

DEPARTMENT OF BIOLOGICAL AND ENVIRONMENTAL SCIENCES

Evaluation of photo-ID technique to estimate harbour seal numbers in Gullmarsfjorden, Sweden.



Simon Isaksson

Degree project for Bachelor of Science with a major in Biology BIO603, Bachelor in biology, exams, 30hp. First cycle Semester/year: Spring 2022 Supervisors: Karin Hårding (BioEnv, GU), Karl Lundström (SLU) & Daire Carroll (GU) Examiner: Johan Höjesjö (BioEnv, GU)

Photograph on front page by myself, Simon Isaksson

Table of content

Sammanfattning
Abstract
1 Introduction
1.1 Harbour seals in Kattegat-Skagerrak and conflict with fishery
1.2 Scientific aim
2 Materials and methods
2.1 Preparation
2.2 Field work
2.3 Wild-ID
2.4 Creating a data base
2.5 Evaluation of correct identification
2.5.1 How cropping size influence matching score9
2.5.2 Estimation of false acceptance and rejection rate9
2.6 An estimation of population sizes10
3 Results
3.1 Unique individuals11
3.2 Misidentification error
3.2.1 Cropping size affection on matching score
3.2.2 FAR and FRR
3.3 Populations magnitude14
4 Discussion
4.1 About the data base
4.2 About the unique individuals16
4.3 About the misidentification error16
4.4 About the estimated population
4.5 Preservation of seal and fish in the Gullmar fjord19
Conclusion
Acknowledgements
References

Sammanfattning

För att kunna bevara naturområden på ett varsamt sätt krävs en god förståelse för den biologiska mångfalden som lever där. Knubbsälen är en toppredator som påverkar och som påverkas av hela födokedjan och som spelar en avgörande roll i ekosystemet. Vissa sälar migrerar långa sträckor medan andra är relativt stationära, det är därför väldigt viktigt att vi lär oss förstå populationerna på en individuell nivå. Tidigare har man identifierat individer genom att fysiskt märka dem, i denna uppsats utforskades möjligheten att använda det naturliga pälsmönstret för att identifiera knubbsälsindivider genom fotoidentifiering. Bilderna processades i programmet Wild-ID. Dessutom uppskattades populationsstorlekarna på två olika lokaler i Gullmarsfjorden, Sverige. Identifikationsprogrammet fungerade bra och under projektets gång lyckades vi identifiera 58 potentiellt unika individer varav 22 identifierades med både dess högra och vänstra sida av ansiktet. Likhetspoängen som programmet gav de matchade bilderna var till 99% under 0.3 (0 = ingen likhet, 1 = exakta kopior) vilket är en relativt låg poäng, trots detta lyckades programmet ranka 75% av matchningarna som dess bästa förslag och 95% rankades som en av dess topp 10 förslag. Detta visar att även om likhetspoängen var väldigt låg kunde programmet ändå vanligtvis hitta den korrekta matchningen. Med den insamlade informationen och fångst-återfångststatistik kunde även populationsstorlekarna på två olika lokaler i fjorden uppskattas till 26 (12 - 60) individer respektive 73 (47 – 129) individer. Detta är första gången som identifiering av knubbsälsindivider utförs i Sverige med hjälp av foto-ID och är ett viktigt steg för att bättre förstå migrationer och spatiell struktur av knubbsälspopulationen. Denna information kan bidra till en bättre förvaltning och bevarande av våra marina ekosystem.

Abstract

To be able to preserve areas in nature in a careful way a good understanding of the biodiversity living there is needed. The harbour seals are a top predator that affect and is affected by the whole food chain and plays a vital role in the ecosystem. Some seals migrate long distances and some are relatively stationary, it is therefore very important that we can get to know the population on an individual level. Earlier, the way to identify an individual has been through physically tag them, this thesis explored both the possibility to use the natural patterns in the fur to identify individual harbour seals by Photo-ID. The images were processed using the software Wild-ID, the population sizes were estimated in two different locations in Gullmarsfjorden, Sweden. The software worked well, 58 potentially unique individuals were managed to be identified during the project, 22 of these with both their left and right side of their face. The similarity scores the software gave the matched images were to 99% scored beneath 0.3 (0 = no similarity, 1 = exact copies) which is a relatively low score. Still, the software ranked about 75% of the matches as the best suggestion and 95% as one of the top 10 suggestions. This shows that even though the scores were very low, the software could still usually pick out the correct match. With the information gathered and capture-recapture statistics, population sizes of two different locations in the fjord was estimated, 26 (12 - 60) individuals on one site and to 73 (47 - 129) on the other. This is the first time identification of harbour seal individuals in Sweden has been done with the help of photo-ID and it's a important step for better understand migration and spatial structures by the harbour seal population. This information can lead to better management and conservation of our marine ecosystems.

1 Introduction

Before the 20th century, marine mammals were seen as an important resource, but since the 20th century, seals have by many been regarded as competitors to commercial fishery (Hansen & Harding, 2006). Because of the perceived competition, bounty hunting was introduced, and the Kattegat-Skagerrak population of harbour seals (Phoca vitulina) decreased from about 18 000 individuals in the beginning of the 20th century to a size of 1 500 in the late 1960s (Heide-Jorgensen & Härkönen, 1988). Since then, marine mammals have been protected in Sweden, and many European countries, and the harbour seals in Skagerrak and Kattegat have therefore been able to recover to a population size of about 15 000 (Sokolova et al., 2018, Silva et al. 2021). The cod population in Kattegat on the other hand reached a spawning stock biomass of over 30 000 tonnes during the 1970s but have decreased with over 90% since then due to overfishing and environmental degradation of the marine ecosystem (SLU, 2020). According to Chen et al. (2005), cod migration is usually density-dependent, i.e. they move from an area when the abundance in the area gets too big. The migration can also be caused by currents, this is especially common for age 1 cod (Chen et al., 2005). Bjørge et al. (2002) found that a harbour seal eats on average 4 kg of fish per day and that cod make up about 20% of that diet during some seasons. However, the prey choice is highly opportunistic and varies from season to season, among regions and years (Scharff-Olsen et al., 2020). Hansson et al. (2017) writes in an article about the competition for fish between fisheries, seals and birds in the Baltic Sea that the pressure on the largest fish stocks, i.e. herring, sprat and cod, is relatively small by seals and birds if you compare it to commercial fishery. They also note that the fishery for these species is not much affected by seals or birds, while seals and birds may be exposed to competition from the fishery (Hansson et al., 2017). Seals in general do cost the fishing industry economic loss locally, grey seals in the Baltic sea is known both by damaging gear and by eating fish from the gear at specific fish traps (Westerberg et al., 2006).

The harbour seal in Scandinavian waters has recovered in abundance which have led to increased hunting quotas. Harbour seals are known to have high site fidelity (Härkönen et al., 1999) and the spatial distribution of where seals are allowed to be hunted is not regulated. A high local hunt on single colonies can have a large effect on genetic diversity in the local seal colonies and long-term viability (Silva et al., 2021). Therefore it is of uttermost importance to study spatial migrations and site fidelity, and to develop non-invasive efficient methods. Earlier studies have applied freeze brands (Härkönen et al., 1999), flipper tags (Hastings et al., 2012) and GPS-trackers (Thomas et al., 2011) which force you to physically mark the animals which all are invasive methods that puts the seals under stress. Several studies have developed photo-ID as a non-invasive tagging technique which works well on some species such as grey seals, where pelage patterns are sufficiently unique. Recently, software for photo-ID on harbour seals was examined (Langley et al., 2021) and had promising results with image analysis software to find matching patterns and thus re-observations. The data could be treated as mark-recapture statistics and provide information on seal abundance and migration patterns. The inventors of the software that later was decided to be used here, Wild-ID, wrote in their paper that the photographic mark-recapture method has three conditions that it requires. First requirement is that photographing the individuals can be made by either remote cameras, while captured or free ranging. Second, they have some individual pattern that make them distinguishable among others. Lastly, the pattern need to be stable during the projects period and that differing conditions doesn't hinder recognition when photographed (Bolger et al., 2012). This study was focusing on a local population at the Swedish west coast outside Lysekil.

1.1 Harbour seals in Kattegat-Skagerrak and conflict with fishery

Since the 70s, the cod population has declined drastically along the western coast of Sweden as well as in the region of the Gullmar fjord (Länsstyrelsen Västra Götaland, 2022). The fjord got protected 2012 by banning fishing of species like cod, haddock and pollock with the belief that this would make the cod population to return (Havs- och vattenmyndighetens författningssamling, 2011). Now, many years after, the population still have not recovered, and seals are sometimes blamed for eating all the cod. To be able to manage the area in a good way, we must get an understanding of not only the diet composition of the seals in the area, but also of how many harbour seals that search for prey inside the fjord and if it's the same individuals all the time or if it shifts. Depending on the result, the way of conservation could vary a lot. If it, for example, is the same 10 individuals that live in the fjord all the time, shooting 10 would eliminate the whole population. If the 10 individuals on the other hand shifts every now and then and the real number of individuals that is coming in and out of the area is a lot bigger, shooting 10 might not help reduce cod predation. Some people will prefer to keep the area free of seals to protect the cod while others want to conserve the population of seals that the area contains. Either way, it is critical to collect more knowledge about the population and make an inventory of the individuals in the area to be able to make a successful local management plan. For this, a trustworthy method needs to be found for identification on harbour seals and with it collect more knowledge of the population.

1.2 Scientific aim

The first aim of this study is to investigate the feasibility of using the photo-ID method on Scandinavian harbour seals. Another ambition is to set a protocol for how to collect, store and analyse photos of the animal. A further aim is to use the photos to identify individuals and with this estimate population sizes on some locations inside the fjord.

2 Materials and methods

2.1 Preparation

To be able to evaluate photo-ID as a method some first photos were collected and different softwares were evaluated. The project started with literature search of photoidentification in general and photo-ID on harbour seals in particular. Langley et al. (2021) found that among three softwares, ExtractCompare, Wild-ID and I³S Pattern, Wild-ID worked best for harbour seals (Langley et al., 2021). They ranked the quality of their images between 1 and 4 and used only those ranked \geq 3. Another study used amphibians to compare software's ability to recognize individuals and the software's they used was Wild-ID, I³S Pattern+, APHIS, and AmphIdent and with different qualities of the photos, AmphIdent performed the best (Matthé et al., 2017). After some research Wild-ID was chosen as the first choice in this study but another software might have worked just as well. As tools for photographing, a Canon EOS 90d camera with a Sigma 150-600 mm f/5-6.3 lens and a sturdy tripod was used. Game cameras were also used to see if that was an possible approach to take as well. Some photos were collected in Slottsskogen, a zoo in Gothenburg, of captive harbour seals. Photographing in Slottsskogen was done to try out how the software worked and get an appreciation in beforehand of what kinds of photos were needed for the identification to work, i.e. the quality of the photo and if whole body images, side of face or simply front face photos, were needed. The photos were decided to be saved as RAW-files and later converted into JPEG. This to save as much information as possible to better have the possibility to edit the photos in case of e.g. overexposure. With this material, an attempt to identify the individuals was made by analysing the images with the selected computer recognition software.



Figure 1: The two camera types. Three game cameras and one system camera were used during the project.

2.2 Field work

When in field, two locations in Gullmarsfjorden were selected where seals have been seen hauling numerous of times. These were Hågarnsskären, located in the outer fjord of Kristineberg, and Ärten, located in the inner fjord about 15 km from Hågarnsskären (figure 2). Ärten is a relatively high island with some vegetation which creates an opportunity to hide out of sight from the seals, making it more likely for the seals to come close to the island and be comfortable enough to lie on the rocks. Hågarnsskären on the other hand is some rocks pointing up over the surface. This makes it much easier for the seals to spot you which lower the chance of them going up on the rocks but enables you on the other hand to get closer to the water level.



Figure 2: Selected locations. Kristineberg marine station is located at the brink where the fjord goes out to the sea. Hågarnsskären and Ärten is located about 1 Nm and 9 Nm inside the fjords gap (image source: Google Earth).

According to local knowledge (unpublished) the seals are most active inside the fjord from spring until about midsummer, after midsummer observations are less common. This is the time when they give birth (SLU, n.d.) and the time when people start getting out more on their boats moving through the archipelago, and therefore a decision was made to only do the field work until midsummer and then focus on analysing the data to not stress the seals more than necessary. Three weeks were spent out in field, spread out over the projects first half, the 26-28th of April, 18-20th of May and the 10-12th of June. The weeks were spread like this to lower the risk of only getting the same individuals if they hung out there for only a short period and then moved on. When in field, the seals were photographed sometimes from the boat but most often from a nearby island. Photographing from boats worked when the water was relatively calm and it was possible to come at least within 50 meters from the seals, the closer the better. Most photos were instead taken from a nearby island, approximately 20-30 meters from their anticipated haul out rock. Usually, the photographer was dropped off in the morning on an island and then picked up again in the afternoon, in the best of cases without getting any attention from the seals. In these cases, several seals and several parts of them could be photographed on individuals on a pretty close range. This was found to be the most successful way of photographing and where therefore later chosen as preferred method. Game cameras was also rigged on some island by tying the cameras to steady rocks on the island and placed the rocks a couple of meters from the water's edge. This creates a possibility to photograph for a much longer period and decrease the risk of disturbing the seals as much by our presence.

2.3 Wild-ID

Between the weeks in field, time was used to go through the images taken previous and analyse them. The photos were sorted into different folders depending on date and location. The ones that were seen as good enough for analysing were cropped so most of the surrounding was removed and the RAW-file were converted into JPEG. The software Wild-ID could now be used to find the same individual on different photos. Wild-ID uses a SIFT (Scale Invariant Feature Transform) algorithm to compare images and identify patterns in photographs (Bolger et al., 2012). The algorithm takes information from the image and transform different features of the image into coordinates independent of scale (Lowe, 2004). According to Lowe, the inventor of SIFT, a 500x500 sized image usually generate about 2000 stable features. The software then saves these features in a database and when a new image is added, the software compares these new features with the database and tries to find a matching candidate based on the Euclidean distance of the features (Lowe, 2004). Another asset of SIFT is that it is relatively unaffected by scale, rotation etc., this is very important since it's usually very hard to get photos in the same angle and scale when it comes to photographing wild animals (See Bolger et al., 2012). This enables us to use a big number of images, and therefore to have a big sample and identify big populations, without having to go through and compare every single photo which would both be very time consuming and be practically impossible to do by eye. Instead, the software quickly compares the focal image with the database and takes forward only the top 20 best matches according to the software itself.



Figure 3: SIFT example. A self-made image for pedagogic reason to show how SIFT works. The two upper images represent the focal image, they are exactly the same. The lower images represent two different comparing images, the one to the left is a different image of the same individual as the ones on top and the one to the right is another individual. The red lines symbol some matching features between the images. Because the two images to the left is the same images and the two to the right is not, the two to the left should hopefully get a higher similarity score.

Since the software shows the top 20 best matches, these still had to been compared and paired together manually. In figure 4a and 4b you can see an example from when the software selects the best match. The upper and lower left photo is the same image and is the one that the software is trying to find a match to, the focal image. The square in the upper right contains the 20 suggestions that the software have chosen starting with the number 1 ranked suggestion furthest to the left and the lower ranked photos follows to the right. In the lower right appears the active comparison window showing the image that the focal image is compared to. Under

this photo sits the matching score as the goodness of fit score between the candidate matching and the focal image. If it's a match, the accept button is used, if not, the next images is examined and if no image match the focal image, the press 'reject all' option is used. Goodness of fit is a measurement of how well a predicted probability distribution matches observations, i.e. how well a focal image matches a comparison image (Kang et al., 2021).

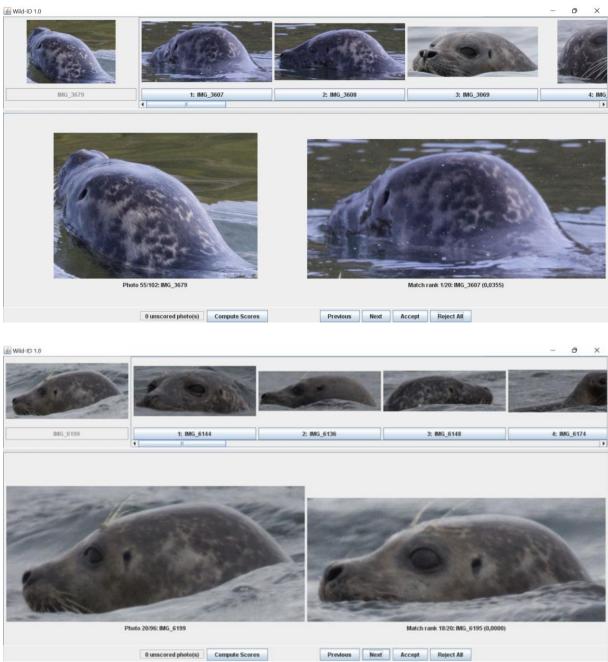


Figure 4a & 4b: Wild-ID. In figure 4a, the software takes forward a candidate that matches the focal image nicely and the observer have no real problem matching these two together. In figure 4b on the other hand is an example that shows that this is not always the case, here the only real match is ranked 18 of 20 and has a score of 0.0000 meaning that the software can't find any similarities worth noting (the score is actually 0.000001) meanwhile a observer can easily see that these the same individual.

2.4 Creating a data base

After going through all images in the software a txt-file was created noting the images matching score. In addition to self-notes taken in field, different images of the same

individuals could be paired together creating a data base with the various unique individuals. These files were named with the species name, part photographed, date, location and file name. An example of this could be PV00001_L_28_04_22_GÄ_IMG_3914. PV stands for Phoca vitulina. 00001 is the given number of this individual, 5 numbers were given because of the total population size of the species in our waters. The part photographed is shorted as L (left), R (right), F (front) of the face, N (neck), T (throat), BL (body left) or BR (body right). The date is written as dd_mm_yy. GÄ stands for Gullmarsfjorden Ärten and is the location for the photograph, Hågarnsskären would be called GH. IMG_3914 is simply the image name and helps while analysing.

The unique individuals were first sorted into folders for "unique for this day", later to "unique for this week" and finally into a folder of "unique for total time of project". This last folder could in the end be used to compare with the unique for each day and each week to compile data on which individuals that had been observed in each day or week and should be able to cover all individuals caught on camera during this period. An example of this could be seen in figure 5a and 5b where in fig. 5a the scores are from the 12th of June and show some representative matching scores (numbers to the right), some images where it was a match but the goodness of fit score is, as it often is, under 0.1 or 10%, and some where matching images from this folder could not be found ("NONE"). Figure 5b shows some of the scores from txt-file for comparing the unique individuals with images from the 28th of May. When the score is 1.000000 it means that the focal image is identical to the candidate matching image, proposing it is the exact same image. The score of e.g. 0.002002 indicates that it is two different images where the user have chosen to match the two, meaning that the photos match each other for about 0.2% according to the software but a user can still pair the two.

19	IMG_7604.jpg	7	IMG_7552.jpg	1	0,036502
20	IMG_7611.jpg	7	IMG_7552.jpg	1	0,005503
21	IMG_7616.jpg	7	IMG_7552.jpg	1	0,001003
22	IMG_7624.jpg	5	IMG_7532.jpg	1	0,073532
23	IMG_7627.jpg	20	IMG_7611.jpg	1	0,005503
24	IMG_7628.jpg	21	IMG_7616.jpg	1	0,073749
25	IMG_7632.jpg	18	IMG_7600.jpg	1	0,242721
26	IMG_7639.jpg	-1	NONE Ø NaN		
27	IMG_7641.jpg	-1	NONE Ø NaN		

240	28-04-22\IMG_3856_3680-2.jpg	165	PV00001_BL_28_04_22_GÄ_IMG_3856.jpg 1	1,000000
241	28-04-22\IMG_3866_3836.jpg	172	PV00001_L_28_04_22_GÄ_IMG_3914.jpg	0,002002
242	28-04-22\IMG_3876_3634.jpg	167	PV00001_BR_28_04_22_GÄ_IMG_3876.jpg 1	1,000000
243	28-04-22\IMG_3892_3687.jpg	166	PV00001_BL_28_04_22_GÄ_IMG_3892.jpg 1	1,000000
244	28-04-22\IMG_3896_3836.jpg	164	PV00001_BL_28_04_22_GÄ_IMG_3836.jpg 1	0,001502
245	28-04-22\IMG_3900_3634.jpg	171	PV00001_L_28_04_22_GÄ_IMG_3900.jpg	1,000000
246	28-04-22\IMG_3901_3687.jpg	168	PV00001_F_28_04_22_GÄ_IMG_3901.jpg	1,000000
247	28-04-22\IMG_3904_3634.jpg	2	PV00001_R_28_04_22_GÄ_IMG_3904.jpg	1,000000
248	28-04-22\IMG_3909_3680-2.jpg	172	PV00001_L_28_04_22_GÄ_IMG_3914.jpg	0,008003

Figure 5a & 5b: Matching score generated by the software. Fig. 5a shows some of the scores from the 12th of June meanwhile fig. 5b show some matching scores between the unique individuals and images from the 28th of May.

2.5 Evaluation of correct identification

To be able to evaluate the credibility of the software, the softwares ability to do the right choices was tested when identifying the different images. This was done by estimating scoring differences in cropping sizes and estimating the false acceptance rate (FAR) and the

false rejection rate (FRR). The FAR is an estimation of how likely the software is to match two images together that is not the same individual. The FRR is therefore the opposite, i.e. the probability that the software will miss to match two images that is the same individual.

2.5.1 How cropping size influence matching score

To evaluate the differences in matching scores depending on the images cropped sizes, a folder for unique seals for each day was used. There, every fifth seal were selected started on seal 2 and 5, this to randomize which images that are selected but not risk that two following images is selected that is a bit too alike. In this way 85 images are selected and two folders is made with copies of these in both, one where the images is normally cropped, i.e. with some background left, and one where the background is gone all together. The differences in scoring between these could then be tested. Note that only the highest scored match for each focal image were selected because of the difficulty of keeping track of every match. The software was run four times. One time with the folder containing cropped images where still some background was left and a second time with all the background outcropped. The matches made were noted for both folders before the software were used a third and fourth time when the folders were re-run but this time with the scores from the first two rounds at hand as well to compare with, this to make sure that no match that was available as a suggestion was missed on one but found on the other. The final number of matches as well as the average and median score for the different folders were then put together.

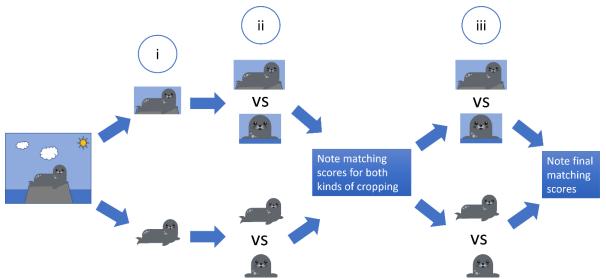


Figure 6: Flow diagram for the effect of cropping size. Selected images were transferred to two folders where (i) they in one folder was cropped with some background left and one without background. (ii) These were then each run in Wild-ID and the score and matched images were noted before (iii) the same folders were re-run in the software but this time with the results from earlier at hand. The final scores could then be put together.

2.5.2 Estimation of false acceptance and rejection rate

For an evaluation of FAR and FRR, inspiration was taken from the method of the software's creators (Bolger et al., 2012). Two images of the same side of 26 pairs of unique seals were selected. The photos were selected that you were able to identify the seal but not necessarily the exact same angle or only the same part of body, some could be that one side of face should be matched with the same side but of both face and body. In addition to these, another 18 images of different seals were added, making a stock of in total 70 images. This created an opportunity to make 26 false rejections (rejecting an image that is known to be a match) and

at least 18 false acceptance (matching two different individuals). One testing round was made to ensure that all pairs were at least ranked top 20 with its matching image, if it wasn't, one image was swapped with another of the same individual (examples of such occasions can be seen in the results) so that no FARs or FRRs were forced. The testing round was also made to note the matching images ranking and the score of them. Two persons that hasn't been included in the project were asked to do the matching after getting a small introduction in how the software works. With their results, a rate of false acceptance and rejection could be estimated. However, a problem with this method is that it more tests the user of the software than the software itself. Matching scores were therefore also collected from the whole project and their ranking, in total 295 scores, in an excel-file to calculate the software's FRR depending on different thresholds. This was done by selecting a range of thresholds, 0.9, 0.5, 0.25, 0.2, 0.1, 0.01, 0.001, 0.0001 and 0.000001, where the last one is the lowest score the software can give. Excels "IF"-function was used where every match with a score above or at the threshold was marked with a 0 and every match with a score below the threshold was marked with a one. A total for each threshold was summed and a false rejection rate of the software was calculated by dividing the sum with the total number of samples. Did was made for all matches with a ranking of top 20 as well as for top 10 and top 1. As done by Langley et al., (2021) the cumulative density function (CDF) were also calculated, which simply is the cumulated sum of matches found at one rank divided by the total number of matches available.

2.6 An estimation of population sizes

The third aim of the project was to get some population information about the studied seals in the fjord. A capture-recapture method could then be applied by the use of notes taken about days and locations each seal had been observed to calculate the population sizes for each island. Capture-recapture, or mark-recapture as it also is called, is a method where you collect a sample of individuals on one encounter, mark them and then later collect a new sample. Between the sampling occasions there must be enough time for the individuals to once again mix with the rest of the population, this is to not accidentally capture some individuals more often than you usually should in a natural setting (Urian et al., 2014). With a number of captured and recaptured individuals, Schnabels method could be used to calculate an estimation of the population size as follows (Krebs, 2014):

$$N' = \frac{\sum C_t M_t}{\sum R_t}$$
Variance $\left(\frac{1}{N'}\right) = \frac{\sum R_t}{(C_t M_t)^2}$
Standard error of $\left(\frac{1}{N'}\right) = \sqrt{Variance\left(\frac{1}{N'}\right)}$

where N' is the estimated population, C_t is the quantity of captured individuals at time t, M_t is the quantity of total marked individuals at time t, and R_t is the quantity of marked individuals at time t. With the help of a table in the book Ecological Methodology by Krebs (2014), ΣR_t in the equation for estimated population can be replaced with a number derived from the table and in that way calculate the upper and lower confidence limits as well.

3 Results

3.1 Unique individuals

Individual seals in Gullmarsfjorden were managed to successfully be identified through both manual recognition by researchers and by the automated software. All images that were used was taken with the system camera, the game cameras did unfortunately turn out not to work as well as aspirated. While you could get a rough estimate of how many seals that had been there, the images from the game cameras weren't sharp enough to enable us to pick out any individuals. During the project 59 'unique' individuals have been observed, photographed and saved to the data base. Of these, 22 have been identified with both the left and right side of the head, meanwhile of the remaining 37 individuals, 15 have been identified with only the left side, 17 with only the right, 5 individuals is solely of the front and not with either the left or the right side. Some of these 37 could still be two sides of the same individual but this without the possibility of determination.

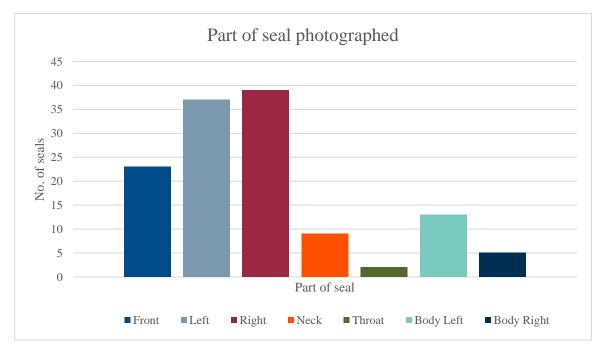


Figure 7: Photographed parts. This shows the number of the 'unique' seals on the y-axis and what specific part of the body that has been photographed on the x-axis. If the same individual has both been photographed on the left and right side, it will be included in both these columns. Note: Left and right side have been used to identify the seals and is therefore photographed on almost every seal except 5 where only the front was photographed. Neck, throat, and body pictures is almost only used on individuals that already have been identified with another image.

3.2 Misidentification error

3.2.1 Cropping size affection on matching score

As mentioned in section 2.5.1, the software was run four times. In the first round with the folder containing cropped images where still some background is left, 23 pairs were spotted. The second time the software was run it contained the folder with all the background outcropped and there 25 pairs were observed. The third and fourth times, these two folders were re-run but with the scores from the first two rounds at hand. The final number of matches for the different folders became 23 for the folder with some background and 25 for

the one without any background remaining. The average score for these were 0.017 for those with some background left and 0.021 for the ones without background. The median for these were 0.0010 vs 0.0015.



Figure 8: Many angels of one seal. The figure above shows different images of seal individual PV00001. All sides of this seal have successfully been identified and with this set of photos, any photo taken of this individual in the future should be able to be matched.

3.2.2 FAR and FRR

FAR and FRR were calculated to evaluate the credibility of Wild-ID as an identification software. During the first test round for the false rejection and acceptance rate, a total of 9 of our 26 pairs didn't show up as one of the top 20 suggestions. These are gathered in figure 9 which shows both the two first images that didn't show up as a match and what image that was added instead of one of the originals. In the first round, 10/26 matches were ranked as the top 1 suggestion, 15/26 were top 10 and 17 of our 26 pairs were ranked top 20. After the change of images for those not ranked as top 20 suggestions, 17/26 were ranked top 1, 23/26 as top 10 and all of them were at least top 20. The test persons had a cumulated false rejection rate of 15 percent and a false acceptance rate of 8 percent. The software's FRR by different thresholds is shown in figure 10. The number of top 20 matches were 295, 279 was ranked as top 10 and 219 was ranked as top 1. The similarity score for the thresholds 0.5 and 0.9 was excluded in the figure meanwhile 0.3 and 0.05 was added for esthetical reasons. The CDF for the top 1 ranked matches was calculated to 219/295 = 0.74, top 5 to 0.91 and top 10 as 0.95.

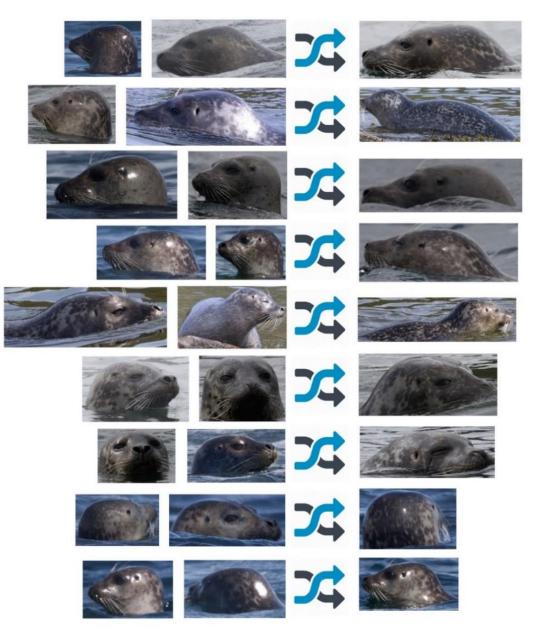


Figure 9: Exchanged images. To the left is the original match that didn't show up as a suggested match among the top 20 candidates, the right one of these got swopped against the image to the right. This gives an appreciation of what differences the software can cope with and, maybe more important, what it can't.

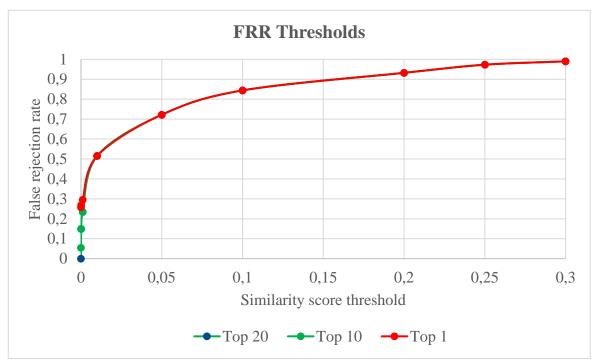


Figure 10: Thresholds for FRR. The rate of falsely rejected matches (y-axis) depending on different matching score thresholds (x-axis).

3.3 Populations magnitude

In total, 17 seals were observed and saved to the data base at Ärten and 44 individuals at Hågarnsskären. With this, the total population sizes were estimated to about 26 individuals at Ärten, with a variance of the reciprocal of the population (1/N) of 0.00025 and a standard error of 0.016. The 95% confidence limits for Ärten were calculated to 12 - 60. The island of Hågarnsskären on the other hand was estimated to a population of about 73 individuals, a variance of 1.1×10^{-5} , a standard error of 3.3×10^{-3} and a 95% confidence limit of 47 - 129 individuals.

Ärten	# caught	# recaptures	# new caught	Total tagged
1st capture	8	0	8	0
2nd capture	9	3	6	8
3rd capture	6	3	3	14
Total	23	6	17	17

Hågarnsskären	# caught	# recaptures	# new caught	Total tagged
1st capture	4	0	4	0
2nd capture	1	0	1	4
3rd capture	16	3	13	5
4th capture	7	1	6	18
5th capture	13	3	10	24
6th capture	13	8	5	34
7th capture	7	2	5	39
Total	61	17	44	44

Table 1a & 1b: Population size. The tables above contain the numbers of individuals observed at each sampling occasion and the numbers needed for estimating the populations sizes of Ärten and Hågarnsskären.

4 Discussion

This is the first time that software identification of harbour seal individuals has been done in Sweden and it is a big break through that the left and the right side of 22 unique seals have successfully been linked together. The project has given a lot of useful knowledge about photo-identification and hopefully this will be the beginning of a large data base of our Swedish harbour seal population.

4.1 About the data base

To begin with, one learned experience is that the side of the seals face is a good part to aim for and there is plenty of reasons for this. This area of the seal contains a lot more patterns than the front face and it got both eyes and ears which is two practical referencing points that the body doesn't have. Furthermore, the seals get scared more easily when they're on land and you will therefore usually not get as close to them as when they are in the water, and if they are in the water, the head is the only part of the seal that you will be able to photograph and the fur will be wet. Dry versus wet fur changes the colours of the fur and the patterns visibility, they tend to have a lighter colour when they are dry and the patterns isn't as visible meanwhile the suns reflection is more comprehensive when the fur is wet which can make it impossible to see any patterns at all. Important to mention here is to keep your distance to the seal, don't put unnecessary stress on them by going to close, let them come to you instead, many of them are curious by nature and are likely to come close enough if you only show some patience. A preferred method was also to sit on an island instead of trying to photograph from the boat. Reasons for this was that it is much easier to get a sharp image when using the tripod on land than when you are on a boat that is almost never completely still. On a boat the seals will also for sure know where you are all the time and will therefore mostly look straight at you, not giving you the chance to photograph the sides, on land you may have the possibility to hide or at least not seem threatening. The seals alternate between popping up their head over the water and swimming beneath the surface all the time so you have to stay alert and don't be too confident that it is the same seal that keeps popping up but if it is and you succeed on photographing different sides of it, be sure to make a note so that you remember it for later, these notes can later be decisive when you want to link different sides of the seals together. Another important part of this is the fact that the image name was saved in the filename. At first this wasn't the plan and it was only kept for the moment to easier know where to find the image again but it soon turned out to be a critical help many times and was therefore later decided to be a part of the name of the file for the unique seals images. A suggestion moving forward would be to focus on photographing the sides of the head but of course photograph the body as well if possible. Also make sure to do the homework in beforehand regarding islands the seals like to visit and if there are any possible observation spots nearby to photograph from. To spend some hours on one location gives you a better appreciation of the true number of seals that lives in that area than if you were to go with the boat from location to location and only stay for a couple of minutes. While sitting and hiding on the island Ärten during this project it got demonstrated several times that when a boat approached the island, the seals would usually disappear and then stay gone until the boat had moved on again. One example of this was on the 28 of April when the seals had been pretty active in the morning, laying for short times on the rocks. At 11:47 a man in a boat approached the island and stopped just by the rocks that the seals had explored not long before. No seals were observed during the mans visit except a couple of heads popping up some 300 meters away. At 12:13 the boat left again and the seals returned at 12:45. The bottom line with this is that the chance for you to get closer to the true number of individuals in an area is generally higher if you stay put on one location and try to get as little attention as possible instead of going from location to location hunting the seals.



Figure 11: The rock at Ärten. This photo is taken at 11:32 the same day as described in the text above and the boat put down the anchor about 5-10 meters to the right of the rock. On the image PV00001 is laying on the rock and looking at PV00009.

4.2 About the unique individuals

As mentioned in the results, not all the 59 "unique" individuals can for sure be said to be unique. 22 of these are unique with both their left and right side identified but for the remaining 37 only one side is photographed and therefore there is a chance that some of these could be the same individual after all and this must be considered both in the future and in the calculations made in this paper. Next part is to expend the data base and maybe in that way also pair together more left and right sides of the seals.

4.3 About the misidentification error

As seen in section 3.2.1, the cropping size do affect the software's matching ability. Because the software looks at the similarities for the whole image, it is important to erase as much unimportant parts of the image as possible. In our test, both kinds of images are cropped, but some still have a little part of the background left and some don't have any background at all. The result that the ones without any background had both more matches and a higher score shows that the software works better the less background the image has. The same thing was observed by Halloran et al. (2014) who writes in their article about photo-ID on giraffes that the only variable that has a significant effect on identification success is the complexity of the background and that the act of removing this reduced their identification error by over 50%. My recommendation for further studies is therefore to crop the images as much as possible,

preferably remove the background all together if the time is at hand but if not, just cropping down to a narrow square of the seal with only a small portion of the background left will probably be enough for the software to do its job.

Two other important factors for the ability of the software to match the images is the quality of the image and the difference in angle. Examples of this is demonstrated in figure 9. The first image shows the importance of quality where the one that got switched is almost identical to the one used instead in all aspects except the quality of the photo, the one removed wasn't even suggested as top 20 meanwhile the replacement got ranked top 1. The importance of angle is shown in many of them, but one is image 6 where the quality is almost better in the one that got rejected but the angle makes it too hard for the software to see the similarities, also here did the ranking of the match go from not even appear in top 20 to becoming the top 1. The importance of image quality has been observed in other articles as well, one of these is made by Bendik et al. where they compared the image quality effect on FRR using Wild-ID on amphibians. They found that the FRR decreased from about 16% with lower quality images to lower than 1% when using images of high quality, i.e. high resolution, good focus and a usable angle of the subject (Bendik et al., 2013). To get a better understanding of how angle and quality is affecting the identification of harbour seals it would be wise to explore this with further studies.



Figure 12: Kissing pair. It was a common sight to see the seals play around with each other, chasing each other and smack their flippers on the water surface. It seemed as some of them stayed together and at this moment PV00005 (one of two seals that has been observed at both locations) to the left and PV00003 to the right gave each other a gentle kiss.

The graph in figure 10 show us that 99% of the matching images gets a score lower than 0.3 and more than 50% gets a score lower than 0.01, i.e. 1% similarity. The fact that the top 1 ranked only has a false rejection rate of 0.26 at 0.000001 shows that about 75% of all matches is covered by only looking at the top 1 suggestions, this implicates that the software works good, even though it may not be perfect. One way to speed up the process could be to programme the software to pick out fewer matching candidates, this would maybe save some time but could also lead to that a few matches fails to be detected. The feeling is the same from using it, it may have some problems finding the similarities sometimes, even missing it completely at times, but it is easy to use and it does make the work a lot easier. It for sure wouldn't be possible to go through the images and find the matches without it but with the help of a software like this the project can now easily be scaled up and involve a lot more seals and get a larger sample size. The software is made using the patterns of giraffes and maybe that is where it works the best. During this project a new software was launched called SealNet, specifying on harbour seals (see Birenbaum et al., 2022). Unfortunately, it hasn't been possible to test SealNet as well, but it would be interesting as comparison and the software looks very promising. But the fact remains that you only have to crop the image a bit and then import the image before Wild-ID is good to go, meanwhile most other softwares, SealNet included, needs to get the eyes, nose and mouth marked before it is ready which, of course, also takes some time. The CDF for Wild-ID of 74% for top 1 ranked and 91% for top 5 can be compared with about 53% for top 1 ranked and 60% for top 5 ranked in the article by Langley et al. (2021) and the SealNets scores of 88% for top 1 ranked and 96% for top 5 (Birenbaum et al., 2022). The comparison may on the other hand be a bit misleading since the CDF and FRR is based on matches made, therefore top 20 ranked suggestions will for sure consist of 100% of the matches, this is certainly not the case for Wild-ID in Langley et al. and perhaps not in Birenbaum et al. either, the latter is a bit unknown of how they have proceeded. But it offers an appreciation on the proportion that has been ranked top 1 or top 5 and it provides us, nonetheless, with a perception of how well the software is working.

4.4 About the estimated population

As earlier noted, the estimated population sizes for the islands were 26 (12 - 60) for Ärten and 73 (47 - 129) for Hågarnsskären. Important to mention is that this is only a first estimation, more observations and sampling occasions are needed for making these numbers more trustworthy. One reason for this is the fact that at least two of the individuals have been observed at both locations which has been disregarded in the calculations. Another uncertainty is that only 22 of our 59 individuals have been identified with both the left and right side, meaning that some of the remaining 37 individuals still could be the same individual without our current knowledge. Ways to avoid the risk of identifying the same individual as different seals is to restrict the images to an adequate quality to lower the risk of misidentification, only use individuals that's been observed with both sides or abiding by one side of the seal though this will decrease the size of the data base (Hiby et al., 2012). But even though our data base creates an uncertainty for our estimations, it still gives us some appreciation of the population sizes and it retains a lot more information than the ones that only was identified on both or a particular side were to be eliminated. A last uncertainty to mention is the actuality that a various of time were spent on each location at different capture occasions. Some days, a whole day could be spent on the same island photographing all the seals that was observed. Other days the whole capture occasion composed of moving by with the boat, on such a day maybe only five minutes would be spent at the location trying to photograph the observed seals but probably without being able to capture all individuals present. Moving forward, a suggestion would be to pick some interesting locations, only spend full days on a single location or at least choose for how long you should stay there each time and make sure to cover the occasions fully and equally. If you have a great lens and a steady boat you could visit haul-out islands to be able to catch a bigger portion of the population but if not, the recommendation would be to pick an observation spot and wait for the seals to come to you. This takes a bit more preparation and beforehand knowledge but is, in my opinion, the more successful way to do it.

4.5 Preservation of seal and fish in the Gullmar fjord

With the data collected in this project an assumption can be made that the seal migrates between the archipelago and the outer parts of the fjord often. Proof of seals moving from the inner parts of the fjord to the outer parts is also collected and it is therefore also likely that they may move from the archipelago all the way to islands as Ärten. Observations for longer periods are necessary to be able to show if some individuals are stationary for the bigger parts of the year or if they all move in and out of the fjord. The seals are, as earlier mentioned, a highly opportunistic hunter and will eat the fish that is most common (Scharff-Olsen et al., 2020). For the moment, this isn't the cod and the seals impact on the cod population should, according to Hansen and Hårding (2006), be negligible. This suggests that a shooting of seals probably wouldn't have the desired impact.



Figure 13: Dinner time. The seals didn't play all the time, here PV00031 enjoyed a freshly caught dinner consisting of what seems to be a garfish.

Conclusion

Overall, the software has proven its use and that it is a massive help when you have to sort through a large data base for matching images, it surely isn't flawless but it simplifies the job enormously, is easy to use and gets most matches right. A project like creating a data base of unique individuals would feel completely overwhelming without a software like this. You can't pick out all unique individuals in one go with the software because it will miss some and you will miss some but by narrowing down the data base you will eventually get there. The time it will take will also decrease with experience, the number of images that got processed in a whole day in the beginning took only an hour or two after some time and this with a lower amount of false rejections. The risk of falsely accepting an individual because of the software is practically zero because of the fact that you have to accept the matches manually. As mentioned earlier, it would be interested to compare Wild-ID to the new and harbour seal specific SealNet. Maybe SealNet is the way to go but Wild-ID certainly gets the job done and is user-friendly. A better ability to match images with different quality or angles would be preferable but it is better that the software gives too low scores to the matches than the other way around. Unfortunately, a similarity score or ranking threshold where all suggestions over/under that could be disregarded couldn't be found, but it was shown that 99% of the matches got a score under 0.3 and that even though the similarity scores were so low, about ³/₄ of the matches got ranked as the top 1 suggestion and 95% as the top 10. A critical next step is to revisit the locations once the seals have moulted again to make sure that the patterns remain and can be used for identification in all future years. The aims of creating a data base and setting a protocol for the implementation was fulfilled and once we know that this can be used for the years ahead, it can then easily be extended to more individuals, evaluation of the effect of image quality, photographing angles, how dry and wet fur influence matching ability, differences between colonies, of cameras and so on. All this would expand our understanding of what is needed and preferred in the work of identifying individual harbour seals and is critical knowledge when it comes to understand the populations and to better care for nature.

Acknowledgements

I want to thank Karin Hårding for agreeing on being my supervisor and connecting me to other knowledgeable people and Johan Höjesjö for being my examinator. I also want to thank Karl Lundström for being my other supervisor and for always helping me out in field, thank you for all your time, your willingness to share your knowledge and everything else you have done. Last but certainly not least, I want to thank Daire Caroll for helping me with the whole analysis part, always there as a sounding board, for your never-ending patience when I had problem understanding and for all your encouraging words.

References

Bendik, N. F., Morrison, T. A., Gluesenkamp, A. G., Sanders, M. S., & O'Donnell, L. J. (2013). Computer-Assisted Photo Identification Outperforms Visible Implant Elastomers in an Endangered Salamander, Eurycea tonkawae. *PLoS ONE*, *8*(3), e59424. <u>https://doi.org/10.1371/journal.pone.0059424</u>

Birenbaum, Z., Do, H., Horstmyer, L., Orff, H., Ingram, K., & Ay, A. (2022). SEALNET: Facial recognition software for ecological studies of harbor seals. *Ecology and Evolution*, *12*(5). <u>https://doi.org/10.1002/ece3.8851</u>

Bjørge, A. (2002). Interactions between harbour seals, Phoca vitulina, and fisheries in complex coastal waters explored by combined Geographic Information System (GIS) and energetics modelling. *ICES Journal of Marine Science*, 59(1), 29–42. https://doi.org/10.1006/jmsc.2001.1137

Bolger, D. T., Morrison, T. A., Vance, B., Lee, D., & Farid, H. (2012). A computer-assisted system for photographic mark-recapture analysis. *Methods in Ecology and Evolution*, *3*(5), 813–822. <u>https://doi.org/10.1111/j.2041-210x.2012.00212.x</u>

Chen, Q. X., Chan, K. S., Lekve, K., Torstensen, E., Gjosaeter, J., Ottersen, G., & Stenseth, N. C. (2005). Population dynamics of cod Gadus morhua in the North Sea region: biological density-dependent and climatic density-independent effects. *Marine Ecology Progress Series*, *302*, 219-232. <u>https://doi.org/10.3354/meps302219</u>

Halloran, K. M., Murdoch, J. D., & Becker, M. S. (2014). Applying computer-aided photoidentification to messy datasets: a case study of Thornicroft's giraffe (*Giraffa camelopardalis thornicrofti*). *African Journal of Ecology*, *53*(2), 147–155. <u>https://doi.org/10.1111/aje.12145</u>

Hansen, B. J. L., & Harding, K. C. (2006). On the potential impact of harbour seal predation on the cod population in the eastern North Sea. *Journal of Sea Research*, 56(4), 329-337. https://doi.org/10.1016/j.seares.2006.06.004

Hansson, S., Bergström, U., Bonsdorff, E., Härkönen, T., Jepsen, N., Kautsky, L., Lundström, K., Lunneryd, S. G., Ovegård, M., Salmi, J., Sendek, D., & Vetemaa, M. (2017). Competition for the fish – fish extraction from the Baltic Sea by humans, aquatic mammals, and birds. *ICES Journal of Marine Science*, *75*(3), 999–1008. <u>https://doi.org/10.1093/icesjms/fsx207</u>

Hastings, K. K., Small, R. J., & Pendleton, G. W. (2012). Sex- and age-specific survival of harbor seals (Phoca vitulina) from Tugidak Island, Alaska. *Journal of Mammalogy*, *93*(5), 1368-1379. <u>https://doi.org/10.1644/11-mamm-a-291.1</u>

Havs- och vattenmyndighetens författningssamling. (2011). Havs- och vattenmyndighetens föreskrifter om ändring i Fiskeriverkets föreskrifter (FIFS 2004:36) om fiske i Skagerrak, Kattegatt och Östersjön.

https://www.havochvatten.se/download/18.571a313c1331de20a5f80002089/HVMFS2011-27-ev.pdf

Heide-Jorgensen, M. P., & Harkonen, T. J. (1988). REBUILDING SEAL STOCKS IN THE KATTEGAT-SKAGERRAK. *Marine Mammal Science*, 4(3), 231–246. https://doi.org/10.1111/j.1748-7692.1988.tb00204.x Hiby, L., Paterson, W. D., Redman, P., Watkins, J., Twiss, S. D., & Pomeroy, P. (2012). Analysis of photo-id data allowing for missed matches and individuals identified from opposite sides. *Methods in Ecology and Evolution*, 4(3), 252–259. <u>https://doi.org/10.1111/2041-210x.12008</u>

Härkönen, T., Harding, K. C., & Lunneryd, S. G. (1999). Age- and sex-specific behaviour in harbour seals Phoca vitulina leads to biased estimates of vital population parameters. *Journal of Applied Ecology*, *36*(5), 825–841. <u>https://doi.org/10.1046/j.1365-2664.1999.00434.x</u>

Kang, T.-H., Sharma, A., & Marshall, L. (2021). Assessing Goodness of Fit for Verifying Probabilistic Forecasts. *Forecasting*, *3*(4), 763–773. <u>https://doi.org/10.3390/forecast3040047</u>

Krebs, C. J. (2014). Ecological Methodology (3rd Edition). Benjamin Cummings.

Königson, S., Hemmingsson, M., Lunneryd, S. G., & Lundström, K. (2007). Seals and fyke nets: An investigation of the problem and its possible solution. *Marine Biology Research*, *3*(1), 29–36. <u>https://doi.org/10.1080/17451000601072596</u>

Langley, I., Hague, E., & Civil, M. A. (2021). Assessing the performance of open-source, semi-automated pattern recognition software for harbour seal (P. v. vitulina) photo ID. *Mammalian Biology*. <u>https://doi.org/10.1007/s42991-021-00165-8</u>

Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), 91–110. https://doi.org/10.1023/b:visi.0000029664.99615.94

Länsstyrelsen Västra Götaland. (2022). Fiskeregler i havet i Västra Götalands län. <u>https://www.lansstyrelsen.se/download/18.6ae610001636c9c68e5179bc/1653907000960/fisk</u> <u>eregler-havet.pdf</u>

Matthé, M., Sannolo, M., Winiarski, K., Spitzen - Van Der Sluijs, A., Goedbloed, D., Steinfartz, S., & Stachow, U. (2017). Comparison of photo-matching algorithms commonly used for photographic capture-recapture studies. *Ecology and Evolution*, 7(15), 5861–5872. https://doi.org/10.1002/ece3.3140

Scharff-Olsen, C. H., Galatius, A., Teilmann, J., Dietz, R., Andersen, S. M., Jarnit, S., Kroner, A. M., Botnen, A. B., Lundström, K., Møller, P. R., & Olsen, M. T. (2018). Diet of seals in the Baltic Sea region: a synthesis of published and new data from 1968 to 2013. *ICES Journal of Marine Science*, *76*(1), 284–297. <u>https://doi.org/10.1093/icesjms/fsy159</u>

Silva, W. T. A. F., Bottagisio, E., Härkönen, T., Galatius, A., Olsen, M. T., & Harding, K. C. (2021). Risk for overexploiting a seemingly stable seal population: influence of multiple stressors and hunting. *Ecosphere*, *12*(1). <u>https://doi.org/10.1002/ecs2.3343</u>

SLU Artdatabanken. (n.d.) *Knubbsäl*. Retrieved 2022-08-24 from https://artfakta.se/naturvard/taxon/phoca-vitulina-102708

SLU Artdatabanken. (2020). Torsk. Artdatabanken. <u>https://artfakta.se/naturvard/taxon/gadus-morhua-206142</u>, 2022-03-31

Sokolova, M., Buchmann, K., Huwer, B., Kania, P., Krumme, U., Galatius, A., Hemmer-Hansen, J., & Behrens, J. (2018). Spatial patterns in infection of cod Gadus morhua with the seal-associated liver worm Contracaecum osculatum from the Skagerrak to the central Baltic Sea. *Marine Ecology Progress Series*, 606, 105–118. <u>https://doi.org/10.3354/meps12773</u>

Thomas, A., Lance, M., Jeffries, S., Miner, B., & Acevedo-Gutiérrez, A. (2011). Harbor seal foraging response to a seasonal resource pulse, spawning Pacific herring. *Marine Ecology Progress Series*, *441*, 225–239. <u>https://doi.org/10.3354/meps09370</u>

Urian, K., Gorgone, A., Read, A., Balmer, B., Wells, R. S., Berggren, P., Durban, J., Eguchi, T., Rayment, W., & Hammond, P. S. (2014). Recommendations for photo-identification methods used in capture-recapture models with cetaceans. *Marine Mammal Science*, *31*(1), 298–321. <u>https://doi.org/10.1111/mms.12141</u>

Westerberg H., Lunneryd S.-G., Wahlberg M., Fjälling A. Reconciling fisheries activities with the conservation of seals through the development of new fishing gear: a case study from the Baltic fishery – grey seal conflict, *Transactions of the American Fisheries Society (World Fisheries Congress Proceedings)*, 2006 pp. 587–597