SMRU Consulting

understand ♦ assess ♦ mitigate

Aerial survey data for monitoring harbour porpoise population health

Authors:	Sinclair, RR & Booth, CG
Report Code:	SMRUC-BEI-2019-003
Date:	Thursday, 17 October 2019

THIS REPORT IS TO BE CITED AS: SINCLAIR, RR & BOOTH, CG. (2019). AERIAL SURVEY DATA FOR MONITORING HARBOUR PORPOISE POPULATION HEALTH. REPORT NUMBER SMRUC-BEI-2019-003 PROVIDED TO BEIS, OCTOBER 2019 (UNPUBLISHED).

Document Control

Please consider this document as uncontrolled copy when printed

Rev.	Date.	Reason for Issue.	Prep.	Chk.	Apr.	Client
1	19/03/2019	Draft 1	RRS	CGB		
2	27/03/2019	Edits post QA	RRS			
3	14/10/2019	Edits following external review	RRS	CGB		

CONTENTS

FIGURES

TABLES

1 Executive Summary

Typically, marine mammal populations are monitored via surveys to determine population size or density. However, these are expensive and, particularly in the case of cetacean populations, tend to provide imprecise estimates which only have the power to detect the large declines. Additionally, for long lived species, it can take a long time before changes in vital rates manifest themselves as changes in population size. As such, it is important to explore alternative approaches for monitoring the early warning indicators of population change.

There are a number of challenges associated with assessing the effects of impact at both an individual and a population level as significant data is required on the baseline population foraging behaviour, life-history strategies, and demographics of populations, which are often lacking for marine mammal species. One alternative to monitoring population size, is monitoring for changes in the age frequency distribution of a population in order which provide early warning signs of changes (i.e. changes in population demographic rates that would later result in changes in population size). Photogrammetric analysis can be conducted using aerial photographs of animals in order to determine animal length and where supporting information on age/length at sexual maturity to assess stage class (immature or mature) of individuals.

The feasibility of this approach for application to the North Sea harbour porpoise population was tested using existing data obtained from aerial surveys conducted by APEM Limited and HiDef Aerial Surveying Limited. Both companies provided a sample of images of harbour porpoise which were measured in order to estimate the immature proportion of the population. Length measurements were obtained for 398 animals and an attempt was made to use these data to assess the stage structure of the porpoise population.

This feasibility study identified the advantages and disadvantages of such an approach. For harbour porpoises, there are a number of considerations. The first is the quality of the images for photogrammetric analysis. The porpoise in the images were often not in the most suitable correct orientation for measuring, their body was curved as they surfaced, obscured by splashes at the surface or images were too pixilated or lacked the contrast to be able to confidently locate the tip of the rostrum and the notch in the tail fluke for measurement. Therefore, images were graded for quality. Of 398 images, 55% were classified as grade 1 or 2; the remaining images were categorised as grade 3 (considered less suitable for photogrammetric analysis. Most of the animals measured

were between 110 and 150 cm in length, with a range between 69 cm and >200 cm. We determined that using images to measure body condition was not viable.

The second limitation of this approach for harbour porpoise was a consequence of the sexual dimorphism in the species, where females grow to larger maximum sizes and mature at larger sizes than male porpoise. This was exacerbated by a wide range of length-at-sexual maturity estimates are in the published literature (varying with both sex and geographically). This made it challenging to determine a suitable threshold for identifying mature vs immature animals. Based on the published literature, it was assumed that animals <127 cm in length were all immature, animals ≥152 cm were all mature, but that animals between 127 and 152 cm would be comprised of a mixture of large immature males, mature males, immature females and small mature females. Since the sex of the animal in each image is unknown, it was not possible to determine the stage class of those animals between 127 and 152 cm (44% of the sample). Depending on the length threshold used, the proportion of immature animals in the sample photographs provided ranged between 34 – 80%.

It is important to stress that the purpose of this project was a feasibility study of whether body condition or the proportion of juveniles could be assessed using digital aerial survey methods. Uncertainties also remain more generally about the proportion of immature animals in healthy and unhealthy populations.As with any monitoring methods, sources of error were identified and, where possible, accounted for in the measurement process. For example, not all images were provided with a calibrated Ground Sampling Distance (taking account of the exact camera height at the time of the photograph). This is important since it is the Ground Sampling Distance that is used to scale the measurements from pixel length to animal length. Going forward, it is essential that all measurement errors are reduced as much as possible, and therefore image-specific ground sampling distance would be required to minimise this source of error. Another error source is human based, and is driven by how well the user identifies the tip of the rostrum and the notch of the tail fluke for measurement. Therefore, multiple users independently measured a subsample of images in order to assess inter-user measurement error. For most images, the difference in measured length between the users was low, however, images of poorer quality resulted in larger inter-user measurement differences.

Even when measurement error is minimised, quantified and accounted for, there still remains the difficulty in assigning stage structure to the sample due to the considerable variation in the literature in the estimated age or length of sexual maturity of harbour porpoise. It was therefore concluded from this feasibility study that while this method cannot currently be used to reliably assess the stage structure of the North Sea harbour porpoise population, it can provide a means by which to measure animal length and estimate the minimum proportion of juveniles in a population which may be a potential way towards monitoring population health.

2 Introduction

Typically, animal populations are monitored via surveys to determine population size or density and for marine mammal populations, methods such as telemetry-corrected haulout counts for pinnipeds (Thompson and Harwood 1990) and line-transect surveys for cetaceans (Wade and Gerrodette 1993) are employed. However, these are expensive and, particularly in the case of cetacean populations, tend to provide imprecise estimates which only have the power to detect the most drastic declines (Taylor et al. 2007, Jewell et al. 2012). Additionally, for long lived species, it can take a long time before changes in vital rates manifest themselves as changes in population size. As such, it is important to explore alternative approaches for monitoring the early warning indicators of population change, to provide more sensitive metrics of declines (Booth et al. 2017).

Assessing the impact of noise on marine mammal species is of critical importance to various stakeholders including industry, who must balance project objectives with the potential environmental impacts of proposed activities. However, there are many difficulties associated with assessing the effects of impact at both an individual and a population level. In order to implement frameworks to assess the population level consequences of impacts, substantial pre-existing knowledge is required of the baseline foraging behaviour, life-history strategies, and demographics of populations. However, for many marine mammal populations, current knowledge is lacking and such 'data poor' situations mean that any such forecasts of population level impacts have significant uncertainty associated with them. Given these uncertainties there is merit in identifying the data gaps that need to be filled in order to better parameterise population models. However it may take decades to fill these gaps and, in the meantime, undetected population declines may occur. Therefore we are interested in finding variables we can monitor to provide an early warning of decline.

Holmes and York (2003) used time-varying matrix models of Steller sea lion (*Eumatopias jubatus*) populations to determine whether age-structure information could be used to detect changes in survival, fecundity and population status. Incorporating the 'juvenile fraction' (akin to the proportion of immatures) along with count data improved their ability to detect changes in demographic parameters.

Booth et al. (2017) identified methods for monitoring populations that might be subject to disturbance that may also provide insights into the processes through which disturbance may affect these populations. In addition, Booth et al. (2017), using existing PCoD models, determined that changes in certain demographic variables are strongly correlated with changes in abundance or population status, and can therefore provide some early warning of future changes in abundance. The study simulated a harbour porpoise (*Phocoena phocoena*) population of 10,000 animals, exposed to an intermittent disturbance for 10 years and reported the relationship between the maximum reduction in population size and the proportion of immature animals. In 90% of the simulations where a decline in the proportion of juveniles was detected in year 5, the population had declined by more than 40% by the end of the disturbance period (Figure 1). While environmental variation affects the strength of this signal, there is a general pattern of increasing reduction in population size associated with decreasing proportion of immatures in the population. The data suggest that a reduction in population size is reasonably small for most scenarios where the proportion of juveniles was ~0.25 or higher (though there were still cases where a larger decline occurred, but this was rare). Therefore it was concluded that the proportion of immature animals in a population might provide a reasonable early indicator of population decline. Booth et al. (2017) note that in practice, the number of false positives will be reduced by monitoring for multiple demographic parameters simultaneously.

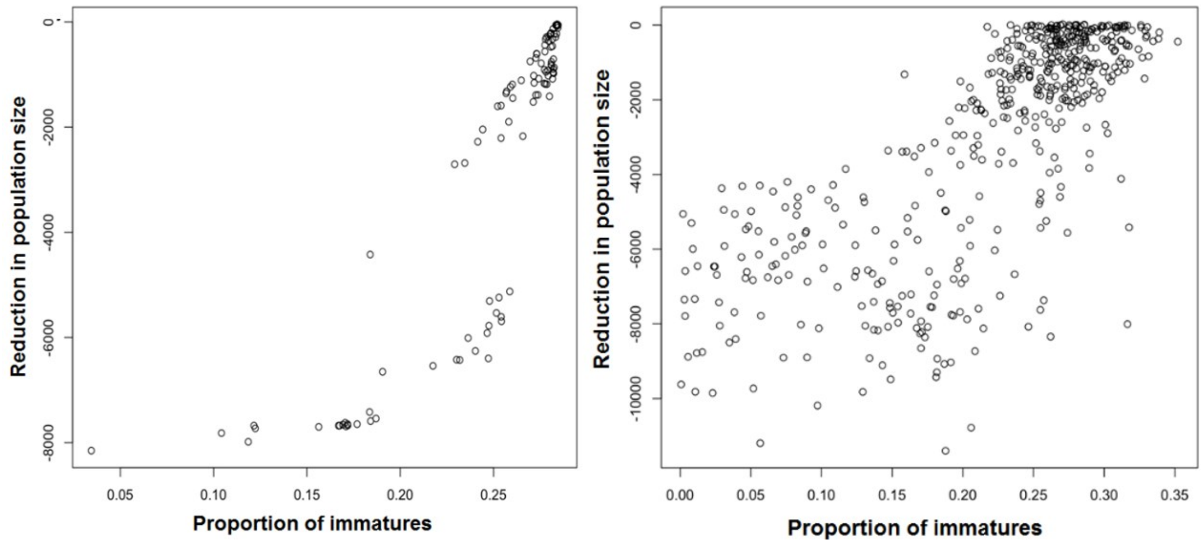

Figure 1 Relationship between the maximum predicted reduction in harbour porpoise population size the proportion of immature animals. Left = without environmental variation in the simulation and right = with environmental variation in the simulation.

Age frequency distributions (also known as population pyramids) are used to assess the age structure of populations that have differing trajectories – typically in humans. For example, an expanding (growing) population is likely to have a large proportion of individuals in the younger age groups, compared to a contracting (declining) population which is likely to have a smaller proportion of the population in the younger age groups (Figure 2) (Staetsky and Boyd 2015). In bird species, studies have shown differences in age frequency distributions between decreasing and increasing populations. For example, snail kite populations in Florida have shown differing trends in different regions, where the population has been increasing in the northern region and declining in the southern region. A study on the stage structure between the two regions has shown that in the southern region the proportion of older senescent animals has increased as the population size has decreased, while the northern region had a larger proportion of immature animals and this generally increased as the population size increased (Reichert et al. 2016). Thus monitoring for changes in the age frequency distribution of a population can be used to identify early warning signs of changes in population demographic rates that would likely lead to changes in population size.

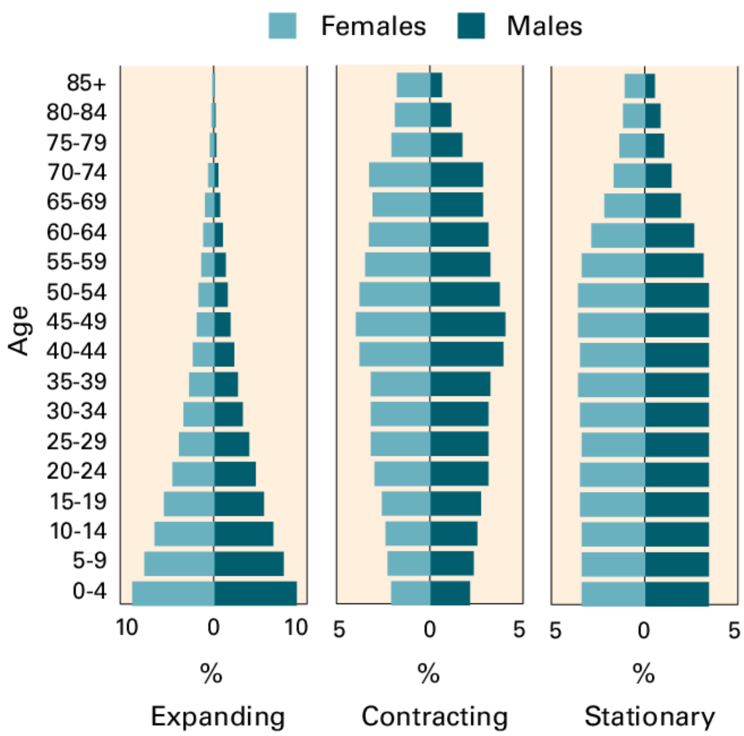

Figure 2 Examples of different types of population pyramids for humans. Obtained from Staetsky and Boyd (2015).

Booth et al. (2017) observed that marine mammal demographic parameters tend to be most commonly estimated from monitoring that uses established approaches such as visual surveys. Visual and digital aerial surveys have become an established method for both baseline and post-consent monitoring in the UK, mainland Europe and off the east coast of the USA. In recent years, these surveys are often carried out using photographic or videographic methods providing a top-down image or video of sighted animals whilst animals are close to the surface. The resolution of the images/videos from these surveys typically range between 2 and 3 cm. These 'digital aerial' survey approaches use high-resolution cameras and therefore provide high definition images/video which are sufficient to identify many species of marine mammals to a species level to be used in the generation of density estimates. However, the data collected during these surveys have not been investigated as a source of data for monitoring individual and population health via assessments of stage structure and body condition (via morphometric measurements).

Therefore, the aim of this pilot study was to explore the viability of digital aerial survey sightings data as a method by which stage-structure (including the proportion of immature animals in the population) can be derived and whether it is possible to generate robust estimates of body condition via this method.

3 Methods

3.1 Digital aerial survey images

There are two companies in the UK that provide digital aerial surveying of seabirds and marine mammals. These are APEM Limited[1] and HiDef Aerial Surveying Limited[2]. The two companies differ in their methodologies, for example APEM collect digital still images while HiDef collects digital video.

Both APEM Limited and HiDef Aerial Surveying Limited were asked to provide 200 good quality aerial survey images of harbour porpoise. Since HiDef collect digital video during their surveys, they provided single frames from the video footage. In order for an image to be suitable for measurement it had to be clear and un-pixilated (as far as possible), for the animal to be at the

[1] https://www.apemltd.co.uk/

[2] https://hidef.bioconsult-sh.de/

surface dorsal side up with a straight body axis and for there to be no splashes or glare obscuring the animal in the image.

Both companies found it difficult to provide 200 images that met the specified quality in terms of animal orientation, clarity and without any splashes or glare obscuring view of the animal. For example, one company examined 580 images of harbour porpoise but only selected 23 images as being of the required quality (2.7%). Therefore the requirements were relaxed slightly and both companies provided the best quality images available. In order to account for this, when the images were analysed a simple grading system was implemented based on animal orientation, contrast and pixilation level (see examples in Figure 3), as follows:

- Grade 1 - Images in which the animal was flat, clearly defined and unobscured.
- Grade 2 - Images in which the animal was slightly pixilated / blurred and the measurer found it slightly more difficult to pinpoint the best pixels to measure.
- Grade 3 - Images in which the animal was off angle or there was particularly bad pixilation/blurring/obscuring of the animal.

It was noted where the animal was slightly off angle which would influence the length measurement and where the measurer found it particularly difficult to obtain a measurement due to pixilation or contrast issues.

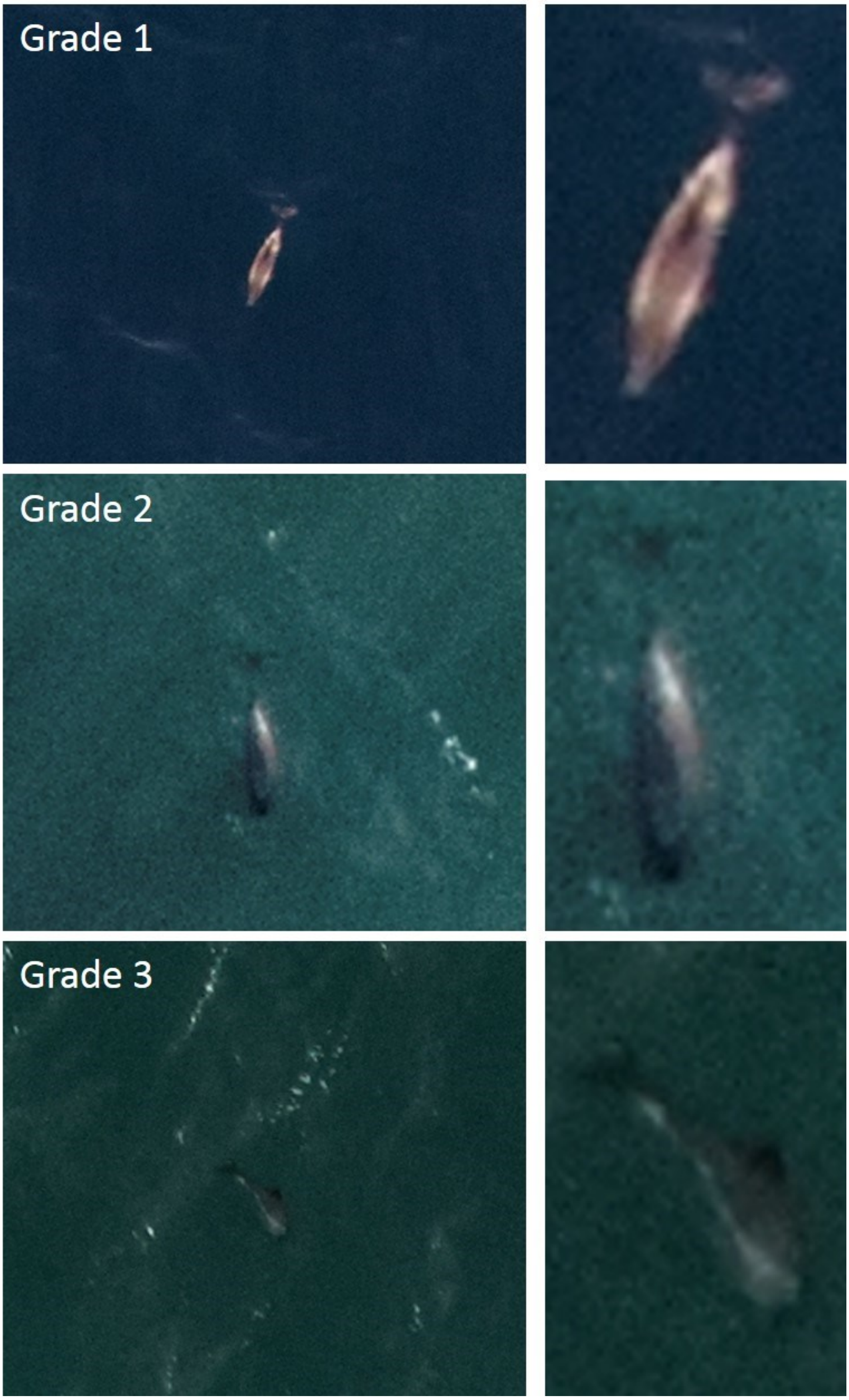

Figure 3 Examples of different grade images. Grade 1 is ideal with a flat animal and little blurring/pixilation. Grade 2 is more blurry/pixilated and so more difficult to pinpoint measurement locations. Grade 3 is both blurry/pixilated and the animal is off angle.

3.2 Morphometric measurements

In order to assess the stage structure of the population, the lengths of animals were measured and combined with length-at-maturity data to categorise animals as either mature or immature based on length. In addition, width measurements were taken to attempt to estimate body condition. Two different software packages were trialled to obtain animal measurements: 'whale.morpho' (written in R[3] from Christiansen et al. (2016)) and ImageJ[4]. Whale-morpho is a function in R specifically designed to obtain animal morphometric data from aerial photographs, while ImageJ is an image processing toolkit in Java.

3.2.1 Length

To determine the most appropriate (and feasible) morphometric measurements to extract for harbour porpoises, an exploratory analysis using the 'whale.morpho' function in R was used. This approach was developed to extract morphometric measurements of humpback whales from aerial photographs (Christiansen et al. 2016). As a standard, the following metrics are extracted from each image (see Figure 4):

- *Rostrum.X:* The distance in X between the rostrum and the bottom-left corner of the full picture (in % of total width (X) of the photo).
- *Rostrum.Y:* The distance in Y between the rostrum and the bottom-left corner of the full picture (in % of total height (Y) of the photo).
- *Fluke.X:* The distance in X between the notch of the tail fluke and the bottom-left corner of the full picture (in % of total width (X) of the photo).
- *Fluke.Y:* The distance in Y between the notch of the tail fluke and the bottom-left corner of the full picture (in % of total height (Y) of the photo).
- *Total.length.pix:* Total length of the whale (in pixels) from rostrum to notch of the tail fluke.
- *Length.to.blowhole.pix:* The length between the rostrum and the blowhole of the whale (in pixels).
- *Length.to.start.of.dorsal.fin.pix:* The length between the rostrum and the beginning of the dorsal fin of the whale (in pixels).
- *Length.to.end.of.dorsal.fin.pix:* The length between the rostrum and the end of the dorsal fin of the whale (in pixels).
- *Length.to.start.of.fluke.pix:* The length between the rostrum and the start of the tail of the whale (in pixels).
- *Length.to.eyes.pix:* The length between the rostrum and the line of the eyes of the whale (in pixels).
- *Width.at.eyes.pix:* The width between the eyes of the whale (in pixels).

[3] https://www.r-project.org/
[4] https://imagej.nih.gov/ij/

- *Width.fluke.pix:* The length between the two extremities of the tail fluke (in pixels).
- *Width.5.proc.pix to Width.95.proc.pix:* Width measurements of the body of the whale at 5% intervals along the entire body (in pixels).

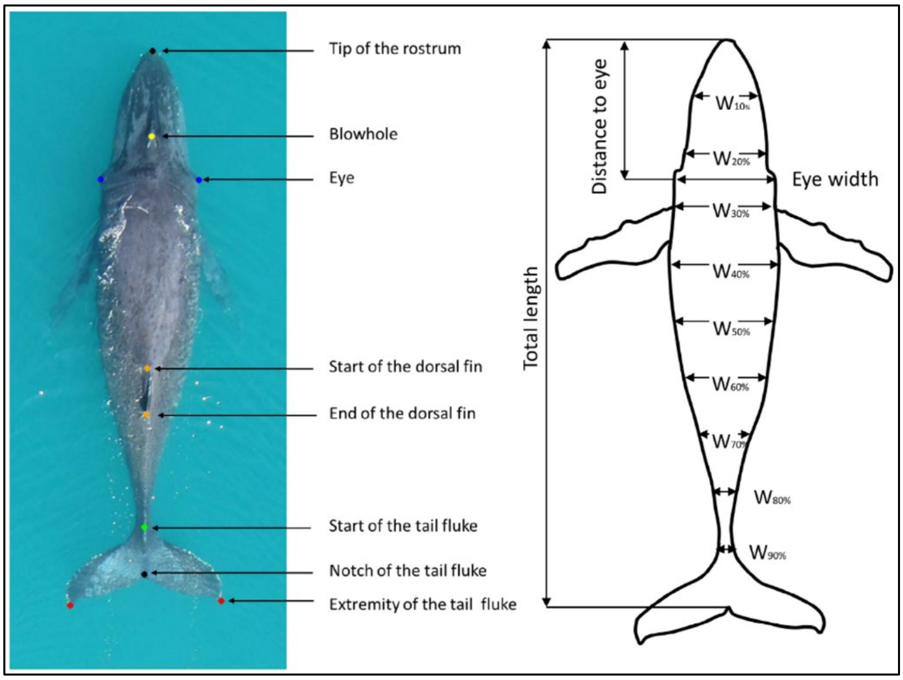

Figure 4 Position of measurement sites of humpback whales recorded using the whale.morpho function R scripts. Only width (W) measurement sites located at 10% increments along the body axis are shown. Figure taken from Christiansen et al. (2016).

Not all of these measurements were required (or possible) for the images of harbour porpoise in this project. For example, the location of the eyes, the blowhole and the start and end of the dorsal fin was not discernible in the majority of images due to the morphology of porpoise and the resolution of the images (Figure 5).

Whilst the whale.morpho code script provided a number of measures, only required length and girth was required for this study. In addition, making measurements was found to be easier in ImageJ. Therefore it was decided that all images would be processed in ImageJ.

ImageJ has previously been used for image-processing in photogrammetry studies (e.g. Rowe and Dawson 2009, Cheney et al. 2017). Each image was scaled according to the Ground Sampling Distance (GSD) at the time of the shot (in pixels/cm). Therefore each resulting measurement represented a true distance (at the surface).

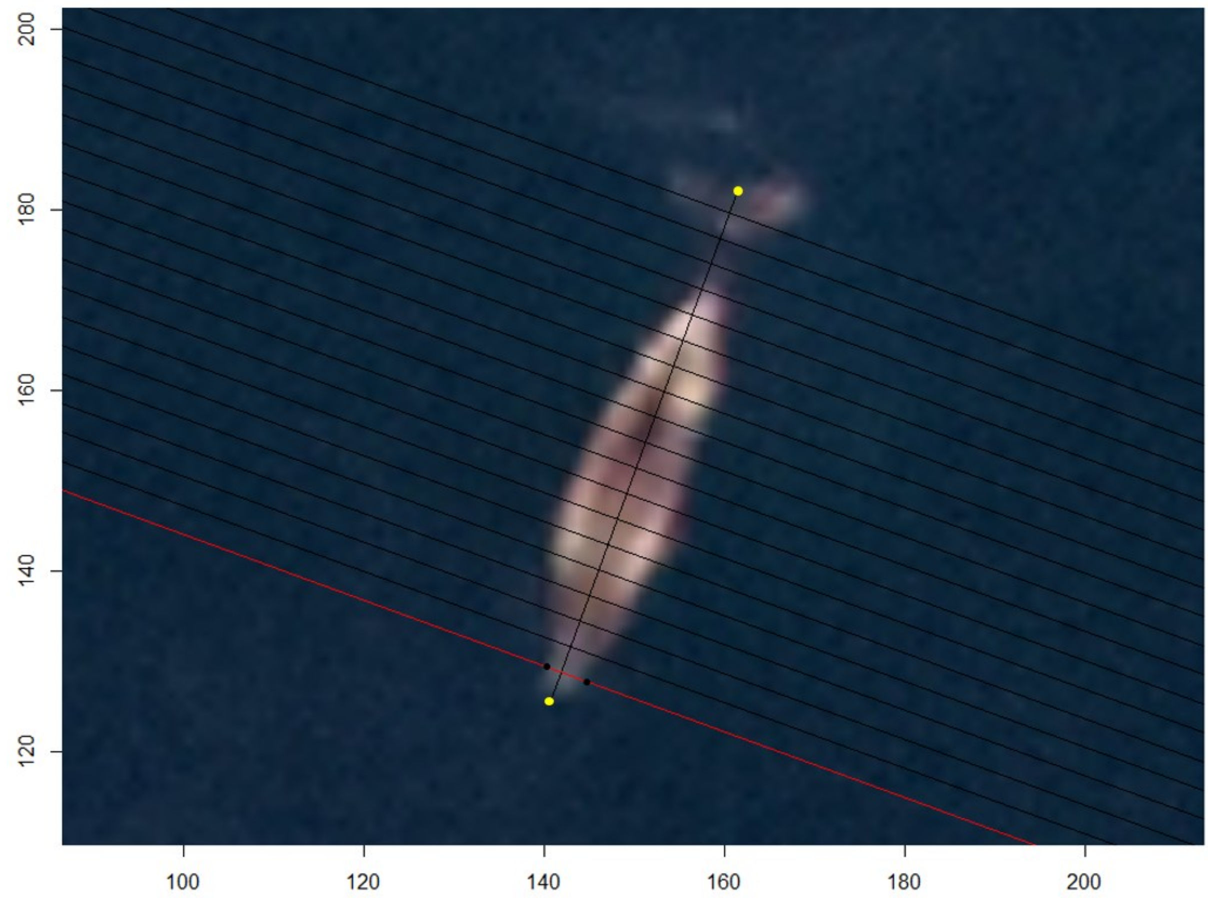

Figure 5 Example image analysis in R using the whale.morpho function. Yellow dots represent marks for the tip of the rostrum and notch of fluke, lines represent width measurements at 5% intervals, red line represents the current width under analysis.

3.2.2 Width

Previous studies have used girth at axilla (immediately behind the pectoral fin) (e.g. Kastelein et al. 2018b) as a measurement for assessment of body condition in captive porpoise (and other morphometric measurements including body mass). However, in this study, the pectoral fins were not clearly visible in the majority of the images and therefore it was hard to identify the axilla of the porpoise. In order to obtain width measurements of the porpoise in the images, a total of 24 images were measured in 'whale.morpho' in order to determine the region of the body in which porpoises were widest. This provided the result that most porpoise in the sample (67%) were widest at the 30 or 35% along the body length from the rostrum (corresponding to *Width.30.proc.pix* or *Width.35.proc.pix* in 'whale.morpho'). This therefore identified approximately where to take a width measurement for each porpoise when measured in ImageJ, to obtain the widest point for the assessment of body condition.

For each image (where possible), the length (defined as the distance in cm between the tip of the rostrum and the notch in the fluke - Figure 6) and width (defined as the widest part of the body between the pectoral fins and the dorsal fin, in cm – using the 30-35% measure as a guide) were measured.

3.2.3 Measurement Error

All length and width measurements were made by a single observer. In order to assess measurement error, a random sample of 20 images (21 animals as one image contained two animals) were independently measured by five users (the single observer and four additional users) using ImageJ. Since we do not have absolute measures of the animals (i.e. none of the porpoise were caught and measured in the field), the true length of each animal is unknown. Therefore the error measurements are not error from truth, but show how measurements can differ between different users when using the same image and the same software to measure an image.

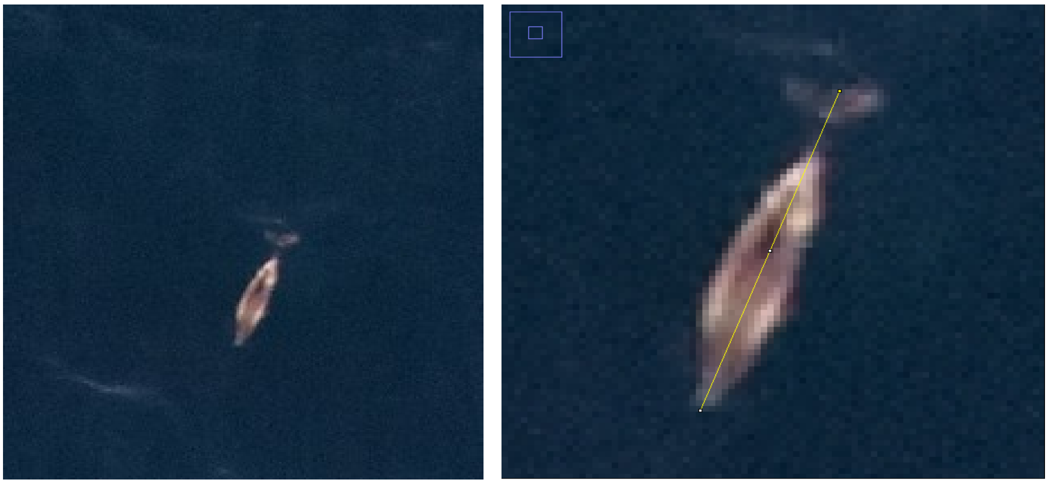

Figure 6 Example of a good quality image (grade 1) measured in ImageJ. Left: image extent, right: zoomed in showing measurement points for length.

3.2.3.1 Ground Sampling Distance

GSD essentially sets the scale for each image as it provides the true distance in cm that is represented by 1 pixel in the image. As flight height varies slightly both within and between surveys, the GSD will therefore change slightly between images. Both companies were asked to provide the GSD for each image, in order to correctly set the scale for measurements.

3.2.3.2 Comparison to automated software

Some of the images provided came with a length measurement which had been obtained by the data-provider's automated software. Therefore this provided the opportunity to assess the differences in length measurements obtained by hand (i.e. in this study) and those by the automated software (from the data provider). A total of 186 images had a length measurement obtained by hand in ImageJ and by the automated software.

3.3 Assessing stage-structure

In order to assess the stage structure of the harbour porpoises sampled, information is required on 'length-at-age' and estimates of the age of sexual maturity. A review of the published and grey literature was carried out to better understand the length and age of sexual maturity in harbour porpoise in the North-east Atlantic. The review identified suitable data sources for harbour porpoise from Scottish waters (Learmonth et al. 2014), British waters (Lockyer 1995a), Danish waters (both North Sea and Kattegat/Skaggerak regions) (Lockyer and Kinze 2003), Icelandic waters (Ólafsdóttir et al. 2003), waters west of Greenland (Lockyer et al. 2001), Norwegian waters (Karstad 1993) and Swedish waters (Karstad 1993, Kull and Berggren 1993, Berggren 1995). While there are data available for harbour porpoise in other areas, such as Newfoundland, the Sea of Azov and the Black Sea, it was judged that these locations were too distant from UK waters to be considered representative of UK harbour porpoise.

The studies were mainly conducted on stranded and bycaught animals, though the Lockyer and Kinze (2003) study also included direct catches of porpoises. In each case the carcass was recovered, sex was established and morphometric measurements were obtained. The length of the animal was measured in a straight line between the tip of the rostrum and the notch in the tail fluke. The age of animals was estimated by obtaining teeth from the middle of the lower jaw and by counting the growth layer groups where one year's growth consists of two complete layers, one opaque and one transparent. Using these two metrics, the studies produced length-at-age curves for both male and female porpoise. These length-at-age curves are presented in 'Appendix 1: Length-at-age curves'.

Female porpoise were categorised as mature if the ovaries contained at least one corpus luteum/albicans. Male porpoise were categorised as mature based on testes weight (mature = testes weight ≥200 g). By combining this data with the length data, the studies provided Length at Sexual Maturity (LSM) and Age at Sexual Maturity (ASM).

4 Results

4.1 Morphometric measurements

4.1.1 Length

Not all 400 images were suitable for length measurement, and therefore a length measurement was obtained for a total of 398 animals. Of these, 219 were grade 1 or 2 (55%). All grade 1-3 images were included in the analysis in order to maximise the sample size, however results are divided into groups for grade 1-2 and grades 1-3 to assess the effect of image quality on the results. There were only nine animals in the sample that had a measured length <100 cm (and two of them were obviously calves alongside an adult (Figure 7) (the smallest length recorded was 69 cm from the calf in Figure 7a). Most of the porpoise measured (71%) were between 110 and 150 cm (Figure 8) when all grades were considered. For those images graded as 1 or 2, a similar pattern was seen, where 69% of the images were between 110 and 150 cm (Figure 8). The largest length measured was 257 cm. This is larger than previous maximum length measurements obtained from the literature review (Table 1) where the maximum length recorded from stranded and bycaught animals was 189 cm for female porpoise in British and Danish waters. Only 35 of the animals in this sample (8.8%) were measured as >190 cm, and of these 16 were considered to be low grade images where length measurements were more challenging to make, and thus confidence in the resulting length measurement was low. The unusually high length measurements in these images may have been due to misclassification of species identification, measurement error in the photogrammetric data processing or error in the GSD used to scale the image.

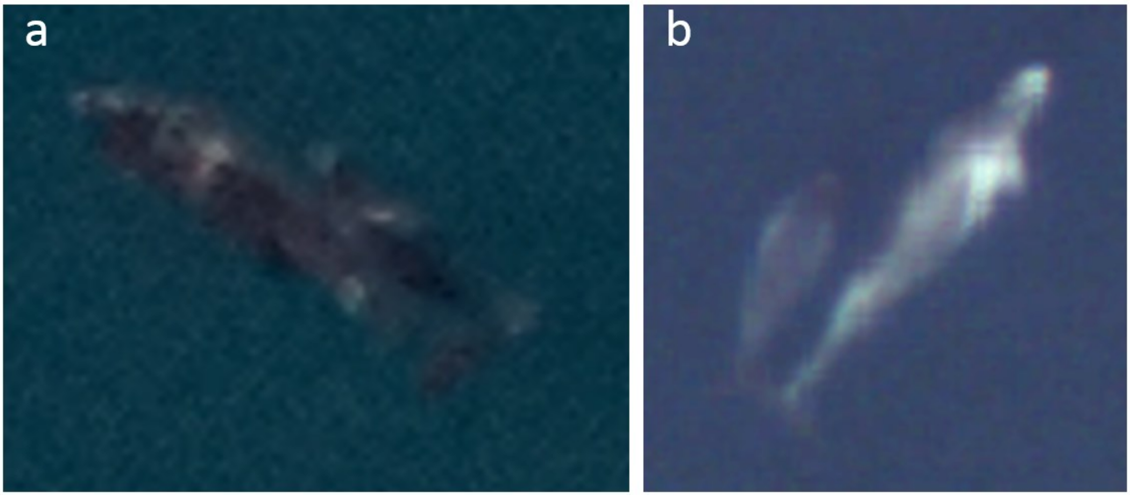

Figure 7 Images of a porpoise calf alongside an adult. (a) calf = 69 cm, (b) calf = 90 cm.

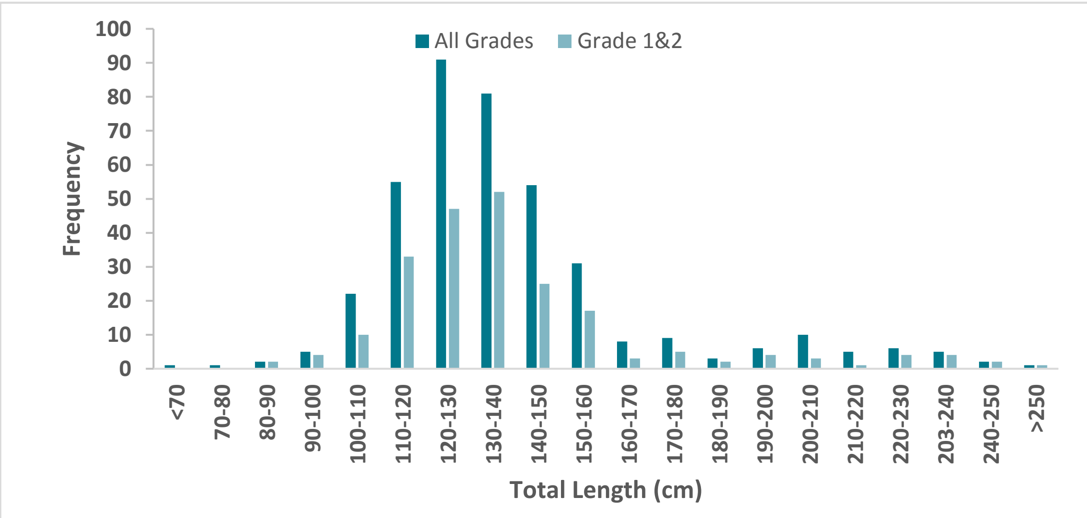

Figure 8 Length measurements for 398 harbour porpoise (all grades), of which 219 animals were grade 1 or 2.

Table 1 Maximum harbour porpoise lengths obtained via the literature review. F – female, M – male.

Reference	Location	Maximum Length (cm)	
		F	M
Lockyer (1995a)	British waters	189	163
Lockyer and Kinze (2003)	Inner Danish waters + North Sea	189	167
Berggren (1995)	Swedish Skagerrak and Kattegat	167	154
Lockyer et al. (2001)	West Greenland	166	158
Ólafsdóttir et al. (2003)	Iceland	174	165
Møhl-Hansen (1954) & van Bree (1973) – referenced in Lockyer (2003)	Baltic Sea	189	169

4.1.2 Width

A width measurement was obtained for 353 of the porpoise images. The remaining images were not measured for width due to incorrect orientation of the animal or insufficient image quality to determine the edge of the animal with any confidence. Width ranged between 18 and 67 cm, with an average of 33 cm, and most animals had a width measurement between 25 and 40 cm (Figure 9). As expected, animal width increased with increasing animal length and a linear trend line (with a resulting R^2 value of 0.70 or 0.73) was fitted to Grade 1-3 or Grade 1&2 data respectively (Figure 10).

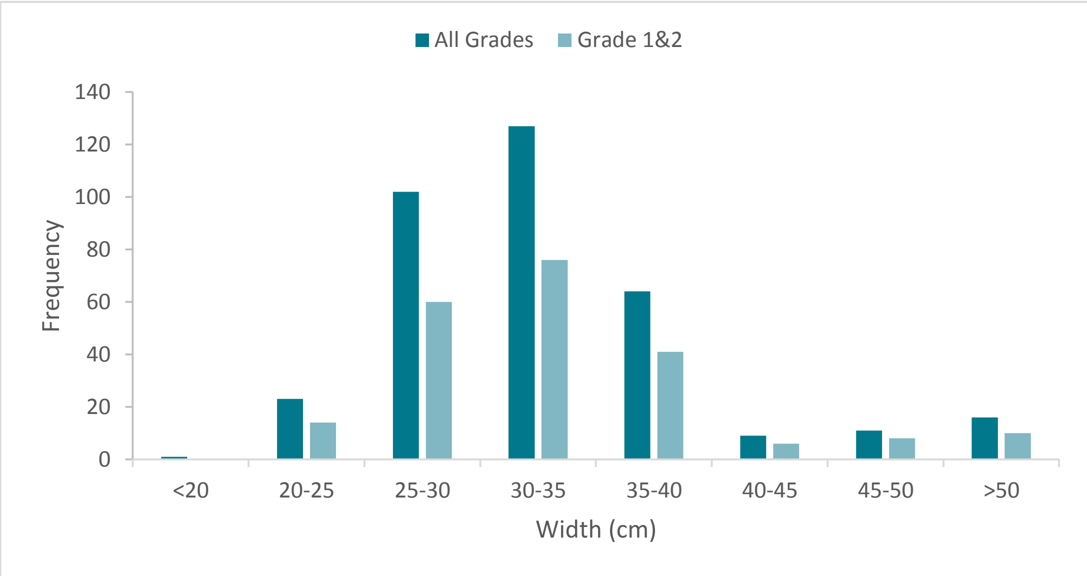

Figure 9 Width measurements for 353 harbour porpoise (215 of which were Grade 1&2).

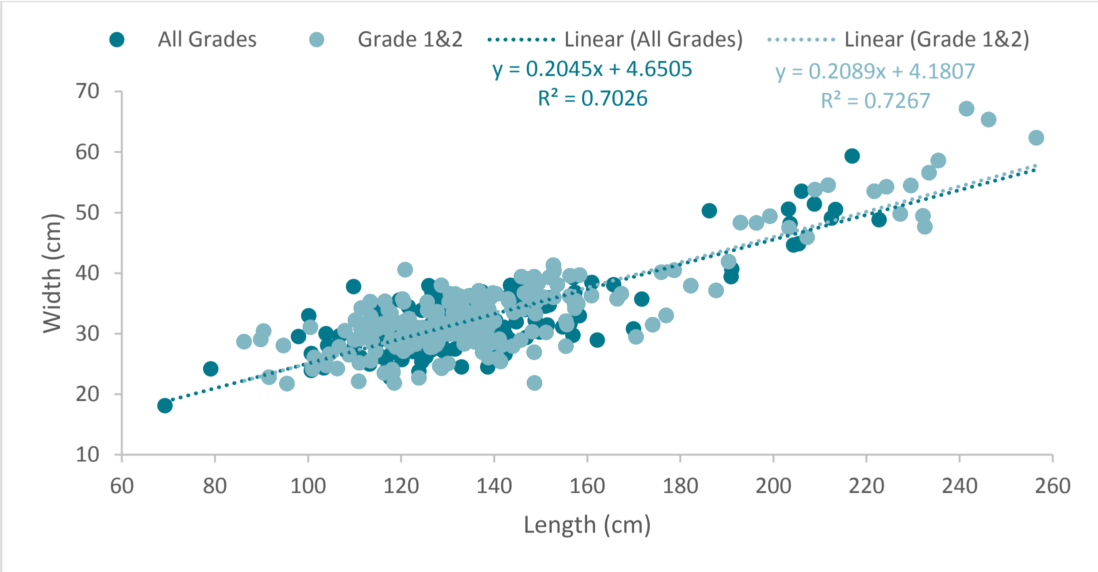

Figure 10 Relationship between length and width.

4.2 Measurement error

There are considered to be several sources of measurement error in the measures taken from the images. Errors can arise from the quality of the image itself, if contrast, pixilation or obscuring features affect the measurers ability to accurately place measurement positions. In addition, the surfacing behaviour of harbour porpoise resulting in body curvature will result in errors in the length measurements. Details on the limitations due to image quality and animal orientation/position are discussed in section 5.3. Error can also occur where the observer was incorrect in their placement positions for the measurements (resulting in variation within and between individual observers). Another source of error is in the scaling for the images, resulting from an inaccurate ground sampling distance (e.g. driven typically by varying flying height). Inter-user measurements and GSD errors and their effects on the resulting measurements are outlined below.

4.2.1 Inter-user error

For most images the length measurement error was reasonably low, but there were some images in which the measurement error was markedly higher (Figure 11). The average range in length measurement between the five users across the 21 animals was 6.6 cm, and 43% of the animals had an inter-user measurement error range within 5 cm. The most consistent measurement across observers resulted in a range in length measurements of 2.5 cm (animal 18) and the widest range in length measurement among observers was 15.0 cm (animal 21) see images in Figure 31 and Figure 32 in Appendix 2). The discrepancy between users for animal 21 was due to the presence of a splash/bow wave at the head of the porpoise as it surfaced, which partially obscured visibility of the animal and made the rostrum point difficult to identify (also see animals 6 and 20 where the length range was high in Table 6 in Appendix 2).

We explored further the images with a wide range in inter-user measurements and this range in measurements was found to be driven by one user providing a length measurement that was considerably different to the other four users; however it was not consistently the same user that was an outlier. By removing any outlying measurements (those >3 cm from the next closest measurement) significantly improved the precision of the measurements by observers as the average range in length measurement across the 21 animals was 3.8 cm (range between 1.2 and 7.4 cm), and 86% of the animals had an inter-user measurement error range within 5 cm (Figure 12).

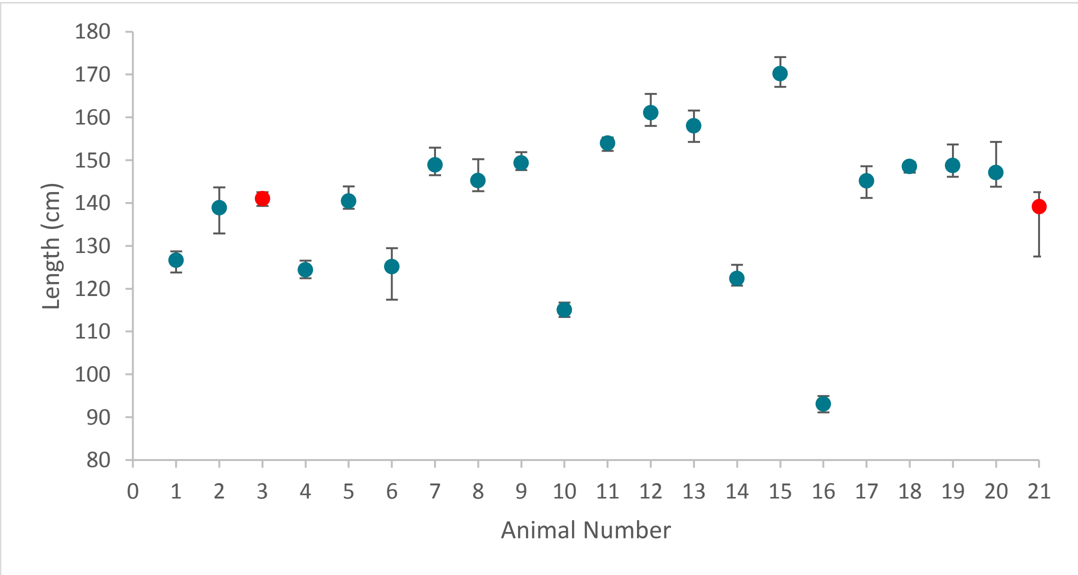

Figure 11 Average length measurement for each of the 21 animals measured. Error bars show the range in measurement for the five users for each animal. All images were Grade 2 except animal number 3 and 21 which were Grade 3 (coloured red).

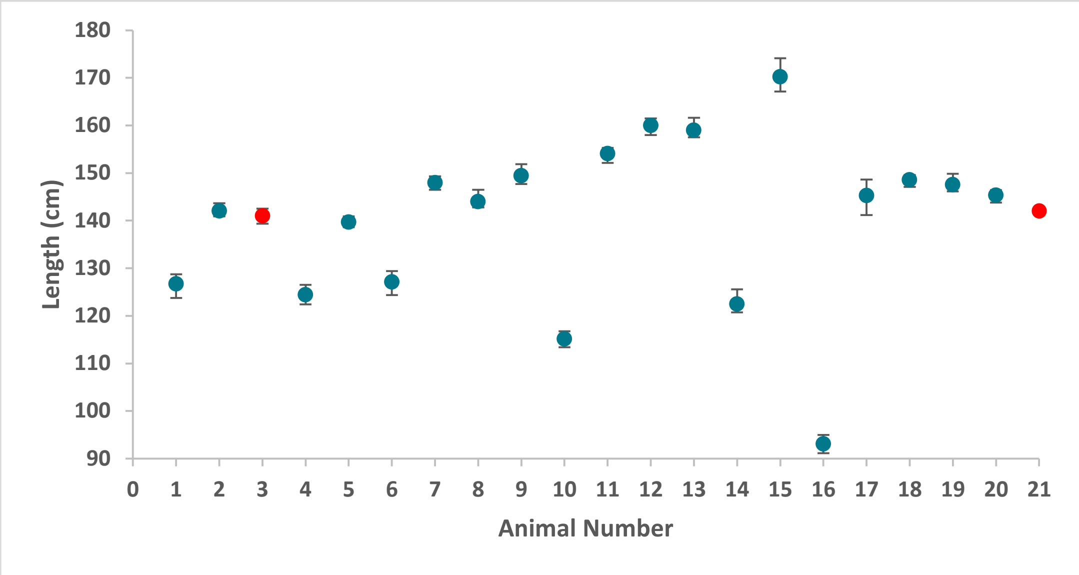

Figure 12 Average length measurement for each of the 21 animals measured, excluding the outlying measurements (defined as being >3cm from the next closest measurement). Error bars show the range in measurement between the users for each animal. All images were Grade 2 except animal number 3 and 21 which were Grade 3 (coloured red).

As per the length measurements, measurement error was investigated with the width measurements and in general, the inter-user measurement error was greater in width measurements than in length (Figure 13). The most consistent measurements were 2.0cm (animal 10) and widest range was 9.7 cm (animal 17). In addition, all observers reported that they found the width measurements harder to obtain than the length measurements. This was primarily because the length measurement required a more defined and obvious point of measurement: the tip of the rostrum and the notch of the fluke. Conversely, the target width measurement was between ~30-35% along the body length from the head (approximately between the pectoral fins and the dorsal fin), and this position was difficult to determine in many images as neither the pectoral fins nor the dorsal fin were clearly visible, or the body orientation was unsuitable (see **Error! Reference source not found.**in Appendix 2).

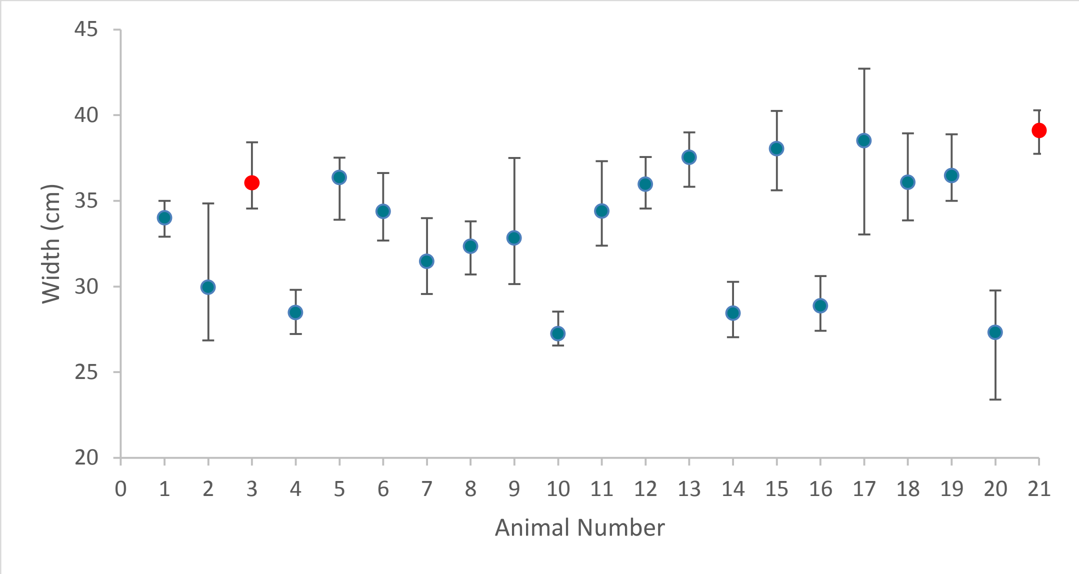

Figure 13 Average width measurement for each of the 21 animals measured. Error bars show the range in measurement between the five users for each animal. All images were Grade 2 except animal number 3 and 21 which were Grade 3 (coloured red).

4.2.2 Ground Sampling Distance

One company provided GSDs in discrete categories (as integers) and the other company provided a calibrated GSD (to 2 decimal places) for each image (calibrated using the aircraft flight height and which camera was used during the survey).

An inaccurate GSD is a source of error in the measurements as it is the scale by which the image is measured. Therefore the effect of differing GSD on resulting length measurements was explored (Table 2)**Error! Reference source not found.**. Differences in measurement length with varying GSD were measured for 20 images, using both the calibrated GSD and the 2 cm integer GSD. The resulting difference in length measurement in these 20 images ranged between 0.64 – 10.19 cm, and where the calibrated GSD varied considerably from 2 cm (e.g. 2.16 cm), the resulting difference in length measurement was large (10.19 cm).

Table 2 Differences in length measurements when assuming GSD=2cm and when GSD is calibrated.

Measured length (pixels)	Calibrated GSD	Porpoise length (cm)		
		GSD = 2cm	GSD = calibrated	Difference
81.56	1.93	163.12	157.41	5.71
72.33	1.98	144.66	143.21	1.45
63.76	2.01	127.51	128.15	0.64
63.70	2.16	127.40	137.59	10.19
77.19	1.98	154.37	152.83	1.54
50.94	2.12	101.89	108.00	6.11
90.23	1.95	180.45	175.94	4.51
68.51	1.9	137.01	130.16	6.85
73.75	2.12	147.49	156.34	8.85
66.77	1.88	133.53	125.52	8.01
70.58	1.88	141.17	132.70	8.47
68.55	1.92	137.11	131.62	5.48
61.68	1.91	123.36	117.81	5.55
70.36	1.93	140.73	135.80	4.93
63.43	1.91	126.86	121.15	5.71
74.53	1.88	149.05	140.11	8.94
60.54	1.94	121.08	117.45	3.63
72.28	1.92	144.55	138.77	5.78
71.56	1.98	143.12	141.68	1.43
80.09	1.98	160.17	158.57	1.60

4.3 Comparison to automated software

For the subset of images measured by hand for which there were also automated measurements (n=186), 53% resulted in hand measured lengths that were within 5 cm of the automated length measurement, and 80% were within 10 cm (Figure 14). A further examination of the images with >10 cm difference found that, in general, these images had been assigned a low grade due to issues with distortion, pixilation, contrast, body orientation etc. Therefore there was a low level of confidence associated measurements. A selection of the images where there was >10 cm difference in length measurement are illustrated in Figure 15, which clearly shows that the quality of the images was low, with contrast and obstruction issues.

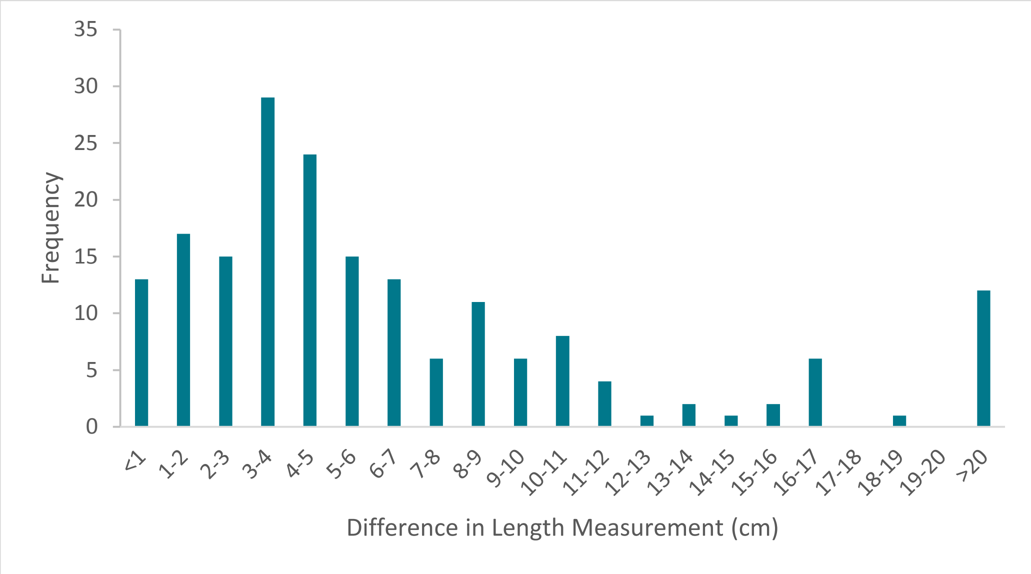

Figure 14 Difference in length measurement (cm) between the automated software measurement and the single observer hand-measured lengths obtained in this project (n= 186).

Figure 15 Example images where the difference between the single observer hand-measured length and the automated length measurement differed by >10 cm. These images were all Grade 3.

4.4 Stage structure

The review of the length and age at maturity literature clearly showed that there is sexual dimorphism in harbour porpoise, where female porpoise grow to larger maximum sizes and mature at larger sizes than male porpoise. In addition, the length at which animals become sexually mature differs between both sexes and between studies (Table 3 and 4). For example, the mean LSM for females ranged between 138 and 152 cm between these studies, and for males ranged between 127 and 135 cm. In most studies female porpoise were estimated to reach sexual maturity at a slightly older age than males, and overall porpoise are estimated to reach sexually maturity between age 2 and 5 years old. Since there are insufficient data on the population structure or the ranging patterns of harbour porpoise in the North Atlantic and the fact that there may be sampling biases within each of the different studies, it is unknown which of these estimates of LSM are likely to be most applicable to harbour porpoise round the UK where the images in this study were obtained. Therefore the results below present a variety of stage class assessments based on this range of LSM estimates.

The difficulties when assigning stage/age class to each animal is the sexual dimorphism in harbour porpoise, where females mature at larger sizes (138-152 cm) compared to males (127-135 cm) and because there is significant geographic variation in length-at-age. Since the sex of the animal in each image is unknown, it is not be possible to determine the stage class for many of the images. For example, it is unknown if an animal of 135 cm was a mature male or an immature female. Therefore depending on the length threshold for maturity used, different estimates of the proportion of immatures are generated (Figure 16) below.

Table 3 Length at sexual maturity (LSM) information obtained from the literature review. F – female, M – male, na – not available.

Reference	Location & Sampling Period	Source	Sample size	Average LSM (cm)		Notes
				F	M	
Learmonth et al. (2014)	Scotland (1992-2005)	Stranded & bycaught	F=190 M=145	138.8 (95% CI: 135.9-141.6)	132.2 (95% CI: 129.1-135.6)	The smallest mature female was 3 years old and 119 cm long, which was considered by the authors to be unusual. Excluding this animal, mature female body length ranged from 137 cm to 173 cm. For males, excluding the youngest (and smallest) animal, which was considered by the authors to be unusually precocious, the pubescent male length range was 123–153 cm and the active mature male length range was 135-157 cm.
Lockyer (1995a)	British waters	Stranded & bycaught	M=144	Na	130	For male porpoise, it was assumed that puberty occurred at 130 cm length and that sexual maturation followed that.
Lockyer and Kinze (2003)	Danish waters & North Sea (1996-1998)	Caught, bycaught & stranded	M=135	143	135	The age at first reproduction corresponded to a length of ~143 cm. Males were considered to be pubertal until the testes weight exceeded 200 g which corresponded to 135 cm length.
Ólafsdóttir et al. (2003)	Iceland (1991-1997)	Bycaught	F=335 M=664	146	135	Mean length of first time ovulators was 147.6 cm (n=30). Mean length when 50% of the females are mature was 146 cm. Mean length when 50% of the males were mature was 135 cm.
Lockyer et al. (2003b)	West Greenland (1988-1995)	Bycaught	187	138 – 142	127	Male maturation range = 123 - 130 cm
Karstad (1993)	Norway (N. Norway, North Sea, Skagerrak & Kattegat)	na	na	152.2	>130	Referenced in Lockyer (2003)
Lockyer (1995c), Lockyer (1995a, 1995b), Karakosta et al. (1999)	British Isles (mainly North Sea)	na	na	140 - 145	130 - 135	Referenced in Lockyer (2003)
Karstad (1993), Berggren (1995)	Sweden (Skagerrak, Kattegat & western Baltic)	na	na	145.3	na	Referenced in Lockyer (2003)

Table 4 Age at sexual maturity (ASM) information obtained from the literature review. F – female, M – male, na – not available.

Reference	Location	Average ASM (years)		Notes
		F	M	
Learmonth et al. (2014)	Scotland	4.35	5.00	A binomial GLM for females estimated an ASM of 4.35 (95% CI=3.93–4.71, 83.6% of deviance explained). A binomial GLM for males estimated an ASM of 5.00 (95% CI=4.03–5.88, 82.3% of deviance explained).
Lockyer (1995a)	British waters	na	3	For male porpoise, it was assumed that puberty occurred at age 3 and that sexual maturation followed that.
Lockyer and Kinze (2003)	Inner Danish waters + North Sea	3.3	3 - 4	Based on testes weight, the change from immature to mature occurs between 3-4 years.
Ólafsdóttir et al. (2003)	Iceland	2.5	2.9	Females: Pubertal females had an average age of 1.5, implying an ASM of 2.5 (s=0.971, n=62). Average age of first time ovulators was 2.8 years (s=1.02, n=21). Estimated age when 50% of the females are mature was 3.20 years (n=269). Regression of the number of corpora against age estimated ASM of 2.1 years (n=51). The DeMaster (1978) method estimated ASM as 4.4 years (s = 0.318, n= 293). Males: Pubertal females had an average age of 1.9 years and therefore an ASM of 2.9 years (s=1.108, n=21). Estimated age when 50% of the males were mature was 1.9 years (n=493). The DeMaster (1978) method estimated ASM as 2.6 years (s=0.099, n=526).
Lockyer et al. (2003b)	West Greenland	2.2 - 3.6	2.0 - 2.45	Using the DeMaster method resulted in an average ASM of 2.45 years for males. Using the DeMaster method for age at first ovulation resulted in an average ASM of 3.6 years for females.
Karstad (1993)	Norway (N. Norway, North Sea, Skagerrak & Kattegat)	3.9 ±0.4	2.0 - 3.0	Referenced in Lockyer (2003)
Addink et al. (1995)	Dutch waters	5.0	na	Referenced in Addink et al. (1995), Lockyer (2003)
Karstad (1993), Kull and Berggren (1993), Berggren (1995)	Sweden (Skagerrak, Kattegat and western Baltic)	4.3 +0.5	na	Referenced in Lockyer (2003)
Murphy et al. (2015)	UK	4.73	na	Average age of females at sexual maturity was 4.73 years (standard error = 0.03, n = 250).
Kesselring et al. (2017)	German North Sea and Baltic Sea	4.95	na	The threshold age at which >50% had one corpus or more was used as a sign of former ovulation. The threshold was determined at 4.95 years or higher (95% CI: 4.2±4.8 years).

Different length-thresholds could be used to assess the proportion of immatures (given the sex and geographic variations observed in Table 3 and 4) and depending on the length threshold used, the proportion of immature animals in the sample ranged between 34 – 80% (Figure 16).

Based on the length at sexual maturity studies outlined in Table 3 and Table 4**Error! Reference source not found.**, it can be assumed that all animals <127 cm in length are very likely immature animals and all animals ≥152 cm are very likely mature. The animals classed as unknown (≥127 - <152 cm) will be a mix of large immature males, mature males, immature females and small mature females. From the length measurements obtained for Grade 1&2 images, it is estimated that at least 36% of the sample were immature animals (<127 cm), 20% were mature animals (≥152 cm) and 44% were unknown (≥127 - <152 cm) (Table 5).

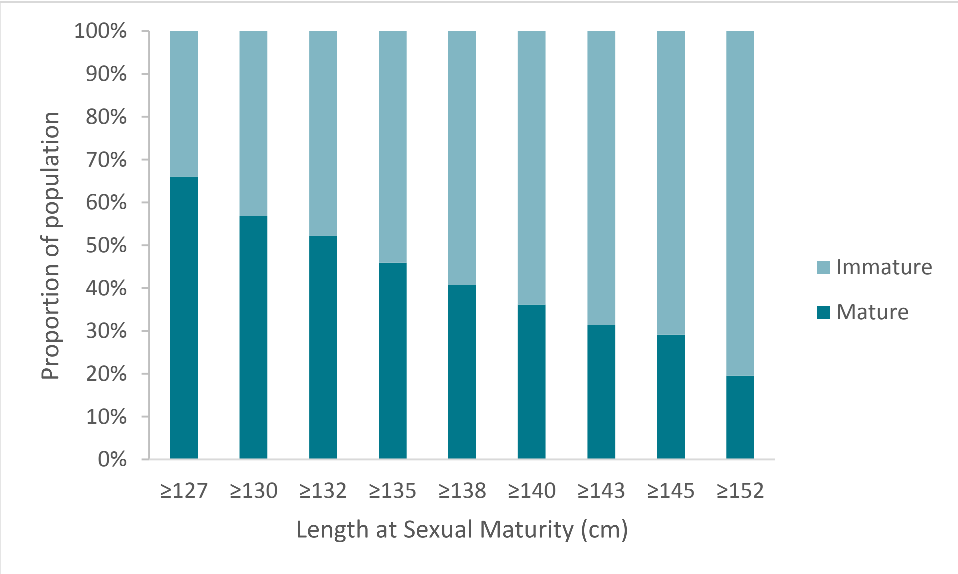

Figure 16 Proportion of the sample estimated to be mature by varying length at sexual maturity estimates (all Grades).

Table 5 Proportion of the sample estimated to be in each age class.

Length	Stage Class	% Sample (Grade 1-3)	% Sample (Grade 1&2)
<127 cm	Immature	34%	36%
≥127 & <152 cm	Unknown	46%	44%
≥152 cm	Mature	20%	20%

5 Discussion

5.1 Stage structure

Previous assessments have determined that studying the juvenile fraction of the population can improve the detection of demographic changes and population status. The juvenile fraction of the population is often used in the fisheries literature as a measure of population health, however there is little in the marine mammal literature about it. One example is for the declining Alaskan Steller sea lion population, where demographic parameters changed over time as the population declined, starting with severe effects on juvenile survival (Holmes and York 2003). Holmes and York (2003) used time-varying matrix models to determine whether age-structure information could be used to detect changes in survival, fecundity and population status. Incorporating the 'juvenile fraction' (akin to the proportion of immatures) along with count data improved their ability to detect changes in demographic parameters. If the juveniles in the population had been monitored before and during the period of early population decline then a change in the proportion of juveniles could potentially have been detected to trigger appropriate conservation or management actions. Therefore monitoring such a demographic parameter (in combination with count data) demonstrates one approach to early warning monitoring.

As outlined in section 4.4, the sexual dimorphism and the geographic variations in length at maturity estimates of harbour porpoise make it difficult to reliably assign a stage class to the animals in many of the images. That is, there is no single threshold for length at maturity to guide assessments of population health. However, there is some promise in using a fixed threshold for length at sexual maturity of ≥127 cm (the smallest mean length at maturity estimate for a male porpoise from Table 3**Error! Reference source not found.**), where animals below this size are classified as juveniles, then the proportion of juveniles can be monitored in order to identify any changes over time. Such an approach is conservative as it will likely underestimate the proportion of juveniles in the population as it does not account for the number of juveniles in the length range between ≥127 – 152 cm where the stage class of the animals is an unknown mix of juvenile males, mature males, juvenile females and mature females.

The current status of the North Sea harbour porpoise population is unclear but there is no evidence that it is declining. The abundance estimates obtained from the three SCANS surveys of the North Sea in 1994, 2005 and 2016 suggest a slight increase over this period (Hammond et al. 2017) (Figure

17)(though the power of these surveys to detect changes remains unclear). The UK harbour porpoise has been assessed as having a Favourable Conservation Status with medium confidence, and the species is expected to survive and prosper under the current conservation approach (JNCC 2013). However, monitoring of the proportion of the population that is immature could provide an early warning of any changes in population demographics and abundance. Whilst we identify and attempt to account for different sources of error and uncertainty in this photogrammetric approach, there is uncertainty in estimates of all marine mammal monitoring, including the SCANS surveys (see uncertainty estimates in Figure 17). The benefit of monitoring demographic parameters is that they can provide an early warning of population change, before they are reflected in abundance or density estimates.

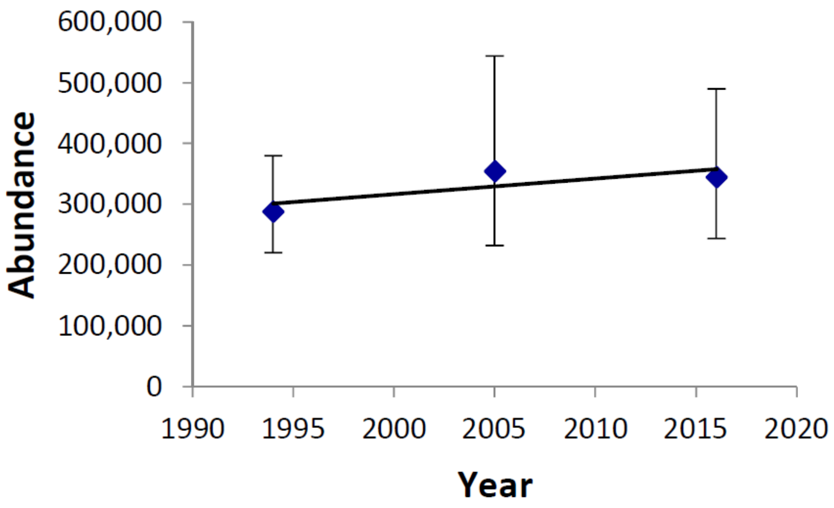

Figure 17 Abundance estimates and trend line for the North Sea harbour porpoise population. Estimated rate of annual change = 0.8% (95%CI: -6.8; 9.0%), p = 0.18. Error bars are log-normal 95% confidence intervals. Figure taken from Hammond et al. (2017).

5.1.1 Age frequency distribution

A critical element in assessing the proportion of immatures (or other demographic parameters) in a population is an understanding of the threshold at which a conservation or management action should be triggered. Unfortunately, limited data are available for marine (or terrestrial mammal species to inform this.

Age frequency distributions have been reported for European harbour porpoise using stranded or bycaught animals (Lockyer 1995a, Lockyer and Kinze 2003, Luque et al. 2009, Learmonth et al. 2014, Murphy et al. 2015, Kesselring et al. 2018) (see Appendix 3 for summary information). However, understanding how the age-frequency distributions of stranded or bycaught animals differs requires an estimate of the probability of animals of different age classes to strand or be bycaught. Therefore without this, there is a risk that such distributions will be biased and thus not representative of the population. For example, a large proportion of juveniles in a sample of bycaught or stranded animals may be as a result of higher mortality rates for juveniles or a difference in the distribution between juvenile and adults, potentially resulting in juvenile carcasses being more likely to wash ashore and be recovered. If strandings are biased towards young animals, then the age-structure derived from stranding data is not a true representation of the population, which means that strandings data may be more representative of the age structure of deaths rather than the age structure of the population. In addition, while the general expectation is that growing populations have higher proportions of juveniles, it is also possible that populations with large proportions of juveniles in strandings and bycatch datasets may be as a result of low survival rates for younger age classes which may not necessarily be associated with an overall growing population size.

5.2 Measurement error

5.2.1 Inter-user error

Measurement error was calculated by determining the range of measurements obtained by five independent users (the main observer and four other users). There was reasonable agreement in the measured lengths between the five users; however, once the outliers were removed the agreement in length measurements improved, with 86% of the measured lengths being within a 5 cm range. Given that four of the users were less experienced at measuring images than the main observer and that these were the first 20 images they had ever measured, it is highly likely that with a small amount of additional training and some more experience viewing images from aerial surveys that the overall range in measurements between users would reduce. Should digital aerial survey data be used to make morphometric measurements going forward, it is highly recommended that a sub-sample of images are measured by multiple independent experienced users in order to quantify and account for measurement error in the resulting lengths and stage class allocation. In addition, given the likely sample sizes in available existing datasets from digital aerial surveys, it will likely not be practical to measure each image by hand. Therefore the use of automated approaches (e.g. such as those in development by Duke Marine Laboratory[5]) should be considered in combination with human observers (See section 5.5).

It is important to distinguish between the inter-user measurements and the comparison to the automated software. The comparison to the automated software was for a single observer compared to a single measurement from the automated software. There was no information provided about how the automated software made its measurements, nor were there any resampling of the automated software and therefore there is no quantification of error range. It would be beneficial to understand more about the data-providers automated software, such as how it detects the animal, where it positions its points of measurement and how/if it accounts for curvature in the animal in any way.

5.2.2 Ground Sampling Distance

This study has demonstrated that an accurate and calibrated GSD is vital in order to reduce error in the measurements. Should digital aerial survey data be used to make morphometric measurements going forward, it is essential that all measurement errors are reduced as much as possible, and therefore image-specific GSD (incorporating flight height at the time of the image) would be required to minimise this source of error.

5.3 Limitations

This study has shown that photogrammetry analysis of images of porpoise obtained from aerial surveys has the potential to provide a means by which to measure animal length and can therefore provide an estimate of the minimum proportion of juveniles in a population and thus provide insights into the population stage structure and therefore an indication of the population health. However, like most marine mammal monitoring methods, it is not without its limitations and the approaches require further development before they can be employed in population monitoring. For example, the way in which harbour porpoise move both at the surface and underwater, combined with the quality of the images meant that there were several challenges to obtaining reliable length

[5] https://sites.nicholas.duke.edu/uas/

and width measurements for many images. These are summarised in the sections below with example images to illustrate the limitations.

5.3.1 Width / Body condition

As outlined in section 4.2.1, all users found it difficult to measure animal width and had considerably less confidence in these measurements compared to the length measurements. While a width measurement was taken for each image (where possible), due to the lack of confidence and level of difficulty in obtaining this measurement, it was not possible to obtain any sensible estimate of body condition based on measured animal width.

Other studies that have investigated measures of body condition have recommended that width/girth measurements are not used for female animals since this measure would be confounded by the reproductive status of the animal (pregnancy, lactating, resting) which would inevitably affect width/girth measurements and therefore the resulting body condition estimate (Kershaw et al. 2017). In addition, captive harbour porpoise have shown variations in body mass and girth (both indicators of body condition) by season (Lockyer et al. 2003a, Kastelein et al. 2018a) (Figure 18**Error! Reference source not found.** and Figure 19**Error! Reference source not found.**). Therefore a change in body condition may be linked to (natural) seasonal changes in thermoregulatory requirements (rather than a change in physical condition due to exposure to anthropogenic stressor), and as such, this approach is not considered to be suitable for assessing the body condition of porpoise in these aerial images.

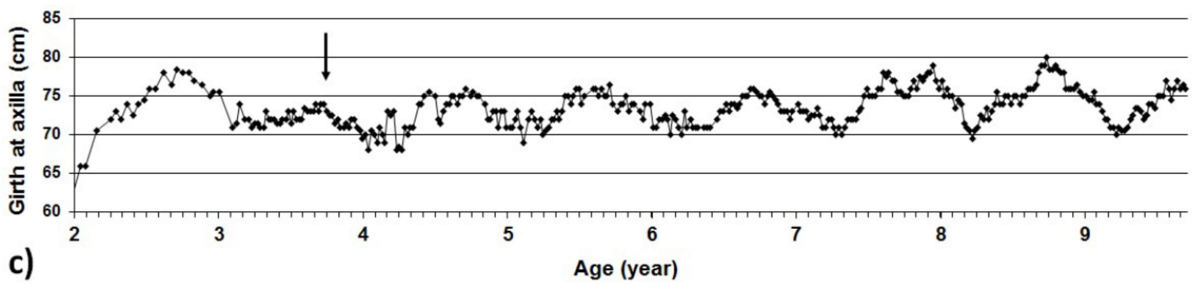

Figure 18 Seasonal changes in girth for a male captive porpoise. The arrows indicate the end of the period of rapid growth in body length and the start of the period of relative slow growth and stability in body length. Figure taken from Kastelein et al. (2018a).

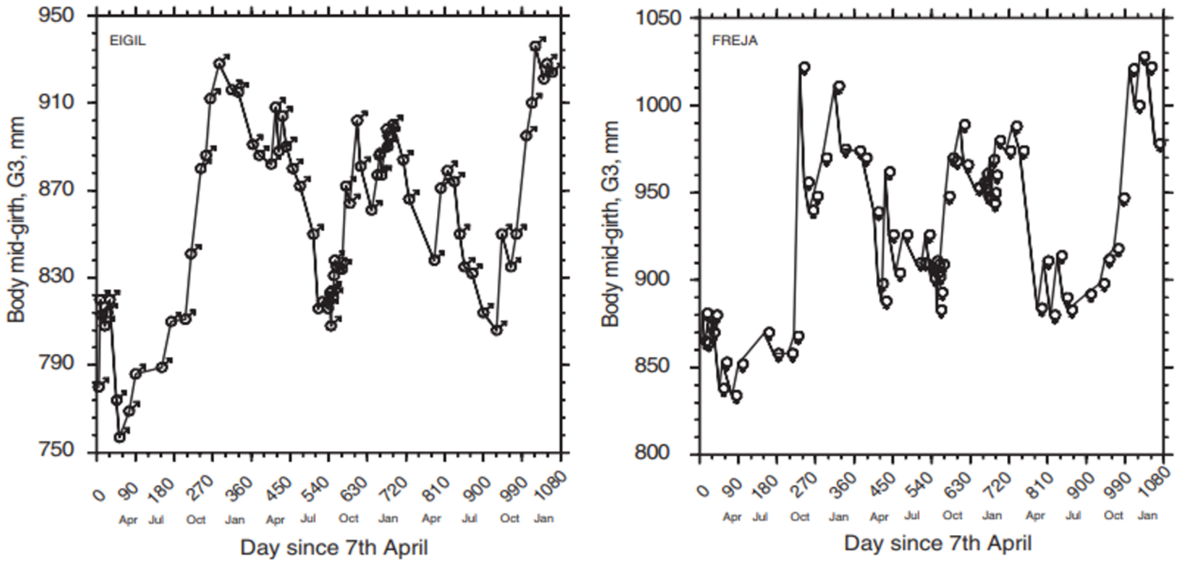

Figure 19 Seasonal changes in girth for a male captive porpoise (left) and a female captive porpoise (right). Figure taken from Lockyer et al. (2003a).

5.3.2 Body angle

In order to obtain an accurate length measurement, the animal in the image needs to be straight and flat (i.e. parallel to the water surface). However, it is well known that harbour porpoise tend to roll as they surface to breathe, so that often either the head or the tail will be angled downwards. In cases where the animal did not have a straight, flat body axis, length measurements were likely to have been underestimates. For example, Figure 20a shows a porpoise that has a reasonably flat, straight body axis and it is possible to identify both the rostrum and the fluke, while Figure 20b (Grade 3) shows the animal in clearly mid-surfacing with the dorsal fin above the water surface and the head angled downwards.

In addition, porpoise sometimes surface off axis (i.e. with a twist to their body) such that the ventral side can be seen at the surface. They also curve and bend their body so that they do not present a flat body axis for measurement. For example, Figure 21a is an example of a porpoise showing a flat, straight body axis, Figure 21b (Grade 4 – not measured) shows an animal with a curved body axis and Figure 21c (Grade 4 – not measured) shows an animal twisted on its side so that the ventral side is visible.

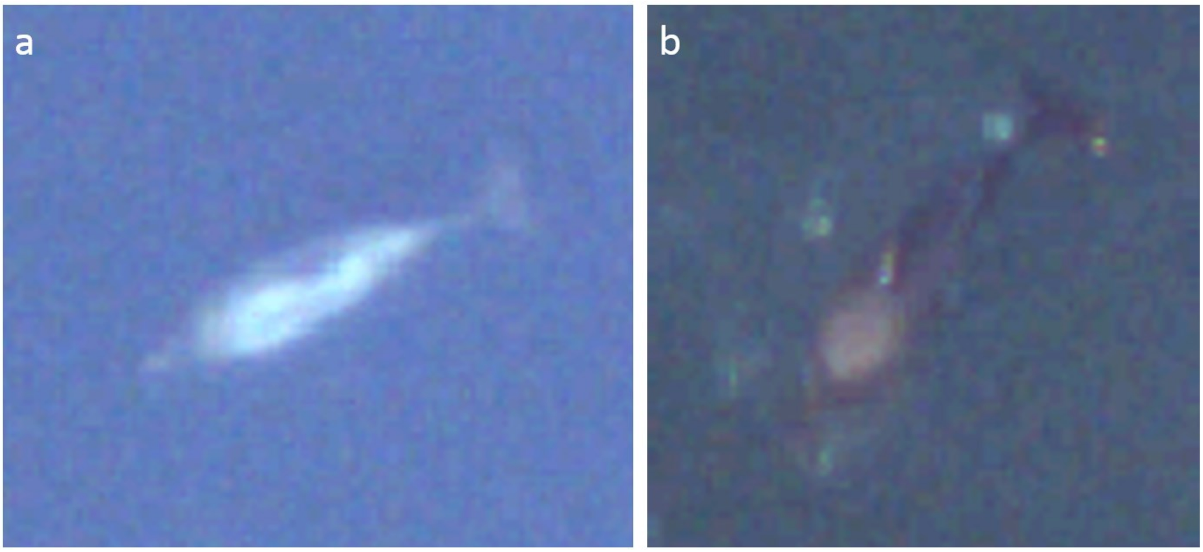

Figure 20 Example of body-axis in the porpoise images. (a) shows a reasonably flat, straight body-axis while (b) shows a porpoise mid-surfacing with the head angled downwards (Grade 3).

Figure 21 Examples of body-axis in the porpoise images. (a) shows a flat, straight body-axis, (b) shows a curved body axis (Grade 4 – not measured) and (c) shows an animal twisted on its axis so the ventral side is visible (Grade 4 – not measured).

The length measurements obtained in this pilot study were straight line measurements from the tip of the rostrum to the notch of the tail fluke and therefore did not account for curvature of the body axis. It is possible to use the segmented line tool in ImageJ to draw a curved line between the rostrum and the notch in the fluke in order to obtain a length measurement (e.g. see Figure 22), however due to time constraints this was not developed or included in this project. Should this photogrammetry method be adopted to monitor the porpoise population, this segmented line measurement approach should be developed further in order to obtain more accurate length measurements for animals displaying a curved body axis.

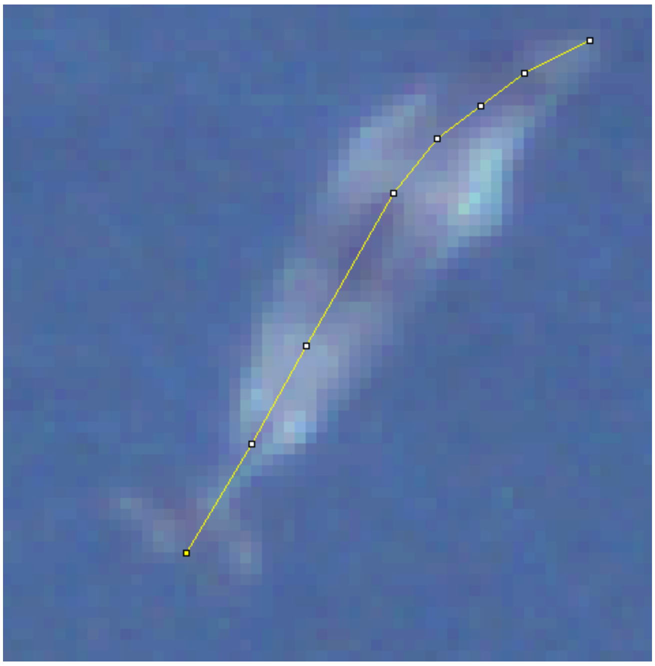

Figure 22 Example of a segmented line measurement for a porpoise with a curved body axis.

5.3.3 Obscured images

As a porpoise breaks the surface it can cause bow waves and splashes at the water surface which can be visible in the images and can obscure the view of the animal. Figure 23 shows examples of images where bow waves and splashes obscure view of the rostrum which will have resulted in inaccurate length measurements as the measurer would have been unable to correctly identify the tip of the rostrum for measurement.

Figure 23 Examples of images where bow waves and splashes obscure view of the rostrum. Left to right image grades assigned were: G2, G3, G3, G2, G3.

5.3.4 Contrast and pixilation

The contrast between the animal and the sea was often low which, in addition to the pixilation in the image, made it difficult to accurately measure the length and width of the animal. For example, Figure 24 (top) is an example of an image where the left hand side of the animal has a particularly low contrast to the surrounding water and so the edge of the animal was difficult to identify and as such the width measurement was difficult to obtain. Figure 24 (bottom) is an example of an image where the animal was underwater and there was little contrast between the colour of the animals body and the surrounding sea, making it difficult to pinpoint the exact locations required for both length and width measurement.

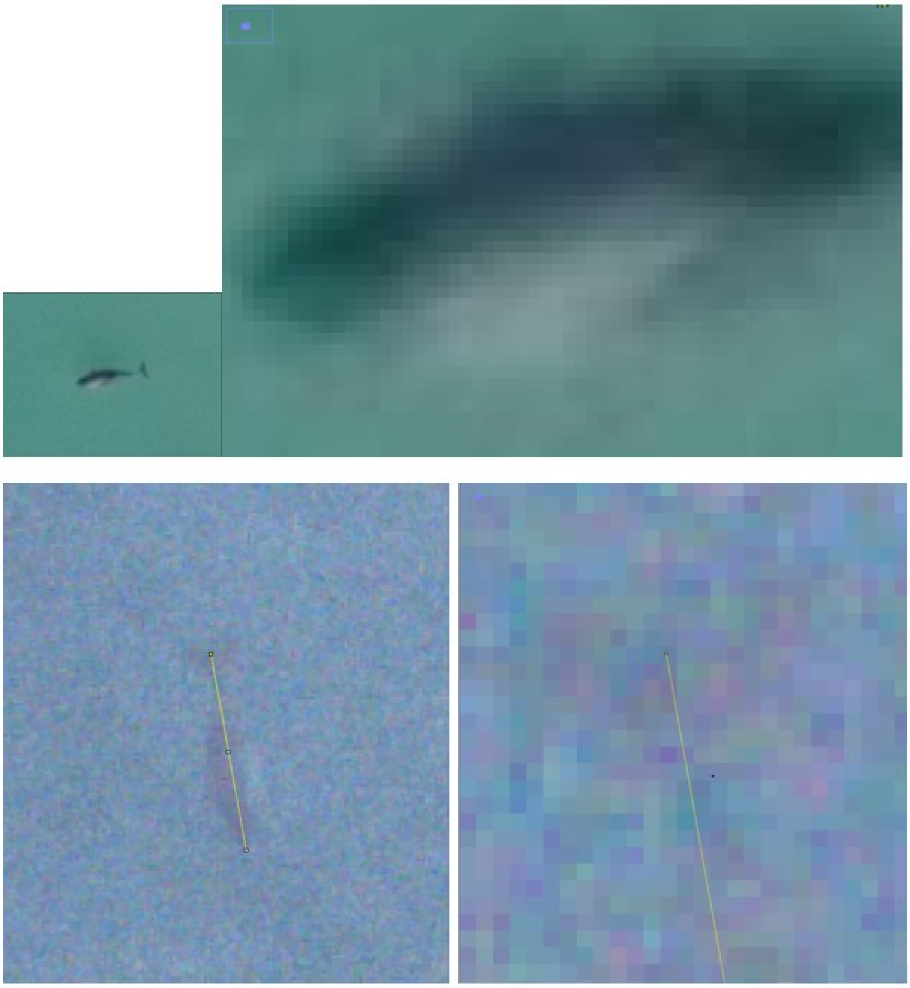

Figure 24 Examples of images with poor contrast and high pixilation. Top: example of an image with high levels of pixilation and low contrast on the animals left side (grade 3). Bottom: example of an image with very low contrast in colour between the body and the sea (grade 3).

It is important to note that the images obtained from the APEM and HiDef aerial surveys were not obtained from a survey design that was intended for photogrammetric analysis. For example, previous photogrammetric studies such as Christiansen et al. (2016) obtained photographs from a drone positioned 30-50 m above the whale which could be remote controlled to move with the animal until images of a particular quality were obtained. By contrast, the images obtained from HiDef and APEM are obtained from a moving aircraft from >500 m altitude, not necessarily moving in the same direction as the animal and, in the case of APEM, taking snapshot images at a set time interval.

Therefore it was never expected that the images obtained from these surveys would be perfect for photogrammetric analysis, but given the likely sample sizes of sighted porpoises held by aerial survey companies across NW Europe – this is an area for further assessment for the potential to monitor for early warnings of decline and utilise pre-existing datasets (and those collected for other primary means, representing added value of such datasets). While decades of data are held by both companies, it is important to understand what proportion of the images obtained from these aerial surveys are of suitable quality for use in this type of photogrammetric analysis. Indications from this study show that image quality is likely to be a limiting factor since one of the companies considered here found that only 23 of a sample of 580 images examined were of ideal quality for photogrammetric analysis (2.7%). It is unknown if these 580 images were a random sample from their total database or if they were all from the same survey site since information on location or timing of images was not provided. It is possible that if the images examined were all from the same survey site then they may not be representative of image quality obtained from other sites, if, for example, the environmental conditions during the survey were suboptimal (higher sea state, decreased water clarity etc.).

Furthermore, there is a need to better understand the temporal and spatial distribution over which past digital aerial surveys have been conducted. This will allow an assessment of the kinds of sample sizes that could be available for future photogrammetric analysis.

5.4 Conclusions

This study has shown that photogrammetry analysis of images of porpoise obtained from aerial surveys can provide a means by which to measure animal length and estimate the minimum proportion of juveniles in a population and thus may have potential towards monitoring population health. However, like most marine mammal monitoring methods, it is not without its limitations. While it is not possible to define a threshold for the proportion of juveniles in a population that is considered to be "healthy"; it is important to note that a change in the proportion of juveniles in a population will reflect a change in the demographic rates and can therefore be considered as an early warning sign of a change in population trends.

This study has demonstrated that an accurate and calibrated GSD is vital in order to reduce error in the measurements and we would recommend a combination of automation and human observers if digital aerial survey images are to be used for monitoring population health.

This study has determined that this photogrammetric approach is unlikely to be a suitable method for determining body condition in harbour porpoise due to the difficulties in obtaining width measurements and because porpoise girth measurements are likely change seasonally (driven by thermoregulatory need) and that girth measurements can be confounded by reproductive status in females.

Both HiDef and APEM have undertaken aerial surveys around the UK to characterise bird and marine mammal presence for over 10 years, and therefore they hold between them a huge wealth of data that may have the potential to be used as a way to measure and monitor the stage structure of harbour porpoise populations around the UK.

5.5 Next steps

This study has demonstrated that whilst this approach shows promise, there are a number of research topics that require further assessment in order for stage-structure to be assessed in this way. These include: a better understanding of length and age of sexual maturity in male and female porpoises in the North Sea; improvements to measurement processes; and, improved GSD calibration of images. These are discussed further below.

There may be potential for ground-truthing of these methods for harbour porpoise. For example, researchers at the Marine Mammal Research Program at the University of Hawaii are working on a project to obtain ground truthed UAS morphometric measurements of captive bottlenose dolphins and false killer whales (Vivier et al. 2019). This study will obtain aerial images of animals both lying flat and when performing surfacings in order to measure lengths for differing levels of body curvature in known length animals. A similar study on captive porpoise could potentially be conducted using similar focal lengths (determined by camera systems and flight heights) as used in APEM and HiDef surveys. This would help to assess the accuracy of the photogrammetric measurements and therefore provide further insight into the suitability of these methods to assess the stage-structure of harbour porpoise in the UK.

Another potential approach to assessing how body curvature at the surface affects resulting length measurements is to select multiple still frames for a HiDef video sequence and measure the animal from multiple sequential images in order to obtain a range of measurements over the course of a surfacing. This may provide information on the range of potential length measurements that can be obtained during one surfacing, and therefore how much surfacing body curvature affects the length measurements.

In order to obtain a better understanding of the current/recent stage structure of the harbour porpoise population around the UK, it is recommended that further harbour porpoise images obtained from APEM and HiDef surveys are processed for photogrammetric purposes. Given the vast number of images that each company holds, it is recommended that images are measured using an automated system with a portion of these verified by hand rather than processing each image manually. This will drastically reduce the time and cost of processing the images. For example, a deep neural network approach is being developed at the Marine Robotics and Remote Sensing (MaRRS) Lab at Duke University using a neural net to scan photogrammetry images and apply the drone altitude information to automatically measure animals. Should further images be assessed, it is recommended that each image is graded for quality by two independent users and any discrepancies discussed and resolved. This could help to prevent bias in the grading and to ensure that a strict set of rules is adhered to. This is common practice in several photo-ID projects where images are graded for quality by multiple users.

If an automated system to measure the images is used in future, then it is recommended that prior to measurement, each image is graded for quality. That way it is known which of the images contain porpoise with a correct body-axis to measure (and thus the resulting measurements should have a high degree of confidence) and which images contain some form of curvature in the animals body axis or obstructed view where resulting length measurements may be less confident. In addition, it would be beneficial to bootstrap the automated measurements to obtain a measure of the error.

Even when measurement error is minimised, quantified and accounted for (ie: calculation of calibrated GSD, bootstrapped automated measurements, human verification), there still remains the difficulty in assigning stage structure to the sample due to the considerable variation in the literature in the estimated age or length of sexual maturity of harbour porpoise. It is therefore considered necessary to collate and analyse more data in order to attempt to refine this estimate, however the sexual dimorphism in harbour porpoise will always result in a confounding factor for certain size classes of animals. More generally we would also recommend a full assessment of the literature to determine other suitable metrics or approaches of how demographic parameters have been used to supplement population health monitoring and help identify early warnings of change (e.g. as in Holmes and York, 2003).

6 Acknowledgements

Many thanks to both APEM Limited and HiDef Aerial Surveying Limited for the provision photographs of harbour porpoises obtained from their previous aerial surveys.

The whale.morpho function R code was developed by Antoine M. Dujon and was accessed as part of the supplementary information provided in Christiansen et al. (2016).

ImageJ was developed at the National Institutes of Health. ImageJ version 1.52a was downloaded from https://imagej.nih.gov.

7 Literature Cited

Addink, M., T. Sørensen, and M. G. Hartmann. 1995. Aspects of reproduction and seasonality in the Harbour Porpoise from Dutch waters. Pages 459-464 Developments in Marine Biology. Elsevier.

Berggren, P. 1995. Stocks, status and survival of harbour porpoises (*Phocoena phocoena*) in Swedish waters. Department of Zoology, Stockholm University.

Booth, C. G., R. Plunkett, and J. Harwood. 2017. Identifying Monitoring Priorities for Population Consequences of Disturbance. Report Code SMRUC-ONR-2017-017, submitted to the Office of Naval Research – Marine Mammal & Biology Program.

Carlén, I., L. Thomas, J. Carlström, M. Amundin, J. Teilmann, N. Tregenza, J. Tougaard, J. C. Koblitz, S. Sveegaard, and D. Wennerberg. 2018. Basin-scale distribution of harbour porpoises in the Baltic Sea provides basis for effective conservation actions. Biological Conservation **226**:42-53.

Cheney, B., R. Wells, T. Barton, and P. Thompson. 2017. Laser photogrammetry reveals variation in growth and early survival in free-ranging bottlenose dolphins. Animal Conservation **21**:252–261.

Christiansen, F., A. M. Dujon, K. R. Sprogis, J. P. Arnould, and L. Bejder. 2016. Noninvasive unmanned aerial vehicle provides estimates of the energetic cost of reproduction in humpback whales. Ecosphere **7**.

DeMaster, D. 1978. Calculation of the average age of sexual maturity in marine mammals. Journal of the Fisheries Board of Canada **35**:912-915.

Hammond, P., C. Lacey, A. Gilles, S. Viquerat, P. Börjesson, H. Herr, K. Macleod, V. Ridoux, M. Santos, M. Scheidat, J. Teilmann, J. Vingada, and N. Øien. 2017. Estimates of cetacean abundance in European Atlantic waters in summer 2016 from the SCANS-III aerial and shipboard surveys.

Holmes, E., and A. York. 2003. Using age structure to detect impacts on threatened populations: a case study with Steller sea lions. Conservation Biology **17**:1794-1806.

Jewell, R., L. Thomas, C. Harris, K. Kaschner, R. Wiff, P. Hammond, and N. Quick. 2012. Global analysis of cetacean line-transect surveys: detecting trends in cetacean density. Marine Ecology Progress Series **453**:227-240.

JNCC. 2013. European Community Directive on the Conservation of Natural Habitats and of Wild Fauna and Flora (92/43/EEC) Third Report by the United Kingdom under Article 17 on the implementation of the Directive from January 2007 to December 2012 Conservation status assessment for Species: S1351 - Harbour porpoise (*Phocoena phocoena*).

Karakosta, C., P. Jepson, H. Ohira, A. Moore, P. Bennett, and W. Holt. 1999. Testicular and ovarian development in the harbour porpoise (*Phocoena phocoena*). Journal of Zoology **249**:111-121.

Karstad, S. 1993. Vekst og reproduksjon hos nise (*Phocoena phocoena*) i norske og i svenske farvann. (Growth and reproduction of porpoises in Norwegian and Swedish waters). University of Oslo.

Kastelein, R. A., L. Helder-Hoek, and N. Jennings. 2018a. Seasonal Changes in Food Consumption, Respiration Rate, and Body Condition of a Male Harbor Porpoise (*Phocoena phocoena*). Aquatic Mammals **44**:76-91.

Kastelein, R. A., S. Van de Voorde, and N. Jennings. 2018b. Swimming Speed of a Harbor Porpoise (*Phocoena phocoena*) During Playbacks of Offshore Pile Driving Sounds. Aquatic Mammals **44**:92-99.

Kershaw, J. L., M. Sherrill, N. J. Davison, A. Brownlow, and A. J. Hall. 2017. Evaluating morphometric and metabolic markers of body condition in a small cetacean, the harbor porpoise (*Phocoena phocoena*). Ecology and Evolution **2017; 7**:3494–3506.

Kesselring, T., S. Viquerat, R. Brehm, and U. Siebert. 2017. Coming of age:-Do female harbour porpoises (*Phocoena phocoena*) from the North Sea and Baltic Sea have sufficient time to reproduce in a human influenced environment? PLoS ONE **12**:e0186951.

Kesselring, T., S. Viquerat, R. Brehm, and U. Siebert. 2018. Correction: Coming of age: - Do female harbour porpoises (*Phocoena phocoena*) from the North Sea and Baltic Sea have sufficient time to reproduce in a human influenced environment? PLoS ONE **13**:e0199633.

Kull, M., and P. Berggren. 1993. Seasonal patterns in dentinal zone formation in harbour porpoises (*Phocoena phocoena*) from the Kattegat and Skagerrak Seas. Abstract, p. 60, 10th Biennial Conference on the Biology of Marine Mammals, Nov. 11-14, 1993, Galveston, Texas, USA.

Learmonth, J. A., S. Murphy, P. L. Luque, R. J. Reid, I. A. P. Patterson, A. Brownlow, H. M. Ross, J. P. Barley, M. Begona Santos, and G. J. Pierce. 2014. Life history of harbor porpoises (*Phocoena phocoena*) in Scottish (UK) waters. Marine Mammal Science **30**:1427-1455.

Lockyer, C. 1995a. Aspects of the biology of the harbour porpoise, *Phocoena phocoena*, from British waters. Pages 443-457 Developments in Marine Biology. Elsevier.

Lockyer, C. 1995b. Aspects of the morphology, body fat condition and biology of the harbour porpoise, *Phocoena phocoena*, in British waters. 0255-2760, International Whaling Commission Special Issues.

Lockyer, C. 1995c. Investigation of aspects of the life history of the harbour porpoise, *Phocoena phocoena,* in British waters. 0255-2760, Reports-International Whaling Commission Special Issues.

Lockyer, C. 2003. Harbour porpoises (*Phocoena phocoena*) in the North Atlantic: Biological parameters. NAMMCO Scientific Publications **5**:71-89.

Lockyer, C., G. Desportes, K. Hansen, S. Labberté, and U. Siebert. 2003a. Monitoring growth and energy utilisation of the harbour porpoise (*Phocoena phocoena*) in human care. NAMMCO Scientific Publications **5**:107-120.

Lockyer, C., M. Heide-Jørgensen, J. Jensen, C. C. Kinze, and T. B. Sørensen. 2001. Age, length and reproductive parameters of harbour porpoises *Phocoena phocoena* (L.) from West Greenland. ICES Journal of Marine Science **58**:154-162.

Lockyer, C., M. Heide-Jørgensen, J. Jensen, and M. Walton. 2003b. Life history and ecology of harbour porpoises (*Phocoena phocoena*) from West Greenland. NAMMCO Scientific Publications **5**:177-194.

Lockyer, C., and C. Kinze. 2003. Status, ecology and life history of harbour porpoise (*Phocoena phocoena*), in Danish waters. 2003 **5**:33.

Luque, P. L., G. J. Pierce, J. A. Learmonth, M. B. Santos, E. Ieno, A. López, R. Reid, E. Rogan, A. Gonzalez, and J. Boon. 2009. Dentinal anomalies in teeth of harbour porpoises (*Phocoena phocoena*) from Scottish waters: are they linked to sexual maturation and environmental events? Journal of the Marine Biological Association of the United Kingdom **89**:893-902.

Murphy, S., J. L. Barber, J. A. Learmonth, F. L. Read, R. Deaville, M. W. Perkins, A. Brownlow, N. Davison, R. Penrose, and G. J. Pierce. 2015. Reproductive failure in UK harbour porpoises Phocoena phocoena: legacy of pollutant exposure? PLoS ONE **10**:e0131085.

Ólafsdóttir, D., G. A. Víkingsson, S. D. Halldórsson, and J. Sigurjónsson. 2003. Growth and reproduction in harbour porpoises (*Phocoena phocoena*) in Icelandic waters. NAMMCO Scientific Publications **5**:195-210.

Reichert, B. E., W. L. Kendall, R. J. Fletcher, Jr., and W. M. Kitchens. 2016. Spatio-Temporal Variation in Age Structure and Abundance of the Endangered Snail Kite: Pooling across Regions Masks a Declining and Aging Population. PLoS ONE **11**:e0162690.

Rowe, L. E., and S. M. Dawson. 2009. Determining the sex of bottlenose dolphins from Doubtful Sound using dorsal fin photographs. Marine Mammal Science **25**:19-34.

Staetsky, L. D., and J. Boyd. 2015. Strictly Orthodox rising: What the demography of British Jews tells us about the future of the community. Institute for Jewish Policy Research. October 2015.

Taylor, B. L., M. Martinez, T. Gerrodette, J. Barlow, and Y. N. Hrovat. 2007. Lessons from monitoring trends in abundance of marine mammals. Marine Mammal Science **23**:157-175.

Thompson, P. M., and J. Harwood. 1990. Methods for estimating the population size of common seals, Phoca vitulina. Journal of Applied Ecology:924-938.

Vivier, F., A. L. Bradford, M. Hill, E. Oleson, K. Yano, J. L. Pawloski, and L. Bejder. 2019. Calibrating Unoccupied Aerial System (UAS) photogrammetry to derive delphinid population demographic parameters. University of Hawaii conference.

Wade, P. R., and T. Gerrodette. 1993. Estimates of cetacean abundance and distribution in the eastern tropical Pacific.

8 Appendix 1: Length-at-age curves

8.1 Learmonth et al. (2014): Scotland

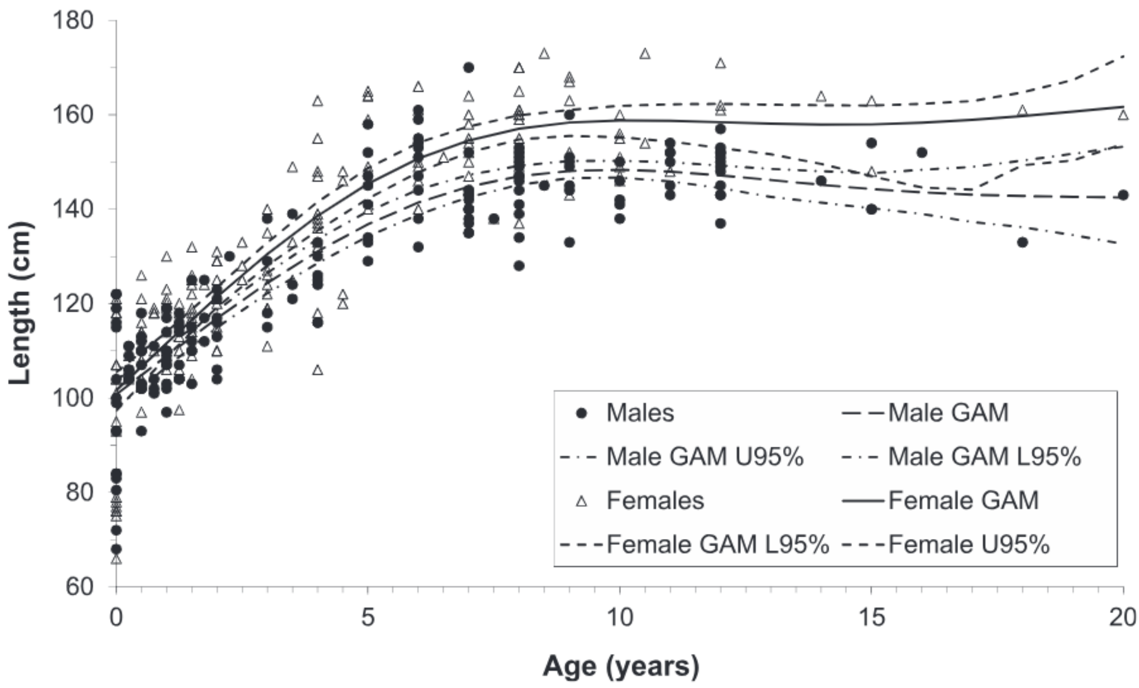

Figure 25 Length at age curves based on predictions from GAMs for male and female harbour porpoises from Scotland, with 95% confidence limits on the predictions.

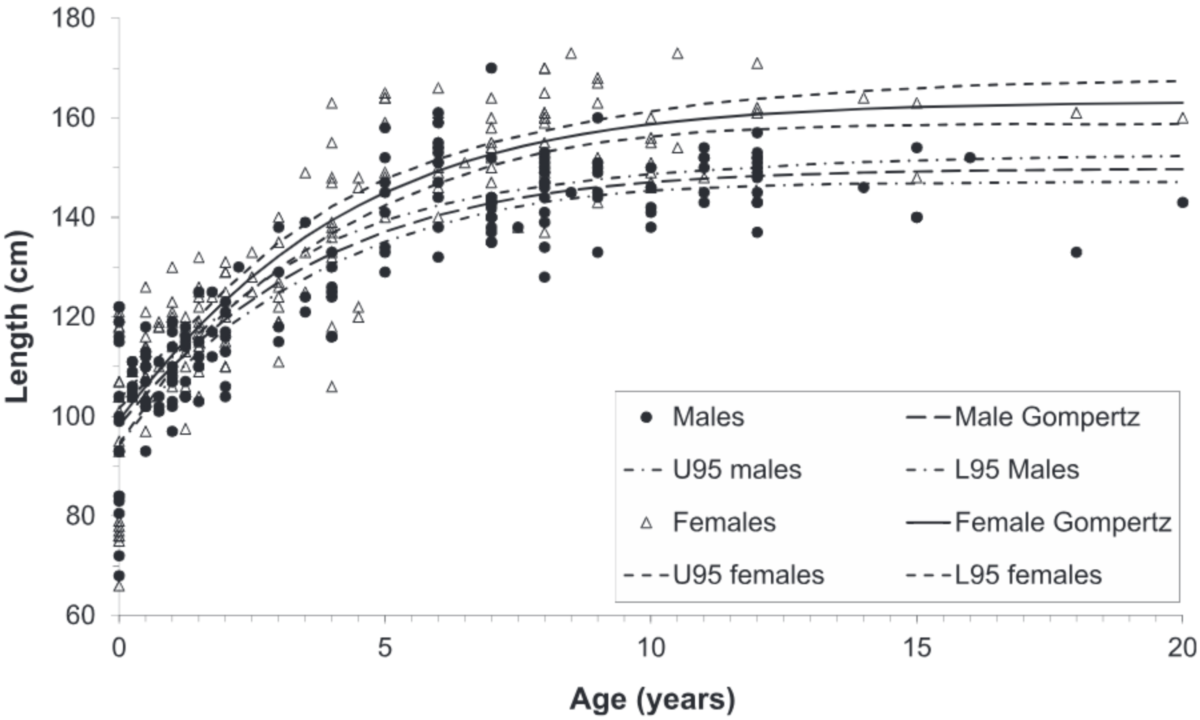

Figure 26 Age at length for male and female harbour porpoises from Scotland, with fitted Gompertz curves (with upper and lower 95% confidence limits).

8.2 Lockyer (1995a): Britain

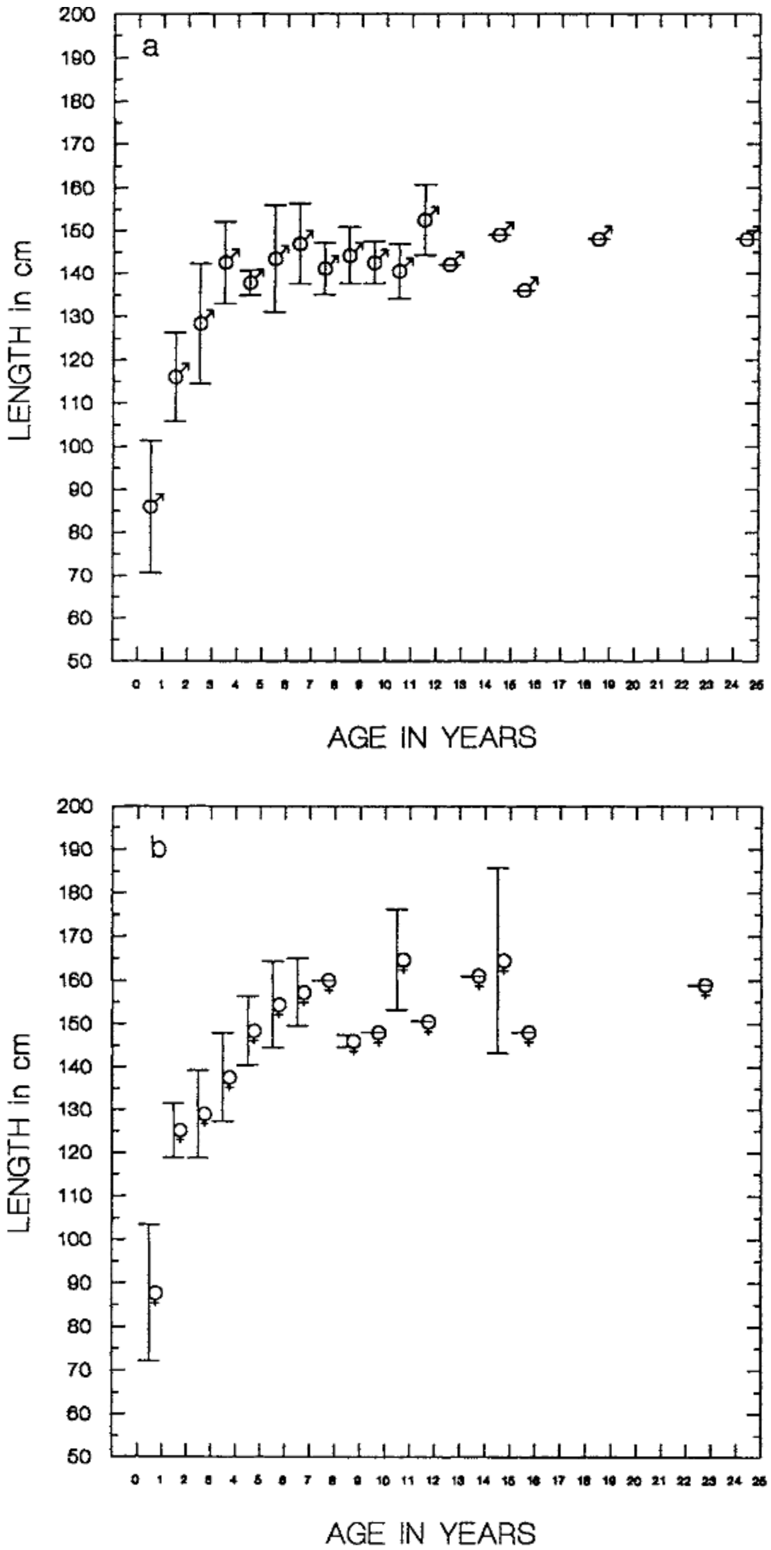

Figure 27 Length (cm) at age for porpoise in British waters, with log-fitted curves (a=males, b=females).

8.3 Lockyer and Kinze (2003): Denmark

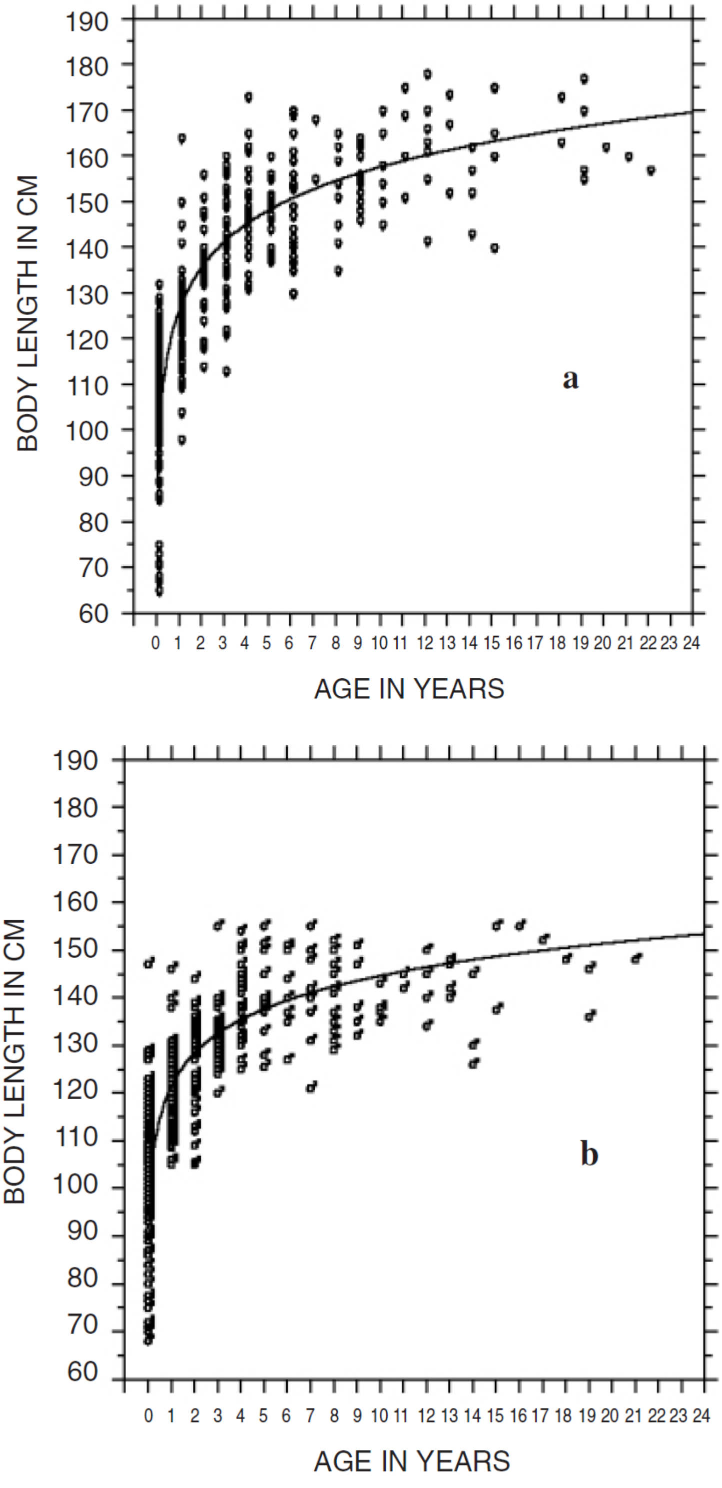

Figure 28 Body length at age of porpoises from strandings and bycatches (a= females, b=males).

8.4 Ólafsdóttir et al. (2003): Iceland

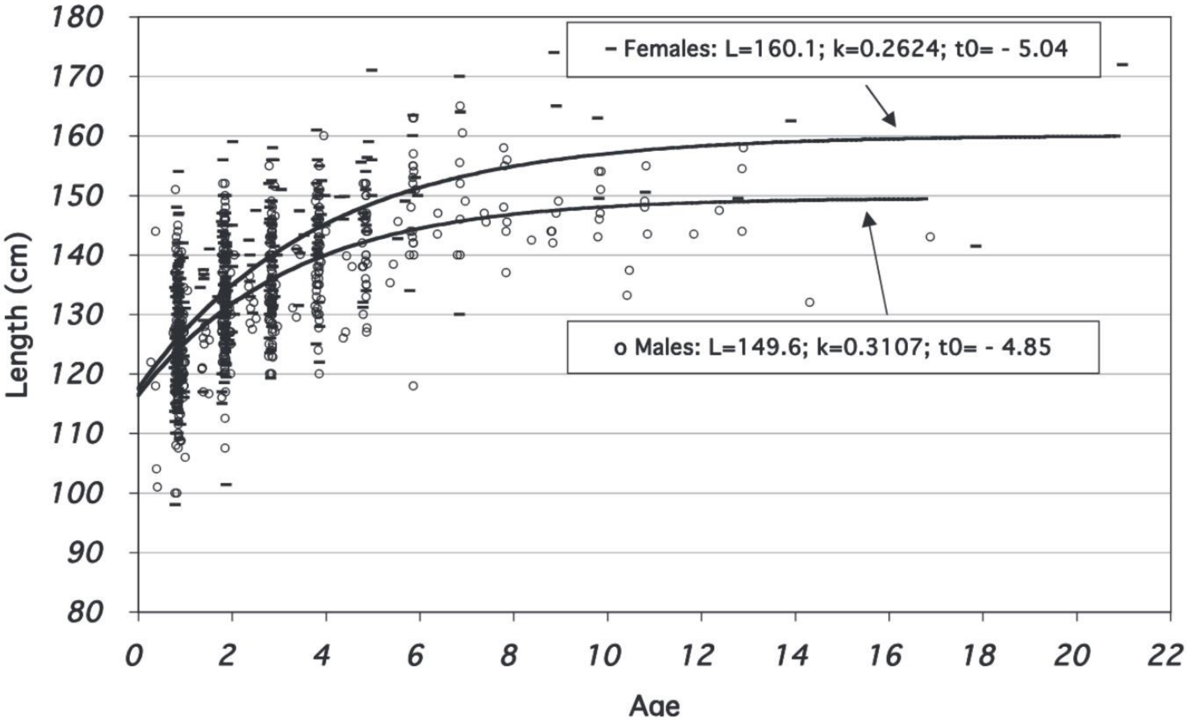

Figure 29 Length at age for harbour porpoises from Icelandic waters.

8.5 Lockyer et al. (2003): West Greenland

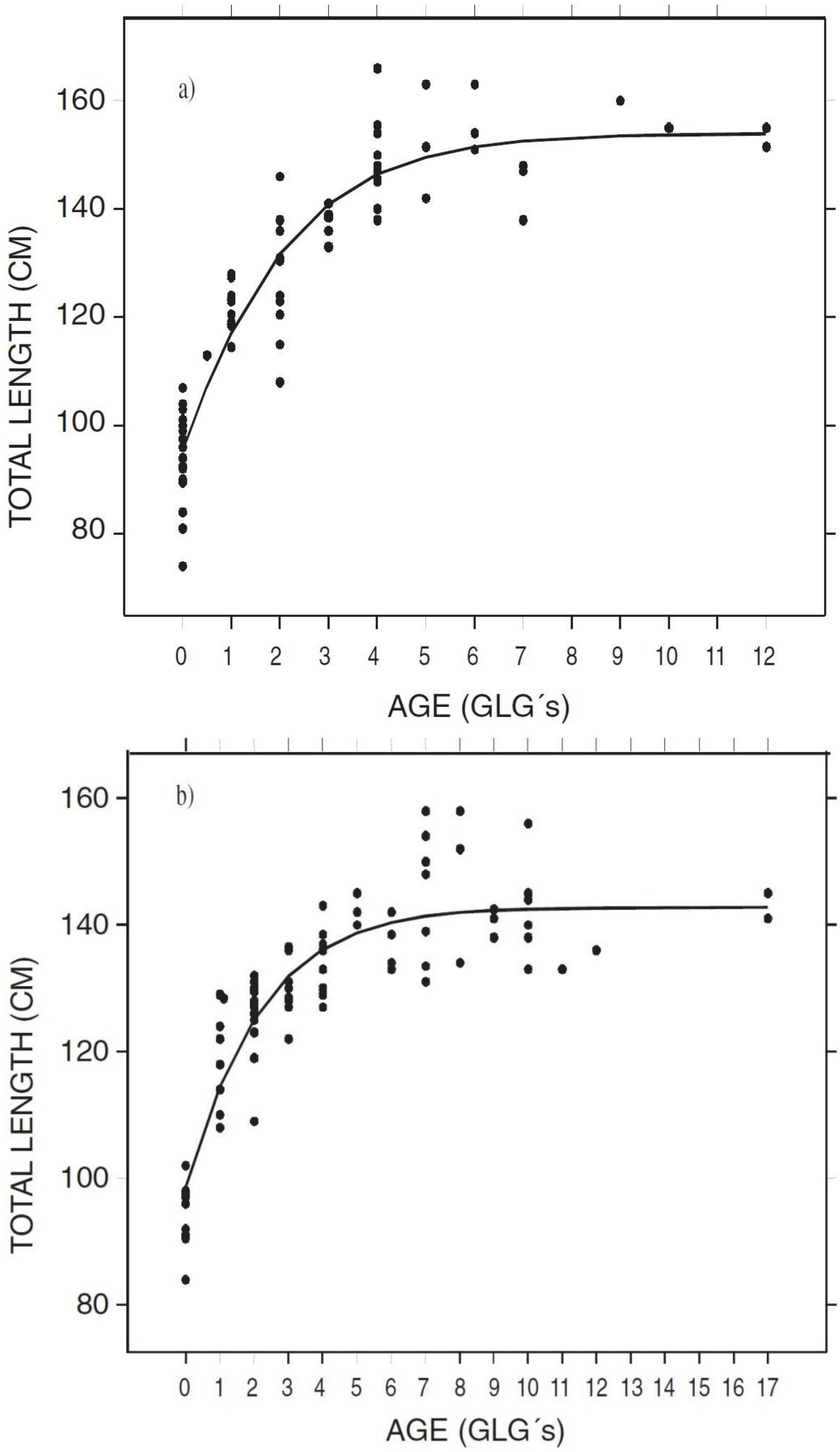

Figure 30 Length at age curve for female porpoises (n=85) (a), and male porpoises (n=91) (b), from 3 regions of West Greenland (1988, 1989 and 1995) with fitted Gompertz growth curves.

9 Appendix 2: Inter-user measurements

Table 6 Estimated harbour porpoise length from five independent users. The length range is the difference between the minimum and maximum lengths measured by the five users.

Animal Number	Length measurement (cm) by user					Length measurement summary			Image Grade
	1	2	3	4	5	Range	Average	Notes	
1	128.8	126.2	123.8	128.5	126.2	5.0	126.7		2
2	140.9	141.6	132.9	135.3	143.6	10.7	138.9	Splash at rostrum but not totally obscuring	2
3	141.9	139.7	142.5	141.8	139.3	3.2	141.0		3
4	125.9	124.3	126.5	122.4	122.8	4.1	124.4		2
5	140.9	139.2	138.6	139.9	143.9	5.3	140.5		2
6	129.4	117.5	124.4	128.1	126.4	12.0	125.1	Glare/reflection at rostrum but not totally obscuring	2
7	149.3	153.0	147.6	148.4	146.5	6.5	149.0		2
8	150.2	146.5	143.1	143.6	142.8	7.4	145.2		2
9	149.1	151.9	147.7	149.2	149.1	4.1	149.4		2
10	115.3	116.7	113.4	116.0	114.1	3.3	115.1		2
11	154.8	154.5	153.4	155.3	152.2	3.1	154.0		2
12	165.5	161.5	160.2	160.2	158.0	7.5	161.1		2
13	161.6	154.3	157.5	159.0	157.7	7.3	158.0		2
14	125.6	120.9	120.7	122.8	122.1	4.8	122.4		2
15	174.1	167.1	170.0	172.1	167.8	7.0	170.2		2
16	93.1	94.9	91.7	94.5	91.1	3.8	93.1		2
17	148.6	146.0	144.3	146.0	141.2	7.4	145.2		2
18	148.7	149.6	147.1	149.4	148.0	2.5	148.6		2
19	153.7	149.9	146.9	147.4	146.2	7.5	148.8		2
20	154.3	146.4	143.8	146.4	144.6	10.5	147.1		2
21	142.0	142.5	127.6	141.3	142.5	15.0	139.2	Obscuring splash at rostrum	3

Figure 31 Animal 18: Location of pixels identified by the multiple users as the rostrum and fluke points of measurement. Difference in length measurement: 2.5 cm.

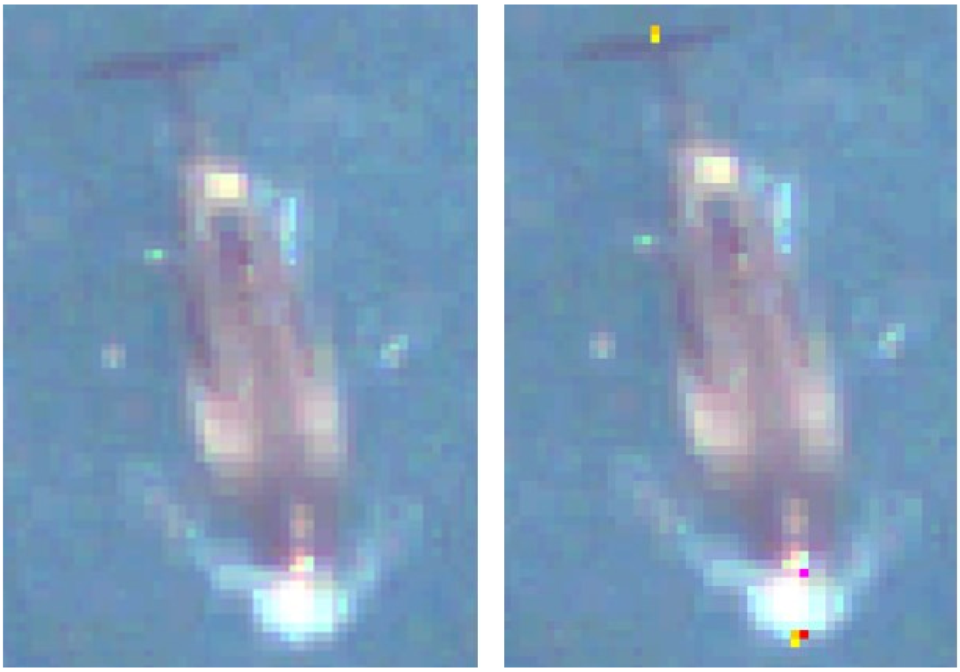

Figure 32 Animal 21: Location of pixels identified by the multiple users as the rostrum and fluke points of measurement. Difference in length measurement: 15.0 cm.

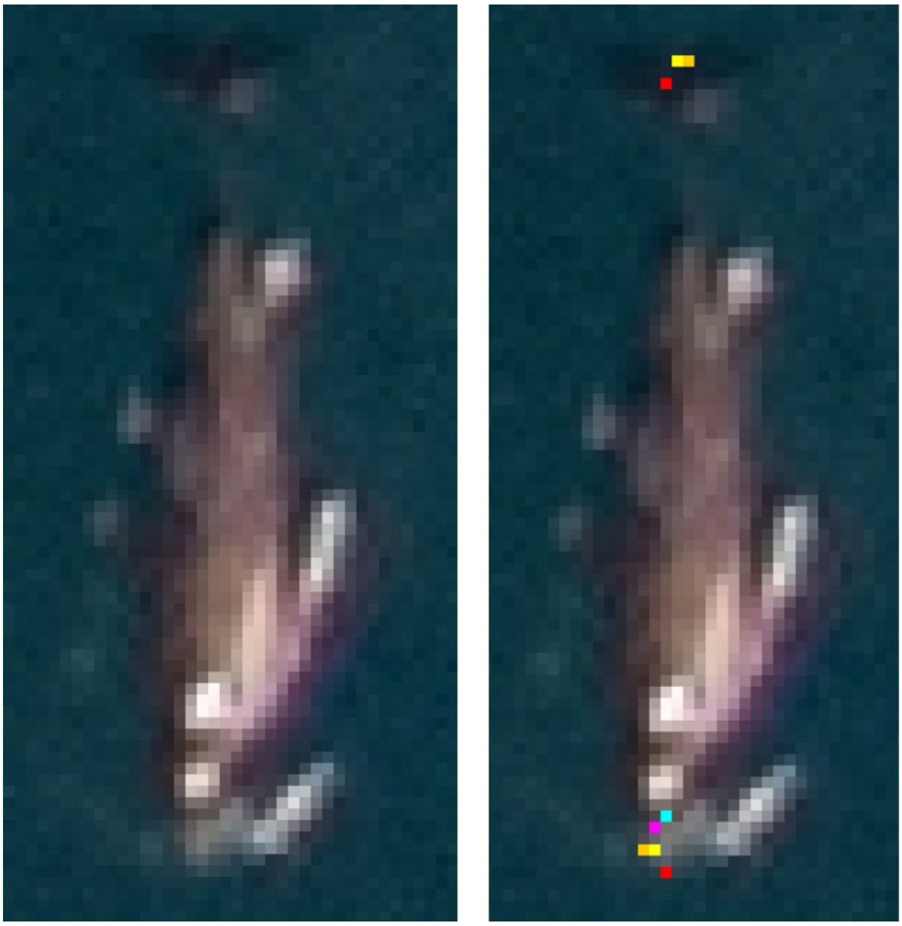

Figure 33 Animal 2: Location of pixels identified by the multiple users as the rostrum and fluke points of measurement. Range of length measurements: 10.7 cm. Issues noted: splash at rostrum.

Figure 34 Animal 6: Location of pixels identified by the multiple users as the rostrum and fluke points of measurement. Range of length measurements: 12.0 cm. Issues noted: glare or reflection at rostrum.

Table 7 Estimated harbour porpoise width from five independent users. The width range is the difference between the minimum and maximum widths measured by the five users.

Animal Number	Width measurement (cm) by user					Width measurement summary			Image Grade
	1	2	3	4	5	Range	Average	Notes	
1	34.2	32.9	33.3	35.0	34.6	2.1	34.0	Right side blurry	2
2	30.0	28.1	34.9	26.8	29.9	8.0	29.9	Ripples/splash at sides	2
3	34.6	35.4	38.4	37.3	34.6	3.9	36.1	Blurry edges	3
4	29.0	27.5	27.2	28.8	29.8	2.6	28.5	Right side blurry	2
5	37.4	35.6	33.9	37.4	37.5	3.6	36.4		2
6	32.7	33.5	35.1	33.9	36.6	3.9	34.4		2
7	30.4	33.3	29.6	30.1	34.0	4.4	31.5	Blurry edges	2
8	32.5	31.1	33.8	33.6	30.7	3.1	32.3		2
9	32.3	30.1	30.6	33.5	37.5	7.3	32.8	Low resolution	2
10	26.6	26.9	28.5	27.6	26.6	2.0	27.2	Low resolution	2
11	33.7	33.7	32.4	34.8	37.3	4.9	34.4	Blurry edges	2
12	35.7	34.6	37.2	37.6	34.8	3.0	36.0		2
13	39.0	35.8	38.4	37.6	36.8	3.2	37.5	Blurry edges	2
14	28.9	27.1	27.0	30.3	28.9	3.2	28.4		2
15	35.6	38.2	38.6	37.5	40.3	4.6	38.0		2
16	28.6	27.9	27.4	29.7	30.6	3.2	28.9		2
17	33.0	42.7	40.4	40.5	35.9	9.7	38.5	Body appears distorted	2
18	38.9	35.8	33.9	36.4	35.4	5.1	36.1	Right side blurry	2
19	35.2	38.9	35.0	37.0	36.3	3.9	36.5		2
20	28.6	27.4	23.4	27.4	29.8	6.4	27.3	White sides difficult to measure	2
21	38.3	39.8	39.3	37.8	40.3	2.5	39.1		3

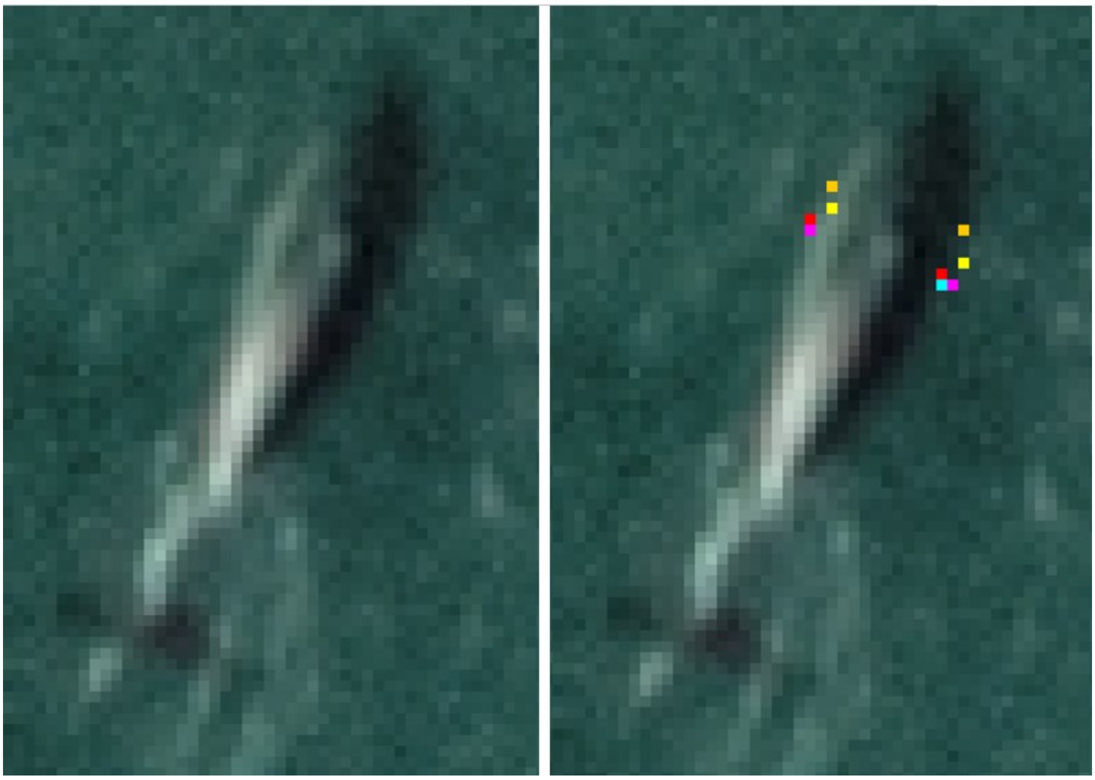

Figure 35 Animal 10: Location of pixels identified by the multiple users as the width points of measurement. Difference in width measurement: 2.0 cm.

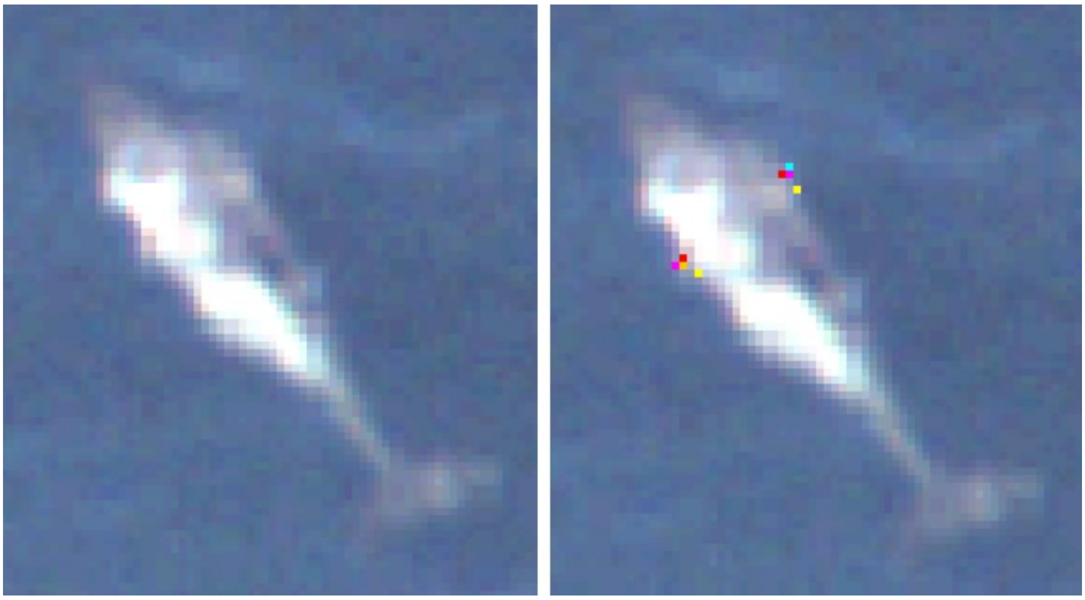

Figure 36 Animal 17: Location of pixels identified by the multiple users as the width points of measurement. Difference in width measurement: 9.7 cm.

Figure 37 Animal 2: Location of pixels identified by the multiple users as the width points of measurement. Difference in width measurement: 8.0 cm.

10 Appendix 3: Porpoise age frequency distributions

The age frequency distribution for bycaught and stranded harbour porpoise in Danish waters is presented in Lockyer and Kinze (2003). This demonstrates that the largest age group for both sexes was age 0-1 years (25% of females and 33% of males) and that less than 5% of the animals were older than 12 years. The same study estimated that age at sexual maturity for females was 3.3 years therefore, assuming the female ASM was 3 years, ~50% of the female sample was juvenile. The study also estimated that males matured at 3-4 years therefore, assuming the male ASM was 3 years, ~60% of the male sample was juvenile.

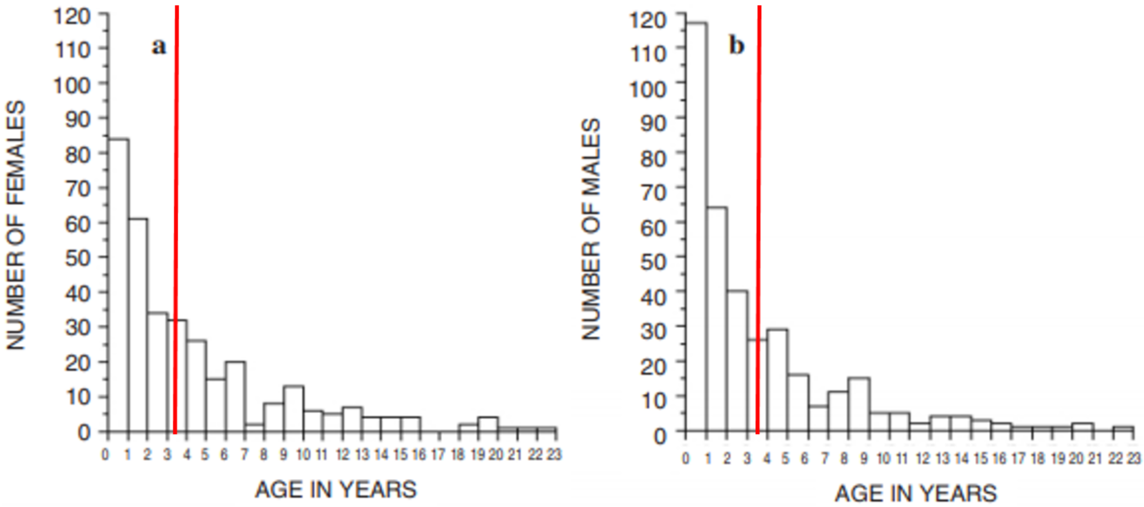

Figure 38 Age frequency distribution for stranded and bycaught harbour porpoise in Danish waters, a=female, b=male. Red line denotes estimated age at sexual maturity from the same study (F=3.3 years, M=3-4 years). Obtained from Lockyer and Kinze (2003).

The population age structure in both the North Sea and the Baltic Sea has recently been estimated by Kesselring et al. (2018) based on data obtained from 526 female harbour porpoise stranded in Germany between 1990 and 2016. The age of the stranded animals ranged between 0 and 22 years, with a mean age of 4.87 ± 0.2 years. As stated previously, there is no evidence that the harbour porpoise population in the North Sea has changed since 1994, and therefore it is considered to be a stable population. By contrast however, the harbour porpoise sub-population in the Baltic Sea is classified as Critically Endangered (Carlén et al. 2018). Therefore we would expect to see a difference in the population age structure between the North Sea (stable) and the Baltic Sea (declining), with a smaller proportion of juveniles expected to be present in the Baltic Sea population. Kesselring et al.

(2018) found that the average age of death for porpoise in the German North Sea was 5.7 ±0.27 years and for the German Baltic Sea was 3.67 ±0.3 years. Assuming that porpoise reproduce at 4.95 years of age, it was estimated that a total of 55% of female harbour porpoise in the North Sea and 27% in the Baltic Sea were adults (thus 45% and 73% juvenile respectively). This result is not as expected. The proportion of juveniles in the Baltic Sea sample here is much higher than the North Sea sample, which does not follow the expected stage structure for a declining population compared to a stable population.

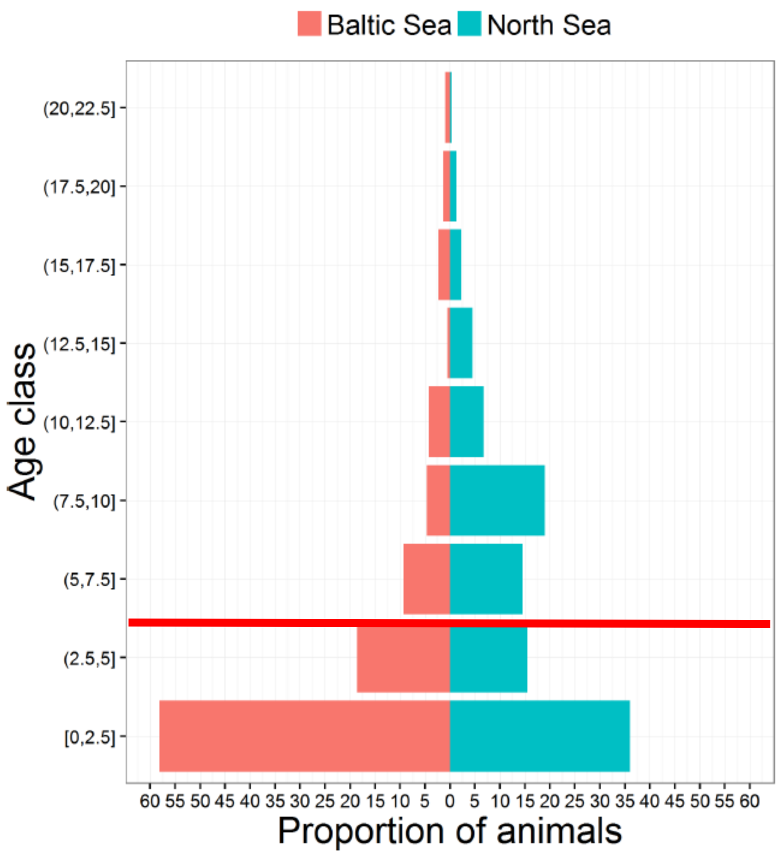

Figure 39 Population structure of female harbour porpoises of the North Sea (green bars) and Baltic Sea (red bars). Red line denotes estimated age at sexual maturity from the same study (4.95 years). Obtained from Kesselring et al. (2018).

A range of estimates for the age frequency distribution of harbour porpoise in UK waters are also available. The age structure has been shown in a study of harbour porpoise in British waters (Lockyer 1995a), where the proportion of the population aged <2 years was 47% for males and 34% for females, resulting in a total of 41% of the total population being <2 years old. A study of stranded harbour porpoise in Scottish waters between 2000 and 2005 found that 47% of the sample was juvenile (assuming ASM was 3 years) (Luque et al. 2009). Murphy et al. (2015) sampled 250 female harbour porpoise stranded around the UK between 1990 and 2012 and determined that 64% of the animals sampled were immature. Similarly, for stranded and bycaught porpoise in Scottish waters, Learmonth et al. (2014) found that ~61% of their sample had not reached 5 years old (assumed ASM).

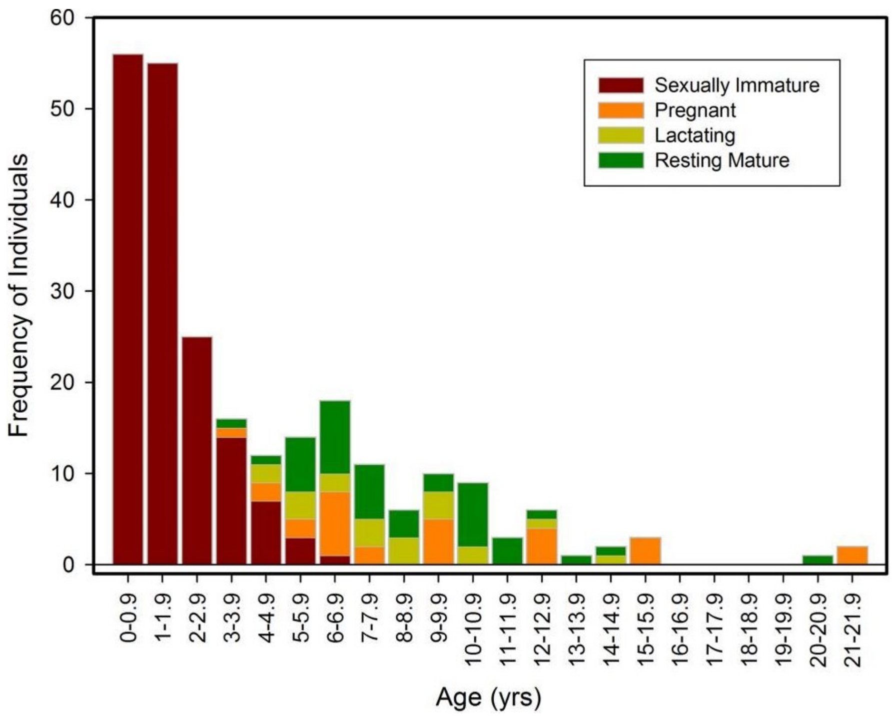

Figure 40 Age frequency distribution for female reproductive status categories in UK stranded harbour porpoises (n = 250). Obtained from Murphy et al. (2015).