimens

As noted above in subsection 1.6.3, much of the Raman spectroscopy-led research that falls within medical and veterinary contexts has tended to focus on the inferences that could be made from changes to univariate spectral band parameters on bone quality. For example, with the aid of Raman spectroscopy, Khalid *et al*. (38) modelled staphylococcal osteomyelitis to determine changes that *in vitro Staphylococcus aureus* (*S. aureus*) infection might produce in human knee bone samples and the effect this infection would have on bone quality. *S. aureus* is the predominant (bacterial) cause of osteomyelitis, the inflammation and destruction of bone and bone marrow. Their investigation found that there was a marked reduction in the bone mineral quality and crystallinity parameters collected from the Raman spectral data from the *in vitro S. aureus*-infected human knee bone sections when compared to control bone sections. There was also altered collagen cross-linking of *in vitro S. aureus*-infected human knee bone sections. The group's results implied that early diagnosis and treatment of osteomyelitis might eventually be possible.

Down a similar line of investigation, de Souza *et al*. (39) exploited Raman spectroscopy to uncover osteoarthritis (OA)-related changes to collagen deposition and tissue remodelling in two well-established experimental rat knee OA models: collagenase-induced OA and treadmill exercise-induced OA. The experimental control group consisted of the preserved menisci-covered tibial joint. They observed that there was a distinct increase in the mineralisation and tissue remodelling-related parameters in both models when compared to the controls, pointing toward the successful induction of OA in the rats. Additionally, noticeably lower phenylalanine (Phe) content and higher crystallinity in the treadmill exercise-induced OA model than in the collagenase-induced OA model, connoted model-dependent OA pathogenesis. These main findings highlighted the potential for the detection and diagnosis of OA-associated cartilage, subchondral bone, and overall joint damage. Such

studies as these demonstrate molecular vibrational spectroscopy's capability in uncovering the differences in bone tissues, even using univariate statistics.

Bonifacio *et al.*'s characterisation of articular cartilage hyperspectral data from a porcine humeral-scapular joint bone section via the combined used of univariate and multivariate statistical techniques (PCA, partial least-squares regression (PLSR), and hierarchical and fuzzy *c*-means cluster analyses) in (40) was one example of tandem Raman-chemometric applications to bone tissue specimens in medical and veterinary contexts. Briefly – as its name indicates – cluster analysis groups spectral data into clusters, based on the level of similarity or difference between spectral attributes, and there are two broadly different types of cluster analysis (34). The outcome of the hierarchical clustering method is akin to a dendrogram or tree where smaller clusters of more similar data give way to larger clusters of more dissimilar data (34); these are distinct clusters (40). From either a randomly chosen or a user-specified initial number of centroids, the fuzzy *c*-means clustering method groups the data iteratively into clusters. The final clusters would be those where the data have partial membership in multiple clusters rather than explicit membership in a distinct cluster (these membership values are normalised to take on values between zero and one, and indicate the level of similarity between a datum and the cluster mean) (40). PCA brought out the overall compositional changes by differentiating between the chief biochemical components (collagenous and non-collagenous proteins, proteoglycans, and nucleic acids). The two cluster analytical techniques, together with the PLSR, enabled detection of changes to these components in each of the cartilage's three zones. From their results, there was again, the potential for further understanding of cartilage degradation processes.

Toledano *et al*. (41) did Raman-cluster analysis work on human postmenopausal fractured femoral trabecular bone. It was a proposal of protocol for modelling trabecular bone degradation, based on microstructural modifications inferred from the specimens' mineral and organic matrix's Raman peak parameter measurements. These measurements also generated false-colour cluster and single-point maps via classical least-squares analysis (CLS), hierarchical cluster analysis (HCA), and PCA images. Clustering information from these images highlighted the differences and similarities in the specimens' two main components' chemical compositions.

Other tandem applications to bone tissue and related specimens have not necessarily fallen into medical and veterinary contexts; many of these have tended to focus on bone microstructure as opposed to bone disease. There have been, for instance, applications from

the forensic sciences like McLaughlin and Lednev's proof-of-principle partial least squares-discriminant analysis (PLS-DA) of chicken, turkey, bovine, and porcine bone tissue specimens as a precursor to discriminating between human and non-human bone tissue specimens in (42). The first two components from the PLS-DA were able to separate the four species of origin, mainly based on contributions from the bone mineral and organic matrix. Likewise, Shimoyama *et al*.'s use of PCA and PLSR analysis (PLSR) to enable discrimination between two subspecies of African elephant, mammoth, hippopotamus, and sperm whale ivories in (43), and Brody *et al*.'s use of PCA and stepwise discriminant analysis to differentiate between ivories of six mammalian species in (44) were both also examples of species-of-origin discrimination of bone tissue. In (43), the first two PCs (accounting for about 84 % of the total spectral variation) were respectively capable of separating the two elephant subspecies' ivories based on band intensity changes from interactions between collagenous proteins, non-collagenous proteins, and water, and sorting the elephant ivories from the other ivories. The PLSR was a predictive tool for specific gravity – based on second derivative spectral data collected from a set of five other ivories – with a correlation coefficient of 0.960 and root mean-square-error-of-cross-validation (RMSECV) of 0.037. In (44), there were two rounds of PCA conducted before the discriminant analysis. The first PCA was of spectral data from African elephant, 'Asian' elephant, hippopotamus, mammoth, sperm whale, and walrus dentine; sperm whale cementum; bovine, porcine, and ovine bones; and netsuke (derivatives of elephant or mammoth ivory). The second PCA was just of spectral data from African elephant, 'Asian' elephant, and mammoth dentine, and netsuke. In both rounds of PCA, the first two PCs accounted for most of the total spectral variation from the two main components in the specimens (about 81 % for the first PCA, and 70.6 % for the second PCA), differentiating to an extent between the dentine, bone, and cementum. The discriminant analysis did not give 100 % classification for each species since there remained an overlap between groups.

Peak parameters – which provide information about bone compositional variation – correlate highly with bone's physical properties and how well (or otherwise) it might cope with applied loads and forces (28). Despite peak ratio data giving insight into how bone quality may be altered, univariate statistical techniques have their limitations. Buckley *et al*. noted in (45) that the more experimental factors there are, the more poorly univariate methods perform. Hence, their power is most apparent with a small total number of factors. The literature debates which spectral bands to use as part of some of these peak parameters: as an example, with the

mineral-to-matrix ratio, the ν_1 PO_4^{3-} and Amide I bands have commonly been used as representatives of the inorganic and organic components of bone (28, 46). Some groups have suggested that, instead, the use of either the ν_2 PO_4^{3-} or ν_4 PO_4^{3-} band with the Amide III band would better represent mineralisation (35). There is the added complexity of some spectral bands' dependence on bone tissue orientation that usually requires deconvolution and derivatisation of the spectral bands into their underlying peaks (17). Kazanci *et al*. have demonstrated the tissue orientation- and laser polarisability-dependence of the ν_1 PO_4^{3-} and Amide I bands in (47, 48). They highlighted the notion that the ν_1 PO_4^{3-}/Amide I ratio provided information about lamellar bone orientation, and the multiple vibration modes of the ν_2 PO_4^{3-}, ν_4 PO_4^{3-}, and Amide III bands meant that they were less susceptible to these two phenomena (47). Within the more readily available literature, if multivariate statistical techniques have been used, they have tended to be combined with univariate statistical techniques – probably to enable assessment of comparable information. Kerns *et al*. combined the peak ratio data with their PCA and PCA-LDA results collated from the subchondral bone spectra of human OA and non-OA tibial specimens in (49), alongside peripheral quantitative computed tomography (pQCT) and biochemical analysis results. Among their findings were a difference between the OA and non-OA specimens – with there being higher mineralisation in the OA specimens (via the phosphate-to-Amide I ratio, and a bioapatite-to-collagen ratio, which was based on the phosphate-to-hydroxyproline ratio). The first three PCs from the PCA found the phosphate and Amide I bands to be the primary contributors to the differences between the OA and non-OA specimens. The PCA-LDA results highlighted the phosphate, proline, hydroxyproline, and Amide III bands as being the discriminators between the OA and non-OA specimens. The majority of these discriminators being organic matrix-associated spectral bands, coupled with the hypermineralisation of the OA specimens, implied that there was a change in the organic matrix composition before subchondral bone thickening. That is, any changes to the collagenous component would alter the mineral component in the progression of OA.

To reiterate, the highly overlapping spectral data collected from inherently complex bone tissue mean that univariate analytical methods may not necessarily give a proper narrative of any alteration, due to the multivariate dependencies of the spectral measurements taken. Univariate methods may give specific details; multivariate methods may clarify how these many specific details simultaneously fit into a bigger picture. From the PCA results that will be presented later in Chapters 3 and 4, both the PC loadings and PC scores' distributions detected localised differences within the bone tissue specimens. These differences could

potentially be highlighting areas of hypermineralisation or some organic matrix degeneration within fracture predilection sites or OA-induced sites that may well be indicative of the early development of these orthopaedic diseases. That may be how information from a multivariate method might be comparable to that gleaned from the changes to univariate band ratios commonly presented in the literature.

Each multivariate statistical technique also has its advantages and limitations and, in some contexts, combining techniques may overcome some of these inherent disadvantages and give more nuanced detection of variations in different regions of the bone tissue's underlying microstructure. Bonifacio *et al*. found the hierarchical cluster analysis was better at separating areas of more distinct biochemical constituents, and the fuzzy *c*-means cluster analysis was better at differentiating between areas of more continuous changes in (40). In other contexts, sometimes following the Occam's razor principle to enable sorting may allow better detection of microstructural variation.

The work contained herein shares with these other multivariate studies the fact that they have been capable of determining – to various degrees – differences and similarities in the spectral information.

1.7 Aims of this thesis

IR spectroscopy and chemometrics have successfully detected differences in equine bone microstructure. It was proposed that Raman microspectroscopy and chemometrics would also be capable of identifying changes in the chemical composition of equine bone. Some of these changes might be indicative of abnormalities that could very well be precursors to bone fracture.

The primary goals of this research were to a) conduct an analogous Raman microspectroscopy study to that done by Nicholson *et al*. on a different set of embedded equine Mt3 bone specimens, as well as b) as-yet uncharacterised (by Raman microspectroscopic methods, at least) fresh equine carpal joint bone specimens, and c) use multivariate statistical analytical techniques – namely PCA or LDA – to provide a classification model for the early stages of equine bone disease. The scope of this work was limited to the exploration of Raman microspectroscopy's ability to determine – rather than quantify – differences.

In contrast to Nicholson *et al*.'s scrutiny of a smaller number of specimens, work carried out as part of this study was on a larger number of specimens. All specimen samples had known

histories, but to minimise bias only details enough to enable Raman spectral data collection were shared.

Chapter 2 – Materials & Methods

2.1 Sample preparation

2.1.1 Equine third metatarsal (Mt3) bone specimens

The third metatarsal (Mt3) bone specimens were from 30 thoroughbred racehorses of three different age groups (ten from newborn foals, ten from five-month-old foals, and ten from two-year-old horses). These samples were supplied prepared by Catherine Nicholson (Massey University), having been embedded in a polymethylmethacrylate (PMMA) resin due to prior analysis with specular reflectance Fourier-Transform infrared (SR FT-IR) spectroscopy. Figure 2.1 is a montage of photographs of the Mt3 bone sections from three of the thoroughbred racehorses (one from each age group).

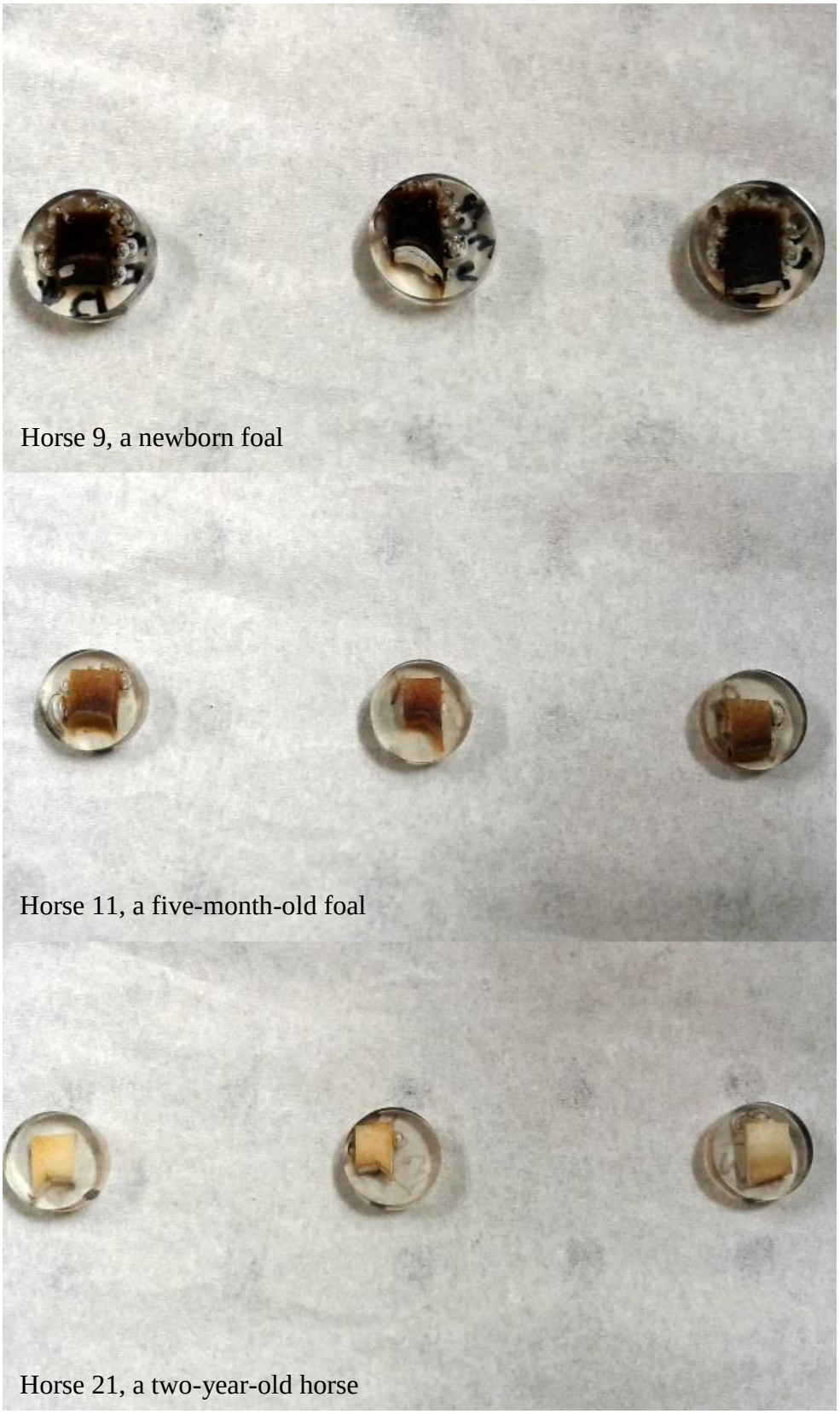

Figure 2.1 Photographs of Mt3 bone specimens from three of the 30 thoroughbred racehorses (one from each age group): horses 9, 11, and 21.
L to R for each horse: lateral parasagittal groove (section 1, prone to fracture), medial parasagittal groove (section 2, prone to fracture), medial condylar surface ('control' site). Image by author.

2.1.2 Equine carpal joint bone specimens

The carpal joint bone specimens were from eight thoroughbred racehorses of two different age groups (seven from two-year-old horses, and one from a three-year-old horse). These samples were supplied prepared by Luca Panizzi (School of Veterinary Sciences (SVS), formerly the Institute of Veterinary, Animal, and Biomedical Sciences (IVABS), Massey University), having been kept in storage at approximately -80 °C until ready for analysis. Figure 2.2 is a montage of photographs of the carpal joint bone sections from one of the thoroughbred racehorses (one of the two-year-olds).

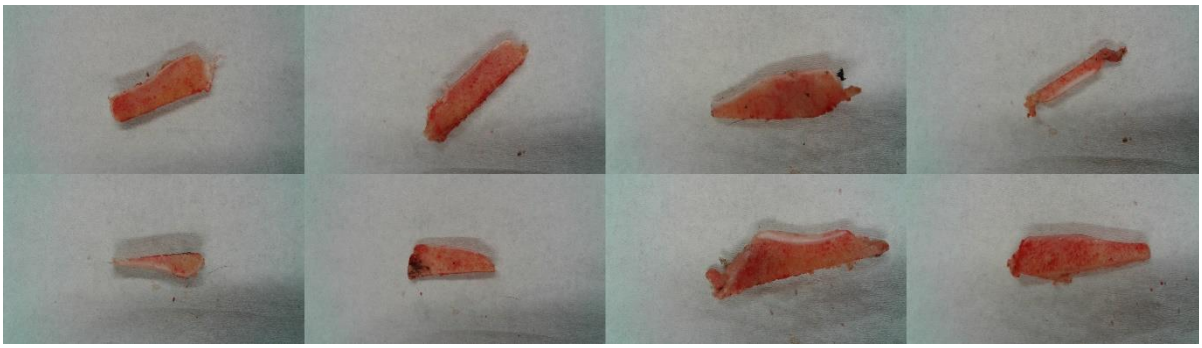

Figure 2.2 Photographs of carpal joint bone sections from horse 17, one of the seven two-year-old thoroughbred racehorses. Image by author.
The upper row consists of carpal joint bone sections from the left leg, whilst the lower row is of those from the right leg.
L to R for both rows: third carpal bone (C_3); fourth carpal bone (C_4); and radial carpal bones (C_R: upper part of the radial carpal bone, R2 and lower part of the radial carpal bone, R3).

2.1.3 Preparation of the third metatarsal (Mt3) bone sections and polymer-embedded polished bone samples

The distal Mt3 bone sections had been prepared as outlined in Nicholson *et al*. (37): a 2 mm thick palmar (posterior) bone slice was cut out at approximately 30° to the bone's long axis. Three 5 mm by 5 mm sections were cut out of each palmar slice to encompass the lateral and medial parasagittal grooves (both known to be fractures sites) and the medial condylar surface ('control' site). In total, there were 90 bone sections (three for each horse).

These bone sections had been previously analysed with backscattered electron microscopy (BSEM) and SR FT-IR spectroscopy. They would have been dehydrated in absolute ethanol before being embedded in PMMA and eventually polished with increasingly fine grades of carborundum paper, 3 μm Al_2O_3 powder and 0.3 μm Al_2O_3 powder.

2.1.4 Preparation of carpal joint bone sections

The carpal joint bone sections had been cut out in slices of approximately 3 – 5 mm thickness from the dorsal aspect and along the long axis of the third, fourth, and radial carpal bones (C_3, C_4, and C_R bones, respectively) to encompass just the articular cartilage and subchondral bone layers. Each horse had had one forelimb operated whereby a fragment or 'chip' had been surgically created. The opposite forelimb was the 'control' limb. That is, one leg was healthy; the other was 'fractured'. Each of the horses had eight bone sections (four from the left leg, four from the right leg). In total, there were 32 bone sections.

2.2 Epi-illumination for optical imaging of third metatarsal (Mt3) bone sections

Visualisation of collection sites was performed just prior to Raman microspectroscopy to ensure (or at the very least, indicate) that an appropriate sampling point had been chosen, as, by nature, the samples were opaque. With the polished side face-down towards the objective, imaging was carried out with a puA1280-54uc digital camera (Basler, Ahrensburg, Germany) via the eyepiece/camera setting of a home-built Raman microscope system and captured with pylon Viewer software (version 5.0.0.6150, Basler, Ahrensburg, Germany). A white LED lamp cage-mounted system (ThorLabs, Newton, New Jersey) provided the collimated illumination source.

2.2.1 No epi-illuminated optical images of carpal joint bone sections

No epi-illuminated optical images of the carpal joint bone sections were taken, as in their natural state they did not have the smooth surface often required for visualisation. Therefore, determination of appropriate sample points before data collection was by the preview of Raman spectra within the LightField software (software details in the next section). Spectra from the articular cartilage (AC) would be expected to show more signals from the collagenous component. In comparison, spectra from the subchondral bone (SCB) were expected to show more signals from the mineralised component. Due to the randomness of the spectral sampling sites, it could not be disregarded that some of the Raman spectral data collected from the AC may have consisted of a mixture of non-mineralised hyaline articular cartilage (HAC) and mineralised articular calcified cartilage (ACC).

2.3 Raman microspectroscopy

2.3.1 Raman microscope system

The home-built Raman microscope system was set up for use with four different monochromatic lasers: 488 nm, 532 nm, 633 nm, and 785 nm. As the samples were of a biological nature, a 785 nm, 50 mW single-mode, 50 μm diameter fibre-coupled diode laser (Warsash Scientific, Sydney, Australia) served as the near-infrared (NIR) monochromatic excitation source, to reduce the fluorescence that is inherent to biological samples.

As mentioned in the preceding sections, an epi-illumination system was necessary for the visualisation of the opaque bone sections under the microscope objective. An Olympus IX-70 inverted fluorescence microscope body fitted with a 10× magnification objective (NA 0.25, Edmund Optics, Singapore) and a custom-manufactured 785 nm filter (Iridian Technologies, Ottawa, Ontario, Canada) passed the scattered light through a series of Volume Bragg Grating (VBG) interference (notch) filters (OptiGrate, Oviedo, Florida) to collect in a LS 785 spectrograph (with Acton 2500i grating for the visible region of the electromagnetic spectrum) (Princeton Instruments, Trenton, New Jersey).

The detector used was a liquid nitrogen-cooled Princeton Instruments PIXIS 400 charge-coupled device (CCD). Slit width allowing light to the spectrograph and detector was set to 50 μm. LightField software (version 6.0.4.1611, Princeton Instruments, Trenton, New Jersey) was used to obtain the Raman measurements.

Figures 2.3 to 2.6 are annotated photographs of the set-up of the Raman microscope system, each highlighting different components:

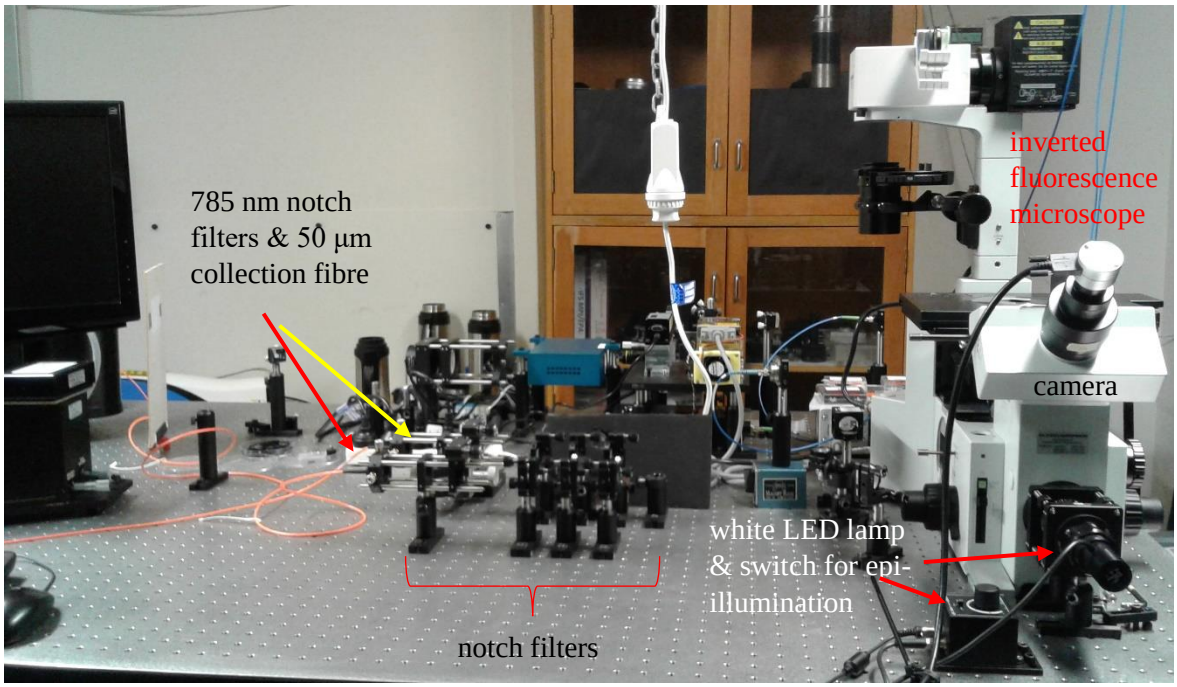

Figure 2.3 Photograph of the home-built Raman microscope system, showing part of the spectral collection side of the set-up. Image by author.

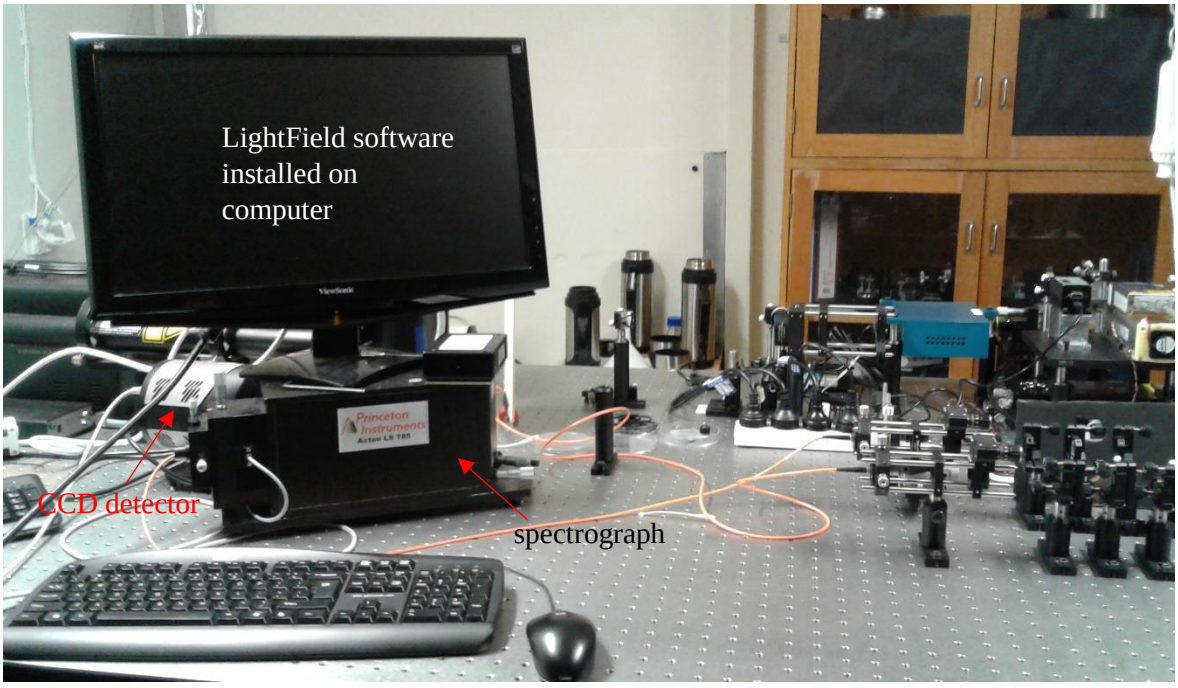

Figure 2.4 Photograph of the home-built Raman microscope, showing the other part of the spectral collection set-up. Image by author.

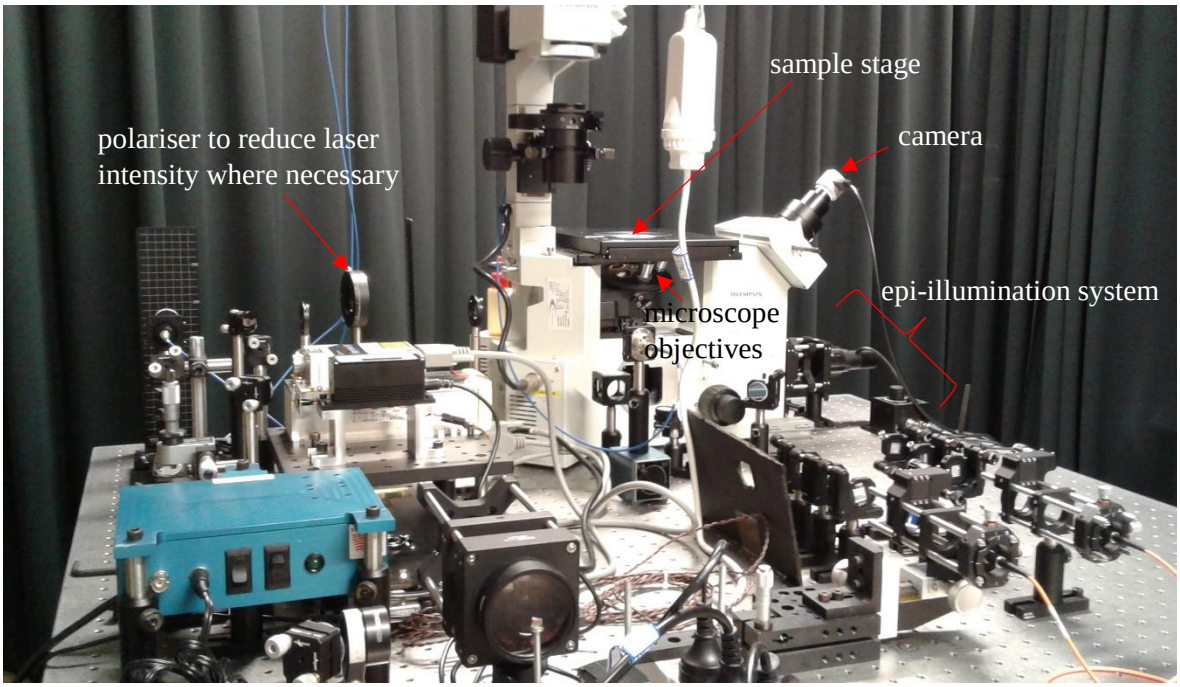

Figure 2.5 Photograph of the home-built Raman microscope system from another angle. Image by author.

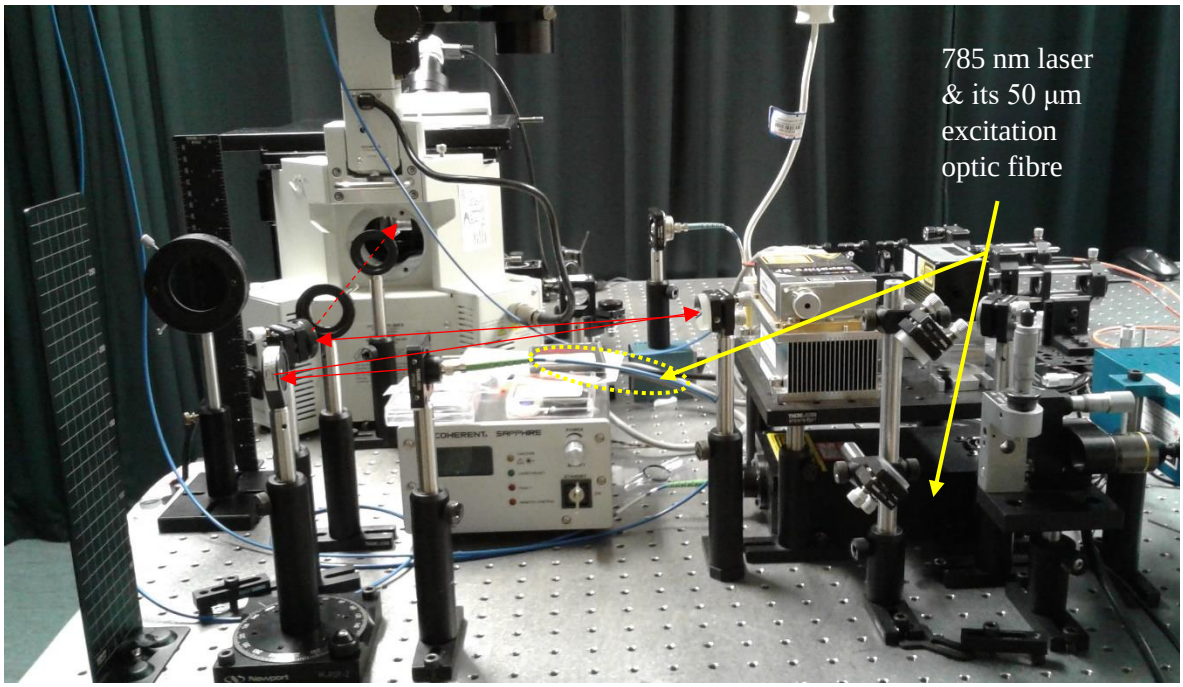

Figure 2.6 Photograph of the home-built Raman microscope system; the red arrows show the path of the 785 nm laser from the 50 µm diameter excitation fibre optic into the inverted microscope. Image by author.

2.3.2 Raman spectral data collection

Each spectrum of Raman scattering would have been collected from an area determined by the laser spot size; this was approximately 10 µm under the 10× objective – a small area at the centre of the microscope camera's 50 µm by 50 µm cross-hairs.

Spectra were collected over the wavelength range -500 – 2000 cm^{-1}, with a spectral resolution of approximately 5 cm^{-1}. Subsequent spectral pre-processing and statistical analysis focused on the wavelength range 300 – 2000 cm^{-1}. The LightField software enabled online corrections such as background subtraction, cosmic ray removal, and CCD hardware orientation to be applied to each Raman spectrum before file exportation in .csv format.

2.3.2.1 Raman spectral data collection for the third metatarsal (Mt3) bone sections

As an aside, Nicholson *et al*. (37) remarked from work done by Firth *et al*. (50) that fractures of the parasagittal grooves of the equine Mt3 bone tend to initiate within the articular calcified cartilage (ACC) before moving into the subchondral bone and beyond. Considering each layer of bone sections from the fracture-prone lateral and medial parasagittal grooves, and the medial condylar surface, in turn, might shed light on any sites of possible microstructural abnormalities.

Initially the plan had been to collect a total of 20 spectra in each of the three 5 mm by 5 mm Mt3 bone sections, to be comprised of (i) 16 in the articular cartilage (AC), and (ii) four in the subchondral bone (SCB), with a collection time of 2.5 minutes. It was this plan that was in mind when the spectral data collection began with bone sections from the two-year-old horses. Due, however, to time pressures, it was decided to collect ten spectra – six from the AC, four from the SCB – for the remaining Mt3 bone sections from the other two age groups, with a collection time of 1 minute. This period was deemed the minimum for still obtaining reasonably high signal-to-noise in the spectral data. Altogether, 60 spectra were collected for each of the two-year-old horses' bone sections, while 30 spectra were obtained for each of the five-month-old foals' and newborn foals' bone sections.

Each of the spectra from the two-year-olds and reference materials (sintered bone, hydroxy carbonated apatite, defatted bone, and pure PMMA; also kindly provided by Nicholson) represented an average of 15 accumulations; each of the spectra from the other horses represented an average of six accumulations.

On a day-to-day basis, laser power would fluctuate between approximately 7 and 17 mW. This variability in laser power may have contributed to the changeable signals' strengths observed in the Raman spectral data; this would also have precluded any calculation of absolute intensities of the Raman spectral bands.

2.3.2.2 Raman spectral data collection for the carpal joint bone sections

A total of four spectra were collected in each of the carpal joint bone sections – two from the AC, two from the SCB – with the same 1-minute integration collection time as for two-thirds of the Mt3 bone sections. Altogether, 32 spectra were collected for each horse.

Decreased laser power was used to minimise the potential burning of the fresh-frozen carpal joint bone sections (fluctuating between approximately 3 and 7.5 mW). There was a trade-off between the laser power and the signal-to-noise ratio (lower laser power, lower signal-to-noise ratio). Despite the shorter integration time, the signal-to-noise ratio for the Raman spectral data from the carpal joint bone sections was surprisingly reasonable, as many of the expected bone spectral signatures were still detectable.

2.4 Pre-processing of Raman spectral data

The fluorescent nature of these samples meant that their Raman spectra had non-linear baselines; these had to be corrected for before any statistical analysis could be carried out, to minimise any extraneous sources of variation. Pre-processing of the Raman spectral data was carried out using the *ChemoSpec* (B. A. Hanson, version 4.4.17), *R.utils* (H. Bengtsson, version 2.5.0), *baseline* (K. H. Liland & B.- H. Mevik, version 1.2-1), *wavelets* (E. Aldrich, version 0.3-0), and *signal* packages within the programme RStudio (The R Foundation for Statistical Computing/R Core Team, version 1.0.136). *ChemoSpec* and *R.utils* enabled the input of the spectral data .csv files into RStudio, while *baseline*, *wavelets*, and *signal* implemented the actual baseline correction and smoothing algorithms. Exemplar R code used to enable the data pre-processing and multivariate statistical analysis can be viewed in Appendix A. Plots of the raw Raman spectra can be viewed in Appendix B.

2.4.1 Pre-processing of the third metatarsal (Mt3) bone sections' Raman spectral data

After much trial-and-error, it became apparent that the most useful algorithm for baseline correction of the Raman spectra collected from the equine Mt3 bone sections looked to be asymmetric least squares (ALS) coupled with Savitzky-Golay smoothing. None of the baseline correction algorithms available within RStudio's *baseline* literature (including iterative restricted least squares, low-pass Fast Fourier Transform filtering, modified

polynomial fit, and simultaneous peak detection and baseline correction) did a perfect job of removing the fluorescence background. Still, gradual fine-tuning of the ALS baseline correction algorithm parameters eventually yielded corrected spectra. ALS is an iterative baseline correction algorithm that applies a second derivative constrained weighted regression to the spectral data matrix. For more details about the original ALS algorithm, refer to (51).

Before baseline calculation and correction, single normal variate (SNV) scaling was done on the truncated and transposed Raman spectral data matrix. The spectral data matrix was truncated to remove the lower frequencies, which were not part of the spectral region of interest, and transposed, so each spectrum occupied a row to be more amenable to later statistical analysis. SNV scaling uses the total intensities within each Raman spectrum to scale each of the raw spectra. The data are then scaled by standard deviation and mean-centred. Mean-centring is required in PCA, as it is not scale-invariant. Scaling to unit standard deviation is useful to avoid samples that are at the "edges" of the data set dominating the principal components. Creation or calculation of the baseline and baseline correction were then applied to the SNV-scaled data frame with the ALS method (with the following parameters: second derivative constraint, $\lambda = 4$; weighting of positive residuals, $p = 0.001$; maximum number of iterations, $maxit = 10$). Lastly, Savitzky-Golay derivation and smoothing were utilised on the baseline-corrected Raman spectral data matrix (with these parameters: 1 = performed on matrix rows (again, one spectrum per row); filter length, $n = 13$, m^{th} derivative of the filter coefficients, $m = 0$).

2.4.2 Pre-processing of the carpal joint bone sections' Raman spectral data

Pre-processing of Raman spectra from the equine carpal joint bone sections required a slightly different approach since the make-up of this data set was more complicated than that of the Mt3 bone data set. The fact that information about each bone section was also dependent on the forelimb (left or right) and the condition of that forelimb ('control' or 'chip') from which it was excised meant that set-up would need to take into account differencing of spectra.

The carpal joint Raman spectral data set was initially subdivided into two to make later aggregation, differencing, and PCA easier. One subset consisted solely of the 'chip' spectral data from all eight horses' carpal joint bone sections, and the other contained the equivalent 'control' spectral data.

As with spectral data from the Mt3 bone sections, finding an acceptable baseline correction algorithm was a trial-and-error procedure and, again, none of those available was perfect; the ALS method seemed the most suitable for this data set, too (parameters: second derivative constraint, $\lambda = 4$; weighting of positive residuals, $p = 0.002$; maximum number of iterations, *maxit* = 10). The same Savitzky-Golay filter parameters as for the Mt3 bone sections' data were used (that is, 1, $n = 13$, $m = 0$).

2.5 Multivariate statistical analysis of Raman spectral data

2.5.1 Multivariate statistical analysis for the third metatarsal (Mt3) bone sections

PCA of the covariance matrix (the Raman spectral data matrix multiplied by its transpose) from the Savitzky-Golay-smoothed, ALS baseline-adjusted Raman spectral data from the Mt3 bone sections was also carried out using the programme RStudio.

Spectral data collected from the Mt3 bone sections were aggregated by computing the means to remove some of the influence of each horse. This so-called 'horse effect' would control for horse-related bias and remove between-object variability to obtain more comprehensible results later. An observed spectrum can be thought of as consisting of various factors or influences; in this context, it would have been those of each horse, section, layer, and any errors. Since the data set was "balanced", with the same number of observations for each combination of horse, section, and layer, an estimate of the 'horse' influence could be made by averaging over the other influences. This balance was required so that the 'section' and 'layer' effects averaged out in the estimated 'horse' effects; subtracting the horse-level averages removed its effect from the remainder of the spectral data structure. This subtraction meant that spectral variations due to the influence of section and layer should have been more apparent.

The collapsed spectral data generated a summary of PCA results, as well as a scree plot, loadings and scores matrices, loadings plots, and scores plots for pairs of the first six PCs labelled by layer, section, and combinations of the two.

2.5.2 Multivariate statistical analysis for the carpal joint bone sections

As had been done for the Mt3 bone specimens' Raman spectral data, horse-related bias from each of the carpal joint spectral data subsets needed to be minimised so that any effects due to section and layer could be more easily seen. Each subset generated a summary of PCA results, as well as the expected scree plot, loadings and scores matrices, loadings plots, and scores plots for pairs of the first six PCs labelled by layer, section, and combinations of the two. Each data subset also produced two PCA result sub-subsets: PCA of spectral data solely from the articular cartilage (AC), and PCA of spectral data from the subchondral bone (SCB). Plots of the PCs were labelled by section, condition, and combinations of the two.

Differenced spectra were also considered to find out if contributing chemical functionalities for the 'condition' of each spectral data subset would be more recognisable. They were created by subtracting the aggregated 'control' data from the aggregated 'chip' data. These differenced spectral data also generated the expected PCA results (such as the summary, scree plot, and loadings and scores matrices).

Another two sets of PCA results were generated from the carpal joint bone sections' spectral data. For the first set of results, the data were created by stacking together the aggregated 'chip' and 'control' data. The usual PCA results were generated (PCA summary, scree plot, loadings and scores matrices, and loadings and scores plots for pairs of the first six PCs labelled by section, layer, and condition, and combinations of the three). For the second set of PCA results, the initial data were formed by again stacking together the aggregated 'chip' and 'control' data. This time, however, mean spectral intensities' values for each combination of horse, section, and layer were computed (to remove any replicate effects). These means were then subtracted from the individual values in the original 'stacked' data (again, to remove their respective 'averages', with the expectation that only the effect of 'condition' might linger). PCA was performed on these mean-adjusted, stacked, and aggregated data; yet again, the usual PCA results were generated (such as the PCA summary, scree plot, and loadings and scores matrices; plots of the PCs were labelled by section, layer, and combinations of the three).

Two PCA result subsets were also produced from these mean-adjusted data: PCA of spectral data solely from the articular cartilage (AC), and PCA of spectral data from the subchondral

bone (SCB). Plots of the PCs were labelled by section, condition, and combinations of the two.

Chapter 3 – Results & Discussion: Embedded samples from the equine third metatarsal (Mt3) bone specimens

3.0 Introduction

As outlined in Chapter 1, a lack of literature exists on tandem Raman microspectroscopy-chemometrics applications to equine bone specimens (particularly that of principal component analysis) for categorisation. This chapter should aid in starting to fill the gap.
This chapter covers the outcome of multivariate statistical analysis performed on Raman spectral data collected from the embedded equine third metatarsal bone specimens, preceded by exemplar epi-illuminated images and pre-processed Raman spectral plots from three of the horses (one from each age group). It also includes two tables – one of expected Raman spectral assignments associated with bone specimens and another summarising the data groupings – before moving into the actual results of the principal component analysis of the data set. Discussion of the results and the limitations of this aspect of the body of work are also presented.

3.1 Epi-illuminated optical images of equine third metatarsal (Mt3) bone sections

Many of the epi-illuminated optical images of the equine third metatarsal (Mt3) bone sections were not completely clear. The basic outline of the osteon within the subchondral bone, however, was readily observable (more so in the five-month-old foals and two-year-old horses). The greenish hue in the images was the result of viewing each of the images through the 785 nm Raman edge filter (which blocked red light but passed most of the green; the low contrast was due to the filter simply blocking a reasonable fraction of the reflected light). Spectroscopic sampling areas were from a small portion – approximately 10 μm under the 10× objective – of the area at the centre of the microscope camera cross-hairs.

Owing to its location in the distal part of long bones, the articular cartilage (AC) cushions joints to minimise friction in the transfer of applied loads and forces to the subchondral bone (SCB). The SCB, conversely, provides structural support for the AC and acts as a shock absorber. To restate from subsection 2.3.2.1, fractures of the parasagittal grooves of the equine Mt3 bone typically work their way inwards from the articular cartilage (37, 50). As the

statistical analysis results later in subsection 3.3.2 attempt to show, considering each layer of bone sections from the fracture-prone lateral and medial parasagittal grooves, and the medial condylar surface, in turn, might shed light on any sites of possible microstructural abnormalities.

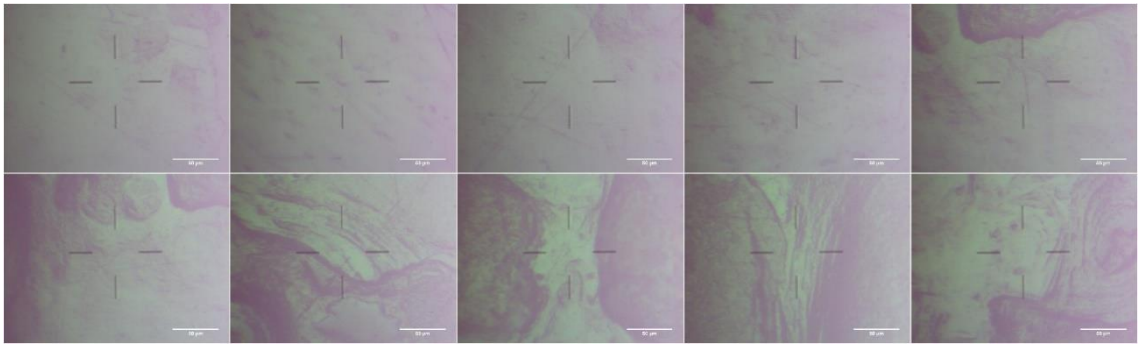

Figure 3.1 Montage of epi-illuminated optical images from horse 1, one of the newborn foals. The upper row and first image in the lower row are images of the articular cartilage (AC). The final four images are of sites within the subchondral bone (SCB). The wavier, more open appearance of cartilage in bone typical of younger individuals and in more recently formed bone is evident in the AC. Scale bar length = 50 μm.
Image by author.

Figure 3.1 is a montage of the epi-illuminated optical images from horse 1, one of the newborn foals. Though not easily visible, the scale bars in each image represent a length of 50 μm. The upper row and the first image in the lower row are images of the articular cartilage (AC), and the final four images in the lower row are of sites within the subchondral bone (SCB). The images from the AC hint at the generally wavier and more open appearance of typical of bone cartilage from younger individuals and in immature, more recently formed bone. This more recently formed bone would have minimal, if any, mineralisation.
As another aside, due to the difficulty in differentiating between the non-mineralised and mineralised sublayers of the AC, it could not be discounted that many of both the optical images and the Raman spectral data collected from the AC may have consisted of a mixture of non-mineralised hyaline articular cartilage (HAC) and mineralised articular calcified cartilage (ACC).

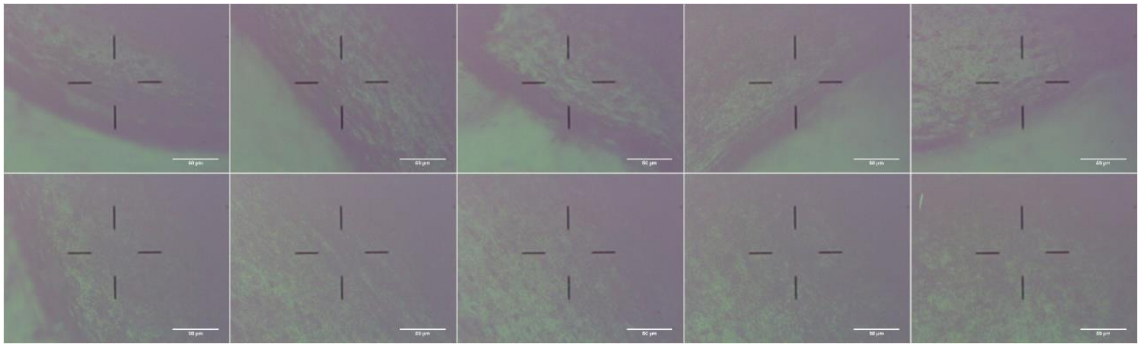

Figure 3.2 Montage of epi-illuminated optical images from horse 11, one of the five-month-old foals. The upper row and first image in the lower row are images of the AC. The final four images are of sites within the SCB. Both the AC and the SCB look denser, compared to specimens from a younger foal such as horse 1. Scale bar length = 50 μm.
Image by author.

Figure 3.2 is a montage of the epi-illuminated optical images from horse 11, one of the five-month-old foals. As with Figure 3.1 previously, the first six images are of sites within the AC, and the last four are of the SCB. Both the AC and SCB look denser when compared to that from a younger horse such as horse 1. This denser appearance is due to the continued development of the equine skeletal system, notably the progressive mineralisation of the AC.

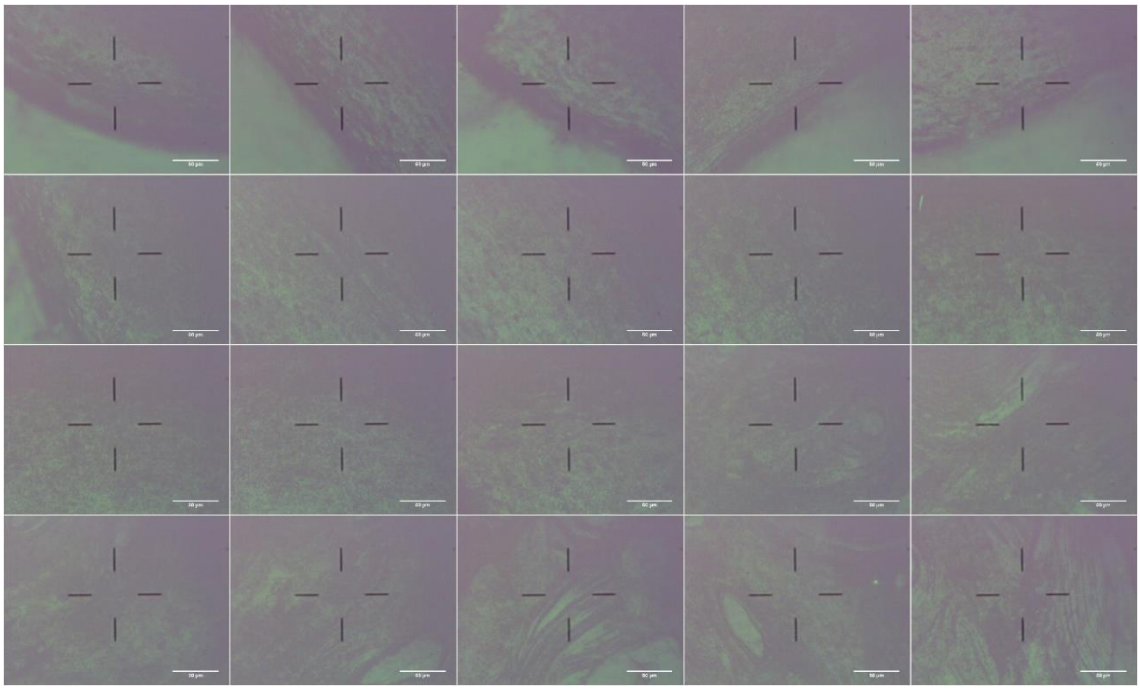

Figure 3.3 Montage of epi-illuminated optical images from horse 21, one of the two-year-old horses. The upper two rows and first three images in the third row are of sites in the AC. The last seven are of sites within the SCB. The images from the two-year-old horses probably best depict the denser appearance of the SCB, with its highly lamellar structure.
Image by author.

Figure 3.3 is a montage of the epi-illuminated optical images from horse 21, one of the two-year-old horses. The two upper rows and the first three images in the third row are of sites in the AC, and the last seven are of sites within the SCB. The images from the two-year-old

horses probably best depict the denser appearance of the SCB with its highly lamellar structure. The apparent lamellar structure is indicative of the remodelling process (bone formation and resorption) typical of bone in response to applied loads and forces.

3.2 Raman spectral data from the equine third metatarsal (Mt3) bone sections

The following information in this section is provided to help with understanding the loadings plots as part of the principal component analysis (PCA) results later in subsection 3.3.2.

3.2.1 Typical Raman spectral assignments of bone

Bands from several different functionalities are expected to fall into certain ranges, as outlined in Table 3.1:

Table 3.1 Some of the typical functionalities that tend to appear roughly in the 400 – 1800 cm^{-1} region of a Raman spectrum of bone (adapted from (52, 53)).

Vibration	Wavenumbers (cm^{-1})	Raman intensity[1]
ν_2 PO_4^{3-}	~ 422 – 454	m, sh
ν_4 PO_4^{3-}	~ 578 – 617	m, sh
hydroxyproline	~ 855 – 876	w, sh
ν_1 PO_4^{3-}	~ 857 – 962	vs
proline	~ 921	sh
phenylalanine (Phe)	~ 1004	w
ν_3 PO_4^{3-}	~ 1006 – 1055	sh
ν_1 CO_3^{2-}/ν_3 PO_4^{3-}	~ 1065 – 1071	sh
Amide III	~ 1242 – 1340	br
$\delta(CH_2)$, scissoring/deformation	~ 1447 – 1452	m
Amide II	~ 1540 – 1580	w, br
Amide I	~ 1595 – 1700	br

[1] The intensities of the peaks or bands that appear in a Raman spectrum tend to be characterised as any of the following: weak (w), medium (m), strong (s), very strong (vs), shoulder (sh), broad (br).

Raman signals from inorganic functionalities (phosphate (PO_4^{3-}) and carbonate (CO_3^{2-})) tend to be sharper than those from organic functionalities are (such as proline, hydroxyproline, phenylalanine (Phe), lipids, and Amides I, II, and III; cf. IR spectra where the opposite is true). The Amide III band from proteins tends to be more complicated because it involves several bonds. The Amide I band (which appears anywhere between about 1595 and 1700 cm^{-1}) is due to the carbonyl (C=O) stretch.

3.2.2 Raman spectral data from the reference materials

Figure 3.4 shows the processed Raman spectral data from the reference materials: defatted bone, hydroxy carbonate apatite, the embedding material polymethylmethacrylate (PMMA), and sintered bone. Defatted bone included collagen and apatite; the sintered bone was 'pure' mineral. These spectra from the defatted bone, hydroxy carbonate apatite, and sintered bone were coloured turquoise, violet-red, and blue, respectively. As can be seen in the figure, PMMA's Raman spectral bands – in black – tended to intersect with many of those from bone components, highlighting the need for a quantitative method like chemometrics to extract the subtle variations in bone microstructure that would not be readily discernible by merely relying on qualitative assignments.

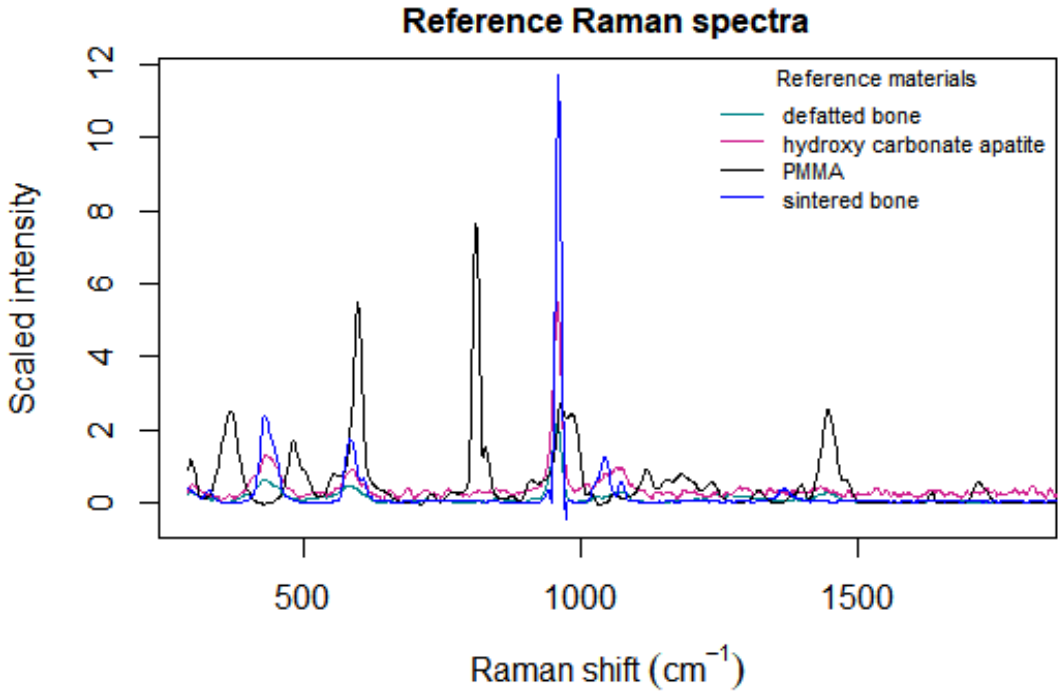

Figure 3.4 Processed Raman spectra of the reference materials: defatted bone (included collagen and apatite; turquoise), hydroxy carbonate apatite (violet-red), polymethylmethacrylate (PMMA; black), and sintered bone ('pure' mineral; blue).

3.2.3 Raman spectral data from the equine third metatarsal (Mt3) bone sections

Figures 3.5 to 3.7 are a few examples of the processed Raman spectral data from the Mt3 bone sections – one from a horse in each of the three age groups (horses 1, 11, and 21). The SCB spectra are shown as their band assignments were less ambiguous than the AC spectra, which tended to have bands that appeared to have been overlaid with those from the embedding material, PMMA.

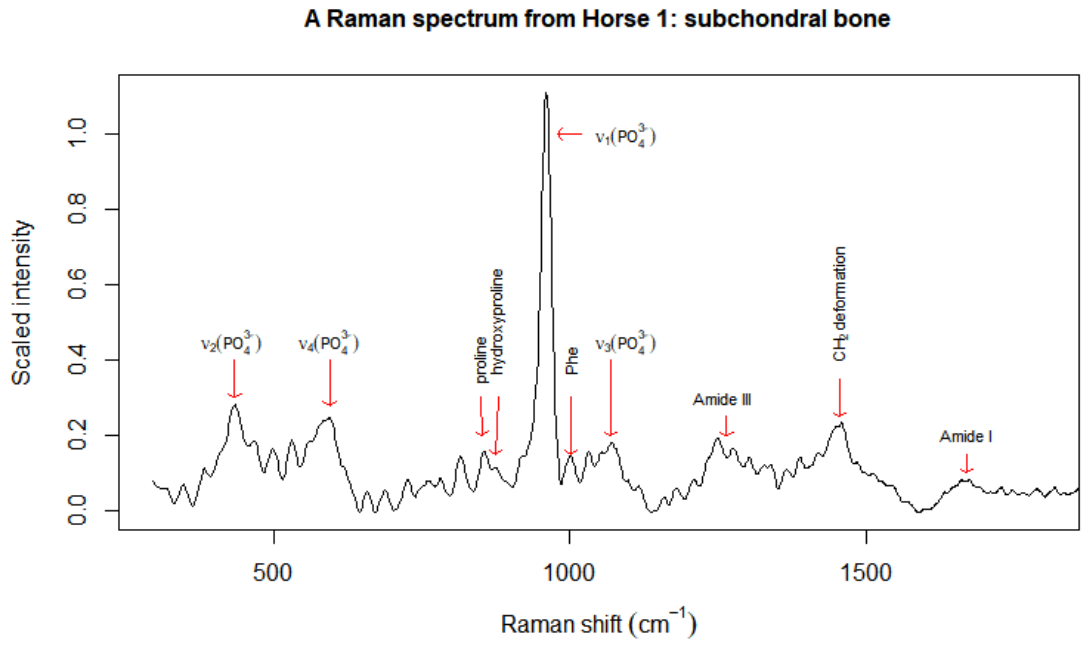

Figure 3.5 A Raman spectrum of the subchondral bone (SCB) from horse 1, one of the newborn foals. Typical band assignments from the mineral and matrix components are highlighted.

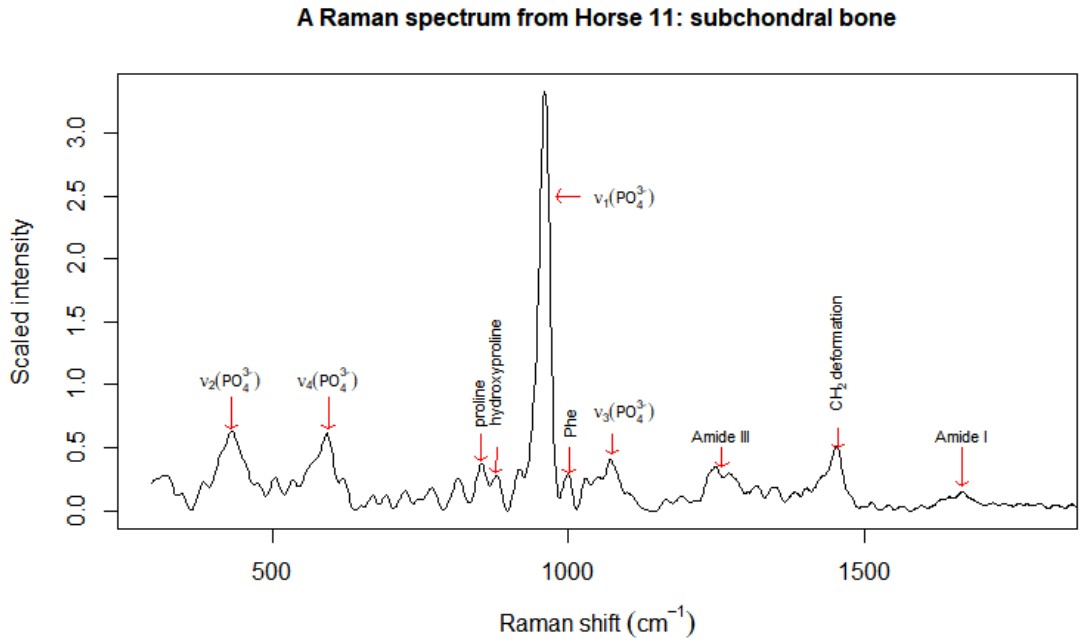

Figure 3.6 A Raman spectrum of the SCB from horse 11, one of the five-month-old foals.

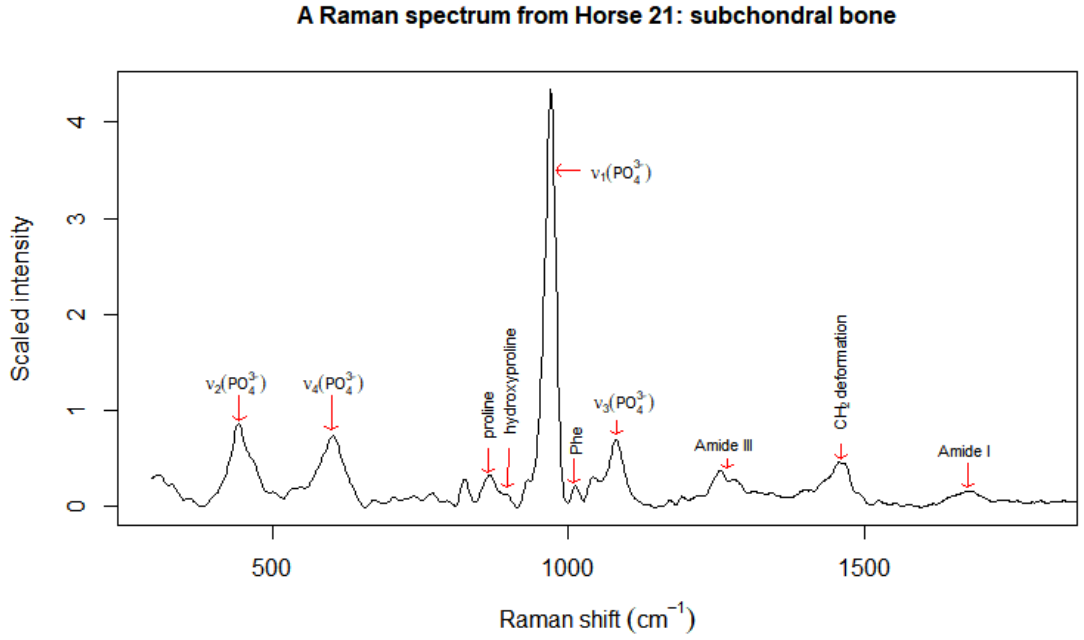

Figure 3.7 A Raman spectrum of the SCB from horse 21, one of the two-year-old horses.

3.3 Multivariate statistical analytical results

3.3.1 Summary table

Table 3.2 below outlines the Raman spectral data groupings that follow in subsection 3.3.2:

Figure	Spectral data grouping
3.8	scree plot of aggregated spectral data from equine Mt3 bone sections
3.9	first six PCs' loadings plots from aggregated spectral data
3.10	PC4 loadings plot from aggregated spectral data
3.11	PC2 vs PC1 scores plot from aggregated spectral data, by section and layer
3.12	montage of PC2 vs PC1 scores plot from aggregated spectral data, by section (columns) and layer (rows)
3.13	upper panel of PC2 vs PC1 scores plot from aggregated spectral data: AC layer of all three equine Mt3 bone sections
3.14	lower panel of PC2 vs PC1 scores plot from aggregated spectral data: SCB layer of all three equine Mt3 bone sections
3.15	PC3 vs PC1 scores plot from aggregated spectral data, by section and layer
3.16	PC3 vs PC2 scores plot from aggregated spectral data, by section and layer
3.17	paired scores plot matrix for the first six PCs from aggregated spectral data, by section and layer

3.3.2 PCA

Following are results from the PCA of the aggregated Raman spectral data from the equine Mt3 bone sections. The spectral data were average by horse (again, having incorporated extra details about the horses). These average spectra were subtracted from the data to remove a so-

called fixed 'horse effect' (containing horse-related bias and between-object variability). Its removal should have made within-object variability more understandable (refer to subsection 2.5.1, if necessary). Figure 3.8 shows the scree plot of the variances of the first 10 PCs. The first six PCs accounted for approximately 97.2 % of the total spectral variation, rapidly tailing off after PC3. PC1 had the greatest variance, explaining about 64.9 % of the total variation while PCs 2 through six explained approximately 17.8 %, 10.8 %, 2.3 %, 1.3%, and 0.824 %, respectively.

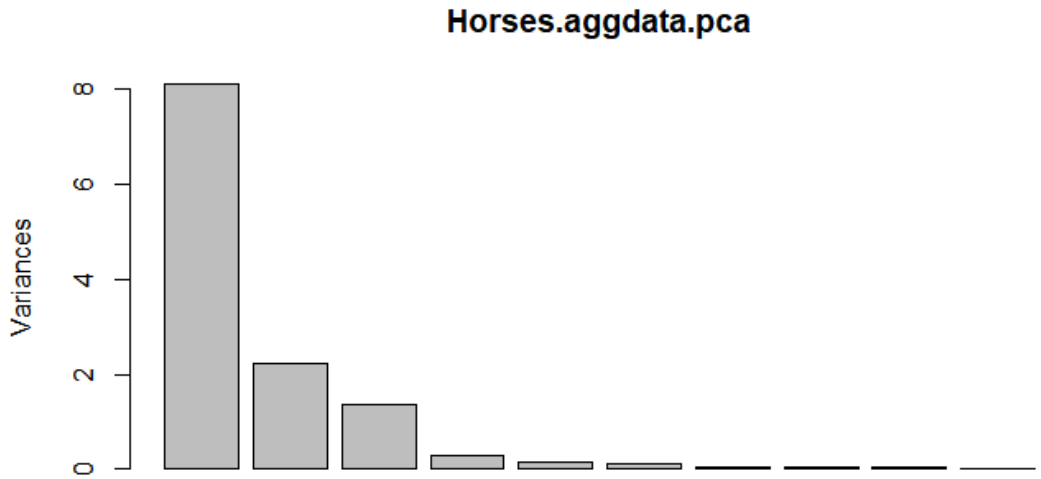

Figure 3.8 Scree plot of the aggregated Raman spectral data from the equine Mt3 bone sections. Cumulatively, the first six PCs accounted for approximately 97.2 % of the total spectral variation; individually, they explained about 64.9 %, 17.8 %, 10.8 %, 2.3 %, 1.3 %, and 0.8 %, respectively.

The loadings plots for the first six PCs (based on the covariance matrix) are shown in Figure 3.9. The loadings plot for PC1 appeared to represent the contribution from the mineral component of bone (like the original spectra), with the most dominant loading originating from the ν_1 PO_4^{3-} band within the 960 cm^{-1} region. The other noticeable loadings for PC1 were also phosphate-related bands: the ν_2 PO_4^{3-} and ν_4 PO_4^{3-} bands within the 400 – 620 cm^{-1} region. The loadings plot for PC2 appeared to be a mixture of the mineral and matrix components (with the major loadings again coming from the phosphate bands, alongside loadings from the Amide III, alkyl (CH_2) deformation, and Amide I functionalities within the 1240 – 1700 cm^{-1} region). The derivative-like shape of the ν_1 PO_4^{3-} band in the 960 cm^{-1} region of the PC2 loadings plot suggested some environmental change occurring for the bone mineral phosphate. The loadings plot for PC3 – also a mixture – showed more readily

distinguishable 'features' from the collagenous component around the 1240 – 1700 cm^{-1} region (this time, those from the hydroxyproline, proline, and phenylalanine functionalities around approximately 855 – 877 cm^{-1}, 921 cm^{-1}, and 1004 cm^{-1}, respectively, were also visible). It is interesting to note that apart from the scale of the loadings and the derivative-like shape of the ν_1 PO_4^{3-} band within the 960 cm^{-1} region observed in the PC2 loadings plot, the appearance of the loadings plot for PC3 was somewhat similar to that for PC2.

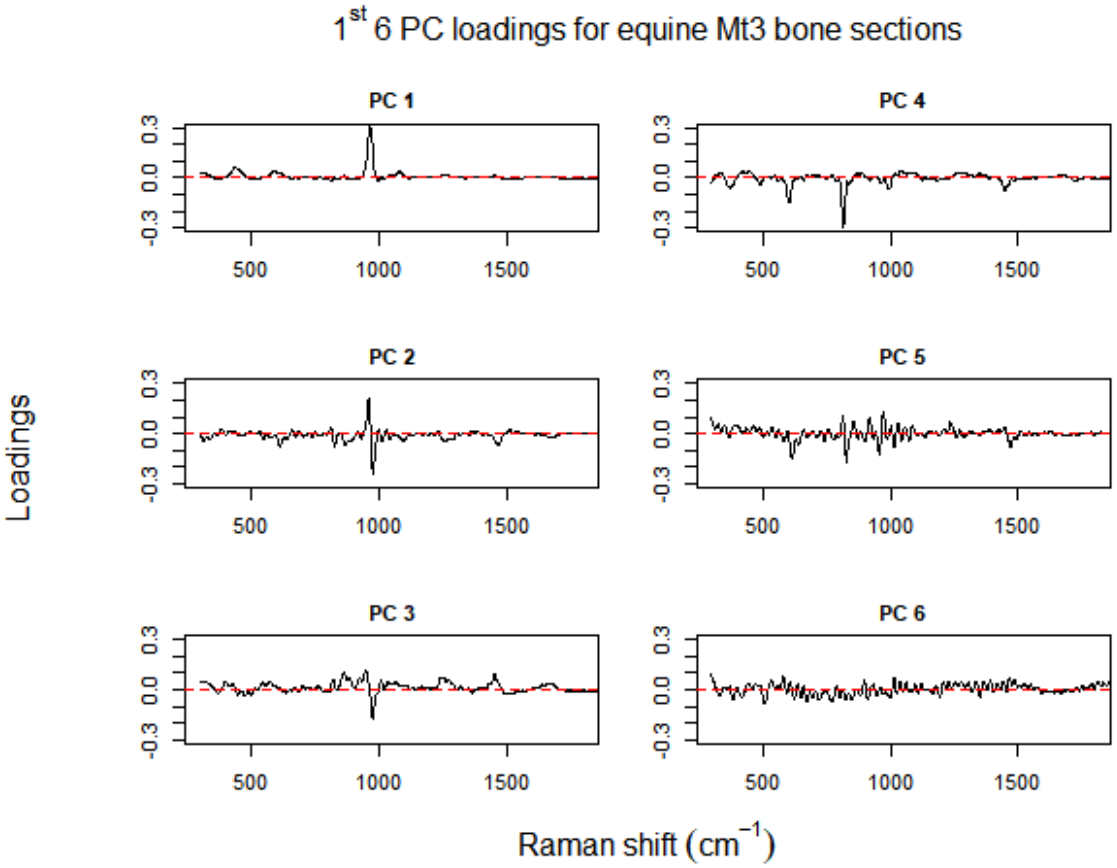

Figure 3.9 The first six PCs' loadings of the aggregated Raman spectral data from the equine Mt3 bone sections. The loadings plot for PC1 closely resembled the original spectra with respect to the mineral component. The loadings plots for PCs 2 and three appeared to be mixtures of contributions from the mineral and matrix components; the PC3 loadings seemed to show slightly more readily distinguishable features from the matrix component. PC4 seemed to be the contribution from the embedding material, polymethylmethacrylate (PMMA). The loadings for PC5 were likely also PMMA-related, but the loadings for PC6 were difficult to interpret.

Figure 3.10 shows the loadings plot for PC4, which resembled the mirror image of a Raman spectrum of the embedding material, PMMA (refer to the black spectrum in Figure 3.4). Though the PMMA did not disappear even after subtracting off the so-called 'horse' average, to some extent, PCA could separate it from the rest of the Raman spectral data. Since it occupied an entire principal component loading, it was likely that its contribution to the overall variation within the Raman spectral data was another indication of its infiltration into

the equine Mt3 bone sections' void and marrow spaces. As can be noted from the loadings plots of PCs 5 and six in Figure 3.9, identification of the functional groups that contributed to the higher PCs appeared to be noisier. The loadings plot for PC5 was probably also PMMA-related; it was difficult to determine what information could be garnered from the loadings plot for PC6.

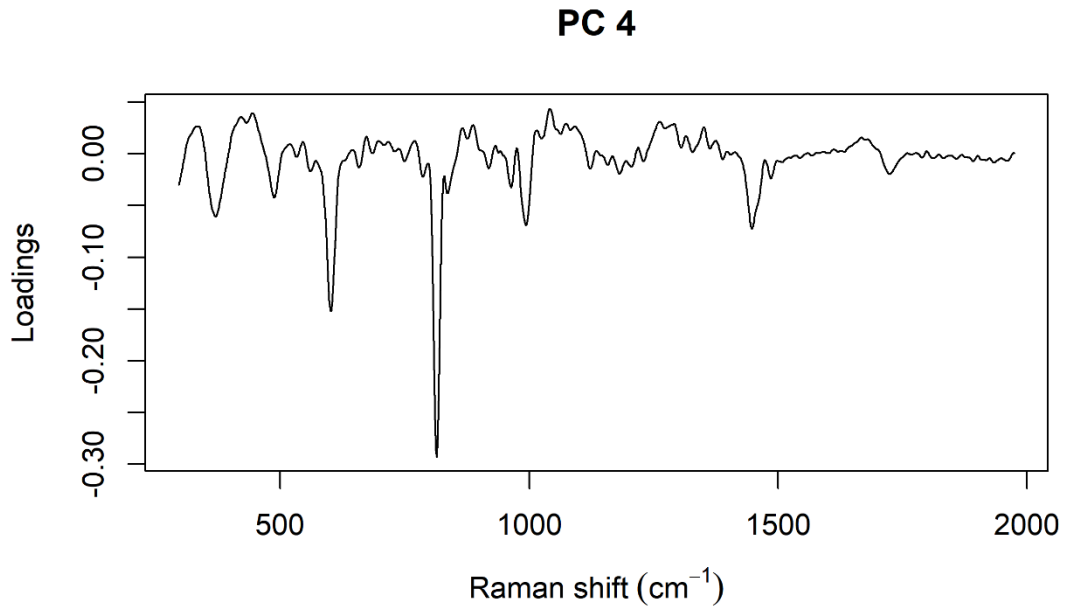

Figure 3.10 PC4 loadings plot of the aggregated Raman spectral data from the equine Mt3 bone sections appears to be the contribution from the embedding material, polymethylmethacrylate (PMMA), reminiscent of its Raman spectrum (cf. Figure 3.4).

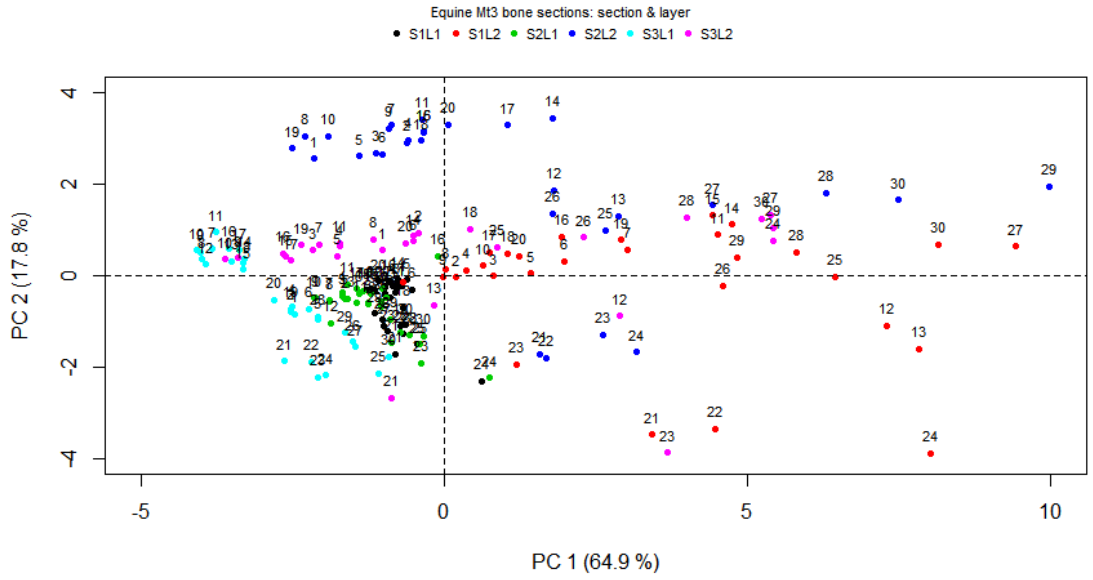

Figure 3.11 Scores plot for PC2 against PC1 from the aggregated equine Mt3 bone Raman spectral data (by section and layer): S1L1 (black); S1L2 (red); S2L1 (green); S2L2 (blue); S3L1 (cyan); S3L2 (magenta).
S1 = lateral parasagittal groove (prone to fracture); S2 = medial parasagittal groove (prone to fracture); S3 = medial condylar surface ('control' site);
L1 = articular cartilage (AC); L2 = subchondral bone (SCB);
horses 1 – 10 = newborn foals; horses 11 – 20 = five-month-old foals; horses 21 – 30 = two-year-old horses.

Figure 3.11 shows the scores plot for the second PC against the first for the six combinations of section and layer from the aggregated Raman spectral data collected from the articular cartilage and subchondral bone of the equine Mt3 bone specimens. It also draws attention to the fact that PCA could separate the sections and layers into clusters to bring out the within-object variations. The samples' scores showed a higher degree of separation along PC2; it raised the question of similarities between individuals concerning the matrix component. Most of the scores for the AC (L1) in all three sections (lateral parasagittal groove (S1), medial parasagittal groove (S2), and medial condylar surface (S3); coloured black, green, and cyan, respectively) tended to cluster towards the left-hand, negative side of the plot along PC1 (the exception being S1 and S2 scores from horse 24, a two-year-old).

There was greater scatter from the SCB (L2) in all three sections (coloured red, blue, and magenta, respectively) along PC1; this spread raised the question of the extent of heterogeneity between individuals in the mineral component of this layer of bone.

Scores from the SCB for S1 and S2 (coloured red and blue, respectively) were divided somewhat sparsely along PC1. For S1, all the newborn foals gathered closer to the centre of the plot. Of the five-month-old foals, horses 12 to 14, and 20 fell further to the right of centre.

The most spread along PC1 for S1 was from the two-year-olds, raising the question of the possibility of an age effect for the mineral component for this particular bone section.

In contrast to their 'counterpart' scores from the AC, scores from the SCB for S1 and S2 were located somewhat more haphazardly along PC2. For S1, the scores from most of the foals formed something of a gradient around the centre of the plot (whilst horses 11, 14, and 15 bunched in near two-year-old horses 26 and 29 slightly further down the horizontal axis). Many of the two-year-olds each seemed to be somewhat isolated within PC2.

Similarly, for S2, the scores from most of the foals formed a 'trend' along the uppermost part of the plot (the exceptions being horses 12 and 13, found slightly further down). Apart from horse 20, which was closer to most of the foals along PC2, scores from the two-year-old horses were far more spread out.

Most of the foals gathered loosely centre-left (excepting horses 12, 13, and 18) along PC2 for S3 (coloured magenta). Scores from many of the two-year-old horses tended to fall within the upper part of the scores plot for the PC2 dimension.

The scores plots of PC2 against PC1 in Figures 3.12 through 3.14 show an overview of the aggregated equine Mt3 bone Raman spectral data according to section (columns) and layer (rows). In the upper row are scores plots of the AC (L1) for the lateral parasagittal groove (S1), medial parasagittal groove (S2), and medial condylar surface (S3). In the lower row are the corresponding scores plots of the SCB (L2). Figures 3.13 and 3.14 are simply the respective halves of Figure 3.12.

The score for one of the two-year-old horses, horse 21, appeared slightly separated from the others along PC1 – towards the right-hand side of the plot – for the AC of both the lateral and medial parasagittal grooves (S1L1 and S2L1, respectively). The score from horse 16, a five-month-old foal, was also slightly separated from the others in S2L1 along PC2. Superficially, there appeared to be two clusters in the scores plot for S3L1 along PC2, seemingly with most of the foals in the upper cluster and most of the older horses in the lower cluster.

There was greater scatter in the scores for the SCB of all three sections along both PC1 and PC2. There also appeared to be something of a 'trend' along PC2 in all three sections, with looser clusters along PC1. Generally, it was the older horses (horses 21 to 30) that tended to spread out the most, implying there were more considerable microstructural differences within this age group. Interestingly, some of the newborn and five-month-old foals clustered

together with some of the two-year-olds; this suggests that perhaps some of the two-year-olds had areas of bone that were more recently formed, similar in composition to the immature bone of younger individuals. These areas of bone were not readily discernible in the epi-illuminated optical images of sampling areas from the older horses. Along PC1 for S3L2, horse 1 (a newborn foal) appeared to have a slightly different profile from the others. Along PC2, some horses from different age groups amassed in proximity to each other: horses 17 and 20 (a five-month-old and a two-year-old, respectively) were bunched with horse 10 (a newborn foal) near the centre of the plot. Scores from horses 7 and 19 (a newborn and a five-month-old foal, respectively) clustered together not too far away from that grouping, and the score for two-year-old horse 29 was rather close to those for horses 11, 14, and 15 (five-month-old foals). Apart from horses 12 to 14, 17, and 20, the foals seemed to separate well from the two-year-olds along both PC dimensions for S2L2; horses 21 to 24 were further separated from the other two-year-old horses along PC2. Likewise, with the exceptions of horses 12, 13, 18, and 21 (five-month-olds and a two-year-old), the foals and two-year-olds were well separated along PC1 for S3L2; scores from horses 12 and 13 (five-month-old foals), and horses 21 to 23 (two-year-olds) were disconnected from the others along PC2.

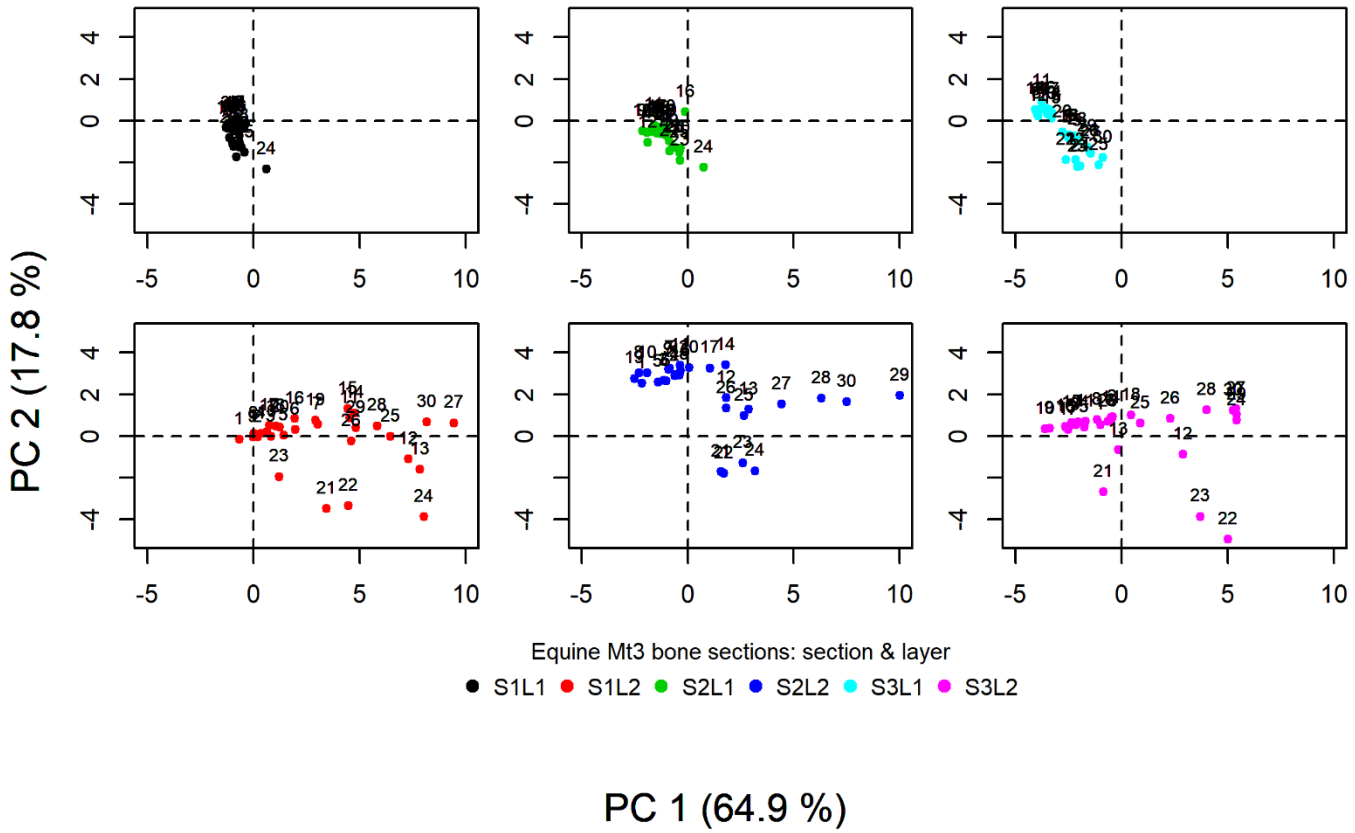

Figure 3.12 Montage of scores plots for PC2 against PC1 from the aggregated equine Mt3 bone Raman spectral data, by section (columns) and layer (rows): S1L1 (black); S1L2 (red); S2L1 (green); S2L2 (blue); S3L1 (cyan); S3L2 (magenta).
S1 = lateral parasagittal groove (prone to fracture); S2 = medial parasagittal groove (prone to fracture); S3 = medial condylar surface ('control' site);
L1 = articular cartilage (AC); L2 = subchondral bone (SCB);
horses 1 – 10 = newborn foals; horses 11 – 20 = five-month-old foals; horses 21 – 30 = two-year-old horses.

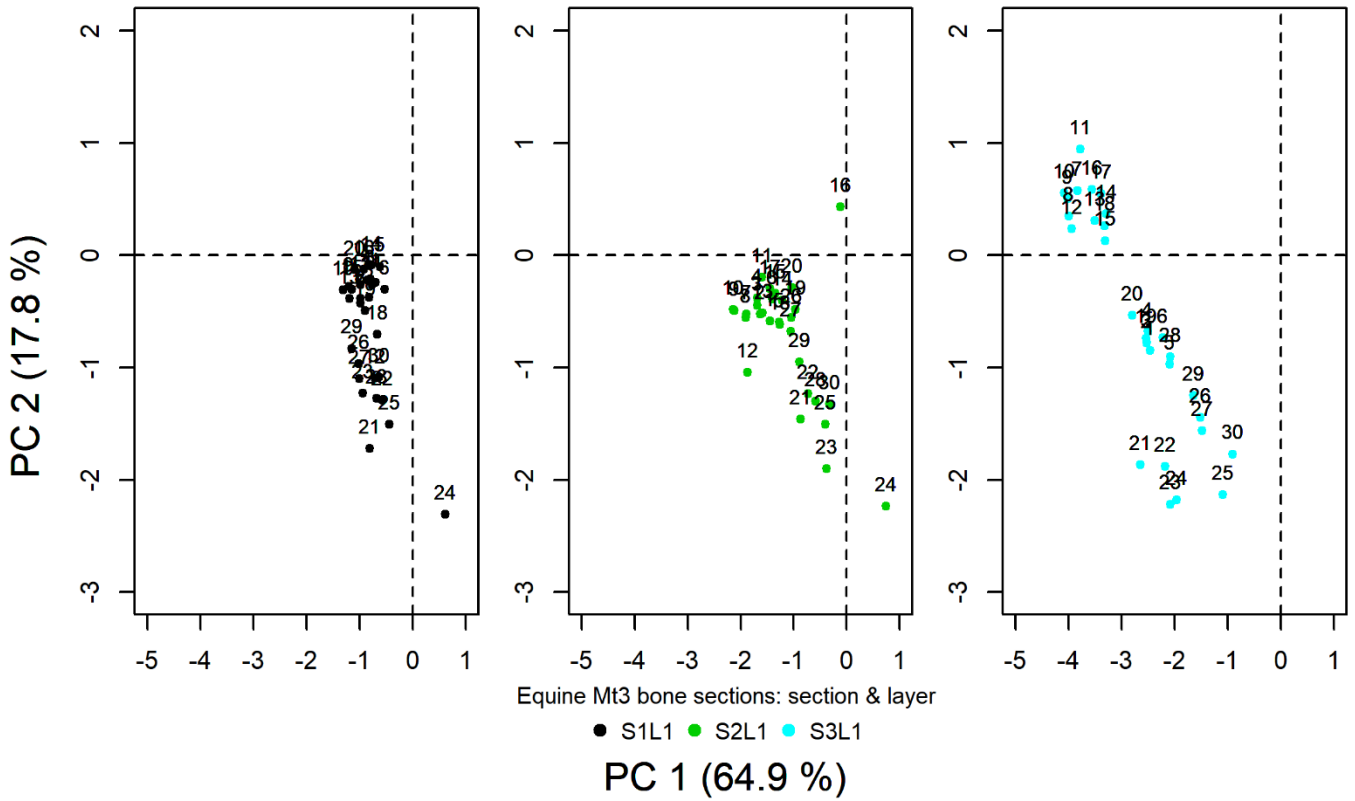

Figure 3.13 Upper panel of scores plots for PC2 against PC1 from the aggregated equine Mt3 bone Raman spectral data, by section (columns) and layer (row): S1L1 (black); S2L1 (green); S3L1 (cyan).
S1 = lateral parasagittal groove (prone to fracture); S2 = medial parasagittal groove (prone to fracture); S3 = medial condylar surface ('control' site);
L1 = articular cartilage (AC);
horses 1 – 10 = newborn foals; horses 11 – 20 = five-month-old foals; horses 21 – 30 = two-year-old horses.

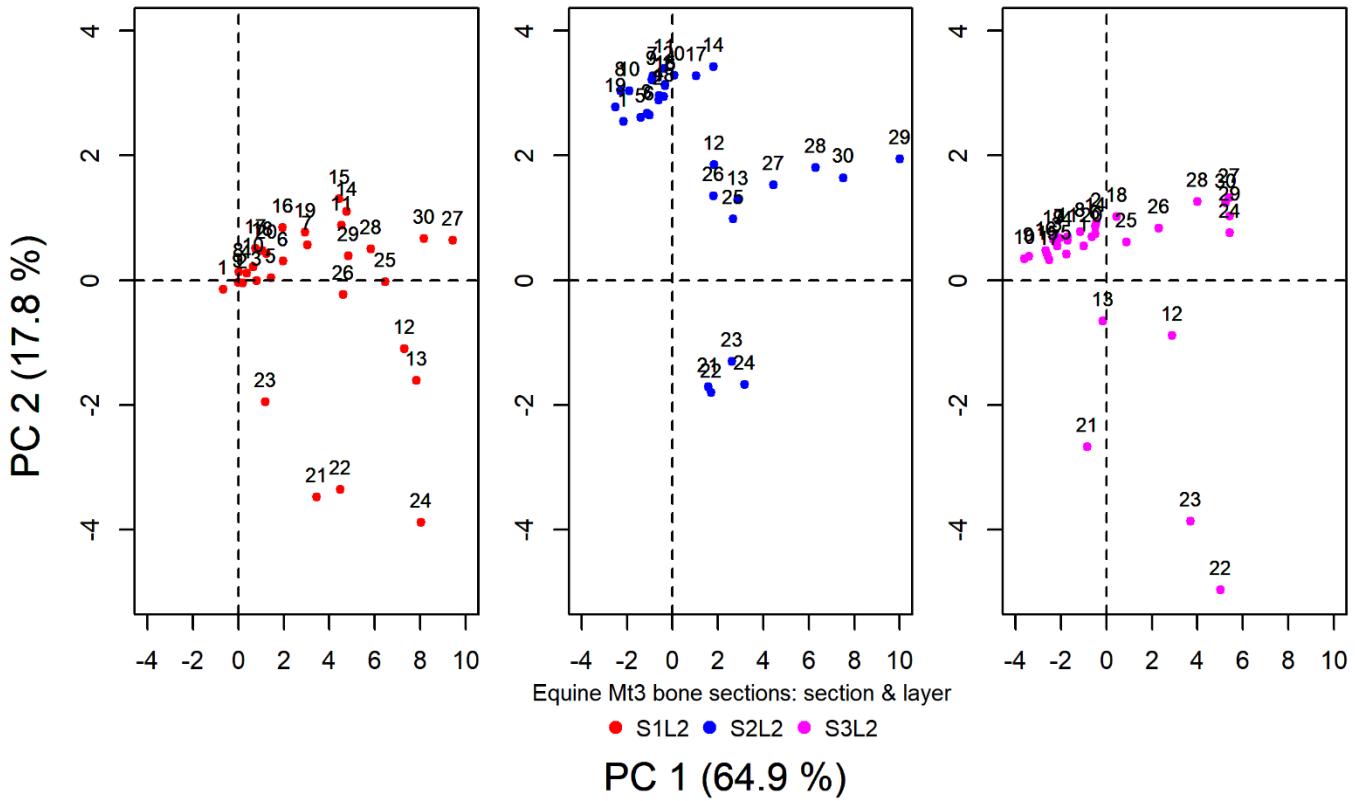

Figure 3.14 Lower panel of scores plots for PC2 against PC1 from the aggregated equine Mt3 bone Raman spectral data, by section (columns) and layer (row): S1L2 (red); S2L2 (blue); S3L2 (magenta).
S1 = lateral parasagittal groove (prone to fracture); S2 = medial parasagittal groove (prone to fracture); S3 = medial condylar surface ('control' site);
L2 = subchondral bone (SCB);
horses 1 – 10 = newborn foals; horses 11 – 20 = five-month-old foals; horses 21 – 30 = two-year-old horses.

Figure 3.15 shows the scores plot of the third PC against the first for the six combinations of section and layer from the aggregated equine Mt3 bone Raman spectral data. Some similar patterns to those in Figure 3.11 exist: a ‘trend’ along PC3 was noticeable. Likewise, many of the scores for the AC in all three sections (coloured black, green, and cyan) tended to cluster towards the left-hand side of the plot along PC1 (the scores from the two-year-old horse 24 and one from the 5-month-old horse 16 again being the odd ones out). S1L1 and S2L1 still coincided, with scores from S3L1 still being a little distinct from them. Scores from the SCB in all three sections (coloured red, blue, and magenta) were more disperse along PC1. There was some mixing between scores from different-aged horses, and scores from the older individuals were generally more diffuse in both dimensions.

Curiously, about half of the scores from the SCB in S2 were contiguous with clusters from the AC of S1 and S2 – most of them from both sets of foals (horses 12 to 14, 17, and 20 appeared around centre-right along PC1). Unlike in Figure 3.11, there was no overlap between scores from the two layers of the medial condylar surface (coloured cyan and magenta, respectively) in either dimension. Some of the scores from the SCB in S1 and S2 (red and blue) collected together: for instance, S2 from horse 12 near S1 scores from horses 6 and 16 along a PC3 gradient in the centre-right of the plot, and S1 from horse 12 near S2 from horse 30 towards the far right of the plot. Most of the scores from the SCB of S3 from the foals could be found near the bottom left-hand side of the plot; horses 12 and 18 appeared in the bottom centre-right. Apart from horse 21 near bottom centre-left, the older horses were towards the far central-right and bottom-right of the plot.

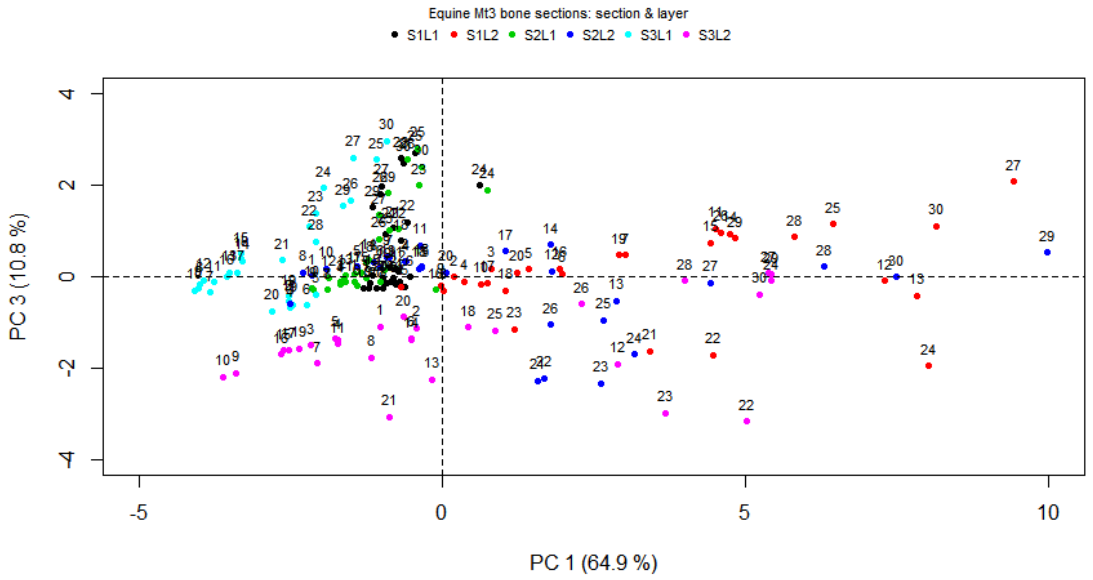

Figure 3.15 Scores plot for PC3 against PC1 from the aggregated equine Mt3 bone Raman spectral data (labelled by section and layer): S1L1 (black); S1L2 (red); S2L1 (green); S2L2 (blue); S3L1 (cyan); S3L2 (magenta).
S1 = lateral parasagittal groove (prone to fracture); S2 = medial parasagittal groove (prone to fracture); S3 = medial condylar surface ('control' site);
L1 = articular cartilage (AC); L2 = subchondral bone (SCB);
horses 1 – 10 = newborn foals; horses 11 – 20 = five-month-old foals; horses 21 – 30 = two-year-old horses.

Figure 3.16 shows the scores of the third PC against the second for the six combinations of section and layer from the aggregated equine Mt3 bone spectral data. Some patterns persisted: the older horses had – yet again – tended to be more spread out than the younger ones. Scores from the articular cartilage of all three sections were more bunched together than those from the subchondral bone, and 'trends' were more evident in the PC3 dimension than in the PC2. There were two groupings of the SCB from S2: scores from all except two of the foals (horses 12 and 13 fell in with the larger group of older horses in the grouping just to the left) were on the far centre-right of the plot. Those from most of the older horses were located near SCB scores for S1 and S3 (those from horses 21 to 24 were far more detached, about the lower left-hand part of the plot).

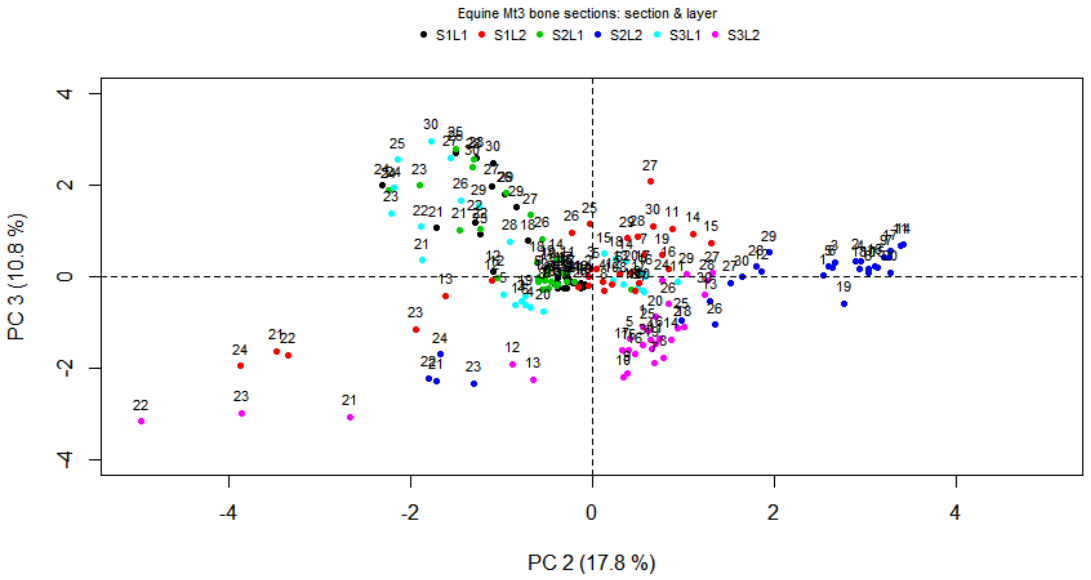

Figure 3.16 Scores plot for PC3 against PC2 from the aggregated equine Mt3 bone Raman spectral data (labelled by section and layer): S1L1 (black); S1L2 (red); S2L1 (green); S2L2 (blue); S3L1 (cyan); S3L2 (magenta).
S1 = lateral parasagittal groove (prone to fracture); S2 = medial parasagittal groove (prone to fracture); S3 = medial condylar surface ('control' site);
L1 = articular cartilage (AC); L2 = subchondral bone (SCB);
horses 1 – 10 = newborn foals; horses 11 – 20 = five-month-old foals; horses 21 – 30 = two-year-old horses.

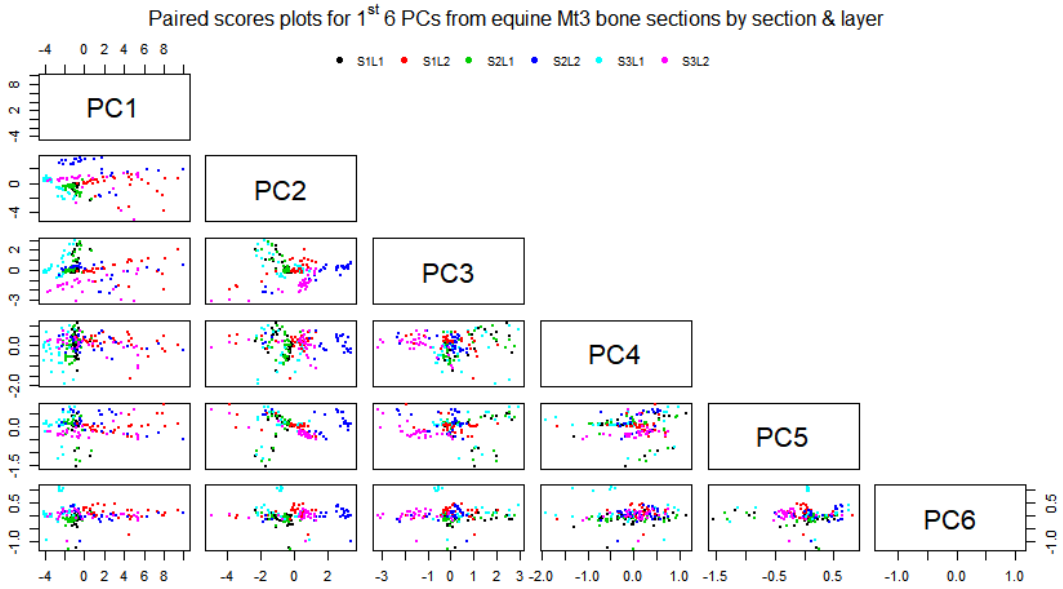

Figure 3.17 Paired scores plot matrix for the first six PCs from the aggregated equine Mt3 bone Raman spectral data (labelled by section and layer): S1L1 (black); S1L2 (red); S2L1 (green); S2L2 (blue); S3L1 (cyan); S3L2 (magenta).
S1 = lateral parasagittal groove (prone to fracture); S2 = medial parasagittal groove (prone to fracture); S3 = medial condylar surface ('control' site);
L1 = articular cartilage (AC); L2 = subchondral bone (SCB).

Figure 3.17 shows a scatterplot matrix of the fifteen combinations of pairs of the first six PCs (from the six combinations of section and layer) from the aggregated equine Mt3 bone Raman spectral data. Surveying the scatterplot matrix, some of the spreading or grouping of the PC scores from the Raman spectral data may have been slightly more apparent with pairs from the lower PCs than from those of the higher PCs. The following general observations were made about PC scores from PC pairs not covered in Figures 3.11 to 3.16. Some of the subchondral bone-associated L2 scores (specifically those coloured blue (medial parasagittal groove) and magenta (medial condylar surface)) seemed to share similar mineral profiles with many of the articular cartilage-associated L1 scores (representing the lateral and medial parasagittal grooves (coloured black and green, respectively)), and medial condylar surface (coloured cyan) along the PC1 dimension. It could almost be said that PC1 roughly separated many of the AC-associated L1 scores (coloured black, green, and cyan, respectively) from SCB-associated L2 scores (coloured red, blue, and magenta, respectively).

As noted earlier in Figure 3.9, loadings for the second and third PC dimensions represented contributions to spectral variation from mineral-matrix component mixtures. The loadings for both PCs were still dominated by mineral phosphate-related functionalities, but also showed noticeable contributions from organic matrix hydroxyproline, proline, phenylalanine, Amide III, CH_2 deformation, and Amide I functionalities. For the most part, the PC2 dimension tended to differentiate many of the S2L2 scores (coloured blue, representing the SCB of the medial parasagittal groove; associated with some of the newborn foals and two-year-old horses in the individual paired PC scores plot, not shown) from the other scores. The PC3 dimension would typically segregate many of the S3L2 scores (coloured magenta, representing the SCB of the medial condylar surface) from the other scores. The fourth to sixth PC dimensions showed a greater overall scattering of the scores. The PMMA-representative PC4 dimension did not seem to show any exceptionally distinctive groupings. The PC5 dimension tended to split some of the AC-associated L1 scores (coloured black, green, and cyan, respectively; associated with some two-year-old horses in individual paired PC scores plot, not shown) from the other scores. The PC6 dimension seemed to isolate all the scores related to one horse – horse 11 (a five-month-old foal; in individual paired PC scores plot, not shown) – and some of the S3L1 scores (coloured cyan, representing the AC of the medial condylar surface) from the other scores.

3.4 Discussion

Bone microstructure reflects any abnormalities within its major components, but these are not necessarily plain in gross clinical appraisals; if left undetected and unmonitored, such anomalies could lead to orthopaedic diseases like osteoporosis. Current clinical risk assessment methods only address one aspect of bone strength and health: quantity; more research attention needs to turn to bone quality to gain a better understanding of overall bone strength. A better understanding would mean more accurate risk assessment and improved general medical practice. Timely detection, interpretation and classification of changes in the chemical composition of bone hold the potential to help with eventual identification of at-risk individuals. Molecular vibrational spectroscopy, in tandem with multivariate statistical analytical techniques, would seem a logical starting point in this pursuit. This field is recent, with infrared (IR) probing of model animal bone specimens being a bit more established than Raman is within the literature.

There was a dearth of (readily available and accessible) literature with previous vibrational spectroscopy-chemometrics applications on equine bone specimens – one of the few studies being that of IR-linear discriminant analysis (LDA) work conducted by Nicholson *et al*. in (37). Exploratory work herein was to determine the ability of coupled Raman microspectroscopy and chemometrics (namely, principal component analysis (PCA)) to identify possible changes in the chemical composition of equine bone specimens; some of these changes could potentially help in building a classification model for the early stages of equine bone disease.

For some time, many researchers have been making a more thorough investigation into the structure of bone, rather than how bone microstructural alterations play a major role in the development of various orthopaedic diseases. Since bone microstructural alterations can provide much information about the individual from which it originated, it has also naturally lent itself to other research fields, including the forensic, anthropological (as touched on in Chapter 1), and archaeological sciences. France *et al*. screened for collagen diagenesis (alterations to collagen that lead to its degradation) from the outer surfaces and inner cross-sections of a collection of archaeological and palaeontological bone specimens with bivariate and multivariate statistics (PCA and partial least-squares-discriminant analysis (PLS-DA)) (54). The condition of the collagen in the bone specimens ranged from poorly preserved to well preserved. The mineral-to-matrix peak ratio (based on the first symmetric phosphate band

and the carbonyl backbone of the Amide I band) was found to best distinguish between the poorly preserved and well-preserved collagen. Specimens that had well-preserved collagen had a lower ratio – especially those from the inner cross-sections. The PCA was able to separate most of the well-preserved specimens from the poorly preserved specimens, with collagen-associated bands in both first two principal components' (PCs) loadings being highlighted as the most important sources of spectral variation. These indications, together with the ratio values, then informed predictions of collagen quality for a training set. The PLS-DA gave sensitivity rates of 95 % and 68 % for the inner cross-sectional and outer surface spectra, respectively. These rates also pointed towards the greater likelihood of correct classification for the well-preserved specimens. Thus, their findings eventually established a method of accurate collagen quality determination in bone specimens before conducting other more destructive tests. The PC loadings in their work, like the work here, highlighted the functionalities that were the more likely sources of spectral variation separating the bone specimens.

It must be stated from the outset that with minimal *a priori* knowledge about the horses' backgrounds (utterly separate from Nicholson who, having provided the bone samples, did have prior knowledge), the following interpretations of PCA results in this chapter are at best speculative. Raman spectral data were collected from the articular cartilage (AC) and subchondral bone (SCB) of the lateral and medial parasagittal grooves (fracture predilection sites) and medial condylar surface ('control' site) of the third metatarsal (Mt3) bone specimens of 30 thoroughbred racehorses (ten newborn foals, ten five-month-old foals, and ten two-year-olds). The data were aggregated via means, labelled by section and layer, to minimise a so-called 'horse effect' that might have obscured other patterns within the chemical data; PCA was then conducted across the whole Raman spectral range on the condensed data set. PCA was used to identify the primary sources of variation across the entire Raman spectral range of bone tissue within microstructural dimensions, enabling identification of structural make-up features that differ between the bone tissue layers.

PCA results yielded six principal components (PCs) relating to the bone components, which accounted for approximately 97.2 % of the total spectral variation. These six PCs represented a contribution from the mineral component, as well as some opposing inputs from mineral-matrix component mixtures, and the embedding material, PMMA. Additionally, the represented 'mixture' contributions highlighted some compositional changes to the mineral component. That is, these PCA results suggested that localised microstructural differences

were detectable – especially within the bone mineral. From a statistical point of view, this type of PCA was an analysis of group means, having taken out averages and analogous to performance of analysis-of-variance (ANOVA) on a massively multivariate data set. The overall variation from multiple correlated response variables was considered simultaneously, assessing whether differences in combinations of correlated attributes among samples were detectable.

One of the main patterns observed in the distribution of aggregated PC scores in the subspace was that the PC scores relating to the articular cartilage of all three equine Mt3 bone specimen sections tended to cluster more closely together. PC scores from the subchondral bone consistently showed greater scatter. There could have been to some extent, some similarity between individuals concerning the matrix component. There are different developmental timescales for an equine bone joint's substructures (AC and SCB, in this case) such that its collagenous matrix develops and matures earlier than the mineral component (55, 56). There do not tend to be any significant biochemical and structural modifications to the collagen network after about six months to one year of age; the mineral component of equine subchondral bone, however, continues development until a horse is about four-years-old (55, 56). These different development timescales – as well as structural alterations due to recurring bone remodelling – for the two major bone components would affect the bone's overall load-bearing capacities (55-57). There may also have been the possibility of variation from the mineral component masking that of the organic matrix component. Inorganic functionalities like phosphates and carbonates' Raman spectral signals are inherently stronger and sharper than those from organic functionalities such as amides are.

There was also the question of the degree of heterogeneity in the bone mineral of individuals, especially in the subchondral bone. Of the three age groups, the two-year-old horses tended to have the most spread in their PC scores. Interestingly, both the PC scores for the representative mineral contribution to the AC from the fracture-prone lateral and medial parasagittal grooves from two-year-old horse 24, were separate from its co-aged group members, raising the question as to whether some abnormality was starting to manifest itself within this particular horse's mineral substructures.

The subchondral bone (SCB) is more heavily mineralised than the articular cartilage (which has some degree of mineralisation, particularly in its calcified sublayer (the articular calcified cartilage, ACC)). It is thought that fracture predilection sites such as the lateral and medial parasagittal grooves in the equine hind limb have a higher degree of mineralisation. Doube *et*

al. (58) point out a likely indication of early defect development is highly mineralised ACC and SCB layers – that are stiffer and less stiff, respectively – when compared to control sites within the same joint. Delayed chondroclastic resorption of the ACC in earlier life leads to improper fusion of the cartilage and SCB, thus, increasing the probability of expanding linear defects (58).

PC scores from some of the older horses tended to cluster together with some of the younger horses' scores. This PC score distribution implied that perhaps aggregates of sampled sites from the older horses were those of more recently-formed or remodelled bone somewhat similar in composition to those from younger individuals. Both shifts in wavenumber and changes to the relative intensities of sub-bands in the PC loadings plots suggested subtle bone tissue compositional differences. Shifts were indicative of different chemical environments and intensity changes of amounts of bone tissue component present. Phosphate functionalities relayed information about bone mineral present in the sample; amides pointed at the amount and properties of the organic matrix. The additional presence of carbonate functionalities in the Raman spectra of bone tissue would be indicative of the mineral crystals' maturation status, as they are associated with older mineral crystals (35, 53). Tissue age can often be a confounding factor with specimen age since there are various ongoing biological processes within each individual (35). Interstitial lamellar bone is older than osteonal bone since the remodelling process triggers osteon formation.

As an indirect comparison, there were 'contrasting' chemometric results obtained by infrared (in (37)) and Raman spectroscopies due, in part, to the complementary nature of the two vibrational spectroscopic techniques. IR would show groupings based on the matrix, Raman would show groupings based on the mineral. Again, amide spectral bands were more intense in the former and phosphates stronger in the latter. Contrasting results were also likely attributable to the chosen chemometric technique: PCA will create 'natural' separation of unlabelled data with no initial knowledge of the expected differences; LDA, however, attempts to group data into predefined or *a priori* classes (59). The spectral collection methodologies differed, too: spectra from an amalgamated AC layer here as opposed to distinct spectra from each of the hyaline articular cartilage (HAC) and ACC layers in (37).

Limitations

There were a few limitations in this study. Fluctuating laser power day-to-day may have affected signal strength of the embedded bone specimens' chemical functionalities, especially the already weaker organics.

The epi-illumination set-up gave a rather low-contrasting view of the cartilaginous layers in the bone specimens, so the majority of the time there was difficulty distinguishing between the HAC and ACC layers, even when collecting test spectra to try to confirm spectral differences. Thus, the decision was made to amalgamate the two into a single AC layer. If it had been easier to identify non-calcified sites, both visually and spectrally, there might have been different and perhaps more apparent clustering results from the aggregated spectral data.

There were also timeline alterations and contractions to data collection procedures from pressure due to competing uses of the multi-laser inverted Raman microscope system. Due to the need to generate some results for two data sets, the PCA conducted on the spectral data from the Mt3 bone specimens was abbreviated; perhaps subsetting the aggregated spectral data may have also generated more understandable clustering results.

Chapter 4 – Results & Discussion: Fresh samples from the equine carpal joint bone specimens

4.0 Introduction

Likewise, a shortage of literature covering fresh equine carpal bone specimens that have been characterised and analysed by way of Raman microspectroscopy and chemometrics exists; any inferences that could be made from this chapter should relieve some of the shortage. This chapter covers the application of multivariate statistical analysis to Raman spectral data collected from the equine carpal joint bone specimens. It is again prefaced with exemplar plots of the pre-processed Raman spectral plots from two of the horses (one from each age group) and a table summarising the data groupings. It then moves towards the actual findings of the multivariate statistical analysis of the two data subsets and their 'layered' sub-subsets, the differenced data set, the 'stacked' data set (that is, the combination of the two subsets), the mean-adjusted, 'stacked' data set, and 'layered' subsets of this whole data set. It discusses each one in turn, before finally rounding off with the limitations of this aspect of the presented body of work.

4.1 Raman spectral data from the equine carpal joint bone sections

Figures 4.1 to 4.8 are some examples of the processed Raman spectral data from the equine carpal joint bone sections – two from each subset (one from a horse in each of the two age groups (horses 4 and 19)). If necessary, refer back to Table 3.1 in subsection 3.2.1 for typical Raman spectral assignments of bone.

Many of the spectral assignments for the articular cartilage (AC) were likely to be those from collagen. Some spectral bands from other functional groups, however, like lipids and macromolecules can overlap with those from smaller proteins. As with the subchondral bone (SCB) spectra from the third metatarsal (Mt3) bone specimens, many of the expected functional groups from both components were present.

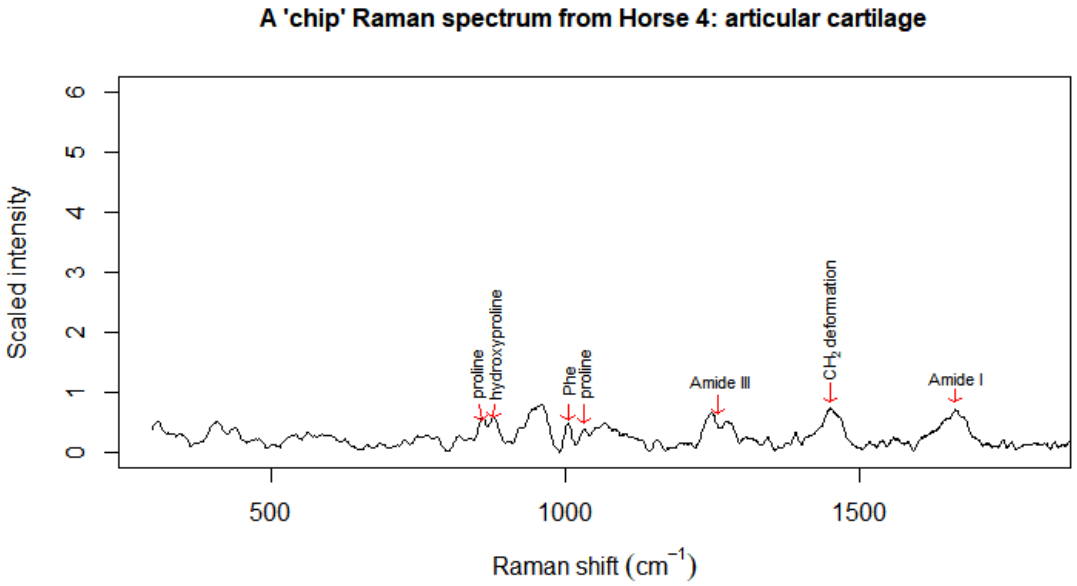

Figure 4.1 An aggregated 'chip' Raman spectrum from the articular cartilage (AC) of the third carpal bone (C_3) from horse 4, one of the two-year-old horses. Typical band assignments from the collagenous component are highlighted.

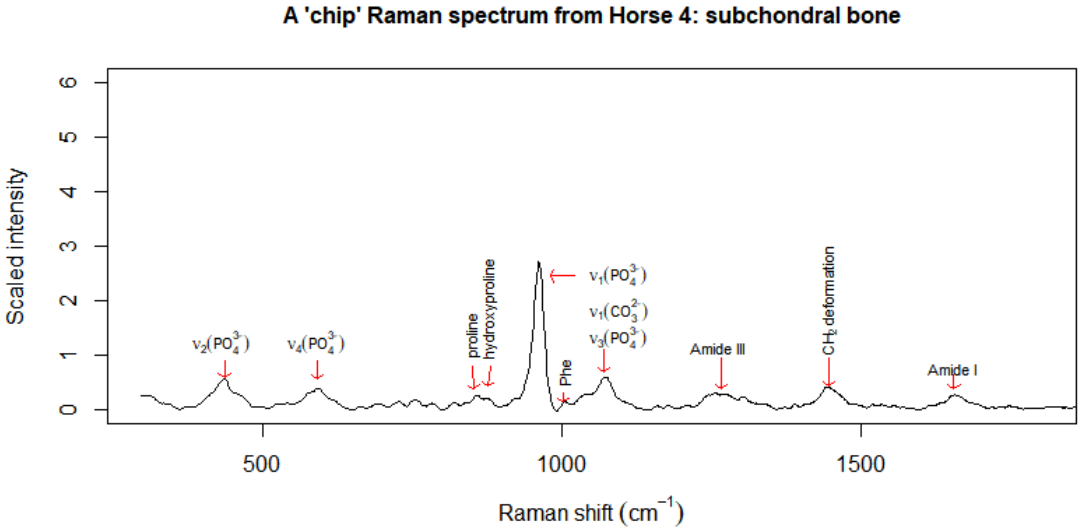

Figure 4.2 An aggregated 'chip' Raman spectrum from the subchondral bone (SCB) of the third carpal bone (C_3) from horse 4. Typical assignments from the mineral and matrix components are highlighted.

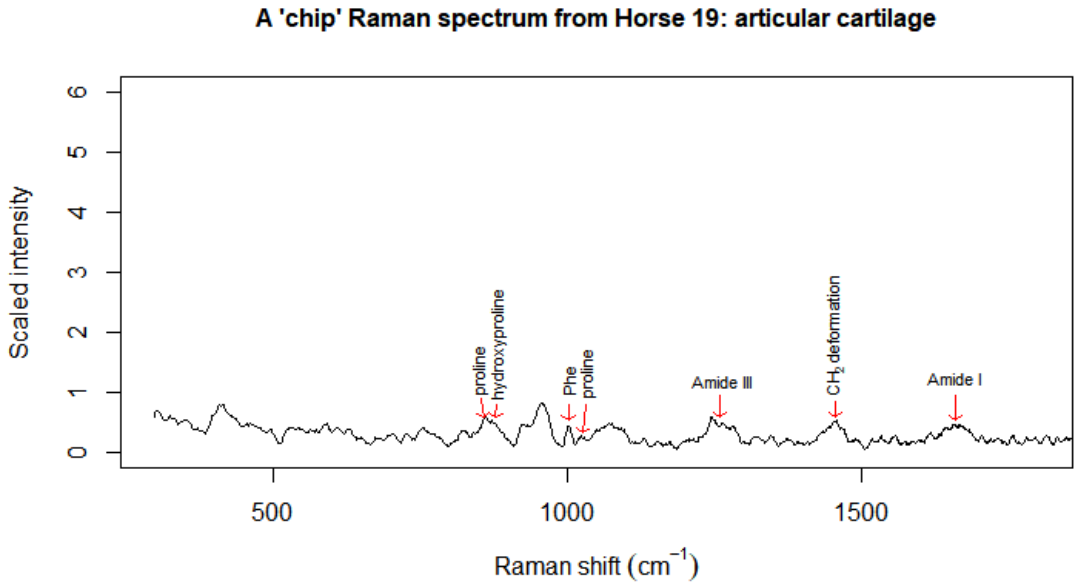

Figure 4.3 An aggregated 'chip' Raman spectrum from the AC of the third carpal bone (C_3) from horse 19, a three-year-old horse.

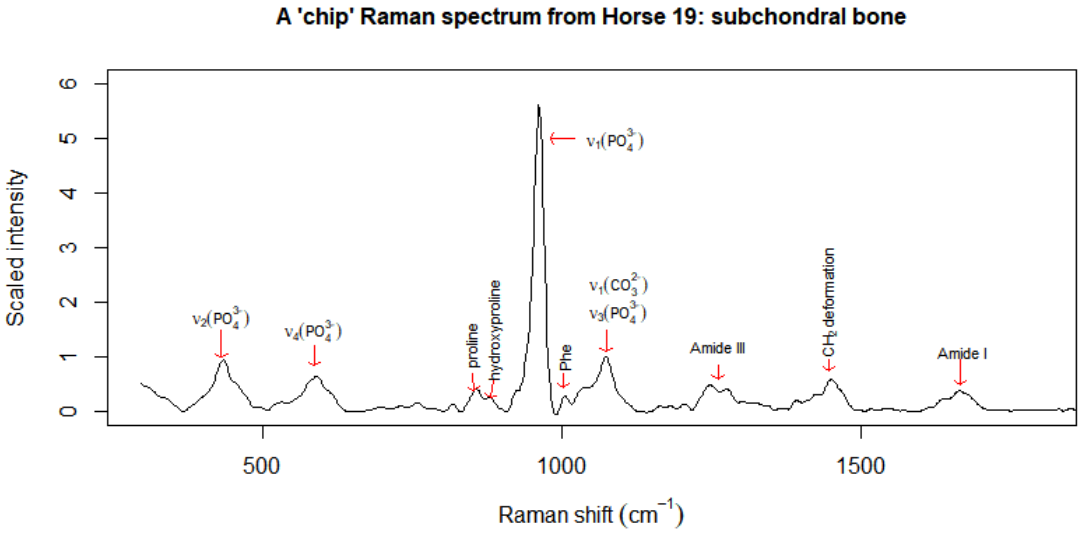

Figure 4.4 An aggregated 'chip' Raman spectrum from the SCB of the third carpal bone (C_3) from horse 19.

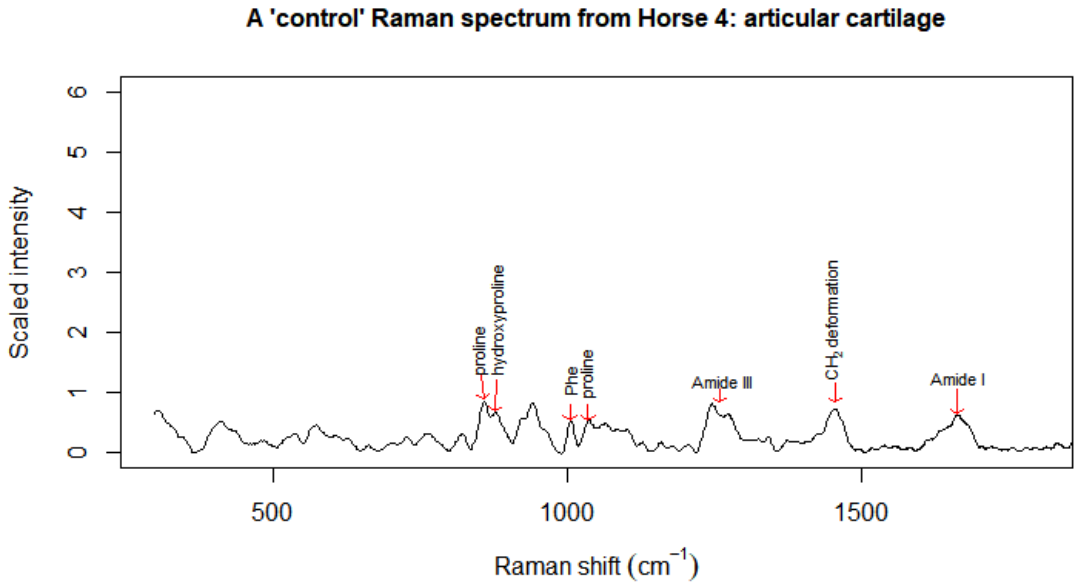

Figure 4.5 An aggregated 'control' Raman spectrum from the AC of the C_3 from horse 4.

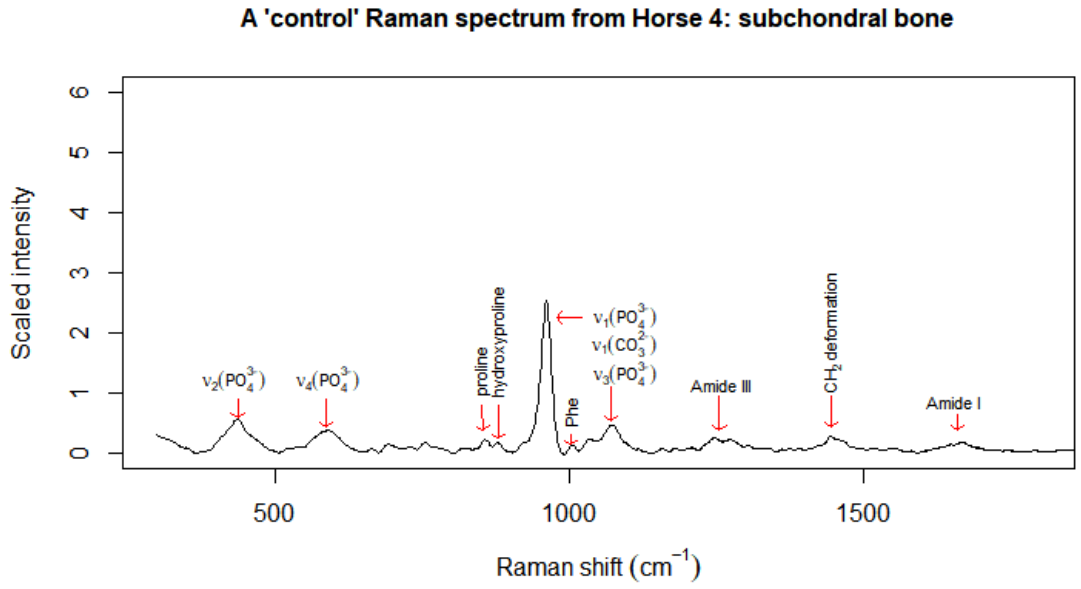

Figure 4.6 An aggregated 'control' Raman spectrum from the SCB of the C_3 from horse 4.

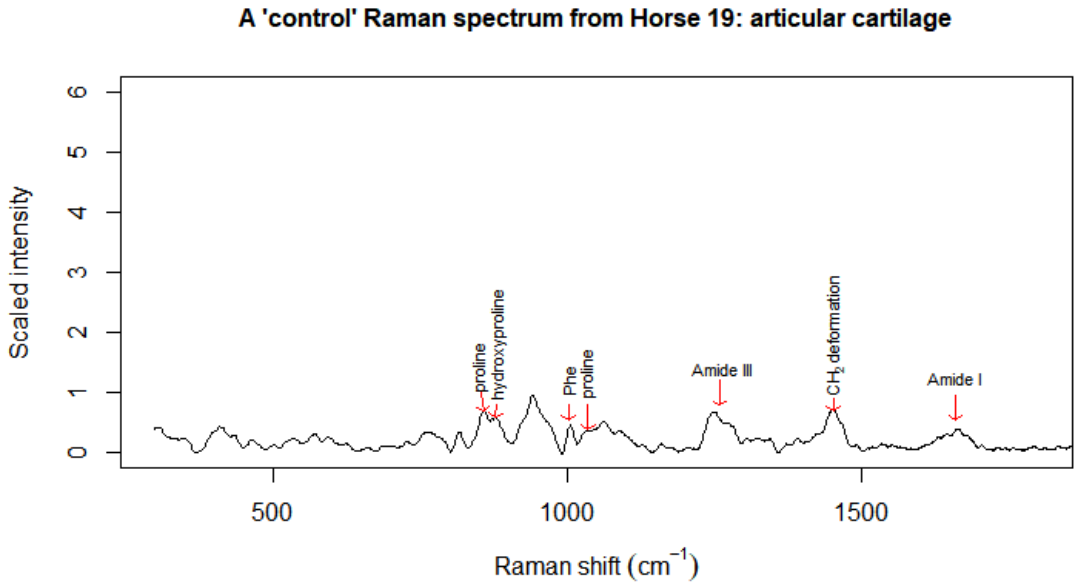

Figure 4.7 An aggregated 'control' Raman spectrum from the AC of the C_3 from horse 19.

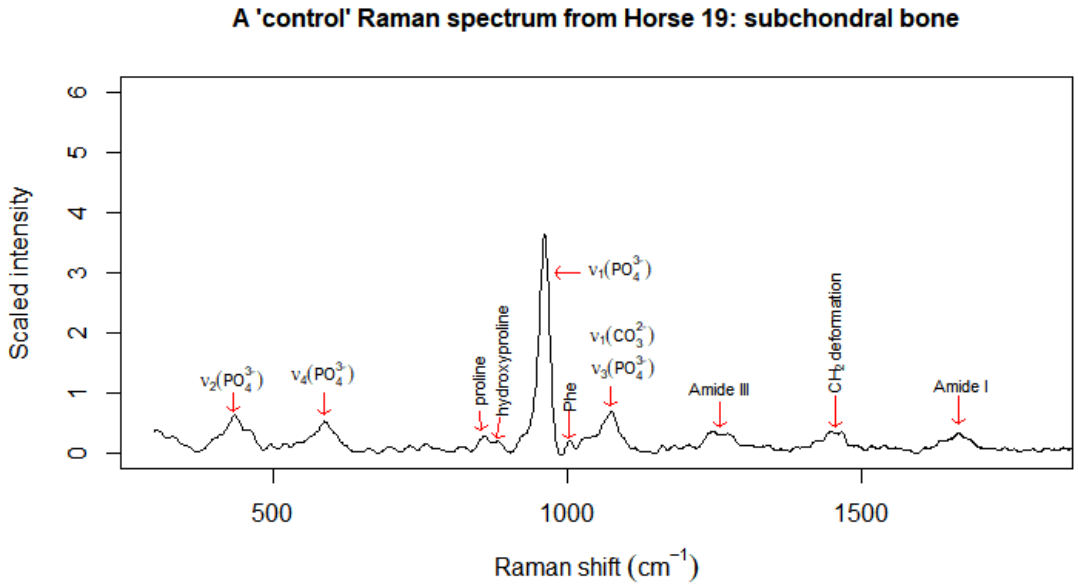

Figure 4.8 An aggregated 'control' Raman spectrum from the SCB of the C_3 from horse 19.

4.2 Multivariate statistical analytical results

4.2.1 Summary table

Table 4.1 outlines the Raman spectral data groupings that follow in subsection 4.2.2:

Table 4.1 Summary of the Raman spectral data groupings used in the PCA of the equine carpal joint bone specimens.

Figure	Spectral data grouping
4.9	scree plot of aggregated ‘chip’ spectral data subset from equine carpal joint bone sections
4.10	first six PCs’ loadings plots from aggregated ‘chip’ spectral data
4.11	PC2 vs PC1 scores plot from aggregated ‘chip’ spectral data subset, by section and layer
4.12	paired scores plot matrix for the first six PCs from aggregated ‘chip’ spectral data subset, by section and layer
4.13	scree plot of the articular cartilage (AC) spectral data subset of aggregated ‘chip’ spectral data subset from equine carpal joint bone sections
4.14	first six PCs’ loadings plots from articular cartilage (AC) spectral data subset of aggregated ‘chip’ spectral data subset
4.15	PC2 vs PC1 scores plot from articular cartilage (AC) spectral data subset of aggregated ‘chip’ spectral data subset, by section and layer
4.16	paired scores plot matrix for the first six PCs from articular cartilage (AC) spectral data subset of aggregated ‘chip’ spectral data subset, by section and layer
4.17	scree plot of subchondral bone (SCB) spectral data subset of aggregated ‘chip’ spectral data subset from equine carpal joint bone sections

4.53	paired scores plot matrix for the first six PCs from mean-adjusted, aggregated, ‘stacked’ subchondral bone (SCB) spectral data subset, by section and condition

4.2.2 PCA

This section covers various results from the PCA of the aggregated Raman spectral data from the equine carpal joint bone sections.

4.2.2.1.1 PCA of the ‘chip’ spectral data subset

Results from the PCA of the aggregated ‘chip’ spectral data subset from the equine carpal joint bone sections follow below. In a similar fashion to the Mt3 bone sections, additional details about the horses were incorporated into the analysis; the subset was also averaged by horse to make within-object variability more intelligible (refer back to subsection 2.5.2, if necessary).

Figure 4.9 shows the scree plot of the variances for the first 10 PCs from the ‘chip’ subset. The first six PCs accounted for approximately 97.4 % of the total variation in the data subset, rapidly tailing off after PC4. PC1 explained about 64.1 % of the total variation, while PCs 2 through six accounted for about 23.8 %, 5.9 %, 2.5 %, 0.677 %, and 0.448 %, respectively.

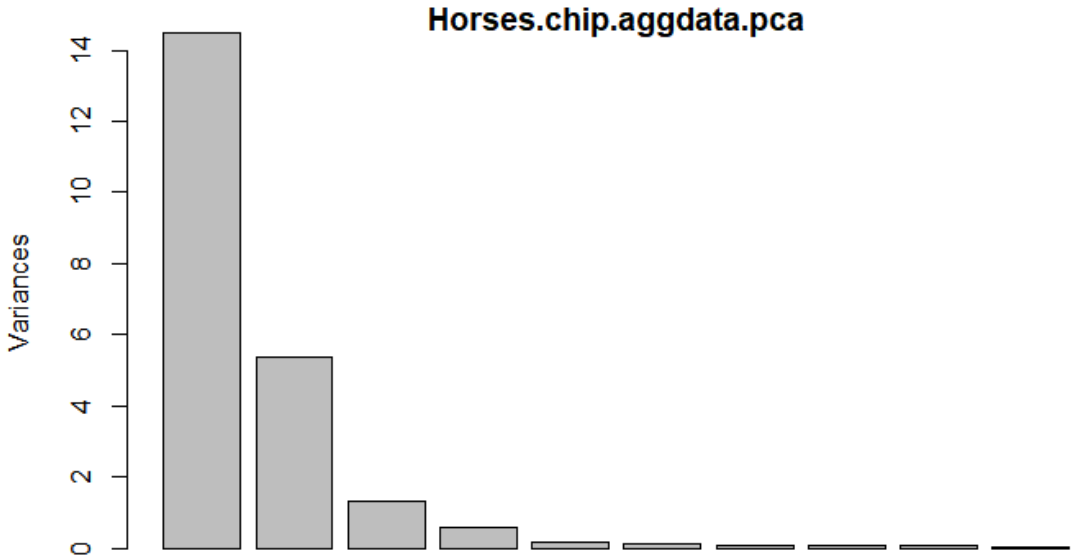

Figure 4.9 Scree plot for the aggregated 'chip' Raman spectral data subset from the equine carpal joint bone sections. Cumulatively, the first six PCs accounted for approximately 97.4 % of the total spectral variation; individually, they explained about 64.1 %, 23.8 %, 5.9 %, 2.5 %, 0.7 %, and 0.5 %, respectively.

The loadings plot for the first six PCs (based on the covariance matrix) is shown in Figure 4.10. The loadings plot for PC1 appeared to represent a mixture of contributions from the mineral and organic matrix components of bone (much like the original spectra from the subchondral bone and hinting at the subchondral bone spectral data's influence on this subset's PCA results). The most prominent loading originated from the ν_1 PO_4^{3-} band in the 960 cm^{-1} region; there were more subdued loadings from the ν_2 PO_4^{3-} band around ~ 430 cm^{-1}, ν_4 PO_4^{3-} band ~ 590 cm^{-1}, ν_3 PO_4^{3-} band ~ 1030 cm^{-1}, Amide III ~ 1250 cm^{-1}, alkyl (CH_2) deformation ~ 1490 cm^{-1}, and Amide I ~ 1600 cm^{-1}. The loadings plot for PC2 also appeared to show a mixture of contributions from both components (with the 'primary' loadings coming from the ν_1 PO_4^{3-} band ~ 960 cm^{-1}, ν_3 PO_4^{3-} band ~ 1030 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1490 cm^{-1}, and Amide I ~ 1600 cm^{-1}). Loadings for PC3 seemed to represent more of a contribution from the collagenous component – especially those originating from the Amide III, CH_2 deformation, and Amide I functionalities, around ~ 1250 cm^{-1}, ~ 1490 cm^{-1}, and ~ 1600 cm^{-1}, respectively. To a lesser extent, there were also loadings from proline, hydroxyproline, and phenylalanine around ~ 920 cm^{-1}, ~ 930 cm^{-1}, and ~ 1004 cm^{-1}, respectively. Likewise, the loadings plot for PC4 seemed to represent more contributions from the collagenous component; loadings from hydroxyproline around ~ 930 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1} stood out more. The loadings plot for PC5 also looked like mixed contributions from both components, with the ν_2

PO_4^{3-} band ~ 430 cm^{-1} and Amide I ~ 1600 cm^{-1} a little more evident in this PC, alongside loadings from hydroxyproline ~ 930 cm^{-1}, Amide III ~ 1250 cm^{-1}, and CH_2 deformation ~ 1480 cm^{-1}. Loadings for PC6 were considerably noisier, likely representing a contribution from the Amide I functionality (~ 1600 cm^{-1}) of the organic matrix.

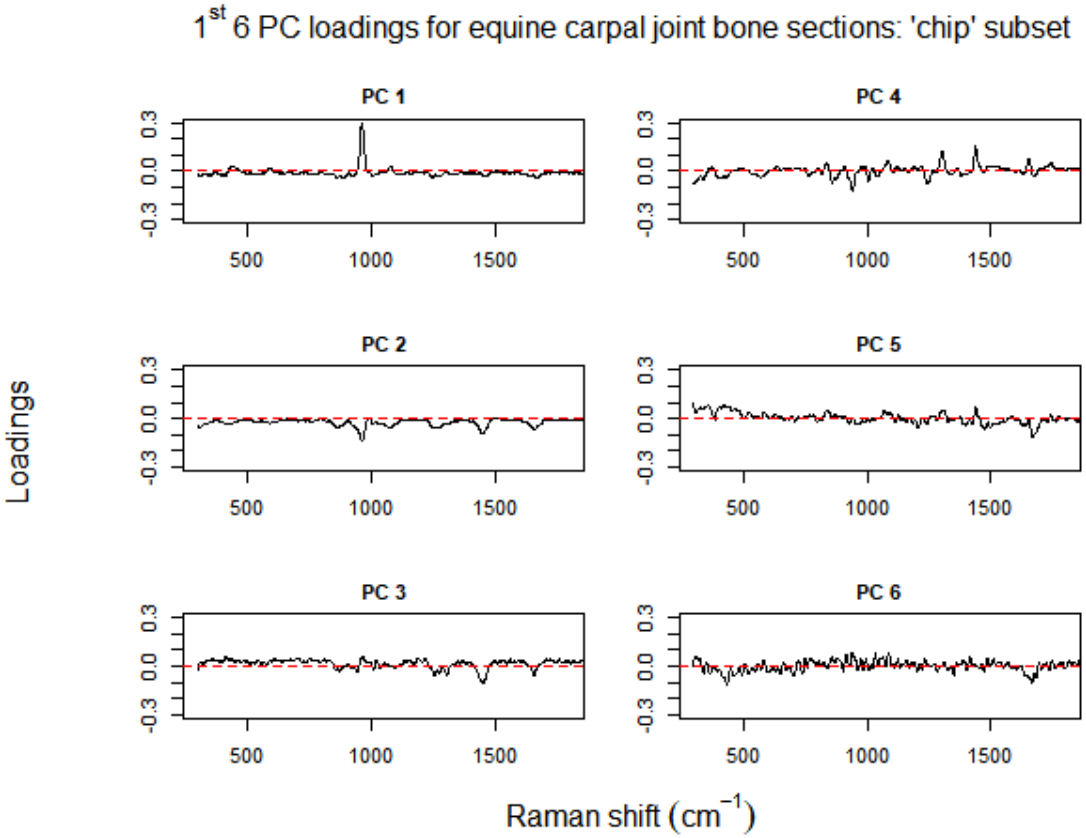

Figure 4.10 The first six PCs' loadings plots for the aggregated 'chip' spectral data from the equine carpal joint bone sections. PC1 closely resembled the original spectra with respect to both the mineral and organic matrix components. PC2 loadings also appeared to be a mixture of contributions from both components; PCs 3 and four seemed to represent more contributions from the matrix component. Though the loadings for PC5 also looked like a mixed contribution from both components, those for PC6 were noisier in appearance, and vaguely suggestive of another contribution from the organic matrix.

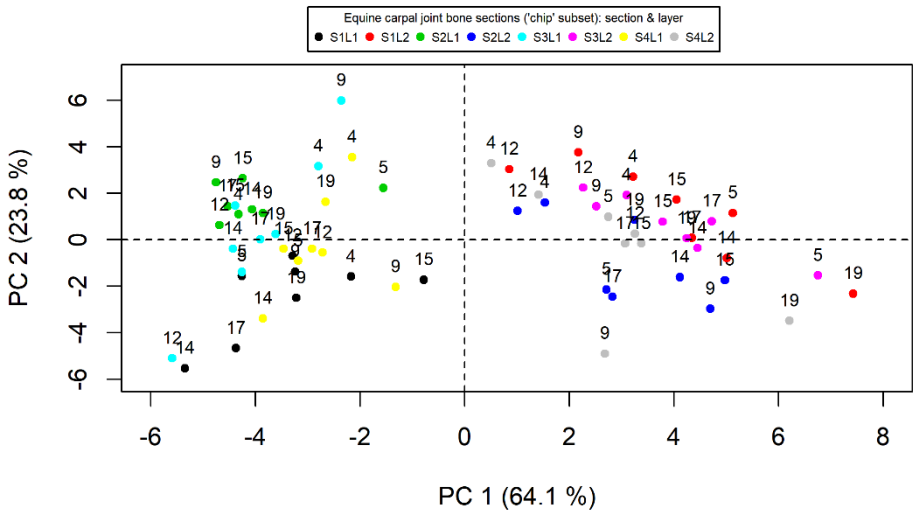

Figure 4.11 Scores plot for PC2 against PC1 from the aggregated 'chip' spectral data subset (by section and layer): S1L1 (black); S1L2 (red); S2L1 (green); S2L2 (blue); S3L1 (cyan); S3L2 (magenta); S4L1 (yellow); S4L2 (grey).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L1 = articular cartilage (AC); L2 = subchondral bone (SCB);
horses 4, 5, 9, 12, 14, 15, & 17 = two-year-old horses; horse 19 = three-year-old horse.

Figure 4.11 shows the scores plot for the second PC against the first for the eight combinations of section and layer from the aggregated 'chip' spectral data subset from the articular cartilage and subchondral bone layers of the equine carpal joint bone specimens. Again, PCA was capable of separating the sections and layers into clusters to bring out some of the within-object variations.

The most immediately noticeable feature was that scores for the AC (L1) in all four sections (third carpal (S1), fourth carpal (S2), the upper part of the radial carpal (S3), and the lower part of the radial carpal (S4); coloured black, green, cyan, and yellow, respectively) tended to cluster on the left-hand, negative side of the plot along the PC1 axis. Scores from the SCB (L2) in all four sections (coloured red, blue, magenta, and grey, respectively) tended to cluster along the right-hand, positive side of the plot along PC1. Score distribution along the PC2 axis was slightly more ambiguous than along the PC1 axis.

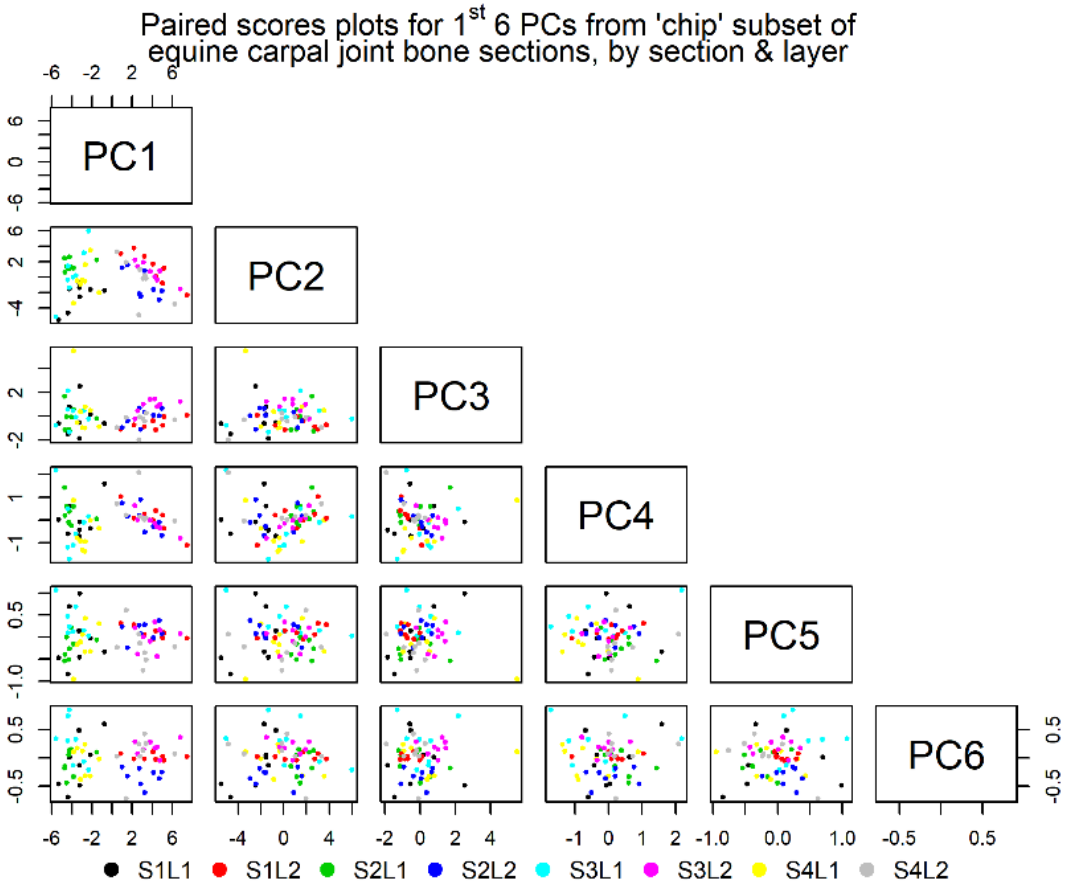

Figure 4.12 Paired scores plot matrix for the first six PCs from the aggregated equine carpal joint 'chip' spectral data subset (by section and layer): S1L1 (black); S1L2 (red); S2L1 (green); S2L2 (blue); S3L1 (cyan); S3L2 (magenta); S4L1 (yellow); S4L2 (grey).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L1 = articular cartilage (AC); L2 = subchondral bone (SCB).

Figure 4.12 shows a scatterplot matrix of the fifteen combinations of pairs of the first six PCs (from the eight combinations of section and layer) from the aggregated equine carpal joint 'chip' Raman spectral data. Surveying the scatterplot matrix, some of the spreading or grouping of the PC scores from the Raman spectral data was more obvious with pairs from the lower PCs than from those of the higher PCs. Like Figure 4.11, the PC scores from each section were more easily separated – by layer – along PC1 than along the other PCs. Overall, score distribution along these other PCs' axes was, again, more unclear.

4.2.2.1.2 PCA of the articular cartilage (AC) spectral data subset of the 'chip' spectral data subset

Results from the PCA of the articular cartilage (AC) subset of the aggregated 'chip' subset from the equine carpal joint bone sections are presented below.

Figure 4.13 shows the scree plot of the variances for the first 10 PCs from the articular cartilage (AC) sub-subset. The first six PCs accounted for approximately 94.1 % of the total variation in the sub-subset, again, rapidly tailing off after PC4. PC1 explained about 56.3 % of the total variation, whilst PCs 2 through six accounted for roughly 18.2 %, 10.6 %, 5.7 %, 2.1 %, and 1.3 %, respectively.

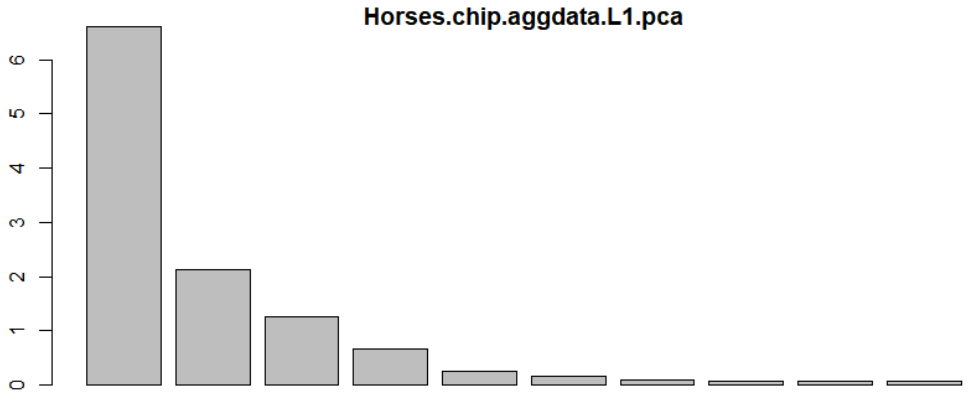

Figure 4.13 Scree plot for the articular cartilage (AC) Raman spectral data subset of the aggregated 'chip' data subset from the equine carpal joint bone sections. Cumulatively, the first six PCs accounted for approximately 94.1 % of the total spectral variation; individually, they explained about 56.3 %, 18.2 %, 10.6 %, 5.7 %, 2.1 %, and 1.3 %, respectively.

Figure 4.14 shows the loadings plots for the first six PCs (again, based on the covariance matrix). Loadings plots for PCs 1, two, and four appeared to represent contributions from the organic matrix. The loadings plot for PCs 3 and five, though, seemed to be denoting a mixed contribution from the mineral and matrix components. Along similar lines to PC6 loadings from the original aggregated 'chip' subset in Figure 4.10, loadings for PC6 were noisier than the previous PCs. The 'major' loadings for PCs 1 and two were from proline around ~ 920 cm^{-1}, hydroxyproline ~ 930 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1}. The ν_1 PO_4^{3-} band ~ 960 cm^{-1} dominated the loadings for PC3, with less intense loadings from the ν_2 PO_4^{3-} band ~ 450 cm^{-1}, ν_4 PO_4^{3-} band ~ 590 cm^{-1}, ν_3 PO_4^{3-} band ~ 1030 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1}. Like the equivalent loadings from the original aggregated 'chip' subset in Figure 4.10, loadings from hydroxyproline ~ 930 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1} were a bit more marked for PC4. The loadings from hydroxyproline ~ 930 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide

I ~ 1600 cm^{-1} for PC5 somewhat resembled those from the original aggregated 'chip' subset in Figure 4.10. The hydroxyproline and Amide I functionalities (~ 930 cm^{-1} and 1600 cm^{-1}, respectively) were the only easily discernible loadings for PC6.

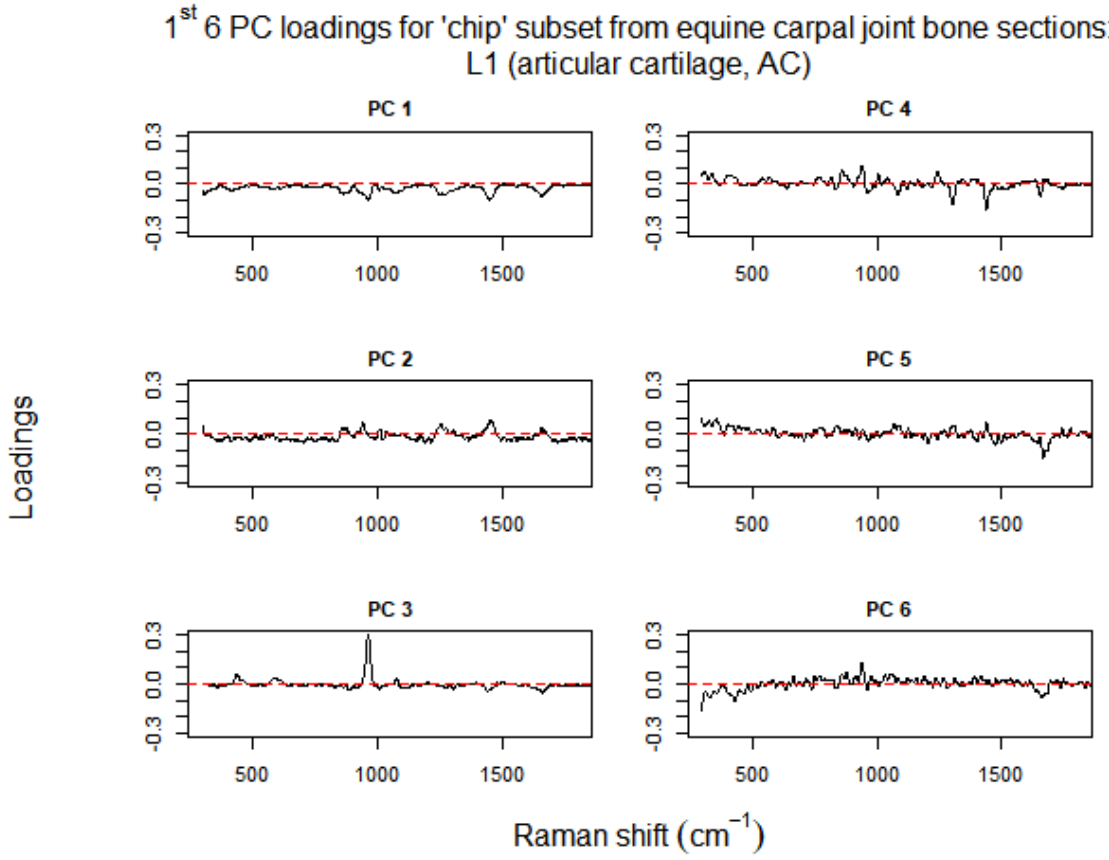

Figure 4.14 The first six PCs' loadings plots for the articular cartilage (AC) Raman spectral data subset of the aggregated 'chip' data subset from the equine carpal joint bone sections. The loadings for PCs 1, two, and four appeared to be representative of contributions from the organic matrix, whereas those from PCs 3 and five seemed to represent mixed contributions from the mineral and matrix components. The loadings for PC6 were much noisier.

Figure 4.15 shows the scores plot for the second PC against the first for the four combinations of section and layer from the AC sub-subset of the aggregated equine carpal joint 'chip' Raman spectral data subset. PCA was capable of separating the sections and layer; the data had been subdivided to the point that there were no longer any recognisable patterns in either PC dimension.

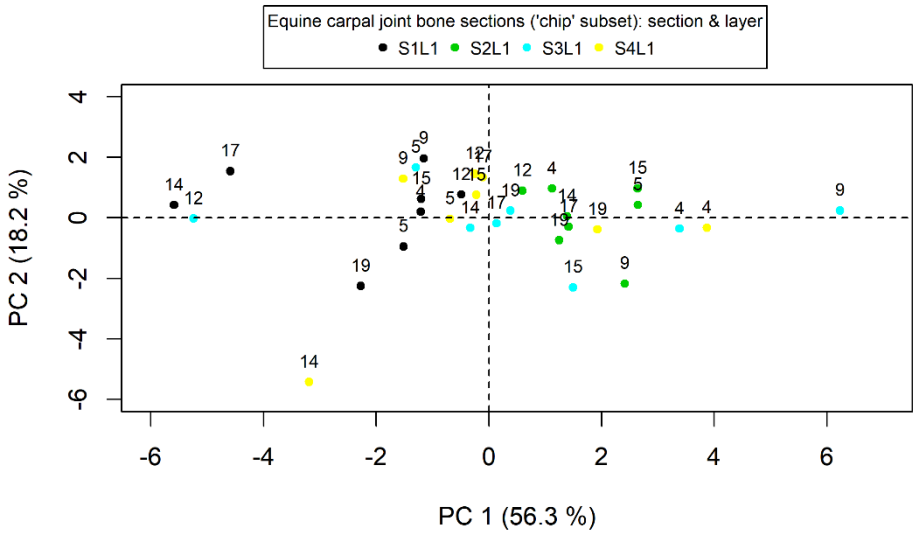

Figure 4.15 Scores plot for PC2 against PC1 for the articular cartilage (AC) Raman spectral data subset of the aggregated 'chip' data subset (by section and layer): S1L1 (black); S2L1 (green); S3L1 (cyan); S4L1 (yellow).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L1 = articular cartilage (AC);
horses 4, 5, 9, 12, 14, 15, & 17 = two-year-old horses; horse 19 = three-year-old horse.

Figure 4.16 shows a scatterplot matrix of the fifteen combinations of pairs of the first six PCs (from the four combinations of section and layer) from the AC sub-subset. Like Figure 4.15, there were no longer any discernible clusters of the PC scores because of the Raman spectral data being extensively subdivided by this point.

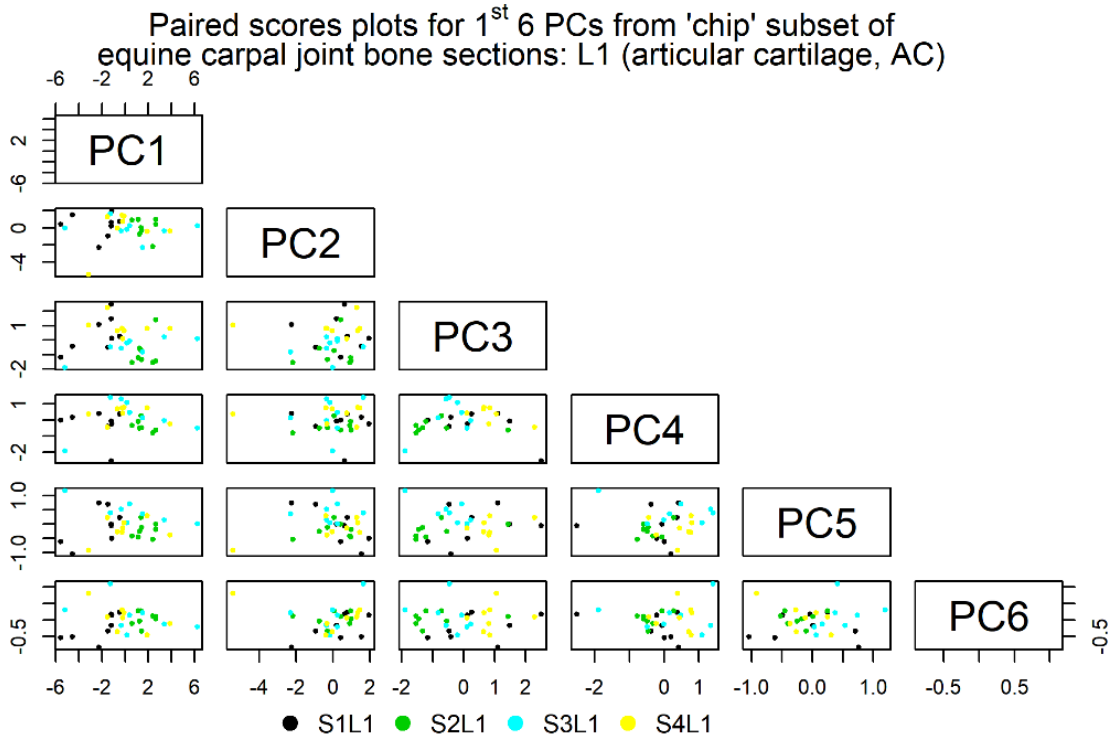

Figure 4.16 Paired scores plot matrix for the first six PCs for the articular cartilage (AC) Raman spectral data subset of the aggregated equine carpal joint 'chip' data subset (by section and layer): S1L1 (black); S2L1 (green); S3L1 (cyan); S4L1 (yellow).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L1 = articular cartilage (AC).

4.2.2.1.3 PCA of the subchondral bone (SCB) spectral data subset of the 'chip' spectral data subset

Results from the PCA of the subchondral bone (SCB) subset of the aggregated 'chip' subset from the equine carpal joint bone sections are displayed below.

Figure 4.17 shows the scree plot of the variances for the first 10 PCs from the subchondral bone (SCB) sub-subset. The first six PCs accounted for approximately 97.3 % of the total variation in the sub-subset, rapidly tailing off after PC3. PC1 explained about 66.3 % of the total variation, while PCs 2 through six accounted for approximately 21.9 %, 5.6 %, 1.6 %, 1.1 %, and 0.799 %, respectively.

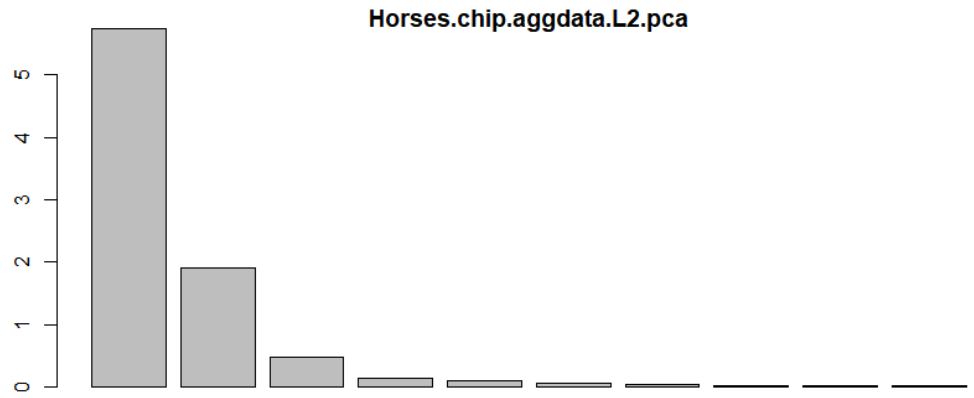

Figure 4.17 Scree plot for the subchondral bone (SCB) Raman spectral data subset of the aggregated 'chip' data subset from the equine carpal joint bone sections. Cumulatively, the first six PCs accounted for approximately 97.3 % of the total spectral variation; individually, they explained about 66.3 %, 21.9 %, 5.6 %, 1.6 %, 1.1 %, and 0.8 %, respectively.

Figure 4.18 shows the loadings plot for the first six PCs (again, based on the covariance matrix). The loadings plot for PC1 appeared to represent a mixed contribution from the mineral and organic matrix components (again, much like the original spectra from the subchondral bone). Like the PC1 loadings plot from the original aggregated 'chip' subset in Figure 4.10, the main loading for PC1 was from the ν_1 PO_4^{3-} band in the 960 cm^{-1} region. There were more subdued loadings from the ν_2 PO_4^{3-} band around ~ 430 cm^{-1}, ν_4 PO_4^{3-} band ~ 590 cm^{-1}, ν_3 PO_4^{3-} band ~ 1030 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1490 cm^{-1}, and Amide I ~ 1600 cm^{-1}. The loadings plot for PC2 also appeared to be a mixed contribution from both components, with the primary loadings originating from the ν_1 PO_4^{3-} band around ~ 960 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1490 cm^{-1}, and Amide I ~ 1600 cm^{-1}. There were more muted loadings from the proline, hydroxyproline, and ν_3 PO_4^{3-} functionalities around ~ 920 cm^{-1}, ~ 930 cm^{-1}, and ~ 1030 cm^{-1}, respectively. The loadings plots for PCs 3, four, and six were rather noisy, though they seemed vaguely suggestive of another mixed contribution from both components. For PC3, the more perceptible loadings looked to be the ν_1 PO_4^{3-} band around ~ 960 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1490 cm^{-1}, and Amide I ~ 1600 cm^{-1}. For PC4, they were the ν_4 PO_4^{3-} band ~ 590 cm^{-1} and CH_2 deformation ~ 1490 cm^{-1}. For PC6, there was an almost derivative-like shape to the ν_2 PO_4^{3-} and ν_4 PO_4^{3-} bands around ~ 480 – 575 cm^{-1}. The loadings plot for PC5 seemed to be yet another mixed contribution from both components, with subtle loadings from the ν_2 PO_4^{3-} band around ~ 430 cm^{-1}, proline

~ 920 cm^{-1}, hydroxyproline ~ 930 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1490 cm^{-1}, and Amide I ~ 1600 cm^{-1}.

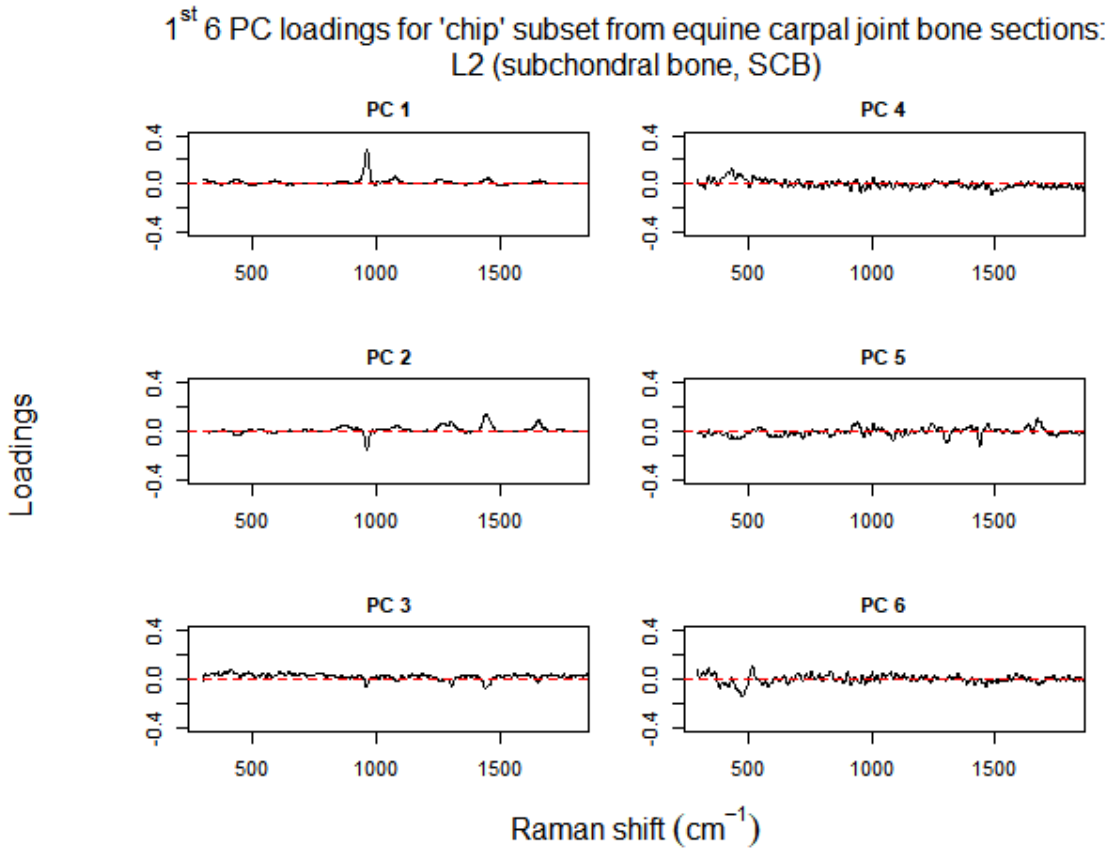

Figure 4.18 The first six PCs' loadings plots for the subchondral bone (SCB) Raman spectral data subset of the aggregated 'chip' data subset from the equine carpal joint bone sections. PC1 closely resembled the original spectra from the SCB with respect to both the mineral and organic matrix components. PCs 2 through six also appeared to be a mixed contribution from both components, though loadings for PCs 3, four, and six were noisier than the others.

Figure 4.19 shows the scores plot for the second PC against the first for the four combinations of section and layer from the SCB sub-subset of the aggregated equine carpal joint 'chip' Raman spectral data subset. As was also the case with the AC sub-subset in subsection 4.2.2.1.2, there were no longer any recognisable patterns in either PC dimension even though PCA was once again capable of separating the sections and layer.

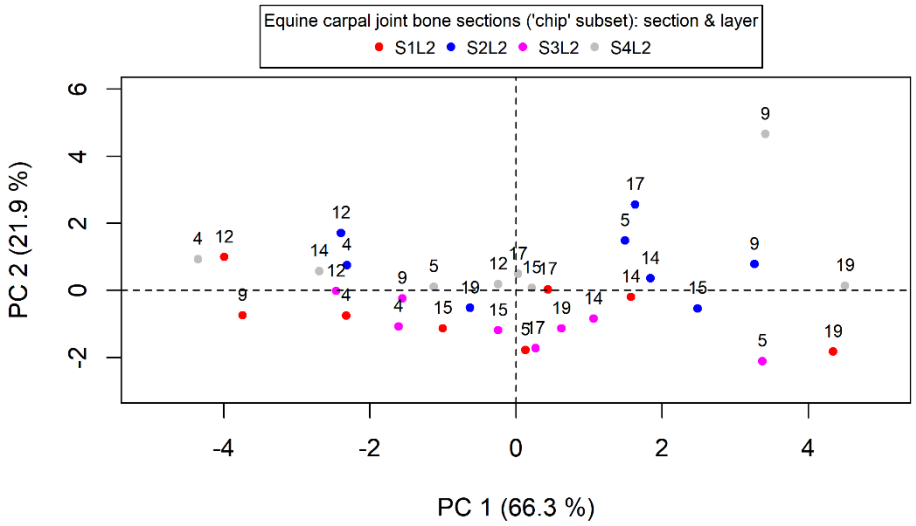

Figure 4.19 Scores plot for PC2 against PC1 for the subchondral bone (SCB) Raman spectral data subset of the aggregated 'chip' data subset (by section and layer): S1L2 (red); S2L2 (blue); S3L2 (magenta); S4L2 (grey).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L2 = subchondral bone (SCB);
horses 4, 5, 9, 12, 14, 15, & 17 = two-year-old horses; horse 19 = three-year-old horse.

Figure 4.20 shows a scatterplot matrix of the fifteen combinations of pairs of the first six PCs (from the four combinations of section and layer) from the SCB sub-subset. Again, like Figure 4.19, there were no longer any distinct clusters of the PC scores because of the extensive subdivision of the Raman spectral data by this point.

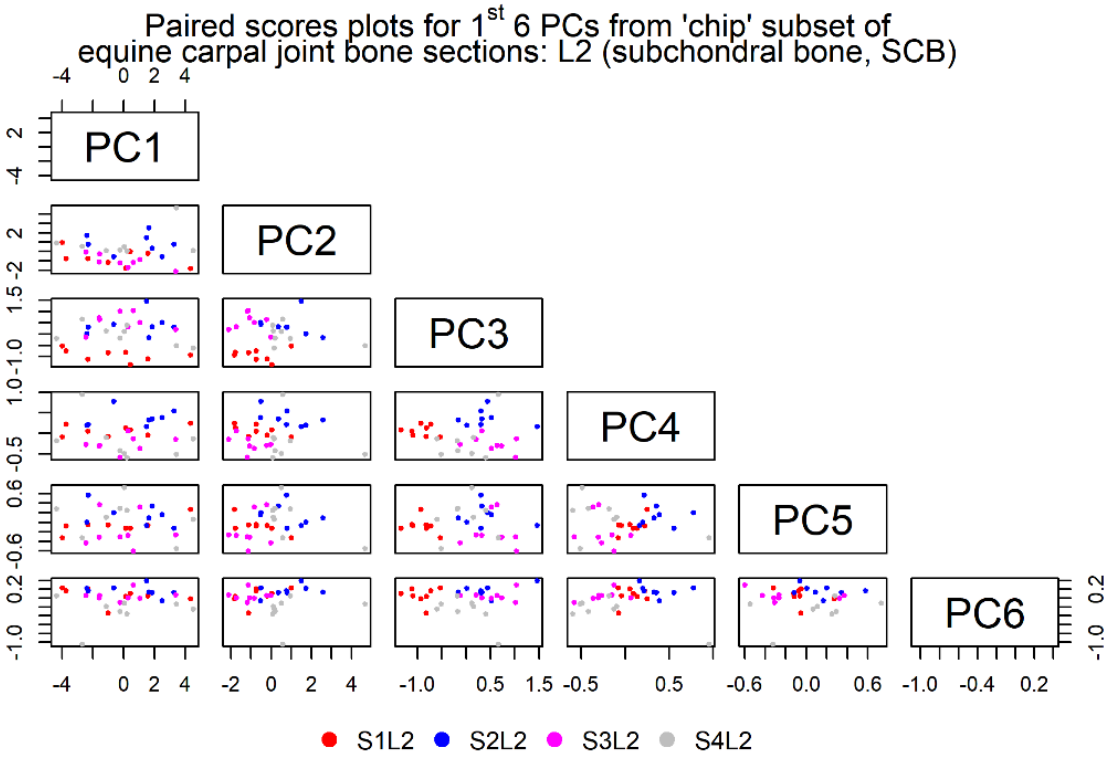

Figure 4.20 Paired scores plot matrix for the first six PCs for the subchondral bone (SCB) Raman spectral data subset of the aggregated equine carpal joint 'chip' data subset (by section and layer): S1L2 (red); S2L2 (blue); S3L2 (magenta); S4L2 (grey).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L2 = subchondral bone (SCB).

4.2.2.2.1 PCA of the 'control' spectral data subset

Results from the PCA of the aggregated 'control' spectral data subset from the equine carpal joint bone sections follow on. As with the 'chip' subset, incorporation and averaging of additional information about the horses before PC analysis was necessary to make within-horse variability more discernible.

Figure 4.21 shows the scree plot of the variances for the first 10 PCs from the 'control' subset. The first six PCs accounted for approximately 97.3 % of the total variation in the data subset, again, rapidly tailing off after PC4. PC1 explained about 71.8 % of the total variation, while PCs 2 through six comprised approximately 16.5 %, 5.2 %, 2.6 %, 0.723 %, and 0.473 %, respectively.

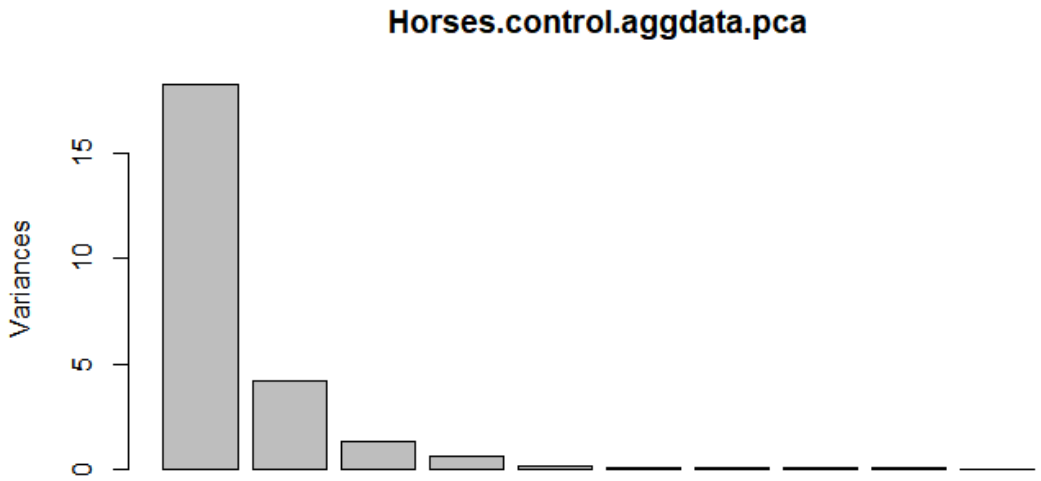

Figure 4.21 Scree plot for the aggregated 'control' spectral data subset from the equine carpal joint bone sections. Cumulatively, the first six PCs accounted for approximately 97.3 % of the total spectral variation; individually, they explained about 71.8 %, 16.5 %, 5.2 %, 2.6 %, 0.7 %, and 0.5 %, respectively.

Figure 4.22 shows the loadings plots for the first six PCs. As with the 'chip' subset, the loadings for PC1 from the 'control' subset also closely resembled the original Raman spectra – especially those from the SCB. The dominant loading originated from the ν_1 PO_4^{3-} band within the 960 cm^{-1} region. The other loadings for PC1 were from the ν_2 PO_4^{3-} band around ~ 430 cm^{-1}, ν_4 PO_4^{3-} band ~ 590 cm^{-1}, ν_3 PO_4^{3-} band ~ 1030 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1490 cm^{-1}, and Amide I ~ 1600 cm^{-1}. The loadings plot for PC2 also appeared to show a mixture of contributions from both components (again, with the 'primary' loadings coming from the ν_1 PO_4^{3-} band ~ 960 cm^{-1}, ν_3 PO_4^{3-} band ~ 1030 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1490 cm^{-1}, and Amide I ~ 1600 cm^{-1}). The loadings plots for PCs 3 and four seemed to represent more of a contribution from the collagenous component. For PC3 these came from the Amide III, CH_2 deformation, and Amide I functionalities in particular, around ~ 1250 cm^{-1}, ~ 1480 cm^{-1}, and ~ 1600 cm^{-1}, respectively. To a lesser extent, there were also loadings from the proline, hydroxyproline, and phenylalanine around ~ 920 cm^{-1}, ~ 930 cm^{-1}, and ~ 1004 cm^{-1}, respectively. Loadings from hydroxyproline around ~ 930 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1} were the 'primary' loadings for PC4. Interestingly, many of the contributory loadings for the second, third, and fourth PCs had opposite signs to their 'chip' equivalents. The loadings plot for PC5 also looked like a contribution from the matrix component; the major loadings came from proline, hydroxyproline, CH_2 deformation, and Amide I, around ~ 920 cm^{-1}, ~ 930 cm^{-1}, ~ 1480 cm^{-1}

and ~ 1660 cm^{-1}, respectively. Much like its 'chip' equivalent, the loadings plot for PC6 was much noisier, likely representing contribution from the ν_4 PO_4^{3-} band (~ 590 cm^{-1}) from the mineral component.

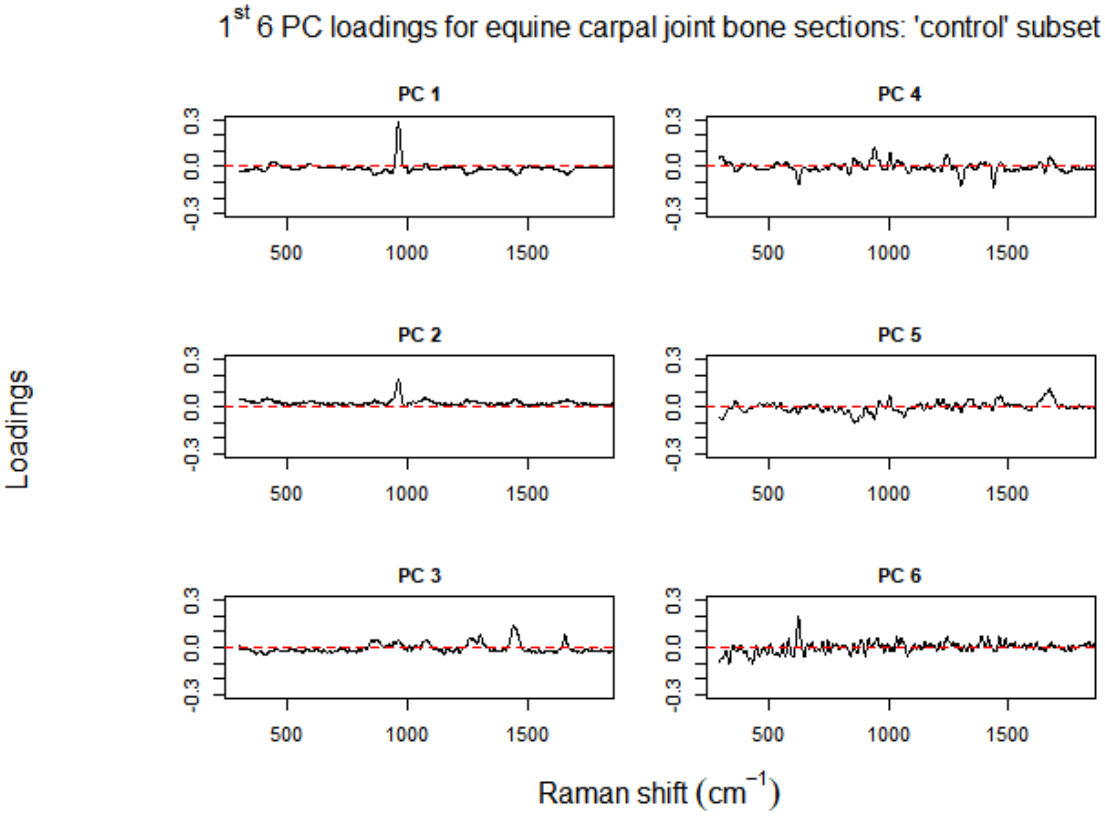

Figure 4.22 The first six PCs' loadings plots for the aggregated 'control' Raman spectral data from the equine carpal joint bone sections. Again, PC1 closely resembled the original spectra with respect to both the mineral and organic matrix components. PC2 loadings also appeared to be a mixture of contributions from both components; PCs 3 and four seemed to represent more contributions from the matrix component. PC5 also looked like a contribution from the organic matrix. Loadings for PC6 were noisier in appearance, and vaguely suggestive of another contribution from the mineral component.

Figure 4.23 shows the scores plot for the second PC against the first for the eight combinations of section and layer from the aggregated equine carpal joint 'control' spectral data subset from the articular cartilage and subchondral bone of the equine carpal joint bone specimens. Again, PCA was capable of separating the sections and layers into clusters to bring out some of the within-object variations.

As with the 'chip' subset, the scores for the AC (L1) in all four sections (third carpal (S1), fourth carpal (S2), the upper part of the radial carpal (S3), and the lower part of the radial carpal (S4); coloured black, green, cyan, and yellow, respectively) from the 'control' subset tended to cluster along the negative side of PC1. Scores from the SCB (L2) in all four sections

(coloured red, blue, magenta, and grey, respectively) clustered along the positive side of PC1. Score distribution along the PC2 axis was, again, vaguer than along the PC1 axis.

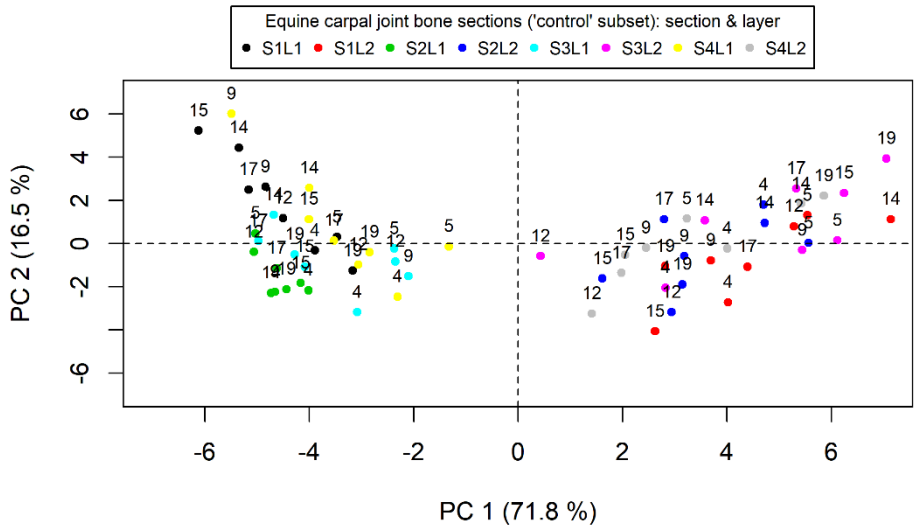

Figure 4.23 Scores plot for PC2 against PC1 from the aggregated 'control' spectral data subset (by section and layer): S1L1 (black); S1L2 (red); S2L1 (green); S2L2 (blue); S3L1 (cyan); S3L2 (magenta); S4L1 (yellow); S4L2 (grey).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L1 = articular cartilage (AC); L2 = subchondral bone (SCB);
horses 4, 5, 9, 12, 14, 15, & 17 = two-year-old horses; horse 19 = three-year-old horse.

Figure 4.24 shows a scatterplot matrix of the fifteen combinations of pairs of the first six PCs (from the eight combinations of section and layer) from the aggregated equine carpal joint 'control' Raman spectral data subset. Spreading or grouping of some PCs, was, once again, more evident with pairs from the lower PCs than from the higher PCs. Like Figure 4.23, the scores from each section were more easily separated – by layer – along PC1 than along the other PCs. Overall score distribution along these other PCs' axes was once more unclear.

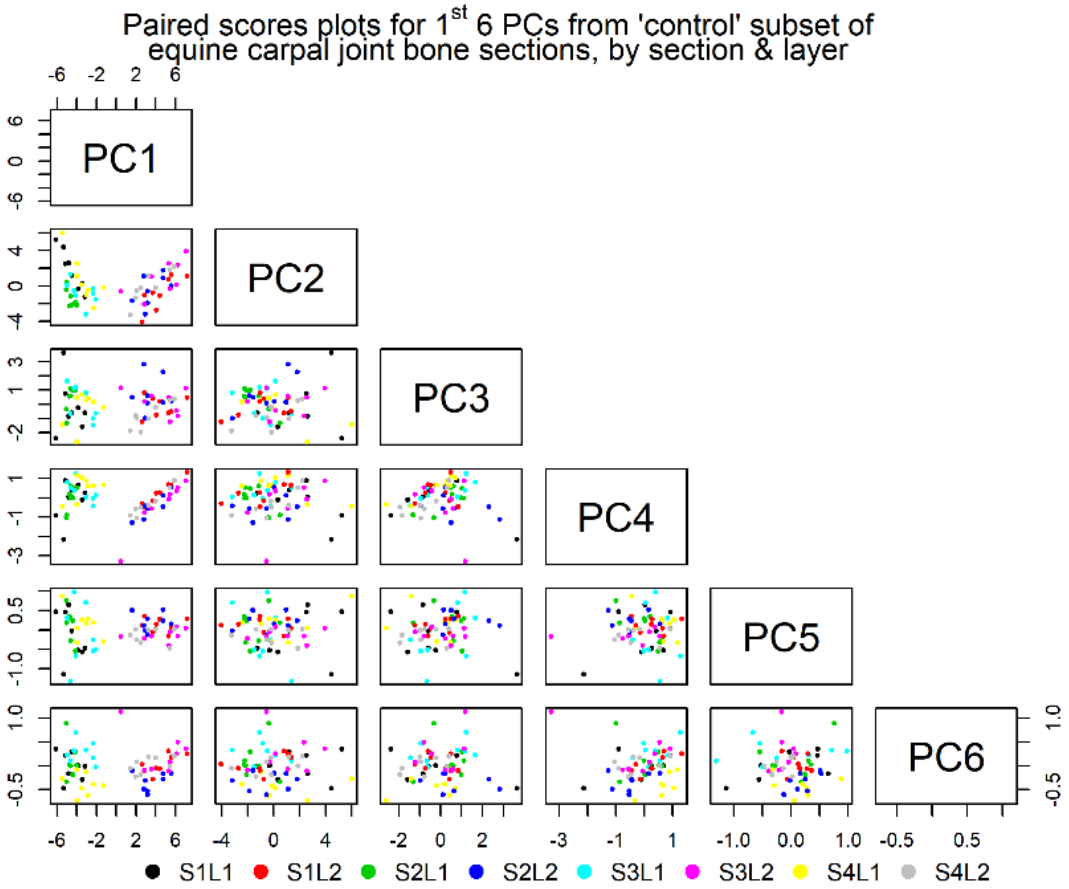

Figure 4.24 Paired scores plot matrix for the first six PCs from the aggregated 'control' spectral data subset (by section and layer): S1L1 (black); S1L2 (red); S2L1 (green); S2L2 (blue); S3L1 (cyan); S3L2 (magenta); S4L1 (yellow); S4L2 (grey).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L1 = articular cartilage (AC); L2 = subchondral bone (SCB).

4.2.2.2.2 PCA of the articular cartilage (AC) spectral data subset of the 'control' spectral data subset

Results from the PCA of the articular cartilage (AC) subset of the aggregated 'control' subset from the equine carpal joint bone sections are set out below.

Figure 4.25 shows the scree plot of the variances for the first 10 PCs from the articular cartilage (AC) sub-subset. The first six PCs accounted for approximately 91.9 % of the total spectral variation in the sub-subset, again, rapidly tailing off after PC4. PC1 explained about 58.4 % of the total spectral variation, whilst PCs 2 through six accounted for roughly 17 %, 7.7 %, 4.5 %, 2.8 %, and 1.5 %, respectively.

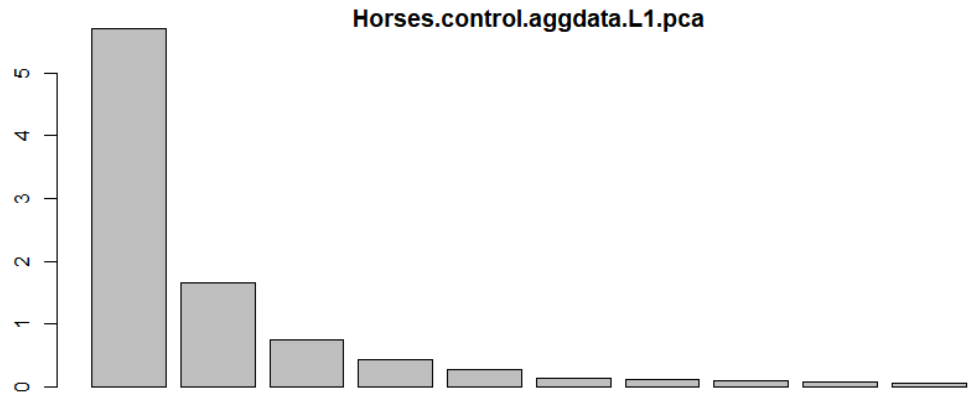

Figure 4.25 Scree plot for the articular cartilage (AC) Raman spectral data subset of the aggregated 'control' data subset from the equine carpal joint bone sections. Cumulatively, the first six PCs accounted for approximately 91.9 %; individually, they explained about 58.4 %, 17 %, 7.7 %, 4.5 %, 2.8 %, and 1.5 %, respectively.

Figure 4.26 shows the loadings plots for the first six PCs. Loadings plots for PCs 1, two, four, and five appeared to represent contributions from the organic matrix. The loadings plot for PC3, though, seemed to be denoting a mixed contribution from both the mineral and matrix components. Loadings for PC6 were much noisier than the preceding PCs; it was difficult to extract any useful information from it. Loadings from PC1 were subtler than for PC2, hinting at contributions from the phenylalanine around ~ 1004 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1}. The 'primary' loadings for PC2 were from the proline around ~ 920 cm^{-1}, hydroxyproline ~ 930 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1}. The ν_1 PO_4^{3-} band in the 960 cm^{-1} region was the most noticeable loading for PC3; the ν_2 PO_4^{3-} band ~ 450 cm^{-1}, ν_4 PO_4^{3-} band ~ 590 cm^{-1}, ν_3 PO_4^{3-} band ~ 1030 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1} loadings were not as substantial. Loadings from phenylalanine around ~ 1004 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1} somewhat resembled the equivalent PC4 loadings from the original aggregated 'control' subset in Figure 4.22. Curiously, like the original aggregated 'control' subset there seemed to be something of a change in the shape of the Amide III band, suggesting a possible change in some part of the collagen's secondary structure. The proline, hydroxyproline, and Amide I functionalities (~ 920 cm^{-1}, ~ 930 cm^{-1}, and ~ 1600 cm^{-1}, respectively) were the more recognisable features in the PC5 loadings. There were also – to a lesser extent – the phenylalanine ~ 1004 cm^{-1}, and Amide III ~ 1250 cm^{-1}.

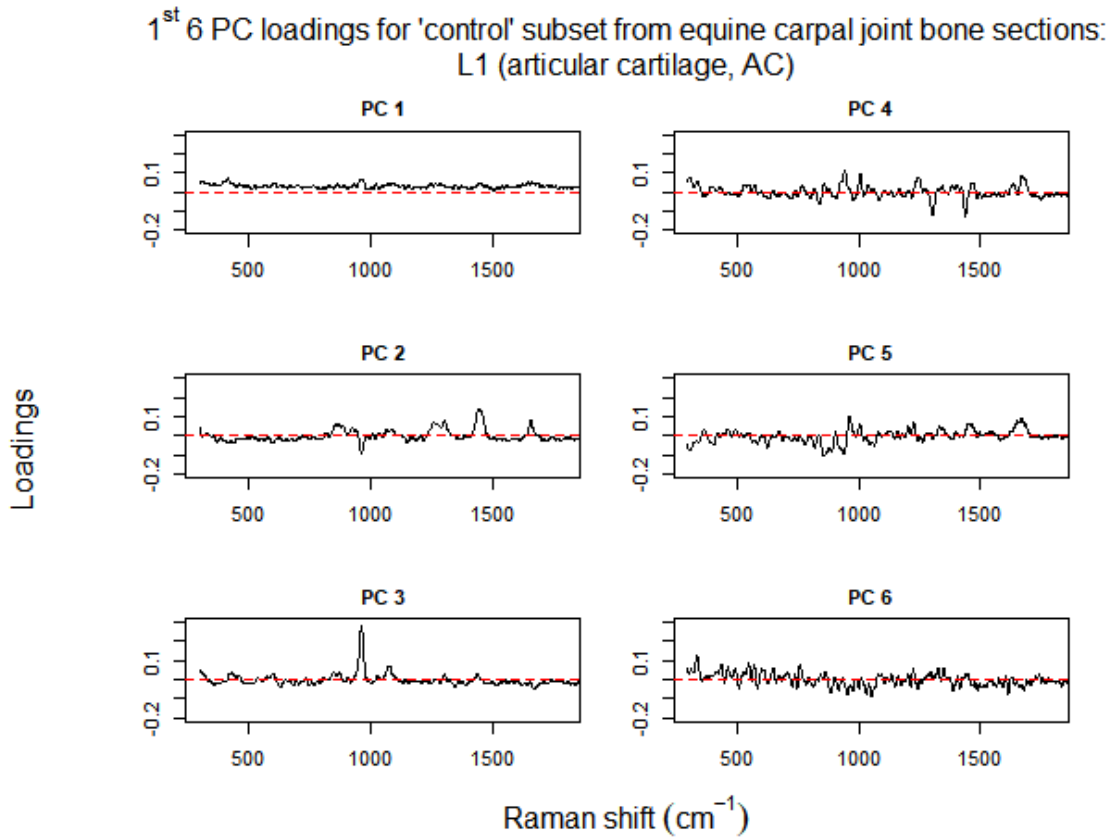

Figure 4.26 The first six PCs' loadings plots for the articular cartilage (AC) Raman spectral data subset of the aggregated 'control' data subset from the equine carpal joint bone sections. PCs 1, two, four, and five appeared to be representative of contributions from the organic matrix, and those from PC3 seemed to represent mixed contributions from both the mineral and matrix components. The loadings for PC6 were much noisier.

Figure 4.27 shows the scores plot for the second PC against the first for the four combinations of section and layer from the AC sub-subset of the aggregated equine carpal joint 'control' Raman spectral data subset. As was the case with the 'chip' subset, PCA was again capable of separating the sections and layer; the data had been subdivided to the point that there were no longer any recognisable patterns in either PC dimension.

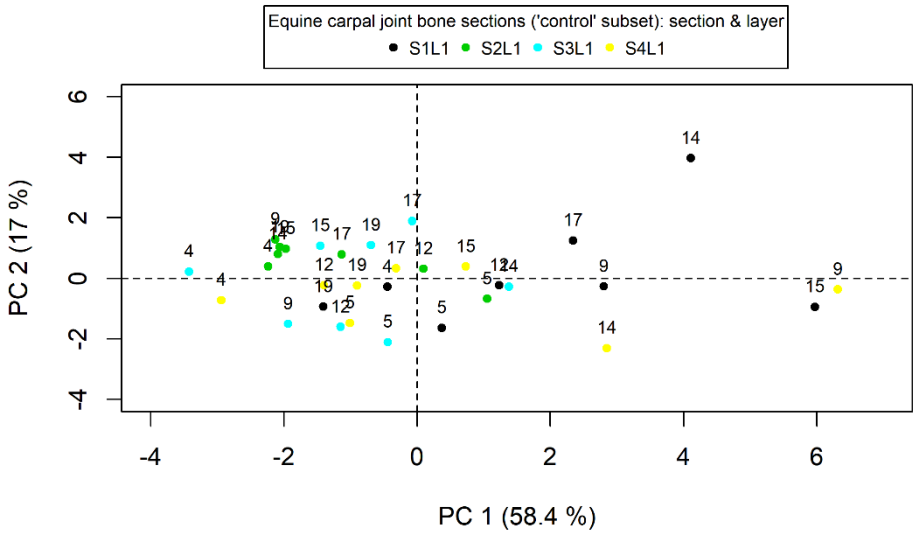

Figure 4.27 Scores plot for PC2 against PC1 for the articular cartilage (AC) Raman spectral data subset of the aggregated 'control' data subset (by section and layer): S1L1 (black); S2L1 (green); S3L1 (cyan); S4L1 (yellow).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L1 = articular cartilage (AC);
horses 4, 5, 9, 12, 14, 15, & 17 = two-year-old horses; horse 19 = three-year-old horse.

Figure 4.28 shows a scatterplot matrix of the fifteen combinations of pairs of the first six PCs (from the four combinations of section and layer) from the AC sub-subset. Like Figure 4.27, there were no longer any discernible clusters of the PC scores because of the Raman spectral data being extensively subdivided by this point.

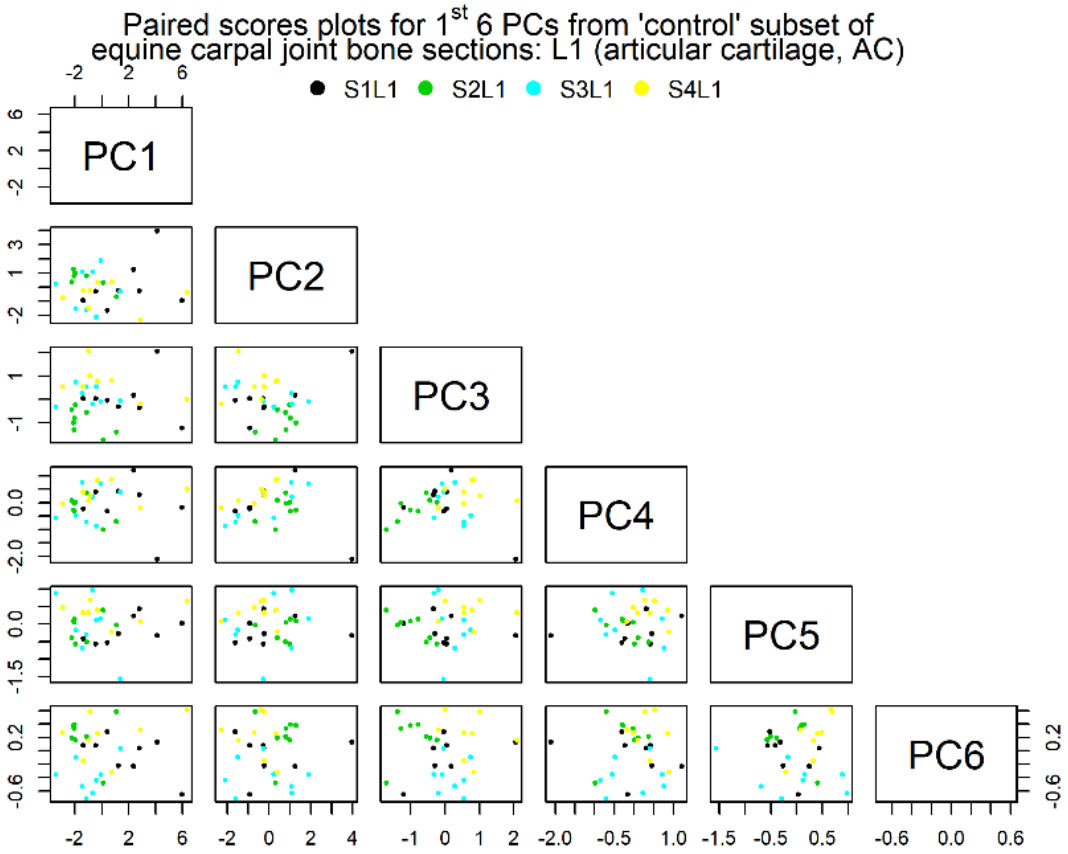

Figure 4.28 Paired scores plot matrix for the first six PCs for the articular cartilage (AC) Raman spectral data subset of the aggregated equine carpal joint 'control' data subset (by section and layer): S1L1 (black); S2L1 (green); S3L1 (cyan); S4L1 (yellow).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L1 = articular cartilage (AC).

4.2.2.2.3 PCA of the subchondral bone (SCB) spectral data subset of the 'control' spectral data subset

Results from the PCA of the subchondral bone (SCB) subset of the aggregated 'control' subset from the equine carpal joint bone sections are set out below.

Figure 4.29 shows the scree plot of the variances for the first 10 PCs from the subchondral bone (SCB) sub-subset. The first six PCs accounted for approximately 96.7 % of the total variation in the sub-subset, rapidly tailing off after PC3. PC1 explained about 65.9 % of the total variation, while PCs 2 through six accounted for around 19.4 %, 7.9 %, 2 %, 0.817 %, and 0.682 %, respectively.

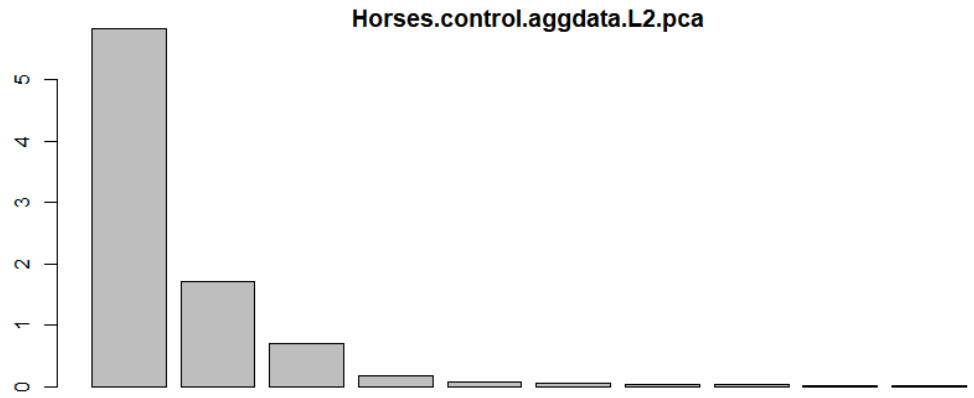

Figure 4.29 Scree plot for the subchondral bone (SCB) Raman spectral data subset of the aggregated 'control' data subset from the equine carpal joint bone sections. Cumulatively, the first six PCs accounted for approximately 96.7 % of the total spectral variation; individually, they explained about 65.9 %, 19.4 %, 7.9 %, 2 %, 0.8 %, and 0.7 %, respectively.

Figure 4.30 shows the loadings plots for the first six PCs. The loadings plots for five of the six PCs appeared to represent – to varying degrees – mixed contributions from both the mineral and organic matrix components (only the fifth PC appeared to represent a contribution from the mineral component). Additionally, many of these PCs' loadings were barely distinguishable from the baseline. The ν_1 PO_4^{3-} band in the 960 cm^{-1} region was the most noticeable loading for PC1; loadings from the ν_2 PO_4^{3-} band ~ 450 cm^{-1}, ν_4 PO_4^{3-} band ~ 590 cm^{-1}, and ν_3 PO_4^{3-} band ~ 1030 cm^{-1} were not as substantial. The Amide III band around ~ 1250 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1} were slightly more noticeable than the loadings from the ν_4 PO_4^{3-} band around ~ 590 cm^{-1}, proline ~ 920 cm^{-1}, hydroxyproline ~ 930 cm^{-1}, ν_1 PO_4^{3-} band ~ 960 cm^{-1}, and ν_3 PO_4^{3-} band ~ 1030 cm^{-1} for PC2. Loadings from the ν_1 PO_4^{3-} band around ~ 960 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1} barely stood out from the baseline for PC3. The loading from the ν_4 PO_4^{3-} band around ~ 590 cm^{-1} was the stand-out for PC4; the ν_2 PO_4^{3-} band ~ 450 cm^{-1}, ν_1 PO_4^{3-} band ~ 960 cm^{-1}, and Amide I ~ 1600 cm^{-1} were barely detectable. PC5 was also very noisy – there appeared to be a very subtle loading from the ν_4 PO_4^{3-} band around ~ 590 cm^{-1}. Lastly, there were subdued loadings from the ν_2 PO_4^{3-} band ~ 450 cm^{-1}, ν_4 PO_4^{3-} band ~ 590 cm^{-1}, proline ~ 920 cm^{-1}, hydroxyproline ~ 930 cm^{-1}, Amide III ~ 1250 cm^{-1}, and Amide I ~ 1600 cm^{-1} for PC6.

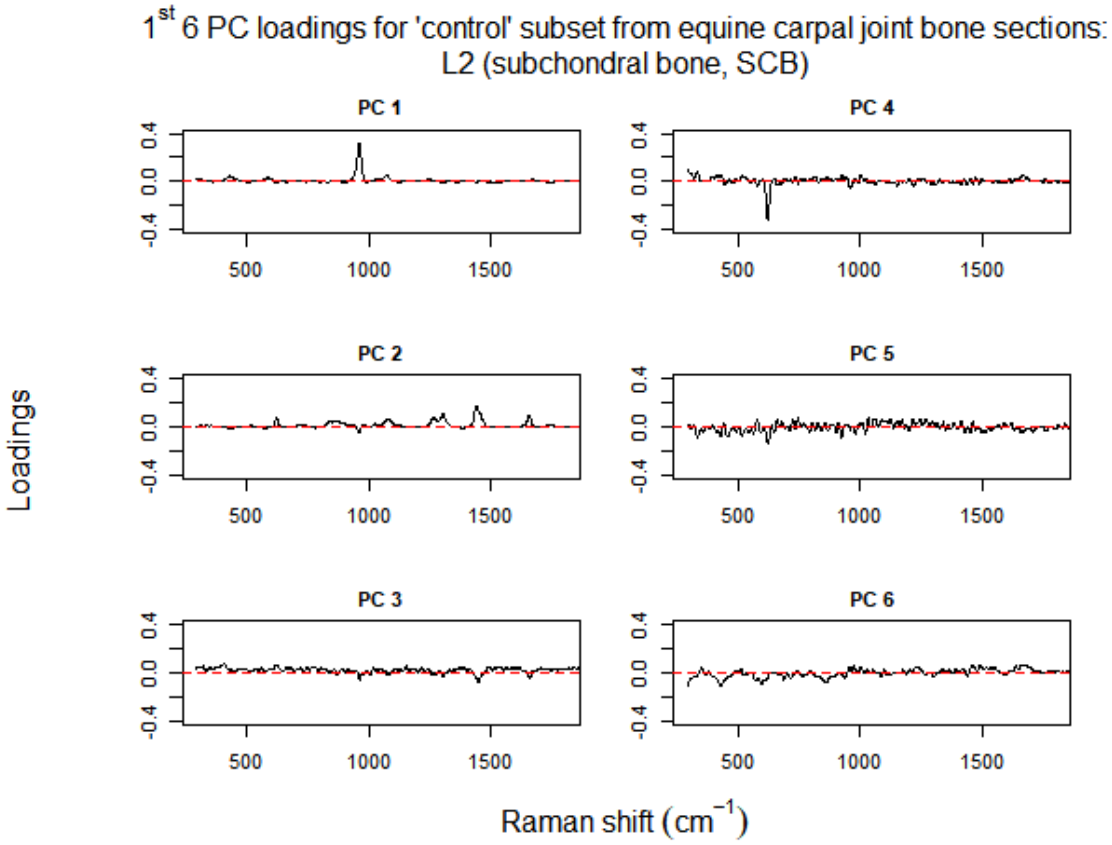

Figure 4.30 The first six PCs' loadings plots for the subchondral bone (SCB) Raman spectral data subset of the aggregated 'control' data subset from the equine carpal joint bone sections. PCs 1, two, three, four, and six seemed to represent mixed contributions from both the mineral and matrix components. PC5 represented a contribution from the mineral component; its loadings were much noisier.

Figure 4.31 shows the scores plot for the second PC against the first for the four combinations of section and layer from the SCB sub-subset of the aggregated equine carpal joint 'control' Raman spectral data subset. Likewise following its 'chip' counterparts (subsections 4.2.2.1.2 and 4.2.2.1.3) where there was an apparent lack of pattern, there were no longer any instantly recognisable patterns in either PC dimension even though PCA was once again capable of separating the sections and layer.

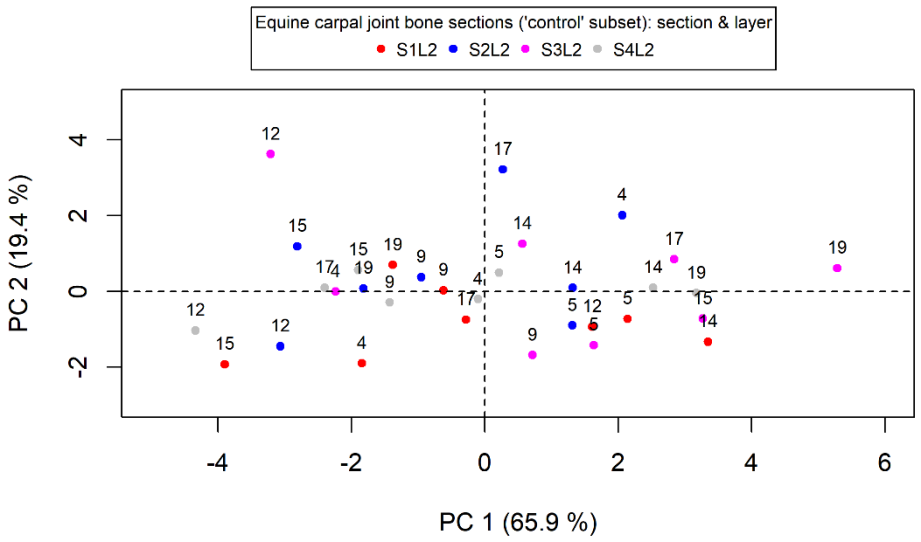

Figure 4.31 Scores plot for PC2 against PC1 for the subchondral bone (SCB) Raman spectral data subset of the aggregated 'control' data subset (by section and layer): S1L2 (red); S2L2 (blue); S3L2 (magenta); S4L2 (grey).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L2 = subchondral bone (SCB);
horses 4, 5, 9, 12, 14, 15, & 17 = two-year-old horses; horse 19 = three-year-old horse.

Figure 4.32 shows a scatterplot of the fifteen combinations of pairs of the first six PCs (from the four combinations of section and layer) from the SCB sub-subset. Again, like Figure 4.27, there were no longer any distinct clusters of the PC scores because of the extensive subdivision of the Raman spectral data by this point.

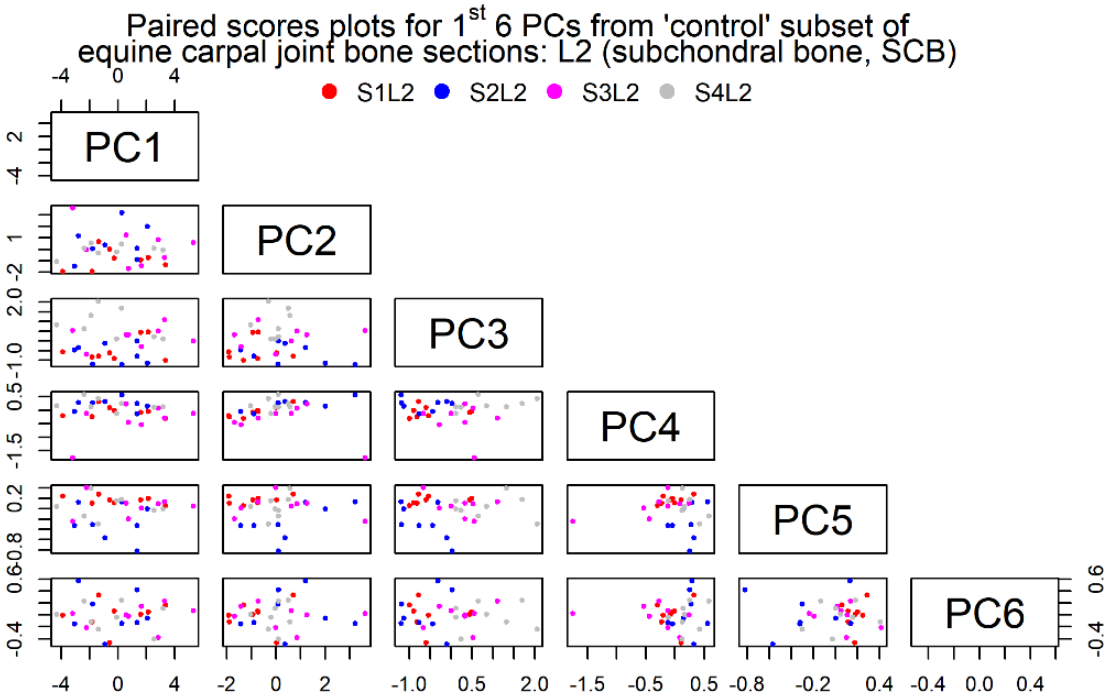

Figure 4.32 Paired scores plot matrix for the first six PCs for the subchondral bone (SCB) Raman spectral data subset of the aggregated equine carpal joint 'control' data subset (by section and layer): S1L2 (red); S2L2 (blue); S3L2 (magenta); S4L2 (grey).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L2 = subchondral bone (SCB).

4.2.2.3 PCA of the differenced spectral data

Results from the PCA of the differenced spectral data from the equine carpal joint bone sections are shown below.

Figure 4.33 shows the scree plot of the variances for the first 10 PCs from the differenced spectral data set. The first six PCs accounted for approximately 92.3 % of the total variation in the spectral data, again, rapidly tailing off after PC4. PC1 explained about 38 % of the total variation whilst PCs 2 through six accounted for approximately 24.7 %, 20.5 %, 5.8 %, 2.1 %, and 1.3 %, respectively.

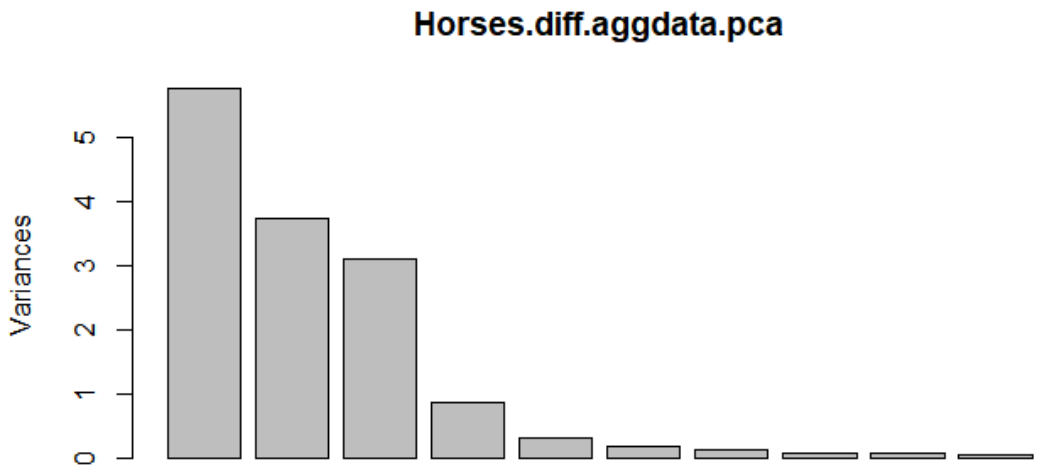

Figure 4.33 Scree plot for differenced Raman spectral data from the equine carpal joint bone sections. Cumulatively, the first six PCs accounted for approximately 92.3 % of the total spectral variation; individually, they explained about 38 %, 24.7 %, 20.5 %, 5.8 %, 2.1 %, and 1.3 %, respectively.

Figure 4.34 shows the loadings plots for the first six PCs from the differenced spectral data set. Loadings for the first two PCs both appeared to be representative of contributions from both components, dominated to some extent by the first symmetric stretch phosphate peak at ~ 960 cm^{-1}. PC1's loadings hinted at the likelihood of there being some functional group dissimilarities between the 'chip' and the 'control' subsets, especially concerning the loading for the first symmetric stretch of the phosphate. Within PC2 only the loadings for the first symmetric stretching and second and fourth bending vibrations from the phosphate (located around ~ 960 cm^{-1}, ~ 450 cm^{-1}, and ~ 590 cm^{-1}, respectively) had a closer resemblance to those from the 'chip' subset; many of the other functional groups' loadings mimicked those from the 'control' subset. The loadings for PCs 3 through five seemed more representative of contributions from the collagenous component, namely from the proline, hydroxyproline, phenylalanine, Amide III, CH_2 deformation, and Amide I functionalities (located around ~ 920 cm^{-1}, ~ 930 cm^{-1}, ~ 1004 cm^{-1}, ~ 1250 cm^{-1}, ~ 1480 cm^{-1}, and ~ 1600 cm^{-1}, respectively). PC3 loadings resembled those from the 'chip' subset, while loadings from PCs 4 and five were more like those from the 'control' subset (the main loadings for PC5 being hydroxyproline, Amide III, CH_2 deformation, and Amide I located around ~ 920 cm^{-1}, ~ 1250 cm^{-1}, ~ 1480 cm^{-1}, and ~ 1600 cm^{-1}, respectively). With another mixed representative contribution from both components, subtle loadings from the ν_4 PO_4^{3-} band around ~ 590 cm^{-1} and Amide I

~ 1600 cm^{-1} seemed to be the more distinguishable features for PC6. Its loadings also somewhat resembled those from the 'control' subset.

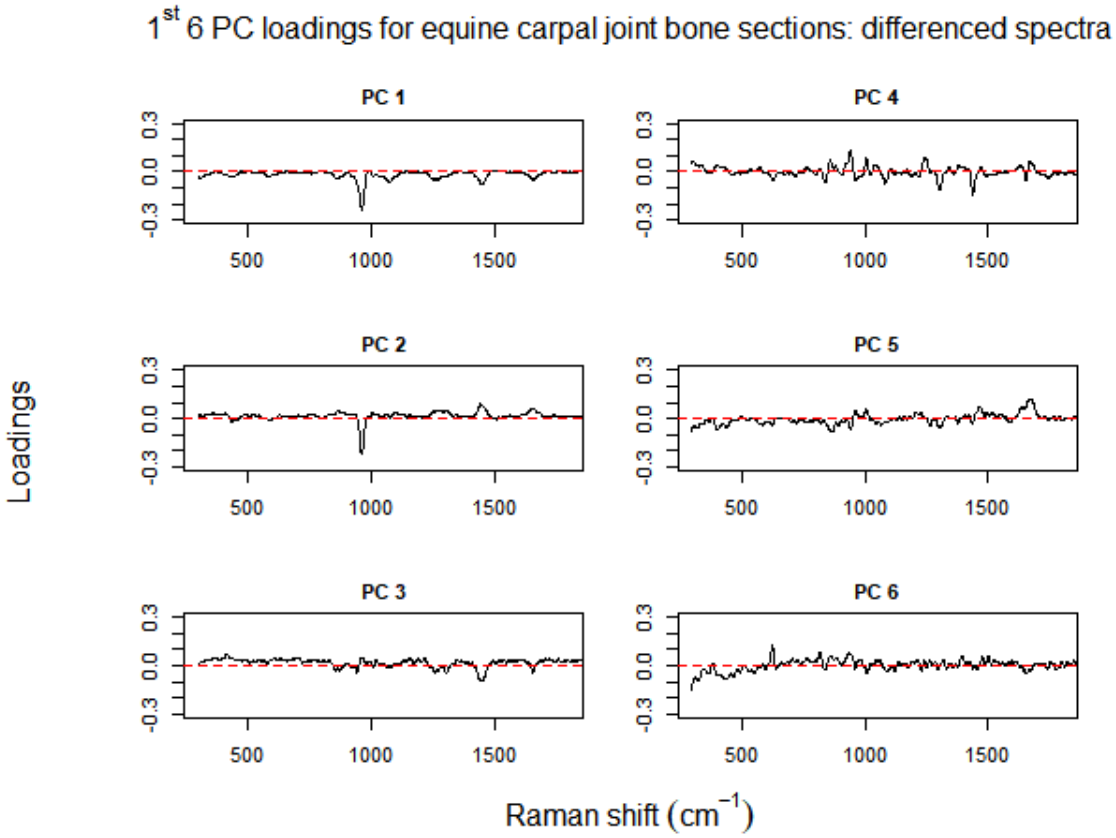

Figure 4.34 The first six PCs' loadings plots for the differenced Raman spectral data from the equine carpal joint bone sections. Again, PCs 1 and two were mixed contributions dominated by the first symmetric stretch phosphate band from the mineral component. PCs 3 through five seemed to represent more contributions from the matrix component. PC6 seemed like another mixed contribution from both components.

Figure 4.35 shows the scores plot for the second PC against the first for the eight combinations of section and layer from the differenced spectral data from the articular cartilage and subchondral bone of the equine carpal joint bone specimens. Again, PCA was capable of separating the sections and layers into clusters to bring out some of the within-object variations. Scores for the AC (L1) in all four sections (third carpal (S1), fourth carpal (S2), the upper part of the radial carpal (S3), and the lower part of the radial carpal (S4); coloured black, green, cyan, and yellow, respectively) showed seemingly random scatter. Likewise, those for the SCB (L2) in all four sections (coloured red, blue, magenta, and grey, respectively) showed no distinct clustering within either PC dimension.

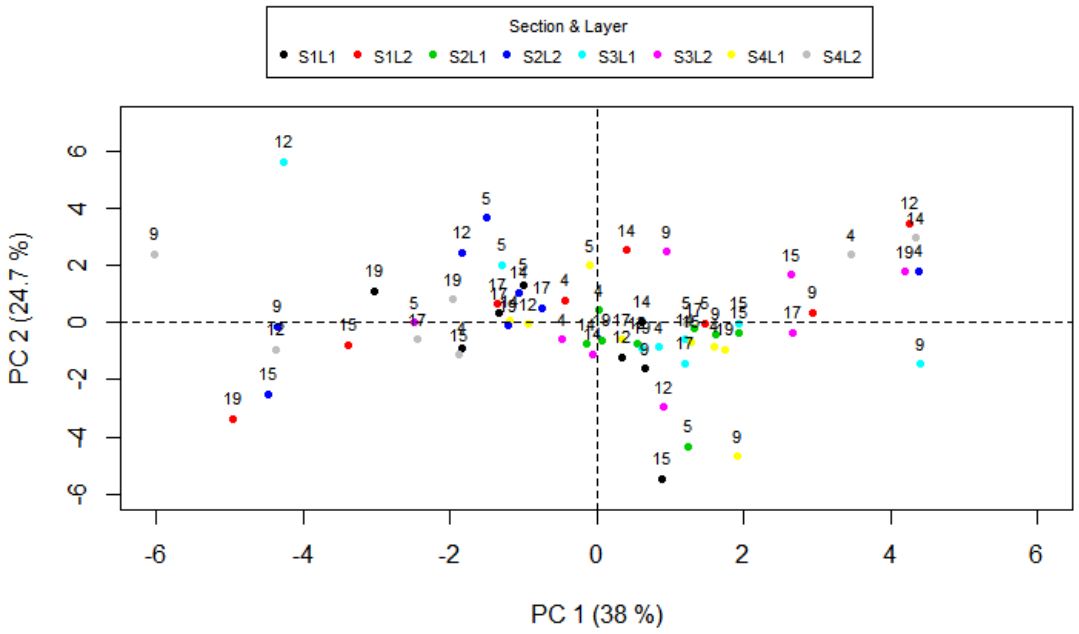

Figure 4.35 Scores plot for PC2 against PC1 from the differenced Raman spectral data (by section and layer): S1L1 (black); S1L2 (red); S2L1 (green); S2L2 (blue); S3L1 (cyan); S3L2 (magenta); S4L1 (yellow); S4L2 (grey).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L1 = articular cartilage (AC); L2 = subchondral bone (SCB);
horses 4, 5, 9, 12, 14, 15, & 17 = two-year-old horses; horse 19 = three-year-old horse.

Figure 4.36 shows a scatterplot matrix of the fifteen combinations of pairs of the first six PCs (from the eight combinations of section and layer) from the aggregated equine carpal joint differenced Raman spectral data. There was no explicit score clustering in any other pairs of PC dimensions upon a survey of the scatterplot matrix.

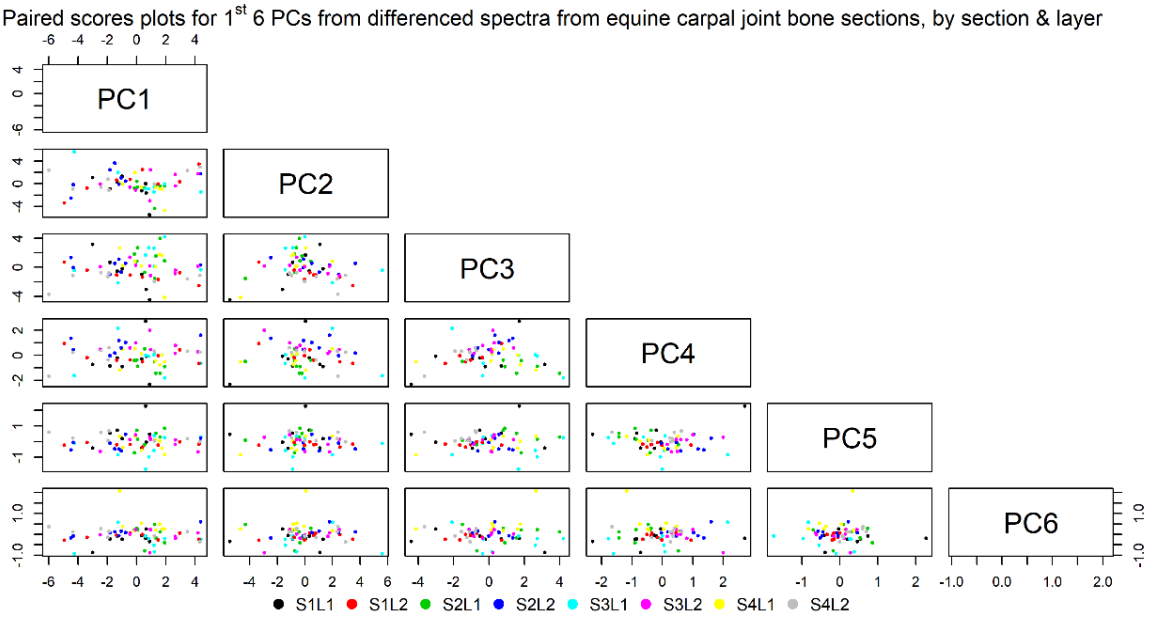

Figure 4.36 Paired scores plot matrix for the first six PCs for the differenced Raman spectral data from the equine carpal joint bone sections (by section and layer): S1L1 (black); S1L2 (red); S2L1 (green); S2L2 (blue); S3L1 (cyan); S3L2 (magenta); S4L1 (yellow); S4L2 (grey).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L1 = articular cartilage (AC); L2 = subchondral bone (SCB).

4.2.2.4 PCA of the 'stacked' spectral data

PCA results for the 'stacked' (that is, the two subsets combined), aggregated spectral data from the equine carpal joint bone sections follow.

Figure 4.37 shows the scree plot of the variances for the first 10 PCs from the stacked, aggregated spectral data set. The first six PCs accounted for approximately 97 % of the total variation in the data, again, rapidly tailing off after PC4. PC1 explained about 67.6 % of the total variation while PCs 2 through six accounted for approximately 19.6 %, 6.2 %, 2.5 %, 0.657 %, and 0.429 %, respectively.

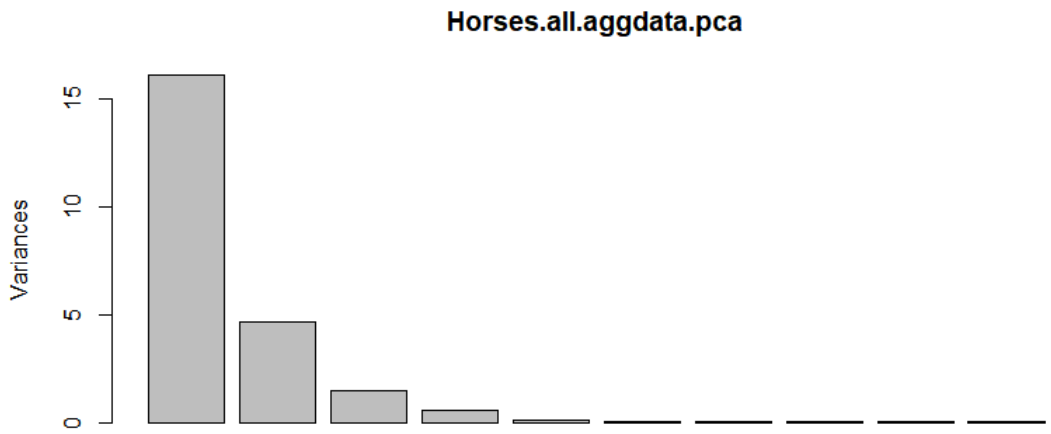

Figure 4.37 Scree plot for stacked, aggregated Raman spectral data from the equine carpal joint bone sections. Cumulatively, the first six PCs accounted for approximately 97 % of the total spectral variation; individually, they explained about 67.6 %, 19.6 %, 6.2 %, 2.5 %, 0.7 %, and 0.4 %, respectively.

Figure 4.38 shows the loadings plots for the first six PCs from the stacked, aggregated spectral data set. The first four PC loadings closely resembled those of the aggregated 'chip' spectral data subset. Loadings for the first two PCs both appeared to be representative of contributions from both components, dominated to some extent by the first symmetric phosphate stretch at ~ 960 cm^{-1}. The other loadings for PC1 came from the ν_2 PO_4^{3-} band around ~ 430 cm^{-1}, ν_4 PO_4^{3-} band ~ 590 cm^{-1}, ν_3 PO_4^{3-} band ~ 1030 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1490 cm^{-1}, and Amide I ~ 1600 cm^{-1}. For PC2, the other loadings were from the ν_3 PO_4^{3-} band around ~ 1030 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1490 cm^{-1}, and Amide I ~ 1600 cm^{-1}. Loadings from the third PC were more suggestive of matrix component contributions, and those from the fourth PC were more suggestive of another mixed contribution. For PC3, these loadings looked to be from the proline, hydroxyproline, Amide III, CH_2 deformation, and Amide I functionalities (located around ~ 920 cm^{-1}, ~ 930 cm^{-1}, ~ 1250 cm^{-1}, ~ 1480 cm^{-1}, and ~ 1600 cm^{-1}, respectively). In addition to these, PC4 had loadings from the ν_2 PO_4^{3-} band around ~ 430 cm^{-1}, ν_4 PO_4^{3-} band ~ 590 cm^{-1}, and phenylalanine ~ 1004 cm^{-1}. The loadings plots for PCs 5 and six were also more suggestive of a mixed contribution. There were more subdued loadings from the ν_2 PO_4^{3-} band around ~ 430 cm^{-1}, proline ~ 920 cm^{-1}, hydroxyproline ~ 930 cm^{-1}, phenylalanine ~ 1004 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1} for PC5. The loadings plot for PC6 was much noisier, likely representing a mixed contribution from the ν_2 PO_4^{3-}

band (around ~ 430 cm^{-1}) of the mineral component, and Amide I (~ 1600 cm^{-1}) from the organic matrix component.

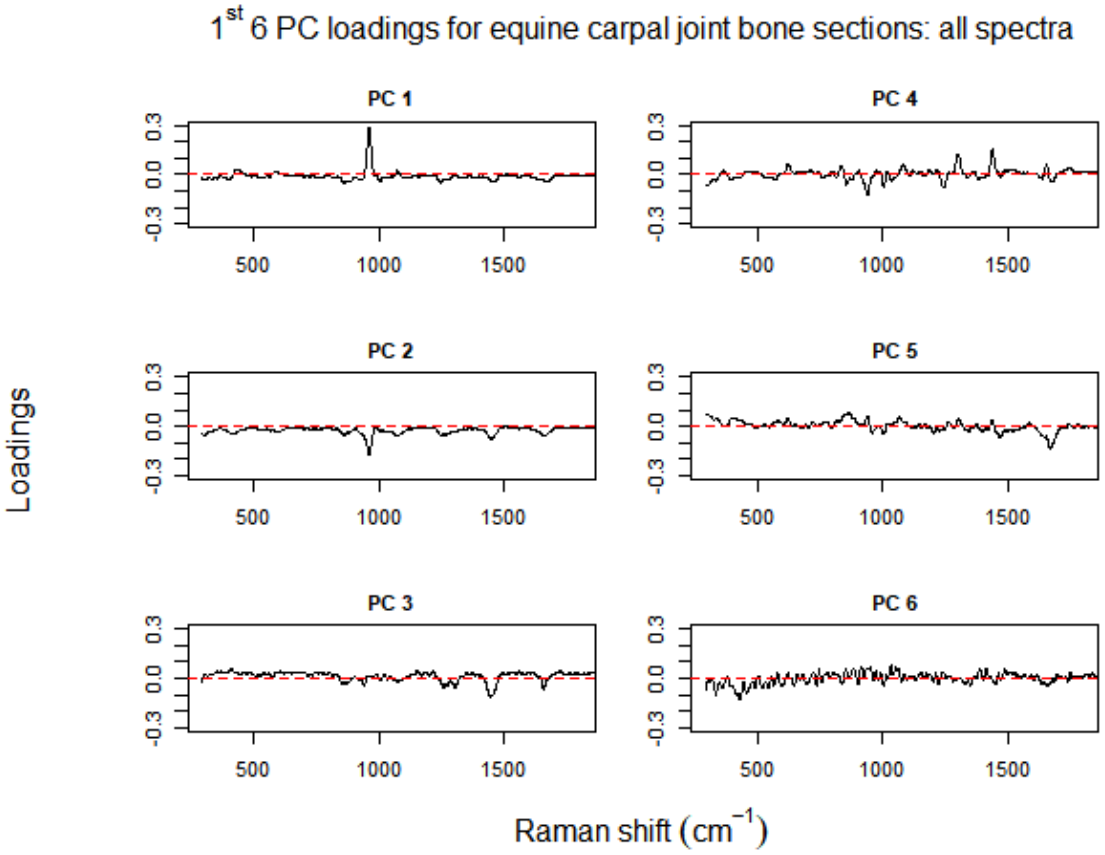

Figure 4.38 The first six PCs' loadings of the stacked, aggregated Raman spectral data from the equine carpal joint bone sections. The first four PC loadings closely resembled those from the original aggregated 'chip' spectral data subset. Most of the six PCs seemed to represent mixed contributions from both components (only the third PC seemed to represent a contribution from the organic matrix. The sixth PC was much noisier in appearance than the others).

Figures 4.39 and 4.40 show the scores plot for the second PC against the first for the sixteen combinations of section, layer, and condition from the stacked, aggregated equine carpal joint spectral data. Again, PCA was capable of separating the sections, layers, and conditions into their respective clusters to bring out some of the within-object variations. Two overall clusters of scores formed within the PC1 dimension based on bone layer: the AC (L1) from each of the four sections (coloured black, green, cyan, and yellow, respectively) on the negative side, and the SCB (L2) from each of the four sections (coloured red, blue, magenta, and grey, respectively) on the positive side. There was, however, considerable overlap between the two conditions' scores. Removing the average 'horse-section-layer' effects did not achieve separation of the PC scores based on the 'condition' of the bone sections, but seemingly still separated the scores based on layer.

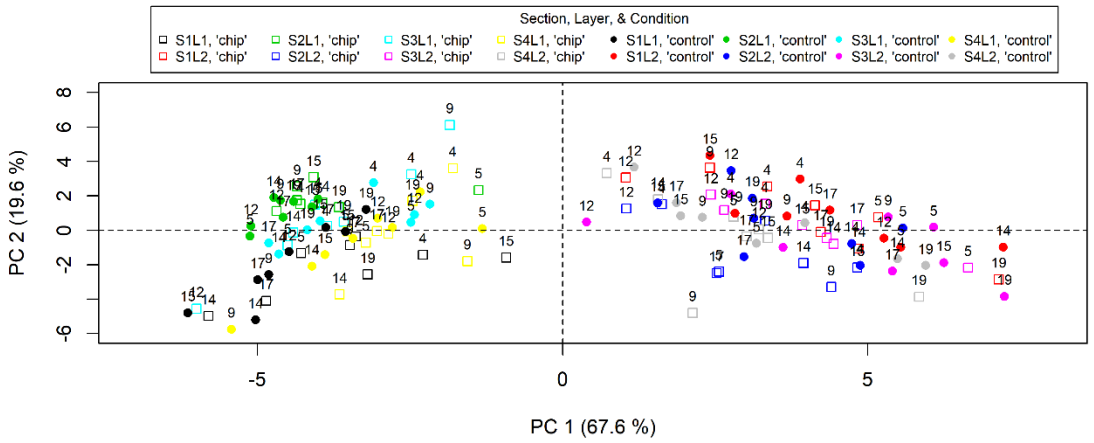

Figure 4.39 Scores plot for PC2 against PC1 from the stacked, aggregated Raman spectral data (by section, layer, and condition): S1L1 (black); S1L2 (red); S2L1 (green); S2L2 (blue); S3L1 (cyan); S3L2 (magenta); S4L1 (yellow); S4L2 (grey).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L1 = articular cartilage (AC); L2 = subchondral bone (SCB);
open square = 'chip' condition, filled circle = 'control' condition;
horses 4, 5, 9, 12, 14, 15, & 17 = two-year-old horses; horse 19 = three-year-old horse.

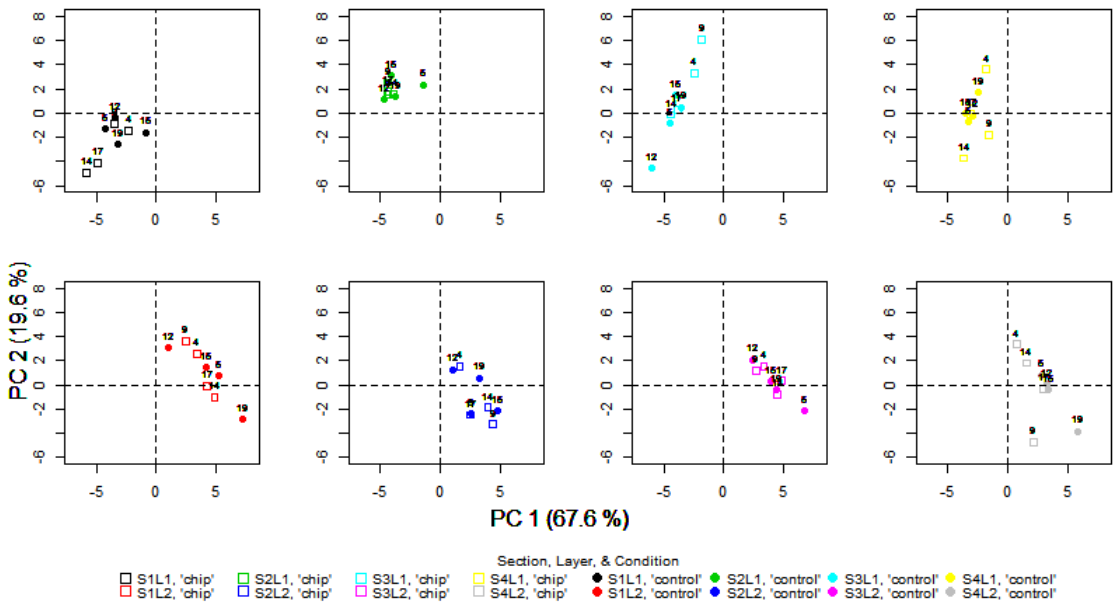

Figure 4.40 Montage of scores plots for PC2 against PC1 from the stacked, aggregated Raman spectral data by section (columns), layer (rows), and condition (open square = 'chip' condition, filled circle = 'control' condition): S1L1 (black); S1L2 (red); S2L1 (green); S2L2 (blue); S3L1 (cyan); S3L2 (magenta); S4L1 (yellow); S4L2 (grey).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L1 = articular cartilage (AC); L2 = subchondral bone (SCB);
horses 4, 5, 9, 12, 14, 15, & 17 = two-year-old horses; horse 19 = three-year-old horse.

Figure 4.41 shows the scatterplot matrix of the fifteen combinations of pairs of the first six PCs (from the sixteen combinations of section, layer, and condition) from the stacked, aggregated equine carpal joint spectral data. Spreading or grouping of some PC scores was, yet again, more evident with pairs from the lower PCs than from the higher PCs. Like Figures 4.39 and 4.40, there was easier 'layer'-influenced separation of each bone section's PC scores along PC1 than along the other PCs. There was also still no PC score separation by 'condition' for the higher PCs, and the overall score distribution along these other PCs' axes was unclear.

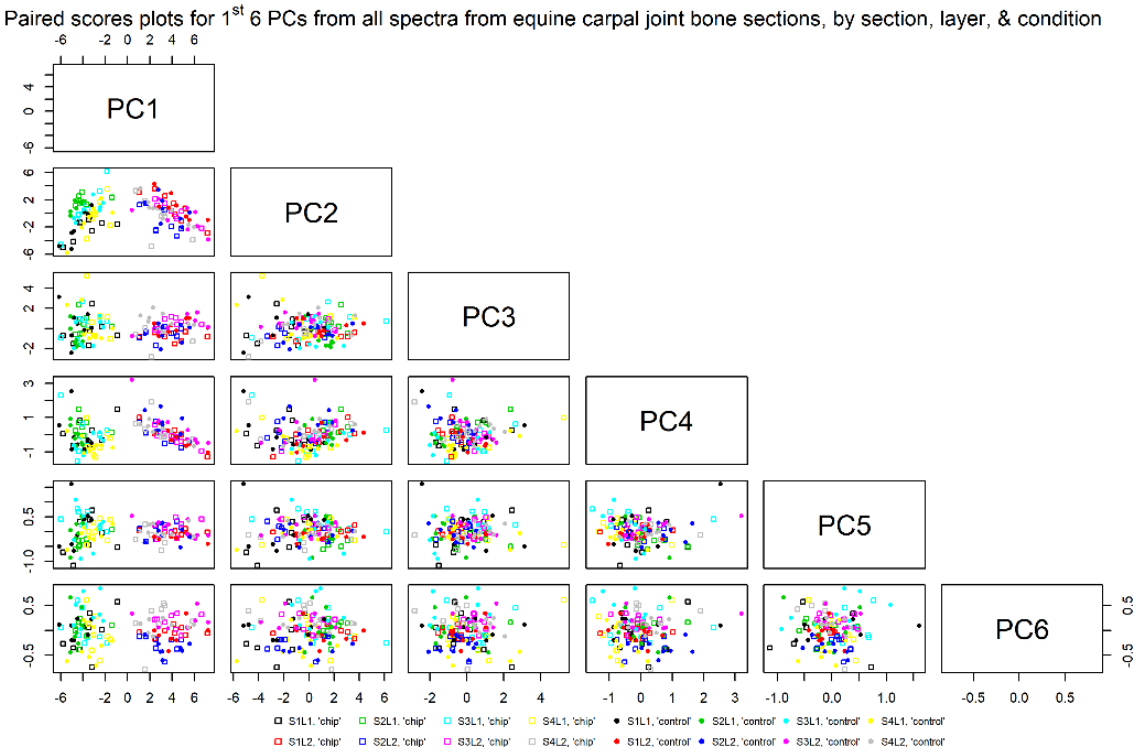

Figure 4.41 Paired scores plot matrix for the first six PCs for the stacked, aggregated Raman spectral data from the equine carpal joint bone sections (by section, layer, and condition): S1L1 (black); S1L2 (red); S2L1 (green); S2L2 (blue); S3L1 (cyan); S3L2 (magenta); S4L1 (yellow); S4L2 (grey).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L1 = articular cartilage (AC); L2 = subchondral bone (SCB);
open square = 'chip' condition, filled circle = 'control' condition.

4.2.2.4.1 PCA of the mean-adjusted, 'stacked' spectral data

PCA results for the mean-adjusted, stacked, and aggregated spectral data from the equine carpal joint bone sections are displayed below.

Figure 4.42 shows the scree plot of the variances for the first 10 PCs from the mean-adjusted, stacked spectral data set. The first six PCs accounted for approximately 96.9 % of the total variation in the data, again, rapidly tailing off after PC4. PC1 explained about 73.2 % of the total variation whilst PCs 2 through six accounted for approximately 15.3 %, 4.7 %, 2.5 %, 0.692 %, and 0.374 %, respectively.

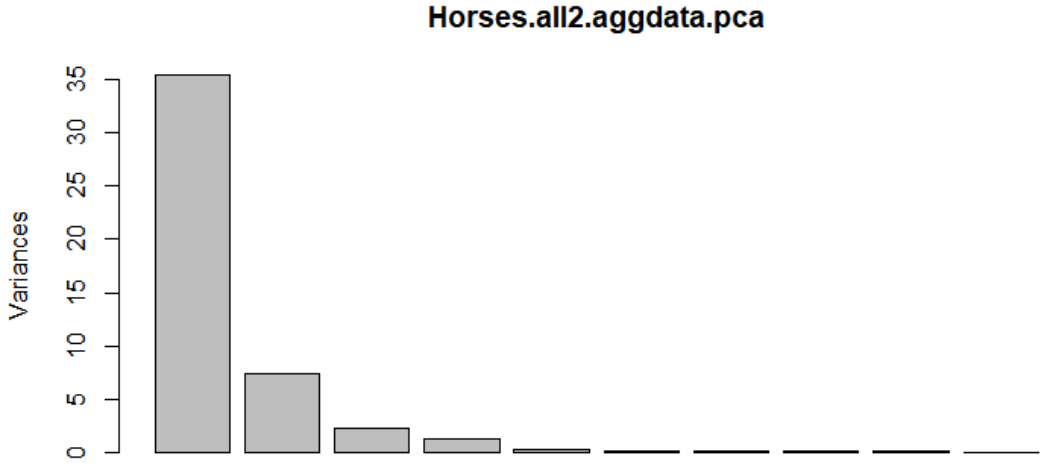

Figure 4.42 Scree plot for the mean-adjusted, stacked Raman spectral data from the equine carpal joint bone sections. Cumulatively, the first six PCs accounted for approximately 96.9 % of the total spectral variation; individually, they explained about 73.2 %, 15.3 %, 4.7 %, 2.5 %, 0.7 %, and 0.4 %, respectively.

Figure 4.43 shows the loadings plots for the first six PCs from the mean-adjusted, stacked and aggregated spectral data set. Loadings for the second, third, fourth, and fifth PCs bore some passing similarities to their respective counterparts from the 'control' subset – especially the latter three PCs. PC1 appeared to represent a contribution from the mineral component. The dominant loading for PC1 was the first symmetric phosphate stretch at ~ 960 cm^{-1}; the other loadings from the ν_2 PO_4^{3-} band ~ 430 cm^{-1}, ν_4 PO_4^{3-} band ~ 590 cm^{-1}, and ν_3 PO_4^{3-} band ~ 1030 cm^{-1} were much more subdued. Loadings for the latter five PCs were of rather low intensity. PC2 seemed to represent a mixture of contributions from both components. The proline, hydroxyproline, ν_1 PO_4^{3-} band, ν_3 PO_4^{3-} band, Amide III, CH_2 deformation, and Amide I (located around ~ 960 cm^{-1}, ~ 1030 cm^{-1}, ~ 1250 cm^{-1}, ~ 1480 cm^{-1}, and ~ 1600 cm^{-1}, respectively) were the noticeable loadings for PC2. The third and fourth PCs seemed to represent more of a contribution from the organic matrix component. PC3 had loadings from Amide III, the CH_2 deformation, and Amide I around ~ 1250 cm^{-1}, ~ 1490 cm^{-1}, and ~ 1600 cm^{-1}, respectively. To a lesser extent, there were also loadings from the proline, hydroxyproline, and phenylalanine around ~ 920 cm^{-1}, ~ 930 cm^{-1}, and ~ 1004 cm^{-1}, respectively. Loadings from hydroxyproline around ~ 930 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1} were the 'primary' loadings for PC4. The fifth and sixth PCs seemed to each represent another mixture of contributions from both components. PC5's loadings were mainly from the ν_2 PO_4^{3-} band around ~ 430 cm^{-1}, proline

~ 920 cm^{-1}, hydroxyproline ~ 930 cm^{-1}, the CH_2 deformation ~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1}. The more easily identifiable loadings for PC6 looked to be from the ν_2 PO_4^{3-} band around ~ 430 cm^{-1}, ν_4 PO_4^{3-} band ~ 590 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1}.

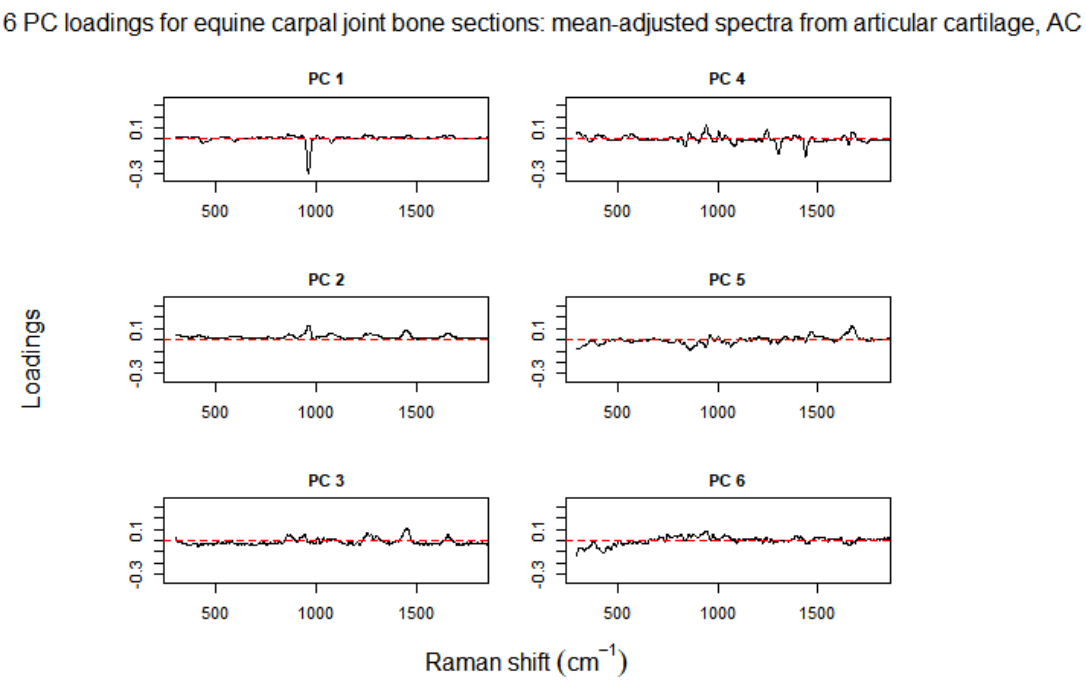

Figure 4.43 The first six PCs' loadings plots from the mean-adjusted, stacked Raman spectral data from the equine carpal joint bone sections. The loadings for PCs 2 to five held some passing similarities to those from the aggregated 'control' spectral data subset. All six PCs, again, showed varying contributions from the mineral and matrix components.

Figure 4.44 shows the scores plot for the second PC against the first for the sixteen combinations of section, layer, and condition from the mean-adjusted, stacked and aggregated equine carpal joint spectral data. Though there seemed to be three overall clusters of the scores around the PC1 dimension, as with the PCA of the original stacked spectral data, there was, unfortunately, still considerable overlap between the two conditions' scores. This overlap suggested that no separation of the two 'conditions' of the carpal joint bone specimens was possible.

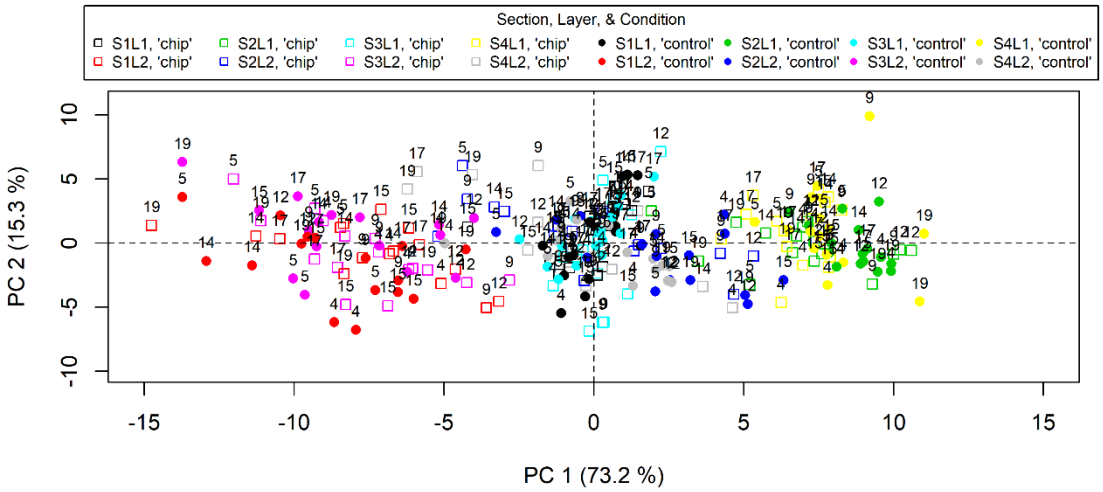

Figure 4.44 Scores plot for PC2 against PC1 from the mean-adjusted, stacked Raman spectral data (by section, layer, and condition): S1L1 (black); S1L2 (red); S2L1 (green); S2L2 (blue); S3L1 (cyan); S3L2 (magenta); S4L1 (yellow); S4L2 (grey).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L1 = articular cartilage (AC); L2 = subchondral bone (SCB);
open square = 'chip' condition, filled circle = 'control' condition;
horses 4, 5, 9, 12, 14, 15, & 17 = two-year-old horses; horse 19 = three-year-old horse.

Figure 4.45 shows the scatterplot matrix of the fifteen combinations of pairs of the first six PCs (from the sixteen combinations of section, layer, and condition) from the mean-adjusted, stacked, and aggregated equine carpal joint spectral data. Spreading or grouping of some PC scores was more evident from the lower PCs than from the higher PCs. Like Figure 4.44, only PC1 seemed to separate the PC scores – by layer – into three main clusters. There was – once more – neither any PC score separation by 'condition' nor any intelligible PC score clustering with any of the higher PCs.

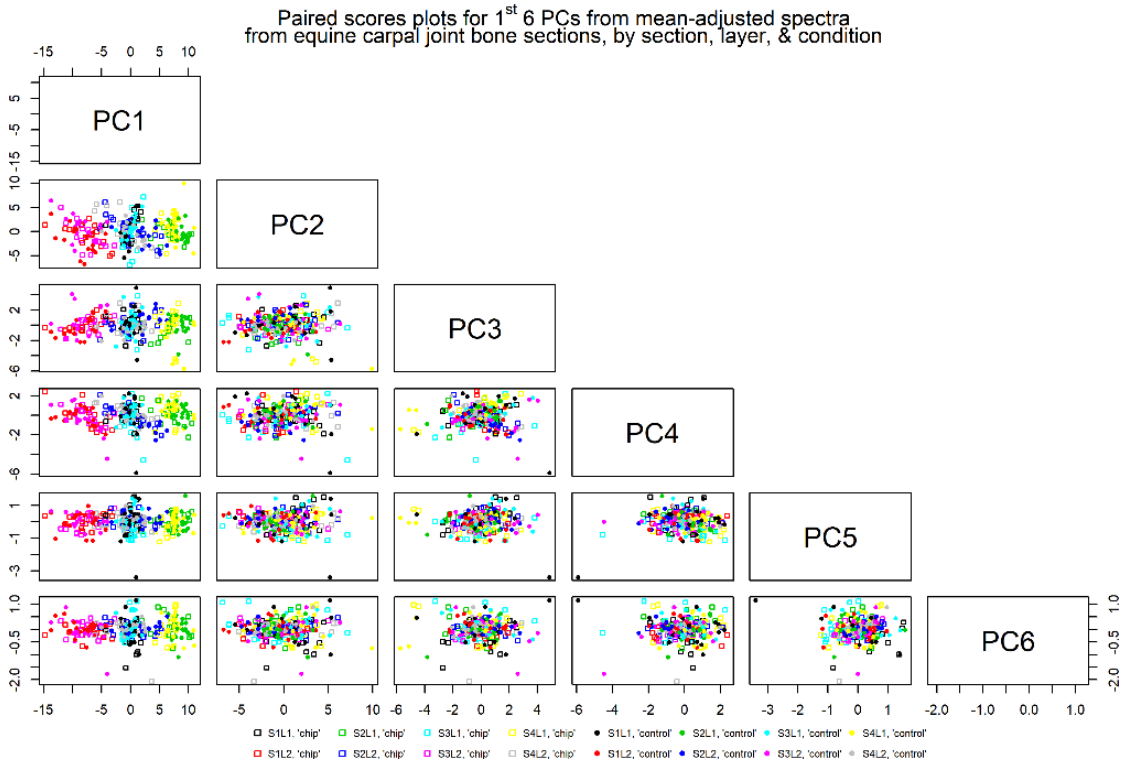

Figure 4.45 Paired scores plot matrix for the first six PCs for the mean-adjusted, stacked, aggregated spectral data from the equine carpal joint bone sections (by section, layer, and condition): S1L1 (black); S1L2 (red); S2L1 (green); S2L2 (blue); S3L1 (cyan); S3L2 (magenta); S4L1 (yellow); S4L2 (grey).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L1 = articular cartilage (AC); L2 = subchondral bone (SCB);
open square = 'chip' condition, filled circle = 'control' condition.

4.2.2.4.2 PCA of the mean-adjusted, 'stacked' spectral data from the articular cartilage (AC)

PCA results for the mean-adjusted stacked, and aggregated spectral data from the articular cartilage (AC) of the equine carpal joint bone sections are presented below.

Figure 4.46 shows the scree plot of the variances for the first 10 PCs from the mean-adjusted, stacked, and aggregated spectral data subset. The first six PCs accounted for approximately 93.9 % of the total variation in the spectral data, again rapidly tailing off after PC4. PC1 explained about 53.2 % of the total variation, while PCs 2 through six accounted for about 24.2 %, 10 %, 4.2 %, 1.6 %, and 0.794 %, respectively.

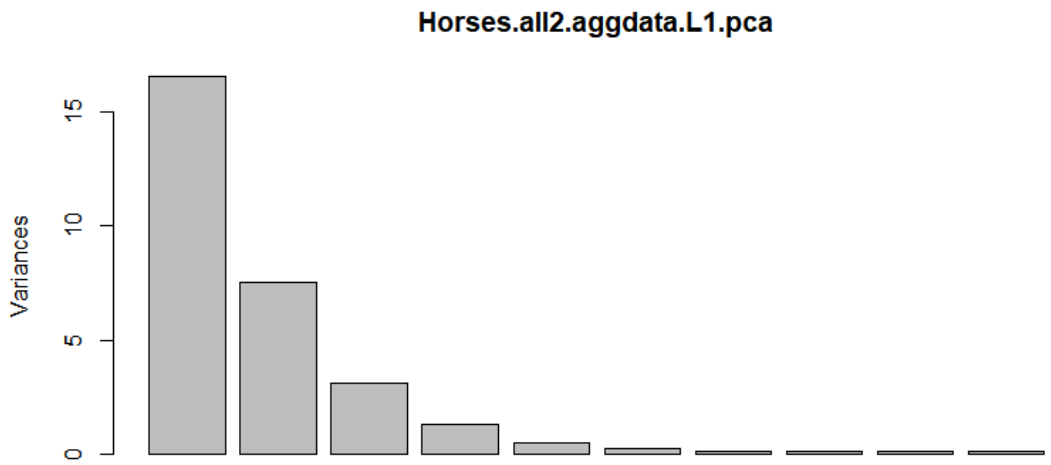

Figure 4.46 Scree plot of the mean-adjusted, stacked Raman spectral data from the articular cartilage (AC) of the equine carpal joint bone sections. Cumulatively, the first six PCs accounted for approximately 93.9 % of the total spectral variation; individually, they explained about 53.2 %, 24.2 %, 10 %, 4.2 %, 1.6 %, and 0.8 %, respectively.

Figure 4.47 shows the loadings plots for the first six PCs from the mean-adjusted, stacked, and aggregated articular cartilage (AC) spectra data subset. Most of these loadings appeared to be nearly identical to those from the original mean-adjusted, stacked, and aggregated data set (namely the first four PCs), suggesting that the variations between the two layers of bone may have overwhelmed any observable differences in this data set. That is, PCs associated with the AC may have dominated the PCA of this original mean-adjusted, stacked, and aggregated data set. PC1 once again appeared to represent a contribution from the mineral component. The dominant loading for PC1 was the first symmetric phosphate stretch at ~ 960 cm^{-1}; the other loadings from the ν_2 PO_4^{3-} band around ~ 430 cm^{-1}, ν_4 PO_4^{3-} band ~ 590 cm^{-1}, and ν_3 PO_4^{3-} band ~ 1030 cm^{-1} were much more subdued. PC2 seemed to represent a mixture of contributions from both components. The proline, hydroxyproline, ν_1 PO_4^{3-} band, ν_3 PO_4^{3-} band, Amide III, CH_2 deformation, and Amide I (located around ~ 920 cm^{-1}, ~ 930 cm^{-1}, ~ 960 cm^{-1}, ~ 1030 cm^{-1}, ~ 1250 cm^{-1}, ~ 1480 cm^{-1}, and ~ 1600 cm^{-1}, respectively) were the noticeable loadings for PC2. The third and fourth PCs seemed to represent more of a contribution from the organic matrix component. PC3 had loadings from Amide III, the CH_2 deformation, and Amide I around ~ 1250 cm^{-1}, ~ 1490 cm^{-1}, and ~ 1600 cm^{-1}, respectively (to a lesser extent there were also loadings from the proline, hydroxyproline, and phenylalanine around ~ 920 cm^{-1}, ~ 930 cm^{-1}, and ~ 1004 cm^{-1}, respectively). Loadings from hydroxyproline around ~ 930 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide I

~ 1600 cm^{-1} were the 'primary' loadings for PC4. The loadings for PCs 5 and six were also more suggestive of a mixed contribution, albeit somewhat closer in appearance to the equivalent loadings from the original stacked and aggregated spectral data set in subsection 4.2.2.4. There were, again, subdued loadings from the ν_2 PO_4^{3-} band around ~ 430 cm^{-1}, proline ~ 920 cm^{-1}, hydroxyproline ~ 930 cm^{-1}, phenylalanine ~ 1004 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1} for PC5. The loadings plot for PC6 was a little noisier, likely representing a mixed contribution from the ν_2 PO_4^{3-} band (around ~ 430 cm^{-1}) of the mineral component, and Amide I (~ 1600 cm^{-1}) from the organic matrix component.

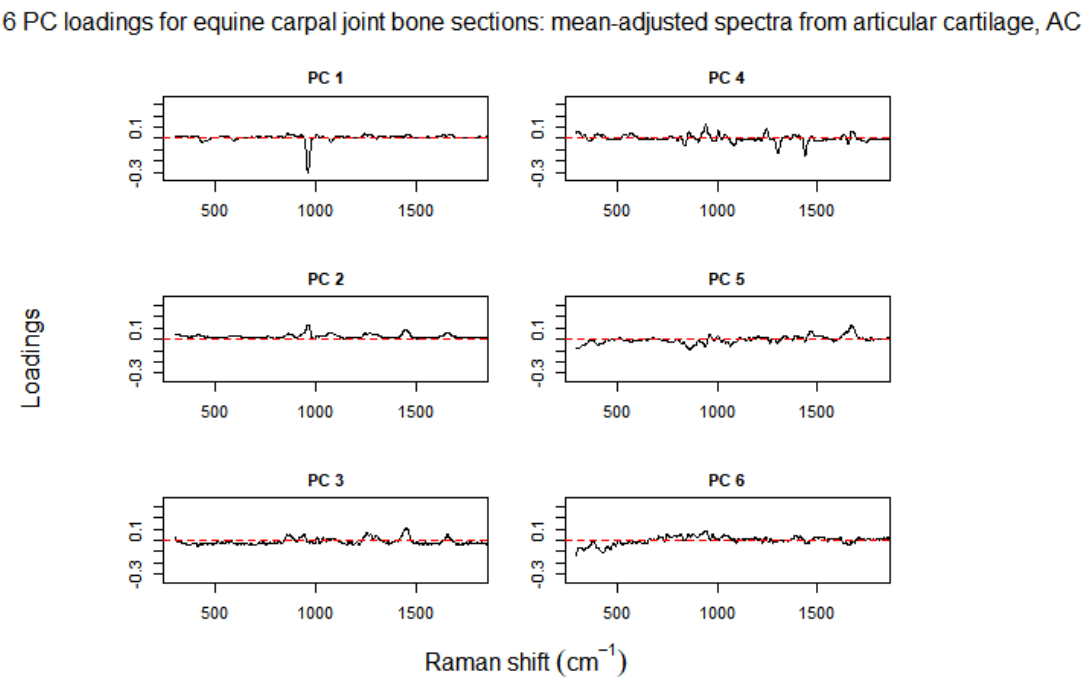

Figure 4.47 The first six PCs' loadings plots of the mean-adjusted, stacked Raman spectral data from the articular cartilage (AC) of the equine carpal joint bone sections. Most of the loadings were nearly identical to those from the original mean-adjusted, stacked Raman data set. PCs 1, two, five, and six appeared to represent mixed contributions from both components, whereas PCs 3 and four were more representative of organic matrix contributions.

Figure 4.48 shows the scores plot for the second PC against the first for the eight combinations of section and condition from the mean-adjusted, stacked, and aggregated articular cartilage (AC) spectral data subset. Two overall clusters formed: the third carpal (S1) with the upper part of the radial carpal (S3) (coloured black and cyan, respectively), and the fourth carpal (S2) with the lower part of the radial carpal (S4) (coloured green and yellow, respectively). As with the PCA of the original mean-adjusted stacked spectral data, there was still considerable overlap between the two conditions' scores.

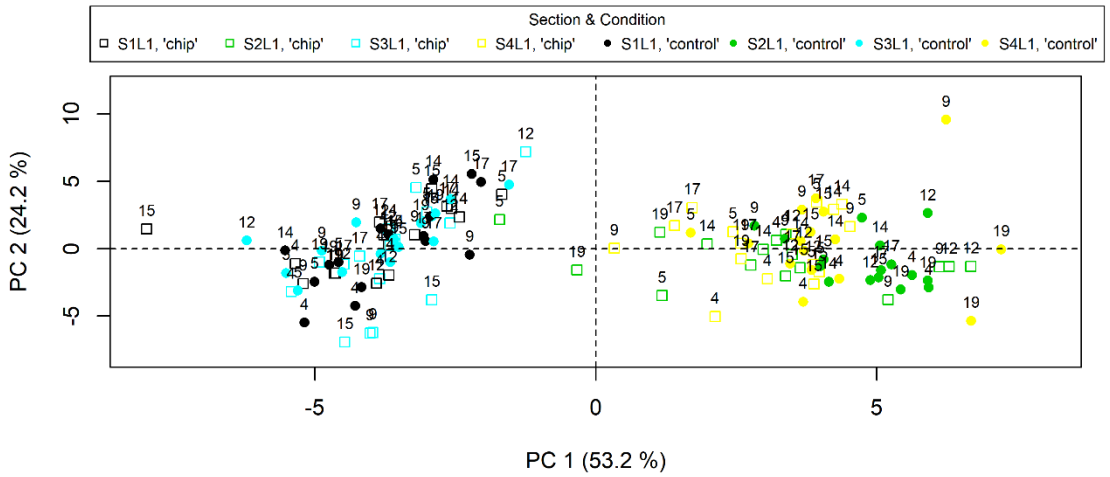

Figure 4.48 Scores plot for PC2 against PC1 from the mean-adjusted, stacked Raman spectral data from the articular cartilage (AC) of the equine carpal joint bone sections (by section and condition): S1L1 (black); S2L1 (green); S3L1 (cyan); S4L1 (yellow).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L1 = articular cartilage (AC);
open square = 'chip' condition, filled circle = 'control' condition;
horses 4, 5, 9, 12, 14, 15, & 17 = two-year-old horses; horse 19 = three-year-old horse.

Figure 4.49 shows the scatterplot matrix of the fifteen combinations of pairs of the first six PCs (from the eight combinations of section, layer, and condition) from the mean-adjusted, stacked, and aggregated equine carpal joint articular cartilage spectral data. Spreading or grouping of some PC scores was more evident with pairs from the lower PCs than from the higher PCs. Like Figure 4.48, only PC1 seemed to separate the PC scores by layer (and to a lesser extent, section) into two main clusters. There was neither any PC score separation by 'condition' nor any other intelligible PC score dispersal with any of the higher PCs.

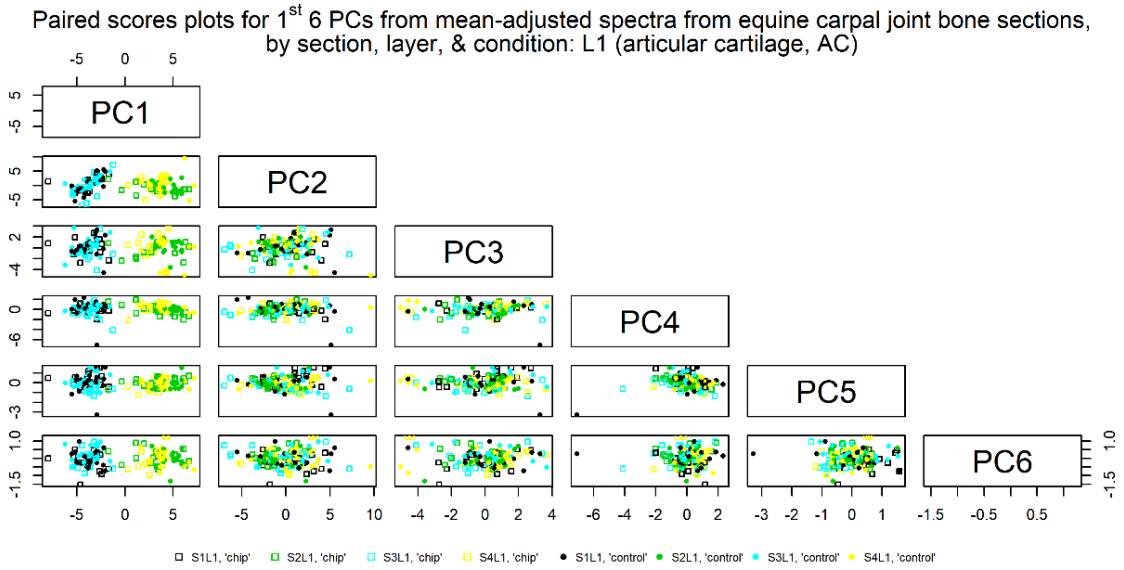

Figure 4.49 Paired scores plot matrix for the first six PCs for the mean-adjusted, stacked, and aggregated Raman spectral data from the articular cartilage (AC) of the equine carpal joint bone sections (by section and condition): S1L1 (black); S1L2 (red); S2L1 (green); S3L1 (cyan); S4L1 (yellow).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L1 = articular cartilage (AC);
open square = 'chip' condition, filled circle = 'control' condition.

4.2.2.4.3 PCA of the mean-adjusted, 'stacked' spectral data from the subchondral bone (SCB)

In this final subsection, PCA results for the mean-adjusted, stacked, and aggregated spectral data from the subchondral bone (SCB) of the equine carpal joint bone sections are set out below.

Figure 4.50 shows the scree plot of the variances for the first 10 PCs from the mean-adjusted, stacked, and aggregated spectral data subset. The first six PCs accounted for approximately 97.5 % of the total variation in the spectral data, again, rapidly tailing off after PC4. PC1 explained about 71.6 % of the total variation, and PCs 2 through six accounted for approximately 18.2 %, 4.6 %, 2 %, 0.64 %, and 0.498 %, respectively.

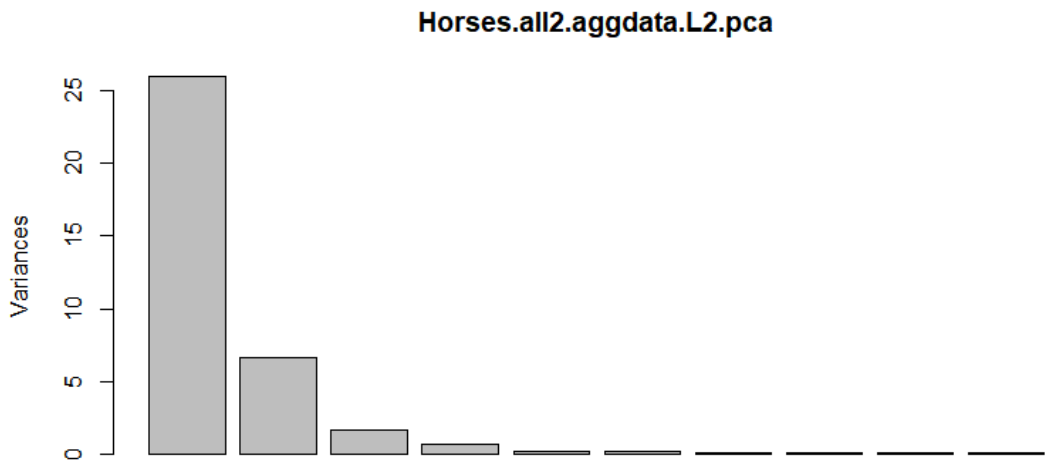

Figure 4.50 Scree plot of the mean-adjusted, stacked Raman spectral data from the subchondral bone (SCB) of the equine carpal joint bone sections. Cumulatively, the first six PCs accounted for approximately 97.5 % of the total spectral variation; individually, they explained about 71.6 %, 18.2 %, 4.6 %, 2 %, 0.6 %, and 0.5 %, respectively.

Figure 4.51 shows the loadings plots for the first six PCs from the mean-adjusted, stacked, and aggregated subchondral bone (SCB) spectral data subset. Though the loadings for PC1 were nearly identical to those from the original mean-adjusted, stacked, and aggregated data set, there were, however, some subtle differences in the loadings for PC2 – particularly in the ~ 855 – 925 cm^{-1} region where hydroxyproline and proline Raman bands usually reside. The dominant loading for PC1 was the first symmetric phosphate stretch at ~ 960 cm^{-1}; there were more minor loadings from the ν_2 PO_4^{3-} band around ~ 430 cm^{-1}, ν_4 PO_4^{3-} band ~ 590 cm^{-1}, and ν_3 PO_4^{3-} band ~ 1030 cm^{-1}. There were also barely perceptible loadings from the Amide III, CH_2 deformation, and Amide I functionalities (found around ~ 1250 cm^{-1}, ~ 1480 cm^{-1}, and ~ 1600 cm^{-1}, respectively). Though the loadings for PCs 3 through five also showed some mixed contributions from the mineral and matrix components within the bone, they did not seem to bear much resemblance to their equivalent PC loadings from the original mean-adjusted, stacked, and aggregated data set. The loadings for PC3, rather, more closely resembled the equivalent PC loadings from the original stacked and aggregated data set in subsection 4.2.2.4. Again, these loadings looked to be from the proline, hydroxyproline, Amide III, CH_2 deformation, and Amide I functionalities (located around ~ 920 cm^{-1}, ~ 930 cm^{-1}, ~ 1250 cm^{-1}, ~ 1480 cm^{-1}, and ~ 1600 cm^{-1}, respectively). PC4 loadings seemingly originated from the ν_4 PO_4^{3-} band around ~ 590 cm^{-1}, proline ~ 920 cm^{-1}, hydroxyproline ~ 930 cm^{-1}, phenylalanine ~ 1004 cm^{-1}, Amide III ~ 1250 cm^{-1}, CH_2 deformation

~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1}. PC5's 'main' loadings were from the ν_2 PO_4^{3-} band ~ 430 cm^{-1}, ν_4 PO_4^{3-} band ~ 590 cm^{-1}, proline ~ 920 cm^{-1}, hydroxyproline ~ 930 cm^{-1}, CH_2 deformation ~ 1480 cm^{-1}, and Amide I ~ 1600 cm^{-1}. Like the previous three PCs, the loadings plot for PC6 did not look much like its equivalent from the original mean-adjusted, stacked, and aggregated data set. PC6 loadings, instead, somewhat resembled their equivalent PC loadings from the aggregated 'control' data subset in subsection 4.2.2.2.1, with a 'large' loading from the mineral component's ν_2 PO_4^{3-} band around ~ 430 cm^{-1}.

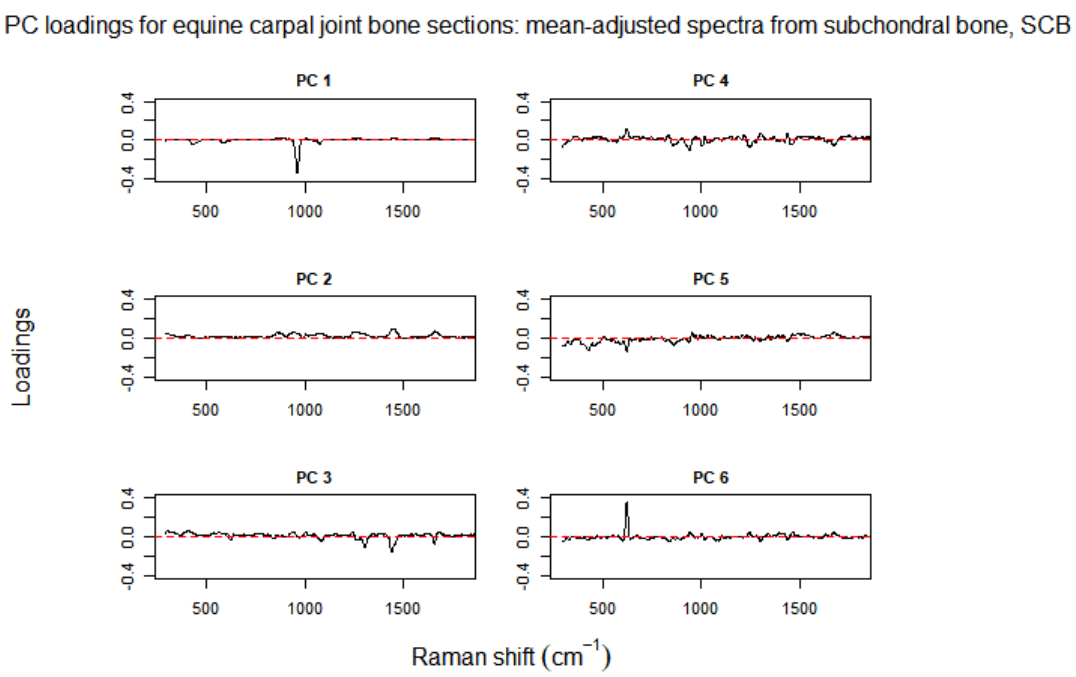

Figure 4.51 The first six PCs' loadings plots of the mean-adjusted, stacked Raman spectral data from the subchondral bone (SCB) of the equine carpal joint bone sections. The loadings for PC1 were nearly identical to those of the original mean-adjusted, stacked Raman spectral data; there were, however, some subtle differences in some of the loadings for PC2 – particularly in the ~ 855 – 925 cm^{-1} region. The loadings for PCs 3 through five also showed some mixed contributions from the mineral and matrix components within the bone. PC6 looked like a representative contribution from the bone mineral component.

Figure 4.52 shows the scores plot for the second PC against the first for the eight combinations of section and condition from the mean-adjusted, stacked, and aggregated subchondral bone (SCB) spectral data subset. Again, two overall clusters formed: the third carpal (S1) with the upper part of the radial carpal (S3) (coloured red and magenta, respectively), and the fourth carpal (S2) with the lower part of the radial carpal (S4) (coloured blue and grey, respectively). As had been the case with the PCA of the original stacked and aggregated spectral data, there was still considerable overlap between the two conditions' scores. SCB PC scores were generally more scattered than PC scores from the articular cartilage (AC) in Figures 4.48 and 4.49.

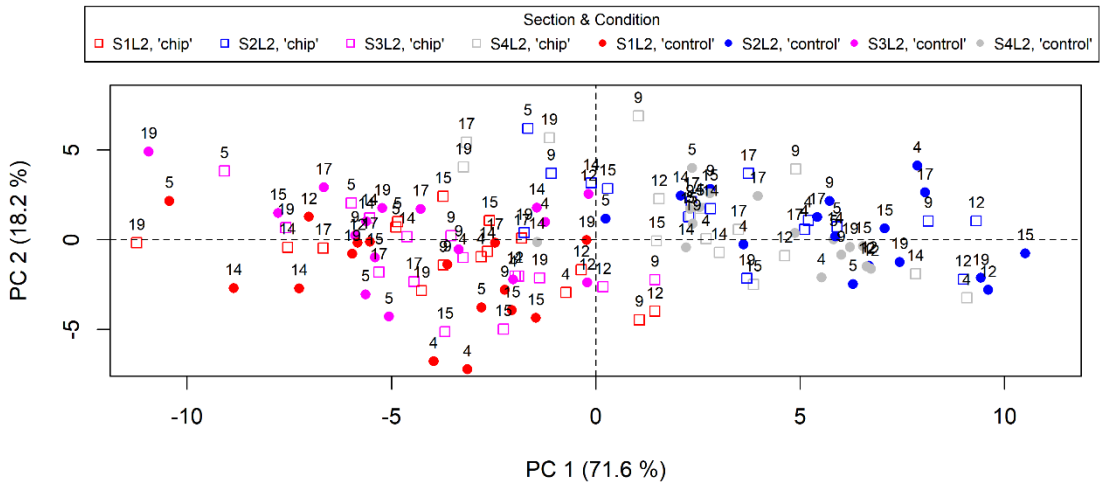

Figure 4.52 Scores plot for PC2 against PC1 from the mean-adjusted, stacked Raman spectral data from the subchondral bone (SCB) of the equine carpal joint bone sections (by section and condition): S1L2 (red); S2L2 (blue); S3L2 (magenta); S4L2 (grey).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L2 = subchondral bone (SCB);
open square = 'chip' condition, filled circle = 'control' condition;
horses 4, 5, 9, 12, 14, 15, & 17 = two-year-old horses; horse 19 = three-year-old horse.

Figure 4.53 shows the scatterplot matrix of the fifteen combinations of pairs of the first six PCs (from the eight combinations of section, layer, and condition) from the mean-adjusted, stacked, and aggregated equine carpal joint subchondral bone spectral data. Spreading or grouping of some PC scores was slightly more evident with pairs from the lower PCs than from the higher PCs. Like Figure 4.52, only PC1 seemed to partially separate the PC scores (again, more so by layer than by section) into two main clusters. There was neither any PC score separation by 'condition' nor any other intelligible PC score dispersal with any of the higher PCs.

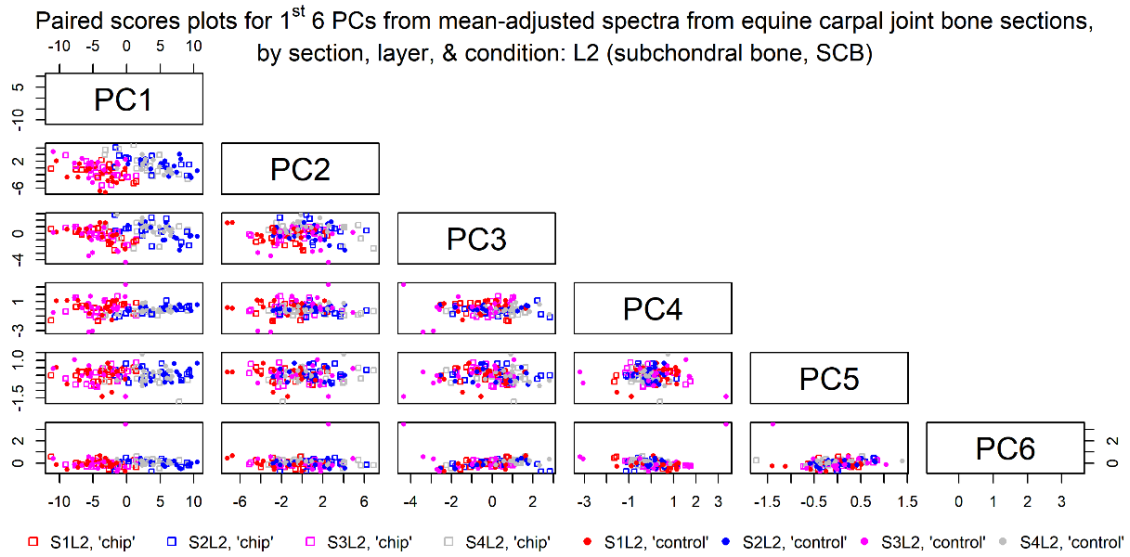

Figure 4.53 Paired scores plot matrix for the first six PCs for the mean-adjusted, stacked, and aggregated spectral data from the subchondral bone (SCB) of the equine carpal joint bone sections (by section and condition): S1L2 (red); S2L2 (blue); S3L2 (magenta); S4L2 (grey).
S1 = third carpal bone, C_3; S2 = fourth carpal bone, C_4; S3 = upper part of the radial carpal bone, C_R/R2; S4 = lower part of the radial carpal bone, C_R/R3;
L2 = subchondral bone (SCB);
open square = 'chip' condition, filled circle = 'control' condition.

4.3 Discussion

As noted earlier in Chapters 1 and 3, it is hoped that information provided by the combined use of molecular vibrational spectroscopy (particularly Raman in this context) and chemometrics would continue to broaden understanding of bone quality (and therefore overall bone strength and health). The desired outcome is the pinpointing of individuals most at risk of developing musculoskeletal diseases like osteoporosis and osteoarthritis (OA).

The equine carpal joints are among the most commonly OA-affected joints, as they are high motion joints. Since joint microstructures act together as single functional units, any change in one substructure of the joint affects the others (60). One of the characteristics of joint diseases like OA is subchondral bone sclerosis (that is, stiffening of the SCB), particularly in its later stages (61). This sclerosis accompanies articular cartilage volume loss, a sign of its degradation (60).

Groups such as Frisbie *et al.* (62) have conducted studies of experimentally induced equine carpal OA as part of animal models for human OA. As was the case with the equine Mt3 bone specimens, though, little of the readily available and accessible literature had examples of previous tandem Raman-chemometrics applications (especially PCA) to equine carpal joint bone specimens. There were, however, a few of these tandem applications with other animals.

Ahmed *et al*. (63), for instance, used PCA (and subsequently LDA, along with Raman peak parameters and other characterisation techniques) as part of their assessment of a calvarial defect-healing model in murine parietal bone specimens. The PCA results generated in (63) suggested that the first two PCs, in particular, were able to separate Raman spectral data from the three treatment conditions: *in vivo* defects, control defects, and normal surfaces (areas within 2 mm of defect sites). Both of the first two PCs distinguished between *in vivo* defects and the latter two conditions; the first symmetric phosphate stretch (958 cm^{-1}) was the major source of variation from PC1, whereas it was the CH_2 deformation (1448 cm^{-1}) for PC2.
Fu *et al*. (64) also used PCA (and later, LDA, along with Raman peak parameters and receiver operating characteristic curves (ROC curves)) to determine the effects of ovariectomy on rat mandibular cortical bone specimens as an experimental model for human postmenopausal osteoporosis. As an aside, these ROC graphs determine the performance of a classification model in discriminating between positive and negative tests (65). According to Fawcett (66), such graphs show the compromise between classifiers' 'hit' and 'false alarm' rates. Additionally, quantification of the diagnostic accuracy of a test lies in the measurement of sensitivity (sometimes called the 'true positive rate') and specificity (sometimes called the 'true negative rate') and the trade-off between the two (65-67). With the first three PCs together accounting for more than 85 % of the total variance, their PCA results suggested that the first two PCs more clearly delineated between average Raman spectral data from the two treatment groups – an ovariectomised group (OVX) and a sham-operated group (SHAM) – by the latest post-operation time-point. The bone specimens from each group were harvested at three time-points: two months, four months, or eight months post-operation. There were no PC loadings plots in (64). It could be surmised, though, from the difference spectra and other aspects of their results that the first symmetric phosphate stretch (959 – 960 cm^{-1}) – related to a gradual decrease in the relative mineral content – was likely the main determining factor for separation of the two groups. Mangueira *et al*. also used PCA in their investigation of low-level laser therapy (LLLT)-associated changes to damaged cartilage in collagenase-induced OA murine tibial specimens, alongside Raman spectroscopy and histomorphometry in (68). There were four experimental groups: a 'control' group (GCON), a collagenase-without-treatment group (GCOL), a collagenase-660 nm LLLT group (G660), and a collagenase-780 nm LLLT group (G780). The first three PCs – especially the second and third – highlighted types II and III collagen-related Raman spectral bands (between ~1245 and 1460 cm^{-1}) as the leading sources of LLLT-induced fibrocartilage synthesis (which aids in the collagen-rich cartilage's repair process) for the two treated groups, best separating them from the other two

experimental groups. Such work as done by these parties, and the PCA results therein, demonstrate the capabilities of tandem Raman-chemometrics applications to differentiate between and classify healthy and diseased tissues within biomedical contexts.

The PCA results from the differenced spectral data above in subsection 4.2.2.3, again suggested that some localised microstructural differences were detectable. What was observed in the distribution of aggregated PC scores in the subspace between the two spectral data subsets was their clear separation by layer (articular cartilage and subchondral bone) along the PC1 axis. As noted above, separation by section (third, fourth, upper radial, and lower radial carpal bones) along other PC axes was ambiguous.

PCA results from the differenced data also suggested the likelihood of some disparity between the 'chip' and 'control' subsets within the PC1 dimension; the latter PCs had loadings from one or more functional groups simulating respective loadings from either original spectral subset. Presumably, on the one hand, those loadings that mimicked loadings from the 'chip' subset may have been hinting at some of the OA-associated degenerative processes within the carpal joint bones. On the other hand, if there had been, for instance, a more intense loading from the carbonate-phosphate functionality (~ 1065 – 1071 cm^{-1}) concomitant with a less intense loading from the first symmetric phosphate stretch (~ 857 – 962 cm^{-1}), perhaps more appreciable carbonate accumulation could be inferred (69). Increased carbonate substitution of apatite structures within the SCB of OA bone tissue is also seen as a compensatory occurrence due to its hypomineralisation from abnormal remodelling (69). As such, the first symmetric phosphate stretch being the major source of variation for the first two PCs suggested that there might be the start of some adverse mineralisation occurring within the carpal joint in that respect.

As indicated above, loadings from functional groups associated with the organic matrix had a comparatively more mixed outcome: within PCs 2 and four, many of their loadings were more like those from the 'control' subset, and those from PC3 were more closely related to the 'chip' subset. These results suggested that some aspects of the protein backbone of the carpal joint specimens might have been pointing to the beginning of cartilage degeneration, even as others may still have been maintaining cartilage homeostasis.

The possibility of some organic matrix masking remained, due to the inherent characteristics of the Raman technique – as was seen with the Mt3 bone specimens.

The PCA results from both the stacked spectral data and the mean-adjusted, stacked spectral data above suggested that variations between the two layers of bone (the AC and the SCB)

may have obscured any otherwise observable differences between the two 'conditions' within the data set. As alluded to above in subsection 4.2.2.4, removing the average 'horse-section-layer' effect did not separate the PC scores based on the 'condition' of the bone sections. Nevertheless, it seemingly did so on 'layer', as a 'layer effect' was seen along the PC1 axis in the PC scores plots for the original stacked spectral data. This outcome suggested the possibility that the 'condition effect' might show up in one layer of bone, but not the other. The next analytical step was to determine if adjusting for this would better separate the two conditions.

Despite adjustments for this 'layer effect' and 'condition effect' in the stacked spectral data, the desired separation did not occur. There may have been some other (not-yet-considered) underlying feature that led to the outcome observed in the PCA of the mean-adjusted, stacked spectral data.

It was interesting to note that the PC scores for the SCB were much more scattered than those from the AC were, suggesting that there might have been some phenomenon occurring in the AC that was not yet or as strongly evident in the SCB.

PC scores from the upper part of the radial carpal bone tended to cluster together with those from the third carpal bone, whereas those from the lower part of the radial carpal bone tended to cluster closer to the fourth carpal bones' PC scores. The former was understandable, as the radial and third carpal bones are in anatomic proximity to each other, the latter, however, was a little puzzling, as the fourth carpal bone is not as close to the radial carpal bone. Had the PCA results from the Raman spectral data herein been more unambiguous, some comparison could have been made with similar research work, as in (49), for instance. One part of Kerns *et al*.'s PCA results indicated that there was a more marked distinguishing between OA and non-OA specimens when the PCA was distinctly applied in turn to spectral data from the medial and lateral compartments of tibial specimens. The phosphate and Amide I bands best aided in this partition for the medial compartment; it was the hydroxyproline, proline, Amide I and phosphate-related bands for the lateral compartment. Although different microstructure-related Raman spectral bands contributed to some division between the different specimens in work presented herein, a complete OA-to-non-OA split of bone sections remained elusive.

Relative principal component analysis (RPCA) – seemingly closely related to LDA – is a new method developed by such groups as Ahmad *et al*. (70), which seeks to extract information that best explains changes due to the differences between two data sets. This method contrasts with PCA's extraction of the most sizable variations within one data set. This technique holds

the promise of adaptability for determining differences between sets of spectral data. It would require further understanding before it could be implemented.

Limitations

Laser power may have had some influence over signal strength from the organic functionalities within the carpal joint bone specimens; there was a trade-off between signal strength and fresh tissue preservation for other characterisation methods.

There were also unforeseen time contractions that precluded a more thorough PCA of the Raman spectral data from the equine carpal joint bone sections.

PCA was an appropriate method to use in this context because the real issue in the data sets presented both in Chapter 3 and here in Chapter 4 was the presence of multiple levels. This thesis focused on approaches to correctly analysing multi-level multivariate data sets. In future work, once the methodology for handling multi-level multivariate data sets has been refined, other multivariate methods like *k*-means cluster analysis and soft-independent modelling of class analogy or analogies (SIMCA) could be investigated. SIMCA is somewhat similar to fuzzy *c*- and *k*-means cluster analyses in that an object can simultaneously belong to more than one class – after performing PCA on each class of objects – so that each class is independent of the others (29, 71). Both data sets were too small to have performed something like PLS. The number of observations in both data sets were too small to have given reliable results with PLS because PLS incorporates the observations into the feature construction process, as it inherently aims to find the maximum covariance between the original data matrix, X, and Y (a matrix consisting of some property within the sample or samples that needs prediction) (72). There were a very small number of observations in both data sets, to reiterate. The data presented in this body of work are likely to have much potential within them, but to make another repetition, until the issue of having multiple levels within multivariate data is adequately addressed, that potential cannot be fully realised by interrogating the data with more sophisticated multivariate statistical techniques.

Chapter 5 – Summary

5.1 Summary

Both quantity and quality influence bone strength and health; there needs to be greater attention paid to the nuances of bone quality to understand better the factors that might lead to the development of musculoskeletal diseases (whose pervasiveness are increasing due to greater longevity among more of the world's population). Study of microstructural changes in bone (by way of molecular vibrational spectroscopy) may enhance interpretation of the impact that interactions between and alterations to bone's major components have on its overall quality. A greater grasp of both facets of bone strength and health should eventually encourage better-targeted medical practices: animal models of human medicine seem a reasonable starting point.

Owing to the intricate nature of biological samples such as bone tissue, and the often-complex results of probing its structure, multivariate statistical analytic techniques – chemometrics in the context of analytical chemical and biological data – assist in simplifying, extracting, and categorising or classifying the underlying patterns within the data. Principal component analysis (PCA) and linear discriminant analysis (LDA) are just two of the available quantitative methods.

A prior tandem infrared spectroscopy-chemometrics study was successful in detecting differences in equine bone disease. An analogous, exploratory Raman study of other sets of equine bone specimens should also have been capable of determination. Principal component analysis of aggregated Raman spectral data collection from the fracture-prone, embedded equine third metatarsal (Mt3) bone specimens suggested that some localised microstructural differences were detectable – especially within parts of the subchondral bone. What was unclear, however, was the likely cause of these differences. Similarly, PCA of aggregated Raman spectral data collected from the fresh equine carpal joint bone specimens did not provide any clear underlying factors for potential differences observed between the 'control' and 'induced' osteoarthritic 'chip' subsets. Some of the common questions the PCA results raised included the extent of similarity between individuals with respect to the organic matrix component, and the extent of heterogeneity between individuals with respect to the mineral component.

5.2 Future work

Any potential applicable clinical predictions for orthopaedic disorders like osteoporosis and OA necessitate validation and classification of the Raman spectral data beforehand. It was thought that PCA could have classified the Raman spectral data from the embedded equine Mt3 bone specimens according to the 'status' of the bone sections (since no assumptions were made about there being differences between the groups).
It was thought that PCA could also have classified the Raman spectral data from the equine carpal joint bone specimens according to the 'condition' of the bone sections. The results did not truly reflect this. Widening the scope for future data analysis to include LDA (which would process spectra to detect any differences relating to the groups to see if this will help clarify the classification), would be one of the first steps. The issue of proper handling of multi-level multivariate data sets needs to be fully addressed, though, before LDA and other more sophisticated multivariate analytical techniques can be applied to the data.

Some recommendations for future research include the following: refined adaptation of the multi-laser Raman microscope system set-up to better handle spectral data collection from biological samples like the opaque bone sections (perhaps with an upright mode). A higher-contrast epi-illumination set-up with either or both of a wider range of magnification objectives and cameras such that it would include those that can provide a view of the bone section layer features in greater detail.

As to continuation and expansion of this work, validation and classification of the Raman spectral data set would be necessary in order for any potential predictions to be applicable. If not readily available, condensing instrumentation into fibre optics might enable trialling of this technique in a practical, clinical setting. Perhaps – if feasible practically – even instrumentation that combines the two vibrational spectroscopic techniques in tandem with chemometrics, in order to provide simultaneous information about groupings of samples.

References

1. Khan K, McKay H, Kannus P, Bailey D, Wark J, Bennell K. Physical activity and bone. Champaign, IL: Human Kinetics; 2001.

2. Piliner S, Davies Z. Equine science. 2nd ed. Oxford, England: Blackwell Publications; 2004.

3. Wojnar R. Bone and cartilage – its structure and physical properties. 2010. In: Biomechanics of hard tissues: modeling, testing, and materials [Internet]. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA.; [1-75].

4. Garip S, Boskey AL. Diagnosis of bone and cartilage diseases. 2012. In: Vibrational spectroscopy in diagnosis and screening [Internet]. Amsterdam, Netherlands: IOS Press; [272-303].

5. Riggs CM. Fractures – a preventable hazard of racing Thoroughbreds? The Veterinary Journal. 2002;163(1):19-29.

6. Pasquini C, Spurgeon T, Pasquini S. Anatomy of domestic animals: systemic & regional approach. Pilot Point, TX: Sudz Publications; 2007.

7. Sinclair C, Birch HL, Smith RKW, Goodship AE. Skeletal physiology: responses to exercise and training. 2014. In: Equine sports medicine and surgery: basic and clinical sciences of the equine athlete [Internet]. London, England: Elsevier Health Sciences UK. 2nd. [145-66].

8. Boskey AL, Camacho NP. FT-IR imaging of native and tissue-engineered bone and cartilage. Biomaterials. 2007;28(15):2465-78.

9. Zhang G, Chan KLA, Flach CR, Mendelsohn R. Interplay of univariate and multivariate analysis in vibrational microscopic imaging of mineralized tissue and skin. 2008. In: Biomedical vibrational spectroscopy [Internet]. Hoboken, NJ: John Wiley & Sons, Inc.; [357-78].

10. Choo-Smith L-P, Ko AC-T, Hewko M, Popescu DP, Friesen J, Sowa MG. Combining optical coherence tomography and Raman spectroscopy for investigating dental and other mineralized tissues. 2008. In: Biomedical vibrational spectroscopy [Internet]. Hoboken, NJ: John Wiley & Sons, Inc.; [263-90].

11. Kanis JA, McCloskey EV, Johansson H, Odén A, Melton III LJ, Khaltaev N. A reference standard for the description of osteoporosis. Bone. 2008;42(3):467-75.

12. Fox C, Edwards MH, Dennison EM, Cooper C. Personal and societal burden of osteoporotic fractures. Clinical Reviews in Bone and Mineral Metabolism. 2015;13(2):53-60.

13. Johnell O, Kanis JA. An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. Osteoporosis International. 2006;17(12):1726-33.

14. Oleson CV. Osteoporosis rehabilitation: a practical approach. Cham, Switzerland: Springer International Publishing; 2017.

15. Emer MÖ, İnce S, Arslan N. Bone mineral densitometry: measurement and evaluation methods. 2016. In: Musculoskeletal research and basic science [Internet]. Cham, Switzerland: Springer International Publishing; [197-212].

16. Fonseca H, Moreira-Gonçalves D, Coriolano H-JA, Duarte JA. Bone quality: the determinants of bone strength and fragility. Sports Medicine. 2014;44(1):37-53.

17. Gamsjaeger S, Mendelsohn R, Boskey AL, Gourion-Arsiquand S, Klaushofer K, Paschalis EP. Vibrational spectroscopic imaging for the evaluation of matrix and mineral chemistry. Current Osteoporosis Reports. 2014;12(4):454-64.

18. Ruppel ME, Miller LM, Burr DB. The effect of the microscopic and nanoscale structure on bone fragility. Osteoporosis International. 2008;19(9):1251-65.

19. Cooper C, Cole ZA, Holroyd CR, Earl SC, Harvey NC, Dennison EM, et al. Secular trends in the incidence of hip and other osteoporotic fractures. Osteoporosis International. 2011;22(5):1277-88.

20. Eastell R. Osteoporosis. Medicine. 2013;41(10):586-91.

21. Innes JF, Clegg P. Comparative rheumatology: what can be learnt from naturally occurring musculoskeletal disorders in domestic animals? Rheumatology. 2010;49(6):1030-9.

22. McIlwraith CW, Frisbie DD, Kawcak CE. The horse as a model of naturally occurring osteoarthritis. Bone and Joint Research. 2012;1(11):297-309.

23. Smith RKW, Garvican ER, Fortier LA. The current 'state of play' of regenerative medicine in horses: what the horse can tell the human. Regenerative Medicine. 2014;9(5):673-85.

24. Yu C, Gestl E, Eckert K, Allara D, Irudayaraj J. Characterization of human breast epithelial cells by confocal Raman microspectroscopy. Cancer Detection and Prevention. 2006;30(6):515-22.

25. Rehman IU, Movasashi Z, Rehman S. Vibrational spectroscopy for tissue analysis. Boca Raton, FL: CRC Press; 2013.

26. Moore DS, Jepsen PU, Volka K. Principles of vibrational spectroscopic methods and their application to bioanalysis. 2014. In: Handbook of spectroscopy [Internet]. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. 2nd. [1037-78].

27. Moore DS. Raman spectroscopy fundamentals. 2014. In: Handbook of spectroscopy [Internet]. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. 2nd. [1813-29].

28. Morris MD, Mandair GS. Raman assessment of bone quality. Clinical Orthopaedics and Related Research. 2011;469(8):2160-9.

29. Brereton RG. Applied chemometrics for scientists. Chichester, England: John Wiley & Sons, Ltd.; 2007.

30. Severcan F, Akkas SB, Turker S, Yucel R. Methodological approaches from experimental to computational analysis in vibrational spectroscopy and microspectroscopy. 2012. In: Vibrational spectroscopy in diagnosis and screening [Internet]. Amsterdam, Netherlands: IOS Press; [12-52].

31. Ural A, Vashishth D. Hierarchical perspective of bone – from molecules to fracture. International Materials Reviews. 2014;59(5):245-63.

32. Ager III JW, Balooch G, Ritchie RO. Fracture, aging, and disease in bone. Journal of Materials Research. 2006;21(8):1878-92.

33. Lavine BK, Workman Jr. J. Chemometrics. Analytical Chemistry. 2013;85(2):705-14.

34. Chen P, Shen A, Zhou X, Hu J. Bio-Raman spectroscopy: a potential clinical analytical method assisting in disease diagnosis. Analytical Methods. 2011;3(6):1257-69.

35. Gamsjaeger G, Mendelsohn R, Klaushofer K, Paschalis EP. Vibrational spectroscopic analysis of hard tissues. 2014. In: Infrared and Raman spectroscopic imaging [Internet]. Wiley-VCH Verlag GmbH & Co. KGaA. 2nd. [153-80].

36. Buchwald T, Kozielski M, Szybowicz M. Determination of collagen fibers arrangement in bone tissue by using transformations of Raman spectra maps. Journal of Spectroscopy. 2012;27(2):107-17.

37. Nicholson CL, Firth EC, Waterland MR, Jones G, Ganesh S, Stewart RB. Innovative approach to investigating the microstructure of calcified tissues using specular reflectance Fourier transform-infrared microspectroscopy and discriminant analysis. Analytical Chemistry. 2012;84(7):3369-75.

38. Khalid M, Bora T, Al Ghaithi A, Thukral S, Dutta J. Raman spectroscopy detects changes in bone mineral quality and collagen cross-linkage in Staphylococcus infected human bone. Scientific Reports. 2018;8(1):9417.

39. de Souza RA, Xavier M, Mangueira NM, Santos AP, Pinheiro ALB, Villaverde AB, et al. Raman spectroscopy detection of molecular changes associated with two experimental models of osteoarthritis in rats. Lasers in Medical Science. 2014;29(2):797-804.

40. Bonifacio A, Beleites C, Vittur F, Marsich E, Semeraro S, Paoletti S, et al. Chemical imaging of articular cartilage sections with Raman mapping, employing uni- and multi-variate methods for data analysis. Analyst. 2010;135(12):3193-204.

41. Toledano M, Toledano-Osorio M, Guerado E, Caso E, Aguilera FS, Osorio R. Biochemical assessment of nanostructures in human trabecular bone: proposal of a Raman microspectroscopy based measurements protocol. Injury. 2018;49(S2):S11-S21.

42. McLaughlin G, Lednev IK. Spectroscopic discrimination of bone samples from various species. American Journal of Analytical Chemistry. 2012;3(2):161-7.

43. Shimoyama M, Morimoto S, Ozaki Y. Non-destructive analysis of the two subspecies of African elephants, mammoth, hippopotamus, and sperm whale ivories by visible and short-wave near infrared spectroscopy and chemometrics. Analyst. 2004;129(6):559-63.

44. Brody RH, Edwards HGM, Pollard AM. Chemometric methods applied to the differentiation of Fourier-transform Raman spectra of ivories. Analytica Chimica Acta. 2001;427(2):223-32.

45. Buckley K, Kerns JG, Vinton J, Gikas PD, Smith C, Parker AW, et al. Towards the *in vivo* prediction of fragility fractures with Raman spectroscopy. Journal of Raman Spectroscopy. 2015;46(7):610-8.

46. Nyman JS, Makowski AJ, Patil CA, Masui TP, O'Quinn EC, Bi X, et al. Measuring differences in compositional properties of bone tissue by confocal Raman spectroscopy. Calcified Tissue International. 2011;89(2):111-22.

47. Kazanci M, Roschger P, Paschalis EP, Klaushofer K, Fratzl P. Bone osteonal tissues by Raman spectral mapping: orientation-composition. Journal of Structural Biology. 2006;156(3):489-96.

48. Kazanci M, Wagner HD, Manjubala NI, Gupta HS, Paschalis EP, Roschger P, et al. Raman imaging of two orthogonal planes within cortical bone. Bone. 2007;41(3):456-61.

49. Kerns JG, Gikas PD, Buckley K, Shepperd A, Birch HL, McCarthy I, et al. Evidence from Raman spectroscopy of a putative link between inherent bone matrix chemistry and degenerative joint disease. Arthritis & Rheumatology. 2014;66(5):1237-46.

50. Firth EC, Doube M, Boyde A. Changes in mineralised tissues at the site of origin of condylar fracture are present before athletic training in Thoroughbred horses. New Zealand Veterinary Journal. 2009;57(5):278-83.

51. Eilers PHC, Boelens HFM. Baseline correction with asymmetric least squares smoothing. Leiden University Medical Centre Report. 2005;1(1):5.

52. Morris MD, Finney WF. Recent developments in Raman and infrared spectroscopy and imaging of bone tissue. Spectroscopy. 2004;18(2):155-9.

53. Mandair GS, Morris MD. Contributions of Raman spectroscopy to the understanding of bone strength. BoneKEy Reports. 2015;4.

54. France CAM, Thomas DB, Doney CR, Madden O. FT-Raman spectroscopy as a method for screening collagen diagenesis in bone. Journal of Archaeological Science. 2014;42:346-55.

55. van der Harst MR, van de Lest CHA, Degroot J, Kiers GH, Brama PAJ, van Weeren PR. Study of cartilage and bone layers of the bearing surface of the equine metacarpophalangeal joint relative to different timescales of maturation. Equine Veterinary Journal. 2005;37(3):200-6.

56. Isaksson H, Harjula T, Koistinen A, Ivarinen J, Seppänen K, Arokoski JPA, et al. Collagen and mineral deposition in rabbit cortical bone during maturation and growth: effects on tissue properties. Journal of Orthopaedic Research. 2010;28(12):1626-33.

57. Kobrina Y, Turunen MJ, Saarakkala S, Jurvelin JS, Hauta-Kasari M, Isaksson H. Cluster analysis of infrared spectra of rabbit cortical bone samples during maturation and growth. Analyst. 2010;135(12):3147-55.

58. Doube M, Firth EC, Boyde A, Bushby AJ. Combined nanoindentation testing and scanning electron microscopy of bone and articular calcified cartilage in an equine fracture predilection site. European Cells and Materials. 2010;19:242-51.

59. Wang L, Mizaikoff B. Application of multivariate data analysis techniques to biomedical diagnostics based on mid-infrared spectroscopy. Analytical and Bioanalytical Chemistry. 2008;391(5):1641-54.

60. Vincent TL, Watt FE. Osteoarthritis. Medicine. 2018;46(3):187-95.

61. Li G, Yin J, Gao J, Cheng TS, Pavlos NJ, Zhang C, et al. Subchondral bone in osteoarthritis: insight into risk factors and microstructural changes. Arthritis Research & Therapy. 2013;15(6):223-34.

62. Frisbie DD, Ghivizzani SC, Robbins PD, Evans CH, McIlwraith CW. Treatment of experimental equine osteoarthritis by in vivo delivery of the equine interleukin-1 receptor antagonist gene. Gene Therapy. 2002;9(1):12-20.

63. Ahmed R, Wang W, Zia AW, Lau C. Collagen formation observed from healing calvarial defects with principal component analysis of Raman scattering. Analyst. 2018;143(19):4614-22.

64. Fu X, Chen J, Wu D, Du Z, Lei Q, Cai Z, et al. Effects of ovariectomy on rat mandibular cortical bone: a study using Raman spectroscopy and multivariate analysis. Analytical Chemistry. 2012;84(7):3318-23.

65. Lasko TA, Bhagwat JG, Zhou KH, Ohno-Machado L. The use of receiver operating characteristic curves in biomedical informatics. Journal of Biomedical Informatics. 2005;38(5):404-15.

66. Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters. 2006;27(8):861-74.

67. van Erkel AR, Pattynama PM. Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology. European Journal of Radiology. 1998;27(2):88-94.

68. Mangueira NM, Xavier M, de Souza RA, Salgado MAC, Silveira Jr. L, Villaverde AB. Effect of low-level laser therapy in an experimental model of osteoarthritis in rats evaluated through Raman spectroscopy. Photomedicine and Laser Surgery. 2015;33(3):145-53.

69. Buchwald T, Niciejewski K, Kozielski M, Szybowicz M, Siatkowski M, Krauss H. Identifying compositional and structural changes in spongy and subchondral bone from the hip joints of patients with osteoarthritis using Raman spectroscopy. Journal of Biomedical Optics. 2012;17(1):017007.

70. Ahmad M, Helms V, Kalinina OV, Lengauer T. Relative principal components analysis: application to analyzing biomolecular conformational changes. Journal of Chemical Theory and Computation. 2019;15(4):2166-78.

71. Krafft C, Steiner G, Beleites C, Salzer R. Disease recognition by infrared and Raman spectroscopy. Journal of Biophotonics. 2009;2(1-2):13-28.

72. Wehrens R. Chemometrics with R: multivariate data analysis in the natural sciences and life sciences. Heidelberg, Germany: Springer-Verlag; 2011.

Appendices

Appendix A – Exemplar of the type of R code used for multivariate statistical analysis from the equine third metatarsal (Mt3) and carpal joint bone specimens

Below is only a fraction of the R code used to enable multivariate statistical analysis – namely principal component analysis – of Raman spectral data collected from the two sets of equine bone specimens (third metatarsal (Mt3) and carpal joint).

The first block of code was to enable R to read the spectral data .csv files correctly – via the *ChemoSpec* package – after having loaded libraries of various packages that would be necessary to enable data inputting, pre-processing, plotting, and analysis. These 'spectral objects' were saved to avoid constantly repeating the read-in process:

```
library(baseline)
library(ChemoSpec)
library(R.utils)
library(wavelets)
library(signal)
library(ggplot2)
library(ggrepel)
library(gtools)
library(MASS)
library(pBrackets)


# Read in & convert spectral data csv files into a 'SpectraObject' in ChemoSpec package
```

```
 files2SpectraObject(gr.crit = c("RSL\\d+\\s{1}C1", "RSL\\d+\\s{1}C2", "RSL\\d+\\s{1}C3",
                                 "98B\\d+\\s{1}C1", "98B\\d+\\s{1}C2", "98B\\d+\\s{1}C3",
                                 "M(\\d+|\\d+PE)\\s{1}C1", "M(\\d+|\\d+PE)\\s{1}C2", "M(\\d+|\\d+PE)\\s{1}C3"),
                     gr.cols = c("olivedrab1", "olivedrab3", "olivedrab4",
                                 "orange1", "orange3", "orange4",
                                 "firebrick1", "firebrick3", "firebrick4"),
                     freq.unit = "Raman shift (cm^-1)", int.unit = "intensity",
                     descrip = "Thoroughbred racehorse Mt3 bone specimen study",
                     fileExt = "\\.(csv|CSV)$", header = TRUE, sep = ",", dec = ".", out.file = "bone2")
 bone2.Raman <- loadObject("bone2.RData")


# Inspect several ways
sumSpectra(bone2.Raman)
str(bone2.Raman)
levels(bone2.Raman$groups)


# Fix the names of the horse groups; need to keep groups as a factor
levels(bone2.Raman$groups) <- list(
"RSL_C1"   =   "RSL\\d+\\s{1}C1",
"RSL_C2"   =   "RSL\\d+\\s{1}C2",
"RSL_C3"   =   "RSL\\d+\\s{1}C3",
"98B_C1"   =   "98B\\d+\\s{1}C1",
"98B_C2"   =   "98B\\d+\\s{1}C2",
"98B_C3" = "98B\\d+\\s{1}C3",
```

```
"M_C1" = "M(\\d+|\\d+PE)\\s{1}C1",
"M_C2" = "M(\\d+|\\d+PE)\\s{1}C2",
"M_C3" = "M(\\d+|\\d+PE)\\s{1}C3")

levels(bone2.Raman$groups)

sumSpectra(bone2.Raman)

# Fix alt symbols
bone2.Raman$alt.sym <- ifelse(bone2.Raman$group == "RSL_C1", "a", bone2.Raman$alt.sym)
bone2.Raman$alt.sym <- ifelse(bone2.Raman$group == "RSL_C2", "b", bone2.Raman$alt.sym)
bone2.Raman$alt.sym <- ifelse(bone2.Raman$group == "RSL_C3", "c", bone2.Raman$alt.sym)

bone2.Raman$alt.sym <- ifelse(bone2.Raman$group == "98B_C1", "d", bone2.Raman$alt.sym)
bone2.Raman$alt.sym <- ifelse(bone2.Raman$group == "98B_C2", "e", bone2.Raman$alt.sym)
bone2.Raman$alt.sym <- ifelse(bone2.Raman$group == "98B_C3", "f", bone2.Raman$alt.sym)

bone2.Raman$alt.sym <- ifelse(bone2.Raman$group == "M_C1", "g", bone2.Raman$alt.sym)
bone2.Raman$alt.sym <- ifelse(bone2.Raman$group == "M_C2", "h", bone2.Raman$alt.sym)
```

```
bone2.Raman$alt.sym <- ifelse(bone2.Raman$group == "M_C3", "i", bone2.Raman$alt.sym)


# Check results
sumSpectra(bone2.Raman)


unique(bone2.Raman$alt.sym)

# Now use the ifelse strategy to fix the colours and symbols
# Samples:

bone2.Raman$colors <- ifelse(bone2.Raman$group == "RSL_C1", "olivedrab1", bone2.Raman$colors)
bone2.Raman$sym <- ifelse(bone2.Raman$group == "RSL_C1", 1, bone2.Raman$sym)


bone2.Raman$colors <- ifelse(bone2.Raman$group == "RSL_C2", "olivedrab3", bone2.Raman$colors)
bone2.Raman$sym <- ifelse(bone2.Raman$group == "RSL_C2", 2, bone2.Raman$sym)


bone2.Raman$colors <- ifelse(bone2.Raman$group == "RSL_C3", "olivedrab4", bone2.Raman$colors)
bone2.Raman$sym <- ifelse(bone2.Raman$group == "RSL_C3", 3, bone2.Raman$sym)


bone2.Raman$colors <- ifelse(bone2.Raman$group == "98B_C1", "orange1", bone2.Raman$colors)
bone2.Raman$sym <- ifelse(bone2.Raman$group == "98B_C1", 4, bone2.Raman$sym)
```

```
bone2.Raman$colors <- ifelse(bone2.Raman$group == "98B_C2", "orange3", bone2.Raman$colors)
bone2.Raman$sym <- ifelse(bone2.Raman$group == "98B_C2", 5, bone2.Raman$sym)


bone2.Raman$colors <- ifelse(bone2.Raman$group == "98B_C3", "orange4", bone2.Raman$colors)
bone2.Raman$sym <- ifelse(bone2.Raman$group == "98B_C3", 6, bone2.Raman$sym)


bone2.Raman$colors <- ifelse(bone2.Raman$group == "M_C1", "firebrick1", bone2.Raman$colors)
bone2.Raman$sym <- ifelse(bone2.Raman$group == "M_C1", 7, bone2.Raman$sym)


bone2.Raman$colors <- ifelse(bone2.Raman$group == "M_C2", "firebrick3", bone2.Raman$colors)
bone2.Raman$sym <- ifelse(bone2.Raman$group == "M_C2", 8, bone2.Raman$sym)

bone2.Raman$colors <- ifelse(bone2.Raman$group == "M_C3", "firebrick4", bone2.Raman$colors)
bone2.Raman$sym <- ifelse(bone2.Raman$group == "M_C3", 9, bone2.Raman$sym)


# Be sure to run sumSpectra when done as it checks things internally
sumSpectra(bone2.Raman)


# To save bone.Raman for future use without re-running above code
library(R.utils)

saveObject(bone2.Raman, file = "bone2.Raman.RData")
```

```
# To retrieve bone.Raman when starting up at a later point
bone2_Raman <- loadObject(file = "bone2.Raman.RData")
```

Here was the baseline correction procedure itself, beginning with conversion of the spectral data into matrix form. This conversion was followed by transposition of the truncated spectral data to row form to be amenable to later statistical analysis. The spectral intensities were scaled by single normal variate (SNV; that is, scaled by standard deviation and mean-centred) before the asymmetric least squares (ALS) baseline correction algorithm was created and calculated to the SNV-scaled data frame. As a reminder, the actual parameters for the ALS method were: second derivative constraint, λ; weighting of positive residuals, *p*; and maximum number of iterations, *maxit*. The final step was Savitzky-Golay filtering of the baseline-corrected Raman spectral data matrix (parameters: 1 = performed on matrix row, with one spectrum per row; filter length, *n*; and m^{th} derivative of the filter coefficients, *m*). The column of wavelengths (Raman frequencies) was created as a separate data frame:

```
## Baseline correction

# Convert Raman spectral data into a matrix

bone2.Raman.frame <- as.matrix(bone2_Raman$data)
rownames(bone2.Raman.frame) <- as.character(c(bone2_Raman$names))

.

.

.

# To save bone2.Raman.frame for future use without re-running above code
library(R.utils)

saveObject(bone2.Raman.frame, file = "bone2.Raman.frame.RData")
```

```
# To retrieve bone2.Raman.frame when starting up at a later point
bone2.Raman.frame <- loadObject(file = "bone2.Raman.frame.RData")


freq2 <- as.data.frame(bone2_Raman$freq)



# SNV scaling of transposed Raman spectral data matrix (transposed so samples/cases as rows)
# Baseline correction using asymmetric least squares (ALS) method

bone2.Raman.SNV <- scale(t(bone2.Raman.frame[, -(1:300)]), center = TRUE, scale = TRUE)
bone2.cor <- baseline.als(t(bone2.Raman.SNV), lambda = 4, p = 0.001, maxit = 10)


# Baseline creation

bone2.baseline <- as.data.frame(bone2.cor$baseline)
bone2.baselinecorrected <- as.data.frame(bone2.cor$corrected)


# Savitzky-Golay derivation/smoothing of baseline-corrected Raman spectral data matrix
bone2.cor.als.SG <- apply(bone2.baselinecorrected, 1, sgolayfilt, n = 13, m = 0)

.

.

.
```

The pre-processed Raman spectral data were subsequently saved as a new .csv file and read back into R before additional details about the horses were incorporated into the data frame:

```
## Prepare the Savitzky-Golay-smoothed, ALS baseline-corrected Raman spectral data for PCA
# by including extra information about horses; save as csv file

View(bone2.cor.als.SG)

Horses <- t(bone2.cor.als.SG)
View(Horses)

write.csv(Horses, "C:/…/Horses.csv")



# Make a copy of 'Horses' csv file to be able to add in details about
# age, horse ID, section & location

# Read csv back into R

Horses <- read.csv("C:/…/Horses2.csv", header = TRUE, sep = ",")
View(Horses)


# Assign numeric values to age groups for horses
Horses$Age <- 1

Horses$Age[substr(Horses$Horse, start = 1, stop = 1) == "9"] <- 2
Horses$Age[substr(Horses$Horse, start = 1, stop = 1) == "M"] <- 3


# Assign values '1', '2' & '3' to sections C1, C2 & C3, respectively
```

```
Horses$Section <- as.numeric(Horses$Section)



# Check the storage mode of the 'Horse' column
mode(Horses$Horse)


# Create a new variable, 'Horse ID'
Horses$HorseID <- as.numeric(Horses$Horse)


Horses$HorseID <- ifelse(Horses$Age == 1, Horses$HorseID - 20, Horses$HorseID)
Horses$HorseID <- ifelse(Horses$Age == 2, Horses$HorseID + 10, Horses$HorseID)
Horses$HorseID <- ifelse(Horses$Age == 3, Horses$HorseID + 10, Horses$HorseID)


# Fix order of 'Horse ID' for 98Bs (5-month-old foals)


 Horses$HorseID[Horses$Horse == "98B4"] <- 11
 Horses$HorseID[Horses$Horse == "98B6"] <- 12
 Horses$HorseID[Horses$Horse == "98B7"] <- 13
 Horses$HorseID[Horses$Horse == "98B8"] <- 14
 Horses$HorseID[Horses$Horse == "98B9"] <- 15
 Horses$HorseID[Horses$Horse == "98B11"] <- 16
 Horses$HorseID[Horses$Horse == "98B16"] <- 17
```

```
 Horses$HorseID[Horses$Horse == "98B18"] <- 18
 Horses$HorseID[Horses$Horse == "98B25"] <- 19
 Horses$HorseID[Horses$Horse == "98B30"] <- 20
 .
 .
 .
# Create another variable, 'Layer',
# by splitting up 'Location' into specific layers: 'AC' & 'SCB'
Horses$Layer <- 1
Horses$Layer[Horses$Age == 1 & Horses$Location > 6] <- 2
Horses$Layer[Horses$Age == 2 & Horses$Location > 6] <- 2
Horses$Layer[Horses$Age == 3 & Horses$Location > 13] <- 2
.
.
.
```

The fully-detailed data frame was saved as a separate .csv file and read back into R. These data were aggregated by computing the means to minimise horse-related bias and between-object variability to obtain multivariate statistical analytical results that should have been more comprehensible later on. As pointed out in subsection 2.5.1 of Chapter 2, having some ‘balance’ in the data set, with the same number of observations for each combination of horse, section, and layer, an estimate of the ‘horse’ influence could be made by averaging out over the other influences (section and layer). Subtracting out the horse-level averages from the data structure should have made spectral variations

due to the influence of section and layer more apparent. These aggregated data were saved as a final .csv file to be read back into R, ready for principal component analysis (PCA):

```
# Save 'Horses' data frame as another csv file

# containing all above added details about horses
write.csv(Horses, "C:/…/Horses3.csv")


# Read in fully detailed csv file of 'Horses'

Horses <- read.csv("C:/…/Horses3.csv", header = TRUE, sep = ",")


# Compute the averages for the variables in 'Horses',
# grouped according to the Horse, Section & Layer that
# each spectrum comes from

Horses <- Horses[, -(2), drop = FALSE]

Horses.aggdata.HSL <- aggregate(Horses, by = list(Horses$HorseID, Horses$Section, Horses$Layer), FUN = mean)
View(Horses.aggdata.HSL)


# Remove columns 'Group.1', 'Group.2', 'Group.3' & 'Location'
Horses.aggdata.HSL.new <- Horses.aggdata.HSL[, -c(1:3, 7)]
View(Horses.aggdata.HSL.new)


# To average spectra by 'Horse'
```

```
Horses.aggdata.H <- aggregate(Horses.aggdata.HSL.new, by = list(Horses.aggdata.HSL.new$HorseID), FUN = mean)
Horses.aggdata.H.new <- Horses.aggdata.H[, -c(1:2, 4:5)]

View(Horses.aggdata.H.new)



# Create another data frame that contains only information about 'HorseID'
Horses.aggdata.HSL.Honly <- Horses.aggdata.HSL.new[, 1:4]
View(Horses.aggdata.HSL.Honly)


# Merge 'Horses.aggdata.HSL.Honly' & 'Horses.aggdata.H.new' data frames

Horses.aggdata.HSLmerged <- merge(Horses.aggdata.HSL.Honly, Horses.aggdata.H.new, by = "HorseID")
View(Horses.aggdata.HSLmerged)

Horses.aggdata.matrix <- as.matrix(Horses.aggdata.HSL.new) - as.matrix(Horses.aggdata.HSLmerged)

View(Horses.aggdata.matrix)

Horses.aggdata <- cbind(Horses.aggdata.HSL.new[, 1:4], Horses.aggdata.matrix[, 5:1044])
Horses.aggdata <- as.data.frame(Horses.aggdata)

View(Horses.aggdata)



# Save 'Horses.aggdata' as csv file
write.csv(Horses.aggdata, "C:/…/Horses.aggdata.csv")
```

```
# Read 'Horses.aggdata' back in as csv file

Horses.aggdata <- read.csv("C:/…/Horses.aggdata.csv", header = TRUE, sep = ",")
View(Horses.aggdata)
```

This final block of code shows PCA was conducted on the covariance matrix of this aggregated data frame. The PCA generated a summary of the first six principal components (PCs), the scree plot, and loadings and scores matrices and their respective plots (with the scores plots being labelled by horse, and combinations of section and layer that the aggregated data PC scores came from). Pairwise PC scores plots and a PC scores plot matrix could also be constructed:

```
## Perform PCA on collapsed data frame, 'Horses.aggdata',
# labelling by Section &/or Layer

Horses.aggdata.pca <- prcomp(Horses.aggdata[, -(1:4)])
print(summary(Horses.aggdata.pca)$importance[, 1:6], digits = 4)


par(mfrow = c(1, 1))
screeplot(Horses.aggdata.pca)


# Generate Loadings & Scores Matrices

LoadMat3 <- Horses.aggdata.pca$rotation
ScoreMat3 <- predict(Horses.aggdata.pca)


# Look at the structure of the Loadings & Scores Matrices
str(LoadMat3)
```

```
str(ScoreMat3)

n3 <- dim(ScoreMat3)[1]; n3



# Generate Loadings plots for 1st 6 PCs

 layout(matrix(c(1, 1, 4, 4, 2, 2, 5, 5, 3, 3, 6, 6), nrow = 3, byrow = TRUE), respect = TRUE)

 par(oma = c(3, 2, 4, 2), mar = c(4, 2, 2, 2))

 plot(freq2[-(1:300), ], LoadMat3[, 1], type = "l",

      xlab = "", ylab = "", main = "PC 1", col = 1, lty = 1,

      cex.lab = 1, cex.axis = 1, cex.main = 1, cex.sub = 1, xlim = c(300, 1800), ylim = c(-0.3, 0.3))
 abline(h = 0, lty = 2, col = "red")

 plot(freq2[-(1:300), ], LoadMat3[, 2], type = "l",

      xlab = "", ylab = "", main = "PC 2", col = 1, lty = 1,

      cex.lab = 1, cex.axis = 1, cex.main = 1, cex.sub = 1, xlim = c(300, 1800), ylim = c(-0.3, 0.3))
 abline(h = 0, lty = 2, col = "red")

 plot(freq2[-(1:300), ], LoadMat3[, 3], type = "l",

      xlab = "", ylab = "", main = "PC 3", col = 1, lty = 1,

      cex.lab = 1, cex.axis = 1, cex.main = 1, cex.sub = 1, xlim = c(300, 1800), ylim = c(-0.3, 0.3))
 abline(h = 0, lty = 2, col = "red")

 plot(freq2[-(1:300), ], LoadMat3[, 4], type = "l",
```

```
     xlab = "", ylab = "", main = "PC 4", col = 1, lty = 1,
     cex.lab = 1, cex.axis = 1, cex.main = 1, cex.sub = 1, xlim = c(300, 1800), ylim = c(-0.3, 0.3))
abline(h = 0, lty = 2, col = "red")
plot(freq2[-(1:300), ], LoadMat3[, 5], type = "l",
     xlab = "", ylab = "", main = "PC 5", col = 1, lty = 1,
     cex.lab = 1, cex.axis = 1, cex.main = 1, cex.sub = 1, xlim = c(300, 1800), ylim = c(-0.3, 0.3))
abline(h = 0, lty = 2, col = "red")
plot(freq2[-(1:300), ], LoadMat3[, 6], type = "l",
     xlab = "", ylab = "", main = "PC 6", col = 1, lty = 1,
     cex.lab = 1, cex.axis = 1, cex.main = 1, cex.sub = 1, xlim = c(300, 1800), ylim = c(-0.3, 0.3))
abline(h = 0, lty = 2, col = "red")
mtext(text = expression("Raman shift" ~ (cm^{-1})), side = 1, line = 0, outer = TRUE)
mtext(text = "Loadings", side = 2, line = -1, outer = TRUE)
mtext(text = (expression(paste("1"^"st", " 6 PC loadings for equine Mt3 bone sections"))), side = 3, line = 1,
outer = TRUE)


# PC1 = 64.9 %, PC2 = 17.8 %, PC3 = 10.8 %, PC4 = 2.3 %, PC5 = 1.3 %, PC 6 = 0.824 %


# Generate Scores plots for pairs of 1st 6 PCs
# Create a colour vector corresponding to levels
```

```
# of Layer, Section & combinations of the 2
# in 'Horses.aggdata'

Lcol <- Horses.aggdata$Layer
Scol <- Horses.aggdata$Section
LScol <- 2*(Scol - 1) + Lcol

# Scores plots by Layer
par(par(mfrow = c(1, 1)))

plot(x = ScoreMat3[, 1], y = ScoreMat3[, 2], xlab = "PC 1 (64.9 %)", ylab = "PC 2 (17.8 %)", pch = 20, col =
Lcol, frame = TRUE)

legend(x = 0, y = 8, xpd = TRUE, horiz = TRUE, legend = c("1", "2"), col = 1:2, title = "Layer", pch = 20,
pt.cex = 1, cex = 0.7)

abline(h = 0, v = 0, lty = 2)

text(ScoreMat3[, 1:2], label = Horses.aggdata$HorseID, cex = 0.7, pos = 3)

.

.

.

# Scores plots by Section
par(par(mfrow = c(1, 1)))

plot(x = ScoreMat3[, 1], y = ScoreMat3[, 2], xlab = "PC 1 (64.9 %)", ylab = "PC 2 (17.8 %)", pch = 20, col =
Scol)

legend(x = 0, y = 8, xpd = TRUE, horiz = TRUE, legend = c("1", "2", "3"), col = 1:3, title = "Section", pch =
20, pt.cex = 1, cex = 0.7)
```

```
abline(h = 0, v = 0, lty = 2)

text(ScoreMat3[, 1:2], label = Horses.aggdata$HorseID, cex = 0.7, pos = 3)

.

.

.

# Scores plots by Section & Layer
par(par(mfrow = c(1, 1)))

plot(x = ScoreMat3[, 1], y = ScoreMat3[, 2], xlab = "PC 1 (64.9 %)", ylab = "PC 2 (17.8 %)", pch = 20, col =
LScol)

legend(x = -4, y = 8, xpd = TRUE, horiz = TRUE, legend = c("S1L1", "S1L2", "S2L1", "S2L2", "S3L1", "S3L2"), col

= 1:6, title = "Section & Layer", pch = 20, pt.cex = 1, cex = 0.7)
abline(h = 0, v = 0, lty = 2)

text(ScoreMat3[, 1:2], label = Horses.aggdata$HorseID, cex = 0.7, pos = 3)

.

.

.

# Scores plots matrix by Section & Layer

par(par(mfrow = c(1, 1)))

pairs(ScoreMat3[, 1:6], pch = 20, cex = 0.7, col = LScol, upper.panel = NULL)

mtext(text = (expression(paste("Paired scores plots for ", "1"^"st", " 6 PCs from equine Mt3 bone sections by
section & layer"))), side = 3, line = 12, outer = TRUE)
```

```
legend(x = -2, y = 5.8, xpd = NA, horiz = TRUE, legend = c("S1L1", "S1L2", "S2L1", "S2L2", "S3L1", "S3L2"), col =
1:6, bty = "n", pch = 20, pt.cex = 1.5, cex = 0.85)
```

Appendix B – Raw Raman spectra from the equine third metatarsal (Mt3) and carpal joint bone specimens

Figure B.1 is a plot of the raw Raman spectra from the equine third metatarsal (Mt3) bone specimens, showing the non-linear, fluorescent baseline.

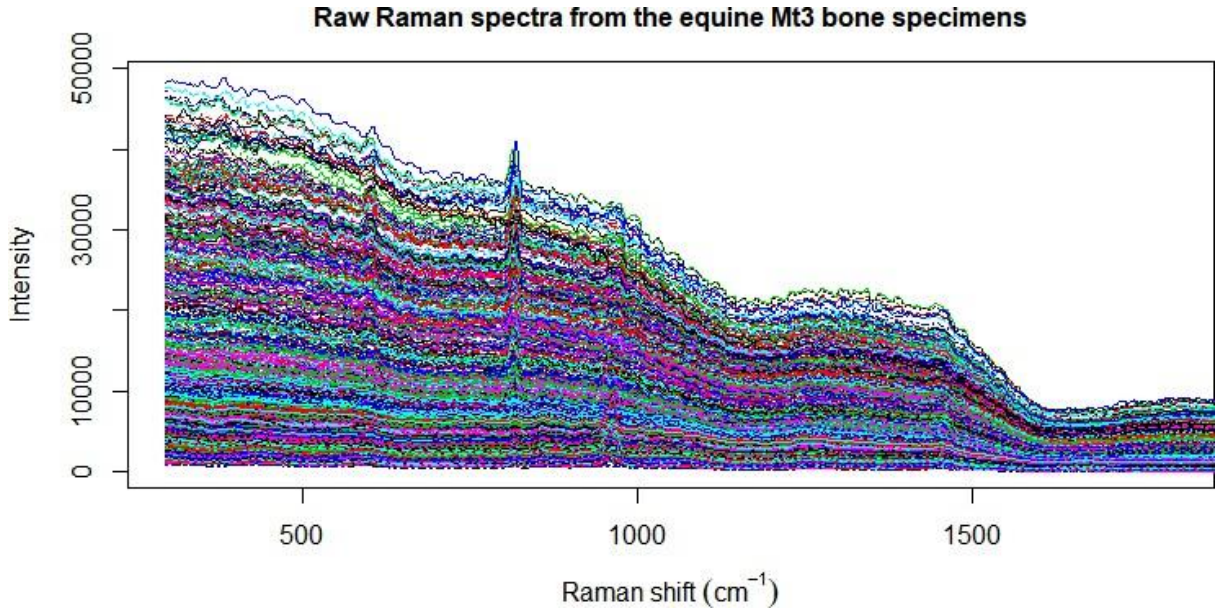

Figure B.1 Raw Raman spectra from the equine third metatarsal (Mt3) bone specimens.

Figure B.2 is a plot of the raw Raman spectra from the equine Mt3 bone specimens depicting one of the pre-processing steps. The truncated and transposed spectra were scaled by standard deviation and mean-centred (that is, the spectra were subjected to single normal variate (SNV) scaling). Again, the fluorescent background could still be seen.

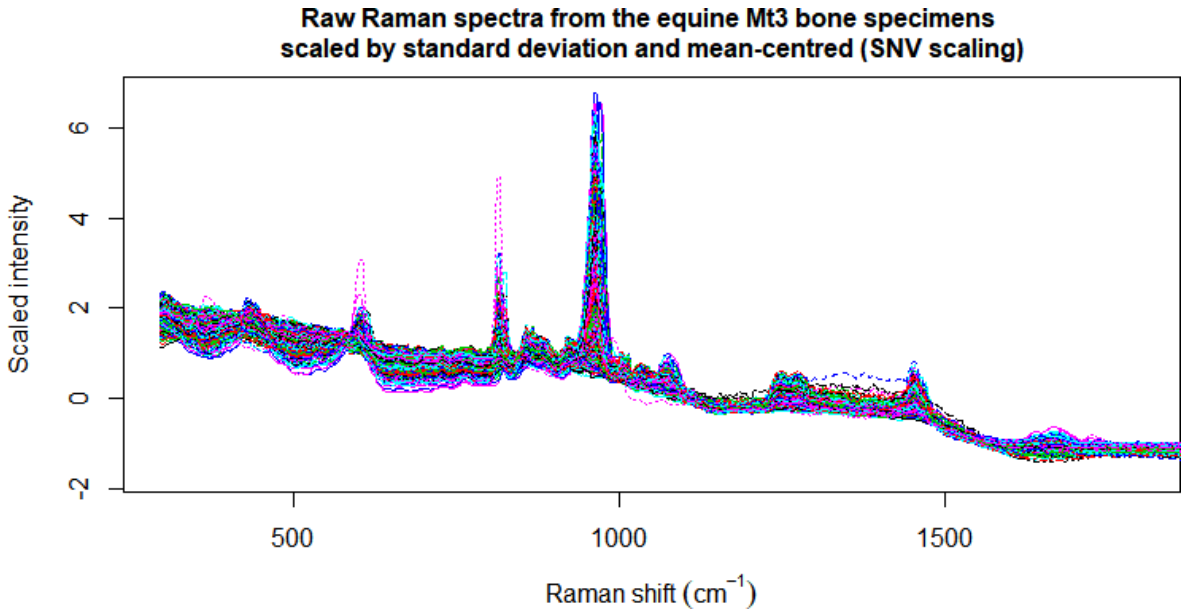

Figure B.2 Raw Raman spectra from the equine Mt3 bone specimens, scaled by standard deviation and mean-centred (i.e. SNV scaling).

Figures B.3 and B.4 are plots of the raw and SNV-scaled Raman spectra from the equine 'chip' carpal bone specimens where the fluorescent background was also still evident.

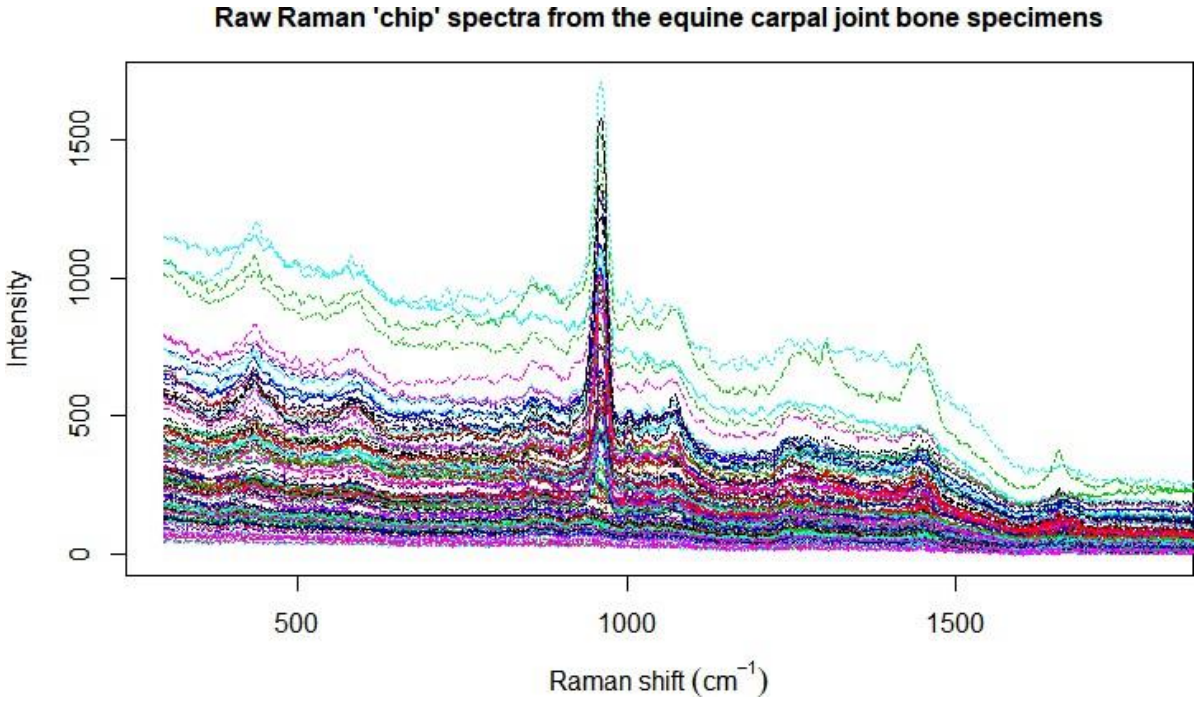

Figure B.3 Raw Raman 'chip' spectra from the equine carpal joint bone specimens.

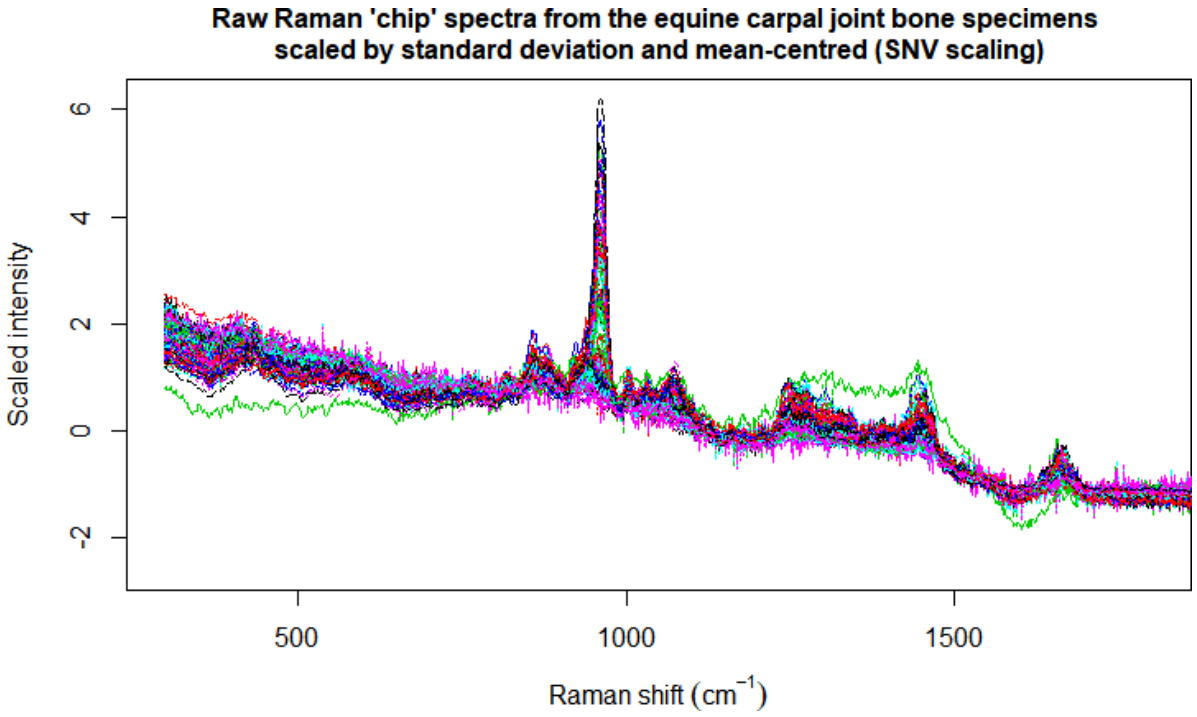

Figure B.4 SNV-scaled raw Raman 'chip' spectra from the equine carpal joint bone specimens.

Figures B.5 and B.6 are plots of the raw and SNV-scaled Raman spectra from the equine 'control' carpal joint bone specimens where the fluorescent background was again still evident.

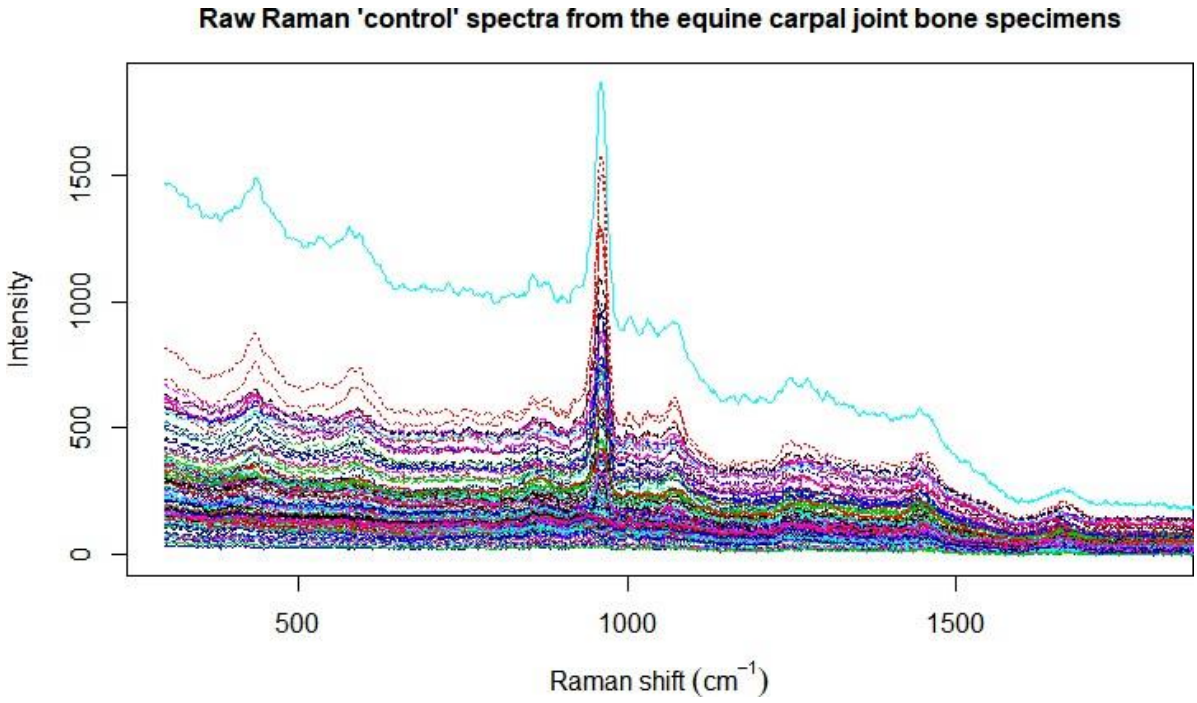

Figure B.5 Raw Raman 'control' spectra from the equine carpal joint bone specimens.

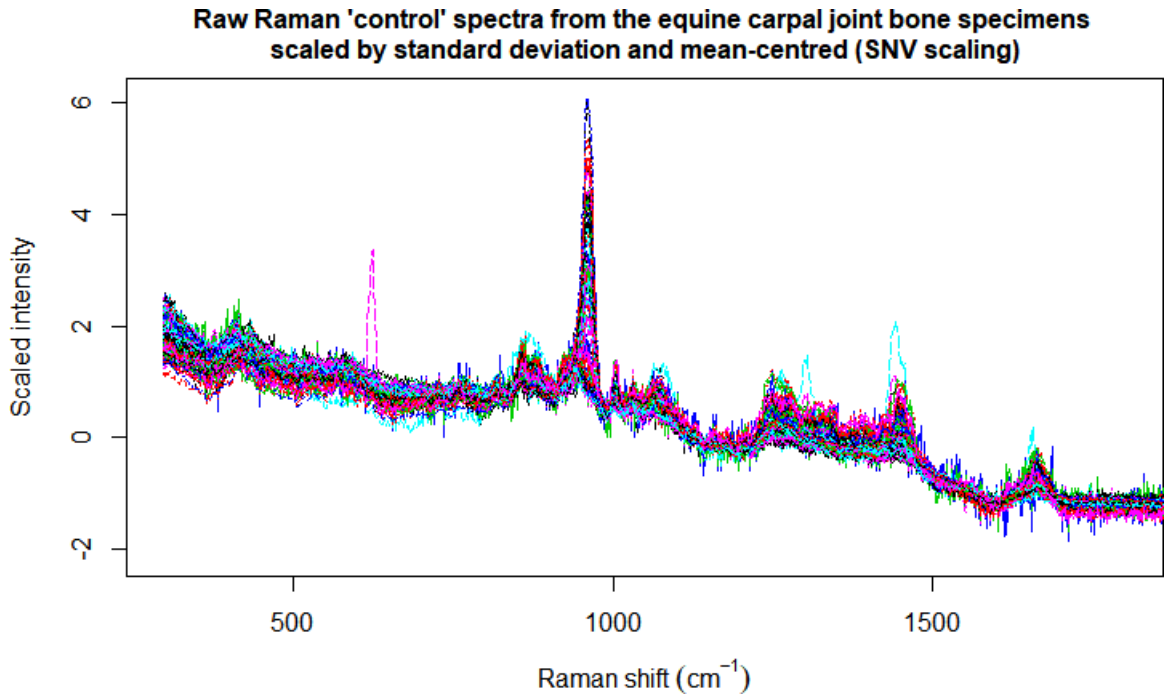

Figure B.6 SNV-scaled raw Raman ‘control’ spectra from the equine carpal joint bone specimens.