

A comparison of statistical techniques
for evaluating body condition in New
Zealand leopard seals (*Hydrurga
leptonyx*) using citizen science data

Jodie Warren

MSc (by research)

University of York
Environment and Geography
January 2021

Abstract

Unlike other southern continents where leopard seals (*Hydrurga leptonyx*) are considered vagrant visitors from Antarctica, sightings of leopard seals in New Zealand have been increasing, resulting in a re-classification from a vagrant to resident species in 2019. However, their ability to adapt to northerly environments still remains relatively unknown. Assessments of leopard seal body condition within this northern part of their range is one way in which scientists can understand their health status and thereby adapt strategies to protect populations. Preceding this study research investigating body condition of leopard seals often-employed invasive and costly techniques that carried risk to seals and researchers. This thesis presents four non-invasive procedures to examine body condition of leopard seals found in New Zealand waters, designed to assist with understanding their health status following such a substantial habitat shift. A body condition scoring system allocated leopard seal sighting records into body condition groups based upon presence/absence of bony protrusions and identified that sighting records of New Zealand leopard seals ($n=80$) were predominantly in Good condition (71.25%). Using these body condition groups, machine learning classifiers were successful in predicting sighting records of New Zealand leopard seals into Good, Moderate and Poor body condition using differences in body shape defined by photogrammetry. Whilst highest classification accuracy was obtained using photographic measurements of body width and Linear Discriminant Analysis (87.5%), an Artificial Neural Network based on leopard seal silhouettes (81.25%) was also identified as being suitable for examining body condition New Zealand leopard seals due to its ability to utilise large, complex datasets and flexibility with lower quality images. Methodologies developed here were enabled by a large photograph library collated by citizen scientists and volunteer researchers and can potentially be applied to assess the body condition of leopard seals in other regions as well other pinniped species.

List of Contents

Abstract	2
List of Figures	7
List of Tables	11
Acknowledgements	13
Declaration	15
Chapter 1 : Introduction to leopard seal ecology with a focus on their population in New Zealand and methods for assessing body condition	16
1.1 Leopard seal morphology and ecology	17
1.2 Leopard seal diet and feeding methods	19
1.3 Leopard seal distribution and occurrence within and outside of Antarctica	20
1.4 Leopard seal population estimates	22
1.5 Leopard seal anthropogenic impacts and conservation status	23
1.6 Leopard seal presence in New Zealand	25
1.7 Review of methods to assess leopard seal body condition	29
1.7.1 Unverified methods	29
1.7.2 Visual scales	29
1.7.3 Measurements using imagery.....	30
1.7.4 Morphometrics to estimate mass.....	31
1.7.5 Unmanned aerial systems and photogrammetric models to estimate mass	35
1.7.6 Evaluation of body condition assessment methods	36
1.8 Thesis rationale	37
Chapter 2 : Body condition scoring as a non-invasive method for examining condition in sighting records of free-ranging New Zealand leopard seals from photographs	39
2.1 Introduction	40
2.1.1 Photo-identification	40
2.1.2 Photograph assessment.....	40
2.1.3 Body condition scoring.....	42
2.1.4 BCS origin and applications.....	43
2.1.5 BCS applications to marine mammals.....	44
2.1.6 BCS applications to leopard seals	46
2.2 Aims	46
2.3 Methods	47
2.3.1 Data collection	47
2.3.2 Photograph assessment.....	47
2.3.3 Body condition scoring.....	49

2.4	Results	50
2.5	Discussion	52
2.5.1	Photograph assessment.....	52
2.5.2	Body condition scoring.....	53
2.6	Conclusions	55
Chapter 3 : Dimensionality reduction of photogrammetric measurements of body width extracted from sighting records of New Zealand leopard seals using Principal Component Analysis.....		
3.1	Introduction	58
3.1.1	Photogrammetry.....	58
3.1.2	Principal Component Analysis.....	60
3.1.3	Applications of Principal Component Analysis.....	60
3.1.4	Applications of Principal Component Analysis to marine mammal studies	61
3.1.5	Applications of Principal Component Analysis to leopard seal studies	61
3.2	Aims	63
3.3	Methods	63
3.3.1	Photogrammetry.....	63
3.3.2	PCA	66
3.4	Results	67
3.4.1	Photogrammetry.....	67
3.4.2	PCA models	69
3.5	Discussion	74
3.5.1	Photogrammetry.....	74
3.5.2	PCA	75
3.6	Conclusions	76
Chapter 4 : The application of Linear Discriminant Analysis and Random Forest Classifiers for assigning body condition groups to sighting records of New Zealand leopard seals		
4.1	Introduction	78
4.1.1	Linear Discriminant Analysis	78
4.1.2	Random Forest Classifiers	79
4.1.3	Comparing applications of LDA and RFC.....	80
4.2	Chapter Aims	83
4.3	Methods	83
4.3.1	Linear Discriminant Analysis	83
4.3.2	Random Forest Classifiers	84
4.3.3	Repeat Sighting Records	85
4.4	Results	85

4.4.1	Linear Discriminant Analysis	85
4.4.2	Random Forest Classifiers	87
4.4.3	Comparison of Linear Discriminant Analysis and Random Forest Classifiers	90
4.5	Discussion	91
4.5.1	Linear Discriminant Analysis	91
4.5.2	Random Forest Classifiers	92
4.5.3	Comparison of Linear Discriminant Analysis and Random Forest Classifiers	92
4.6	Conclusions	94
Chapter 5 : The application of Artificial Neural Networks for assigning body condition groups to sighting records of New Zealand leopard seals		95
5.1	Introduction	96
5.1.1	Artificial Neural Networks	96
5.1.2	Applications of Artificial Neural Networks	98
5.1.3	Application of Artificial Neural Networks to marine mammal studies	99
5.2	Chapter aims	101
5.3	Methods	101
5.3.1	Extraction pixel data from leopard seal silhouettes	101
5.3.2	ANN-1 and ANN-2	102
5.3.3	Added-Data	104
5.4	Results	104
5.4.1	ANN-1 and ANN-2	104
5.4.2	Added-Data	107
5.5	Discussion	109
5.5.1	ANN-1 and ANN-2	109
5.5.2	Added-Data	110
5.6	Conclusions	111
Chapter 6 : Discussion of statistical and machine learning techniques used to assess body condition in New Zealand leopard seals		113
6.1	Overview	114
6.2	Chapter 2: Body condition scoring as a non-invasive method for examining body condition in sighting records of free-ranging New Zealand leopard seals from photographs	115
6.2.1	Key research findings	115
6.2.2	Contribution and significance	115
6.2.3	Limitations and improvements	116
6.3	Chapter 3: Dimensionality reduction of photographic measurements of body width extracted from sighting records of New Zealand leopard seals using Principal Component Analysis	118

6.3.1	Key research findings	118
6.3.2	Contribution and significance	118
6.3.3	Limitations and improvements	119
6.4	Chapter 4: The application of Linear Discriminant Analysis and Random Forest Classifiers for assigning body condition groups to sighting records of New Zealand leopard seals	119
6.4.1	Key research findings	119
6.4.2	Contribution and significance	120
6.4.3	Limitations and improvements	120
6.5	The application of Artificial Neural Networks for assigning body condition groups to sighting records of New Zealand leopard seals	121
6.5.1	Key research findings	121
6.5.2	Contribution and significance	121
6.5.3	Limitations and improvements	121
6.6	Comparison of machine learning methods	123
6.7	Suggestions for further research	126
6.8	Management considerations	127
6.9	Concluding statement	127
	Appendix I: Details on how to plot a data observation onto a biplot using linear discriminant scores	129
	Appendix II: R script for ANN-2.....	130
	Bibliography.....	133

List of Figures

Chapter 1

Figure 1.1. Photograph extracted from Acevedo et al., (2017) of female leopard seal and pup seen in Parry Fjord (Chile) in 2012, showing the pigmentation patterns. 18

Figure 1.2. Heat map extracted from Hupman et al., (2019) that shows leopard seal sightings around the New Zealand coastline (denoted by grey circles) and areas of high density represented by the colour gradient. From top to bottom of the map, these locations from top to bottom are; Auckland, Christchurch and Dunedin..... 28

Figure 1.3. The body condition classification system extracted from Gray et al., 2009, whereby: (a) excellent; (b) good; (c) fair/thin, and; (d) poor..... 30

Figure 1.4. The body condition scoring system extracted from Hupman et al., (2019), their Table 1. 31

Figure 1.5. An image showing locations of bony protrusions, extracted from Hupman et al., (2019), their Figure 3. 31

Figure 1.6. The theoretical relationship between girth (G), length (L) and volume in phocid seals extracted from Hofman (1975). 32

Figure 1.7. Photographs extracted from van den Hoff et al., (2005) that illustrates two individual leopard seals belonging to two of the visual body condition groups (BCGs), whereby a) shows a leopard seal judged to be in medium body condition (BC) based upon “a thick neck and well-rounded body” in addition to a snout-tail length of 180cm and a girth of 230cm. Photograph b) shows a leopard seal judged to be in poor BC, based upon a “wasted body appearance, a thin neck and protruding pelvic region” and snout-tail length to height ratio of 5.5:1 (van den Hoff et al., 2005). 34

Figure 1.8. A leopard seal photographed by an unmanned aerial system (UAS) complete with the 15 photographic measurements; PW1-10 (width 1-10), PAW (axillary width), PUW (umbilical width), PSL (standard length) and POL (overall length) extracted from Krause et al. (2017), their Figure 3. 35

Chapter 2

Figure 2.1. The six photograph views of leopard seals (*Hydrurga leptonyx*) used for photo-identification in a study by Forcada and Robinson (2006). From top to bottom (left to right) the photographs are as follows: head left (HL), throat (TH), head right (HR), body left (BL), ventrum (BV) and body right (BR). 42

Figure 2.2. Body condition scoring system developed for North Atlantic right whales (*Eubalaena glacialis*) by Pettis et al., (2011; their Figure 1). Photograph a identifies an individual with a distinct fat roll in the nuchal crest region (as shown by the arrow), while photograph b shows a whale with a flat nuchal crest, both of which were identified to be in good (score 1) condition. Photographs c and d show individual right whales examined to be moderate (score 2) condition, with slight concavity behind the nuchal crest (as shown by the arrows), and photographs e – g show individuals that were characterised as poor (score 3) condition.

For the three poor condition photographs the large arrows denote areas of concavity while small arrows identify distinct “humps” often seen in very underweight whales. 45

Figure 2.3. Number of individual New Zealand leopard seal sighting records categorised to each body condition group (BCG) based upon a body condition scoring system (BCSS). 51

Figure 2.4. Frequency the five distinct bony protrusions were visible using a body condition scoring system (BCSS), distinguished for each New Zealand leopard seal sighting record. 52

Chapter 3

Figure 3.1. Examples of photographs of an individual southern right whale (*Eubaleana australis*) taken from the unmanned ariel vehicle (UAV) extracted from Christiansen et al., (2019). Photograph (a) shows the dorsal side of the whale complete with the photographic measurement of total body length (BL) from the tip of the rostrum to the tip of the tail notch (as shown by white arrows) and the body width measurements at 5% increments (W). Photograph (b) shows the same whale photographed from the lateral side, used to extract photographic measurements of body height (*i.e.* the dorsal to ventral distance; H) at the same 5% increments, the solid lines represent girths (G) at 25% (pectoral fin), 50% (umbilicus) and 72% (anus) body length. Photograph (c) shows a 3D model of the same whale photographed in (a) and (b), as a series of ellipses used to estimate body volume, where each ellipse represents the variation in height-width ratio (Christiansen et al., 2019). 59

Figure 3.2. Leopard seal side profile displayed in ImageJ software exhibiting ten equidistant points along the overall length of the body (*i.e.* 1-10 shown on the ventral side of the seal), between the start point at the tip of the snout (in blue) and the end point at the tip of the hindflippers (in red), using the ten widths macro developed by Krause et al., (2017). Locations of the three manual measurements (A, AW and UW shown on the dorsal side of the seal; see methods for definitions). Here, all 13 photographic measurements are shown as examples of how measurements were taken from the ten widths line. Photo by Giverny Forbes, LSO (LeopardSeals.org unpublished data). 65

Figure 3.3. Body shape of 80 New Zealand leopard seals using average (and standardised, indicated by ‘S’) widths, divided into body condition groups (BCG); Good ($n=57$), Moderate ($n=15$) and Poor ($n=8$), based upon body condition scores (BCS), at 12 positions along the body (S1 – 10, SAW and SUW). 68

Figure 3.4. Boxplot displaying variation in the measurement of the neck angle (A) in degrees, between the three body condition groups (BCG) for 80 New Zealand leopard seals, where Good=1 ($n=57$), Moderate=2 ($n=15$) and Poor=3 ($n=8$). 69

Figure 3.5. Corroplot from the first Principal Component Analysis model (PCA-1) where ‘Dim’ along the top represents the principal components (PC) and the photographic measurements of body width are displayed on the left. The larger and darker a circle represented the higher the variable contributed to each of the PCs (*i.e.* Dim.1... Dim.7), as shown by the colour scale on the right. 70

Figure 3.6. Loadings plot from the first Principal Component Analysis model (PCA-1) where ‘Dim1’ and ‘Dim2’ represent principal components (PC) PC1 and PC2. The colour (denoted by ‘contrib’ on the right) and the length of the vector indicate how much each of the variables contributed to the PCs. 71

Figure 3.7. Biplot of the first Principal Component Analysis (PCA) model (PCA-1) using all normally distributed photographic measurements: S1; S2; S3; SAW; S4; S6, and; S8 with body condition groups Good, Moderate and Poor superimposed. Each point on the plot represents the measurement profile of an individual leopard seal sighting record ($n=80$), where values of all seven photographic measurements have been projected as one data point, onto the first two principal components (PC1 and PC2, where var = variation) using coordinates. The body condition groups are denoted by three different colours, with ellipses drawn to encompass the majority of the grouping. 72

Figure 3.8. Biplot of the fifth Principal Component Analysis (PCA) model (PCA-5) using photographic measurements: S2; S3, and; S6 with body condition groups Good, Moderate and Poor superimposed. Each point on the plot represents the measurement profile of an individual leopard seal sighting record ($n=80$), where values of all three photographic measurements have been projected as one data point, onto the first two principal components (PC1 and PC2, where var = variation) using coordinates. The body condition groups are denoted by three different colours, with ellipses drawn to encompass the majority of the grouping. 73

Chapter 4

Figure 4.1. Biplot of the most accurate Linear Discriminant Analysis (LDA) model LDA-4 which used photographic measurements S2, S3 and S6 to distinguish between pre-assigned body condition groups (BCG) of sighting records containing Good (red), Moderate (green) and Poor (blue) condition leopard seals within the training dataset ($n=64$). Each point on the plot represents the measurement profile of an individual leopard seal sighting plotted by the discriminant scores of the linear discriminants LD1 and LD2, that were obtained by the equation $V_{1i} = b_0 + b_1DV_{1i} + b_2DV_{2i}$ 87

Figure 4.2. Classification tree for the most accurate Random Forest Classifier (RFC) model RFC-1, used predict New Zealand leopard seal sighting records into one of three body condition groups (BCGs); Good, Moderate or Poor using photographic measurements. For this RFC model all 13 photographic measurements were used for classification of BCGs (see methods for further details). 89

Chapter 5

Figure 5.1. A template of a basic, binary, feed-forward Artificial Neural Network (ANN; adapted from Krogh 2008) with an input layer containing three input units, one hidden layer containing four input units and one output layer containing one input unit. Each arrow represents weight and biases, which pass an input forwards to the next layer with the associated weight. 97

Figure 5.2. A template of a basic, multi-class, feed-forward Artificial Neural Network (ANN; adapted from Krogh 2008) with an input layer containing three input units, one hidden layer containing four input units and one output layer containing three input units. Each arrow represents weight and biases, which pass an input forwards to the next layer with the associated weight. 98

Figure 5.3. Leopard seal silhouette being extracted using ImageJ software where: A is the original image with the traced polygon; B is the polygon printed onto the new blank image; C is the scaled polygon to fit the image, and; D is the polygon inverted into a silhouette. 103

Figure 5.4. Proportion New Zealand leopard seals with a Good ($n=57$), Moderate ($n=15$) and Poor ($n=8$) body condition within the Artificial Neural Network (ANN) dataset ($n=80$). 104

Figure 5.5. Optimisation learning curves (loss) and performance learning curves (accuracy, in %) for training and validation data within Artificial Neural Network (ANN) model ANN-1 which classified body condition groups (BCGs) of sighting records containing photographs of New Zealand leopard seals. The plots visualise training (red points) and validation (blue points) data, where after 100 epochs loss=3.75 and accuracy=81.25% for training data. 106

Figure 5.6. Optimisation learning curves (loss) and performance learning curves (accuracy, in %) for training and validation data within Artificial Neural Network (ANN) model ANN-2 which classified body condition groups (BCGs) of sighting records containing photographs of New Zealand leopard seals. The plots visualise training (red points) and validation (blue points) data, where after 100 epochs loss=3.21 and accuracy=90.625% for training data. 107

Figure 5.7. Comparison between body condition group (BCG) estimates denoted 'ESTIMATE' (dark grey) and BCG predictions made by the two Artificial Neural Network (ANN) models ANN-1 (light grey) and ANN-2 (mid-grey) for sighting records containing photographs of New Zealand leopard seals in the database 'Added-Data', where the x-axis signifies the 48 sighting records within this dataset. For the purpose of this plot BCGs were modified whereby Good=0, Moderate=1 and Poor=2. 108

Figure 5.8. Sighting-ID HL127 from Added-Data..... 111

Figure 5.9. Sighting-ID HL125 from Added-Data..... 111

Chapter 6

Figure 6.1. A leopard seal silhouette used in Artificial Neural Network (ANN) models, of a Good body condition (BC) individual (sighting-ID HL036). 122

Figure 6.2. A leopard seal silhouette used in Artificial Neural Network (ANN) models, showing a Moderate body condition (BC) individual (sighting-ID HL029). 122

List of Tables

Chapter 2

- Table 2.1. Photograph Assessment Criteria (PAC) used to examine images of New Zealand leopard seals. Examples of photographs which passed and failed, whereby reasons for failing were as follows: no full body side profile visible due to obstruction by water (photograph B); no full body side profile visible due to cropping (photograph D), and; no full body side profile visible due to body position of seal being not parallel to the ground (photograph F) and not perpendicular to the camera (photograph H). Photographs A, B, D and G were taken by Thelma Wilson, Gary Mathers, Kerry Butterfield and Katharina Achterberg (LeopardSeals.org unpublished data), while photographs C, E, F and H were taken by the author. 48
- Table 2.2. Photograph Quality Criteria (PQC) used to examine images of New Zealand leopard seals. A score of 1 was assigned to photographs with sufficient quality for each of the three criteria (resolution, lighting and contrast) while a score of 0 was designated for photographs which were of insufficient quality with regard to the criteria. Photographs were taken by Giverny Forbes (A, C and E), Joanne Thwaites (B), Thelma Wilson (D) and Neil Morrison (F; LeopardSeals.org unpublished data). 49
- Table 2.3. Body condition group (BCG) chart and classification for New Zealand leopard seals based upon visibility of bony protrusions, adapted from Hupman et al., (2019). Examples of body condition classifications are outlined to illustrate how body condition scores were calculated. 50
- Table 2.4. Body condition groups (BCG) of New Zealand leopard seals classified as Good (example image A), Moderate (example image B) and Poor (example image C) using a body condition scoring system (BCSS) following methods modified by Hupman et al., (2019). Photographs A and C were taken by Giverny Forbes and Francis Murphy (LeopardSeals.org unpublished data) while photograph B was taken by the author. 50

Chapter 3

- Table 3.1. Photographic measurements used for each Principal Component Analysis (PCA). Cells where information was not applicable were denoted NA. 67

Chapter 4

- Table 4.1. Advantages and disadvantages of Linear Discriminant Analysis (LDA) and Random Forest Classifiers (RFC). 82
- Table 4.2. Photographic measurements of body width of New Zealand leopard seals used in four Linear Discriminant Analysis (LDA) models designed to predict body condition groups (BCG). 84
- Table 4.3. Photographic measurements of body width of New Zealand leopard seals used in four Random Forest Classifier (RFC) models designed to predict body condition groups (BCG). 85
- Table 4.4. Coefficients (*b*-values) of linear discriminants (LD1 and LD2) for each of the three photographic measurements used for the most accurate Linear Discriminant Analysis (LDA) model; LDA-4, where the coefficients give information about the relationship of the predictor variable (photographic measurement)

to each linear discriminant. Negative values portray negative relationships while positive values identify positive relationships. 87

Table 4.5. Correct body condition groups (BCG) and BCGs as predicted by the four Linear Discriminant Analysis (LDA; LDA-1, LDA-2, LDA-3, LDA-4) and three Random Forest Classifier (RFC; RFC-1, RFC-2, RFC-3) models (columns). The bottom row provides a score of correct classification rates (CCRs) and the * indicates the most accurate model for each method (*i.e.* the models that scored the most correct BC classifications). Correct predictions are highlighted in blue whilst incorrect predictions are visible in orange. Cells where information was not applicable were denoted NA. 90

Chapter 5

Table 5.1. Comparison of machine learning techniques. 100

Table 5.2. Values for accuracy and loss for Artificial Neural Network (ANN) Models ANN-1 and ANN-2 which classified body condition groups (BCG) of sighting records containing images of New Zealand leopard seals, within both training data (trainx) and testing data (testx). 105

Table 5.3. Four confusion matrices showing predicted vs correct body condition groups (BCG) for both Artificial Neural Network models (ANN-1 and ANN-2) which classified BCGs of sighting records containing images of New Zealand leopard seals, within the data subsets trainx and testx. 105

Table 5.4. Proportion (%) of New Zealand leopard seal sighting records containing Good, Moderate and Poor condition individuals from both; Added-Data (n=48) and the Body Condition – Leopard Seals Database (BCLSD; n=80), in addition to both databases combined (Average). 109

Chapter 6

Table 6.1. Comparison of actual (Correct) against predicted body condition groups (BCG) of sighting records containing images of New Zealand leopard seals (Sighting-ID) across the three most accurate models from each machine learning technique; Linear Discriminant Analysis (LDA-2.5), Random Forest Classifier (RFC-1) and Artificial Neural Network (ANN-2). Correct predictions are highlighted in blue whilst incorrect predictions are visible in orange. Cells where information was not applicable were denoted NA. 124

Table 6.2. Advantages and disadvantages of the machine learning methods trialled in this manuscript for use at predicting body condition groups (BCG) for New Zealand leopard seals. 125

Acknowledgements

So I thought writing acknowledgements was going to be the easiest part of the whole manuscript but as it turns out it is actually quite difficult to convey into words how grateful I am for all the help and support that I have had over the past year (or technically two in some cases) for doing this MSc... but, I have given it a go, so, in no particular order:

The first round of thanks in these acknowledgements go to my all-things-leopard-seal supervisor Dr Krista Hupman and my university, stats and all-round MSc coach Dr Dean Waters, for taking on this project and guiding me through the trials and tribulations of a research degree. Thank you to Krista for the experience and wonderful memories I have of being a NIWA intern for 2019 (although I do *not* miss the pungent aroma of scat analysis, making headlines all over the world for finding the USB was both exciting and amusing!), as well as the insightful pep talks and overall introduction into the field of marine mammal ecology. You are a huge inspiration to girls and women in the field of science and I hope that our paths cross again one day. Thank you also to Dean, not only for all of your guidance and help this year regarding some really complicated stats and machine learning techniques that I originally had no idea about... but also for listening to all of my R woes and struggles (there were a few), and for our weekly catch-ups during lockdown, which were really pleasant and helpful during a time of such “*uncertainty*” and isolation. At least we all know how to use Zoom now.

Thank you round 2 are for my family. Thank you to my Granny Val and Grandad John, who aren't around to read this, but who literally didn't mind what I did with my life, so long as I was happy, and introduced me to my love of nature and animals. Thank you to my Grandad Barry and Carol, who have always delighted me with their stories of travel, which inspired me to move to NZ in the first place, where I had the most amazing year, made some great friends, and saw wild marine mammals so many times! Thank you to mamma and Rob, who are responsible for some of my fondest childhood memories (well, all memories!), who have taught me so much over the years and helped to shape me into the person that I am today, which I am very grateful for – the maths folders on Monday nights have come in handy here haven't they! Thank you to Jade and Charlie for being all-round great siblings, and particularly to Jade for doing a similar degree where we always scrutinised each other's work (most often with a bottle of baileys and pot of filter coffee!), which kept us both motivated and inspired to work with animals. Thank you to James for being your ever-great and helpful self, by helping me pick a new laptop in the summer so I didn't end up pulling all of my hair out trying to get R to run on my old machine... and finally, thank you to my parents, Sacha and Jason Warren, who have

always loved and supported me in anything I do. Whether it's going to the Lake District to live in a caravan, or moving around the world to New Zealand, you have always been on board and excited along with me. You guys have given me everything I need to be happy and successful in life, I could not have wanted for anything more (you even caved and got me a dog). All of you are wonderful and I am grateful to have such a close and interesting family.

Round 3. Thank you to the Graces, Ann, Cian and Bridie for putting up with me sat in your house with my face in my laptop for months, and for convincing me to take breaks away from the screen to enjoy some of those summer days where we actually had a small amount of freedom to do things during the questionable year that was 2020. And finally, to the most special Grace (to me) Liam. I am going to have to keep this simple, but thank you ever so much for being you. Thank you for moving to the other side of the world with me to throw ourselves into something new, even though we sometimes had to count our pennies, we genuinely had the best time. The first time I ever saw a (live) leopard seal was with you, and it will forever be one of my favourite memories. Thank you for keeping me motivated on the tough days and celebrating with me on the good days, I am very grateful to have you by my side.

Last one. He cannot read, or understand most English, but I want to mention my best friend Jasper. Thank you, little rascal, for being my companion for the past almost 11 years! And particularly for this year when I have spent so much time sat at my desk. Thank you for keeping me entertained and reminding (pestering) me to take breaks through the day to throw your toys for you and take you for a walk. And for coming to tell me the time at 3pm, 3:30pm, 4pm, 4:15pm... all the way until 5pm when it's your dinner time. You're the dog of my life!



Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1 : Introduction to leopard seal ecology with a focus on their population in New Zealand and methods for assessing body condition



Leopard seal hauled out on Wellington's south coast (New Zealand)

1.1 Leopard seal morphology and ecology

The leopard seal (*Hydrurga leptonyx*, de Blainville 1820) is a renowned pagophilic Antarctic predator. The body shape of this species is characteristically slender as compared to other Antarctic phocids, and differences in body shape between individuals can be examined in order to assess condition (Gray et al., 2009; Hupman et al., 2019). This phocid (earless or 'true' seal) has an easily distinguished, almost reptilian appearance, created by their attenuated figure with a long, thick neck and large head with wide gaping jaws (Jefferson et al., 2015). Leopard seal foreflippers are similar to those of otariids (eared seals), as they are both long and wide and set further back on the body, which further creates this long reptilian neck (Jefferson et al., 2015). Their pelage is strongly countershaded, grey dorsally and white ventrally with a broad stripe of darker fur that runs from the back of the head down the spine, with lighter patches above both eyes (Hamilton 1939). In both males and females of all age groups the entire body is covered with irregular, asymmetrical dark and light spots with a pattern that is unique to individuals. Leopard seal pups are born with this same colour pattern (Figure 1.1), although their lanugo (fine hair which covers the foetus of many mammalian species; Riedman 1990) is longer and thicker in a shade of brownish grey and creamy white. Uniquely, leopard seals are one of the few pinniped species born with countershaded lanugo (Jefferson et al., 2015). This pelage pigmentation does not change throughout their lifetime (with the exception of developing a yellow or brown tinge before moulting), thereby spot patterns can be used for individual recognition (Hamilton 1939).

Leopard seals are a solitary species that spend much of their time at sea, only coming ashore for short durations to rest and digest food (Gwynn 1953). Even the moulting process (that occurs in January) does not seem to interfere with normal activity or swimming, as the hairs fall out individually rather than in large sections of epidermis as seen in other pinnipeds (Gwynn 1953). The fatty acid composition of leopard seal blubber analysed by Guerrero et al., (2016) revealed that monounsaturated fatty acids (important for maintaining fluidity under the skin and reducing heat loss; Best et al., 2003) were in high quantities in leopard seal blubber (and particularly abundant the outer layer) which may support thermoregulation and enable this species to spend so much time in the water. They are extremely strong and agile swimmers, propelling themselves through the water both by undulating their hindflippers (typical for phocids) and by sweeping their long foreflippers (more typical of otariids; Jefferson et al., 2015). When hauled out on ice floes leopard seals are typically alone, and when ashore where other leopard seals are present they remain at a distance from one another, although incidences of leopard seal socialising interpreted as "*play*" on land has been documented by Gwynn (1953) on Heard Island.



Figure 1.1. Photograph extracted from Acevedo et al., (2017) of female leopard seal and pup seen in Parry Fjord (Chile) in 2012, showing the pigmentation patterns.

Leopard seals are thought to live for at least 26 years, with females reaching sexual maturity between two and seven years, and males at three to six years old (Jefferson et al., 2015). As adults, females can grow up to 3.8m long and weigh as much as 500kg, while males are smaller at maximum lengths of 3.3m and weights of 300kg (Rogers 2009). Individuals are generally classified into year-classes until they reach the age of maturity, based upon age-length curves researched by Laws (1957), who provided a series of lengths for seals belonging to each age group. Laws (1957) however, stated that not only do females grow larger than males, they also mature slightly later (*i.e.* in their fifth or sixth year). Following this, Forcada and Robinson (2006) published a different set of growth rates for males and females which assumed that the mean age for sexual maturity was four years old. These growth rates were adapted from the age-length classifications from Laws (1957) and categorised individuals into pup (up to 200cm), juvenile (between 200 and 285cm for females and 200 and 275cm for males) or adult (> 285cm for females and > 275cm for males), rather than specific year-classes (Forcada and Robinson 2006). Most recently, research by Krause et al., (2020) assigned individual leopard seals as adult life stage based upon measurements of standard length of >2.7m for females and >2.5m for males.

Leopard seal pups are predominantly born on the Antarctic ice prior to the breeding season, from late October through to mid-November, although the breeding period may extend to early January (Rogers 2009; Jefferson et al., 2015). At birth, pups are approximately 120cm long and they quickly grow in length during the first six months postpartum (Rogers 2009). Little is known about levels of parental care; pups are only accompanied by the mothers (Siniff and Stone 1985). It is suggested that leopard seals have a very short lactation period, varying between 10 – 14 days (suggested due to the advanced development of teeth in new-born pups; Brown 1952) and up to 4 weeks (Rogers 2009; Jefferson et al., 2015), after which point pups are left to fend for themselves. Leopard seals breed from November through to January following pupping. Mating is thought to occur in the water and has never been

formally documented in the wild (Rogers 2009). During the mating season both males and females are highly vocal. They emit low –medium frequency calls underwater which are extremely powerful and can be heard from above water and felt through ice (Rogers 2009). Females produce long-distance acoustic calls at the beginning of the mating season – presumably to attract mates – while males vocalise throughout the day for the entire season as part of an underwater breeding display (Rogers 2007; Rogers 2009).

1.2 Leopard seal diet and feeding methods

The leopard seal is an ambush predator with an extremely broad diet that encompasses a huge variety of species, which allows them to take advantage of seasonal abundances of prey throughout the year and in different regions (Hall-Aspland and Rogers 2004). They are born with large curved incisors and canines that facilitate the capture of bigger prey types like penguins, fish, cephalopods and seals, as well as modified molars similar to those of crabeater seals (*Lobodon carcinophagus*) which allow them to sieve krill from the water column (Rogers 2009; Southwell et al., 2012). The predation of leopard seals on Antarctic species is well documented in the literature across different regions of Antarctica. Antarctic fur seal pups (*Arctocephalus gazella*), demersal notothen fish, penguins and krill (*Euphausia superba*) were identified as the primary prey items for leopard seals inhabiting the South Shetland Islands (Krause et al., 2020). In comparison, in Antarctica leopard seals have also been found to feed on Adelie (*Pygoscelis adeliae*), gentoo (*Pygoscelis papua*) and macaroni (*Eudyptes chrysolophus*) penguins, Antarctic fur and crabeater seals, amphipods, benthic and pelagic fish, squid and krill (Hofman et al., 1977; Øritsland 1977; Stone and Meier 1981; Siniff and Stone 1985; Hall-Aspland and Rogers 2004).

Leopard seals are well known for their predation upon penguins, where they are known to hide amongst ice floes and ambush prey in the water as they enter or depart from colonies (Jefferson et al., 2015). As well as being captured, prey is also eaten in the water as leopard seals will forcibly slap penguins on the water surface to tear them open, the sound of which can be heard from at least 1km away (Jefferson et al., 2015). Although usually intolerant of conspecifics, leopard seals are known to hunt alongside one another in areas of abundant prey, where instances of kleptoparasitism and food caching (perhaps as a response to kleptoparasitism) have both been documented (Penny and Lowry 1967; Rogers and Bryden 1995; Krause et al., 2015). Cooperative feeding between leopard seals has rarely been observed, however a publication by Robbins et al., (2019) describes the first instance of co-feeding between leopard seals in South Georgia, on king penguins (*Aptenodytes patagonicus*). It was suggested that these instances of co-feeding were rare within the leopard seal population and facilitated in this location by a high proportion of predators at the site of large prey colony (Robbins

et al., 2019). On these occasions co-feeding may have occurred as it would have been more energetically costly to defend the prey item, than it was to tolerate kleptoparasitism on a large prey species whereby both seals would gain a sufficient portion of food (Robbins et al., 2019). Leopard seals have also been documented feeding from whale carcasses, and predation on crabeater seals is well evidenced by the high proportion of adult seals bearing the scars of leopard seal attacks (Jefferson et al., 2015).

The leopard seal diet varies between sex, with the change of the season. Research by Krause et al., (2020) in the South Shetland Islands found that over three study years (2013, 2014 and 2017) the austral spring diet of males and females both comprised of krill, fish and penguin, however during the transition to summer, the diet of the males remained unchanged whilst female leopard seals began to incorporate more energy-rich foods of penguins and Antarctic fur seal pups. This diet change was likely a response to the increase in energy and nutrient required for gestation and lactation. Previous research has suggested that the leopard seals ability to maximise on so many different prey types is related to a flexible reproductive strategy involving a less defined breeding period and shorter breeding cycle (Siniff and Stone 1985). This particular strategy would involve high levels of dispersal within the population to take advantage of food sources within different areas (Forcada and Robinson 2006).

1.3 Leopard seal distribution and occurrence within and outside of Antarctica

Leopard seals have broad home range which encompasses the Antarctic coast and circumpolar pack ice, as well as being distributed throughout the Antarctic and sub-Antarctic islands between 50°S and 80°S (Gwynn 1953; Bonner 1994; Bester et al., 2002; Bester et al., 2017). This species exhibits a movement pattern that follows the seasonal expansion and contraction of sea-ice. During the austral spring and summer when ice cover is at its lowest, leopard seals congregate further south towards the Antarctic continent, then when the sea-ice reforms in the subsequent autumn and winter, leopard seals disperse northwards following the ice-edge (Rounsevell and Eberhard 1980; Bester et al., 2002). Population density is inversely related to the size of the pack ice, where highest densities of leopard seals have been found in areas of abundant pack ice (2–20m in diameter ice floes) and brash ice (>2m in diameter ice floes; Rogers 2009). In recent years it has been observed that a part of the leopard seal population ventures north of the polar front to continents in the Southern Ocean. Thus far leopard seals have been recorded in; Argentina (Rodríguez et al., 2003), Australia (Rounsevell and Eberhard 1980; Rice 1998; King 1983), Brazil (De Moura et al., 2011), Chile (Aguayo-Lobo et al., 2011); the Cook Islands (the most northerly sighting to date at approximately 21°S; Berry 1960; Hupman et al., 2019),

New Zealand (Hupman et al., 2019), South Africa (Vinding et al., 2013) and Uruguay (Vaz Ferriera 1984).

Reports exist of a cyclic fluctuation in the number of leopard seals visiting northerly regions; every four to five years at Macquarie Island and Australia (Rounsevell and Eberhard 1980; Elliot 1982), and every three to four years at Bird Island, South Georgia (Walker et al., 1998). Large numbers of adult leopard seals have been recorded at Heard Island in August, followed by an increase in juveniles in the following months (Gwynn 1953). The reverse was documented at Macquarie Island whereby young and “*sickly*” leopard seals were most commonly observed from late winter to early summer (June to December), whilst adults were only present later in the year towards summer and in smaller numbers (Gwynn 1953). In years when leopard seals are abundant at Macquarie Island, a significant number are also sighted on the Australian coast (Rounsevell and Eberhard 1980). A study by Rounsevell and Eberhard (1980) between 1949 to 1979 using hindflipper tags revealed that the seals that visited Macquarie Island were highly mobile as most individuals left after hauling out once. Of the 278 seals tagged at Macquarie Island within this period two were re-sighted in other regions; one four-year-old female was tagged on 11th July 1977 and re-sighted at Campbell Island on 24th August 1977, whilst a four-year-old male tagged in the same year on the 4th August was found dead in Tasmania two months later on 26th October. Rounsevell and Eberhard (1980) further reported that each year seals were observed returning to the island that had been tagged in the previous year, and thereby suggested that this movement is a periodic dispersal of non-breeding individuals in relation to shortage of food resources.

In comparison to the Antarctic Islands, a constant increase of leopard seals has been recorded in Chile, where Aguayo-Lobo et al., (2011) reports that of the 115 individual leopard seals observed in the country between 1927 and 2010, 96 were sighted between 2000 and 2010. Furthermore, Aguayo-Lobo et al., (2011) discovered that leopard seals of both sexes and varied ages (adult and juvenile) were sighted in all seasons in Chile, South America. A very small (estimated at ~180cm in length) individual was also seen at Parry Fjord, leading the authors to suggest that leopard seals might be breeding in Chile, and therefore should be considered as a resident species (Aguayo-Lobo et al., 2011). More recently further evidence has been published supporting this, as two occurrences of births in Parry Fjord were documented (Acevedo et al., 2017) and indication of philopatry of one leopard seal for five months in the Magellan Strait was observed (Acevedo and Martinez 2012). The regions in which leopard seals were observed most commonly in Chile (the Feuguian Channels) are situated in close proximity to the Antarctic peninsula and share similar climatic characteristics with the Antarctic

habitat, including; similar temperatures of seawater and air, as well as presence of pack-ice (Aguayo-lobo et al., 2011), offering a favourable habitat for leopard seals. This is reinforced by research by Gwynn (1953) who found that presence of snow on stony beaches caused an increase in the number of leopard seals hauled out, hypothesising that the presence of the white snow “*attracted*” them, and by Jessopp et al., (2004) who discovered a negative correlation between sea surface temperatures and the number of leopard seals observed at Bird Island in Antarctica, whereby higher numbers of leopard seals were recorded on days when the sea surface temperature was lower.

A number of publications to date have suggested that the northward dispersal of leopard seals from their core home range is a result of highly mobile juveniles being pushed out of Antarctica during a time of food stress during the austral winter. At this time crabeater seal pups (a primary food source for leopard seals; Øritsland 1977; Siniff et al., 2008) are older, and while adults are more proficient in catching them juvenile seals lose their advantage (Siniff and Stone 1985; Rounsevell and Eberhard 1980). Antarctic krill are also an important source of food for juvenile seals (Siniff and Bengston 1977; Siniff et al., 2008); however, the winter distribution of krill in the Southern Ocean is infamously patchy and resides in deeper waters (Marr 1962). Leopard seals are shallow water hunters which lack the anatomical and physiological adaptations that are used to increase the transport, storage and conservation of oxygen during diving, meaning that they must compete with more adept species at foraging for krill such as Adelie penguins, crabeater seals and Antarctic minke whales (*Balaenoptera bonaerensis*; Siniff and Stone 1985; Lynch et al., 1999; Lynch and Bodley 2007; Vogelnest et al., 2010). It has also frequently been proposed that the northerly displacement of leopard seals occurs in juveniles as a result of aggressive behaviour displayed by adults when parturition and mating occur (Siniff and Stone, 1985; Øritsland 1970). Since juveniles do not require the pack-ice habitat for birthing and nursing, they are able to disperse northwards into alternative territories.

1.4 Leopard seal population estimates

Due to the extreme and remote conditions in which they live, population studies of leopard seals are enormously difficult and expensive to conduct (population density in the pack ice has been reported to be as low as 0.003 – 0.151 seals/km²; Rogers 2009), which has resulted highly variable population estimates over the years. The first estimate was published by Laws (1953), who gave a figure of 40 000 individuals. However, Laws (1953) cautioned on his own work as he stated that the figure was largely based upon guess work. Later, a circumpolar population estimate of 100 000–300 000 individuals was calculated by Scheffer (1958), however, the methods used to obtain these estimates were not described (Southwell et al., 2012). Erickson and Hanson (1990) published a similar value for circumpolar abundance of 300 000 leopard seals, calculated by aggregating results from a multitude

of surveys within different regions around the continent between 1968 and 1983. Although this research employed rigorous survey techniques which were an improvement on the earlier surveys, the data collection only occurred in summer, and across several years, during which time leopard seal abundances may have changed both on the continent and within regions (Erickson and Hanson 1990). The most recent and robust large-scale survey to date is named the Antarctic Pack-Ice Seal (APIS) project, directed by use of vessel based and aerial surveys along designated transects in Antarctica between the years of 1996 and 2001 (Southwell et al., 2012). Estimates from the APIS project designate a leopard seal abundance ranging largely between 3500–65 000 individuals (Southwell et al., 2012), much lower than the estimates made in previous years, although more in line with the original population estimates by Laws (1953). The authors of the APIS project do however report that because leopard seals are a solitary species that spend the majority of their time in the water, individuals were difficult to observe, and therefore the estimate was likely to be highly uncertain (Southwell et al., 2012). Presently, all abundance estimates have been based in Antarctic waters. There are no published population estimates for leopard seals in northern regions outside of their core home range.

1.5 Leopard seal anthropogenic impacts and conservation status

Antarctica has undoubtedly been affected by climate change as result of climate change due to increased carbon emissions. The Antarctic Peninsula is one of the most rapidly warming regions of the world, increasing by almost 3°C in the past 50 years (Vaughan et al., 2003; Turner et al., 2005). The climatic changes in the Antarctic continent are thought to have the potential to further alter sea-ice cover, increase the length of the summer season and cause a shift in the architecture of flora and fauna of the Antarctic ecosystem (Lamers et al., 2011). While the extent of the sea ice has been decreasing in the Antarctic Peninsula, the ice cover has been increasing in the Ross Sea (Siniff et al., 2008). The result of this is that summer sea ice in the Antarctic Peninsula has become effectively absent in recent years, whilst in the Ross Sea the sea ice persists into spring and takes a longer period of time to recede towards the continent (Siniff et al., 2008). These variations in sea-ice as result of climate change may potentially affect the life cycle and populations of Antarctic krill, whose reproductive success is be related to extent, duration and timing of the winter sea ice (Quentin and Ross 2001). Furthermore, it has been identified that the early development of krill is vulnerable when exposed to higher levels of CO₂, and growth is inhibited when temperatures are 0.5°C above optimum temperatures (Kawagushi et al., 2013; Atkinson et al., 2006). As a keystone species of Antarctica which forms the base of the food chain, Antarctic krill are of an important food resource for leopard seals, principally juveniles which feed upon them directly (Siniff and Stone 1985; Siniff et al., 2008; Rogers 2009; Krause et al., 2020). Furthermore, leopard seals predate upon mesopredators (such as Adelie

penguins) that also rely upon the krill to some degree (Lamers et al., 2011). Although the reduction of sea-ice may be detrimental to leopard seals which are reliant upon this environment for critical stages in their life-history (*i.e.* birth of pups), the impact is thought to be less than that of the other Antarctic pinnipeds due to their diverse diet and ability to inhabit varied environments in the Southern ocean (Siniff et al., 2008). However, the environmental changes catalysed by changes in sea ice cover and ice thickness will likely increase competition for food resources and resting areas and will thereby affect leopard seals to some extent (Siniff et al., 2008).

The Antarctic continent is often thought of as the last area of true wilderness on the planet, relatively untouched by human activity when compared to the other six continents. In recent times however, Antarctica has become a destination for travel and research, in addition to a huge resource for commercial seafood. The upsurge in human activity within the region has led to concerns of increased noise pollution from vessels and aircraft, and it has recently been demonstrated that leopard seals suffer from noise disturbance in the presence of vessels (Mallet 2020). In a study by Mallet (2020) the acoustic behaviour of leopard seals was studied during the arrival and departure of the German research vessel Polarstern in Atka Bay, to determine how underwater noise emitted from the increased amount of Antarctic tourism boat traffic would affect the local fauna. Mallet (2020) found leopard seal calls decreased abruptly with the arrival of the vessel and increased immediately following the departure of the vessel, as well as reporting that the call repertoire altered as lower frequency calls decreased markedly with the presence of the vessel, during which time high frequency calls were more often used (Mallet 2020). Since leopard seals are reliant upon sound in their underwater environment for communication (Mallet 2020), this intensification in underwater noise pollution from increased boat traffic has the potential to negatively impact the species, particularly during the breeding period when leopard seals - that are solitary and known to have an extensive home range - use specific calls to communicate to one another.

In some regions, it appears leopard seals have been at risk from shooting by fishermen, fish farm operators and even members of the public (Aguayo-Lobo et al., 2011; Rounsevell and Pemberton 1994; LeopardSeals.Org unpublished data). In Chile, Torres (1987) reported that leopard seals were hunted by fisherman as they were viewed as being dangerous at sea.. Additionally, Torres (1987) stated that the number of leopard seals observed in the southern fjordlands of Chile may have only been a portion of the total number of seals in the area due to illegal sealing operations by local fishermen (Aguayo-Lobo et al., 2011). In Tasmania salmonid fish farms are licenced to shoot Australian fur seals (*Arctocephalus pusillus*) which attack and steal penned fish (Pemberton and Shaughnessy

1993), and the presence of leopard seals in this region overlaps with the peak time that Australian fur seals are known to 'attack' these pens in late winter (Rounsevell and Pemberton 1994). Since these attacks usually happen during the night, it is possible that leopard seals in the area may be mis-identified as fur seals and shot by the farm operators (Rounsevell and Pemberton 1994). In Western Australia, a leopard seal that died from being shot, was in an area that was likely to have active fishing operations (Mawson and Coughran 1999). By comparison, in New Zealand there are multiple reports of leopard seals having been shot in the head by members of the public (LeopardSeals.org, unpublished data).

In Chile, there are three known instances of leopard seals being rehabilitated and re-released into the wild, one of which was reported to be in poor condition and had multiple injuries, that included a deep cut on the tongue which was presumed to have been caused by a hook, from which the seal survived (Aguayo-Lobo et al., 2011). In contrast, in Western Australia, 15 out of 27 recorded leopard seals between 1986 and 1996 were discovered already deceased or died whilst receiving treatment (Mawson and Coughran 1999). Of these 15 seals, the causes of death were noted as: shooting ($n=1$); fishing line/hook related ($n=1$); immune system failure ($n=1$); euthanasia ($n=4$) and; anaesthesia (during relocation to a rehabilitation centre or treatment; "several" out of $n=8$); and only one was successfully released (Mawson and Coughran 1999).

Leopard seals are classified as 'Least Concern' on the IUCN Red List following assessment in 2015 and under the Antarctic Treaty's Convention for the Conservation of Antarctic Seals up to 12 000 animals can be legally harvested annually (Hückstädt 2015), although it is unknown if this takes place. In the past, leopard seals were frequently killed for scientific research purposes and in some cases they were harvested for dog food (Reijnders et al., 1993), however this is not currently practiced today.

1.6 Leopard seal presence in New Zealand

In New Zealand, LeopardSeals.Org (hereafter referred to as LSO) was founded by Dr Krista Hupman and Dr Ingrid Visser in 2016, with the intent to conduct research on leopard seal biology and ecology and contribute to management and conservation of this species within this region. Leopard seals were only recently (May 2019) classified as 'Resident' species under the Department of Conservation New Zealand Threat Classification System (Baker et al. 2019), based on evidence of long-term presence in New Zealand, annual sightings throughout all seasons, and records of three births that support evidence of breeding (Hupman et al., 2019). An extensive database (the New Zealand Leopard Seal Database; NZLSD) was collated by LeopardSeals.org (Hupman et al., 2019) from the LSO sighting network (a collation of sighting records obtained via an online sighting form, toll free phone number

0800 LEOPARD and social media) as well as: published resources (e.g. books and scientific journals); unpublished information (museum records and data from zoos and aquariums); grey literature (e.g. online newspaper articles, wildlife, photography and fishing forums), and; the Department of Conservation marine mammal database (more information can be found in Hupman et al., 2019). The NZLSD included 2711 unique records of leopard seal sightings and/or specimens from around the New Zealand coastline up until the end of 2018 (Hupman et al., 2019). Their presence in this region has been dated back to 1200 from remains in Maori middens in Abel Tasman National Park (Smith 1985). Contrary to other regions and islands north of the polar front, data from LSO suggest that the number of leopard seal sightings in New Zealand has been increasing annually, most significantly from years 2000 to 2019 (Hupman et al., 2019).

At present, it is unclear if this increase in leopard seal sightings in New Zealand represents an increase in the number of leopard seals in New Zealand waters or an increase in the number of sightings that are reported (Hupman et al., 2019). The pelage and morphology of the leopard seal is visually unlike that of the two other resident pinnipeds in New Zealand; the New Zealand fur seal (*Arctocephalus forsteri*) and the New Zealand sea lion (*Phocarctos hookeri*) which are predominantly shades of brown in colouration. This dissimilarity might affect reporting regularity as sightings of 'unusual' species are more likely to be reported (Hupman et al., 2019). Moreover, it could be expected that the increase in sightings could be a reflection of the increase in human population in New Zealand, and/or the ability to capture and share data by means of digital equipment such as cameras and smartphones. Logically, some or all of these points may not be mutually exclusive. Despite this, the number of leopard seals visiting and residing in New Zealand is likely to be imprecise due to the fact that many seals may have been un-noticed and un-reported due to: its small national human population; remote locations where seals may haul-out; the time of day seal haul out (i.e. the sightings may be missed at night); lack of information provided on who to report sightings to, and; possible misidentification of the species. Otago's largest city, Dunedin, was identified as one of three sightings 'hotspots' of leopard seals on the New Zealand coast, along with Christchurch and Auckland (Figure 1.2; Hupman et al., 2019). These sighting hotspots represent areas with the most frequent leopard seal sightings and correspond with the most densely populated coastal cities in the country (Hupman et al., 2019). For these reasons, it is difficult to know if leopard seals have always been present in New Zealand in large numbers and seals just haven't been reported, or if the frequency of leopard seal sightings are increasing.

A recent publication by Hupman et al., (2019) analysed age-classes of leopard seals seen in New Zealand waters. These age-classes were divided into three groups: pup; non-adult, and; adult and 930 records of leopard seal sightings were assessed. Of these sightings 319 (34.3%) were determined to be non-adults. Moreover, an evaluation of body condition was conducted on the same records, and it was found that the majority of the juveniles found within New Zealand were classified as having 'Good' or 'Excellent' body condition (Hupman et al., 2019). Given the extensive distance between mainland New Zealand and Antarctica, it would be expected that these juvenile individuals would arrive in New Zealand in poorer health condition. It is therefore reasonable to suggest that some leopard seal births might take place outside of the Antarctic, either within New Zealand or neighbouring islands, and that due to remoteness of these regions and deficiency of research into the species, these occurrences have not yet been recorded (Hupman et al., 2019).

In New Zealand leopard seals are protected under the New Zealand Marine Mammals Protection Act 1978 (MMPA) where it is an offence to disturb, harass, harm, injure or kill any seal. The Department of Conservation is the government authority tasked with managing populations of marine mammals within New Zealand, which includes making decisions regarding euthanasia of individuals that are judged to be in poor body condition or show presence of injury. A number of seals that arrive in New Zealand waters have been euthanised in the past due to such judgements, however, without thorough assessment of leopard seal health and condition in this region, it is difficult to accurately assess what 'poor condition' looks like. A scale for assessing body condition for this species was developed in 2019 based upon photo-identification, the accuracy of which has not been quantitatively examined (Hupman et al., 2019). Therefore, decisions on euthanasia of this species are being made without scientific basis for which they are being assessed.

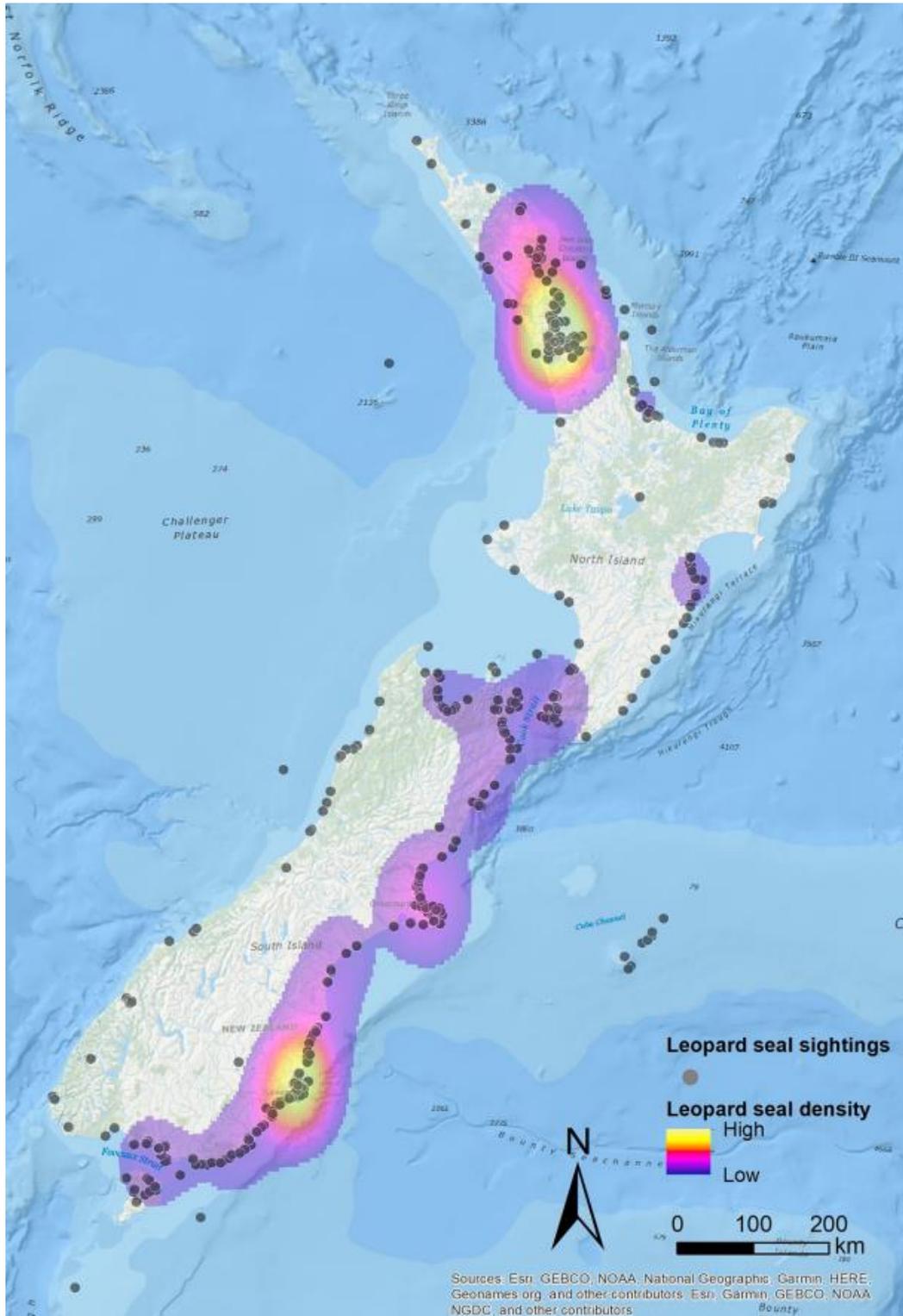


Figure 1.2. Heat map extracted from Hupman et al., (2019) that shows leopard seal sightings around the New Zealand coastline (denoted by grey circles) and areas of high density represented by the colour gradient. From top to bottom of the map, these locations from top to bottom are; Auckland, Christchurch and Dunedin.

1.7 Review of methods to assess leopard seal body condition

Due to the remoteness and harshness of the Antarctic climate and solitary nature of wide-ranging leopard seals assessment of health and body condition (hereafter BC) in this species presents a substantial task. In spite of the challenges there are several studies which have investigated and reported on the BC of leopard seals using a range of techniques, as discussed below.

1.7.1 Unverified methods

To date, there are several studies that classify individual leopard seals into BC categories (*i.e.* poor, fair, good or excellent) without providing an explanation on how this classification was made. For example, Aguayo-Lobo et al., (2011) examined BC of leopard seals in Chile and stated that at least five leopard seals were reported as being in poor condition and some were observed with injuries, however no BC classification methodology was described. Similarly, Rounsevell and Pemberton (1994) described the BC of leopard seals in Tasmania, Australia, stating that 18 leopard seals were judged as being in good condition, while 7 of the remaining 16 seals were assessed as being thin, however, no formal BC classification techniques were outlined. In another example Bester et al., (2017) commented on a leopard seal at Tristan da Cunha Island in the South Atlantic, stating that the leopard seal appeared to be in poor condition and had minor injuries, yet specific methodologies for this assessment were again, not described. The shortcomings with such methods are that classifications were reliant upon observations of experts as well as being entirely subjective based upon individual assessment. Moreover, these comments were not supported by quantitative measures to substantiate these classifications.

1.7.2 Visual scales

A visual BC index was first published for leopard seals by Gray et al., (2009). Within this research qualitative material was described in the form of a figure where four black and white photographs were provided to illustrate the four BC categories (see their Figure 1.3, their Chapter 9; Gray et al., 2009). From this visual scale individuals could be assessed, *i.e.* Gray et al., (2009) stated that leopard seals sampled in years 1997/1998 were identified as being in excellent condition as well as being larger than seals observed in 1999/2000. Although morphometric measures were also taken (standard length, axillary girth, weight and blubber thickness for example), these measures were not reported alongside nor in support for the BC index, with the exception of stating that of the leopard seals sampled in New South Wales, many were too weak to return to the sea and were euthanised due to judgement of poor body condition, which was later evident in examination of sternal blubber thickness (Gray et al., 2009). It was further conveyed that the blubber thickness of these individuals was distinctly less than the blubber thickness of the two leopard seals sampled in Antarctica (Gray et

al., 2009). However, while this study was the first to provide a visual scale for BC assessment, no additional morphological descriptions were provided to assist in allocating individuals into each category (Figure 1.3). A benefit to this visual scale means that by comparing leopard seal side-profiles to these four images, this technique can be applied to leopard seals in other studies and regions.

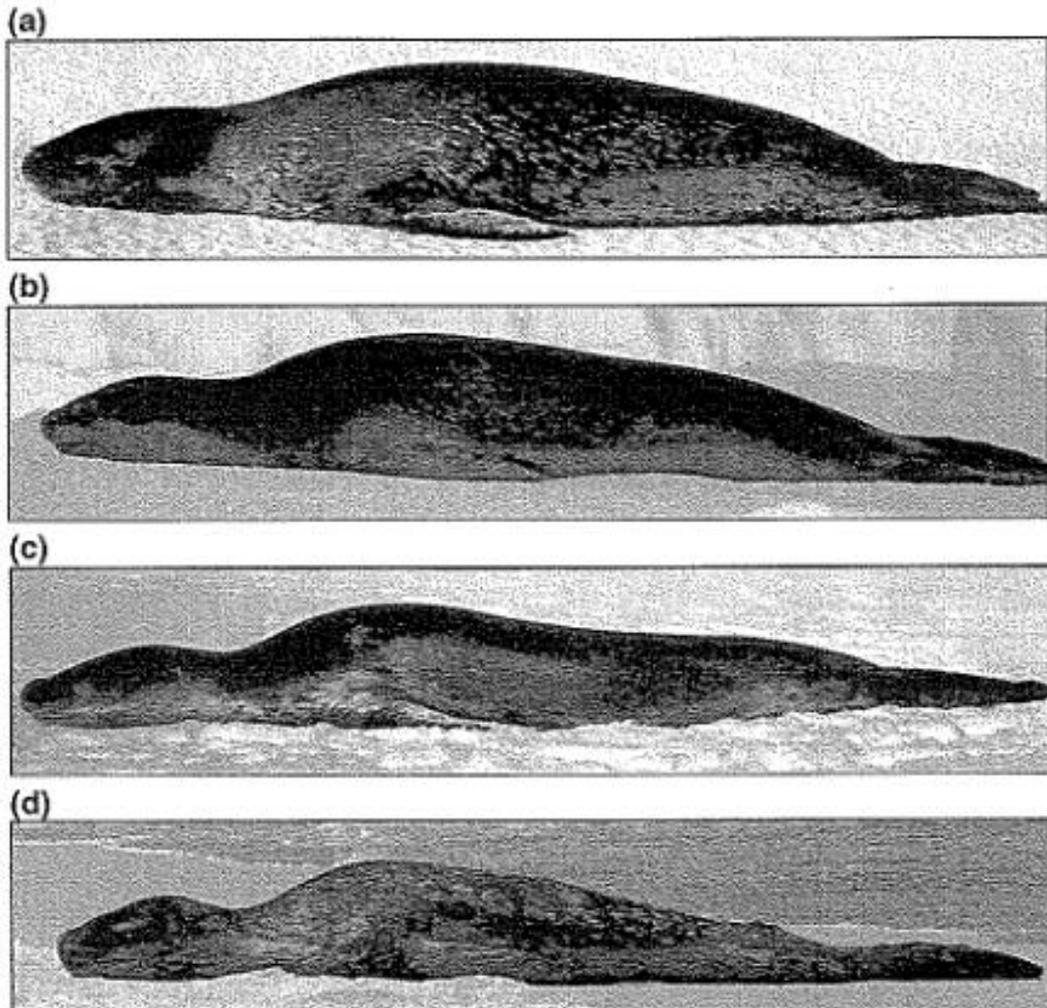


Figure 1.3. The body condition (BC) classification system extracted from Gray et al., 2009, whereby: (a) excellent; (b) good; (c) fair/thin, and; (d) poor.

A publication by Hupman et al., (2019) used a body condition scoring system (hereafter BCSS) to categorise leopard seals in New Zealand. Using images of a full-body profile, leopard seals lying in ventral recumbency, parallel to the ground were assessed for visible bony protrusions. Following methods modified from Gray et al., (2009), research by Hupman et al., (2019) assessed each animals BC by providing qualitative scores to visible bony protrusions including the sagittal crest, zygomatic arch, neural spines, rib bones and pelvic bones (Figure 1.4 and Figure 1.5; their Table 1 and Figure 3, Hupman et al., 2019). When a bony protrusion was visible it was scored as 1 and when it was not

visible it was scored 0 (Figure 1.4 and Figure 1.5; their Table 1 and Figure 3, Hupman et al., 2019). Scores were summed, and the BC of each animal was classified as Severe (score of 5; all bony protrusions were visible), Poor (score of 3-4; most bony protrusions were visible), Good (score of 1-2; some bony protrusions were visible) or Excellent (score of 0; no bony protrusions were visible). This allowed Hupman et al., (2019) to classify leopard seal BC into four distinct categories; 'Excellent', 'Good', 'Poor' and 'Severe' (Figure 1.4). Employing this method, Hupman et al., (2019) suggests that sightings of leopard seals in New Zealand are primarily individuals of Good and Excellent BC.

Sagittal crest	Zygomatic arch	Neural spines	Rib bones	Pelvic bones	Score	Classification
1	1	1	1	1	5	Severe – All clearly visible
e.g. 1	e.g. 0	e.g. 0	e.g. 1	e.g. 1	3-4	Poor – Most clearly visible
e.g. 0	e.g. 0	e.g. 1	e.g. 0	e.g. 0	1-2	Good – Some slightly visible
0	0	0	0	0	0	Excellent – None visible

Figure 1.4. The body condition scoring system extracted from Hupman et al., (2019), their Table 1.

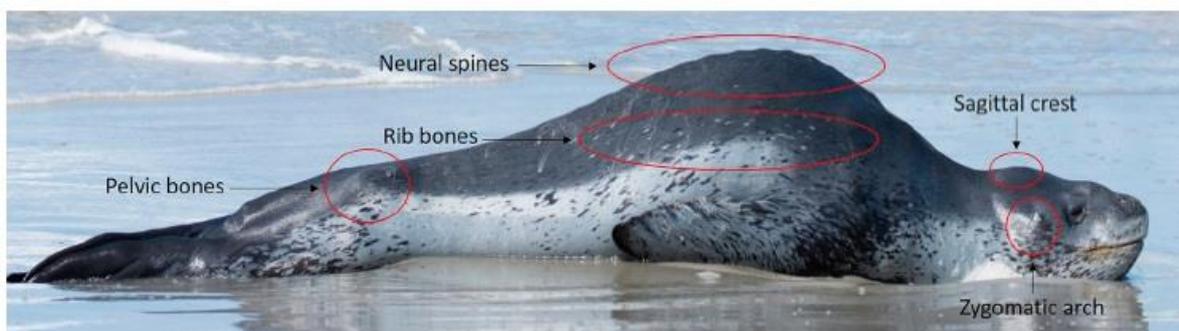


Figure 1.5. An image showing locations of bony protrusions, extracted from Hupman et al., (2019), their Figure 3.

1.7.3 Measurements of blubber volume

An approach used by Kuhn et al., (2006) adopts the truncated cones method, using measurements of length, girth and blubber depth (obtained by an ultrasound scanner) to determine blubber volume. Kuhn et al., (2006) applied this method to examine the body composition of a studied leopard seal which was comprised of approximately 70% lean tissue and 30% adipose tissue. This approach however was only opportunistically trailed on a single juvenile male (Kuhn et al., 2006). As such, more observations would be required in order to compare blubber volumes between individuals, and/or provide an estimate of BC within leopard seal population(s).

1.7.4 Morphometrics to estimate mass

An alternative method is using morphometrics to estimate mass. Hofman (1975) studied the relationship between length (L), girth (G) and mass in Antarctic seals, using linear models to estimate

the weight of anaesthetised leopard seals. He proposed that the geometric form of a seal is similar to that of two adjoining cones with a common base (Figure 1.6), and that equation LG^2 should represent volume of the seal. Hofman (1975) found that although seal weight correlated highly with the volume index, the index was sensitive to G measurements and the regression equation had tendency to overestimate weights of smaller leopard seals, whilst underestimating weights of larger leopard seals.

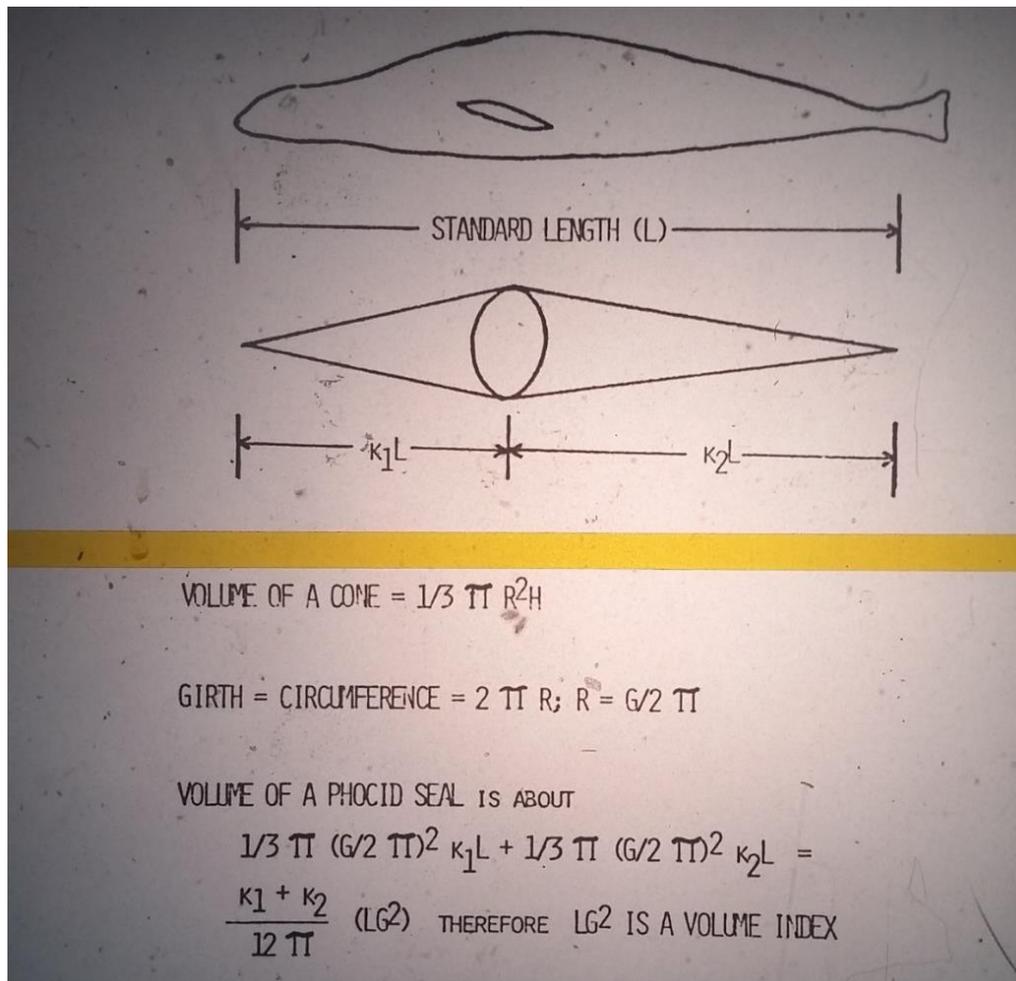


Figure 1.6. The theoretical relationship between girth (G), length (L) and volume in phocid seals extracted from Hofman (1975).

More recently, research by van den Hoff et al., (2005) aimed to develop similar equations that would predict body mass using morphometric measurements. In this study the leopard seals observed at Heard Island, Antarctica, were pre-assigned into five subjective visual body condition groups (hereafter BCGs): very poor; poor; moderate; good, and; very good. The seals were chemically immobilised, and measurements of snout-tail length (STL) and girth were input into the Smirnov Index ($STL * G^2 * 10^{-5}$) and Fitness Ratio (the ratio between seal length and seal height) to estimate seal mass. The results of these indices were compared to determine which would be the most suitable for future

use, both however, were found to have high levels of uncertainty particularly with larger seals. van den Hoff et al., (2005) reported that from using the Smirnov Index a seal with a volume of 40 could potentially have a mass that ranges between 150kg and 210kg, and larger seals were subject to more variability as a seal with a volume of 80 could weigh from 280kg and up to 400kg. Using Fitness Ratio, a seal 2.8m long could have a mass between 240kg and 480kg, double the size. Employing ANOVA, van den Hoff et al., (2005) reported that significant differences existed between the means of Fitness Ratio and Smirnov Index values between the pre-assigned BCGs. Further post-hoc testing however revealed not all means were significant and therefore BCGs were amended to: poor (=very poor + poor); medium (=moderate + good), and; good (=very good; van den Hoff et al., 2005). The morphology of two leopard seals belonging to the new BCGs were described, whereby one individual that was judged as medium condition had a thick neck and a “*well-rounded*” body, whilst the other who was assessed as being in poor condition had a thin neck, protruding pelvic region and “*wasted*” appearance. Both descriptions were accompanied by images (Figure 1.7), and subsequent ANOVA and post-hoc testing was able to significantly delineate Smirnov Index and Fitness Ratio values of seals in poor condition, to those in moderate and good condition. The amended BCGs (derived from visual examination of leopard seals at Heard Island) were then applied to leopard seals observed in three distinct regions; Antarctic (Palmer Station), sub-Antarctic (Macquarie and Heard Island) and sub-Tropical (New Zealand) in order to make comparisons of BC. van den Hoff et al., (2005) reported that the BC of leopard seals observed in the sub-Tropical region were recorded as 1.06 index groups (on average) less than leopard seals observed in the sub-Antarctic and 0.5 index groups (on average) lower than leopard seals observed in the Antarctic.



Figure 1.7. Photographs extracted from van den Hoff et al., (2005) that illustrates two individual leopard seals belonging to two of the visual body condition groups (BCGs), whereby a) shows a leopard seal judged to be in medium body condition (BC) based upon “a thick neck and well-rounded body” in addition to a snout-tail length of 180cm and a girth of 230cm. Photograph b) shows a leopard seal judged to be in poor BC, based upon a “wasted body appearance, a thin neck and protruding pelvic region” and snout-tail length to height ratio of 5.5:1 (van den Hoff et al., 2005).

Research by both Hofman (1975) and van den Hoff et al., (2005) both reported large degrees of error in estimations of leopard seal mass, rendering estimates of mass unreliable. For van den Hoff et al., (2005) the addition of the G measurement in the Smirnov Index yielded slightly more accurate results than the Fitness Ratio, although this measurement was difficult to obtain in the field as it requires anaesthesia, of which leopard seals are particularly vulnerable (Vogelnest et al., 2010; Pussini and Goebel 2015). Furthermore, the dataset of van den Hoff et al., (2005) was comprised of several, smaller datasets from different regions, which employed different methods to obtain measurement data. For example, in this study leopard seals were weighed via different equipment depending upon the region they were observed in, meaning that results were not directly comparable. Based upon the images in Figure 1.7, a standardised method for visually assessing BC was also lacking, as leopard seals

in these images were photographed from different angles and under different conditions (the leopard seal in 1.7a was in the process of being administered an intramuscular injection of sedative; van den Hoff et al., 2005). The implication of this is that despite the visual BC index developed in this research having been successfully validated by two condition indices, due to variations in the method the visual BC index cannot be reliably applied to studies of leopard seal BC in other regions. Moreover, Hofman (1975) stated that while length is related to weight, it does not effectively express the condition of an animal. Similarly, estimates of weight and body mass of leopard seals do not directly provide an evaluation of body condition or individual health, since these values also represent sex and life-stage of an individual.

1.7.5 Unmanned aerial systems and photogrammetric models to estimate mass

Krause et al., (2017) developed a non-invasive method that will enable researchers to determine size and mass of leopard seals using photographs taken from an UAS (unmanned aerial system). Images obtained by the UAS were processed in ImageJ software (<https://imagej.nih.gov/ij/>) where a customised Java script (also known as a macro called 'ten widths') was loaded to enable semi-automatic processing. The script marks 10 equidistant points on a line drawn from the tip of the snout to the tip of the tail, representing photographic standard length (PSL; Figure 1.8). A total of 15 photographic measurements were taken (as seen in Figure 1.8) and were measured in pixels. Using the height of the drone (distance from lens to the object) and focal length of the lens, the photographic pixel measurements were converted into accurate ground units (Krause et al., 2017).

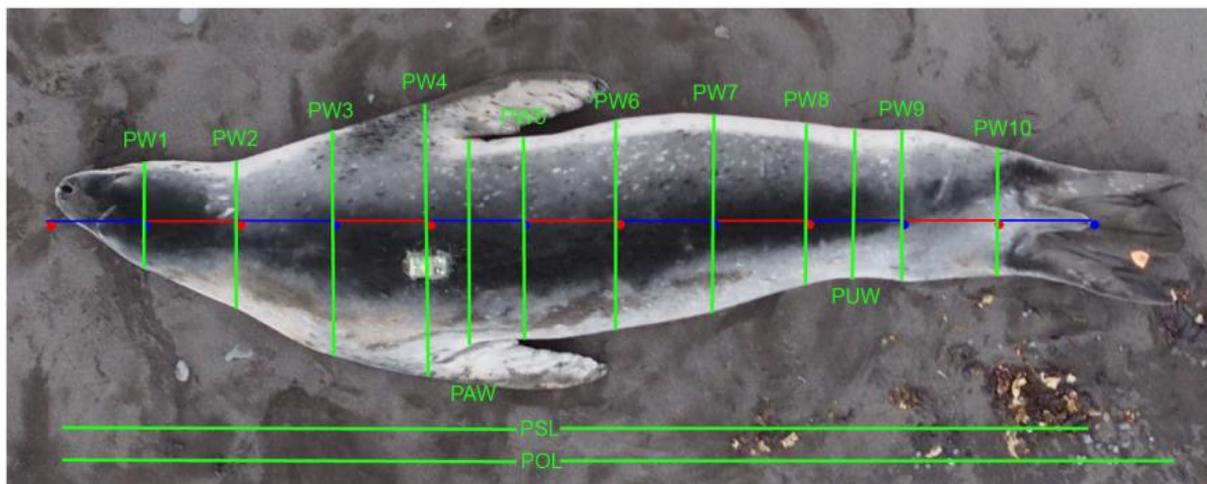


Figure 1.8. A leopard seal photographed by an unmanned aerial system (UAS) complete with the 15 photographic measurements; PW1-10 (width 1-10), PAW (axillary width), PUW (umbilical width), PSL (standard length) and POL (overall length) extracted from Krause et al. (2017), their Figure 3.

Ordinary least squares, linear and power regression, and linear mixed effects models were all utilised to determine leopard seal mass, using both photographic and manual measurements. This study reported the first instance in which photogrammetric models were more accurate at predicting seal body mass than models using manual measurements, since all photogrammetric models contained considerably lower residual error than manually derived models (Krause et al., 2017). This was likely a result of reduced measurement error and addition of body width measurements taken from overhead photos from the UAS (Krause et al., 2017). This included a measure of photographic umbilical width (PUW; Figure 1.8), which was found to correlate most highly with seal mass (Krause et al., 2017). The authors further suggested that this particular measurement should be considered when developing a BC index for leopard seals (contrary to the commonly used girth measurement (van den Hoff et al., 2005) labelled PAW or 'photographic axillary width'; Figure 1.8), and that the index should be $\text{condition index (CI)} = \text{photographic overall length (POL)} / \text{PUW}$ (Krause et al., 2017).

No significant differences were detected between mean values for photographic and manual measurements, supporting the further use of photogrammetry in estimating seal mass in future research. It should be noted however, that these findings are based upon a small sample size ($n=15$) of adult female leopard seals, possibly in a state of reduced body fat following the breeding season which occurs from November to January (Stirling and Siniff 1979). It therefore cannot be assumed that similar findings and relationships would be observed in leopard seals of the opposite sex, in different age classes and in different regions. Nevertheless, these findings offer an alternative to the use of risky and costly manual procedures and provide promise for the further use of photogrammetry for future studies of leopard seal mass and BC.

1.7.7 Evaluation of body condition assessment methods

Research involving chemical immobilisation of leopard seals is frequently associated with high mortality rates due to incorrect estimates of seal mass with respect to volume of administered sedative (Mitchell and Burton 1991; Higgins et al., 2002). Although leopard seals are large phocids (the second largest species in Antarctica after the Southern elephant seal, *Mirounga leonina*), they are characteristically slender and as a consequence their mass is commonly overestimated resulting in higher doses of sedative, which causes anoxia by means of apnoea (Hofman 1975). In research by Hofman (1975), of the 23 seals that were sedated were four mortalities and two which required assistance in forcing open the nasal plates to ensure breathing and prevent anoxia (Hofman 1975). Leopard seals have also been observed to become agitated by the drugs and retreat to the water, which may lead them to drown (Vogelnest et al., 2010). Mortality rates of leopard seals in such studies ranges between 5 and 38% (Mitchell and Burton 1991; Higgins et al., 2002). Research was conducted

by Vogelnest et al., (2010) on the anaesthesia of leopard seals, where the mass of individual seals was estimated using the Fitness Ratio published by van den Hoff et al., (2005). Here Vogelnest et al., (2010) reported that mass was overestimated on 9 occasions up to 110kg, and underestimated on 11 occasions by 80kg, stating that this was a significant error margin within their field operations.

Despite being described as non-invasive, methods described from Kuhn et al., (2006) and van den Hoff et al., (2005) were both disturbance procedures which carried substantial risk to both seals and researchers and required specialised, often non-viable equipment that would have entailed training in order to use. For example, anaesthesia involved the use of oxygen masks which are both heavy and impractical for researchers to carry on slippery ice (Vogelnest et al., 2010). Similarly, weighing seals required large slings (or nets) together with a weighing system (such as a hydraulic crane) which may be sufficient for use on the Antarctic islands (van den Hoff et al., 2005), but are perhaps not suitable for the pack-ice where leopard seals are most commonly found. For these reasons, sample sizes in such studies can be small ($n=1$; Kuhn et al., 2006, $n=7$; Hofman 1975). Within Hofman (1975) it was stated that the small sample size did not represent juvenile individuals, therefore results are often not generally representative of leopard seals as a species.

1.8 Thesis rationale

In the Antarctic bionetwork leopard seals are at the top of the food web with orca (*Orcinus orca*) as their only known predator (Jefferson et al., 2015). As result the ability to interpret their BC within this area and other regions they occupy is important for understanding the health of the population as well as serving as a reflection of the health of their oceanic environment (Krause et al., 2017). Assessment of BC is also imperative in research of marine mammals as it can provide information regarding which individuals are suitable for research without employing invasive techniques (*i.e.* avoiding examination of poor BC individuals). To date, much of the research investigating BC of leopard seals throughout the southern hemisphere has involved a range of techniques being employed across many different studies. At present these methods include: unverified methods (Rounsevell and Pemberton 1994; Aguayo-Lobo et al., 2011; Bester et al., 2017), visual scales (Gray et al., 2009; Hupman et al., 2019), measurements examining blubber thickness (Kuhn et al., 2006) morphometrics to estimate mass (Hofman 1975; van den Hoff et al., 2005) and photogrammetric models to estimate mass (Krause et al., 2017). As each of these studies employ different techniques indices of BC are not comparable between populations or across regions. Moreover, only a small number of the techniques used quantitative measures, some of which have been shown to have high degrees of error (van den Hoff et al., 2005) and most of which use procedures which carry substantial risk to leopard seal mortality (Hofman 1975; van den Hoff et al., 2005; Kuhn et al., 2006).

This study intends to provide validation for a BCSS by extracting measurements from photographic data obtained by citizen science, which can be used for conservation initiatives in the region. This thesis further aims to provide non-invasive, minimum-effort, objective and accurate methods as a standardised method for measuring BC of this species. The manuscript intends to combine BC scores (hereafter BCS) as employed in Hupman et al., (2019) with statistical modelling and machine learning techniques, which are practical for a citizen science approach. Moreover, these methods can be applied to leopard seals in other regions which would enable comparisons of BC across populations inhabiting the Southern hemisphere.

Chapter 2 : Body condition scoring as a non-invasive method for examining condition in sighting records of free-ranging New Zealand leopard seals from photographs



Leopard seal photographed resting on Owhiro Bay beach (New Zealand)

2.1 Introduction

2.1.1 Photo-identification

Photo-identification (hereafter photo-ID) is a low-risk, cost effective and minimally invasive research technique, that uses photographs to identify distinctive markings for individual recognition within a target species (Würsig and Würsig 1977). This method has been applied across a huge array of species for assessing abundance, distribution and movement patterns. Key assumptions of photo-ID are that: markings are unique to individuals; marks can be read without error, and; marks do not alter or get lost over time (Urian et al., 2014). Many pinniped studies have applied photo-ID for individual recognition. This is done by identifying distinctive natural or anthropogenic markings and/or pelage patterns (*i.e.* using spot patterns on seal pelage) and includes studies of: grey seals (*Halichoerus grypus*; Hiby et al., 2007; Sayer et al., 2019); harbour seals (*Phoca vitulina*; Mackey et al., 2007); leopard seals (Forcada and Robinson 2006; Hupman et al., 2019); Hawaiian (*Neomonachus schauinslandi*) and Mediterranean (*Monachus monachus*) monk seals (Harting et al., 2004; Forcada and Aguilar 2000), and; Saimaa ringed seals (*Pusa hispida*; Koivuniemi et al., 2016).

2.1.2 Photograph assessment

A key element to photo-ID is photograph assessment, *i.e.* the manual characterisation of which photographs are suitable for use. Assessing the quality of photographs for photo-ID is important to maintain uniformity within the data and to ensure that the same features are visible within each photograph for comparative purposes. Factors such as clarity and focus, contrast and orientation can affect the appearance and visibility of natural markings being identified, which can create false positive and false negative errors (Urian et al., 2014). To exemplify this method, a recent study of common dolphins by Hupman et al., (2018) which utilised photo-ID assessed quality of photographs by classifying images into categories, for example poor, fair, good or excellent (adapted from Urian et al., 1999 and Nicholson et al., 2012). First, each image was assessed to determine the proportion of the dorsal fin within the photo frame, and when the fin occupied less than 10% of the whole image then these photographs were discarded (Hupman et al., 2018). Secondly, images were assessed for photo quality, whereby each image was assigned a score based upon: clarity and focus; contrast; orientation, and; the visibility of the dorsal fin edge (Hupman et al., 2018). In this instance these scores were weighted, so that deficiency in any of the criteria outlined would result in a poor score for that particular image, which were subsequently discounted from further use (Hupman et al., 2018). The use of photograph assessment thereby eliminates photographs from examination which could potentially lead to false results due to imprecisions within the images caused by lack of detail in photographs with reduced resolution, poor lighting and/or obstructions.

Photograph assessment methods were employed in a photo-ID study of leopard seals combined with traditional tagging methods (Forcada and Robinson 2006). During a leopard seal observation multiple digital photographs were taken of the individual from six specific angles (Figure 2.1; their Figure 1, Forcada and Robinson 2006), in addition to notations of scars and wounds present for identification purposes. In this study, good quality photographs were described as those that were both in focus and with good definition, and in using these good quality photographs both tagged and untagged leopard seals were considered as photo identified (Forcada and Robinson 2006). When leopard seals were untagged, only individuals with an incomplete photograph series (Figure 2.1) that did not contain the most frequently photographed sides were discarded from further use (Forcada and Robinson 2006). Of the 269 occasions when leopard seals were observed, 59 were identified by these methods (Forcada and Robinson 2006). Using identifiable individuals from the photo-ID catalogue analysis of leopard seal population abundance, structure and estimates of turnover during the austral winter was conducted of leopard seals at Bird Island (Forcada and Robinson 2006). This research by Forcada and Robinson (2006) suggested that photo-ID techniques using leopard seal pelage patterns could be used to recapture known individuals as effectively as traditional tagging methods.

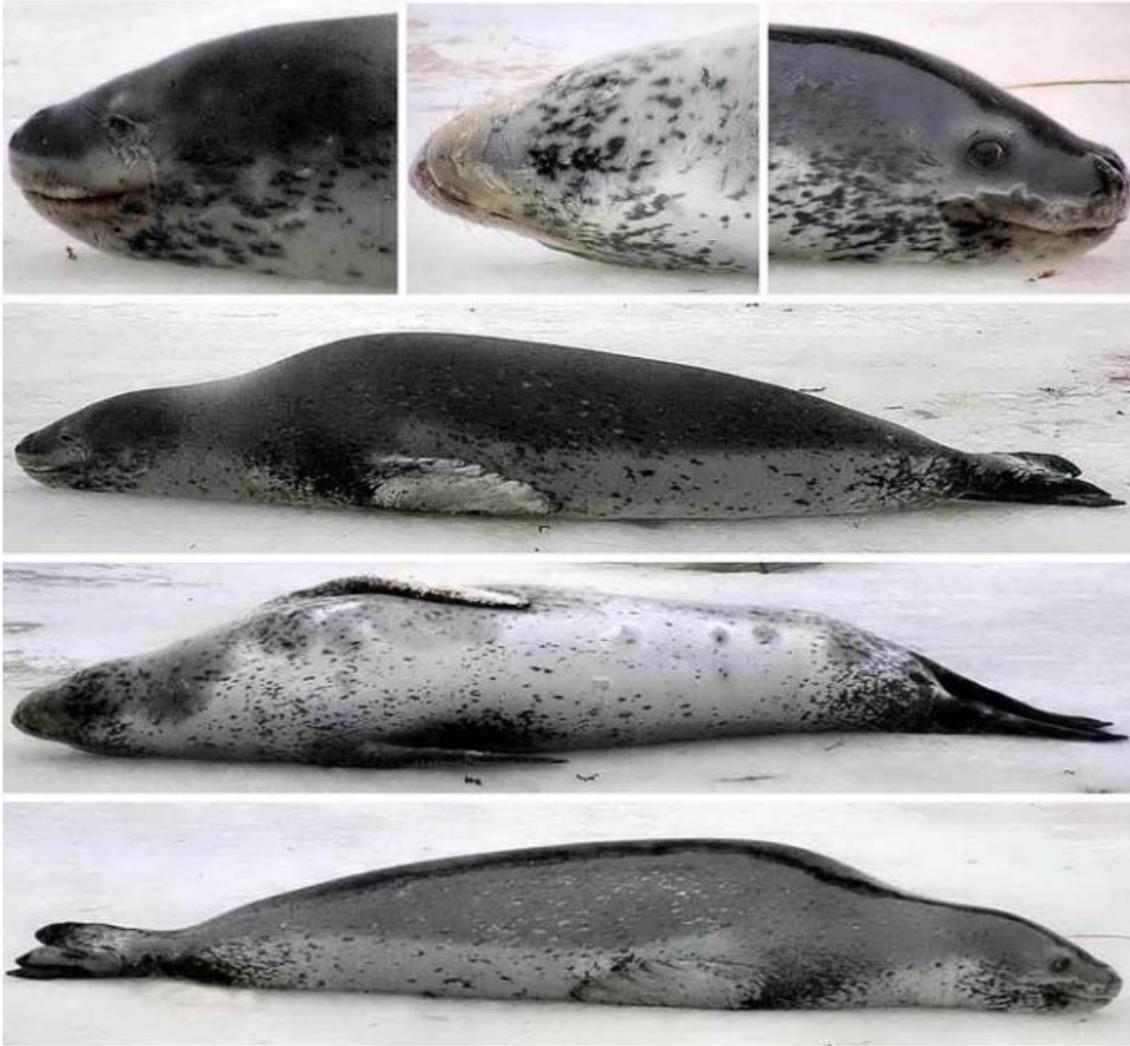


Figure 2.1. The six photograph views of leopard seals (*Hydrurga leptonyx*) used for photo-identification in a study by Forcada and Robinson (2006). From top to bottom (left to right) the photographs are as follows: head left (HL), throat (TH), head right (HR), body left (BL), ventrum (BV) and body right (BR).

2.1.3 Body condition scoring

Body mass and condition parameters are vital in the understanding of pinniped population health (Krause et al., 2017). Not only this, but as an upper trophic marine predator the health of pinniped species is reflective of the health of the ecosystem they occupy and thus serve as an indicator of habitat quality (Boyd and Murray 2001). One simple method for estimating body condition (BC) is using a system called body condition scoring (BCS). BCS is a subjective technique that visually (and sometimes physically) assesses the quantity of metabolisable energy stored on live animals, by comparing the amount of fat covering bony structures (e.g. the pelvis and the spine; Edmonson et al., 1988; Zeilke et al., 2018). It is usually done by looking at each key morphological feature individually and allocating a score of 'fatness'. Then, the average or sum of scores for the entire body can be used to allocate a BC category for an individual (Zielke et al., 2018). Since bony structures on the body

become more visible with decreasing fat density, individuals in a slimmer condition will receive lower BCS, whilst more rounded individuals (where bony structures are barely visible due to increased fat deposition) would receive higher BCS.

2.1.4 BCS origin and applications

The BCS method was originally introduced in the farming industry and has been described in the literature from as early as the 1960's when the method was trialled in ewes (Jefferies 1961). The ewes were allocated a score between 0 ("*the point of death*") and 5 ("*very fat*") based upon physical assessment, whereby the backbone and lumbar portion of the animals were examined to detect the sharpness of the bones as felt through the skin (Jefferies 1961; Edmonson et al., 1988). This technique was subsequently adapted to cows in both the beef and dairy industries in both the United States and the United Kingdom (Wildman et al., 1982; Mulvany 1977), as well as Australia and New Zealand whom combined the method with photographs (Earle 1967; Grainger and McGowan 1982). The use of photographs modernised the practice of assigning BC and allowed a complete hands-off approach, which reduced risk and stress to the animals as well as increasing efficiency, and today the management of welfare and fertility of dairy cows are entirely reliant upon this method (Edmonson et al., 1988; Zielke et al., 2018). BCS is also extensively practiced in veterinary medicine, where scoring charts exist for a number of animals including: dogs (Dorsten 2004); horses (Carroll and Huntingdon 1988); rabbits (Prebble et al., 2015), and; donkeys (Pearson and Oussat 2000).

In recent years, visual BCS systems (BCSS) have been applied to a variety of species, both wild and in captivity, these include; Asian elephants (*Elephas maximus*; Chusyd et al., 2019), mule deer (*Odocoileus hemionus*), elk (*Cervus canadensis*), moose (*Alces alces*; Cook et al., 2010) and European bison (*Bison bonasus*; Zeilke et al., 2018). In a number of studies, the relationship between the visual BCS system and quantitative physiological traits (thickness of rump fat, for example; Cook et al., 2010) was examined in order to validate the system. For European bison, the BCS system was developed with the purpose of monitoring the health of herds resettled in north-eastern Europe (Zeilke et al., 2018). To do this, Zielke et al., (2018) developed a standard scale for the species that relied on "*a few clearly detectable characteristics*" (i.e. bones that were easy to distinguish on the body) ranging from 1 (emaciated) to 5 (obese) by examining photographs of animals both in semi-wild conditions and from zoos. It was essential that the BCSS was both reliable and easy to use, so that it could be taught to other organisations involved in the resettling program, and to people of different skill levels (Zielke et al., 2018). The successful creation of the BCSS would enable comparisons of individuals in herds across different countries, in different habitats and on similar re-introduction schemes (Zielke et al., 2018).

2.1.5 BCS applications to marine mammals

BCSS have also been applied to a number of marine mammal studies worldwide, including Gray whales (*Eschrichtius robustus*; Bradford et al., 2012), North Atlantic right whales (*Eubalaena glacialis*; Pettis et al., 2011) and short-beaked common dolphins (*Delphinus delphis*; Joblon et al., 2008). Research by Joblon et al., (2008) used a BCSS to determine condition and potential survivability of individuals within a population of short-beaked common dolphins. In this study a 4-point visual BCS scale was designed based upon anatomical landmarks which were indicative of emaciation and body condition and was created to determine potential of survivability in live-stranded dolphins, as well as a standardised measure to compare BC across all delphinid species (Joblon et al., 2008). In order to identify these landmarks, photographs of 802 short-beaked common dolphins were assessed in conjunction with morphometric measurements (as means of validation for the BCSS), and results showed a significant difference in measurements obtained from dolphins which were successfully released, and those that died or were euthanised as they were judged unfit for release (Joblon et al., 2008). Similarly, (Pettis et al., 2011) describes the first instance when a visual BCSS was developed for North Atlantic right whales by analysing body and skin condition, rake marks and blow hole cyamids using over 200 000 photographs. In this particular study, BC was evaluated based upon estimates of the quantity of subcutaneous fat in the area behind the blowhole (the nuchal crest), whereby individuals were allocated a score of 1 if (good condition) when the nuchal crest was flat or rounded, a score of 2 (moderate condition) if the nuchal crest was slight to moderately concaved or a score of 3 (poor condition) if the individual showed evident “humps” with significant concavity posterior to the hump in the nuchal crest (Figure 2.2; their Figure 1 Pettis et al., 2004). Using this technique Pettis et al., (2011) reported that adult female whales were significantly thinner in their calving year as well as the subsequent year afterwards as compared with the year prior to calving. Moreover, in assessment of right whales that were “presumed dead” during the course of the study, evaluation of images in the 5 years prior revealed that these individuals received lower BCS in this time, indicating compromised health. Pettis et al., (2011) further reported that the methodology used in this research was objective (assessed by comparing scores by multiple experienced researchers) and could be utilised in future studies of BC assessment for the species.

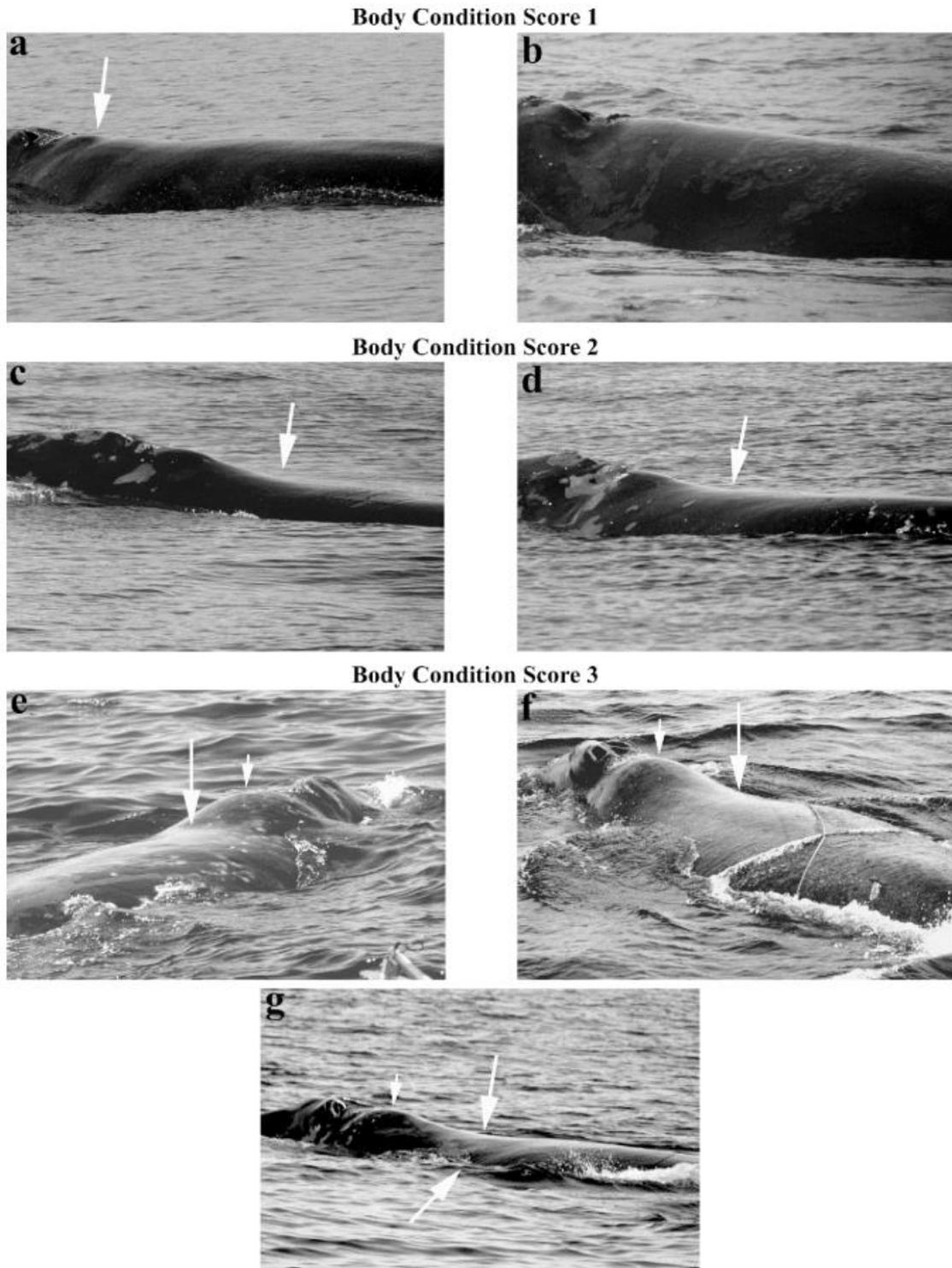


Figure 2.2. Body condition scoring system developed for North Atlantic right whales (*Eubalaena glacialis*) by Pettis et al., (2011; their Figure 1). Photograph a identifies an individual with a distinct fat roll in the nuchal crest region (as shown by the arrow), while photograph b shows a whale with a flat nuchal crest, both of which were identified to be in good (score 1) condition. Photographs c and d show individual right whales examined to be moderate (score 2) condition, with slight concavity behind the nuchal crest (as shown by the arrows), and photographs e–g show individuals that were characterised as poor (score 3) condition. For the three poor condition photographs the large arrows denote areas of concavity while small arrows identify distinct “humps” often seen in very underweight whales.

2.1.6 BCS applications to leopard seals

BCS was applied by Hupman et al., (2019) to New Zealand leopard seal data using methods outlined in Chapter 1. Despite the Hupman et al., (2019) study providing the most detailed and robust methodology to date for assigning BC from photographs, a number of shortcomings of these methods should be noted. The first is that the four BC categories were unequally weighted, as the BC groups Severe (a score of 5) and Excellent (a score of 0) were classified by one point each, whilst the BC groups Poor (scores of 3-4) and Good (scores of 1-2) were both characterised by two points each (refer to Figure 1.4, Chapter 1). This created a bias towards the BC groups of Poor and Good as individual records were more likely to fall into those categories than that of Severe and Excellent. Secondly, the BC categories progress from Good to Poor condition with no Moderate condition in between. In reality, an individual would not transfer directly from Good to Poor BC without going through an intermediate stage, however this was not recognised in the BC methodology of Hupman et al., (2019). Thirdly; the BC assessment in Hupman et al., (2019) was solely based upon the presence or absence of bony protrusions, and there was no quantification provided in the form of morphometric measurements (for example weight, blubber thickness, or measurements of girth). Bony protrusions can be difficult to detect making the BCS subjective as seals may be allocated different scores by different observers. It was therefore important to design a BCS system in which BC categories were of equal weighting with an intermediate Moderate category of BC. Furthermore, it was essential to provide a means of validation for the BCSS in order to support the subjective, visual BC categories.

2.2 Aims

This chapter aimed to describe the methods used to develop a robust BCSS using data from the New Zealand Leopard Seal Photograph Library (NZLSPL). In addition, this chapter aimed to develop a more robust BCSS to assess the BC of leopard seals in New Zealand waters to assist in the understanding of the health of these individuals and the marine environment they occupy.

Specifically, the objectives were to:

1. Describe the methods employed for both collection and recording of data;
2. Describe the procedures used when assessing photograph quality to be used in the BCSS;
3. Conduct photograph assessment on sighting records in the NZLSPL to identify photographs suitable for applications to the BCSS and successive photographic measurement (Chapter 3);
4. Develop a BCSS which is unbiased towards BC categories and included a 'Moderate' BC category, and;
5. Conduct an assessment of BC using suitable photographs identified by the two-tiered photo assessment criteria, using the BCSS.

2.3 Methods

2.3.1 Data collection

Leopard seal sighting records were derived from the NZLSPL which contains photograph sets of individual leopard sightings collected by an individual (member of the public or volunteer researcher, for example) organised by date and location, *i.e.* a sighting record labelled “20140915 Leopard Seal, Aramoana, Otago – Derek Cox” was collected on the 15th September 2014, at Aramoana in Otago, by Derek Cox. An individual leopard seal which has been the focus of ongoing study since 2015 was excluded from this analysis in order to reduce the impact of over-representation of individual seals and therefore limit bias within the results.

2.3.2 Photograph assessment

All sighting records within the NZLSPL were submitted to photograph assessment to ensure that images of high resolution were used where bony protrusions could be easily detected (when visible). Firstly, each photograph in a sighting record was graded according to Photograph Assessment Criteria (PAC). The PAC was used to determine if the seals’: full body was visible; full (left or right) side profile was visible; body positioning was parallel to the ground in ventral recumbency, and; body positioning was perpendicular to the camera (Table 2.1). Only sighting records which contained at least one photograph that met all of the described PAC were retained for further analysis. Secondly, each photograph in a photograph set was graded according to Photograph Quality Criteria (PQC). Photograph quality was assessed by assigning a score of 0 or 1 for resolution, lighting and contrast, whereby a 0 would be assigned to a photograph with insufficient quality and a 1 would be assigned to a photograph with sufficient quality for each of these criteria. Based on this, each photograph in a sighting record would thereby receive a score of 1–3, whereby Good=3, Fair=2 and Poor=1. Only sighting records which contained at least one photograph that achieved a score of 2 or 3 in the PQC was retained for further analysis.

Table 2.1. Photograph Assessment Criteria (PAC) used to examine images of New Zealand leopard seals. Examples of photographs which passed and failed, whereby reasons for failing were as follows: no full body side profile visible due to obstruction by water (photograph B); no full body side profile visible due to cropping (photograph D), and; no full body side profile visible due to body position of seal being not parallel to the ground (photograph F) and not perpendicular to the camera (photograph H). Photographs A, B, D and G were taken by Thelma Wilson, Gary Mathers, Kerry Butterfield and Katharina Achterberg (LeopardSeals.org unpublished data), while photographs C, E, F and H were taken by the author.

Photographic criteria	Pass	Fail
Full body visible		
Full (left or right) side profile visible		
Body position parallel to the ground in ventral recumbency		
Body position perpendicular to the camera		

Table 2.2. Photograph Quality Criteria (PQC) used to examine images of New Zealand leopard seals. A score of 1 was assigned to photographs with sufficient quality for each of the three criteria (resolution, lighting and contrast) while a score of 0 was designated for photographs which were of insufficient quality with regard to the criteria. Photographs were taken by Giverny Forbes (A, C and E), Joanne Thwaites (B), Thelma Wilson (D) and Neil Morrison (F; LeopardSeals.org unpublished data).

Photograph Quality Criteria	Score of '1'	Score of '0'
Resolution		
Lighting		
Contrast		

2.3.3 Body condition scoring

Using sightings records from the NZLSPL that met both the PAC and PQC, a visual assessment of BC was conducted using all photographs for each record. Scores were stored in the Body Condition – Leopard Seals Database (hereafter referred to as the BCLSD) and assigned a sighting-ID for the purpose of differentiating between sighting records (*i.e.* sighting-ID HL001 referred to *Hydrurga leptonyx* sighting record number 1). Following Hupman et al., (2019), BC was graded by detecting the presence of visible bony protrusions including the: zygomatic arch; sagittal crest; neural spines; rib bones, and; the pelvis (Table 2.3). If bony protrusions were visible then a score of 1 was recorded and the sum of the scores allocated the seal into a category named a body condition group (BCG) whereby: 0-1 = Good; 2-3 = Moderate, and; 4-5 = Poor (Table 2.3, adapted from Hupman et al., 2019). Here, BCGs described by Hupman et al., (2019) were modified so that there was an equal chance of belonging to each category (*i.e.* each BCG has increments of 2, thus as scores can range from 0 to 5, there is a 1/3 or 33.3% chance of an individual belonging to one of the categories). Sighting records of leopard seals extracted from the NZLSPL were organised as ‘encounters’ and not photo-ID’d ‘known individuals’, however there were nine occasions ($n=18$ sighting records) when sighting records were identified having the same date and location, and therefore were likely to contain photographs of the same

leopard seal, and thus were denoted as ‘repeat encounters’. Each repeat encounter sighting record was assessed to determine if these records would be categorised into the same BCG using the BCS. Photo-ID of individuals was not a method used within this chapter as it is the focus of another study (LeopardSeals.org unpublished data).

Table 2.3. Body condition group (BCG) chart and classification for New Zealand leopard seals based upon visibility of bony protrusions, adapted from Hupman et al., (2019). Examples of body condition (BC) classifications are outlined to illustrate how body condition scores (BCS) were calculated.

BCG	Score	Example				
		Sagittal crest	Zygomatic arch	Neural spines	Rib bones	Pelvic bones
<i>Good</i>	0 - 1	e.g. 0	e.g. 0	e.g. 0	e.g. 0	e.g. 1
<i>Moderate</i>	2 - 3	e.g. 1	e.g. 0	e.g. 0	e.g. 1	e.g. 1
<i>Poor</i>	4 - 5	e.g. 1	e.g. 1	e.g. 1	e.g. 1	e.g. 1

Table 2.4. Body condition groups (BCG) of New Zealand leopard seals classified as Good (example image A), Moderate (example image B) and Poor (example image C) using a body condition scoring system (BCSS) following methods modified by Hupman et al., (2019). Photographs A and C were taken by Giverny Forbes and Francis Murphy (LeopardSeals.org unpublished data) while photograph B was taken by the author.

BCG	Example image
<i>Good</i>	
<i>Moderate</i>	
<i>Poor</i>	

2.4 Results

A total of 80 sighting records from the BCLSD met both the PAC and PQC. Sighting records were more frequently classified as Good ($n=57$; 71.25%), compared to those classified as Moderate ($n=15$;

18.75%) or Poor ($n=8$; 10%; Figure 2.3), identifying that leopard seals observed within New Zealand are predominantly in good condition. Of the 57 sighting records categorised as Good, many ($n=41$) were assigned a score of 0; *i.e.* no bony protrusions visible, representing just over 50% of sighting records. When bony protrusions were visible ($n=39$), the pelvic bones were the most frequent protrusion that was identifiable, observed in 84.6% of the sighting records ($n=33$), followed by the sagittal crest ($n=21$; 53.8%), neural spines ($n=11$; 28.2%), rib bones ($n=10$; 25.6%) and zygomatic arch ($n=7$; 17.9%; Figure 2.4). The zygomatic arch was only seen in Poor BC sighting records where seals were categorised 4 or 5 for BCS. Although the neural spines and rib bones were observed a similar number of times ($n=11$ and $n=10$; Figure 2.4) they were only seen in conjunction with one another on four occasions, again within sightings records that contained Poor BC individuals.

Regarding the ‘repeat encounter’ sighting records, on eight out of nine occasions the seal within the sighting record was allocated the same BCS, specifically by identifying the presence of the same bony protrusions on each occasion. In the one instance where two different scores given, the sightings records were split into two BCG; Good and Moderate. The sighting record with the sighting-ID HL072 (*i.e.* *Hydrurga leptonyx* sighting record number 72) was identified as containing a leopard seal in Good condition (a score of 0), whereas the record HL029 (*i.e.* *Hydrurga leptonyx* sighting record number 29) was reported as BCG Moderate (a score of 2) where the neural spines and pelvic bones were both observed.

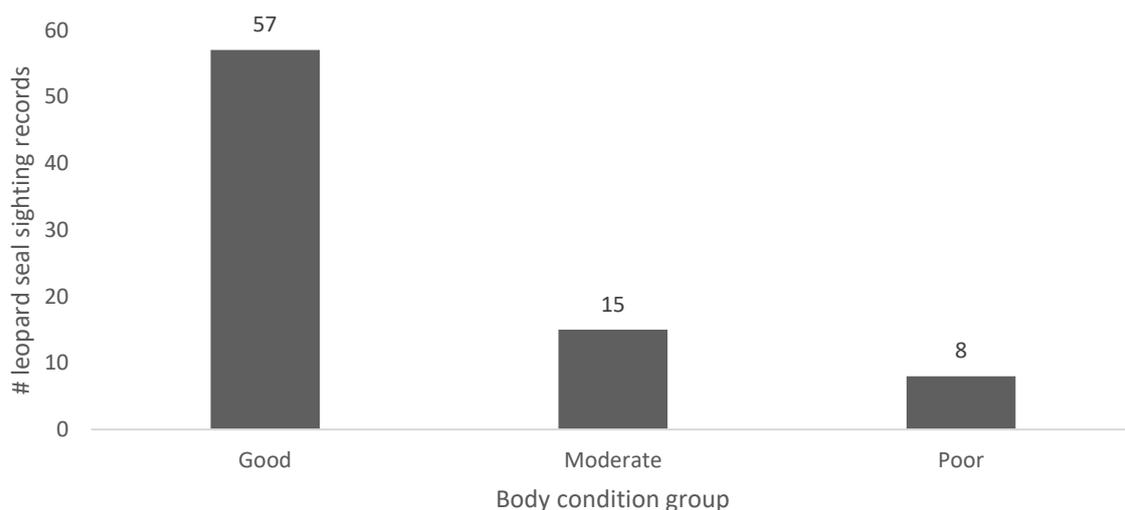


Figure 2.3. Number of individual New Zealand leopard seal sighting records categorised to each body condition group (BCG) based upon a body condition scoring system (BCSS).

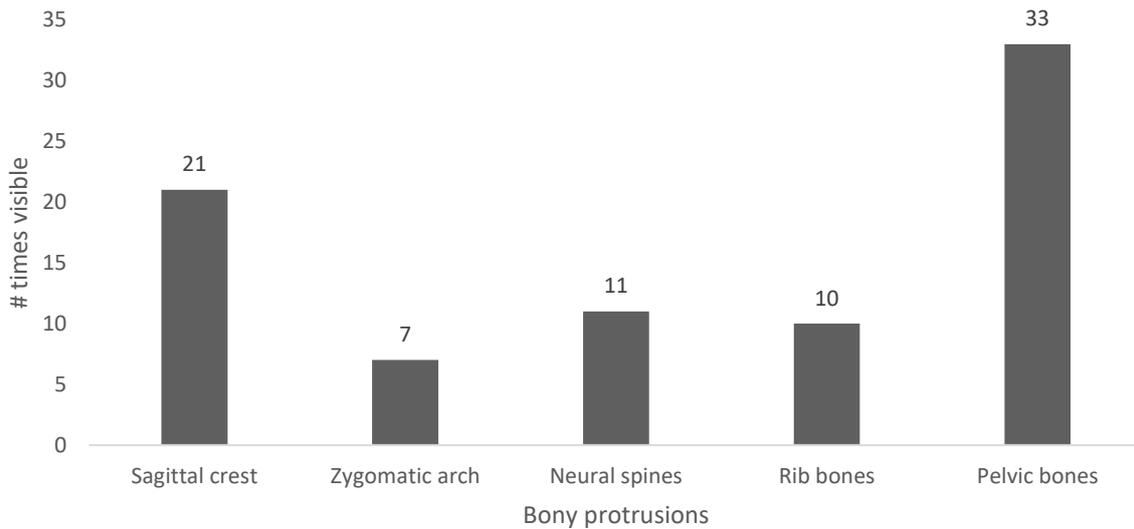


Figure 2.4. Frequency the five distinct bony protrusions were visible using a body condition scoring system (BCSS), distinguished for each New Zealand leopard seal sighting record.

2.5 Discussion

2.5.1 Photograph assessment

The number of sighting records containing photographs identified as suitable for BCS by means of photograph assessment within this chapter ($n=80$) were relatively low as compared with other studies using BCS methods based on photographic data. For instance, research by Joblon et al., (2008) on short-beaked common dolphins involved photographs of 802 individuals, while research on North Atlantic right whales by Pettis et al., (2011) utilised a database of over 200,000 photographs of 435 photo-ID'd known individuals. In comparison, the number of sighting records examined within this chapter ($n=80$) contained less than 5.7% of the total number of sighting records available in the NZLSPL, possibly as result of stringent photograph assessment criteria which eliminated the use of many photograph records. Regarding photograph assessment, limited information was provided in the literature concerning how photographs were sorted into those that were of sufficient quality to be used for analysis vs those that were not. Hupman et al., (2019) reported that photographs were examined for quality and categorised as either; poor, fair, good or excellent using methods outlined in Hupman et al., (2018), and of the 1408 photographic records of New Zealand leopard seals, 368 photographs were identified as suitable for analysis of BC using these categories. Comparatively, research by Zielke et al., (2018) on European bison states describes the challenge of photographing animals, stating that visibility of the animals was an issue as they were living semi-free, however details on how this was overcome were not conveyed. The study by Pettis et al., (2011) outlined difficulties with using photographic data for BC assessment, stating that 38.9% of the 200 000 photographs available were identified as not suitable for assessment of North Atlantic right whale

health parameters (including BC, skin condition, rake marks and cyamids). Furthermore, of the remaining photographs 68.4% were unsuitable for assessment of BC due to the body positioning of the whale, as well as the angle the photograph was taken, as it was difficult to make BC assessments on images that were not taken laterally parallel to the whale (Pettis et al., 2011). The body position of the whale was identified an important consideration when allocating BCS as photographs of whales engaged in feeding or head lifting behaviours created a “*temporary concavity*” which meant these images were not suitable for BC assessment, however, no further information or scales were provided to illustrate how photographs were sorted into suitable or unsuitable for use (Pettis et al., 2011). Within this chapter photograph assessment was an important element to BCS, to ensure that images of leopard seals were standardised so that when bony protrusions were protuberant, they could be clearly detected from high resolution images with good quality and lighting (Table 2.2). Correspondingly to limitations outlined by Pettis et al., (2011) the body position of the leopard seal was also identified within this study as being important for BC assessment, and thus only photographs of the whole left or right side-profile while the animal was lying in ventral recumbency, parallel to the ground and perpendicular to the camera should be used in future analysis of leopard seal BC (Table 2.1).

2.5.2 Body condition scoring

BCSS have not been thoroughly explored as means of assessing population health for pinnipeds, despite being described as early as the 1960s for applications for agriculture (Jefferies 1961), as well as being employed for other marine mammal studies in recent years (*i.e.* short-beaked common dolphins; Joblon et al., 2008 and North Atlantic right whales; Pettis et al., 2011). The BCSS developed for leopard seals in this chapter was based upon five clearly detectable bony protrusions; sagittal crest, zygomatic arch, neural spines, rib bones and pelvic bones. Three of these bony protrusions (neural spines, rib bones and pelvic bones) were common with three of the five bony protrusions (ribs, spine and hips) used in a BCSS by Zielke et al., (2018) for assessing BC in European bison. The five bony protrusions identified by Zielke et al., (2018) were scored separately between 1 and 5 based upon the gradation of visibility, and the average of all five scores equated to the BCS for that individual. In comparison, BCS in this assessment ranged between 0 and 5 where a score of one was allocated for each bony protrusion visible. Research by Pettis et al., (2011) and Joblon et al., (2008) utilised simpler BCS methods. For North Atlantic right whales only one morphological area was assessed (the nuchal crest) as this was the only area consistently photographed above water, where BC was categorised based upon the degree of concavity or convexity in this one morphological region (Pettis et al., 2011). The study of short-beaked common dolphins also assessed concavity in the nuchal crest, in addition to examining the epaxial section (ventrolateral to the dorsal fin) and the thoracic wall (rib bones) as

well as overall shape of the trunk as part of a more robust 4-point BC scale (Joblon et al., 2008). Alike methods employed within this chapter, however, Joblon et al., (2008) had the advantage of photographing dolphins out of water, in ventral recumbency and on a firm surface, while Pettis et al., (2011) utilised images of free-ranging North Atlantic right whales whereby the majority of the animal was submerged underwater in each photograph.

While BCS is not always the most accurate method due to its subjectivity between observers, it is an extremely useful, non-invasive and low-risk technique for investigating the ecology of wild and free-ranging animal populations, without requiring specialised or costly equipment. BCS is particularly useful in studies of marine mammals, where investigating BC is often difficult due to restricted visibility of an animal's body while submerged underwater (Pettis et al., 2011). BCS can therefore be opportunistically employed to species of pinniped, which spend a proportion of their time on land where the entire of the body is visible to observers. As result, all morphological areas can be assessed for bony protrusions which gives a more accurate representation of individual health. The BCS method is also particularly useful with regard to citizen science due to its application to photographic data, as seen in research by Hupman et al., (2019), where photographs taken of leopard seals by members of the public were also incorporated into the database alongside photographs taken by volunteer researchers. This significantly increased the size of the dataset which portrayed a more accurate representation of BC within the leopard seal population in New Zealand.

2.5.2.1 Bony protrusions

Of all five bony protrusions within the BCSS the pelvic bones were most frequently observed suggesting the umbilical region may be significant for determining body condition in the leopard seal species. This is in agreement with research by Krause et al., (2017) who found that the umbilical width (located in front of the pelvic bones, as discussed in Chapter 1) correlated strongly with seal mass. As result of this the umbilical region was identified as a morphological area of interest for the subsequent chapter of this study during extraction of photographic measurements. The bony protrusions neural spines and rib bones were not always seen in conjunction with one another when present. On 13 occasions one of the two protrusions were observed compared to only four occasions where they were both recorded, despite these regions being located around the same body area. This suggests that perhaps not all leopard seals body fat is distributed throughout the body in the same way. There may be differences between unique individuals, between males and females and between different life stages which influenced the cover of fat on the leopard seal body, meaning BCS may not be directly indicative of health but also reflective of variation in morphology between individuals, which should be taken into consideration in future studies.

2.5.2.2 Validation

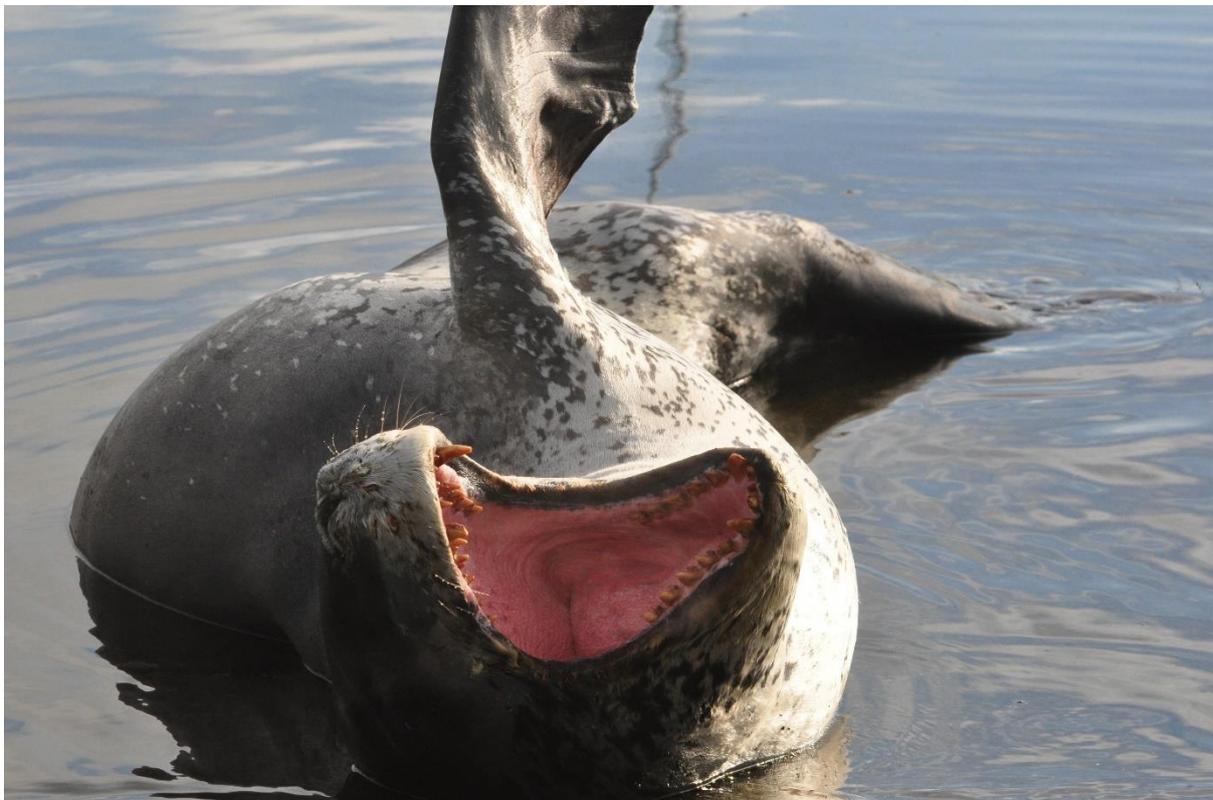
Similar techniques have been employed for assessing consistency in BCS between observers in research by Joblon et al., (2008), Pettis et al., (2011) and Zielke et al., (2018), who all utilised variations of blind studies, which were also a means of providing validation for the subjective BCS methods. Research by Joblon et al., (2008) found a significant level of agreement between assessors, by using a random assortment of 30 photographs sent to five assessors (trained stranding reporters) who independently used the BCSS to allocate a BCS to each animal. Similarly, a double-blind study was used by Pettis et al., (2011) which confirmed a “*strong agreement*” between the primary researcher and two independent researchers (experienced right whale biologists) for assessments of BC using 100 randomly selected photographs. Although an attempt at assessing consistency in BCS was conducted within this chapter by using sighting records containing repeat encounters, BCS were allocated by the same observer on each occasion (the author) and was only conducted on a small scale compared to other studies (Joblon et al., 2008; Pettis et al., 2011; Zielke et al., 2018). All but one of the repeat encounter sighting records were assigned the same BCG by identifying the same bony protrusions, reinforcing the reliability of the method. For the records where two different BCG were allocated for the same individual leopard seal, it was identified that this could be due to the number of photographs within each sighting. Sighting record HL072 contained 18 images, whereas sighting record HL029 contained 37 images. Naturally, with sightings that had larger photograph sets there was a higher chance that if a bony protrusion was visible, then it was identified - this factor may be responsible for the variation in BCS and subsequent BCG of the two sightings. In future assessments using BCS it may be useful to focus only on one photograph from each sighting record (*i.e.* the photograph of the leopard seal lying in ventral recumbency, parallel to the ground and perpendicular to the camera), rather than the entire set of images. Although this would mean that that bony protrusions present may be missed, error between each sighting record would be equal.

2.6 Conclusions

Research involving photo-ID and photograph assessment for pinnipeds exists for a range of species, including leopard seals (Forcada and Robinson 2006; Hupman et al., 2019), as methods for assessing population abundance, distribution and movement patterns. In contrast, BCSS have rarely been developed for pinnipeds in spite of their extensive applications and success for studies involving land mammals in agriculture (Jefferies 1961; Edmonson et al., 1988) and veterinary science (Carroll and Huntingdon 1988; Dorsten 2004) as well as wild and captive mammals (Cook et al., 2010; Zielke et al., 2018; Chusyd et al., 2019). Using photograph assessment methods outlined in this chapter 80 sighting records from the NZLSPL were identified as sufficient quality using two tiered assessment criteria; PAC

and PQC, and thus were extracted for application to BCS in order to determine composition of BC within the New Zealand leopard seal population. Although this dataset was substantially smaller than the dataset of Hupman et al., (2019) whom identified 368 photographs suitable for BCS, employing the BCSS 71.25% of sighting records from the NZLSPL were identified as containing images of Good condition leopard seals, with a further 18.75% identified as containing Moderate and 10% containing Poor condition seals. This was consistent with findings by Hupman et al., (2019), whom reported that the composition of BC within New Zealand leopard seals was; 33.34%=Excellent condition ($n=119$) 57.88% =Good condition ($n=213$), 4.89%=Poor condition ($n=18$) and 0.54%=Severe condition seals ($n=2$), which disputes existing publications from other northerly regions (regions where leopard seals have been observed outside of their core home range of the circumpolar pack ice of Antarctica) whom report that individuals observed in these areas of predominantly of poor condition (Rounsevell and Pemberton 1994; Mawson and Coughran 1999; Bester et al., 2017). In future research using BCSS in this species it would be advantageous to run BCS alongside photo-ID, so that individual leopard seals can be compared in addition to monitoring condition of individuals over time. Moreover, during BC assessment, it is suggested that only one photograph per sighting record is used to assess BC, to maintain the levels of error between sightings.

Chapter 3 : Dimensionality reduction of photogrammetric measurements of body width extracted from sighting records of New Zealand leopard seals using Principal Component Analysis



Leopard seal photographed at Wellington's boat sheds on Oriental bay (New Zealand)

3.1 Introduction

3.1.1 Photogrammetry

Photogrammetry is a technique used to obtain measurements from photographs. This method is widely used in the field of science and has been applied for studies in obtaining measurements from species that are large and/or impractical to physically capture *i.e.* scalloped hammerhead sharks (*Sphyrna lewini*; Klimley and Brown 1983), fin whales (*Balaenoptera physalus*; Ratnaswamy and Winn 1993) as well as North Pacific (*Eubalaena japonica*) and southern right whales (*Eubalaena australis*; Christiansen et al., 2019; Dawson et al., 2017). Research by Christiansen et al., (2019) employed photogrammetry in a study aimed to estimate body mass of free-ranging right whales (*Eubalaena* sp.), using an unmanned aerial vehicle (UAV) equipped with a camera, compared against historical catch records (Omura et al., 1969). Both dorsal and lateral angles of free-ranging southern right whales were photographed using an UAV, where the camera was facing directly down (Christiansen et al., 2019). From the dorsal photographs, total body length (BL; the tip of the rostrum to the tip of the tail notch) were measured in addition to measurements of body width at 5% increments, while measurements of height (H; dorsal to ventral distance) were measured at the same 5% intervals using lateral photographs (Figure 3.1; their Figure 1, Christiansen et al., 2019). All photographic measurements were converted from pixels to metric values using values for the relative size of the whale (in pixels), the size of the camera sensor, ratio of the altitude of the UAV (measured by a range finder) and camera focal length (Christiansen et al., 2018). Using these data, body volume of 86 individual whales were estimated by incorporating their condition (by modelling each individual as a series of ellipses; Figure 3.1) and size (employing a linear model which made use of body girth and length data; Figure 3.1). This model was applied to data from eight lethally caught North Pacific right whales (where actual body mass was measured) to obtain a body volume-body mass conversion factor (Omura et al., 1969; Christiansen et al., 2019). Results by Christiansen et al., (2019) reported that this non-invasive photogrammetry approach yielded accurate estimates of body mass, which fell within close range to actual values of body mass. Furthermore, it was suggested that this technique can be applied to other free-ranging marine mammals, simply by adjusting model parameters. Photogrammetric methods were also applied by Krause et al., (2017) to estimate mass of leopard seals in Antarctica using an unmanned aerial system (UAS; refer to Chapter 1). Results from this study revealed the first instance where photogrammetric methods were more precise than traditional manual measurements for estimating mass (Krause et al., 2017), reinforcing the reliability of photogrammetric measurements.

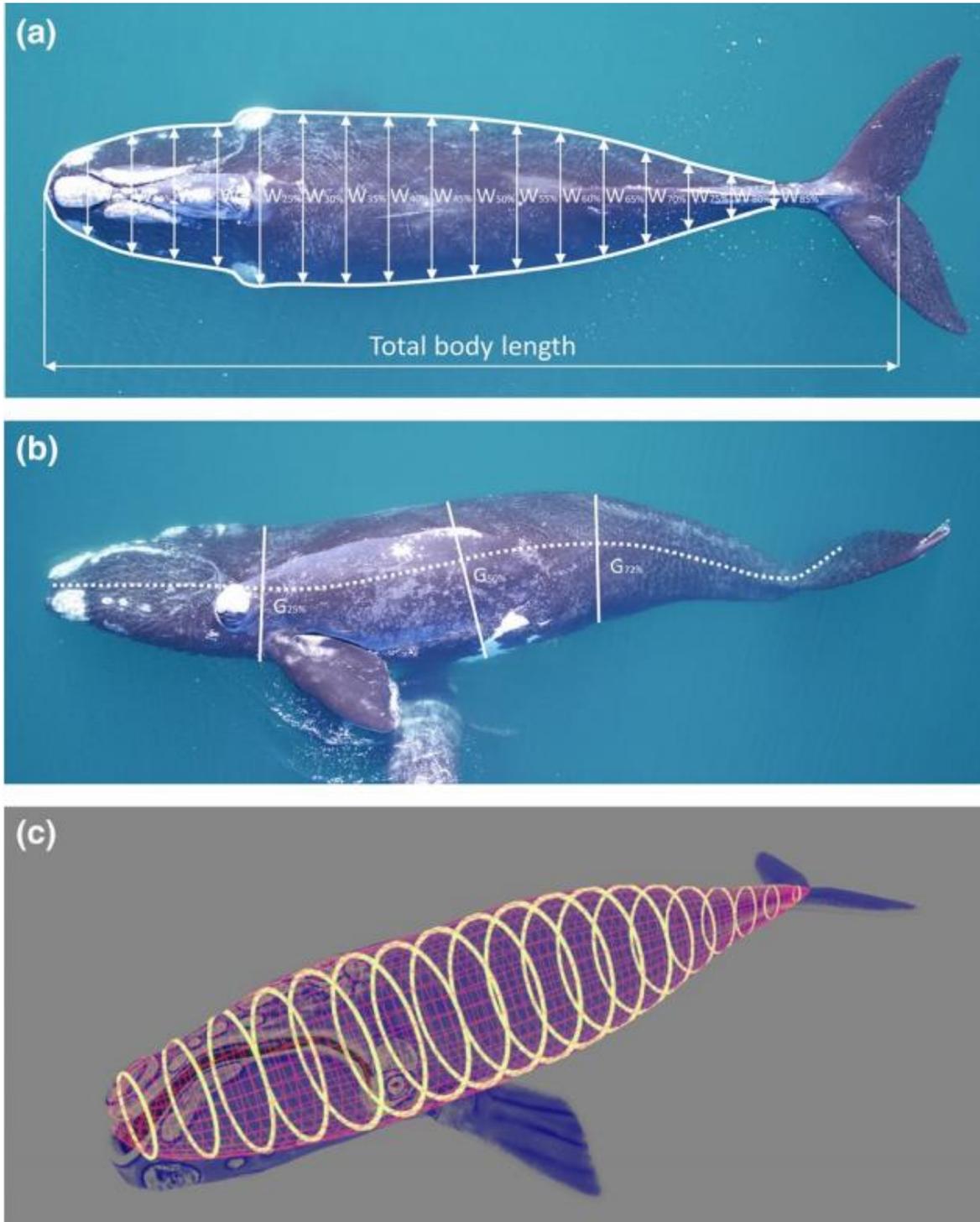


Figure 3.1. Examples of photographs of an individual southern right whale (*Eubaleana australis*) taken from the unmanned ariel vehicle (UAV) extracted from Christiansen et al., (2019). Photograph (a) shows the dorsal side of the whale complete with the photographic measurement of total body length (BL) from the tip of the rostrum to the tip of the tail notch (as shown by white arrows) and the body width measurements at 5% increments (W). Photograph (b) shows the same whale photographed from the lateral side, used to extract photographic measurements of body height (i.e. the dorsal to ventral distance; H) at the same 5% increments, the solid lines represent girths (G) at 25% (pectoral fin), 50% (umbilicus) and 72% (anus) body length. Photograph (c) shows a 3D model of the same whale photographed in (a) and (b), as a series of ellipses used to estimate body volume, where each ellipse represents the variation in height-width ratio (Christiansen et al., 2019).

3.1.2 Principal Component Analysis

Principal Component Analysis (hereafter PCA) is a statistical technique of feature extraction used for dimensionality reduction, when there are a large number of variables (Manly 1994; Kim and Kim, 2012). Dimensionality reduction works to decrease the number of variables, without losing the original variation and trends within the data (Tharwat et al., 2017). As a result, fewer relationships exist between variables, therefore making it easier for the user to identify these relationships as well as trends within the data (Czarnowski and Jędrzejowicz 2018). PCA works by transforming the original correlated variables into a smaller set of uncorrelated variables (artificial variables) named principal components, which account for a large proportion of the variability within the data (Manly 1994; Kim and Kim 2012). These new principal components are ordered in such a way that the first PC (PC1) commonly contains the most variation within the independent variables, and the subsequent PCs account for the majority of the remaining variation, that is not contained in the former PCs (Jolliffe 1968). This statistical method is also used for observing clusters of observations which are similar to one another using biplots (which can be denoted as separate groups). In simplified terms, PCA can be used to identify a set of similar variables which cause observations to differentiate from one another, and is therefore often used in classification techniques, to identify which variables are useful for explaining differences between observations.

3.1.3 Applications of Principal Component Analysis

PCA has been employed in biological and ecological research, including: using morphometric features extracted from bones for taxonomy, palaeontology and sex classification (Marrama and Kriwet 2017; Percy and Wijitten 2011; Lewis et al., 2014); species recognition through images obtained by camera traps (Giraldo-Zuluaga et al., 2017); species identification via acoustics (Binder and Hines 2012), and; identifying differences in diet across a species (Lowry et al., 2004). As an example of biological application, Marrama and Kriwet (2017) used PCA as part of a taxonomic classification study for identifying which features of a tooth were responsible for differentiating between different taxonomic groups of living and extant shark species. Using photogrammetry, 14 measurements plus two angles were extracted from images of teeth using image manipulation software TPSdig2.19 (Rohlf 2005), after standardisation and log-transformation these measurements were input into a PCA. The PCA identified the seven most important variables (*i.e.* features of shark teeth) for explaining 93.6% of the total variability, consequently eliminating the need for future analysis or measuring of the remaining nine variables, thus saving time in data collection. Of these variables, it was found that PC1 (76% of the variation) was predominantly associated with “*degree of slant*” *i.e.* how curved the tooth is, therefore suggesting that this feature alone is a key distinguisher between taxonomic groups (Marrama and Kriwet 2017). Following the PCA, discriminant analysis and Hotelling’s t^2 -test (Hotelling

1951) were employed to determine if the tooth measurements were useful for assigning indeterminate shark teeth to a specific taxon, and reported that these indeterminate teeth were significantly different to three of the four extant shark species, but showed non-significant separation to teeth belonging to *Brachycarcharias*, and therefore were assigned to the *Brachycarcharias* taxon (Marrama and Kriwet 2017).

3.1.4 Applications of Principal Component Analysis to marine mammal studies

The application of PCA to biological questions have been used in a number of marine mammal studies, including; detecting and classifying North Atlantic right, humpback (*Megaptera novaeangliae*), bowhead (*Balaena mysticetus*), minke (*Balaenoptera acutorostrata*) and sperm (*Physeter macrocephalus*) whales using acoustics (Binder and Hines 2012); analysing fatty acid signatures within blubber of North Atlantic orca (*Orcinus orca*; Bourque et al., 2018), and; examining the origin of high-frequency hearing in whales (Churchill et al., 2016). Research by Lowry et al., (2004) used PCA for comparing the diet of bowhead whales between three locations in the Beaufort Sea of Alaska: Barrow (western Beaufort Sea); Kaktovik (eastern Alaskan Beaufort Sea), and; Nuiqsut (central Alaskan Beaufort Sea), using examination of stomach contents. In research by Lowry et al., (2004), PCA was applied to a dataset within which the importance of 16 unique prey types were ranked for each individual whale, where importance was defined by the ratio between the volume of that prey type to the total volume of the sample. The analysis resulted in three PCs that were responsible for explaining the largest amount of variance within the dataset (48.9%), where PC1 (18% of total variability) was defined by a contrast of copepods and euphausiids in the diet of the whale (Lowry et al., 2004). These three PCs were then used as the dependent variables in a multiple regression analysis which assessed the relationships between Alaskan bowhead whale diet and: location; season; length; sex, and; year. The multiple regression found a significant difference between diet and season for all three PCs, as well as a significant effect between diet and location for PC1 (where copepods dominated the eastern Beaufort Sea and euphausiids were most frequently observed in the western Beaufort Sea; Lowry et al., 2004). Therefore, of the 16 prey types observed, the difference between the ratio of copepods and euphausiids in the stomach contents was highlighted by the PCA as being an important factor for identifying differences between bowhead whale diet at different locations and within different seasons in the Alaskan Beaufort Sea.

3.1.5 Applications of Principal Component Analysis to leopard seal studies

PCA has also been applied to studies involving leopard seals, some of which include; analysing the evolution of the pinniped jaw (Jones et al., 2013); examining differences between carnivoran tooth sharpness with reference to filter-feeding capabilities (Hocking et al., 2017); measuring characteristics

of leopard seal calls (Kreiss et al., 2013), and; determining the fatty acid composition of leopard seal blubber (Guerrero et al., 2016). Within research by Kriess et al., (2013) underwater vocalisations of leopard seals were examined between three distinct Antarctic locations (Drescher Inlet, Atka Bay and Davis Sea) during the breeding season (the austral summer), in order to investigate acoustic ecology. In this instance, PCA was used to determine if high double trill (hereafter HDT) calls could be distinguished according to the location in which they were recorded. HDTs are a broadcast call produced by both male and female leopard seals of all age groups (Rogers et al., 1996; Rogers 2007), and are composed of two short duration pules at ~3.5s each (Rogers et al., 1995) recorded in this study at each of the three locations using underwater hydrophones (Kreiss et al., 2013). The Lower and upper HDT call frequencies together with durations of HDT call part 1 and 2 and five pule repetition rate (PRR) parameters were input into the PCA which revealed that principal components 1 and 2 (PC1 and PC2) were explained by 67% of the variance within the data. It was deduced that PC1 reflected the PRR parameters while PC2 was represented by the upper frequency of call parts 1 and 2 (Kreiss et al., 2013). Using PCA, HDTs recorded at Drescher Inlet were separated to HDTs recorded at both Atka Bay and Davis Sea, with differences in characteristics of HDT calls were strongest between Drescher Inlet and Atka Bay despite being located in closest proximity (~500km; Kreiss et al., 2013). Characteristics of HDT calls recorded in Atka Bay and Davos Sea were highly similar with respect to PC1, however differentiated slightly with respect to PC2 (*i.e.* upper frequencies of both HDT call parts; Kreiss et al., 2013). Using PCA in this application, Kreiss et al., (2013) provided support for previous studies which suggest movements of leopard seals between breeding groups to avoid the occurrence of genetically isolated leopard seal populations, by showing a high degree of similarity between HDT calls recorded at all three testing sites in Antarctica.

As can be seen from these examples, PCA is a useful method that allows the user to take a large array of variables and reduce the number of these variables down to those which are important for explaining variation within a dataset, as well as for distinguishing between observations (Kreiss et al., 2013). PCA is also applicable for differentiating between features responsible for variation and features which represent background noise. As result, 'background noise' features can be discarded from further analysis, thus saving time and effort in future research. Features identified by PCA as being 'most important' be used for statistical analysis to identify relationships (Lowry et al., 2004) and to allocate observations into classification groups (Marrama and Kriwet 2017).

3.2 Aims

The aim of this chapter was to employ PCA to refine the number of variables used to categorise body condition groups (BCGs) from photographs and observe any relationship between photographic measurements of body width and pre-assigned BCGs derived in Chapter 2. Photographic measurements were also extracted to provide validation for the body condition scoring system (BCSS) discussed in the preceding Chapter 2, in addition to being applied to machine learning methods in the following Chapter 4.

Specifically, the objectives were to:

1. Use ImageJ software to extract 13 photographic measurements from images identified in Chapter 2 for meeting Photographic Quality Criteria (PQC);
2. Apply these photographic measurements to a PCA to identify the 'most important' photographic measurements for explaining the variation within the data with respect to pre-assigned BCGs;
3. Plot individual leopard seal sighting records to observe if records allocated to the same BCG form clusters, and;
4. Use ANOVA and post-hoc testing to observe whether statistically significant differences exist between the means of photographic measurements identified as 'most important' by PCA with regard to pre-assigned BCGs.

3.3 Methods

3.3.1 Photogrammetry

For all 80 sighting records in the Body Condition – Leopard Seals Database (BCLSD) one image was selected for each sighting record to be used for extraction of measurements, resulting in a sample size of 80 images. These images were chosen for meeting all criteria in photograph assessment (refer to Tables 2.1 and 2.2 in Chapter 2), whereby the full body of the left or right side profile was visible within the frame, and the leopard seal was lying in ventral recumbency, parallel to the ground and perpendicular to the camera. Independently all images were loaded into ImageJ and cropped tightly around the body of the seal. Using the straight-line tool, a line was drawn from the tip of the snout to the tip of the hind flippers representing total body length. This line was then measured in pixels. A macro called 'ten widths' designed by Krause et al., (2017) was then implemented to create 10 individual equidistant points along the line, marked red and blue in between the two points denoting the beginning and end of the line. These beginning and end points can be seen at the tip of the snout (in blue) and over the tip of the hindflippers (in red) in Figure 3.2 and were not measured but served as markers for the beginning and end of the line to generate the equidistant points. Measurements of body width were taken using the straight-line tool at each of the ten points from the ventral side to

the dorsal side of the seal (Figure 3.2). An additional three manual measurements were taken; axillary width (AW), umbilical width (UW) and the angle between the head and the neck (A; which was measured in degrees; Figure 3.2). The additional measurements of AW and UW were selected as both have been previously identified as important measurements for estimating body condition (BC) in studies that undertake manual measurements (Hofman 1975; Van den Hoff et al., 2005; Krause et al., 2017), as well as the umbilical region being identified as a morphological area of interest in Chapter 2. The 13 total photographic measurements from each image were input into an excel spreadsheet and assigned a sighting-ID relating to the folder ID of the sighting record, *i.e.* row HL001 in the database corresponded to the sighting folder “HL001 - 20140915 Aramoana, Otago - Derek Cox” (see Chapter 2 methods for details).

The 12 body width measurements (*i.e.* the ten width measurements from the ten widths macro plus AW and UW) were standardised by dividing each photographic measurement by total body length and log-transformed to reduce effects of size and ontogeny (*i.e.* the difference in size of animals occupying the image frame and difference in size of animals observed at different developmental stages; Marrama and Kriwet 2017). Standardising the measurements was an essential step as photographs of leopard seals at different developmental stages were used, as well as the pixel resolution of photographs varied between 600 and 7000 pixels, therefore standardising in this way meant that only body shape was compared between individual sightings. These new standardised widths were given the prefix ‘S’ in the database and were so named ‘S1-10’, ‘SAW’ and ‘SUW’. Standardising also enabled comparisons of measurements between individual seals. Measurement A was not standardised (and therefore remained as “A”) as this variable was recorded in degrees, therefore was already directly comparable between sighting records. Averages of all standardised photographic measurements were calculated for the 80 leopard seal sighting records and plotted to show difference in seal silhouettes belonging to each BCG (Figure 3.3). As testing normality is a fundamental requirement of PCA, all 13 photographic measurements were also tested for normality using Shapiro-Wilk tests (Shapiro and Wilk 1965). Despite standardisation and log-transformation, 6 of the 13 measurements were non-normal (S5, S7, SUW, S9, S10 and A), and therefore were not used in further analysis in this chapter.

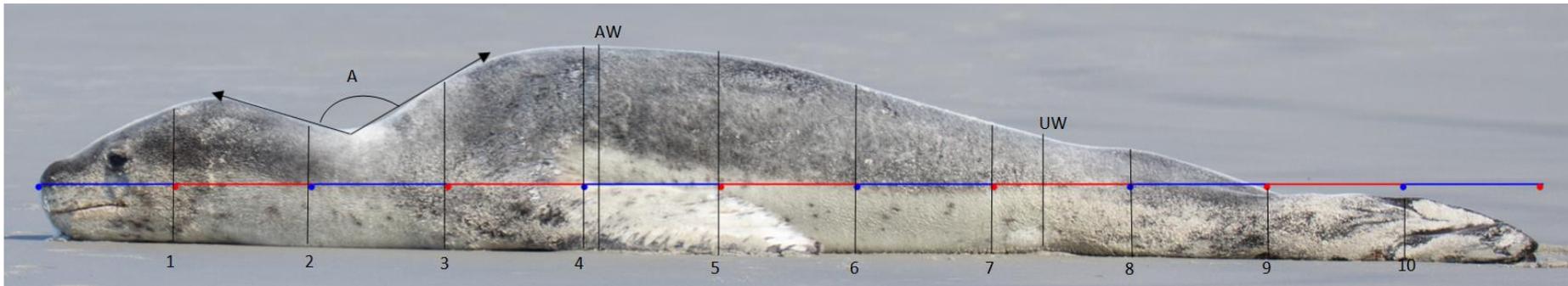


Figure 3.2. Leopard seal side profile displayed in ImageJ software exhibiting ten equidistant points along the overall length of the body (i.e. 1-10 shown on the ventral side of the seal), between the start point at the tip of the snout (in blue) and the end point at the tip of the hindflippers (in red), using the ten widths macro developed by Krause et al., (2017). Locations of the three manual measurements (A, AW and UW shown on the dorsal side of the seal; see methods for definitions) are also identified. Here, all 13 photographic measurements are shown as examples of how measurements were taken from the ten widths line. Photo by Giverny Forbes, LSO (LeopardSeals.org unpublished data).

Using standardised photographic widths from ImageJ and PCA (S1, S2, S3, SAW, S4, S6 and S8), a one-way ANOVA was applied to determine if any statistically significant differences existed between average body width measurements of Good, Moderate and Poor condition seals. Of all normally distributed photographic measurements, only width S1 returned with a statistically non-significant result ($p > 0.05$), indicating there were no differences between the different conditioned seals at this width location, and thus this variable was not tested further in this chapter. The remaining photographic measurements all produced a statistically significant result (*i.e.* $p < 0.05$ in all six instances), signifying that significant differences did exist between sighting records of leopard seals belonging to the different BCGs, at these width locations. Thereafter, Tukey multiple comparison of means (Tukey 1977) tests were employed on these photographic measurements to detect the differences between BCGs.

It was also hypothesised that measurements of A (the angle between the head and the neck, in degrees) would be larger in seals identified as being in Good BC, than those in both Moderate and Poor BC as a wider angle would represent a 'thicker' neck with increased body fat, whilst a smaller angle would signify reduced body fat thus a sharper, more angled body profile. Therefore, measurement A for all 80 records were visualised in a boxplot for each BCG (Figure 3.4). A Kruskal-Wallis (Kruskal and Wallis 1952) test showed that statistically significant differences existed between measurements of A for leopard seal sighting records belonging to the three BCGs ($p < 0.05$), therefore Wilcoxon pairwise comparison tests (Wilcoxon 1945) were also employed to determine which BCGs were statistically different with respect to A.

3.3.2 PCA

The 80 sightings from the BCLSD were applied to six PCA models for dimensionality reduction (reducing the number of photographic measurements) and for identifying relationships between the photographic measurements and BCGs assigned in Chapter 2, using a BCSS. The first PCA (PCA-1) included all normally distributed photographic measurements (S1, S2, S3, SAW, S4, S6 and S8) and a corrplot (a visual display of a correlation matrix; Wei and Simko 2017) was used to show the contribution of each variable to the principal components (Figure 3.5). The variable which contributed the least was 'dropped' from further analysis within this chapter, and this method was repeated, each time refining the number of variables. In doing this, width S8 was identified in PCA-1 as the first photographic measurement which contributed to the PCs the least, and thereby was eliminated from further analysis. Subsequent PCAs eliminated SAW (PCA-2), S4 (PCA-3), S1 (PCA-4) and S3 (PCA-5) in that order (Table 3.1).

Biplots were also drawn to visualise clustering of observations between the three pre-assigned BCGs, where points close to one another represent sighting records with similar measurement values. These biplots were useful for visualising the data as each point on the biplot represented the whole photographic measurement profile for each individual sighting record. Thus, instead of several plots comparing photographic measurements one-by-one, the biplots compared all photographic measurements for all leopard seal sighting records in one figure (Figure 3.7 and Figure 3.8).

Table 3.1. Photographic measurements used for each Principal Component Analysis (PCA). Cells where information was not applicable were denoted NA.

	Photographic Measurements Used	'Dropped' Measurement
<i>PCA-1</i>	S1, S2, S3, SAW, S4, S6, S8	S8
<i>PCA-2</i>	S1, S2, S3, SAW, S4, S6	SAW
<i>PCA-3</i>	S1, S2, S3, S4, S6	S4
<i>PCA-4</i>	S1, S2, S3, S6	S1
<i>PCA-5</i>	S2, S3, S6	S3
<i>PCA-6</i>	S2, S6	NA

All analysis was completed using R 4.0.2 (R Core Team 2020) using packages; pastecs (Grosjean et al., 2018), ggplot2 (Wickham 2016), plyr (Wickham 2011), scales (Wickham and Seidel 2020), grid (R Core Team 2020), ggbiplot (Vu 2011), FactoMineR (Le et al., 2008), factoextra (Kassambara and Mundt 2020), rlang (Henry and Wickham 2020), reshape (Wickham 2007) and corrplot (Wei and Simko 2017).

3.4 Results

3.4.1 Photogrammetry

It was evident that photographic width measurements for sighting records containing Good BC leopard seals were larger on average than records of those containing Moderate and Poor BC seals (Figure 3.3). The Moderate BC profile appears to follow the shape of Good (with smaller average width measurements) whilst the Poor BC profile has a distinctly larger upper torso (*i.e.* measurements S3, SAW and S4) leading to a slimmer umbilical region (*i.e.* measurements S5, S6, S7, SUW and S8), revealing a distinct difference in average body shape between these three BCGs.

The Tukey tests found that for photographic body width locations: S2; S3; SAW, and; S4 there was a statistically significant difference between sighting records containing Good and Moderate BC individuals, whilst for width S8 there was a statistically significant difference between sighting records containing Good and Poor BC individuals. Width S6 was the only width that achieved statistically

significant differences between two groups; Good – Moderate, and Good – Poor. These results signify that the Good BCG was most distinguished from the other two BCGs with regard to the photographic measurements of body width. Unexpectedly, for the photographic measurements that were able to differentiate between Good and Moderate BC sighting records, there wasn't a corresponding difference between Good and Poor BC sighting records, with the exception of S6. For this reason, photographic measurement S6 was identified as an important width for differentiating between seal sightings belonging to different BCGs.



Figure 3.3. Body shape of 80 New Zealand leopard seals using average (and standardised, indicated by 'S') widths, divided into body condition groups (BCG); Good (n=57), Moderate (n=15) and Poor (n=8), based upon body condition scores (BCS), at 12 positions along the body (S1 – 10, SAW and SUW).

With respect to measurement A, pairwise Wilcoxon testing identified that statistically significant differences existed between BCGs Good – Moderate, as well as Good – Poor ($p < 0.05$). This supports the hypothesis that the measurement of A was larger in seals belonging to the BCG Good, which is also visible in Figure 3.4. No significant difference existed for the size of A between Moderate and Poor BCGs.

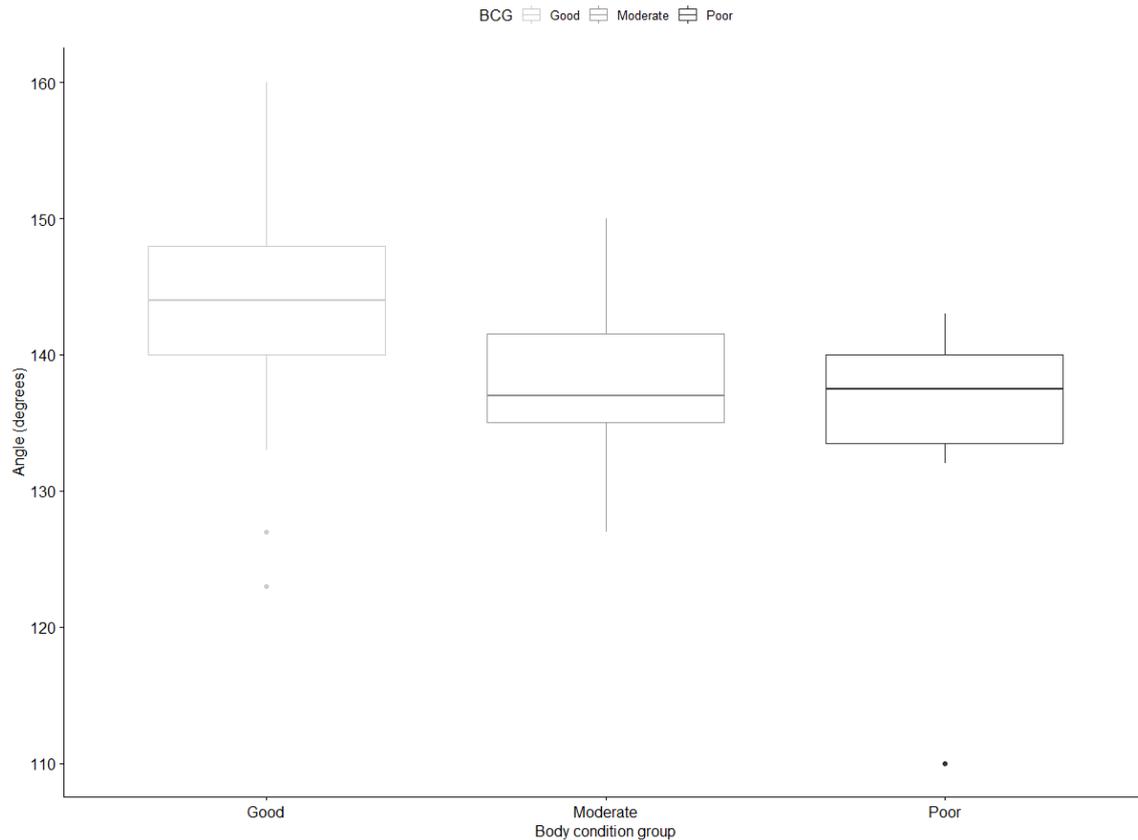


Figure 3.4. Boxplot displaying variation in measurement of the neck angle (A) in degrees between the three body condition groups (BCG) for 80 New Zealand leopard seals, where Good=1 (n=57), Moderate=2 (n=15) and Poor=3 (n=8).

3.4.2 PCA models

In total, six different PCA models were trialled. Each time the number of variables were reduced by ‘dropping’ the variable which contributed to PC1 and PC2 the least. This reduction of variables was done primarily using a corrplots (correlation matrices), illustrated in Figure 3.5 where photographic measurement S8 can be observed as being less important than the other six variables for explaining the variation in BC in both PC1 and PC2 (represented by Dim.1 and Dim.2) in PCA-1, as shown by the much smaller and paler blue circles. Similarly, Figure 3.6 (an alternative representation of the corrplot in Figure 3.5, whereby each photographic measurement is represented by a vector) shows how photographic measurement S8 contributed the least out of all seven photographic measurements within PCA-1, indicated by the shortest length, and blue-green colour of the vector within the plot, compared to the other variables symbolised by longer vectors in shades of yellow (S1, SAW, S4 and S6) and orange (S2 and S3).

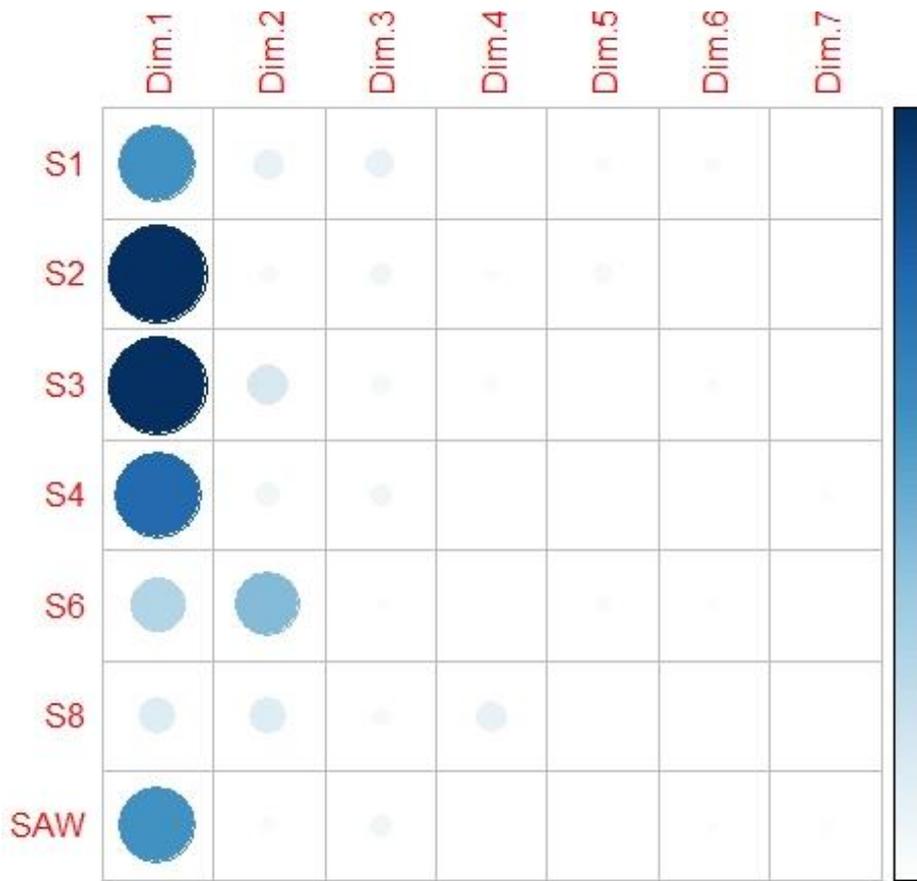


Figure 3.5. Corrplot from the first Principal Component Analysis model (PCA-1) where 'Dim' along the top represents the principal components (PC) and the photographic measurements of body width are displayed on the left. The larger and darker a circle represented the higher the variable contributed to each of the PCs (i.e Dim.1... Dim.7), as shown by the colour scale on the right.

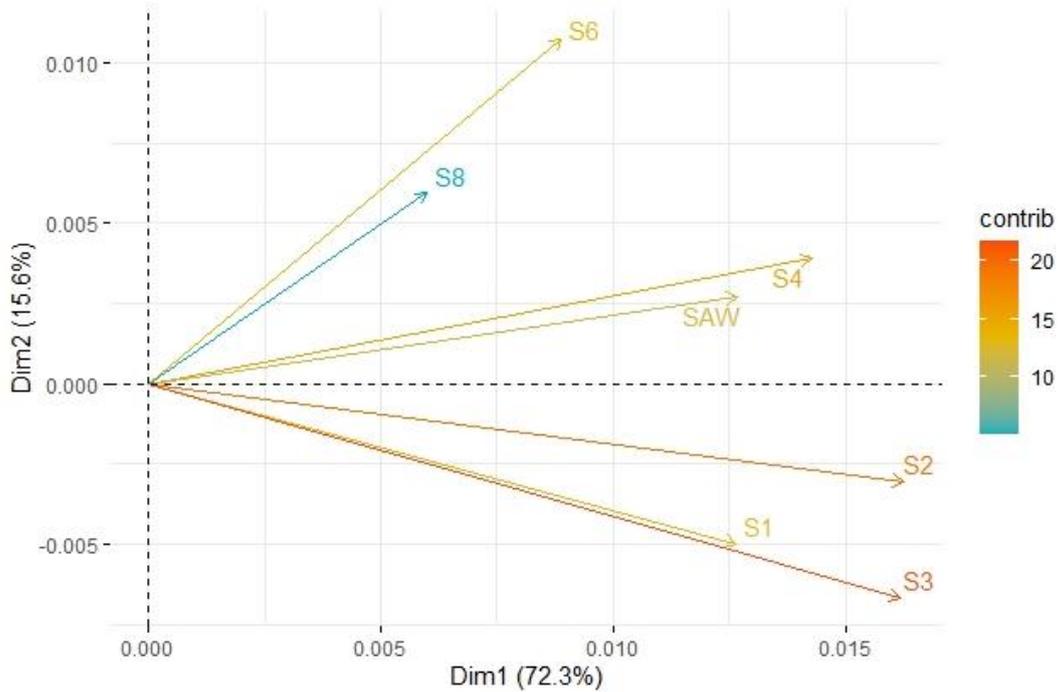


Figure 3.6. Loadings plot from the first Principal Component Analysis model (PCA-1) where 'Dim1' and 'Dim2' represent principal components (PC) PC1 and PC2. The colour (denoted by 'contrib' on the right) and the length of the vector indicate how much each of the variables contributed to the PCs.

The first biplot of PCA-1 (Figure 3.7) showed that the majority of sighting records containing Good BC individuals occupied the centre and upper right portion of the biplot, while Moderate BC records occupied the centre to upper left portion of the biplot and Poor BC records were visible across the lower portion of the biplot. There was considerable overlap of records belonging to all three BCGs within the centre, although the ellipses created by the biplot revealed some distinction between BCGs which indicated that these photographic measurements were related to pre-assigned BCGs. When the photographic measurements were refined down to three in PCA-5 (S2, S3 and S6) the biplot (Figure 3.8) showed less overlap in the centre and individual observations were more dispersed, revealing clearer distinction between BCGs than Figure 3.7. As PCA-6 began to show less grouping structure compared to previous PCAs within the biplot, the photographic measurements in PCA-5 were identified as the three most important variables for explaining variation between pre-assigned BCS in New Zealand's leopard seals.

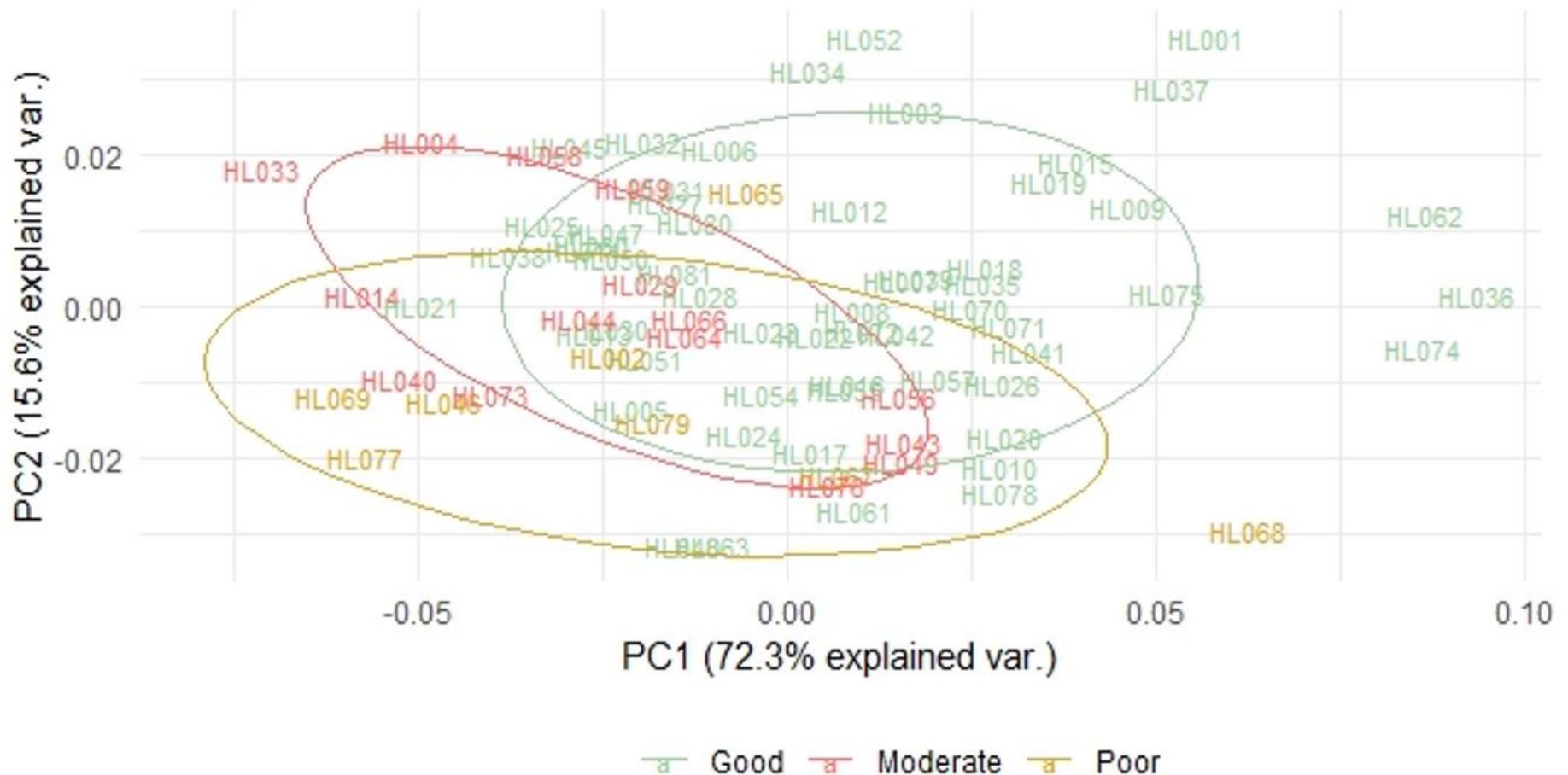


Figure 3.7. Biplot of the first Principal Component Analysis (PCA) model (PCA-1) using all normally distributed photographic measurements: S1; S2; S3; SAW; S4; S6, and; S8 with body condition groups Good, Moderate and Poor superimposed. Each point on the plot represents the measurement profile of an individual leopard seal sighting record (n=80), where values of all seven photographic measurements have been projected as one data point, onto the first two principal components (PC1 and PC2, where var = variation) using coordinates. The body condition groups are denoted by three different colours, with ellipses drawn to encompass the majority of the grouping.

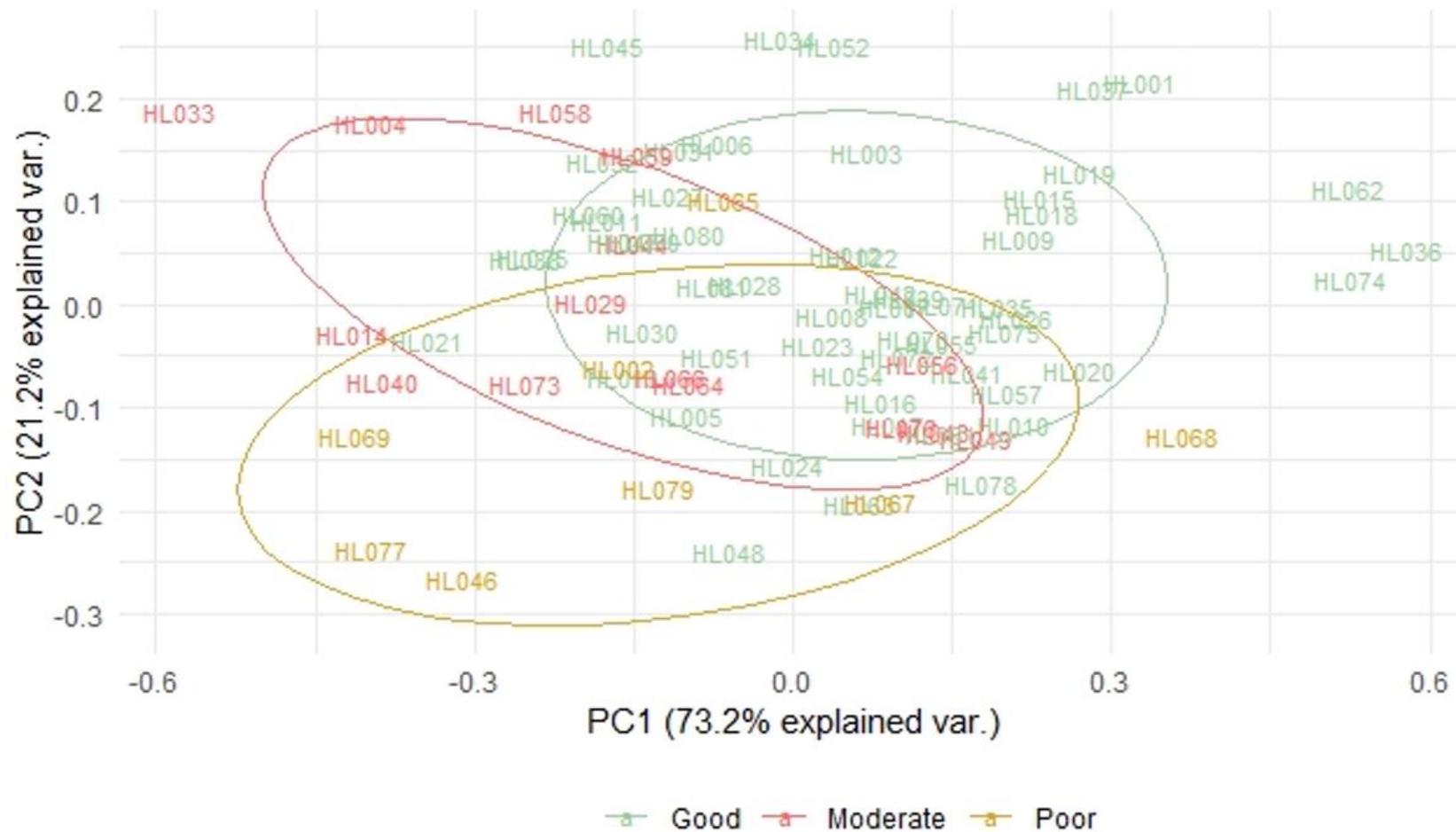


Figure 3.8. Biplot of the fifth Principal Component Analysis (PCA) model (PCA-5) using photographic measurements: S2; S3, and; S6 with body condition groups Good, Moderate and Poor superimposed. Each point on the plot represents the measurement profile of an individual leopard seal sighting record ($n=80$), where values of all three photographic measurements have been projected as one data point, onto the first two principal components (PC1 and PC2, where var = variation) using coordinates. The body condition groups are denoted by three different colours, with ellipses drawn to encompass the majority of the grouping.

3.5 Discussion

3.5.1 Photogrammetry

Measurements of body width were successfully extracted from photographic sighting records of New Zealand leopard seals in this chapter by means of photogrammetric methods. Using photographic measurements, the average body shapes of New Zealand leopard seals belonging to each BCG were compared revealing distinct differences between the profiles of the three groups, particularly for BCG Good, which showed a markedly larger profile than the BCGs Moderate and Poor (Figure 3.3). Supporting this observation, when photographic measurements were applied to ANOVA, the BCG Good was more frequently distinguished from both Moderate and Poor BCGs. It was further found that body width S6 represented the most important photographic measurement for distinguishing between BCGs as statistically significant differences between both Good – Moderate and Good – Poor sighting records were identified for this measurement, which marks the half-way point between AW and UW (Figure 3.2). Therefore, with respect to this dataset photographic measurement S6 represented the morphological region on the body where New Zealand leopard seals differed the most between the three pre-assigned BCGs and was therefore recognised as being the most important photographic width for discriminating between the BC of individual leopard seals.

The UW was recognised as an important measurement by Krause et al., (2017) in addition to the umbilical region being identified as a morphological area of interest in Chapter 2 of this manuscript. Despite this, measurement SUW was unable to be applied to PCA within this chapter for also not conforming to rules of normality. Although there may have been a number of reasons why SUW did not meet normality assumptions, a factor which may have contributed was fact that the UW was attained manually, and unlike AW (the widest area of girth) which can be taken from the crease of the foreflippers (Figure 3.2), UW is the slimmest part of the body in front of the pelvic bones and was therefore more difficult to consistently locate on the seal body (especially if the pelvic bones were not visible). As a result, the UW was likely not reliably measured in the correct place, giving inaccurate measures of photographic body width. In future applications of photogrammetry and PCA, it would be essential to determine a method to accurately characterise the exact point of UW, to ensure that it is consistently measured from precisely the same location.

Disadvantages to photogrammetry used in this chapter were that due to the citizen science nature of the project, data collection was carried out by a number of different observers utilising a variety of different camera types. The consequences of this were that vigorous photograph assessment techniques were employed (refer to Chapter 2 methods) to refine the data, and that photographic measurements of body width were maintained in pixels, rather than being converted to metric units

as seen in other photogrammetric studies (Krause et al., 2017; Christiansen et al., 2019). However, alike other research which utilised photogrammetry (Krause et al., 2017; Christiansen et al., 2019), methods utilised within this chapter offer a non-invasive and low-risk and low cost alternative to obtaining measurements of free-ranging animals, which can be applied to leopard seals in other regions as well as other pinniped species. Similar to BCS techniques outlined in Chapter 2, photogrammetric methods are suitable for application to leopard seals (as well as other pinniped species) which are often difficult to observe since they spend large proportions of their time at sea, and only come ashore for short periods of time for the purpose of resting and digestion (Gwynn 1953). Using this technique, individual seals can be opportunistically photographed by researchers as well as members of the public, and photographic data can be subject to photogrammetry making this method ideal for citizen science projects, in addition to projects with limited funding. While photogrammetric methods were validated by using actual measurements of body mass for each individual leopard seal by Krause et al., (2017), and the volume-mass equation by Christiansen et al., (2019) obtained for southern right whales were validated by individuals of known body mass, photogrammetry in this application was used as means for validation for a subjective BCSS.

3.5.2 PCA

The application of PCA within this chapter confirmed a relationship between pre-assigned BCGs (Good ($n=57$), Moderate ($n=15$) and Poor ($n=8$), derived from a BCSS in Chapter 2) and measurements extracted from 80 photographs of leopard seal sighting records in New Zealand. The use of PCA enabled the seven useable photographic measurements to be refined down three (S2, S3 and S6) that were identified as most important in explaining the variation in photographic body width measurements between leopard seal sighting records assigned to the three BCGs. Of the three most important widths, both S2 and S3 were located directly next to one another on the neck of the seal (Figure 3.2). Although measurement A was not able to be used (as these measurements did not conform to rules of normality), the identification of S2 and S3 (in addition to the results of the Wilcoxon pairwise comparisons that identified significant differences between the size of A for BCGs Good – Moderate and Good – Poor) indicated that the neck region was an important morphological area for assessing and comparing body condition in New Zealand leopard seals. These findings do not concur with previous research using photogrammetric methods for leopard seals, as Krause et al., (2017) recognised the umbilical width as being the most important morphological region for correlating with leopard seal mass. Moreover, previous studies using morphometrics to estimate mass of leopard seals have primarily focused upon axillary girth (photographic measurement AW and SAW in this study), however this photographic measurement was not found to be 'important' as SAW was

dropped from analysis after PCA-2 (Table 3.1). Estimates of mass derived in studies using axillary girth measurements however, were found to have high degrees of error (refer to Chapter 1; Hofman 1975) particularly for larger seals where estimates of mass ranged between 240 and 480kg – double the weight (van den Hoff et al., 2005).

The input data used within this chapter (13 photographic measurements) was consistent with the amount of input data from other studies employing the PCA technique e.g. 16 tooth measurements (Marrama and Kriwet 2017), 16 prey types (Lowry et al., 2004) and nine call attributes (Kreiss et al., 2013). However, when compared to these studies, the variation explained by the first two PCs (PC1 and PC2) were higher in this chapter (87.9% for PCA-1 and 94.4% for PCA-5) than the study investigating shark teeth (88%; Marrama and Kriwet 2017), and the study assessing leopard seal vocalisations (67%; Kreiss et al., 2013), as well as being substantially higher than study investigating the diet of bowhead whales (23.4%; Lowry et al., 2004). Within these studies, variables identified by PCA were used for additional analysis, including multiple regression (Lowry et al., 2004) and classification (Marrama and Kriwet 2017) which was not explored within this chapter. Thus far studies involving applications of PCA to leopard seals described within the literature have analysed; morphometric features (Jones et al., 2013), acoustics (Kreiss et al., 2013) and leopard seal blubber content (Guerrero et al., 2016), and have focused on deceased (Jones et al., 2013), as well as anaesthetised (Guerrero et al., 2016) and free-ranging individuals (Kreiss et al., 2013). The use of PCA within this chapter describes the first application of PCA combined with photogrammetric methods, as well as the first instance when PCA has been used for assessing BC of free-ranging leopard seals in New Zealand. The photographic measurements of body width extracted using photogrammetry and identified as being ‘most important’ by PCA within this chapter can be applied to machine learning techniques for classifying sighting records into BCGs.

3.6 Conclusions

The process of extracting measurements of body width from photographic sighting records of leopard seals in New Zealand has proven to be useful as these measurements were found to be related to subjective BCGs based upon a BCSS (discussed in Chapter 2). There is therefore a potential for these photographic measurements to be applied to machine learning classification methods, which could classify unique leopard seal sightings into a BCG based upon photographic measurements alone, in a similar manner to research described previously by Marrama and Kriwet (2017) using shark teeth. In this way, the BC of New Zealand leopard seals would be classified by objective measurements, rather than subjective BCS.

Chapter 4 : The application of Linear Discriminant Analysis and Random Forest Classifiers for assigning body condition groups to sighting records of New Zealand leopard seals



Leopard seal photographed resting on the south Wellington coast (New Zealand)

4.1 Introduction

Biological and ecological datasets frequently encompass highly dimensional and nonlinear data (often with missing values) where complex interactions exist between variables (Cutler et al., 2007). Analysis of these datasets to identify patterns and relationships within them is challenging using traditional statistical methods, such as generalised linear models, however, can be exposed using more sophisticated classification techniques known as machine learning (De'ath and Fabricius 2000; Cutler et al., 2007). In simplified terms, machine learning uses statistical models that are able to assign classification predictions by learning and adapting from experience, thereby improving performance without following explicit instructions (Mohri et al., 2018). In the application of supervised machine learning, 'experience' refers to a set of sample data known as training data, where data observations are correctly labelled with the classification group. During the learning process, supervised learning models use training data to develop rules which separate observations into two or more classification groups and can assign both existing and new observations into these groups based upon predictor variables (Cutler et al., 2007). The field of machine learning currently contains a number of complex statistical techniques that are suitable for analysis of complex biological and ecological datasets (Cutler et al., 2007), and has applications in: vegetation mapping (Steele 2000); modelling species distribution (Guisan and Thuiller 2005); object detection and recognition (Mohri et al., 2018); image and acoustic classification (Affonso et al., 2015; Zhong et al., 2020), and; species identification (Parson and Jones 2000; Satti et al., 2013; Begue et al., 2017). Both Linear Discriminant Analysis (hereafter LDA) and Random Forest Classifiers (hereafter RFC) discussed within this chapter are examples of such techniques, under the branch of supervised learning.

4.1.1 Linear Discriminant Analysis

LDA is a multivariate analysis of variance used to predict group membership (categorical, dependent variables) based upon a linear combination of predictor variables (interval, independent variables). Similar to that of Principal Component Analysis (PCA; outlined in Chapter 3), the LDA algorithm works to identify directions called linear discriminants (a combination of the original predictor variables) that maximises the separation between classification groups (Field et al., 2012). The method involves a set of observations whereby the values of both the predictor variables and group membership are known, which are divided into two subsets known as training and testing data. Using the training data, LDA computes the means of the predictor variables and uses these means to calculate the probability that an observation will belong to each of the classification groups (Field et al., 2012). The group with highest probability is given as the prediction. LDA can generate evaluations of accuracy called correct classification rates (CCR) by comparing the known values of group membership in the testing data, against the values of group membership as predicted by the model. CCR is given as a score between 0

and 1 which can be converted to a percentage. LDA is a parametric method and thereby data must meet the assumptions of normality and covariance homogeneity for the models to be reliable (Field et al., 2012). There are both advantages and limitations to using the LDA technique (see overview in Table 4.1) however overall, this classification method is often highly accurate and superior to most methods commonly used (Cutler et al., 2007).

An example of LDA application is research by Galimberti et al., (2018) who used this machine learning technique to assess whether southern elephant seals could be accurately categorised into age-classes based upon “*nose-metrics*” of the proboscis. Digitalised landmarks that represented both size and shape of the proboscis were applied to photographs taken of 43 individual elephant seals using photogrammetry and geometric morphometrics (a mathematical portrayal of a biological form based upon size and shape definitions; Savriama 2018), which were input into four LDA models. Results showed that classification accuracy ranged between 57 and 77% whereby the two LDA models which analysed landmarks representing proboscis size performed better than the two models that analysed landmarks representing proboscis shape (Galimberti et al., 2018). Research by Galimberti et al., (2018) in this application demonstrated that photogrammetric techniques can successfully be used on large, free-ranging and non-sedated animals within the field and therefore can be similarly applied to other species.

4.1.2 Random Forest Classifiers

RFC are a non-parametric machine learning algorithm built using classification trees to produce accurate classifications (Cutler et al., 2007). This algorithm was designed by Breiman (2001) to cope with the increasing size of modern datasets by scaling with the volume of data whilst sustaining statistical efficiency (Biau and Scornet 2016). Similar to the LDA process, RFC datasets can be split into training and testing data, whereby training data are used for the learning process and testing data are used to evaluate accuracy (*e.g.* CCR) and error of the classifier, however as a non-parametric method this data is not required to conform to rules of normality or covariance homogeneity. Within an RFC each individual classification tree is drawn using a subset of predictor variables and produces a prediction of classification (Cutler et al., 2007). The classification most frequently predicted by all trees combined is output as the RFCs overall prediction (Cutler et al., 2007). For a RFC to perform well it is crucial that each classification tree has low correlation with the other trees in the model, this low correlation between trees accounts for error between trees, *i.e.* while some of the trees will produce incorrect predictions, the majority of trees will produce correct predictions therefore the overall output of the model will be accurate (Biau and Scornet 2016). RFC calculations also produce measures both of variable importance and similarity of data points which can be applied for clustering,

infographics as well as missing value imputation (Cutler et al., 2007). Despite the recent emergence of this technique, RFCs are competitive with other machine learning methods in common use, however, have their own advantages and limitations to be considered (see overview in Table 4.1).

An example of RFCs used in ecological application is outlined in research by Rogers et al., (2017) who used RFCs as means to analyse differences in morphometric ratios amongst both male and female small-spotted catsharks (*Scyliorhinus canicula*) of both mature and immature life stages using manual measurements obtained from 101 animals caught as bycatch in commercial fisheries. Using morphometric ratios calculated from these measurements, RFC models were developed to make life-stage and sex predictions of the bycaught catsharks; one RFC model used all morphometric ratios for classification and the second RFC model focused only on the top 10 ratios (Rogers et al., 2017). The RFCs were successfully able to discriminate between catsharks of different genders and life stages, correctly predicting sex (80% (all ratios) and 79% (top 10 ratios) accuracy) and life stage for both males (80 and 86% accuracy) and females (78 and 83% accuracy; Rogers et al., 2017). Measurements obtained from 46 bycaught catsharks were also used in a Conditional Inference tree (used to highlight the morphometric ratios used for predicting life-stage; Rogers et al., 2017), used to classify life-stages of free-swimming male catsharks ($n=26$) where measurements were extracted by laser photogrammetry (the use of cameras to record projection of a laser onto an animal later analysed in image analysis software to obtain measurements; Webster et al., 2010; Rogers et al., 2017). This technique made accurate life-stage predictions of male catsharks 88% of the time using laser photogrammetric measurements (Rogers et al., 2017). As consequence of these results, Rogers et al., (2017) stated that photogrammetric techniques could replace destructive sampling of catsharks traditionally used to collect essential information on this species' life stages.

4.1.3 Comparing applications of LDA and RFC

Research by Cutler et al., (2007) compared the classification performances of both LDA and RFCs amongst two other supervised learning techniques (Logistic Regression and Classification Trees) using two ecological datasets. The first dataset looked at methods to predict the presence of invasive species in Lava Beds National Monument in California, whilst the second dataset used these techniques to predict habitat utilised by cavity-nesting bird species in the Uinta Mountains of Utah. Accuracy of the four different statistical classifiers were quantified using overall percentage correctly classified (PCC) scores. In both instances' RFC models were the highest scoring technique, particularly with regard to the Lava Beds National Monument dataset for which RFC PCC scores (95.3 and 92.6%) were over 10% higher than those of the LDA model (79.4% and 79.2%; Cutler et al., 2007). It was suggested by Cutler et al., (2007) that RFC models were expected to perform better than linear

methods including LDA, when relationships exist between the predictor variables, which is precisely what was observed in the Lava Beds NM dataset where interactions between variables were known.

In contrast, research by Hupman (2016) compared the classification accuracy of a variety of statistical models (including LDA, RFC, Shrinkage Discriminant Analysis, k-Nearest Neighbours and Naïve Bayes). This study examined the prevalence and variation of pigmentation patterns on 187 individual common dolphin (*Delphinus* sp.) dorsal fins to assess if these patterns could be used to identify unique animals (Hupman 2016). The success of each model was determined by calculating the percentage of individuals whom were correctly identified (via their dorsal fin pigmentation pattern) within the top-1, top-5 and top-10 individuals (*i.e.* the top 1, 5 or 10 individuals that the model predicted as a most likely match). In this instance, LDA outcompeted RFC in all classifications, as well as achieving the highest classification score out of all other classification techniques in both top-5 (70.9%) and top-10 (78.9%; Hupman 2016). In comparison, the RFC scored 51.3% accuracy in top-5 and 62.4% accuracy in top-10. The top-1 LDA classification rate (52.3%) scored 25.5% higher than the RFC (26.8%; Hupman 2016).

Table 4.1. Advantages and disadvantages of Linear Discriminant Analysis (LDA) and Random Forest Classifiers (RFC).

Method	Advantages	Disadvantages
Linear Discriminant Analysis (LDA)	<ul style="list-style-type: none"> ❖ Considered to be the most well-used method for data reduction (Tharwat et al., 2017). ❖ Can be used for classification as well as dimensionality reduction (Feldesman 2002; Tharwat et al., 2017). ❖ Easy, interpretable and robust classification technique (Feldesman 2002). ❖ Complex parametric method where data must conform to normality theory assumptions thereby results are objective (Feldesman 2002). ❖ Interpretations of results utilise traditional statistical techniques (e.g. <i>P</i> values, confidence intervals and ANOVA testing). 	<ul style="list-style-type: none"> ❖ Sensitive to non-normality, outliers, missing values, small sample sizes and unequal groups, any of which make results undependable (Feldesman 2002; Tharwat et al., 2017).
Random Forest Classifiers (RFC)	<ul style="list-style-type: none"> ❖ Limited parameters to tune (Biau and Scornet 2016). ❖ Utilises small sample sizes and copes with missing values (Cutler et al., 2007). ❖ Can be used when the number of predictor variables surpasses the number of observations (Cutler et al., 2007). ❖ A novel technique for determining importance of variables (Cutler et al., 2007). ❖ Able to model complex interactions between variables (Cutler et al., 2007). ❖ Allows correlated variables to be analysed (Rogers et al., 2017). ❖ Can perform a number of types of statistical analysis, <i>i.e.</i> unsupervised learning, regression and classification (Cutler et al., 2007). ❖ RFC infographics provide information as to which variables are important for classification, as well as how they influence which category a datapoint will fall into. ❖ High classification accuracy (Cutlet et al., 2007). 	<ul style="list-style-type: none"> ❖ Interpretations are unsuitable for application to traditional statistical techniques, such as <i>P</i> values, confidence intervals, or ANOVA testing (Cutler et al., 2007). ❖ The importance of predictor variables is subjective and may be a result of its interactions with other variables in the dataset (Liw and Wiener 2002).

4.2 Chapter Aims

This chapter focuses on the use of LDA and RFC to classify individual leopard seals photographed in New Zealand waters, into body condition groups (BCGs) based upon photographic measurements of body width. This chapter also highlights how LDA and RFC contrast using features of both techniques. Specifically, the objectives were to;

1. Assess the effectiveness of using LDA models to allocate predictions of BCG membership to leopard seal sighting records based on photographic measurements of body width;
2. Assess the effectiveness of using RFC models to allocate predictions of BCG membership to leopard seal sighting records based on photographic measurements of body width, and;
3. Compare both LDA and RFC models using correct classification rates (CCRs).

4.3 Methods

4.3.1 Linear Discriminant Analysis

The BCLSD (defined in Chapter 2) which contained 80 sighting records of New Zealand leopard seals was used to test the efficiency of the LDA models. Within this database each record was accompanied by 13 photographic measurements which were extracted from photographs using ImageJ; 12 body width measurements in pixels (denoted 'S1-10', 'SAW' and 'SUW') and one angle measurement in degrees (denoted as 'A'; refer to Chapter 3 methods). LDA is a parametric method and thereby the assumptions of normal distribution and homogeneity of variances had to be met prior to analysis. Using Shapiro-Wilk tests (Shapiro and Wilk 1965) all variables were standardised, log-transformed and tested for normality (refer to Chapter 3 methods). The photographic measurements in the dataset used for LDA models were those that met these assumptions, and included; S1, S2, S3, S4, SAW, S6 and S8.

The dataset was divided into training and testing datasets with an 80%/20% split, following methods outlined in previous machine learning studies (*e.g.* Franklin and Ahmen 2018; Ditria et al., 2020). Four separate LDA models were used; LDA-1, LDA-2, LDA-3 and LDA-4 (Table 4.2). There were differences in the probability that a sighting record would be assigned to Good, Moderate or Poor BCG. This is because there were substantially more sighting record containing Good BC animals than both Moderate and Poor BC animals, which created a bias within the dataset towards Good BC leopard seals. For both LDA-1 and LDA-3 the prior probabilities were automatically calculated in R 4.0.2, based upon the number of sightings assigned to each class within the training dataset and therefore were biased models. To account for this bias, the R script for LDA-2 and LDA-4 contained an additional line to control prior probabilities as equal (*i.e.* 1/3 or 33.33% chance of an individual leopard seal belonging

to either of the three BCGs; Field et al., 2012) and were therefore unbiased versions of models LDA-1 and LDA-3 (Table 4.2).

Table 4.2. Photographic measurements of body width of New Zealand leopard seals used in four Linear Discriminant Analysis (LDA) models designed to predict body condition groups (BCG).

LDA-1	All normally distributed photographic measurements; S1, S2, S3, S4, SAW, S6 and S8.
LDA-2	All normally distributed photographic measurements; S1, S2, S3, S4, SAW, S6 and S8 with a control for bias towards Good BC animals.
LDA-3	Photographic measurements identified as most important by Principal Component Analysis in Chapter 3; S2, S3 and S6.
LDA-4	Photographic measurements identified as most important by Principal Component Analysis in Chapter 3; S2, S3 and S6, with a control for bias towards Good BC animals.

LDA models can be plotted using a scatterplot which visualises how the linear discriminants (LD1 and LD2) separate observations belonging to the classification groups. An example of such a plot can be seen in Figure 4.1, which visualises model LDA-4. The co-ordinates for the plot were determined by discriminant scores, *i.e.* the score for each sighting record for each linear discriminant using the linear regression equation:

$$V_{1i} = b_0 + b_1DV_{1i} + b_2DV_{2i}$$

Where V_{1i} represents the linear discriminant (*i.e.* LD1 and LD2), and DV_{1i} and DV_{2i} represent the predictor variables (*i.e.* photographic measurements S2, S3 and S6 for LDA-4). The b -values within this equation give information about the contribution of each predictor variable to the linear discriminant known as coefficients for the linear discriminants, and b_0 is the y-intercept. For LDA-4 the coefficients for LD1 and LD2 (b -values) can be seen in Table 4.4. Details on how to plot a sighting record into a biplot are outlined in Appendix I.

All LDA models were computed in R 4.0.2 (R Core Team 2020) using packages MASS (Venables and Ripley 2002), caret (Kuhn 2020), car (Fox and Weisberg 2019) and tidyverse (Wickham et al., 2019). CCRs were recorded in a Microsoft Excel spreadsheet.

4.3.2 Random Forest Classifiers

The BCLSD (defined in Chapter 2) which contained 80 sighting records of New Zealand leopard seals was used to test the efficiency of the RFC models. Of these 13 measurements, the 12 body width measurements (S1-10, SAW and SUW) were standardised (refer to Chapter 2 methods) and the angle measurement (A) was continued in degrees. The data was partitioned into training and testing datasets with an 80%/20% split, following methods outlined in previous machine learning studies (*e.g.*

Franklin and Ahmen 2018; Ditria et al., 2020). In total, four RFC models were run (Table 4.3) and all analysis was conducted in R 4.0.2 using the package randomForest (Liaw and Wiener 2002) with reprtree (Dasgupta 2014) used for data visualisation. CCRs for all four RFC models were calculated and recorded in a Microsoft Excel file.

Table 4.3. Photographic measurements of body width of New Zealand leopard seals used in four Random Forest Classifier (RFC) models designed to predict body condition groups (BCG).

<i>RFC-1</i>	All 13 standardised photographic measurements.
<i>RFC-2</i>	All normally distributed photographic measurements; S1, S2, S3, S4, SAW, S6 and S8.
<i>RFC-3</i>	Standardised photographic measurements identified as most important by Principal Component Analysis in Chapter 3; S2, S3 and S6.
<i>RFC-4</i>	A reduced dataset ($n=38$) containing all 13 standardised photographic measurements with more equally weighted BCGs to account for bias of sighting records containing Good BC animals (Good=15, Moderate=15 and Poor=8). This reduced dataset was split into approximately 80% training ($n=31$) and 20% testing ($n=7$).

4.3.3 Repeat Sighting Records

Because the 80 sightings records of leopard seals were ‘sightings’ and not ‘known individuals’ (refer to Chapter 2) an effort was made to reduce repeats where the same seal was photographed on the same day, by different observers. These sightings were divided so that one sighting was part of the training data, and the other was included in the testing data, for both LDA and RFC. For example, two records exist in the BCLSD as;

- ❖ HL043 - 20190723 Oriental bay, Wellington – Jodie Warren
- ❖ HL058 - 20190723 Oriental bay, Wellington – Lana Young & Jodie Warren

In this instance, HL043 was included in the training dataset whilst HL058 was incorporated into the testing dataset.

4.4 Results

4.4.1 Linear Discriminant Analysis

CCRs of the four LDA models ranged between 62.5% (LDA-2) and 87.5% (LDA-4; Table 4.5). Model LDA-4, using only widths S2, S3 and S6 to classify BC achieved the highest accuracy with only two incorrect predictions (CCR=87.5%). Compared to models LDA-1 and LDA-2, where accuracy decreased by 18.75% after introducing a control for prior probabilities, the accuracy of model LDA-4 was 6.25% higher than that of model LDA-3 (Table 4.5). It is clear that while LDA-1 and LDA-3 achieved relatively high accuracy scores (both 81.25%), both models made exactly the same BCG predictions and were biased towards the BCG Good. Although CCR decreased for LDA-2 after a control for bias was introduced, the model made a wider range of BCG predictions (*i.e.* predictions of leopard seal sighting

records into both Good and Moderate BC) than model LDA-1. LDA-4 exhibited a full range of predictions, allocating sighting records of leopard seals into all three BCGs – the only model to do so of the LDA methods.

With respect to the most accurate LDA model (LDA-4) LD1 represented 86.91% of total variation and LD2 signified the remaining 13.09%. Examination of these linear discriminants (Table 4.4) revealed that LD1 was characterised by a difference between S2 and S6 with S3, where LD1 had a negative relationship with both S2 (-2.85) and S6 (-7.53) compared to a positive relationship with S3 (3.96). LD2 was strongly characterised by a positive relationship with S3 (7.82), and a slightly negative relationship with both S2 (-0.80) and S6 (-0.16; Table 4.4). Using LD1 and LD2, Figure 4.1 displays how both variants combined discriminate between the leopard seal sighting records. LD1 appears to distinguish between sighting records containing Good BC leopard seals with records that contain Moderate and Poor BC individuals, where the majority of Good observations (red) can be seen on the left side of the plot (predominantly below 0 on the x-axis), while Moderate (green) and Poor (blue) observations occur on the right side (above -1 on the x-axis). LD2 is not so easily distinguished, as values for sighting records belonging to all three BCGs are widely spread across the y-axis.

Table 4.4. Coefficients (b-values) of linear discriminants (LD1 and LD2) for each of the three photographic measurements used for the most accurate Linear Discriminant Analysis (LDA) model; LDA-4, where the coefficients give information about the relationship of the predictor variable (photographic measurement) to each linear discriminant. Negative values portray negative relationships while positive values identify positive relationships.

Photographic measurement	Coefficients for linear discriminants (b-values)	
	LD1	LD2
S2	-2.85	-0.80
S3	3.96	7.82
S6	-7.53	-0.16

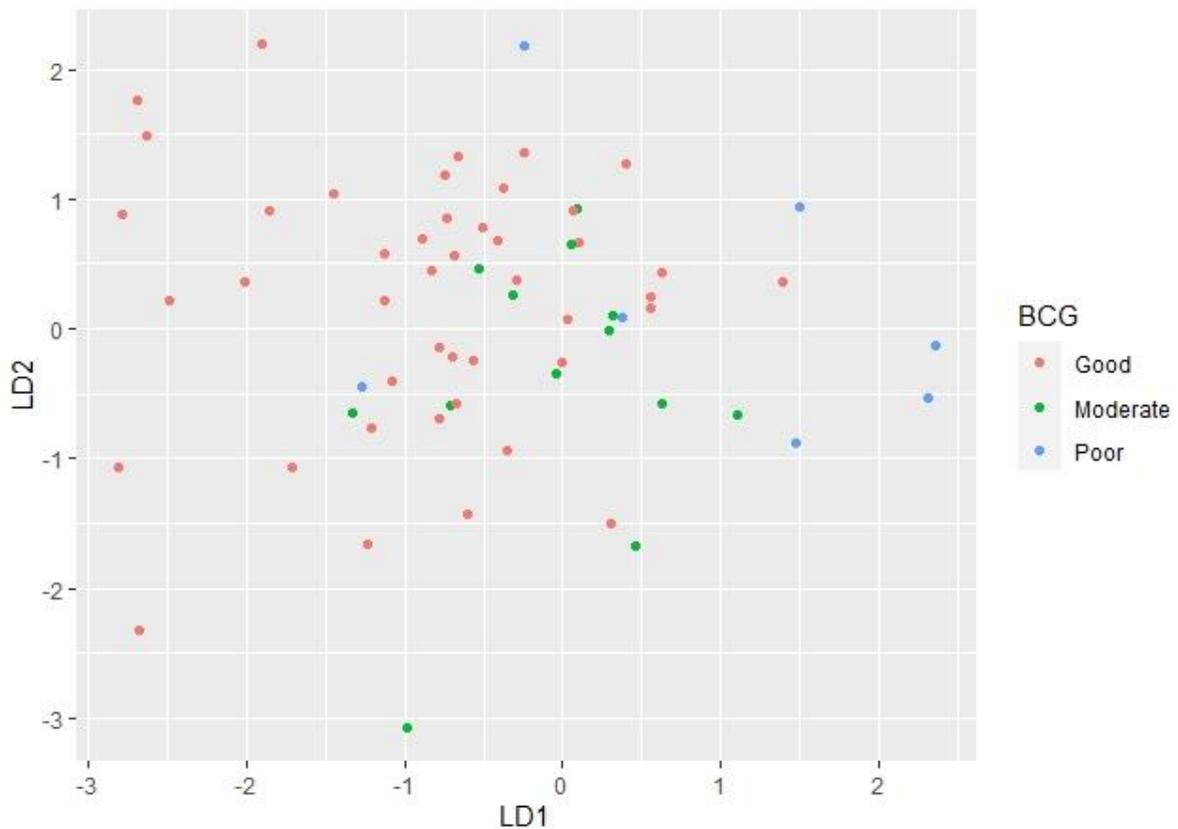


Figure 4.1. Biplot of the most accurate Linear Discriminant Analysis (LDA) model LDA-4 which used photographic measurements S2, S3 and S6 to distinguish between pre-assigned body condition groups (BCG) of sighting records containing Good (red), Moderate (green) and Poor (blue) body condition (BC) leopard seals within the training dataset (n=64). Each point on the plot represents the measurement profile of an individual leopard seal sighting plotted by the discriminant scores of the linear discriminants LD1 and LD2, that were obtained by the equation $V_{1i} = b_0 + b_1DV_{1i} + b_2DV_{2i}$.

4.4.2 Random Forest Classifiers

CCRs in the three RFC models varied between 75% (RFC-1 and RFC-3) and 87.5% (RFC-1; Table 4.5). The RFC-4 model using a subset of the original data did not yield precise results as only two out of seven seals were classified correctly (CCR=28.6%). In RFC-4 all seals were predicted to belong to the

Moderate BCG, when the actual distribution of BC was; Good=4, Moderate=2 and Poor=1. Of all RFC models RFC-1 was the most accurate with only two incorrect predictions (CCR = 87.5%). RFC-2 and RFC-3 both achieved the same level of accuracy (75%) as well as making the same predictions for all leopard seal sighting records in the testing dataset (Table 4.5). Despite the fact that models RFC-1, RFC-2 and RFC-3 all calculate predictions into Good and Moderate BCGs, no predictions of sighting records containing Poor BC individuals were made, and bias of predictions of sighting records to BCG Good was apparent.

The most accurate RF model RFC-1 (87.5% accuracy) can be visualised by the RFC tree in Figure 4.2, which contains labels atop the branches denoting the photographic measurements used as well as direction in classification. The RFC-4 tree shows 23 possible classification outcomes, whereby BCG Good was most frequently predicted ($n=16$), compared to almost equal number of predictions for Moderate ($n=4$) and Poor BC ($n=3$). The branch labels showing photographic measurements used within the RFC-4 tree revealed that of the 13 total measurements applied within this RFC model photographic widths S1 and S7 were most frequently used to separate sighting records into BCGs ($n=4$), followed by SAW ($n=3$), S6, SUW and A ($n=2$) and S4, S5, S8, S9 and S10 ($n=1$; Figure 4.2).

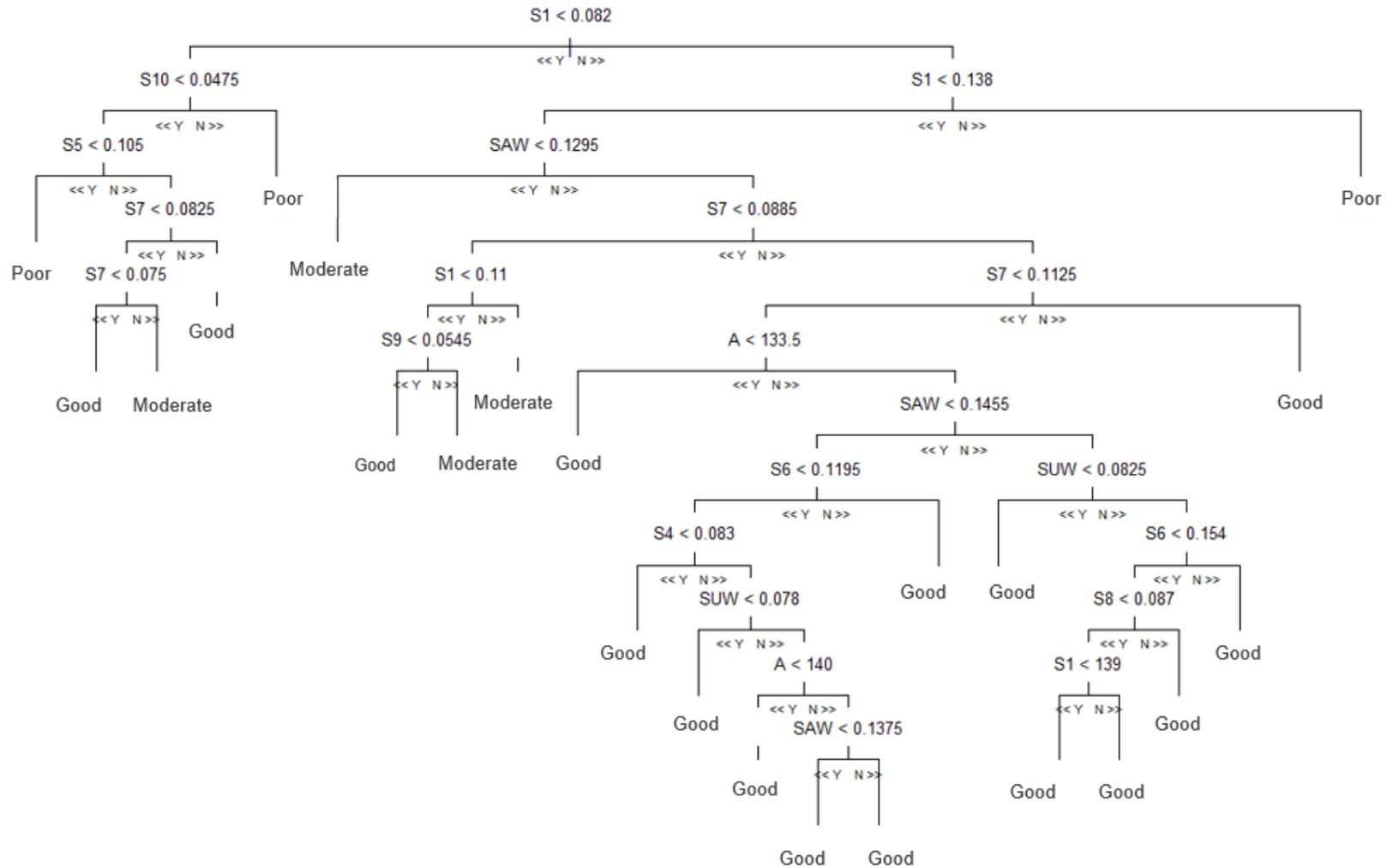


Figure 4.2. Classification tree for the most accurate Random Forest Classifier (RFC) model RFC-1, used to predict New Zealand leopard seal sighting records into one of three body condition groups (BCGs); Good, Moderate or Poor using photographic measurements. For this RFC model all 13 photographic measurements were used for classification of BCGs (see methods for further details).

Table 4.5. Correct body condition groups (BCG) and BCGs as predicted by the four Linear Discriminant Analysis (LDA; LDA-1, LDA-2, LDA-3, LDA-4) and three Random Forest Classifier (RFC; RFC-1, RFC-2, RFC-3) models (columns). The bottom row provides a score of correct classification rates (CCRs) and the * indicates the most accurate model for each method (i.e. the models that scored the most correct BC classifications). Correct predictions are highlighted in blue whilst incorrect predictions are visible in orange. Cells where information was not applicable were denoted NA.

Sighting-ID	Correct BCG	Predicted BCG						
		LDA-1	LDA-2	LDA-3	LDA-4	RFC-1	RFC-2	RFC-3
HL001	Good	Good	Good	Good	Good	Good	Good	Good
HL002	Poor	Good	Moderate	Good	Poor	Moderate	Moderate	Moderate
HL003	Good	Good	Good	Good	Good	Good	Good	Good
HL004	Moderate	Good	Moderate	Good	Moderate	Moderate	Good	Good
HL005	Good	Good	Moderate	Good	Poor	Good	Good	Good
HL006	Good	Good	Good	Good	Good	Good	Moderate	Moderate
HL007	Good	Good	Good	Good	Good	Good	Good	Good
HL010	Good	Good	Good	Good	Good	Good	Good	Good
HL012	Good	Good	Good	Good	Good	Good	Good	Good
HL016	Good	Good	Moderate	Good	Good	Good	Good	Good
HL026	Good	Good	Good	Good	Good	Good	Good	Good
HL050	Good	Good	Moderate	Good	Good	Good	Good	Good
HL052	Good	Good	Good	Good	Good	Good	Good	Good
HL054	Good	Good	Moderate	Good	Good	Good	Good	Good
HL058	Moderate	Good	Good	Good	Good	Good	Good	Good
HL072	Good	Good	Good	Good	Good	Good	Good	Good
CCR (%)	NA	81.25	62.5	81.25	87.5*	87.5*	75	75

4.4.3 Comparison of Linear Discriminant Analysis and Random Forest Classifiers

Both machine learning methods achieved the same level of accuracy at 87.5%. Average accuracy across all four models for each technique revealed LDA was slightly more accurate overall at 78.13% compared to 66.52% for RFC. The LDA model which achieved highest levels of accuracy (LDA-4) used only three photographic measurements that were identified as the ‘most important’ using PCA in Chapter 3, with equal prior probabilities to account for bias. Comparatively, the highest scoring RFC model used all 13 photographic measurements to achieve the same level of accuracy. Despite no correction for bias of sighting records containing Good BC individuals, the three RFC models in Table 4.2 all made predictions into both Good and Moderate BCGs. Comparatively, two of the four LDA models only made predictions into the BCG Good (Table 4.5). Only one model (LDA-4) accurately predicted the Poor BC sighting record (HL002 in Table 4.5) therefore was identified as the most successful machine learning model of those tested here.

4.5 Discussion

4.5.1 Linear Discriminant Analysis

The number of leopard seal sighting records used in this dataset was considered small for application to machine learning techniques and contained fewer input values than similar studies previously described (*e.g.* 22 features extracted from 856 images of 187 individual common dolphin dorsal fins in Hupman 2018 and 8251 data observations used by Cutler et al., 2007). The number of input values in this study was more in line with research by Galimberti et al., (2018), although based on a small sample size of 43 animals, three images of each individual elephant seal were used on which landmarks were drawn twice, resulting in 258 total observations. Comparatively the LDA models explored within this chapter used; 560 input values (*i.e.* seven photographic widths for each of the 80 sighting records; LDA-1 and LDA-2), or 240 input values (*i.e.* three photographic widths for each of the 80 sighting records; LDA-3 and LDA-4). The LDA technique is well-known to be sensitive to small sample sizes (Tharwat et al., 2017), however despite the small sample size used within this chapter three of the LDA models developed (LDA-1, LDA-3 and LDA-4) achieved higher classification accuracy (81.25 and 87.5%; Table 4.5) than the highest performing models in research by Galimberti et al., (2018; 77%), Cutler et al., (2007; 79.4%) and Hupman (2016; 79.8%).

In this application, LDA models which used fewer photographic measurements ($n=3$; LDA-3 and LDA-4) achieved higher CCRs than models that contained a higher number of measurements ($n=7$; LDA-1 and LDA-2; Table 4.5). Research by Galimberti et al., (2018) reported a similar result, whereby LDA classification models based upon elephant seal proboscis size using only four landmarks, outperformed LDA models that classified elephant seals based upon proboscis shape, using 23 landmarks (Galimberti et al., 2018). It was suggested that the reason for this was that a number of the landmarks represented background noise rather than proboscis shape, despite these landmarks representing a better visualisation of overall body shape from the photographs (Galimberti et al., 2018). This inference could be applied to results from this chapter, whereby models that used a higher number of photographic widths might have been less accurate at making BCG classifications due to some of the measurements representing background noise, despite exemplifying more information on overall leopard seal body shape.

4.5.2 Random Forest Classifiers

Similar to the four LDA models, the number of input values were relatively low within this chapter, with the exception of RFC-1 which contained almost double the number of input values ($n=1040$) as compared to the other RFC (and LDA) models (RFC-2 contained 560 input values *i.e.* seven photographic measurements for all 80 sighting records, while RFC-3 contained 240 input values *i.e.* three photographic measurements for all 80 sighting records and RFC-4 utilised 494 input values *i.e.* all 13 photographic measurements for 38 sighting records). This was not expected to be problematic, as RFC are known to cope well with small sample sizes (Cutler et al., 2007), however, the RFC with the smallest sample size within this chapter ($n=38$, RFC-4) performed to much lower classification accuracy than the three remaining RFCs (46.4% lower than RFC-2 and RFC-3 and 58.9% lower than RFC-1). Input values for the most accurate RFC model (RFC-1) were in line with the number of input values of the most accurate RFC models by Rogers et al., (2017; $n=1818$ and 1010) in which case classification accuracy was slightly higher (78-86%) than results within found this chapter (28.6-87.5%). The most accurate RFC (RFC-1) achieved substantially lower classification accuracy than research by Cutler et al., (2007; 92.6-95.3%), however performed to higher levels of accuracy than Hupman (2016; 26.8-62.4%).

Research by Rogers et al., (2017) shared similar aims with those within this manuscript, whereby the use of photogrammetry was used as an alternative method for obtaining essential life-stage information for a species that are difficult to assess, thus potentially replacing traditional morphometric techniques. Their methods differed however, as RFCs that were purposed for application to free-swimming animals were based upon manual measurements of deceased catsharks, placed in a “*natural position*” on a fish board for examination (Rogers et al., 2017). Moreover, live catsharks that were assessed using laser photogrammetry were kept in observation tanks, *i.e.* not examined within their natural environment, as compared to methods utilised in this chapter where all photogrammetric measurements obtained from images of leopard seals that were taken whilst the animal was at rest, in its natural environment. Thereby, data used to train RFC models in this chapter was congruent with data from which the models were tested and intended for future applications.

4.5.3 Comparison of Linear Discriminant Analysis and Random Forest Classifiers

Due to the small sample size (Tharwat et al., 2017) and presence of correlated variables (each photographic measurement was related to all other photographic measurements within the dataset; Rogers et al., 2017) RFCs were expected to perform to higher levels of accuracy than LDA models. This

was not found to be the case as both machine learning techniques achieved similar CCRs, with LDA proving slightly superior for its ability to make predictions across all three BCGs (Table 4.5). Contrary to results from the four LDA models where classification accuracy increased as the number of photographic measurements decreased, the RFC models applied within this chapter generally decreased in performance as the number of measurements decreased (Table 4.5). The RFC (RFC-3) which used the three photographic measurements identified as most important for explaining variability within the dataset using PCA (Chapter 3) achieved a CCR that was lower than the corresponding LDA models (6% for LDA-3 and 13% for LDA-4) using the same dataset. Moreover, LDA-4 allocated sighting records of New Zealand leopard seals into all three BCGs, which was the only machine learning model to do so within this chapter. Predictions made by both LDA and RFC models in this chapter elicited a bias towards sighting records containing Good BC animals (Table 4.5). When this bias was not accounted for within the R script, this was more evident as in some cases only BCG Good (LDA-1 and LDA-3; Table 4.5) or Moderate (RFC-4) were the only prediction made. As RFCs are suitable for modelling complex interactions between variables, as well as allowing variables which are correlated to be analysed (Cutler et al., 2007; Rogers et al., 2017), the full potential of this technique has not been examined within this chapter. Therefore, the development of a RFC with a larger dataset that contains sighting records of New Zealand leopard seals more evenly distributed over each of the three BCGs, could potentially produce more accurate classification predictions.

The LDA technique is sensitive to non-normality, outliers, missing values, small sample sizes and unequal groups, any of which can make results undependable (Table 4.1). This can be a disadvantage to the technique, as data must conform to rules of normality and homogeneity of variances for LDA classifications to be reliable. As consequence of this, photographic measurements identified and utilised in previous study of leopard seal body mass and condition (axillary width (AW); Hofman 1975, van den Hoff et al., 2005 and umbilical width (UW); Krause et al., 2017) could not be examined within this machine learning technique. The addition of AW and UW to LDA in this chapter may have produced LDA models with higher CCRs. In future applications of methods outlined within this chapter, LDA models would be advantageous to use with respect to data collection, as only three of the photographic measurements would need to be measured (S2, S3 and S6) as compared to the most accurate RFC which utilised all 13 measurements. The addition of more measurements also presents more opportunity for human errors in photographic measurement extraction, which again highlights the practicality of the LDA technique.

4.6 Conclusions

The combination of pre-assigned BCGs (obtained by BCS methods discussed in Chapter 2) and photographic measurements of body width (extracted using methods outlined in Chapter 3) were successfully used to train and test two supervised learning techniques; LDA and RFCs. Both methods were able to successfully classify sighting records of New Zealand leopard seals into one of the three BCGs (Good, Moderate or Poor) based upon the photographic measurements body width to similar levels of accuracy. The most successful model was identified to be LDA model; LDA-4 due to highest CCR (together with RFC-1) and ability of the model to make predictions into all three of the BCGs. In machine learning both the size and quality of this training data is fundamental to the success of the model (Mohri et al., 2018). In this particular dataset, there was a bias towards sighting records containing Good BC leopard seals, and although attempts were made to account for this, bias was still evident in the classification predictions of all models. To alleviate the effect of this bias an increased number of sighting records containing both Moderate and Poor BC individuals should be incorporated into the training data, which would improve the learning ability of the models regarding these classification groups. As consequence the classification models have the potential to increase in accuracy and generalise better when presented with new data in future applications of these machine learning techniques.

Chapter 5 : The application of Artificial Neural Networks for assigning body condition groups to sighting records of New Zealand leopard seals



Leopard seal photographed resting on Owhiro Bay beach (New Zealand)

5.1 Introduction

5.1.1 Artificial Neural Networks

Artificial Neural Networks (hereafter ANNs) have proven to be a highly competitive machine learning technique with many applications to real-world problems as compared to more traditional methods of data analysis, which are usually based upon statistical modelling (Prieto et al., 2016). ANNs have been widely used in environmental, medicinal and engineering sciences often providing highly accurate results and out-competing existing machine learning methods (Krogh 2008). This novel machine learning technique has been very successful in image classification (Affonso et al., 2015), face recognition (Taigman et al., 2014) and object detection (Ren et al., 2015). ANNs were created to simulate the learning process of the human brain, modelling the interconnected network of neurons which communicate to one another by transferring electrical signals that are received by dendrites and passed forwards along the axon via synapses (Jian et al., 1996; Priddy and Keller 2005; Krogh 2008). Simply put, a neuron can be viewed as a switch that receives an input from neighbouring neurons and is either excited (activated) or inhibited (inactivated) by the sum of these inputs (McCulloch and Pitts 1943). In a similar fashion, an ANN can be created on a computer from model neurons (otherwise known as a threshold unit), that are interconnected and applied with specific algorithms which emulate the learning process (Krogh 2008). A threshold unit receives an input from one or more neighbouring units, weighs them and calculates the sum of all the input weights. The type of response is determined by the sum of the weights and biases of the input (represented by arrows in Figure 5.1 and Figure 5.2), *i.e.* whether it is higher or lower than the threshold for the neuron to activate; if the total weight of inputs is above the threshold then the unit output is one, whereas if the sum of the input weights is below the threshold then the unit output is zero (Priddy and Keller 2005; Krogh 2008). Positive weights are associated with activated inputs whilst negative weights correspond with inhibited inputs (Jian et al., 1996).

ANNs are built using three main types of layers; input, hidden and output, where the inputs are the predictors (the dataset), the output is the outcome (or classification), and the hidden layer(s) contain units which transform the input data into a form that the output can interpret and classify (Priddy and Keller 2005) and are divided into; feed-forward (most common) and feedback networks (Jian et al., 1996). In a feed-forward network each layer supplies information directly to the next, in one direction and generates one set of output values (Jian et al., 1996). Comparatively, within a feedback neural network an input is repeatedly cycled through the network and an output is computed each time to reduce output error (Zhang 2000). Depending upon the number of units in the output layer, an ANN can be binary (one unit in the output layer, *i.e.* the output will fall into one of two categories; Figure

5.1), or multiclass (multiple units in the output layer to account for multiple classification predictions; Figure 5.2).

Similar to other machine learning methods, ANNs use datasets subdivided into training and testing for classification purposes, where training data are utilised during the learning phase, and test data is used to determine if the ANN can turn input data into the specific anticipated outputs. There are three main types of learning used in ANNs: supervised learning; unsupervised learning, and; hybrid learning (Jian et al., 1996). As discussed in Chapter 4, supervised learning is the type of learning when each input in the neural network is also provided with the correct output. Before training of an ANN using supervised learning, weights and biases (the 'learnable' parameters of a neural network) are all set to random values, which are tuned during training to facilitate learning enabling the network to generate outputs which match (or as close as possible to) the correct output (Jian et al., 1996; Krogh 2008). In the case of unsupervised learning, the neural network is not provided with a correct output for each input, and therefore the network works to identify correlations and patterns within the data, and hybrid learning is a combination of both supervised and unsupervised, whereby the values for weights are determined through a combination of known outputs and underlying data structure (Jian et al., 1996).

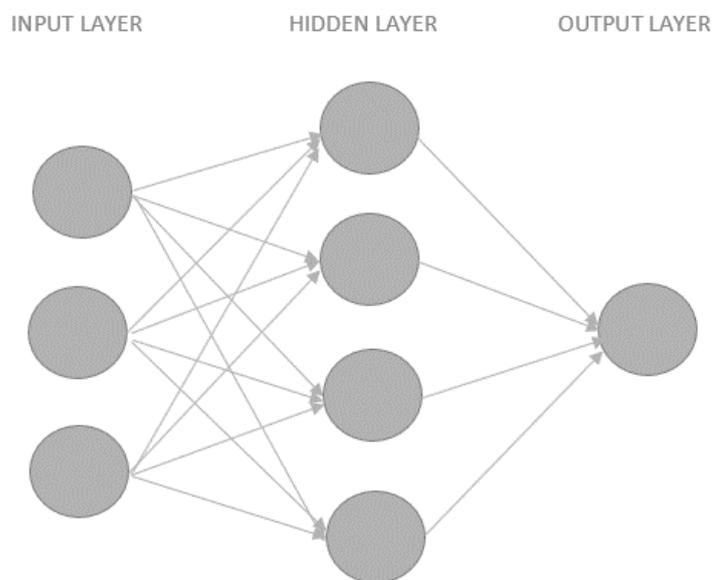


Figure 5.1. A template of a basic, binary, feed-forward Artificial Neural Network (ANN; adapted from Krogh 2008) with an input layer containing three input units, one hidden layer containing four input units and one output layer containing one input unit. Each arrow represents weight and biases, which pass an input forwards to the next layer with the associated weight.

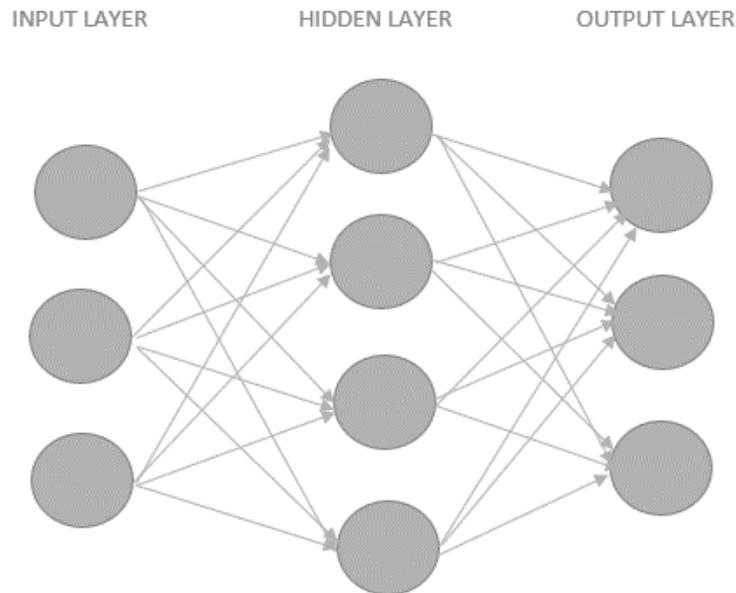


Figure 5.2. A template of a basic, multi-class, feed-forward Artificial Neural Network (ANN; adapted from Krogh 2008) with an input layer containing three input units, one hidden layer containing four input units and one output layer containing three input units. Each arrow represents weight and biases, which pass an input forwards to the next layer with the associated weight.

5.1.2 Applications of Artificial Neural Networks

In ecological research, ANNs have frequently been applied to species classification and detection using images and acoustic data. For example, ANNs were used alongside discriminant function analysis (otherwise known as DFA or LDA; described in Chapter 4) by Parsons and Jones (2000) to classify 12 species of bat using temporal and spectral features from recordings of 641 echolocation calls. They found that ANNs were highly effective in classifying calls to species level (correct classification rates (CCR) of 87% and 85%) and performed better than LDA (CCR of 79% and 81%) at an equivalent level. ANNs have also been adapted in plant recognition. Identification of plant species is fundamental for plant management in agriculture, as well as for medicinal uses in botany (Satti et al., 2013). Research by Satti et al., (2013) offers a simple and efficient plant identification technique making use of digital images using computer vision. The process is comprised of three phases; pre-processing, feature extraction and classification using 1907 samples of 33 plant species. Pre-processing involved image enhancement, which was essential prior to feature extraction which derives features for identification using the colour and shape of the leaf. These features were divided between; colour, shape and geometric, morphological and tooth features, and were used as predictors for two different classifier methods; ANN and Euclidean classifier (also known as K-Nearest Neighbours; KNN) which is based upon distance, where datapoints that are close to one another in a feature space are assumed to be similar (Satti et al., 2013). Both networks were run, and calculations of accuracy were reported as

85.9% (KNN) and 93.3% (ANN), however it was additionally stated that while the ANN outperformed KNN on larger datasets, the KNN was more efficient using smaller datasets (Satti et al., 2013). In a similar study of medical plant recognition by Begue et al., (2017) ANNs were compared against four other machine learning classifiers: Random Forest (defined in Chapter 4); Support Vector Machine (an algorithm that works to find a hyperplane which distinguishes between data points; Noble 2006); Naïve Bayes (an algorithm based upon Bayes' rule which assumes that all data observations are independent of one another and uses learning to estimate the probability that a datapoint belongs to a certain class; McCallum and Nigam 1998), and; KNN. Using a database of 720 leaves, it was surmised that the Random Forest classifier achieved the highest levels of accuracy (90.1%) followed by the ANN (88.2%), Support Vector Machine (87.4%), Naïve Bayes (84.3%) and finally KNN (82.5%; Begue et al., 2017). It was however stated by Begue et al., (2017) that the potential of the ANN had not been fully exploited as result of resource constraints, they concluded by specifying that higher levels of accuracy could potentially be achieved with this machine learning method.

5.1.3 Application of Artificial Neural Networks to marine mammal studies

More recently, ANNs have been used by Zhong et al., (2020) as part of a study that estimated seasonal occurrence of a declining population of beluga whales (*Delphinapterus leucas*) in Cook Inlet of Alaska. A total of 13 000 hours of acoustic recordings of underwater noise were collected using passive acoustic moorings and beluga vocalisations were extracted by semi-automated tonal detectors within PAMGuard software (<https://www.pamguard.org/>) which were manually validated through analysis of spectrograms (Zhong et al., 2020). The recordings were labelled as true (beluga vocalisations present) or false (beluga vocalisations not present) detections and data were input into a convolutional neural network (a version of a ANN primarily used in computer vision; Yamashita et al., 2018) in with a 49%:21%:30% training, validation and testing split. The model achieved high accuracy scores in both training (96.57%) and testing (92.26%) datasets thereby proving an effective model for identifying presence and absence of beluga whale calls (Zhong et al., 2020). It was reported by the authors that this method could save up to 93% of the highly time-consuming and subjective validation process (Zhong et al., 2020).

As can be seen from these examples, ANNs are an effective and versatile classification method that are able to handle large amounts of complex data, and regularly outperform other machine learning methods to high levels of accuracy. A comparison of ANNs with other machine learning techniques can be seen in Table 5.1.

Table 5.1. Comparison of machine learning techniques.

Method	Advantages	Disadvantages
<i>Euclidean Classifier (KNN)</i>	<ul style="list-style-type: none"> ❖ An extremely simple yet effective classifier that requires no training (Satti et al., 2013). ❖ Can be utilised on weak processors (Satti et al., 2013). 	<ul style="list-style-type: none"> ❖ Testing time increases with the size of the training data which limits the scalability (Satti et al., 2013).
<i>Random Forest Classifier (RFC)</i>	<ul style="list-style-type: none"> ❖ Limited parameters to tune (Biau and Scornet 2016). ❖ Utilises small sample sizes and copes with missing values (Cutler et al., 2007). ❖ Can be used when the number of predictor variables surpasses the number of observations (Cutler et al., 2007). ❖ RFC infographics provide information as to which variables are important for classification, as well as how they influence which category a datapoint will fall into. 	<ul style="list-style-type: none"> ❖ Interpretations are subjective, the RFC method is not suitable for application to traditional statistical techniques, such as <i>P</i> values, confidence intervals, or ANOVA testing (Cutler et al., 2007). ❖ The importance of predictor variables is subjective and may be a result of its interactions with other variables in the dataset (Liaw and Wiener 2002).
<i>Linear Discriminant Analysis (LDA)</i>	<ul style="list-style-type: none"> ❖ Easy, interpretable and robust classification technique (Feldesman 2002). ❖ Complex parametric method where data must conform to normality theory assumptions thereby results are objective (Feldesman 2002). ❖ Interpretations of results utilise traditional statistical techniques (e.g. <i>P</i> values and ANOVA testing). 	<ul style="list-style-type: none"> ❖ Data must conform to normality theory assumptions for LDA to be accurate - LDA is sensitive to non-normality, outliers, missing values, small sample sizes and unequal groups, any of which make results undependable (Feldesman 2002).
<i>Support Vector Machine (SVM)</i>	<ul style="list-style-type: none"> ❖ Simple and interpretable classification technique. ❖ Can be applied to large datasets (Noble 2006). ❖ Does not assume normal distribution of data (Noble 2006). 	<ul style="list-style-type: none"> ❖ Assumes that both training and testing data are drawn from the same distribution (Noble 2006). ❖ Only suitable for binary classifications (<i>i.e.</i> distinguishing between two classes; Nobel 2006).
<i>Naïve Bayes (NB)</i>	<ul style="list-style-type: none"> ❖ Very simple classification method that often performs to high levels of accuracy (McCallum and Nigam 1998). ❖ Due to assumption of independent observations, parameters of each class are learnt separately which simplifies the learning process (McCallum and Nigam 1998). 	<ul style="list-style-type: none"> ❖ Makes the assumption that all data observations are independent of one another with regard to the classes being tested, which is often not a realistic assumption for real-life applications (McCallum and Nigam 1998).
<i>Artificial Neural Network (ANN)</i>	<ul style="list-style-type: none"> ❖ Highly accurate and adaptable to varied and complicated datasets (Krogh 2008; Boguki et al., 2018). ❖ Provide state-of-the-art solutions to interesting problems (Krogh 2008). ❖ Able to identify patterns in data that are not clearly visible to human researchers or identifiable using other statistical methods (Masters 1993). ❖ Can be used when non-linear relationships exist between variables (Bradshaw et al., 2012). 	<ul style="list-style-type: none"> ❖ Require vast datasets to facilitate learning, <i>i.e.</i> it is not uncommon for neural network datasets to contain many thousands of observations (Boguki et al., 2018).

5.2 Chapter aims

The aim of this chapter was to evaluate the usability of ANN classification models to both distinguish between and classify sighting records containing Good, Moderate and Poor body condition (BC) leopard seals by comparing silhouettes extracted from side-profile photographs obtained by citizen science data. Specifically, the objectives were to:

1. Extract silhouettes from photographs of New Zealand leopard seals using ImageJ software and create new image files which correspond to the sighting-ID in the Body Condition – Leopard Seals Database (BCLSD; defined in Chapter 2);
2. Compile the new image files of leopard seal silhouettes into a dataset and use these data to design two ANN models; one with class imbalance not accounted for (*i.e.* the higher proportion of Good BC records) and one with class imbalance accounted for (*i.e.* editing the R script to mediate the effect of the high proportion of Good BC records);
3. Compare the accuracy of both ANN models using CCRs (as previously mentioned in Chapters 3 and 4), and;
4. Test the effectiveness of both ANN models when new data is applied using silhouettes extracted from photographs of leopard seals in New Zealand (*i.e.* not part of the original BCLSD).

5.3 Methods

5.3.1 Extraction pixel data from leopard seal silhouettes

Using data from the BCLSD (defined in Chapter 2) a total of 80 photographs of leopard seals were individually loaded into ImageJ and cropped tightly around the body to eliminate background noise caused by the surrounding environment (*e.g.* sand, stones and seaweeds). Using the polygon tool, the body shape outline of each seal was traced, and the polygon was printed onto a new, blank white image file of 8000x2000 pixels (since the width of photographs varied greatly between 600 and 7000 pixels, this resolution size enabled all seal polygons to fit into the new image). In the new image file, the polygon was scaled to fit the dimensions of the new image file and inverted to create a solid black silhouette (Figure 5.3). The silhouettes were saved into a new folder and named according to the sighting-ID that was allocated, *e.g.* HL001.jpg (refer to Chapter 2).

Using packages `keras` (Allaire and Chollet 2020) and `EImage` (Pau et al., 2010) the 80 silhouette images were read into R 4.0.2 (R Core Team 2020) and stored as a list. Each image was resized to 800x200 pixels to reduce processing time and then reshaped into an array comprised of 160 000-pixel values between 0 and 1 which signified lightness, with higher numbers representing darker pixels. All image arrays were combined as a data frame whereby the rows represented the image names (or

sighting-ID's) and the columns represented the pixel values from 1 to 160 000. The data frame was split into 80% training (rows 17-80) and 20% testing (rows 1-16), following methods outlined in previous machine learning studies (e.g. Franklin and Ahmen 2018; Ditria et al., 2020) and the two new data frames named 'train' and 'test' were exported as CSV files, to be used as input data for the ANNs.

5.3.2 ANN-1 and ANN-2

These CSV files were then uploaded into a Kaggle online notebook that allows the use of specialist R scripts using R 3.6.3 (R Core Team 2020). The data frames train and test were imported and converted into matrices and were defined as trainx and testx. Two unique lists of labels (trainy and testy) were assigned to each matrix that represented the pre-allocated body condition group (BCG). All analysis was completed using packages tensorflow (Allaire and Tang 2020) and keras (Allaire and Chollet 2020). Using both packages, a simple three-layer, multi-class, feedforward ANN was designed (ANN-1), compiled and fit with trainx and trainy using a 20% validation split, where the validation data was derived from a set of samples from the training data and used to provide an evaluation of model fit whilst tuning model parameters in the learning phase. The success of the model was determined by values of accuracy and loss for both trainx and testx. Accuracy was scored as a percentage by computing how many occasions the model predicted the correct BCGs for each sighting-ID using CCRs, and was calculated when model parameters were fixed, and no further learning occurred. In contrast, loss was an indicator of how well a model behaved after each iteration and was a summation of errors made in both the training and validation datasets (*i.e.* not a percentage). It was expected that as iterations increased, accuracy of the model would also increase, whilst loss decreased. In addition to this, a list of predicted values for all leopard seal sighting-ID's was obtained for trainx and testx.

As previously described in Chapters 3, 4 and 5, the balance of classes within the BCLSD ($n=80$) was not equal, as there were substantially more sighting records containing Good condition leopard seals than Moderate or Poor condition seals (Figure 5.4). For this reason, a second ANN model (ANN-2) was designed to account for class imbalance by introducing a line of code that controlled class weights. Within the BCLSD, a seal sighting was 7.125 times more likely to be classified as Good than Poor condition, and 3.8 times more likely to be classified as Good than Moderate BC. Considering this, ANN-2 contained an additional line of R script to edit class weights using these values. Calculations of loss and accuracy were determined for both trainx and testx and a list of predicted BCGs were computed for each sighting record.

Both ANN-1 and ANN-2 were plotted to display how accuracy and loss varied through each epoch, where both models contained 100 epochs each. An epoch is the process of passing the entire set of

training data through the network. As the number of epochs in a model increases, the number of times the model parameters (weights) are altered to facilitate learning. Values for loss and accuracy are calculated after each epoch and these values can be plotted to illustrate how well the model fits to the data. ANN-1 can be visualised in Figure 5.5 while ANN-2 was visualised in Figure 5.6.

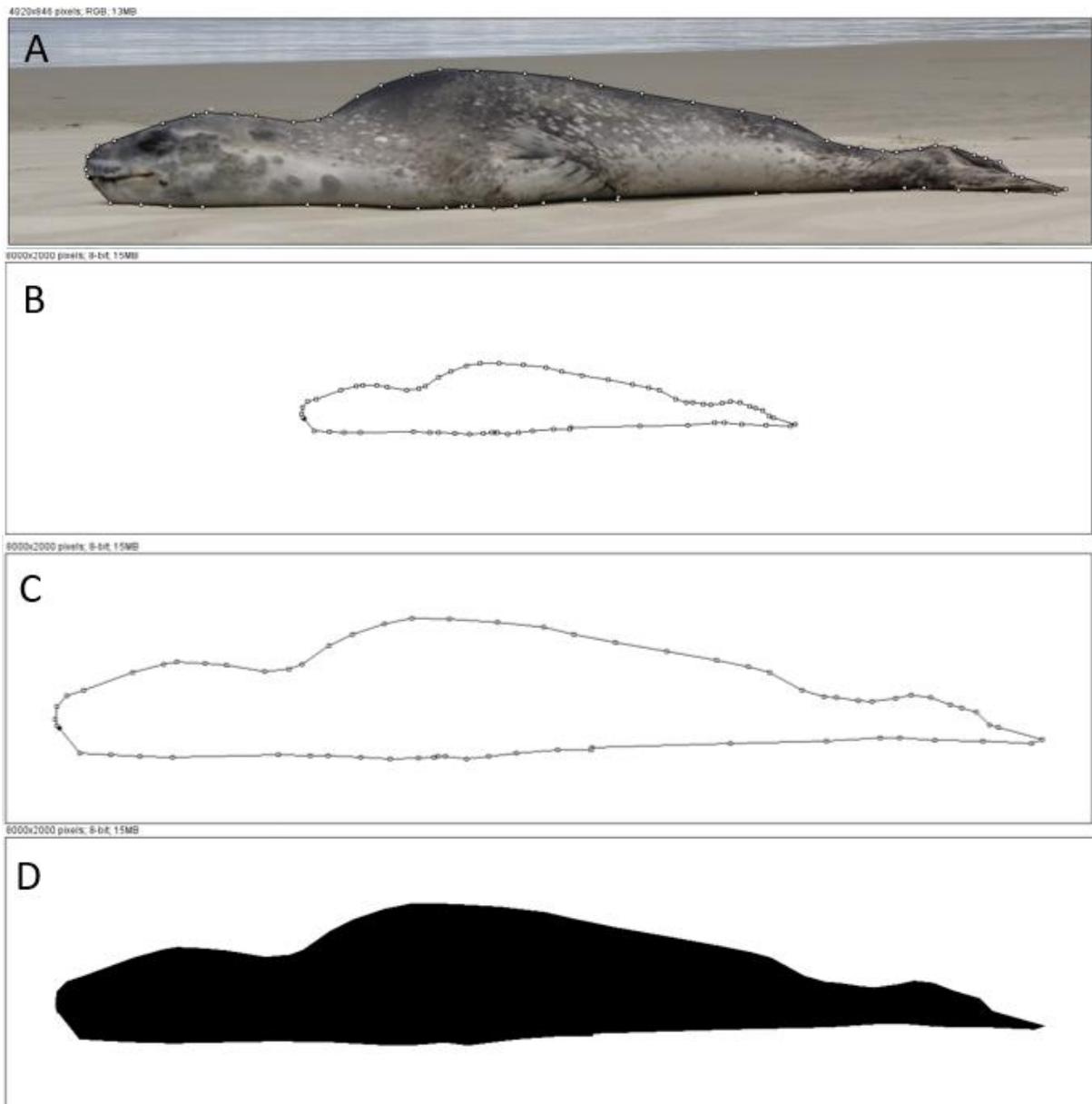


Figure 5.3. Leopard seal silhouette being extracted using ImageJ software where: A is the original image with the traced polygon; B is the polygon printed onto the new blank image; C is the scaled polygon to fit the image, and; D is the polygon inverted into a silhouette.

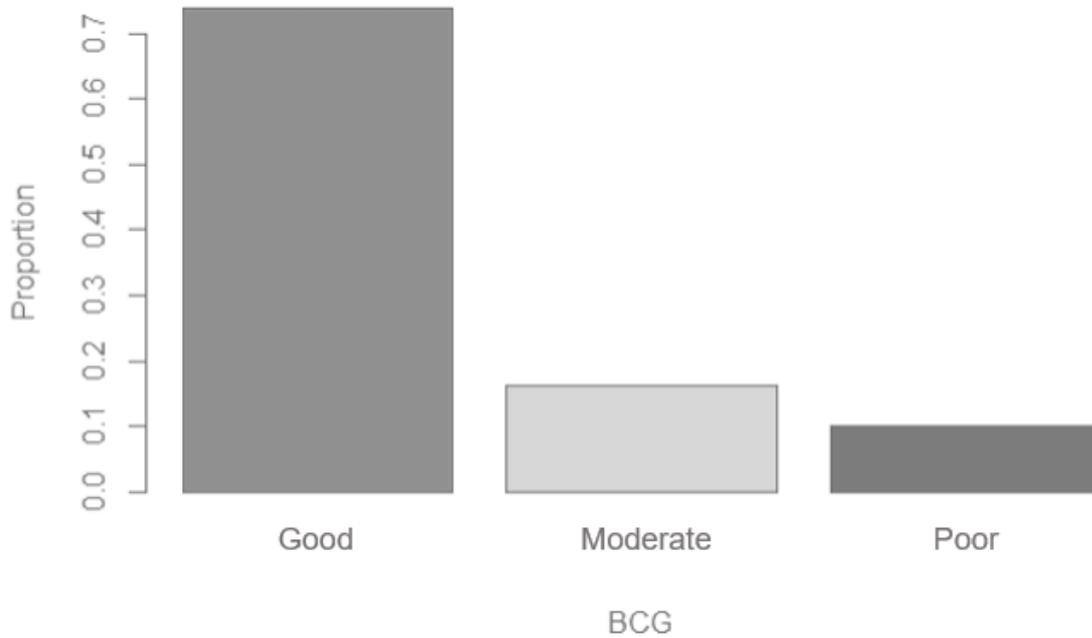


Figure 5.4. Proportion New Zealand leopard seals with a Good ($n=57$), Moderate ($n=15$) and Poor ($n=8$) body condition within the Artificial Neural Network (ANN) dataset ($n=80$).

5.3.3 Added-Data

An additional dataset (named “Added-Data”) was compiled of 48 New Zealand leopard seal sighting records for which a BCG had not been allocated, as this data was comprised of photographs that did not meet the Photograph Assessment Criteria (PAC) or Photograph Quality Criteria (PQC; defined in Chapter 2). These types of images were used to determine how models ANN-1 and ANN-2 would perform with new, less-desirable data, and to generate an estimate of BC composition within the New Zealand leopard seal population. Silhouettes were extracted in ImageJ, transformed in R 4.0.2 and uploaded into the Kaggle R Notebook using the same methods previously outlined. The data was converted into a matrix run through both ANN models where a list of predicted BCGs was derived. These BCGs were compared to one another, and to BCG estimates made by the author following methods outlined in Chapter 2.

5.4 Results

5.4.1 ANN-1 and ANN-2

Accuracy scores for training (trainx) and testing (testx) data within both ANN models varied considerably with ANN-2 scoring substantially higher than ANN-1 in both situations (Table 5.1). In addition to higher accuracy scores, values for loss in ANN-2 were also slightly lower than those for ANN-1 (Table 5.2). Confusion matrices were used to compare correct BCGs and predicted BCGs made by both ANN models in the original data (Table 5.3). Predictions made by ANN-1 in trainx fell under all

three BCG and into two BCG (Good and Moderate) within testx. Predictions made by ANN-2 also fell across all BCG within trainx, however all predictions in testx were projected to be in the Good BCG. Despite ANN-2 achieving higher levels of accuracy, when applied to testing data the ANN failed to accurately classify the two sighting records of Moderate BC leopard seals, and a single sighting record containing a Poor BC leopard seal, compared to test data for ANN-1 which did correctly classify the two records of Moderate BC seals (Table 5.3).

Table 5.2. Values for accuracy and loss for Artificial Neural Network (ANN) Models ANN-1 and ANN-2 which classified body condition groups (BCG) of sighting records containing images of New Zealand leopard seals, within both training data (trainx) and testing data (testx).

Model	Accuracy		Loss	
	trainx (%)	testx (%)	trainx	testx
ANN-1	81.25	50.00	3.75	4.81
ANN-2	90.63	81.25	3.21	3.55

Table 5.3. Four confusion matrices showing predicted vs correct body condition groups (BCG) for both Artificial Neural Network models (ANN-1 and ANN-2) which classified BCGs of sighting records containing images of New Zealand leopard seals, within the data subsets trainx and testx.

Correct BCG			
ANN-1 (trainx)			
Predicted BCG	Good	Moderate	Poor
Good	39	1	2
Moderate	7	10	2
Poor	0	0	3
ANN-1 (testx)			
	Good	Moderate	Poor
Good	6	0	1
Moderate	7	2	0
Poor	0	0	0
ANN-2 (trainx)			
	Good	Moderate	Poor
Good	45	1	3
Moderate	1	10	1
Poor	0	0	3
ANN-2 (testx)			
	Good	Moderate	Poor
Good	13	2	1
Moderate	0	0	0
Poor	0	0	0

Loss values within the optimisation learning curves show how the training and validation data for both ANN-1 (Figure 5.5) and ANN-2 (Figure 5.6) decreased to the point of stability. When looking at the performance learning curves, for model ANN-1 (Figure 5.5) accuracy values within the training and validation datasets are consistently high, rarely dipping below 50%. The accuracy for training data steadily increases through the 100 epochs, however validation data slightly decreases during the learning phase. For ANN-2 (Figure 5.6) accuracy values for both training and validation fluctuate largely between high and low values through each epoch, however within both datasets a gradual increase in model accuracy is evident.

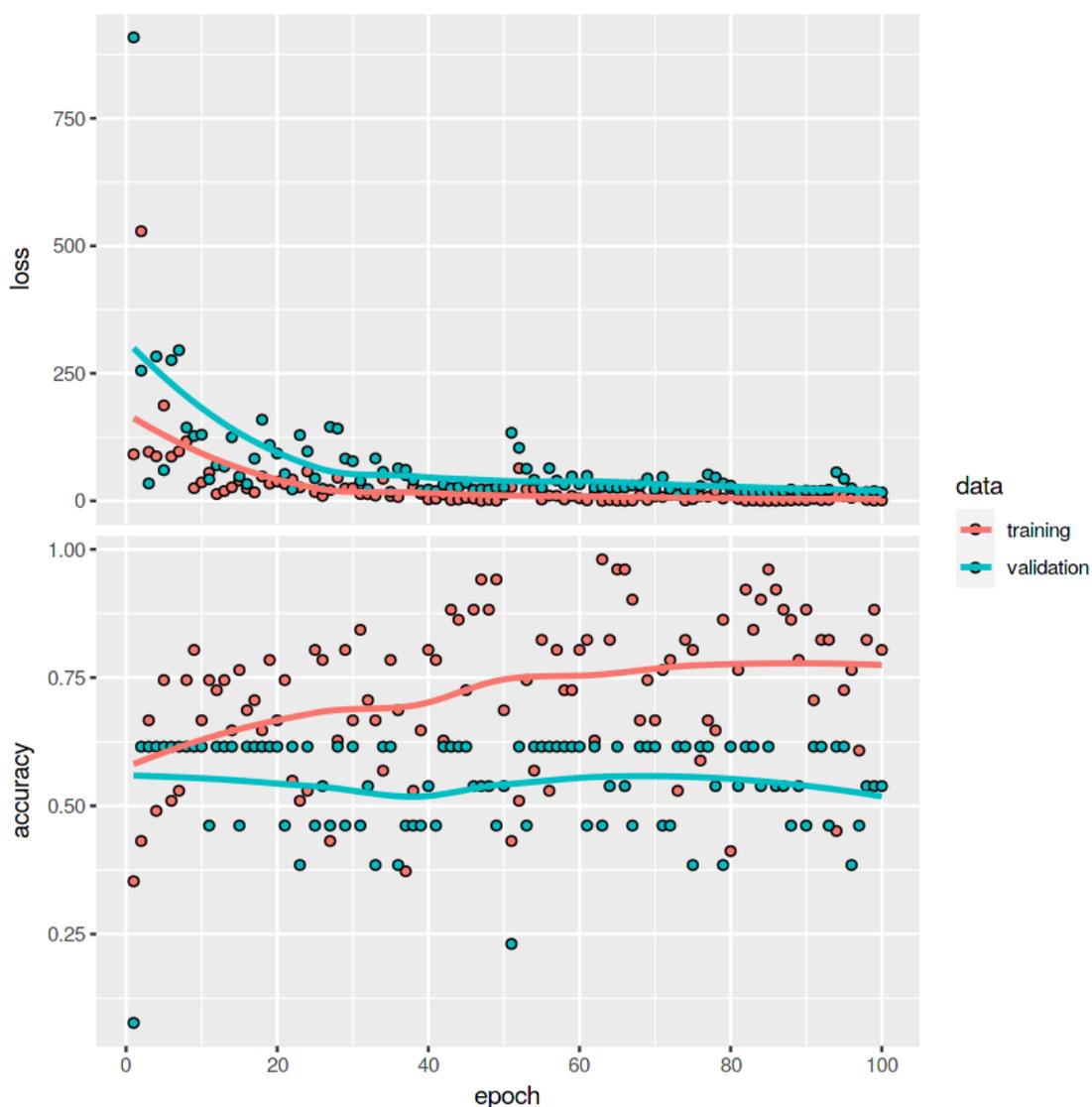


Figure 5.5. Optimisation learning curves (loss, i.e. sum of errors made) and performance learning curves (accuracy, in %) for training and validation data within Artificial Neural Network (ANN) model ANN-1 which classified body condition groups (BCGs) of sighting records containing photographs of New Zealand leopard seals. The plots visualise training (red points) and validation (blue points) data, where after 100 epochs loss=3.75 and accuracy=81.25% for training data.

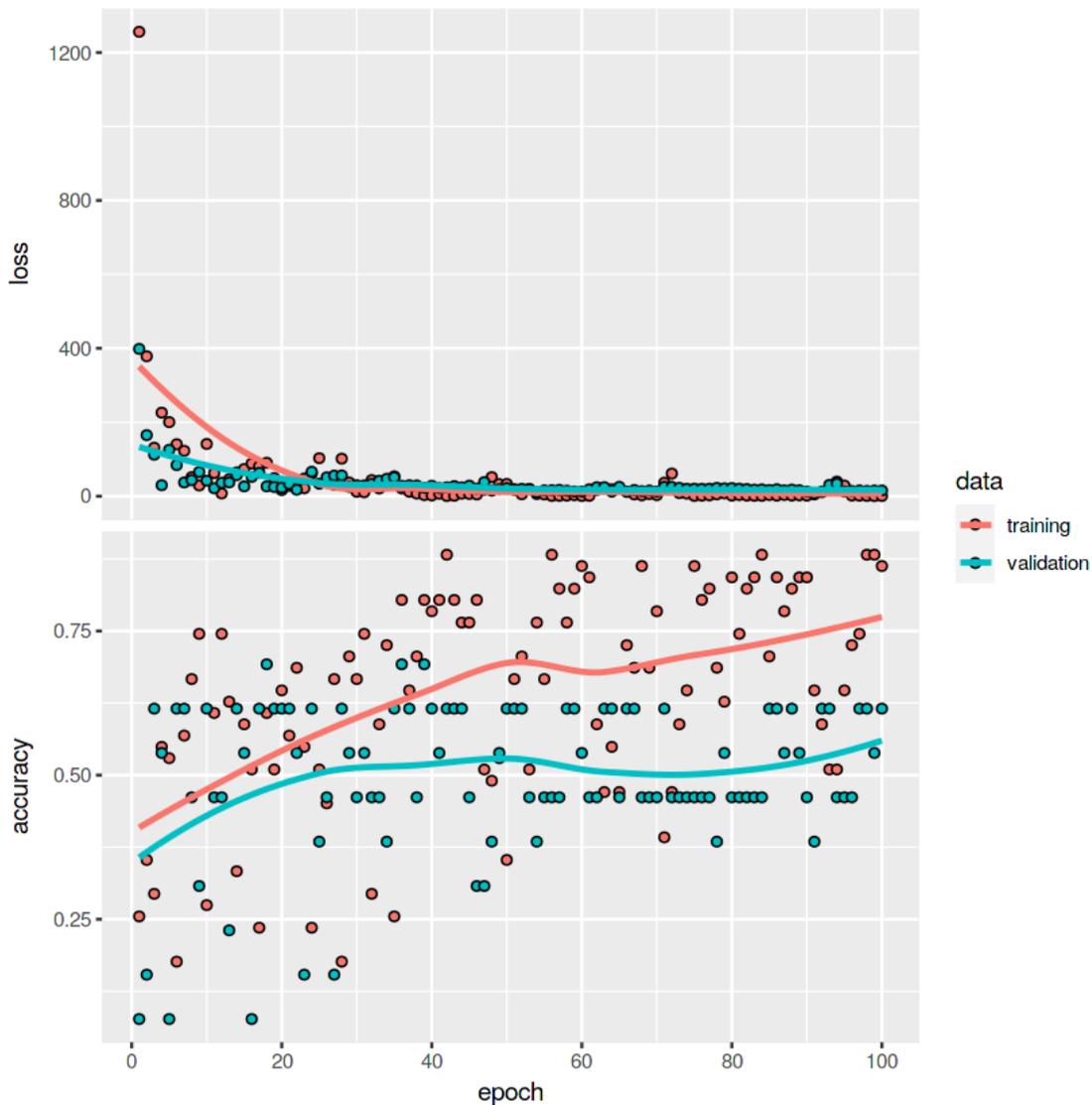


Figure 5.6. Optimisation learning curves (loss, i.e. sum of errors made) and performance learning curves (accuracy, in %) for training and validation data within Artificial Neural Network (ANN) model ANN-2 which classified body condition groups (BCGs) of sighting records containing photographs of New Zealand leopard seals. The plots visualise training (red points) and validation (blue points) data, where after 100 epochs loss=3.21 and accuracy=90.625% for training data.

5.4.2 Added-Data

BCG estimates for New Zealand leopard seal sighting records within Added-Data ($n=48$) were most frequently classified as Good (68.75%; $n=33$), followed by Moderate (20.83%; $n=10$) and Poor BC (10.42%; $n=5$). These estimates were compared with predictions made by models ANN-1 and ANN-2 (Figure 5.5). Compared to BCG estimates ANN-1 was accurate 58.34% of the time (28 out of 48 correct predictions) and predicted BCG of seals to be either BCG Good ($n=30$) or Moderate BC ($n=18$), none were predicted to be as BCG Poor. ANN-2 performed better with 75% accuracy (36 out of 48 correct predictions) and predicted the majority of the dataset to be a BGC Good ($n=41$), with BCGs Moderate

($n=6$) and Poor ($n=1$) also being allocated. Within both models BCG Good was the most frequent classification predicted and on 36 occasions (75% of the time) both models produced the same BCG predictions. Regarding Moderate BC predictions, ANN-1 was marginally better with 6 out of 10 correct predictions compared to 5 out of 10 for ANN-2. Although ANN-2 predicted one Poor BC classification, the leopard seal within this sighting record was estimated by the author to belong to the BCG Good (point 20 on the x-axis of Figure 5.7 which corresponds to sighting record HL102).

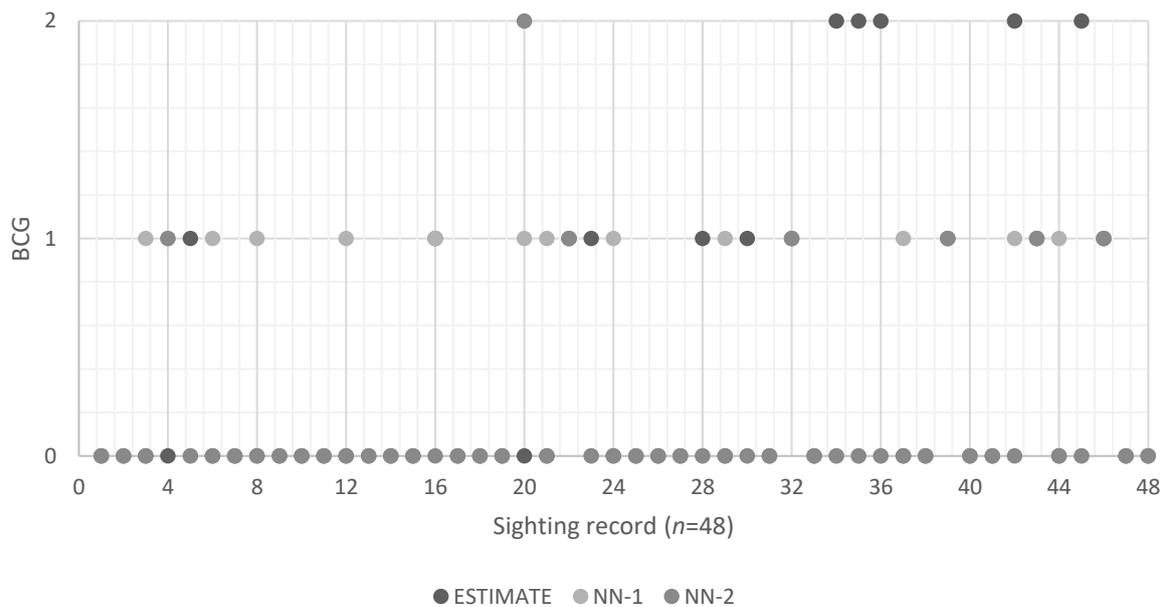


Figure 5.7. Comparison between body condition group (BCG) estimates denoted ‘ESTIMATE’ (dark grey) and BCG predictions made by the two Artificial Neural Network (ANN) models ANN-1 (light grey) and ANN-2 (mid-grey) for sighting records containing photographs of New Zealand leopard seals in the database ‘Added-Data’, where the x-axis signifies the 48 sighting records within this dataset. For the purpose of this plot BCGs were modified whereby Good=0, Moderate=1 and Poor=2.

The proportion of sighting records allocated into each of the three BCGs for Added-Data were consistent with the proportion of sighting records allocated from the BCLSD (Table 5.4). Combining these figures gives an average estimate of the BC composition of New Zealand leopard seals based upon 128 sighting records (*i.e.* BCLSD=80 and Added-Data=48), whereby 69.63% of sightings contain Good BC individuals, while 19.79% of sighting records contain Moderate BC individuals and only 10.21% of sighting records represent animals in Poor BC (‘Average’ in Table 5.4).

Table 5.4. Proportion (%) of New Zealand leopard seal sighting records containing Good, Moderate and Poor condition individuals from both; Added-Data (n=48) and the Body Condition – Leopard Seals Database (BCLSD; n=80), in addition to both databases combined (Average).

BCG	Added-Data (%)	BCLSD (%)	Average (%)
Good	68.75	71.25	69.63
Moderate	20.83	18.75	19.79
Poor	10.42	10.00	10.21

5.5 Discussion

5.5.1 ANN-1 and ANN-2

The process of obtaining silhouettes from photographs of New Zealand leopard seal sighting records enabled an in-depth analysis of BC where each leopard seal image was characterised by 160 000 data points. These data points represented 160 000 pixels (800x200-pixel images for each silhouette) where each pixel value ranged between 0 and 1 which signified lightness. These pixel values were then successfully used to train and test two ANN models; ANN-1 and ANN-2 in classifying sighting records of New Zealand leopard seals belonging to pre-assigned BCGs using data from the BCLSD. Of both ANNs the bias-corrected model ANN-2 achieved the highest number of correct BCG classifications with accuracy scores of 90.63% (training data) and 81.25% (testing data), compared to 81.25% (training data) and 50% (testing data) for ANN-1 (Table 5.2). As can be derived from the accuracy scores, both models performed well on the training data, however ANN-2 out-performed ANN-1 on the testing data by 30.25%. However, with respect to BCG predictions it was ANN-1 that performed slightly better, as a wider range of BCG predictions were made for testing data than for ANN-2 where all BCG predictions were calculated to be Good. It was therefore difficult to determine if the class imbalance correction was effective, in this instance.

Optimisation learning curves for both ANN-1 and ANN-2 revealed good fit models, *i.e.* the models were a good fit to the data, as opposed to overfit (which occurs when a model 'memorises' training data, and incorporates 'background noise' as part of the learning process; Jabbar and Khan 2015) or underfit (the instance where the model was not able to determine the underlying data structure; Jabbar and Khan 2015). Overfitting is generally characterised on optimisation learning curves when the loss values for training data decrease with experience, but the validation loss values decrease initially before increasing again (Jabbar and Khan 2015), compared to underfitting which is visualised by a flat, horizontal line for training data. ANN models that are overfit tend to perform well on training data, but poorly on testing data, however ANN models that are underfit do not perform well on either training or testing data (Jabbar and Khan 2015). Looking at the accuracy figures for both models reinforced that ANN-2 was a good fit model as it performed well on both sets of data, however ANN-

1 may have experienced an element of overfitting as accuracy levels were much lower in the testing dataset.

ANNs designed within this chapter under-performed compared to other ANN models in the literature. Classification accuracy in the testing data for both models was lower (Table 5.2) compared to 85% and 87% for classifying species of bat (Parson and Jones 2000), 93.3% (Satti et al., 2013) and 88.2% (Begue et al., 2017) for classifying species of plants as well as 92.26% for classifying echolocation calls of beluga whales (Zhong et al., 2020). As was acknowledged in Chapter 4, one reason for this could be that the number of pixels (*i.e.* 160 000 pixels per sighting record) were too high and signified background noise rather than true variation within the data (Galimberti et al., 2018). Reducing the dimensions of the silhouettes by half to a smaller image size (400x100-pixels) might produce ANN models with more accurate prediction and generalisation capabilities.

Research by both Satti et al., (2013) and Begue et al., (2017) employed similar techniques of image processing to those outlined within this chapter, where photographs of leaves were converted to binary images and features were extracted to be used as input values for ANNs (*e.g.* length, perimeter and area; Satti et al., 2013). In these applications the sample sizes of the ANN input data were substantially higher the small sample size of 80 sighting records used for ANN models within this chapter, derived from 1907 (Satti et al., 2013) and 720 leaf samples (Begue et al., 2017).

5.5.2 Added-Data

Although it was not possible to predict accurate body condition scores (BCS) and therefore BCG for sighting records within Added-Data, from examining the photographs visually, estimates could be made based upon any bony protrusions that were visible (Figure 5.8 and Figure 5.9). For example, Figure 5.8 shows leopard seal with sighting-ID HL127. Using models ANN-1 and ANN-2 this seal was predicted to belong to BCG Good, however looking at the profile the bony protrusions sagittal crest, spine and pelvis can all be seen despite the low resolution. For this reason, this individual was classed as a BCS of '3' and therefore would be at least Moderate BC. However, due to the low resolution it is not possible to determine if any further bony protrusions were visible, looking at the overall silhouette it would be suspected that this seal was actually Poor BC. Another example is shown in Figure 5.9 which shows leopard seal with sighting-ID HL125. Using models ANN-1 and ANN-2 this individual was predicted to be in BCG Moderate BC. Although the image is in low resolution both the spine and pelvis can clearly be seen, however the shape of the seal is much more robust than the one seen in Figure 5.8, suggesting this could be an adequate classification by the model.

Regarding the singular Poor classification made by ANN-2 (sighting-ID HL102 signified by number 20 on the x-axis of Figure 5.7), this same record was predicted to be BCG Moderate by ANN-1, indicating that the BCG estimate made by the author was incorrect and this leopard seal was in reduced condition than appeared from the photographs. In spite of these inaccuracies, overall BCG estimates made from unsuitable photographs that did not meet the Photograph Assessment and Photograph Quality Criteria (PAC and PQC) were in fact in line with BCGs from the BCLSD (Table 5.3), suggesting that despite the photographs being of reduced quality accurate predictions of BC can be made. It was difficult to assess how the ANN models performed when presented with new data as a significant portion of the Added-Data ($n=33$) were in fact estimated to be Good BC individuals. ANN-1 appeared to over-classify records into Moderate BC and failed to classify any records as Poor BC. Comparatively, predictions made by ANN-2 under-classified Moderate BC records and mis-classified a Poor BC record. In order to observe this more effectively the BCLSD should incorporate an increased number of sighting records that contain Moderate and Poor BC leopard seals which would enable the ANNs to more accurately learn and recognise the parameters of these classification groups.



Figure 5.8. Sighting-ID HL127 from Added-Data.



Figure 5.9. Sighting-ID HL125 from Added-Data.

5.6 Conclusions

The use of silhouettes extracted from New Zealand leopard seal sighting records allowed for an extremely detailed comparison of body shape using ANNs, characterised by differences in pixel data representing shape. The ANN model that accounted for bias of Good BC sighting records within the BCLDS, ANN-2, was denoted the 'best' model due to higher levels of accuracy not only using data from

the BCLSD but within the dataset Added-Data also. From comparing the BCG predictions to estimates made by the author it appeared that model ANN-2 was able to generalise to some degree when presented with new data, however, still exhibited bias towards the BCG Good in the testing data. From the success of this ANN, a similar ANN model using a larger dataset with an increased number of sighting records containing both Moderate and Poor BC leopard seals, and reduced pixel dimensions to reduce the influence background noise, would likely improve the ability of the ANN to produce accurate BCG classifications, as well as being able to generalise when presented with new data.

Chapter 6 : Discussion of statistical and machine learning techniques used to assess body condition in New Zealand leopard seals



Leopard seal photographed trying to get some rest in front of the boat sheds on Oriental Bay(New Zealand)

6.1 Overview

Despite the widespread distribution of leopard seals in the southern hemisphere, knowledge of their northern occupation outside of Antarctica is very limited. While photo-identification (photo-ID) can be used to examine various aspects of their ecology (including health), photo-ID catalogues for leopard seals are scarce and have rarely been created within these northern regions. To date, a number of existing studies have examined leopard seal body condition (BC) both within Antarctica and these northern regions using a range of methods that include: subjective visual observations (Rounsevell and Pemberton 1994; Gray et al., 2009; Aguayo-Lobo et al., 2011; Bester et al., 2017); calculating blubber volume via morphometrics and values of blubber depth (Kuhn et al., 2006), and; estimating seal mass using manual and photographic measurements (Hofman 1975; Van den Hoff et al., 2005; Krause et al., 2017). The use of different techniques across these studies means that baseline information for leopard seal health and condition parameters is lacking, and findings from such studies cannot be directly compared to one another (*i.e.* in making comparisons of leopard seal health across different regions). Without this baseline information, it is difficult to monitor their health condition throughout their range and over time.

The paucity of baseline data on their health condition is also apparent within New Zealand waters, currently only one study has been done to examine their BC within this region. This research focussed on using a visual, qualitative body condition scoring system (BCSS) based on presence/absence of bony protrusions in photographs (Hupman et al., 2019). In New Zealand, no quantitative analysis regarding BC has been published thus far, such data is needed to provide baseline information on their health status to provide effective conservation management initiatives within the region, as well as for comparing BC between individuals and documenting change in BC over time. This study provides the first assessment of BC for leopard seals in New Zealand waters that combines qualitative and quantitative data in the form of body condition groups (BCG) derived using body condition scoring (BCS) and measurements obtained using photogrammetry. This study also provides the first assessment of BC on New Zealand leopard seals using machine learning techniques; Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Random Forest Classifiers (RFC) and analysis of body shape via Artificial Neural Networks (ANN).

Here, key research findings from each chapter are summarised and the significance and contribution of this research is discussed. In addition, the limitations of this study are highlighted followed by suggestions of methodological improvements and questions for future research. Lastly, the management considerations for leopard seals in New Zealand are outlined.

6.2 Chapter 2: Body condition scoring as a non-invasive method for examining body condition in sighting records of free-ranging New Zealand leopard seals from photographs

6.2.1 Key research findings

Results from the BCS identified that the majority of leopard seal sighting records extracted from the New Zealand Leopard Seals Photo Library (NZLSPL) contained images of Good BC individuals (71.25%) compared to Moderate (18.75%) and Poor (10%) BC. Of these Good BC records ($n=57$), 71.9% ($n=41$) were allocated a BCS of 0 (aka no bony protrusions were visible), which represented over half of all sighting records, suggesting that the majority of leopard seals observed within the region are in predominantly very good health. This reinforced conclusions by Hupman et al., (2019). When bony protrusions were visible ($n=39$), the pelvic bones were most frequently observed appearing in 33 sighting records (84.6% of the time) which indicated that this may be an area of interest in their morphology when assessing condition of leopard seals. This finding was congruent with research by Krause et al., (2017) who reported that the same morphological area (the umbilical region) showed a strong correlation with leopard seal mass.

6.2.2 Contribution and significance

Understanding leopard seal health parameters is vital for understanding health of the population as well as health of the marine ecosystem they occupy as upper trophic marine predators (Krause et al., 2017). One way health of an individual can be assessed is by using BCS, which is a well-practiced technique in the fields of agriculture and veterinary science, and has more recently been applied to both captive and wild animals as a hands-off approach for assessing the amount of subcutaneous fat on the body of an animal, as an indicator of individual health (Jefferies, 1961; Edmonson et al., 1988; Dorsten 2004; Pettis et al., 2004; Cook et al., 2010; Zielke et al., 2018). With respect to the New Zealand population of leopard seals, the species was only recently (May 2019) classified as resident under the Department of Conservation New Zealand Threat Classification System (Baker et al. 2019), and therefore it is vital to understand population parameters (*e.g.* health, sex and age-classes) as part of management initiatives. Furthermore, due to their presence in this northern region being so far from their core home range of Antarctica, it is essential to observe health to determine how well they are adapting to the change in environment and resources (for example sea temperature, lack of ice floes and prey items), which may help to understand other research questions regarding their presence in New Zealand, and how they the utilise this environment.

Data from this study identified that leopard seals are overall in very good health, with only a small number of sighting records (10%) being identified as containing Poor BC individuals. This is in stark contrast to publications from other northern regions outside of Antarctica which have reported that

leopard seals observed in these areas are predominantly in poor condition (Rounsevell and Pemberton 1994; Mawson and Coughran 1999; Bester et al., 2017). There is a general trend in the literature that condition is inversely related to latitude, whereby the further north leopard seals are observed the lower the condition they are found in (van den Hoff et al., 2005; Bester et al., 2006). This result overall reinforces that New Zealand is part of the leopard seals current home range and advocates that they are not just able to survive but thrive within this northern region. However, whether this has always been the case and leopard seal presence has been under-reported or whether there is an increase in the number of leopard seals inhabiting New Zealand is not yet clear.

6.2.3 Limitations and improvements

Due to the nature of citizen science data collection many of the sightings that contained photographs were not of desirable quality. This resulted in a very small sample size ($n=80$) compared to the total number of photographic sightings ($n=1408$ until the end of 2018; Hupman et al., 2019) observed in New Zealand, in addition to other studies using photo-ID techniques (Joblon et al., 2008; Pettis et al., 2011) and as a consequence results were not highly generalisable to the New Zealand leopard seal population. One way to enhance sample sizes to make results more generalisable is via data augmentation, which involves altering existing photographs either by flipping, scaling or rotating, and reinstating them back into the dataset as a new data point (Bogucki et al., 2018). Using this method datasets can quickly be magnified to larger sizes (Bogucki et al., 2018). Alternatively, in New Zealand guidelines outlining the type of photographs required for BC assessment (*i.e.* whole left or right-side profile where the seal is lying in ventral recumbency, parallel to the ground and perpendicular to the camera), could be promoted by means of social media and community engagement events to increase the number of suitable photographs obtained from citizen science. As the performance of digital cameras and smartphone technology continues to develop, the availability of high-resolution photographs obtained by citizen scientists may increase, facilitating the citizen science approach to this method. This manuscript focused on records of sighting 'encounters' not known photo-ID'd individuals. The consequence of this is that BC could not be thoroughly assessed between individuals or compared over time for the same animal. In future research, BC assessment should be run in conjunction with photo-ID methods (*i.e.* using spot patterns on leopard seal pelage) to identify individual seals.

BCS itself has limitations that should be identified. Most significantly, BCS is a subjective measure which can vary highly between observers. Even with high resolution photographs it is sometimes difficult to determine whether bony protrusions are visible or not, as thickness of body fat does not always fall into discrete categories, and rather, it is a continuous measure. Herein it was also assumed

that leopard seals maintain, lose and gain body fat at the same rate all over the body, however this has not been confirmed in existing literature. Within this manuscript no differentiation was made between male and female leopard seals, or between age-classes (*i.e.* pup, juvenile or adult) and as such it was assumed that leopard seals of both sexes and all age-classes carry their weight in the same way. However, although leopard seals are not a sexually dimorphic species, it has been reported that some discrete differences do exist in the way weight is distributed between adult males and females, as well as between adults and juveniles (Hamilton 1939; Gwynn 1953). Research on skull morphology by Hamilton (1939) reports that adult females have larger sagittal crests, whilst adult males have wider zygomatic arches, which could potentially have affected BCS as these were two morphological landmarks of interest. Hamilton (1939) also stated that juvenile leopard seals have an attenuated and cylindrical shape with a large head, and that while females develop a large thorax as they mature from juveniles to adults, the morphology of males remains similar to that of a juvenile throughout its life. Similar findings were supported by Gwynn (1953) who reported differences between animals in their first three years and at adulthood. Gwynn (1953) specified that one-year old leopard seals were easily distinguishable by smaller size and smaller heads, while second-year animals had “*noticeable larger heads*” and three-year olds approached the length of adult seals, however had narrow hips “*giving them a long, slim figure*”. This morphological variation between sex and age groups will likely have influenced not only BCS but BC as predicted by photographic measurements (in LDA and RFC) and by pixel information extracted from seal profiles (in ANN).

An assessment of reliability was conducted within this chapter during which ‘repeat encounters’ were independently assessed by the same observer (the author) to determine if equivalent BCS were allocated for sighting records from the same date and location. It was concluded that on 8 out of 9 of these repeat encounters the same BCS were allocated using the same visible bony protrusions. On the single event that different BCS were scored for two sighting records documented on the same occasion, it was suggested that this could have been due to the difference in the number of photographs between these sighting records. The record with sighting-ID HL029 which contained a greater number of photographs was allocated a BCS of 2 (*i.e.* Moderate BC), whereas the record sighting-ID HL027 with a smaller number of photographs was allocated a score of 0 (*i.e.* Good BC), thereby these two bony protrusions (neural spines and pelvic bones) were both missed in this assessment. It was therefore surmised that only photographs which met all criteria from both Photograph Assessment and Photograph Quality Criteria (PAC and PQC) should be used in future assessments of BC. Moreover, only one photograph from each record (of the leopard seal lying in ventral recumbency and parallel to the ground and perpendicular to the camera) should be used for

comparisons of BC. Although this would mean that some bony protrusions may not be identified, the degree of error between sightings would be equal and internal validity would increase.

6.3 Chapter 3: Dimensionality reduction of photographic measurements of body width extracted from sighting records of New Zealand leopard seals using Principal Component Analysis

6.3.1 Key research findings

The use of PCA in this chapter showed that photographic measurements of leopard seal body width were associated to pre-assigned BCGs based upon a BCSS. This provided support for the validity of the BCS technique, enabling the method to be used in future estimates of leopard seal BC within the region. Using PCA, the 13 total photographic measurements extracted from leopard seal images were refined down to three without losing any variation within the data. These three variables were useful on their own for defining BC, thereby potentially decreasing processing time by eliminating the need to measure the remaining ten measurements during potential future study. Of these three measurements, S2 and S3 were both located around the neck area, which highlighted this morphological area as being important for differentiating between BC of New Zealand leopard seals. Photographic measurement S6 was the only measurement of body width where two statistically significant differences were observed between BCGs; Good – Moderate and Good – Poor, signifying S6 as the most important measurement for differencing between the three BCGs. Finally, in this chapter BCG Good was most easily distinguished from both Moderate and Poor BC. This was contrary to observations by van den Hoff et al., (2005) who used manual body measurements of leopard seals to estimate mass and reported that they were able to differentiate poor condition individuals to those in medium and good condition.

6.3.2 Contribution and significance

The third chapter in the manuscript describes the first instance of PCA being applied to New Zealand leopard seal data. The use of this technique revealed important morphological locations on the leopard seal body for differentiating between pre-assigned BCGs, derived from photogrammetry. Furthermore, PCA in this application validated the use of a BCSS which can be used as part of future research to investigate leopard seal population parameters, comparable to other studies of marine mammals where BCSS are applied, including Gray (Bradford et al., 2012) and North Atlantic right whales (Pettis et al., 2011). Similar to other photogrammetric studies (e.g. Krause et al., 2017; Christiansen et al., 2019), methods utilised within this chapter offer a non-invasive and low-risk and low cost alternative to obtaining measurements of free-ranging animals, which is functional for leopard seals in other regions as well as having potential applications for other species of pinniped.

6.3.3 Limitations and improvements

Shortcomings to the application of photogrammetry to data within the BCLSD were that due to the citizen science nature of the data collection, photographs of leopard seals were taken by a number of different observers utilising a variety of different camera types. Hence, robust photograph assessment techniques were employed (refer to Chapter 2 methods) in order to extract useful photographs for photogrammetry. This method largely reduced the number of useable photos to less than 5.7% of the original photograph library. A second consequence of this was that photographic measurements of body width were measured in pixels rather than being converted to metric units as seen in other research employing photogrammetry (Krause et al., 2017; Christiansen et al., 2019), which provide more interpretable data with uses for other applications (*e.g.* estimates of mass; Krause et al., 2017). Independent studies utilising subjective BC categories based upon visual observations have used validation techniques based upon morphometric measurements (van den Hoff et al., 2005; Chusyd et al., 2018) or blind studies evaluating agreement of individual assessors (Joblon et al., 2008; Pettis et al., 2011). Preferably, the BCGs allocated within this study would be validated by similar methods, however as obtaining manual morphometrics carries substantial risk to both leopard seals (due to sedations) and researchers (potential injury), photographic measurements were trialled as a non-invasive alternative. To improve upon the photogrammetric method, photographic body width methods could be measured by multiple independent assessors, to evaluate the accuracy and reliability of the measurements.

6.4 Chapter 4: The application of Linear Discriminant Analysis and Random Forest Classifiers for assigning body condition groups to sighting records of New Zealand leopard seals

6.4.1 Key research findings

The use of photographic measurements of body width within this chapter were successful in predicting BCG membership of sighting records containing New Zealand leopard seals relatively to relatively high levels of accuracy using both LDA (87.5%) and RFC (87.5%) methods, which further supports the relationship between the photographic measurements and pre-assigned BCGs. Overall, both techniques achieved similar correct classification rates (CCR), however the LDA method achieved higher levels of accuracy on average (78.13%) compared to the RFCs (66.52%) using this dataset, as well as making predictions into all three BCGs. The most accurate LDA model (LDA-4) used the three most important widths identified by PCA, as well as containing a control for bias of sighting records containing Good BC individuals by introducing class weights. In comparison, the most accurate RFC model (RFC-1) made use of all 13 photographic measurements and no control for class imbalance.

Therefore, with respect to potential future use, the use of the RFC increases time spent in data collection, making the LDA model more desirable for future application.

6.4.2 Contribution and significance

This research chapter describes the first instance when machine learning techniques have been applied to New Zealand leopard seal data, in order to make predictions of BCG membership using photographic measurements of body width obtained via photogrammetry. Significantly, this chapter identified that photographic measurements of body width (measured in pixels) can be reliably used to differentiate between and classify BC of New Zealand leopard seals as an objective and quantitative alternative to the subjective, qualitative BCSS currently in place. The use of morphometrics also allows unique leopard seal sighting records to be directly compared by metric values, which can be plotted to reveal differences in body shape which may not be so easily distinguished by the human eye. In potential future applications, the three measurements from the LDA-4 model; S2, S3 and S6 can be extracted from photographs in ImageJ and run through the LDA-4 model to get an estimation of BC which would predict correct classifications of BCG 87.5% of the time. These estimation outputs from the model can then be checked using a human observer to moderate model accuracy.

6.4.3 Limitations and improvements

While the most accurate LDA model (LDA-4) developed within this chapter outperformed LDA models in other research studies (Galimerti et al., 2018; Cutler et al., 2007; Hupman 2016), the RFCs generally performed to lower classification accuracy than other models in the literature (Rogers et al., 2017; Cutler et al., 2007), as well as lower than was expected in this chapter despite small sample size of the BCLSD and correlation between variables (Cutler et al., 2007; Tharwat et al., 2017). It is stated by Mohri et al., (2018) that the size and quality of the training data is essential to the success of a model's predictive ability. Within this dataset, as previously discussed there is an evident bias of sighting records containing Good BC individuals, which despite being corrected for within a number of the models in this chapter ($n=3$; LDA-2, LDA-4 and RFC-4) was still apparent when looking at the model predictions. In order to alleviate the effect of this bias within the LDA and RFC models, additional data collection would be required, particularly for of sighting records containing Moderate and Poor BC leopard seals. The addition of these records would reduce the impact of class imbalance. It is likely that as a result of this accuracy, reliability and generalisability of the models would also increase. As the full potential of RFCs were not explored within this chapter, future applications using this novel machine learning technique has the potential to produce more accurate BC predictions using data from the NZLSPL.

6.5 The application of Artificial Neural Networks for assigning body condition groups to sighting records of New Zealand leopard seals

6.5.1 Key research findings

The process of extracting silhouettes from photographs within New Zealand Leopard Seal Photograph Library allowed for an in-depth comparison of leopard seal body shape by means of ANNs. Both ANN models were successful in distinguishing between and classifying sighting records into the three BCGs (Good, Moderate and Poor) based upon differences in pixel values for each silhouette. ANN model ANN-2 (training data=90.65%, testing data=81.25%) which accounted for class imbalance achieved higher levels of accuracy than ANN-1 (training data=81.25%, testing data=50%) using data from the BCLSD. An additional dataset called 'Added-Data' was introduced within this chapter to determine how the ANN models would perform when introduced to new images of leopard seal silhouettes. Again, model ANN-2 outperformed ANN-1 where predictions made by ANN-2 matched estimates made by the author 75% of the time, compared to 58.34% of the time for ANN-1, suggesting that ANN-2 was more adept at generalising to new data. Interestingly, estimates of BCG based upon BCS made by the author on the lower quality Added-Data mirrored BCG calculations from sighting records in the BCLSD (refer to Table 5.3 in Chapter 5), signifying that accurate predictions of BC can be made even in lower resolution photographs.

6.5.2 Contribution and significance

This chapter defines the first application of ANNs to New Zealand leopard seal data. In this manuscript a technique adapted from a study on plant recognition using features extracted from images (Satti et al., 2013) was trialled using side-profile images of New Zealand leopard seals. The procedure used a total of 160 000 pixel values (800x200-pixel images) per sighting record as input values to train and test two ANN models, which were effective in making predictions of BCG that were in line with BCGs allocated by the author using BCS. Although the full potential of ANNs were not explored, classification accuracy of the most accurate ANN model (ANN-2) were relatively high within both training and testing data, which is promising for future applications of this technique. ANN-2 can be employed as an objective, quantitative alternative to the subjective, qualitative BCSS presently used for leopard seals in New Zealand as well as other regions. For this reason, R script used for ANN-2 is provided in Appendix II.

6.5.3 Limitations and improvements

Complications were identified with the ANN method regarding the size of the seal silhouette within the binary image file. Despite the fact that silhouettes were copied, rotated, and shrunk or enlarged to fit the new image file of 8000x2000 pixels, variation in the size and orientation of silhouettes

remained, as can be observed in Figures 6.1 and 6.2. In both figures, the seals are lying slightly at an angle to the observer (*i.e.* not a flat 180 degrees) and the body of seal within sighting-ID HL029 is in an elongated, stretched-out position compared to the seal envisioned in sighting-ID HL036, as well as the silhouette appearing slightly larger. The ANN models therefore may have incorporated size information in addition to shape, which could have affected BCG predictions. However, as ANNs are difficult to interpret, it is unknown which features, or variables, contributed to the classification process. In potential future applications of ANN models using leopard seal side-profile silhouettes efforts to standardise the size and orientation of the seal to ensure that difference in shape is the only parameter being tested.



Figure 6.1. A leopard seal silhouette used in Artificial Neural Network (ANN) models, of a Good body condition (BC) individual (sighting-ID HL036).



Figure 6.2. A leopard seal silhouette used in Artificial Neural Network (ANN) models, showing a Moderate body condition (BC) individual (sighting-ID HL029).

One technique often used in smaller datasets to increase accuracy and ensure results are more reliable is a method called *k*-fold cross-validation, which is often applied to assess the predictive ability of machine learning models (François et al., 2007; Baykan and Yilmaz 2011). In this technique, rather than training models with a training/validation/testing split, the entire dataset is split *k* times, where *k* equals the number of groups the dataset is divided into, resulting in a less biased estimate of model ability (Diamantidis et al., 2000; Baykan and Yilmaz 2011). The process involves randomly shuffling the data, splitting into *k* groups, then extracting one group to use as testing while the remaining groups are used to fit and evaluate the model as training. This method is repeated using all *k* groups and

overall model ability is assessed by the summary of all evaluations (Diamantidis et al., 2000; Baykan and Yilmaz 2011). *k*-fold cross-validation is known to work well with smaller datasets and therefore would be an appropriate method to trial with the New Zealand leopard seal data using ANNs, as an effort to increase the predictive ability and generalisability of the model.

6.6 Comparison of machine learning methods

With respect to testing data, the most accurate ANN model ANN-2 was less accurate at predicting BCGs compared to the most accurate RFC and LDA models (Table 6.1). This ANN model, based upon pixel data derived from images of leopard seal silhouettes, also only made predictions into one BCG (Good), as opposed to the two models based upon photographic measurements of body width which made predictions into two (Good and Moderate; RFC-1) or all three (LDA-4) of the BCGs. Upon review of Table 6.1 it can be inferred that the LDA model LDA-4 was the most successful of all three machine learning techniques in predicting BCG membership of New Zealand leopard seals. This particular model made use of the three 'most important widths' identified by PCA in Chapter 3 of this manuscript for differentiating between pre-assigned BCGs, and performed slightly better than RFC-1 (that made use of all 13 photographic measurements) by correctly classifying the one Poor BC sighting record (HL002; Table 6.1). Although ANN methods were less effective in this application, the potential of this machine learning technique was not fully explored and could provide more promising results in future investigation. It was proposed that the lower accuracy levels of both ANN models as compared to ANNs within the literature could be related to the volume of pixels for each sighting record, and subsequent models using images of silhouettes could perform to higher levels of accuracy if the pixel dimensions of the images were reduced. Advantages and disadvantages to each of the three machine learning techniques employed in this manuscript are outlined in Table 6.2.

Table 6.1. Comparison of actual (Correct) against predicted body condition groups (BCG) of sighting records containing images of New Zealand leopard seals (Sighting-ID) across the three most accurate models from each machine learning technique; Linear Discriminant Analysis (LDA-2.5), Random Forest Classifier (RFC-1) and Artificial Neural Network (ANN-2). Correct predictions are highlighted in blue whilst incorrect predictions are visible in orange. Cells where information was not applicable were denoted NA.

Sighting-ID	BCG			
	Correct	LDA-4	RFC-1	ANN-2
HL001	Good	Good	Good	Good
HL002	Poor	Poor	Moderate	Good
HL003	Good	Good	Good	Good
HL004	Moderate	Moderate	Moderate	Good
HL005	Good	Poor	Good	Good
HL006	Good	Good	Good	Good
HL007	Good	Good	Good	Good
HL010	Good	Good	Good	Good
HL012	Good	Good	Good	Good
HL016	Good	Good	Good	Good
HL026	Good	Good	Good	Good
HL050	Good	Good	Good	Good
HL052	Good	Good	Good	Good
HL054	Good	Good	Good	Good
HL058	Moderate	Good	Good	Good
HL072	Good	Good	Good	Good
CCR (%)	NA	87.5	87.5	81.25

Table 6.2. Advantages and disadvantages of the machine learning methods trialled in this manuscript for use at predicting body condition groups (BCG) for New Zealand leopard seals.

Method	Advantages	Disadvantages
<i>Principal Component Analysis (PCA)</i>	<ul style="list-style-type: none"> ❖ PCA produced distinct values (and plots) which depicted how each variable contributed to the principal components. This was extremely useful for highlighting the three most important variables for differentiating between pre-assigned BCGs, these three widths went on to be the most successful machine learning model in making predictions of BCG based upon the measurements alone. 	<ul style="list-style-type: none"> ❖ Data was required to conform to rules of normality and homogeneity therefore umbilical width was not able to be analysed, despite being identified as a morphological area of interest.
<i>Linear Discriminant Analysis (LDA)</i>	<ul style="list-style-type: none"> ❖ Results of LDA are interpretable, the equation for plotting individual sighting records based upon linear discriminants shows how each predictor variable contributed to the linear discriminant. ❖ A very simple machine learning technique to run and tune. ❖ Utilised only three out of the 13 total photographic measurements to achieve relatively high levels of classification accuracy. 	<ul style="list-style-type: none"> ❖ Data was required to conform to rules of normality and homogeneity therefore umbilical width was not able to be analysed, despite being identified as a morphological area of interest.
<i>Random Forest Classifier (RFC)</i>	<ul style="list-style-type: none"> ❖ As a non-parametric approach data was not required to conform to normality and homogeneity assumptions thereby all 13 photographic measurements were subject for analysis, including umbilical width. ❖ Tolerates small datasets, useful as the BCLSD only contained 80 records. ❖ Plotting RFC trees showed the direction of classification and which variables were used to make certain predictions. ❖ A very simple machine learning technique to run and tune. 	<ul style="list-style-type: none"> ❖ As a non-parametric approach RFC trees - although interpretable – were subjective and could not be assessed by traditional statistical techniques such as <i>P</i> values. ❖ The most successful RFC model utilised all 13 photographic measurements and therefore is more labour intensive to achieve the same levels of accuracy as the LDA model which used only three photographic measurements.
<i>Artificial Neural Network (ANN)</i>	<ul style="list-style-type: none"> ❖ Utilises more information from an image than RFC and LDA as the complete silhouette is input into the model, it therefore took into account the entire body shape as opposed to body shape at select points; this method involved 160,000 data points per individual seal photograph compared to three using LDA or 13 for RFC. ❖ The small dataset in this instance meant that the ANNs could easily handle the large number of pixel values per image file, so image dimensions did not have to be significantly reduced. 	<ul style="list-style-type: none"> ❖ Usually applied to extremely large datasets – the dataset in this manuscript was extremely small. ❖ Difficult to interpret – remains unclear how the ANN differentiated between the silhouettes of leopard seal sighting records based upon pixel values. ❖ More difficult than the other three techniques to design and tune models.

6.7 Suggestions for further research

This thesis has provided baseline information regarding three non-invasive techniques for assessing BC of leopard seals by means of BCS and machine learning methods. It is recommended that future research continues to collect information on BC in New Zealand leopard seals for longitudinal studies. Such data would enable leopard seals in New Zealand to be monitored over time and can provide updated estimates of BC within this region. Moreover, this manuscript focused on records of sighting 'encounters' not known photo-ID'd individuals. In future research, BC assessment should be run in conjunction with photo-ID methods (*i.e.* using spot patterns on leopard seal pelage) to identify individual seals, which would enable assessments between individuals to be compared over time. In addition to completing further investigations on baseline information, this thesis has also identified a number of key research questions which need to be answered. Future key research questions are outlined below.

1. Can photographs of New Zealand leopard seal silhouettes be examined through an ANN using the k-fold cross validation method to improve model predictive ability and reliability?

This first research question is important for exploring the potential of ANNs to identify whether a more accurate machine learning model can be utilised to make objective BC predictions of leopard seals within this region.

2. Can photographs of New Zealand leopard seals be applied to an ANN to predict BCS?

This research question proposes whether ANNs can be employed to make more refined predictions of direct BCS (which range between 0 and 5 *i.e.* six classification groups; refer to Chapter 2) as opposed to BCGs (*i.e.* three classification groups). BCS predictions from an ANN can also be directly compared to BCS derived from manual evaluations made by human assessors to further gauge the reliability of the BCS technique.

3. Can individual leopard seals be identified using photo-ID and then their individual BC be assessed over time?

This research question is particularly useful for determining the BC in which leopard seals are first observed in New Zealand waters as it may help to understand their presence here, *i.e.* do leopard seals arrive in Poor or Good BC? Does BC generally improve or deteriorate whilst they are in this region, due to difference in resources? This is particularly relevant for leopard seals that are judged to be in Poor BC or show presence of injuries for management considerations.

4. How does BC of the overall New Zealand leopard seal population compare to BC of leopard seals in Antarctica, as well as in other northern regions?

This research question is important for defining the difference (if any) in BC of the overall New Zealand leopard seal population, as compared to leopard seals observed within their core home range (Antarctica) and north of the polar front as vagrant visitors to other regions.

6.8 Management considerations

Monitoring BC in New Zealand leopard seals is important not only to help to understand their presence in the region but for monitoring the health of the population as a gauge of overall ecosystem health. As a top marine predator, the health condition of this species is indicated by resource availability and thereby overall habitat quality (Krause et al., 2017). Equally as important, it is vital to monitor the health of leopard seals in a time when pagophilic species are particularly vulnerable to the effects of climate change. Leopard seal population estimates have produced extremely variable figures in the past and IUCN states an unknown population trend at present (Hückstädt 2015), thereby leopard seals may be susceptible to impacts of climate change. Studies such as this that provide baseline information on leopard seal BC within a particular region can therefore assist in conservation efforts of the species.

In New Zealand the presence of leopard seals hauled-out at beaches and marinas can frequently draw in crowds (LeopardSeals.Org unpublished data). On occasions where these individuals are identified as being in Poor BC, it is particularly important that members of the public are made aware of this as disturbance to these individuals can have detrimental impacts. Leopard seals tend to spend much of their time at sea and only haul-out to rest and digest food (Gwynn 1953), therefore, they are forced to waste valuable energy when disturbed back into the water to locate a new haul-out site.

Findings from this manuscript have identified that a large proportion of the sampled leopard seals in New Zealand waters were of good health, as the majority occupied either Good or Moderate BCGs. This indicates that the species can survive well within this region at the northern limits of their range. Data from this thesis should assist management protocols for New Zealand's Department of Conservation for the protection of this species within the region.

6.9 Concluding statement

Prior to this study research investigating BC of leopard seals involved a multitude of methods often employing techniques that carry substantial risk to seals as well as researchers. As result there currently does not exist a standardised and non-invasive way to examine BC for leopard seals in the

literature. This manuscript presents the first non-invasive procedures that were able to distinguish between body shape of Good, Moderate and Poor BC leopard seals in New Zealand from sighting records obtained by citizen scientists and volunteer researchers, using photographic measurements of body width and pixel values from images of leopard seal silhouettes. Of all three methods explored, LDA produced the most accurate classification model, decided by the highest number of correct classification predictions and ability to generalise into all three BCGs. Although less accurate, ANNs were also identified as being suitable for New Zealand leopard seal data, due to their ability to utilise large, complex datasets and flexibility with lower quality images. To improve the ANN models for future use, implementing a k -fold cross validation method to a larger sample size of silhouette images, with smaller pixel dimensions could improve both accuracy and reliability of the ANN. Ideally, for both LDA and ANN methods more sighting records containing Moderate and Poor BC animals should be incorporated into the database to reduce bias thereby improving the ability of the model to generalise when presented with new images and produce outcomes that are generalisable to the New Zealand leopard seal population. Following this, machine learning models could then be used to predict BCG of leopard seals in lower quality images (*i.e.* images that are lower in resolution) where it is not possible to accurately allocate a BCG. Alternatively, the already instated BCSS could be continued since there is now evidence that these categories are validated by photographic measurements of body width. All procedures outlined within this manuscript can be applied to leopard seal populations in other locations of the southern hemisphere, which would enable longitudinal comparisons of BC across regions.

Appendix I: Details on how to plot a data observation onto a biplot using linear discriminant scores.

LDA models can be plotted using a scatterplot which visualises how the linear discriminants (LD1 and LD2) separate observations belonging to the classification groups. An example of such a plot can be seen in Figure 4.1, which visualises model LDA-4. The co-ordinates for the plot were determined by discriminant scores, *i.e.* the score for each sighting record for each linear discriminant using the linear regression equation:

$$V_{1i} = b_0 + b_1DV_{1i} + b_2DV_{2i}$$

Where V_{1i} represents the linear discriminant (*i.e.* LD1 and LD2), and DV_{1i} and DV_{2i} represent the predictor variables (*i.e.* photographic measurements S2, S3 and S6 for LDA-4). The b -values within this equation give information about the contribution of each predictor variable to the linear discriminant known as coefficients for the linear discriminants, and b_0 is the y-intercept. For LDA-4 the coefficients for LD1 and LD2 (b -values) can be seen in Table 4.4.

Using the b -values, the equations for calculating discriminant scores for LD1 and LD2 were:

$$LD1 = b_0 + (-2.85 \times S2_i) + (3.96 \times S3_i) - (7.53 \times S6_i)$$

$$LD2 = b_0 + (-0.80 \times S2_i) + (7.82 \times S3_i) - (0.16 \times S6_i)$$

To translate this into a point on the biplot, example sighting record HL008 was used whereby S2 = 0.095, S3 = 0.138 and S6 = 0.110, therefore:

$$LD1 = b_0 + (-2.85 \times 0.095) + (3.96 \times 0.138) - (7.53 \times 0.110)$$

$$LD1 = -0.41$$

$$LD2 = b_0 + (-0.80 \times 0.095) + (7.82 \times 0.138) - (0.16 \times 0.110)$$

$$LD2 = 0.68$$

Therefore, co-ordinates for leopard seal sighting record HL008 on the biplot were (-0.41, 0.68; Figure 4.1).

Appendix II: R script for ANN-2.

Install packages tensorflow and keras:

```
install.packages("tensorflow")
library(tensorflow)
install_tensorflow()
library(tensorflow)
tf$constant("Hello Tensorflow")
install.packages("keras")
library(keras)
```

Read-in training and testing data and convert to matrix:

```
testx <- read.csv("../input/ls-data/test.csv", row.names = 1)
trainx <- read.csv("../input/ls-data/train.csv", row.names = 1)
trainx <- as.matrix(trainx)
testx <- as.matrix(testx)
```

Add BCG labels (where Good=0, Moderate=1 and Poor=2):

```
trainy <- c(0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,1,0,0,0,0,0,0,1,0,0,1,1,0,2,
0,0,1,0,0,1,0,1,0,0,0,0,0,2,1,2,2,2,0,0,1,0,0,0,2,0,2,0,0)
testy <- c(0,2,0,1,0,0,0,0,0,0,0,0,0,0,1,0)
trainLabels <- to_categorical(trainy)
testLabels <- to_categorical(testy)
```

Design and fit model:

```
model2 <- keras_model_sequential()
model2 %>%
  layer_dense(units = 256, activation = 'relu', input_shape = c(160000)
) %>%
  layer_dense(units = 128, activation = 'relu') %>%
  layer_dense(units = 3, activation = 'softmax')
summary(model2)

model2 %>%
  compile(loss = 'categorical_crossentropy',
          optimizer = optimizer_rmsprop(),
          metrics = 'accuracy')

history2 <- model2 %>%
  fit(trainx,
      trainLabels,
      epochs = 100,
      validation_split = 0.2,
      class_weight = list("0" = 1, "1" = 3.8, "2" = 7.125))
```

Plotting optimisation (for loss values) and performance (for accuracy values) learning curves for ANN-2 (history2):

```
plot(history2)
```

Predict BCGs (classes) for training data using ANN-2 which produces a confusion matrix:

```
pred2 <- model2 %>% predict_classes(trainx)
table(Predicted = pred2, Actual = trainy)
```

Generate a list of BCG predictions for training data against actual BCGs using ANN-2:

```
prob2 <- model2 %>% predict_proba(trainx)
cbind(prob2, Predicted = pred, Actual= trainy)
```

Evaluate loss and accuracy for training data:

```
model2 %>% evaluate(trainx, trainLabels)
```

Predict BCGs (classes) for testing data using ANN-2 which produces a confusion matrix:

```
pred <- model2 %>% predict_classes(testx)
table(Predicted = pred, Actual = testy)
```

Generate a list of BCG predictions for testing data using ANN-2:

```
prob <- model2 %>% predict_proba(testx)
cbind(prob, Predicted = pred, Actual= testy)
```

Evaluate loss and accuracy for training data:

```
model2 %>% evaluate(testx, testLabels)
```

Input Added-Data:

```
adddedata <- read.csv("../input/added-data/added-data.csv")
adddedata <- as.matrix(newdata)
```

Predict BCG of Added-Data using ANN-2:

```
pred3 <- model2 %>% predict_classes(addeddata)
```

Compare predictions of Added-Data with actual BCG estimates (where Good=0, Moderate=1 and Poor=2) using a confusion matrix:

```
addeddata.actual <- c(0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,1,1,0,0,0,0,1,  
0,1,0,1,0,2,2,2,0,0,1,0,0,2,1,0,2,1,0,0)  
table(Predicted = pred3, Actual = addeddata.actual)
```

References

- Acevedo, J. & Martínez, F. (2013) Residence of the leopard seal in the Magellan Strait: a potential sub-Antarctic population inhabiting the waters of Southern Chile? *Polar biology*, 36(3), 453-456.
- Acevedo, J., González, A., Garthe, S., González, I., Gómez, R. & Aguayo-Lobo, A. (2017) Births of leopard seals *Hydrurga leptonyx* in southern Chile. *Polar Biology*, 40(3), 713-717.
- Act, M.M.P. (1978) Number 80, as at 22 December 2005. *Public Act. New Zealand Government*.
- Affonso, C., Sassi, R. J. & Barreiros, R. M. (2015) Biological image classification using rough-fuzzy artificial neural network. *Expert Systems with Applications*, 42(24), 9482-9488.
- Aguayo-Lobo, A., Acevedo, J. R., Brito, J. L., Acuna, P. G., Basso, M., Secchi, E. R. & Dalla Rosa, L. (2011) Presence of the Leopard Seal, *Hydrurga leptonyx*, (De Blainville, 1820), on the Coast of Chile: An Example of the Antarctica – South America Connection in the Marine Environment. *Oecologia Australis*, 15(1), 69–85.
- Allaire, J. J. & Chollet, F. (2020) keras: R Interface to 'keras'. R package version 2.3.0.0.
- Allaire, J. J. & Tang, Y. (2020) tensorflow: R Interface to 'Tensorflow'. R package version 2.2.0.
- Atkinson, A., Shreeve, R. S., Hirst, A. G., Rothery, P., Tarling, G. A., Pond, D. W., Korb, R. E., Murphy, E. J. & Watkins, J. L. (2006) Natural growth rates in Antarctic krill (*Euphausia superba*): II. Predictive models based on food, temperature, body length, sex, and maturity stage. *Limnology and Oceanography*. 51(2), 973-987.
- Baykan, N. A. & Yilmaz, N. (2011) A mineral classification system with multiple artificial neural network using k-fold cross validation. *Mathematical and Computational Applications*, 16(1), 22-30.
- Beck, G. G., Smith, T. G. & Hammil, M. O. (1993) Evaluation of body condition in the Northwest Atlantic harp seal (*Phoca groenlandica*). *Canadian Journal of Fisheries and Aquatic Sciences*, 50, 1372–1381.
- Beck, S., Foote, A. D., Koetter, S., Harries, O., Mandleberg, L., Stevick, P. T., Whooley, P. & Durban, J. W. (2014) Using opportunistic photo-identifications to detect a population decline of killer whales (*Orcinus orca*) in British and Irish waters. *Marine Biological Association of the United Kingdom. Journal of the Marine Biological Association of the United Kingdom*, 94(6).
- Begue, A., Kowlessur, V., Singh, U., Mahomoodally, F. & Pudaruth, S. (2017) Automatic recognition of medicinal plants using machine learning techniques. *International Journal of Advanced Computer Science and Applications*, 8(4), 166-175.
- Berry, J. A. (1960) The occurrence of a leopard seal (*Hydrurga leptonyx*) in the tropics. *Journal of Natural History*. 3(34), 591–591.
- Best, N., Bradshaw, C., Hindell, M. & Nichols, P. (2003) Vertical stratification of fatty acids in the blubber of southern elephant seals (*Mirounga leonina*): implications for diet analysis. *Comparative Biochemistry and Physiology Part B*. 134, 253–263.
- Bester, M. N., Bester, W. A., Wege, M., Schofield, R. A. & Glass, T. A. (2017) Vagrant leopard seal at Tristan da Cunha Island, South Atlantic. *Polar Biology*, 40(9), 1903-1905.
- Bester, M. N., Hofmeyr, G. J. G., Kirkman, S. P., Chauke, L. F., De Bruyn, P. J. N., Ferreira, S. M., Makhado, A. B., Maswime, T. A. M., McIntyre, T., Mulaudzi, T. W. & Munyai, F. M. (2006) The leopard seal at Marion Island, vagrant or seasonal transient? *South African Journal of Wildlife Research-24-month delayed open access*, 36(2), 195-198.
- Bester, M.N., Ferguson, J.W.H. & Jonker, F.C. (2002) Population densities of pack ice seals in the Lazarev Sea, Antarctica. *Antarctic Science*, 14(2), 123-127.
- Biau, G. & Scornet, E. (2016) A random forest guided tour. *Test*, 25(2), 197-227.
- Binder, C. M. & Hines, P. (2012) Applying automatic aural classification to cetacean vocalizations. In *Proceedings of Meetings on Acoustics*. Acoustical Society of America.
- Bogucki, R., Marek, C., Brangwynne Khan, C., Klimek, M., Kanty Milczek, J. & Mucha, M. (2018) Applying deep learning to right whale photo identification. *Conservation Biology* 33(3), 676–684.
- Bonner, N. (1994) *Seals and sea lions of the world*. London: Blandford.
- Bourque, J., Dietz, R., Sonne, C., Leger, J. S., Iverson, S., Rosing-Asvid, A., Hansen, M. & McKinney, M. A. (2018) Feeding habits of a new Arctic predator: insight from full-depth blubber fatty acid signatures of Greenland, Faroe Islands, Denmark, and managed-care killer whales *Orcinus orca*. *Marine Ecology Progress Series*, 603, 1-12.
- Boyd, I. L. & Murray, A. W. A. (2001) Monitoring a marine ecosystem using responses of upper trophic level predators. *Journal of Animal Ecology*, 70(5), 747-760.
- Bradford, A. L., Weller, D. W., Punt, A. E., Ivashchenko, Y. V., Burdin, A. M., VanBlaricom, G. R. & Brownell Jr, R. L. (2012) Leaner leviathans: body condition variation in a critically endangered whale population. *Journal of Mammalogy*, 93(1), 251-266.
- Breiman, L. (2001) Random forests. *Machine learning*, 45(1), 5-32.
- Brown, K. G. (1952) Observations on the newly born leopard seal. *Nature*, 170(4336), 982-983.

- Carroll, C. L. & Huntington, P. J. (1988) Body condition scoring and weight estimation of horses. *Equine veterinary journal*, 20(1), 41-45.
- Christiansen, F., Sironi, M., Moore, M. J., Di Martino, M., Ricciardi, M., Warick, H. A., Irschick, D. J., Gutierrez, R. & Uhart, M. M. (2019) Estimating body mass of free-living whales using aerial photogrammetry and 3D volumetrics. *Methods in Ecology and Evolution*, 10(12), 2034-2044.
- Christiansen, F., Vivier, F., Charlton, C., Ward, R., Amerson, A., Burnell, S. & Bejder, L. (2018) Maternal body size and condition determine calf growth rates in southern right whales. *Marine Ecology Progress Series*, 592, 267-281.
- Cook, R. C., Cook, J. G., Stephenson, T. R., Myers, W. L., Mccorquodale, S. M., Vales, D. J., Irwin, L. L., Hall, P. B., Spencer, R. D., Murphie, S. L. & Schoenecker, K. A. (2010) Revisions of rump fat and body scoring indices for deer, elk, and moose. *The Journal of Wildlife Management*, 74(4), 880-896.
- Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J. & Lawler, J. J. (2007) Random forests for classification in ecology. *Ecology*, 88(11), 2783-2792.
- Czarnowski, I. & Jędrzejowicz, P. (2018) An Approach to Data Reduction for Learning from Big Datasets: Integrating Stacking, Rotation, and Agent Population Learning Techniques. *Complexity*, 2018.
- Dasgupta, A., 2014. reprtree: Representative trees from ensembles. *A package for R*.
- Davidson, A. D., Boyer, A. G., Kim, H., Pompa-Mansilla, S., Hamilton, M. J., Costa, D.P., Ceballos, G. & Brown, J. H. (2012) Drivers and hotspots of extinction risk in marine mammals. *Proceedings of the National Academy of Sciences*, 109(9), 3395-3400.
- Dawson, S. M., Bowman, M. H., Leunissen, E. & Sirguy, P. (2017) Inexpensive aerial photogrammetry for studies of whales and large marine animals. *Frontiers in Marine Science*, 4, 366.
- de Moura, J. F., di Dario, B. P. S. & Siciliano, S. (2011) Occurrence of pinnipeds on the coast of Rio de Janeiro State, Brazil. *Marine Biodiversity Records*, 4.
- De'ath, G. & Fabricius, K. E. (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, 81, 3178-3192.
- Diamantidis, N. A., Karlis, D. & Giakoumakis, E.A. (2000) Unsupervised stratification of cross-validation for accuracy estimation. *Artificial Intelligence*, 116(1-2), 1-16.
- Dorsten, C. M. & Cooper, D. M. (2004) Use of body condition scoring to manage body weight in dogs. *Journal of the American Association for Laboratory Animal Science*, 43(3), 34-37.
- Earle, D. F. (1976) A guide to scoring dairy cow condition. *Journal of the Department of Agriculture of Victoria*, 74.
- Edmonson, A. J., Lean, I. J., Weaver, L. D., Farver, T. & Webster, G. (1989) A body condition scoring chart for Holstein dairy cows. *Journal of dairy science*, 72(1), 68-78.
- Elliot, P. (1982) Comparison of Macquarie Is and NSW leopard seal records with notes on their diet. *Thylacinus*, 7(2), 10-15.
- Erickson, A. & Hanson, M. (1990) Continental estimates and population trends of Antarctic ice seals. In K. Kerry and G. Hemper, (Eds.). *Antarctic ecosystems: ecological change and conservation*. New York: Springer.
- Feldesman, M. R. (2002) Classification trees as an alternative to linear discriminant analysis. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, 119(3), 257-275.
- Field, A., Miles, J. & Field, Z. (2012) *Discovering statistics using R*. Sage publications.
- Forcada, J. & Aguilar, A. (2000) Use of photographic identification in capture-recapture studies of Mediterranean monk seals. *Marine Mammal Science*, 16(4), 767-793.
- Fox, J. & Weisberg, S. (2019) *An R companion to applied regression*. 3rd ed. Sage, California: Thousand Oaks.
- François, D., Rossi, F., Wertz, V. & Verleysen, M. (2007) Resampling methods for parameter-free and robust feature selection with mutual information. *Neurocomputing*, 70(7-9), 1276-1288.
- Galimberti, F., Sanvito, S., Vinesi, M. C. & Cardini, A. (2019) "Nose-metrics" of wild southern elephant seal (*Mirounga leonina*) males using image analysis and geometric morphometrics. *Journal of Zoological Systematics and Evolutionary Research*, 57(3), 710-720.
- Giraldo-Zuluaga, J. H., Salazar, A., Gomez, A. & Diaz-Pulido, A. (2017) Automatic recognition of mammal genera on camera-trap images using multi-layer robust principal component analysis and mixture neural networks. *arXiv preprint arXiv:1705.02727*.
- Grainger, C. & McGowan, A. A. (1982) The significance of pre-calving nutrition of the dairy cow. *Occasional Publication, New Zealand Society of Animal Production*, New Zealand.
- Gray, R. B., Rogers, T. L. & Canfield, P. J. (2009) Health Assessment of the Leopard Seal, *Hydrurga leptonyx*, in Prydz Bay, Eastern Antarctica and NSW, Australia. In K.R. Kerry and M.J. Riddle (Eds.). *Health of Antarctic Wildlife: A Challenge for Science and Policy*. Berlin: Springer, pp.167-192.
- Grosjean, P., Ibanez, F., Etienne, M. & Grosjean, M. P. (2018). Package 'pastecs'.

- Guerrero, A. I., Negrete, J., Márquez, M. E. I., Mennucci, J., Zaman, K. & Rogers, T. L. (2016) Vertical fatty acid composition in the blubber of leopard seals and the implications for dietary analysis. *Journal of Experimental Marine Biology and Ecology*, 478, 54-61.
- Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology letters*, 8(9), 993-1009.
- Gurarie, E., Bengston, J. L., Bester, M. N., Blix, A. S., Cameron, M., Bornemann, H., Nordøy, E. S., Plötz, J., Steinhage, D. & Boveng, P. (2017) Distribution, density and abundance of Antarctic ice seals off Queen Maud Land and the eastern Weddell Sea. *Polar Biology*, 40, 1149-1165.
- Gwynn, A.M. (1953) *The status of the leopard seal at Heard Island and Macquarie Island, 1948-1950*. Antarctic Division, Department of External Affairs.
- Hall-Aspland, S. A., & Rogers, T. L. (2004) Summer diet of leopard seals (*Hydrurga leptonyx*) in Prydz Bay, Eastern Antarctica. *Polar Biology*, 27(12), 729-734.
- Hamilton, J.L. (1939). The leopard seal, *Hydrurga leptonyx* (de Blainville). *Discovery Reports*, 18, 239-264.
- Harting, A., Baker, J. & Becker, B. (2004) Non-metrical digital photo-identification system for the Hawaiian monk seal. *Marine Mammal Science*, 20(4), 886-895.
- Henry, L. & Wickham, H. (2020) rlang: Functions for base types and core R and 'tidyverse' features. R package version 0.4.7.
- Hiby, L., Lundberg, T., Karlsson, O., Watkins, J., Jüssi, M., Jüssi, I. & Helander, B. (2007) Estimates of the size of the Baltic grey seal population based on photo-identification data. *NAMMCO scientific publications*, 6, 163-175.
- Hocking, D. P., Marx, F. G., Fitzgerald, E. M. & Evans, A. R. (2017) Ancient whales did not filter feed with their teeth. *Biology letters*, 13(8).
- Hofman, R. J., Reichle, R. A., Siniff, D. B. & Müller-Schwarze, D. (1977) The leopard seal (*Hydrurga leptonyx*) at Palmer station, Antarctica. *Adaptations within Antarctic Ecosystems*. Smithsonian Institution, Washington DC, USA.
- Hotelling, H. (1951) A generalized T test and measure of multivariate dispersion. In *Proceedings of the second Berkeley symposium on mathematical statistics and probability*. The Regents of the University of California.
- Hückstädt, L. (2015) *Hydrurga leptonyx*. *The IUCN Red List of Threatened Species*, 10.
- Hupman, K. E. (2016) *Photo-identification and its application to gregarious delphinids: Common dolphins (Delphinus sp.) in the Hauraki Gulf, New Zealand* (Doctoral dissertation, Massey University).
- Hupman, K., Stockin, K. A., Pollock, K., Pawley, M. D. M., Dwyer, S.L., Lea, C. & Tezanos-Pinto, G. (2018) Challenges of implementing Mark-recapture studies on poorly marked gregarious delphinids. *PLoS ONE*. 13(7).
- Hupman, K., Visser, I. N., Fyfe, J., Cawthorn, M., Forbes, G., Grabham, A. A., Bout, R., Mathias, B., Benninghaus, E., Matucci, K., Cooper, T., Fletcher, L. & Godoy, D. (2019) From Vagrant to Resident: occurrence, residency and births of leopard seals (*Hydrurga leptonyx*) in New Zealand waters. *New Zealand Journal of Marine and Freshwater Research*, 54(1), 1-23.
- Jabbar, H. & Khan, R. Z. (2015) Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *Computer Science, Communication and Instrumentation Devices*, 163-172.
- Jain, A. K., Mao, J. & Mohiuddin, K. M. (1996) Artificial neural networks: A tutorial. *Computer*, 29(3), 31-44.
- Jefferies, B. C. (1961) Body condition scoring and its use in management. *Tasmanian journal of agriculture*, 32, 9-21.
- Jefferson, T. A., Webber, M. A. & Pitman, R. L. (2015) *Marine mammals of the world: a comprehensive guide to their identification*, 2nd ed. San Diego.
- Jessopp, M. J., Forcada, J., Reid, K., Trathan, P. N. & Murphy, E. J. (2004) Winter dispersal of leopard seals (*Hydrurga leptonyx*): environmental factors influencing demographics and seasonal abundance. *Journal of Zoology*, 263(3), 251-258.
- Joblon, M. J., Pokras, M. A., Morse, B., Harry, C. T., Rose, K. S., Sharp, S. M., Niemeyer, M. E., Patchett, K. M., Sharp, W. B. & Moore, M. J. (2014) Body Condition Scoring System for Delphinids Based on Short-beaked Common Dolphins (*Delphinus delphis*). *Journal of Marine Animals and Their Ecology*, 7(2), 5-13.
- Jolliffe, I. T. (1986) *Principal Component Analysis*. New York: Springer-Verlag.
- Jones, K. E., Ruff, C. B. & Goswami, A. (2013) Morphology and biomechanics of the pinniped jaw: mandibular evolution without mastication. *The Anatomical Record*, 296(7), 1049-1063.
- Kassambara, A. & Mundt, F. (2020) factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0. 7.
- Kawaguchi, S., Ishida, A., King, R., Raymond, B., Waller, N., Constable, A., Nicol, S., Wakita, M. & Ishimatsu, A. (2013) Risk maps for Antarctic krill under projected Southern Ocean acidification. *Nature Climate Change*. 3, 843-847.
- Kim, D. & Kim, S. K. (2012) Comparing patterns of component loadings: Principal Component Analysis (PCA) versus Independent Component Analysis (ICA) in analyzing multivariate non-normal data. *Behavior research methods*, 44(4), 1239-1243.
- King, J.E. (1983) *Seals of the World*. *British Museum Natural History*. Oxford: Oxford University Press.

- Klimley, A. P. & Brown, S. T. (1983) Stereophotography for the field biologist: measurement of lengths and three-dimensional positions of free-swimming sharks. *Marine Biology*, 74(2), 175-185.
- Koivuniemi, M., Auttila, M., Niemi, M., Levänen, R. & Kunnasranta, M. (2016) Photo-ID as a tool for studying and monitoring the endangered Saimaa ringed seal. *Endangered Species Research*, 30, 29-36.
- Krause, D. J., Goebel, M. E. & Kurle, C. M. (2020) Leopard seal diets in a rapidly warming polar region vary by year, season, sex, and body size. *BMC ecology*, 20, 1-15.
- Krause, D. J., Goebel, M. E., Marshall, G. J. & Abernathy, K. (2015) Novel foraging strategies observed in a growing leopard seal (*Hydrurga leptonyx*) population at Livingston Island, Antarctic Peninsula. *Animal Biotelemetry*, 3(1).
- Krause, D. J., Hinke, J. T., Perryman, W. L., Goebel, M. E. & LeRoi, D. J. (2017) An accurate and adaptable photogrammetric approach for estimating the mass and body condition of pinnipeds using an unmanned aerial system. *PLoS ONE*, 12(11).
- Kreiss, C. M., Boebel, O., Bornemann, H., Kindermann, L., Klinck, H., Klinck, K., Plötz, J., Rogers, T. L. & van Opzeeland, I. (2013) Call characteristics of high-double trill leopard seal (*Hydrurga leptonyx*) vocalizations from three Antarctic locations. *Polarforschung*, 83(2), 63-71.
- Krogh, A. (2008) What are artificial neural networks? *Nature biotechnology*, 26(2), 195-197.
- Kruskal, W.H. & Wallis, W.A. (1952) Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*. 47(260), 583-621.
- Kuhn, C. E., McDonald, B. I., Shaffer, S. A., Barnes, J., Crocker, D. E., Burns, J. & Costa, D. P. (2006) Diving physiology and winter foraging behavior of a juvenile leopard seal (*Hydrurga leptonyx*). *Polar Biology*, 29(4), 303-307.
- Kuhn, M. (2020) caret: Classification and Regression Training. R package version 6.0-86.
- Lamers, M., Eijgelaar, E. & Amelung, B. (2011). Last chance tourism in Antarctica – Cruising for change? In H. Lemelin, J. Dawson, and E. Stewart. *Last-Chance Tourism: Adapting Tourism Opportunities in a Changing World*. London: Routledge.
- Laws, R. M. (1957) *On the growth rates of the leopard seal, Hydrurga leptonyx (De Blainville, 1820)*. Originalarbeiten.
- Laws, R.M. (1953) The seals of the Falkland Islands and dependencies. *Oryx*, 2, 87–97.
- Lê, S., Josse, J. & Husson, F. (2008) FactoMineR: an R package for multivariate analysis. *Journal of statistical software*, 25(1), 1-18.
- Levene, H. (1960) Robust Tests for Equality of Variances. In I. Olkin (Eds.). *Contributions to Probability and Statistics*. Stanford: Stanford University Press, pp. 278-292.
- Lewis, P. J., Buchanan, B. & Johnson, E. (2005) Sexing bison metapodials using principal component analysis. *Plains Anthropologist*, 50(194), 159-172.
- Liaw, A. & Wiener, M. (2002) Classification and regression by randomForest. *R news*, 2(3), 18-22.
- Lowman, B. G., Scott, N. A. & Somerville, S. H. (1976) Condition scoring of cattle. The East of Scotland College of Agriculture. *Bulletin*, 6.
- Lynch, M. & Bodley, K. (2007) Phocid seals. *Zoo animal and wildlife immobilization and anesthesia*, 1.
- Lynch, M. J., Tahmindjis, M. A. & Gardner, H. (1999) Immobilisation of pinniped species. *Australian Veterinary Journal*, 77(3), 181-185.
- Mackey, B. L., Durban, J. W., Middlemas, S. J. & Thompson, P. M. (2008) A Bayesian estimate of harbour seal survival using sparse photo-identification data. *Journal of Zoology*, 274(1), 18-27.
- Mallet, M. (2020) *Environmental management of Antarctic cruise tourism: Does ship presence affect marine mammal acoustic behavior? A multi-year case study on the acoustic occurrence of Ross and leopard seals in relation to the presence of the German research icebreaker Polarstern* (Doctoral dissertation, Hochschule Bremerhaven).
- Manly, B. F. & Alberto, J. A. N. (2016) *Multivariate statistical methods: a primer*. CRC press.
- Marr, J. W. S. (1962) The natural history and geography of the Antarctic krill (*Euphausia superba* Dana). *Discovery Rep.* 32, 33-464.
- Masters, T. (1993) *Practical Neural Network Recipes in C++*. San Diego: Academic Press.
- Mawson, P. R. & Coughran, D. K. (1999) Records of sick, injured and dead pinnipeds in Western Australia 1980-1996. *Journal of the Royal Society of Western Australia*, 82.
- McCallum, A. & Nigam, K. (1998) A comparison of event models for naive bayes text classification. In *AAAI-98 workshop on learning for text categorization* (Vol. 752, No. 1, pp. 41-48).
- McCulloch, W. S. & Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115-133.
- Mohri, M., Rostamizadeh, A. & Talwalkar, A. (2018) *Foundations of machine learning*. MIT press.
- Mulvany, P. (1977) Dairy cow condition scoring. Handout No. 4468. *National Institute for Research in Dairying*. Reading, UK.

- Nicholson, K., Bejder, L., Allen, S. J., Krützen, M. and Pollock, K. H. (2012) Abundance, survival and temporary emigration of bottlenose dolphins (*Tursiops sp.*) off Useless Loop in the western gulf of Shark Bay, Western Australia. *Marine and Freshwater Research*, 63, 1059–1068.
- Noble, W. S. (2006) What is a support vector machine? *Nature biotechnology*, 24(12), 1565-1567.
- Omura, H., Ohsumi, S., Nemoto, T., Nasu, K., & Kasuya, T. (1969). Black right whales in the North Pacific. *Scientific Reports of the Whales Research Institute (Tokyo)*, 21, 1–78.
- Øritsland, T. (1970) Sealing and seal research in the South-west Atlantic Pack Ice, Sept-Oct 1964. In: M. W. Holdgate (Eds.). *Antarctic ecology*. Academic Press: London.
- Øritsland, T. (1977) *Food consumption of seals in the Antarctic pack ice*.
- Parson, S. & Jones, G. (2000) Acoustic identification of twelve species of echolocation bat by discriminant function and artificial neural networks. *Journal of Experimental biology*, 2003, 2641-2656.
- Pau, G., Fuchs, F., Sklyar, O., Boutros, M. & Huber, W. (2010) EBImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics*, 26(7), 979-981.
- Pearson, R. A. & Ouassat, M. (2000) *A guide to live weight estimation and body condition scoring of donkeys*. Scotland: Centre for Tropical Veterinary Medicine, University of Edinburgh.
- Pemberton, D. & Shaughnessy, P. D. (1993) Interaction between seals and marine fish-farms in Tasmania, and management of the problem. *Aquatic Conservation: marine and freshwater ecosystems*, 3(2), 149-158.
- Penney, R. L. & Lowry, G. (1967) Leopard seal predation of adelic penguins. *Ecology*, 48(5), 878-882.
- Pettis, H. M., Rolland, R. M., Hamilton, P. K., Brault, S., Knowlton, A. R. & Kraus, S. (2011) Visual health assessment of North Atlantic right whales (*Eubalaena glacialis*) using photographs. *Canadian Journal of Zoology*, 82, 8-19.
- Prebble, J. L., Shaw, D. J. & Meredith, A. L. (2015) Bodyweight and body condition score in rabbits on four different feeding regimes. *Journal of Small Animal Practice*, 56(3), 207-212.
- Priddy, K. L. & Keller, P. E. (2005) *Artificial neural networks: an introduction* (Vol. 68). SPIE press.
- Prieto, A., Prieto, B., Ortigosa, E. M., Ros, E., Pelayo, F., Ortega, J. & Rojas, I. (2016) Neural networks: An overview of early research, current frameworks and new challenges. *Neurocomputing*, 214, 242-268.
- Pussini, N. & Goebel, M.E. (2015) A safer protocol for field immobilization of leopard seals (*Hydrurga leptonyx*). *Marine Mammal Science*. 31(4), 1549-1558.
- Quetin, L. B. & Ross, R. M. (2001) Environmental variability and its impact on the reproductive cycle of Antarctic krill. *American Zoologist*, 41, 74-2001.
- R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Australia.
- Ratnaswamy, M. J. & Winn, H. E. (1993) Photogrammetric estimates of allometry and calf production in fin whales, *Balaenoptera physalus*. *Journal of Mammalogy*, 74(2), 323-330.
- Reijnders, P., Brasseur, S., van der Toorn, J., van der Wolf, P., Boyd, I., Harwood, J., Lavigne, D. & Lowry, L. (1993) *Seals, fur seals, sea lions, and walrus. Status survey and conservation action plan*. IUCN The World Conservation Union.
- Ren, S., He, K., Girshick, R. & Sun, J. (2016) Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6), 1137-1149.
- Rice, D.W. (1998) *Marine mammals of the world: Systematics and distribution. Special Publication Number 4, The Society for Marine Mammalogy*. USA: Allen Press.
- Riedman, M. (1990) *The pinnipeds: seals, sea lions and walruses*. Berkeley: The University of California Press.
- Robbins, J. R., Poncet, D., Evans, A. R. & Hocking, D. P. (2019) A rare observation of group prey processing in wild leopard seals (*Hydrurga leptonyx*). *Polar Biology*, 42(8), 1625-1630.
- Rodríguez, D., Bastida, R., Morón, S., Heredia, S. R. & Loureiro, J. (2003) Occurrence of leopard seals in northern Argentina. *Latin American Journal of Aquatic Mammals*, 2(1), 51-54.
- Rogers, T. & Bryden, M. M. (1995) Predation of Adelie penguins (*Pygoscelis adeliae*) by leopard seals (*Hydrurga leptonyx*) in Prydz Bay, Antarctica. *Canadian Journal of Zoology*, 73, 1001–1004.
- Rogers, T. D., Cambiè, G. & Kaiser, M. J. (2017) Determination of size, sex and maturity stage of free swimming catsharks using laser photogrammetry. *Marine Biology*, 164(11), 213.
- Rogers, T. L., Cato, D. H. & Bryden, M. M. (1995) Underwater vocal repertoire of the leopard seal (*Hydrurga leptonyx*) in Prydz Bay, Antarctica. *Sensory systems of aquatic mammals*, 223-236.
- Rogers, T. L., Cato, D. H., & Bryden, M. M. (1996) Behavioural significance of underwater vocalizations of captive leopard seals, *Hydrurga leptonyx*. *Marine Mammal Science*, 12, 414-427.

- Rogers, T.L. (2007): Age-related differences in the acoustic characteristics of male leopard seals, *Hydrurga leptonyx*. *The Journal of the Acoustical Society of America*, 122, 596-605.
- Rogers, T.L. (2009) Leopard Seal: *Hydrurga leptonyx*. In W.F. Perrin, B. Würsig and J. G. M Thewissen, (Eds.). *Encyclopaedia of Marine Mammals (Second Edition)*. Academic Press, pp. 673-674.
- Rohlf, F. J. (2005) TpsDig, digitize landmarks and outlines, version 2.05. *Department of Ecology and Evolution, State University of New York at Stony Brook*.
- Rounsevell, D. & Eberhard, I. (1980) Leopard Seals, *Hydrurga leptonyx* (Pinnipedia), at Macquarie Island from 1949 to 1979. *Wildlife Research*, 7(3), 403-415.
- Rounsevell, D. & Pemberton, D. (1994) The status and seasonal occurrence of Leopard Seals, *Hydrurga leptonyx*, in Tasmanian Waters. *Australian Mammalogy*, 17, 97-102.
- Savriama, Y. (2018) A step-by-step guide for geometric morphometrics of floral symmetry. *Frontiers in plant science*, 9, 1433.
- Sayer, S., Allen, R., Hawkes, L. A., Hockley, K., Jarvis, D. & Witt, M.J. (2019) Pinnipeds, people and photo identification: the implications of grey seal movements for effective management of the species. *Journal of the Marine Biological Association of the United Kingdom*, 99(5), 1221-1230.
- Scheffer, V.B. (1958) *Seals, sea lions and walruses: A review of the Pinnipedia*. Stanford: Stanford University Press.
- Shapiro, S. S. & Wilk, M. B. (1965) An analysis of variance test for normality (complete samples). *Biometrika*, 52, 91-611.
- Siniff, D. B. & Bengtson, J. L. (1977) Observations and hypotheses concerning the interactions among crabeater seals, leopard seals, and killer whales. *Journal of Mammalogy*, 58(3), 414-416.
- Siniff, D. B. & Stone, S. (1985) The role of the leopard seal in the tropho-dynamics of the Antarctic marine ecosystem. In W.R. Siegfried, W.R. Condy and R.M. Laws (Eds.). *Antarctic nutrient cycles and food webs*. Berlin, Heidelberg: Springer, pp. 555-560.
- Siniff, D. B., Garrott, R. A., Rotella, J. J., Fraser, W. R. & Ainley, D. G. (2008) Opinion: Projecting the effects of environmental change on Antarctic seals. *Antarctic Science*, 20(05), 425-435.
- Smith, I. W. G. (1985) *Sea mammal hunting and prehistoric subsistence in New Zealand* (Dissertation, University of Otago).
- Southwell, C., Bengtson, J., Bester, M. N., Schytte-Blix, A., Bornemann, H., Boveng, P., Cameron, M., Forcada, J., Laake, J., Nordøy, E. & Plötz, J. (2012) A review of data on abundance, trends in abundance, habitat utilisation and diet for Southern Ocean ice-breeding seals. *CCAMLR Science*, 19, 1-49.
- Steele, B. M. (2000) Combining multiple classifiers: an application using spatial and remotely sensed information for land cover mapping. *Remote Sensing of Environment*, 74, 545-556.
- Steele, B. M. (2000) Combining multiple classifiers: An application using spatial and remotely sensed information for land cover type mapping. *Remote sensing of environment*, 74(3), 545-556.
- Stone, S. & Meier, T. (1981) Summer leopard seal ecology along the Antarctic Peninsula. *Antarct. J.U.S.*, 16(5), 151-152.
- Taigman, Y., Yang, M., Ranzato, M. A. & Wolf, L. (2014) Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1701-1708).
- Tharwat, A., Gaber, T., Ibrahim, A. & Hassanien, A.E. (2017) Linear discriminant analysis: A detailed tutorial. *AI communications*, 30(2), 169-190.
- Torres, D. (1987) Antecedentes sobre el lobo fino de Juan Fernández, *Arctocephalus philippii*, y proyecciones para su estudio. In J.C. Castilla (Eds.). *Islas Oceánicas Chilenas: Conocimiento Científico y Necesidades de Investigaciones*. Ediciones de la Universidad Católica de Chile, pp. 353.
- Tukey, J. W., 1977. *Exploratory data analysis* (Vol. 2), 131-160.
- Turner, J., Colwell, S. R., Marchall, G., Lachlan-Cope, T., Carleton, A., Jones, P., Lagun, V., Reid, A. & Iagovkina, S. (2005) Antarctic climate change during the last 50 years, *International Journal of Climatology*, 25, 279-94.
- Urian, K. W., Hohn, A. A. & Hansen, L. J. (1999) Status of the photoidentification catalog of coastal bottlenose dolphins of the western North Atlantic: Report of a workshop of catalogue contributors. National Oceanic and Atmospheric Administration Administrative Report.
- Urian, K. W., Hohn, A. A. & Hansen, L. J. (1999) Status of the photoidentification catalog of coastal bottlenose dolphins of the western North Atlantic. Report of a workshop of catalogue contributors. National Oceanic and Atmospheric Administration Administrative Report, NMFS-SEFSC-425, North Carolina, U.S.A. pp. 1-24.
- van den Hoff, J., Fracarro, R., Mitchell, P., Field, I., McMahon, C., Burton, H., Blanchard, W., Duignan, P. & Rogers, T. (2005) Estimating Body Mass and Condition of Leopard Seals by Allometrics. *Journal of Wildlife Management*, 69(3), 1015-1023.
- Vaughan, D. G., Marshall, G. J., Connolley, W. M., Parkinson, C., Mulvaney, R., Hodgson, D. A., King, J. C., Pudsey, C. J. & Turner, J. (2003) Recent rapid regional climate warming on the Antarctic Peninsula, *Climatic Change*, 60, 243-74.
- Vaz-Ferreira, R. (1984) La foca leopardo *Hydrurga leptonyx* (de Blainville, 1820) (Pinnipedia, Phocidae) en el Uruguay. *Boletín de la Sociedad Zoológica del Uruguay (2da Época)*, 2, 18-21.
- Venables, W. N. & Ripley, B. D. (2020) *Modern Applied Statistics with S*. 4th ed. New York: Springer.

- Vinding, K., Christiansen, M., Hofmeyr, G. J., Chivell, W., McBride, R. & Bester, M. N. (2013) Occurrence of vagrant leopard seals, *Hydrurga leptonyx*, along the South African Coast. *South African Journal of Wildlife Research*, 43(1), 84–86.
- Vogelnest, L., Edwards, N., Clagila, M., Carlini, A., Slip, D. & Rogers, T. (2010) Anaesthesia of leopard seals, *Hydrurga leptonyx*, on the Western Antarctic Peninsula. *Wildlife Society of New Zealand Veterinary Association, Kokako*, 17, 31-33.
- Vu, V. Q. (2011) ggbiplot: A ggplot2 based biplot. *R package*, 342. *Statistician*, 56(316).
- Walker, T. R., Boyd, I. L., McCafferty, D. J., Huin, N., Taylor, R. I. & Reid, K. (1998) Seasonal occurrence and diet of leopard seals (*Hydrurga leptonyx*) at Bird Island, South Georgia. *Antarctic Science*, 10(1), 75-81.
- Webster, T., Dawson, S. & Slooten, E. (2010) A simple laser photogrammetry technique for measuring Hector's dolphins (*Cephalorhynchus hectori*) in the field. *Marine mammal science*, 26(2), 296-308.
- Wei, T. & Simko, V. (2017) R package "corrplots": Visualization of a Correlation Matrix (Version 0.84).
- Wickham, H. (2007) Reshaping data with the reshape package. *Journal of Statistical Software*. 21(12).
- Wickham, H. (2011) The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1), 1-29.
- Wickham, H. (2016) *ggplot2: elegant graphics for data analysis*. New York: Springer-verlag.
- Wickham, H. & Seidel, D. (2020) Scales: Scale functions for visualization. R package version 1.1. 1.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J. & Kuhn, M. (2019) Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43).
- Wilcoxon, F. (1945) Individual comparisons by ranking methods, *Biometrics*. 1, 80-83.
- Wildman, E. E., Jones, G. M., Wagner, P. E., Boman, R. L., Troutt, H. F. & Lesch, T. N. (1982) A dairy cow body condition scoring system and its relationship to selected production characteristics. *Journal of dairy science*, 65(3), 495-501.
- Würsig, B. & Jefferson, T. A. (1990) Methods of photo-identification for small cetaceans. *Reports of the International Whaling Commission*, 12, 43-52.
- Würsig, B. & Würsig, M. (1977) The photographic determination of group size, composition, and stability of coastal porpoises (*Tursiops truncatus*). *Science*, 198, 755–756.
- Yamashita, R., Nishio, M., Do, R. K. G. & Togashi, K. (2018) Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9(4), 611-629.
- Zhang, X. S. (2000) Introduction to artificial neural network. In *Neural Networks in Optimization*. Boston, MA: Springer.
- Zhong, M., Castellote, M., Dodhia, R., Lavista Ferres, J., Keogh, M. & Brewer, A. (2020) Beluga whale acoustic signal classification using deep learning neural network models. *The Journal of the Acoustical Society of America*, 147(3), 1834-1841.