

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL



Echolocation click rates in narwhals (*Monodon monoceros*) as a conservation stepping-stone

Diana Amaral Marques

Mestrado em Biologia da Conservação

Dissertação orientada por:

Dr. Tiago A. Marques

Dr. Susanna B. Blackwell

2021

ACKNOWLEDGEMENTS

First, I would like to greatly thank my advisor, Tiago Marques, for helping me choose a theme for my thesis and for looking for an external co-supervisor to teach me about the acoustics of marine mammals. I am truly thankful for having been part of ACCURATE, an international project, which allowed me to have Susanna Blackwell as my co-supervisor and to get to know other research scientists from around the world.

I also want to thank Susanna Blackwell and Mads Peter Heide-Jørgensen for providing me the narwhals' data for my thesis. This thesis would not be possible without Susanna's support, who taught me how to use the Software MT Viewer and how to interpret narwhal sounds. I am very grateful for everything I learned with you, Susanna!

I was only able to have MT Viewer on my computer because of the help of a computer engineer, Helder Lopes. I really appreciate all the help that you gave me.

Again, I want to give my gratitude to Tiago Marques for helping me with the statistical analysis, and for creating "Espacinho especial", where, once a week, invited people came to talk about their professional path, subjects related to Biology, and discussions about other topics, like mental health. These zoom meetings were important psychology during the lockdown, and for my future career because I got to know more about the research that is being done with marine mammals.

Carolina Marques, a student at FCUL, helped me with some of the statistical analysis, so I want to thank you for your willingness and time to help me.

Finally, I am grateful for all the support from my family and friends.

This thesis was conducted under the ACCURATE project. The data was available via funds from the US Navy Living Marine Resources program (contract no. N3943019C2176).

RESUMO

Os narvais da Gronelândia Oriental estão em declínio, com uma estimativa de cerca de 1000 indivíduos, encontram-se dispersos e são especialmente sensíveis às alterações climáticas. Desta forma, a capacidade de adaptação dos narvais a mudanças que ocorram no seu habitat e na distribuição das suas presas encontra-se limitada. Ainda não há conhecimento acerca do número de subpopulações que poderão existir na Gronelândia Oriental, mas sabe-se que estas se encontram isoladas das subpopulações presentes na Gronelândia Ocidental. Adicionalmente, os narvais encontram-se no apêndice II da Convenção sobre o Comércio Internacional das Espécies Silvestres Ameaçadas de Extinção. Por todas estas razões, é compreensível a necessidade e importância de se obterem estimativas das densidades para as subpopulações de narvais ao longo dos anos, o que permitirá observar as suas tendências populacionais. A estimativa da densidade é fundamental para a gestão e conservação dos narvais. É necessário haver um conhecimento prévio sobre o comportamento acústico da espécie quando a monitorização acústica passiva é usada na estimativa de densidades. A monitorização acústica passiva nunca foi considerada na estimação da densidade de narvais, porém este método é confiável e permite a obtenção de estimativas de densidade precisas devido ao facto de ser um método que não é influenciado por condições climáticas extremas e funcionar corretamente em locais remotos, como o Ártico. A colocação de marcas acústicas (*acoustic tags*) que gravam os sons produzidos pelos narvais permite a deteção e contagem dos sons de ecolocalização. Estes dados são fundamentais na obtenção de uma estimativa média das taxas de cliques, e esta permite converter uma densidade de sons numa densidade de indivíduos. Além disto, a variabilidade das taxas de cliques obtidas entre os indivíduos da população deve ser baixa, permitindo uma estimativa da densidade populacional confiável. Assim, o objetivo desta dissertação é a obtenção, pela primeira vez, de uma taxa média de cliques de ecolocalização produzidos por cada indivíduo e por unidade de tempo para a subpopulação de narvais de Scoresby Sound, o que permitirá no futuro estimar a sua densidade através do método de contagem de indícios (*cue counting*).

Os dados fornecidos pelas marcas acústicas são um dos métodos mais apropriados para estudar os narvais devido à distribuição remota e comportamento vocal ativo destes. Nesta dissertação, os dados das marcas acústicas foram usados na obtenção de uma estimativa média das taxas de cliques a partir de oito indivíduos marcados, dois machos e seis fêmeas. A partir do som gravado foram anotados os períodos de ocorrência de cliques, onde os narvais podiam produzir vários cliques por segundo ou nenhum clique em cada segundo, e em períodos sem ocorrência de cliques, onde foi assumido que nenhum clique era produzido em cada segundo, porém o número exato do número de cliques produzidos por unidade de tempo durante o período de cliques não estava disponível. Assim, desenvolveu-se um esquema de amostragem para os segundos presentes nos períodos de ocorrência de cliques de cada marca, onde se teve em conta as diferentes durações de cada gravação na escolha do tamanho da amostra. Após isto, os cliques de ecolocalização foram contados para cada segundo presente nas amostras a partir do *Software* MT Viewer. Antes de se estimarem as taxas de cliques de ecolocalização para os dados de cada marca, foram removidos os períodos onde os dados acústicos não estavam disponíveis e aqueles que, na sequência direta da colocação das marcas acústicas e da perturbação que este processo representa, se verificaram não ser representativos da atividade vocal das baleias após a observação dos dados.

Com base nos dados fornecidos, cinco hipóteses foram testadas: 1) A probabilidade de produção de cliques não depende da profundidade; 2) O número de cliques produzidos durante o período de ocorrência de cliques não varia ao longo do tempo; 3) O número de cliques produzidos durante o período de ocorrência de cliques não varia com a profundidade; 4) O sexo dos narvais não influenciam as suas taxas de cliques de ecolocalização; e 5) O comprimento dos narvais não influencia as taxas de cliques de ecolocalização.

Após serem modelados todos os dados no R, observou-se que a captura e marcação dos narvais levou a uma abstenção na produção de sons por um longo período de tempo, sendo que cada indivíduo apresentou uma resposta comportamental diferente. Também foi verificado que a produção de sons de ecolocalização foram frequentes a partir de profundidades superiores a 400m. O tamanho da amostra escolhida para as amostragens realizadas verificou-se suficiente e confiável para a obtenção de taxas de cliques por segundo, já que os coeficientes de variação destas foram inferiores a 10% para cada baleia. Com os dados das amostragens realizadas, foi verificado que os narvais ecolocalizaram regularmente ao longo do tempo e que houve uma maior produção de cliques com o aumento da profundidade durante os seus períodos de ocorrência de cliques. Para os dados de cada marca, foram estimadas taxas de cliques de ecolocalização e o valor obtido para a média não ponderada destas taxas foi de 1.40 cliques/s com um coeficiente de variação igual a 11.41% e o valor obtido para a média ponderada (com a ponderação a ser a duração de cada marca) foi de 1.27 cliques/s com um coeficiente de variação igual a 10.20%. As médias das taxas de cliques mostraram-se ser fidedignas e poderão, no futuro, tornar possível a estimação da densidade a partir da monitorização acústica passiva. Recomendamos que a média a usar no cálculo da densidade deva ser a média ponderada, já que esta tem em conta a variabilidade das diferentes durações das marcas. Não foi possível inferir se as taxas de cliques variaram com o sexo ou tamanho dos narvais. No entanto, é necessário verificar se estas taxas não apresentam realmente diferenças entre machos e fêmeas usando um maior número de indivíduos, visto que no total apenas tínhamos oito indivíduos, entre os quais apenas dois machos. Foi ainda possível obter uma taxa de cliques para a segunda metade dos dados de um dos nossos machos, cujos sons ainda não tinham sido processados, ou seja, tendo apenas disponíveis as profundidades sem a parte acústica. Este resultado mostra algo que há partida poderia não ser óbvio: mesmo marcas sem acústica podem conter informação relevante para estimar as taxas de produção de sons, em particular nestes animais cuja profundidade reflete de forma clara o seu estado comportamental. Adicionalmente, no futuro, seria interessante perceber quais os fatores que poderão influenciar as taxas de cliques, nomeadamente a distribuição das presas dos narvais, e perceber se as médias das taxas de cliques obtidas nesta dissertação poderão, ou não, ser generalizadas para outras áreas. Também seria importante confirmar ou refutar a capacidade de adaptação dos narvais perante alterações no seu habitat e distribuição das suas presas durante a sua atividade de alimentação, já que cada população apresenta diferentes dietas, e consequentemente, poderá haver um maior leque de opções na escolha das presas, assim como nas preferências de profundidades para a captura destas. Com base nos resultados desta dissertação, estamos bastante mais próximos de conseguir obter uma estimativa da densidade para a subpopulação de narvais de Scoresby Sound por contagem de sons via acústica passiva. O que fica por estimar é essencialmente a probabilidade de detenção dos sons e a realização de uma amostragem de monitorização acústica passiva.

Palavras-chave: Gronelândia Oriental, marcas acústicas, taxa de cliques, comportamento acústico, estimativa de densidade

ABSTRACT

Narwhals in East Greenland are dispersed, declining, and are particularly sensitive to climate change making their responses limited to adjust to changes in their habitat and prey resources. Density estimation is fundamental for narwhal's management and conservation. If passive acoustic approaches are to be considered, knowledge about the species acoustic behavior is fundamental. Consequently, it is important to obtain a mean cue rate estimate that could be used for converting a density of cues into a population density.

Animal borne tags are one of the best approaches to study narwhals given their remote distribution and active vocal behavior. In this thesis, tag data were used to obtain a mean click rate estimate based on eight tags from two males and six females. The sound had already been processed for *clicking* and *non-clicking periods*, but the exact number of clicks per time unit during *clicking periods* was not available. We therefore developed a sampling scheme and counted the number of echolocation clicks per second with the help of the Software MT Viewer. These data were then further processed in R.

Tag data showed that narwhals were affected by their capture and tagging. It was also verified that narwhals were regularly echolocating, both buzzing and clicking, at depths deeper than 400m. With sampling data, it was verified that narwhals echolocate regularly over time and that there was a tendency to make more clicks with deeper depths during their clicking time. For each tag record, echolocation click rates were estimated. Even if based on a small sample size, a key result is an overall cue rate estimate for the species. The estimated cue rate was 1.40 clicks/s with a coefficient of variation of 11.41% and 1.27 clicks/s with a coefficient of variation of 10.20%, based on unweighted and weighted averages, respectively, where record lengths were the considered weights. Finally it was not possible to show whether click rates vary with narwhals' sex and length. However, it is necessary to confirm these results with a larger sample size, since our data only had eight individuals with just two males. Furthermore, in the future, it would be interesting to know whether click rates are influenced by narwhals' prey distribution.

Keywords: East Greenland, tags, cue rate, acoustic behavior, density estimation

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	I
RESUMO	II
ABSTRACT	III
TABLE OF CONTENTS	IV
LIST OF TABLES	V
LIST OF FIGURES	V
LIST OF EQUATIONS.....	VI
LIST OF ABBREVIATIONS	VII
1 INTRODUCTION.....	1
1.1 Cue rate and density estimates with PAM and tag data	1
1.2 Narwhals characteristics and threats to their populations	2
1.3 The acoustics of narwhals and diving behavior.....	5
1.4 Narwhals from East Greenland	6
1.5 Thesis aim and hypotheses	7
2 METHODS.....	8
2.1 Study area	8
2.2 Acoustic data	8
2.3 Statistical analysis	13
3 RESULTS.....	16
3.1 Effects of tagging	16
3.2 Sample size effect on cue rate estimates	16
3.3 Relationship between depth and probability of foraging sounds produced.....	19
3.4 Trends of the number of clicks over time and depth	19
3.5 Click rate estimates	21

3.6	Test for the influence of narwhals sex and length in echolocation click rates	21
3.7	Click rate for the last part of Jonas' tag.....	23
4	DISCUSSION	23
4.1	Effects of tagging	23
4.2	Sample size effect on cue rate estimates	24
4.3	Relationship between depth and probability of foraging sounds produced.....	24
4.4	Trends of the number of clicks over time and depth	25
4.5	Click rate estimates	25
4.6	Test for the influence of narwhals sex and length in echolocation click rates	26
4.7	Click rate for the last part of Jonas' tag.....	27
4.8	Final considerations.....	27
5	REFERENCES.....	28
6	APPENDIX	32

LIST OF TABLES

Table 1.1:	Characteristics of the individual narwhals considered in this study.....	9
Table 1.2:	Table with respective values for each sampling data	21

LIST OF FIGURES

Figure 1.1:	Representation of 11 narwhal stocks and distribution of their summer and winter grounds by Louis et al. (2020). The numbers after the name of the stock are the number of samples used in the Louis et al. (2020) study.....	3
Figure 1.2:	Representation of the 17 coastal areas where narwhals spend summer in Canada and Greenland by Chambault et al. (2020).	4
Figure 1.3:	Representation of the study area of this thesis. Map of Scoresby Sound fjords and inlets made by Blackwell et al. (2018). The thick red lines delimit the four areas represented in the map (Blackwell et al. 2018).	7
Figure 1.4:	Simplified schematic figure that represents how the clicks and silence are present in MT Viewer and how a cue rate can be calculated.....	12
Figure 1.5:	Presence (in color blue), absence (in color pink) and unknown (in color black) of echolocation behavior by each narwhal as a function of depth and days since capture and tagging. Black lines represent the days from which narwhals start their foraging behavior. After the red line, in Balder's dataset, acoustic information was not available, and in the second part of Jonas' dataset (after the dark blue line) sound was available but not processed.....	14
Figure 1.6:	Presence (in color blue) and absence (in color pink) of echolocation behavior by each narwhal as function of depth and days since capture and tagging. These data were corrected by deleting long periods of silence as well as periods in which acoustic information was not available or processed.	15
Figure 1.7:	Plots showing how the increasing sample sizes change the estimate of the mean obtained and how the variances of the mean and their coefficients of variation change with increasing sample sizes for Freya1 dataset (first sampling made from Freya's record). The green line represents the mean obtained for Freya1 dataset.	17
Figure 1.8:	Plots showing how the increasing sample sizes change the estimate of the mean obtained and how the variances of the mean and their coefficients of variation change with increasing sample	

sizes for the Freya2 dataset (second sampling made from Freya’s record). The green line represents the mean obtained for Freya2 dataset..... 17

Figure 1.9: Plots showing how the increasing sample sizes change the estimate of the mean obtained and how the variances of the mean and their coefficients of variation change with increasing sample sizes for Freya1+2 dataset (the two samplings made from Freya’s record were pooled). The green line represents the mean obtained for Freya1+2 dataset. 18

Figure 1.10: Plot showing how the increasing sample sizes change the estimate of the average obtained by a sub-sample of k samples out of 100 that were selected for the Freya1 and Freya2 datasets, the first and second samplings made from Freya’s record, respectively. The blue line represents the mean obtained for Freya1 dataset, and the red line represents the mean obtained for Freya2 dataset. The blue and red points are the average values obtained for the Freya1 and Freya2 sub-samples, respectively. The grey points are the average values for the Freya1 and Freya2 sub-samples that were overlapped. 18

Figure 1.11: Plot showing the probability of echolocation clicks produced by narwhals (“Prob.Clicking”) with whale variable as a factor along with the depth (in meters). Each narwhal is represented by a different color..... 19

Figure 1.12: Plot showing the trend of the number of clicks during the *clicking period* over time. The time variable is the time differences in minutes between the first sample selected with the other samples selected for each sampling data. Each narwhal is represented by a different color. 20

Figure 1.13: Plot showing the trend of the number of clicks during the *clicking period* with deeper depths in meters. Each narwhal is represented by a different color. 20

Figure 1.14: Plot showing the relationship between the click rates and narwhals’ length, in centimeters, for each sex. Color pink and blue represent the females and males, respectively..... 22

Figure 1.15: Boxplot of the click rates for females and males. The black line (highlighted) represents the click rate medians obtain for females (1.49 clicks/s) and males (1.10 clicks/s). The first quartile is the line after the minimum click rate (present at the lower extreme) and the third quartile is the line before the maximum click rate (present at the upper extreme). In the males' boxplot, the minimum and maximum have the same values as the first and third quartile, respectively, because our data have only two males..... 22

LIST OF EQUATIONS

Equation 1.1: General formula to estimate animal density..... 9

Equation 1.2: Formula to estimate animal density from a cue counting approach..... 10

Equation 1.3: Formula to estimate echolocation click rate..... 11

Equation 1.4: Formula to estimate the mean number of clicks per second in tag i 12

Equation 1.5: Formula to estimate variance of echolocation click rate 12

Equation 1.6: Formula to estimate coefficient of variation of echolocation click rate in percentage.. 12

Equation 1.7: Formula to estimate a simple average cue rate (unweighted average)..... 13

Equation 1.8: Formula to estimate a weighted average cue rate, where the weights are the different tags durations 13

LIST OF ABBREVIATIONS

AIC – Akaike Information Criterion

Freya1 – First sampling made from Freya’s dataset

Freya2 – Second sampling made from Freya’s dataset

Freya1+2 – The combined two samplings made from Freya’s dataset

GAM – Generalized Additive Model

GLM – Generalized Linear Model

PAM – Passive Acoustic Monitoring

\hat{r} – Echolocation click rate

1 INTRODUCTION

1.1 Cue rate and density estimates with PAM and tag data

Sounds produced by cetaceans are essential to their survival because these marine mammals depend on them to interact with the environment, to forage, communicate, escape from dangers, orientate, and navigate (Weilgart 2017). Therefore, cetacean sounds (e.g., echolocation clicks, burst pulses, whistles,...) reflect their behavioral state and are important to conservation since they can be analyzed using PAM (Passive Acoustic Monitoring) to inform about ecology and even to produce population density estimates (Marques *et al.* 2009; Nowacek *et al.* 2016; Blackwell *et al.* 2018). PAM data are collected through recorders deployed on the seafloor or behind a vessel, giving information about the sound present in a habitat (Merchant *et al.* 2015). Density estimates over time allow the detection of population trends, of anthropogenic detrimental effects, and of the impacts of oceanographic and climate changes (Marques *et al.* 2013).

Many cetacean species are perfectly suited for PAM methods since they spend most of their time underwater, produce species specific sounds that can be readily detected, and occur over wide areas at low densities, which make visual survey methods difficult to implement (Marques *et al.* 2009). Contrary to visual surveys, PAM has the advantage of operating at any time of the day, with oceanographic and weather conditions having less influence on data collection (Marques *et al.* 2013). In addition, PAM allows working in remote locations, like the Arctic, where extreme weather conditions along the year make human labor highly challenging and at times impossible (Marques *et al.* 2013). Acoustic monitoring can occur over longer time scales, leading to an automated way of gathering and processing data (Marques *et al.* 2013).

Cues produced by cetaceans, like echolocation clicks, calls, and songs can be used in a type of distance sampling method, referred to as cue counting, where the cues are detected and counted allowing a cue density estimation (Hiby & Ward 1986; Marques *et al.* 2013). Cue rates obtained with this method are fundamental to obtain a cue density estimate, because we need a cue rate to convert a density of sounds into a density of animals (Marques *et al.* 2013).

The collection of acoustic data can be done with acoustic tags attached to the back of cetaceans (Marques *et al.* 2013). Acoustic tags are mobile platforms that record sound for several days and can record additional information, like tracking the movement of the tagged individual (Burgess *et al.* 1998; Johnson & Tyack 2003). These tags can be used for cue rate estimation, in addition to potentially revealing other relevant information, namely the link between cetacean movement and vocal behavior, which is important for conservation as well (Marques *et al.* 2013). In this case, a tag can provide a direct cue rate estimate for the tagged animal, and provided a representative sample of animals is obtained, we can turn those into a population average cue production rate, which in turn can be used to convert a density of cues into a population density (Marques *et al.* 2013). To get a density estimate, it is necessary to associate the tag data with data collected with, for example, fixed acoustic sensors over a given time and, hopefully, with a reasonable design over the space for which one would like to produce density estimates (Marques *et al.* 2013).

Factors that hamper PAM density estimates are low population densities, a wide variation in cue rates, and a low cue duty cycle (little amount of vocalization) because these will lower the odds of detecting cues on recorders, and hence there will be higher variability and lower samples sizes (Marques *et al.* 2013; Blackwell *et al.* 2018; Scheidat *et al.* 2019). In addition, a reliable and adequate—in the sense of being valid for the time and place the survey is conducted in—mean cue rate estimate required for the density estimate is also hard to obtain since it might be dependent on behavior, and on differences

between individuals like their reproductive status, age, sex, etc (Thomas & Marques 2012). Therefore, for reliable density estimation using PAM, it is essential to have good knowledge of the species' acoustic ecology and behavior, as well as understanding its biosonar properties, such as the beamwidth and frequency composition of the sounds used for cue counting (Macaulay 2020).

1.2 Narwhals characteristics and threats to their populations

Narwhals (*Monodon monoceros*) are toothed whales belonging to the Monodontidae family (Heide-Jørgensen 2018). The majority of males present a tusk, a tooth that grows on the left side of the maxillary bones (Porsild 1922; Hay 1984; Graham *et al.* 2020). The tusk can reach 3 m and occurs uncommonly in females (Porsild 1922; Hay 1984; Graham *et al.* 2020). There is some evidence that tusks might be used during male contests and to attract females (Graham *et al.* 2020). However, this is not totally clear since tusks could possess other functions (Graham *et al.* 2020). Narwhals are long-lived, with maximum estimated ages of over 115 ± 10 years (Garde *et al.* 2007). Males have a mean body length of about 4.5 m, and reach maturity at 12-20 years old (Garde *et al.* 2015; Heide-Jørgensen 2018). The maturity for females starts at between 8 and 9 years old, and their mean body length is around 4 m (Garde *et al.* 2015; Heide-Jørgensen 2018). Narwhals are arranged in big herds that present small clusters with 5 to 10 individuals (Heide-Jørgensen 2018). Generally, clusters are composed of individuals of the same sex (Marcoux *et al.* 2009).

It is thought that around 12 subpopulations or stocks of narwhals exist in the world, but some parts of the whales' distribution would require more sampling to confirm this assertion (Figure 1.1) (Lowry *et al.* 2017). Narwhal stocks are defined by fjords and bays where narwhals spend the summer (Lowry *et al.* 2017). In Canada and Greenland, there are 17 coastal areas used during summer by narwhals (Figure 1.2) (Chambault *et al.* 2020). Narwhal stocks are separated during summer since, in late spring, when the mating season occurs, there is a matrilineal site fidelity leading each stock to a different summer ground (Heide-Jørgensen 2018). The birth season occurs between June and August, and the gestation lasts about 11 to 15 months (Heide-Jørgensen & Garde 2011).

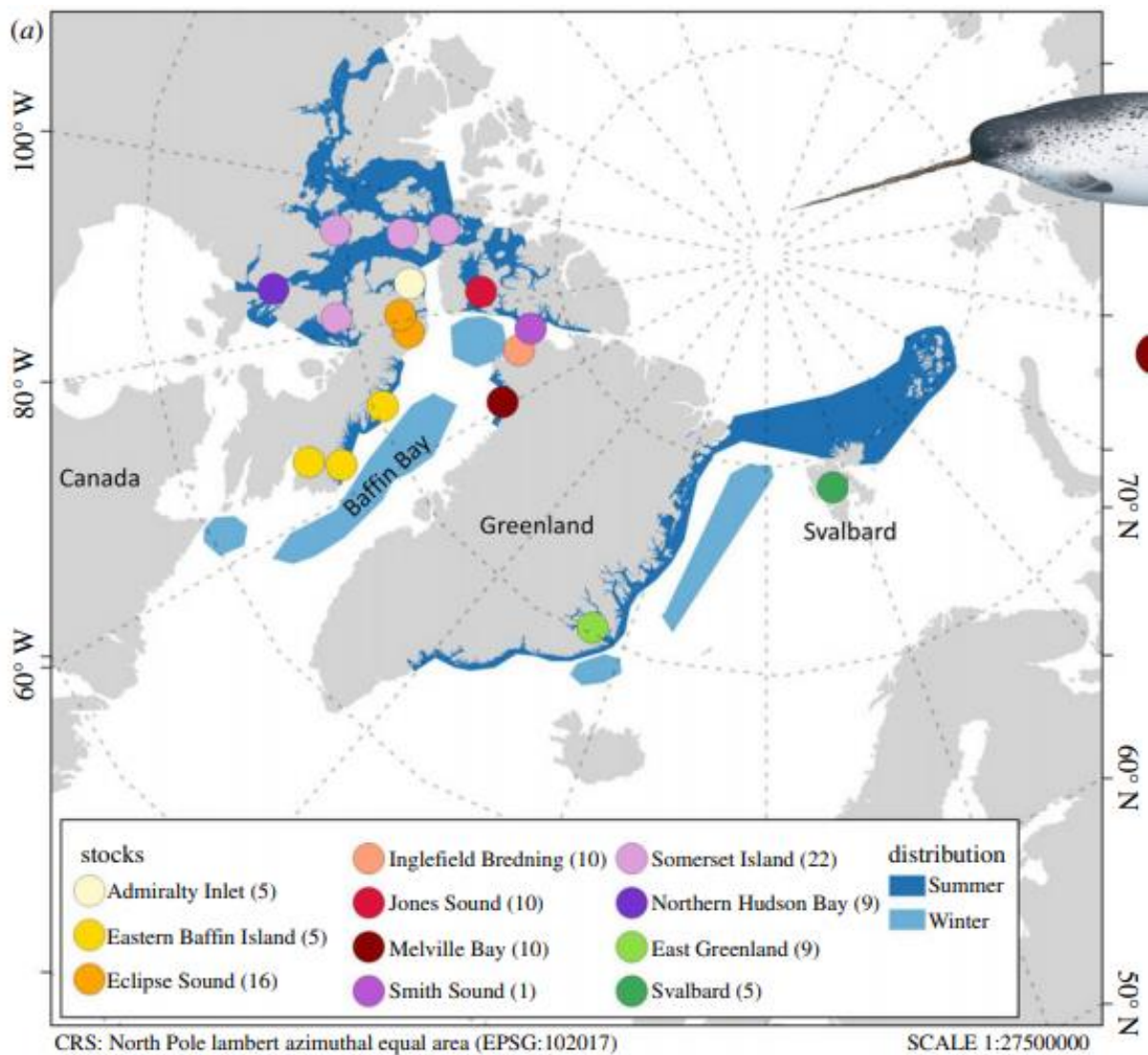


Figure 1.1: Representation of 11 narwhal stocks and distribution of their summer and winter grounds by Louis et al. (2020). The numbers after the name of the stock are the number of samples used in the Louis et al. (2020) study.

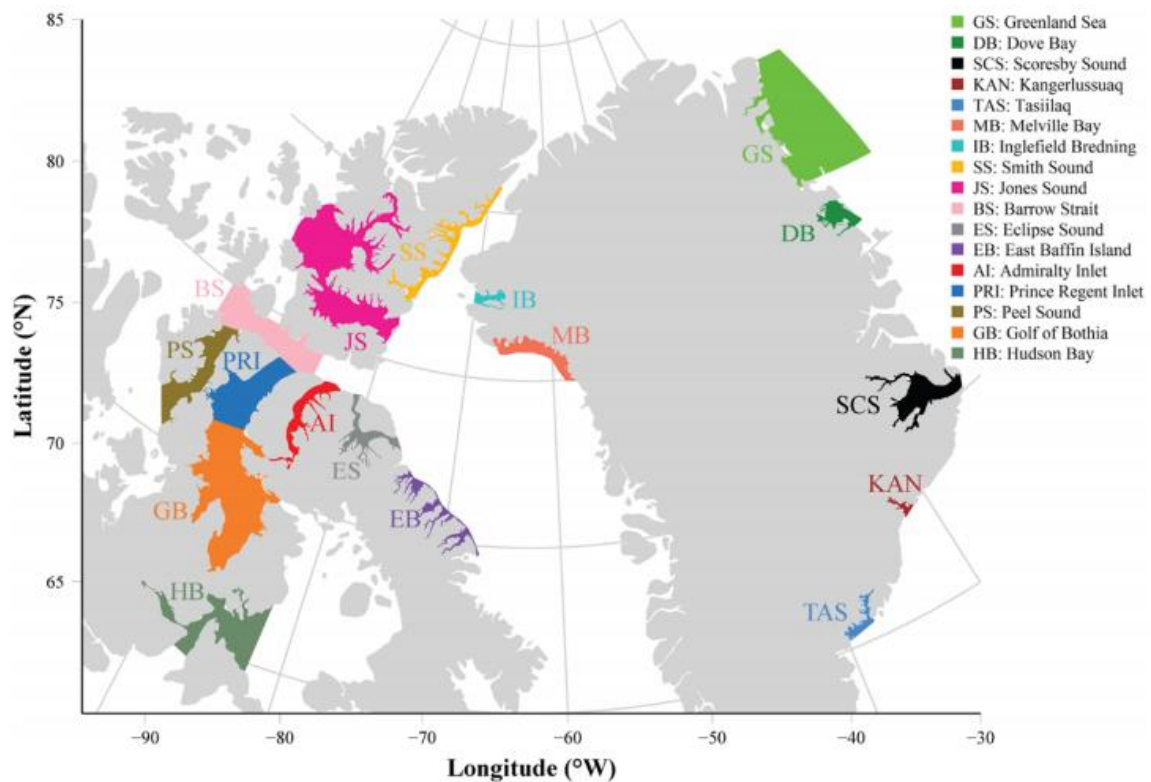


Figure 1.2: Representation of the 17 coastal areas where narwhals spend summer in Canada and Greenland by Chambault *et al.* (2020).

Over time, narwhal subpopulations have slowly decreased in genetic diversity, and if this decline continues, the survival of the species could be threatened (Westbury *et al.* 2019). Narwhals present distinct physiological adaptations that make them have a slow endurance swimming and prolonged diving, in comparison to other odontocetes and pinnipeds (Laidre *et al.* 2008; Williams *et al.* 2011). This means that their ability to adjust to changes in habitat and prey resources are limited by the lack of physiological flexibility (Laidre *et al.* 2008; Williams *et al.* 2011). There are differences in foraging behavior and prey preferences among the different narwhal populations, which may indicate that narwhals have a higher adaptability in their foraging behavior than previously thought (Watt *et al.* 2013; Watt *et al.* 2015).

Before the last glaciation, narwhals were distributed in the North Atlantic however, after retreat of the ice, the distribution of narwhals widened to the Arctic Ocean between Svalbard and Franz Josef Land, to northern Hudson Bay, Baffin Bay, the Eastern Canadian High Arctic Archipelago, Davis Strait, and to the Greenland Sea (Heide-Jørgensen 2018). Narwhals are particularly sensitive to climate change due to their constrained distribution in the Arctic and marked loyalty to their pack-ice habitat, which plays an important role during winter feeding (Laidre *et al.* 2008; Williams *et al.* 2011). Climate change is leading to rises in temperature and ice melting, which will have consequences on narwhal habitats, especially because changes in their endemic habitat are occurring faster than in other earth regions, and will lead to the disappearance of summer ice in the Arctic Sea (Stocker *et al.* 2013; Louis *et al.* 2020). Thus, the availability of some narwhal prey, like cod and benthic species, may decrease, while other prey, like capelin (*Mallotus villosus*), may increase (Watt *et al.* 2013). While biophysical changes will diminish the narwhal's habitat, the lack of ice will give more opportunities for human exploitation in areas that were inaccessible before (Grebmeier 2012; Brown *et al.* 2018; Hauser *et al.* 2018). The

decrease in summer ice coverage has also given killer whales (*Orcinus orca*) access to narwhals (Breed *et al.* 2017).

Furthermore, noise pollution from seismic surveys, offshore drilling and artificial island construction may also increasingly affect their survival (Finley *et al.* 1990; Cosens & Dueck 1993; Heide-Jørgensen *et al.* 2013b; Reeves *et al.* 2014; Simmonds 2017). There is evidence that ship presence and airgun pulses disturb narwhals since they show alterations in their behavior, such as moving closer to the shore and remaining there to avoid the anthropogenic sound (Heide-Jørgensen *et al.* 2021). In addition, narwhals trapped in a closed bay move faster when they are exposed to ships, but when closely exposed to airgun noise they move slower (Heide-Jørgensen *et al.* 2021). These non-normal behaviors occur in the presence of killer whales as well (Laidre *et al.* 2006). It is thought that narwhals remain in silence when they are being disturbed (Podolskiy & Sugiyama 2020). If noise pollution becomes persistent, it could interfere with narwhals' echolocation activity, and the whales may modify their migration patterns and abandon their summering areas (Heide-Jørgensen *et al.* 2013a).

1.3 The acoustics of narwhals and diving behavior

The narwhal is one of the least studied Arctic species regarding behavioral acoustics (Ahonen *et al.* 2019). Narwhal sounds cover a wide breadth of frequencies and include pulsed signals, tonal signals, and whistles, that have communication purposes, and regular echolocation clicks (click trains) and buzzes (lower amplitude bursts of clicks) which serve for prey location and capture, as well as navigation (Blackwell *et al.* 2018; Ahonen *et al.* 2019). Whistles are pure-tone signals that are continuous, present overtones, and occur infrequently (Watkins *et al.* 1971; Ford & Fisher 1978).

Click trains occur when narwhals dive, and sometimes, they can also happen on the surface (Blackwell *et al.* 2018). The high-frequency clicks allow narwhals to determine a more precise position of their prey (Rasmussen *et al.* 2015). High-frequency narwhal clicks can be produced up to 200 kHz, and the lower frequency click peaks observed are between 3.5 kHz and 5 kHz (Stafford *et al.* 2012; Rasmussen *et al.* 2015).

Click trains can transit to buzzes in less than half a second, or rarely, take a few seconds, lowering the amplitudes and turning inter-click intervals from 80-250 ms to between 2 ms and 8 ms (Blackwell *et al.* 2018). Buzzes are used just before prey capture as they provide a more precise position of the intended target (Johnson *et al.* 2004; Miller *et al.* 2004; DeRuiter *et al.* 2009; Wisniewska *et al.* 2014). Reported buzzing rates present a mean between 16-47 buzzes per hour and a maximum of 72-250 buzzes per hour (Blackwell *et al.* 2018). Also, buzzes have a mean duration between 1 s and 3 s, but some longer buzzes lasting 10 s or more can occur (Blackwell *et al.* 2018). Given their lower amplitude, buzz clicks are much less likely to be detected by acoustic sensors and hence typically they are not considered for cue rate estimates (Blackwell *et al.* 2018).

The sonar beam of narwhals is narrow and very directional, which implies a greater intensity of clicks emitted to the forward direction and may reduce echoes from the water surface or pack ice, which is also achieved with the asymmetry of the beam (Koblitz *et al.* 2016). The higher the directionality, the longer the distance at which narwhals can locate their prey (Koblitz *et al.* 2016). A higher directionality is possible to obtain when higher frequencies are produced, and larger narwhals can obtain the same directionality with lower frequencies emitted in comparison to smaller narwhals (Koblitz *et al.* 2016). When narwhals lower the frequency of clicks, they allow their beam to be wider, which is advantageous during foraging activities (Madsen *et al.* 2004; Tervo *et al.* 2021). Similarly, spinning behavior also expands the echolocation beam as a larger area is covered when the body rotates, which helps narwhals forage (Tervo *et al.* 2021).

When narwhals are foraging, they have a higher probability of doing a spinning behavior, and the dive phase where this behavior occurs is always the descent phase (Tervo *et al.* 2021). Another behavior present in this phase is gliding, whose probability of occurrence increases with depth (Tervo *et al.* 2021). In contrast, fast stroking behavior (stroking with caudal fin) is less likely at deeper depths during the descent phase, since this behavior consumes more oxygen, which limits the time submerged (Tervo *et al.* 2021). During the ascent phase it is common to observe a fast stroking behavior with little gliding (Tervo *et al.* 2021). In the final ascent phase, where oxygen reserves are at their minimum, narwhals also behave with a fast stroking behavior, which means they do not need to glide to compensate for the lack of oxygen (Kooyman 2009; Tervo *et al.* 2021). Adult narwhals have a minimum oxygen reserve which allows for 21–24 min of dive duration, and it is unsustainable to maintain a vertical speed of more than 1.5–2.5 m/s because of narwhal’s slow endurance swimming, which narrows its speed range (Williams *et al.* 2011; Tervo *et al.* 2021). Narwhals increase their speed in the descent phase and spend more time gliding when they plan to dive into deeper depths (Tervo *et al.* 2021). Diving activity may be influenced by the physiology, reproductive state, and environmental conditions present at the moment (Ngô *et al.* 2019).

Foraging behavior can be observed on tagged narwhals with surprisingly low sampling rates (25.8 kHz) when tags are attached using suction cups, because of sound transmission from the whales’ bodies through the suction cups to the tag (Blackwell *et al.* 2018). The clicks of the tag bearer have generally higher amplitudes and more low-frequency content than clicks produced by the other (untagged) whales, but from time to time, tagged narwhals may have lower amplitudes, so it is necessary to be careful when listening the records (Johnson *et al.* 2009; Walmsley *et al.* 2020).

1.4 Narwhals from East Greenland

In East Greenland, narwhals are dispersed and declining with an estimated number of individuals of fewer than 1000, listed in Appendix II of Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) (Heide-Jørgensen *et al.* 2010; North Atlantic Marine Mammal Commission (NAMMCO) 2017). Narwhals from East Greenland are isolated from the West Greenland narwhal subpopulations, and it is not known how many narwhal stocks there are in East Greenland (Heide-Jørgensen *et al.* 2015; Ahonen *et al.* 2019). Therefore, the importance of having new density estimates for the management and conservation of the East Greenland subpopulations is understandable (North Atlantic Marine Mammal Commission (NAMMCO) 2019). To do that, PAM and acoustic tags are the most appropriate methods to study narwhals because of the whales’ remote location and active vocal behavior (Watt *et al.* 2015; Ahonen *et al.* 2019).

One of the narwhal stocks from East Greenland spends the summer in fjords and inlets of Scoresby Sound, where the tag data used in this thesis came from (Figure 1.3) (Watt *et al.* 2015). After spending the winter offshore, where they dive deeper, the narwhals from the Scoresby Sound stock migrate to coastal areas, where they spend the summer, show site fidelity, and generally dive to shallower depths (Heide-Jørgensen *et al.* 2001; Heide-Jørgensen *et al.* 2002; Laidre *et al.* 2002; Ahonen *et al.* 2019). In spring and winter, time spent at the surface is greater than in summer, and during these three seasons, identical time is spent in the mid-surface zone, where the pelagic prey of the Scoresby Sound narwhals are found (Watt *et al.* 2013; Watt *et al.* 2015). At depths of less than 50m, where the whales spend longer times and where encounters with other individuals are more likely to happen, there is acoustic communication between individuals (Blackwell *et al.* 2018).

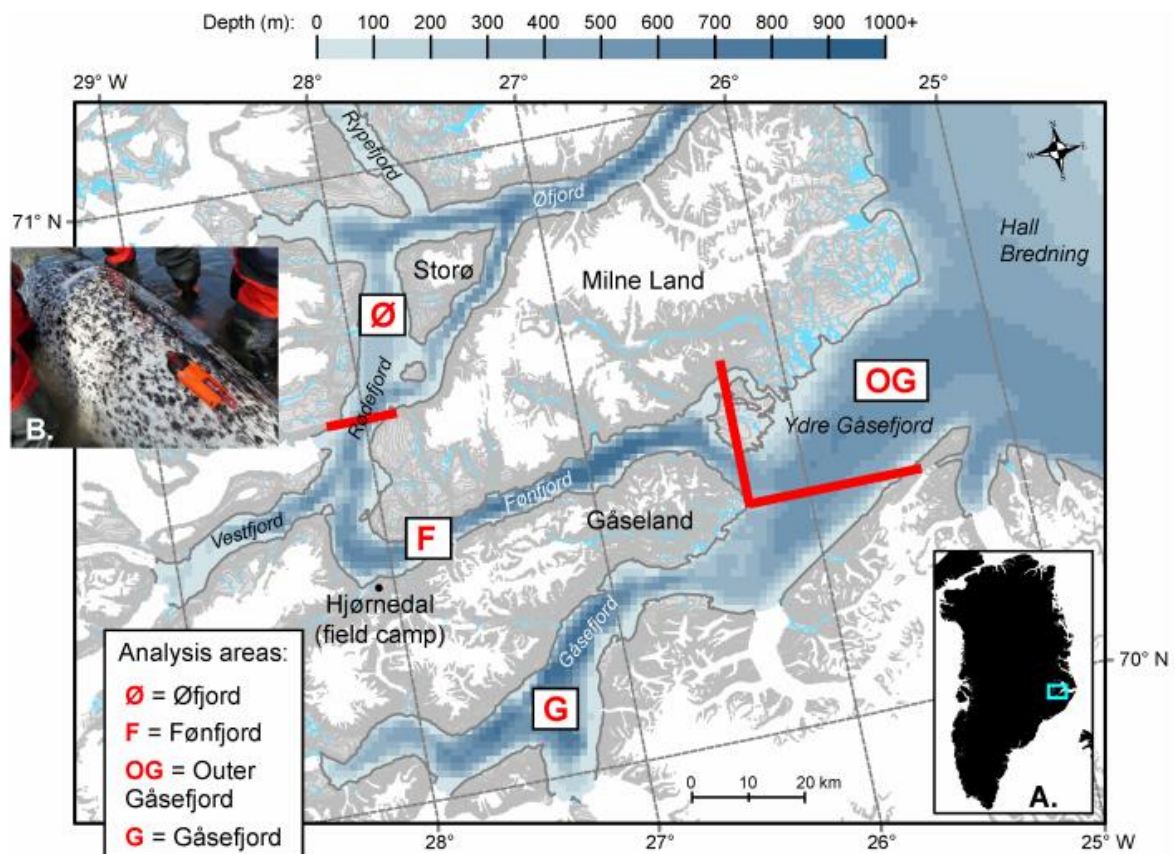


Figure 1.3: Representation of the study area of this thesis. Map of Scoresby Sound fjords and inlets made by Blackwell et al. (2018). The thick red lines delimit the four areas represented in the map (Blackwell et al. 2018).

Narwhals migrate and make more horizontal movements in spring since deep dives rarely occur during that season (Watt *et al.* 2015). During summer and when there is shallow bathymetry, the narwhals dive to the bottom more frequently, and the time spent at the bottom is greater than in winter (Watt *et al.* 2015). Narwhals summer dives have other purposes than foraging, as only 56.4% of summer dives are spent searching for prey, and summer foraging rates are lower than in winter (Laidre *et al.* 2003; Laidre & Heide-Jørgensen 2005; Tervo *et al.* 2021). Dives into the pelagic zone are the most frequent for the Scoresby Sound narwhal subpopulation, where their preferred prey, prey from the mid-surface region, is present (Watt *et al.* 2013; Watt *et al.* 2015). This subpopulation has a foraging preference for capelin (*Mallotus villosus*) (Watt *et al.* 2013). There are different strategies that narwhals use in East Greenland to capture prey, e.g., “transiting foraging” used to capture benthic and solitary prey, and “stationary foraging” used to capture prey present in a school and near the surface (Tervo *et al.* 2021). Narwhals may also forage at the bottom since they spend some time there (Watt *et al.* 2015).

1.5 Thesis aim and hypotheses

This thesis aims to lay the groundwork required to estimate narwhal density from PAM, namely, estimating the rate of echolocation clicks that will be required to turn an estimate of echolocation click density into an estimate of animals density (Marques *et al.* 2013). Hence, the specific objective is to obtain a cue production rate, the average number of echolocation clicks produced per animal per unit of time as well as its precision. To do so, we will use data collected from acoustic tags (Acousonde™—<http://www.acousonde.com/>) on eight narwhals in Scoresby Sound.

The first hypothesis that we tested was whether the probability of producing echolocation sounds is depth dependent. Blackwell *et al.* (2018) showed that the majority of buzzes are made below 250m depth, and so we expected this to be the case. The mean cue rate will be more reliable if we can show that echolocation clicks do not vary much with external and internal factors (Thomas & Marques 2012). Hence, in the second and third hypotheses we tested whether clicking rates, during *clicking periods*, vary over time and depth, respectively. Blackwell *et al.* (2018) showed that females foraged mainly at night, whereas the single male of the study foraged at the end of the afternoon, but it is still unclear if these differences were real, leading to our final objective, to evaluate whether echolocation click rates are influenced by narwhal sex and length, respectively.

2 METHODS

2.1 Study area

Data were recorded in Scoresby Sound, East Greenland (Figure 1.3) (Blackwell *et al.* 2018), an important fjord system during summer for the local narwhal subpopulation (Heide-Jørgensen *et al.* 2020). This system is fed by 12 glaciers strengthened in the north entrance with the cold East Greenland current input, whose way out is in the south entrance (Digby 1953; Heide-Jørgensen *et al.* 2020). The length of Scoresby Sound is about 350 km with shorter ramifications around Milne Land island (Heide-Jørgensen *et al.* 2020). The majority of the fjords' depths reach 1000 m or more, and there are shallow depths in northwestern Jameson Land (Digby 1953; Heide-Jørgensen *et al.* 2020).

The formation of sea ice starts in October and is completed by December (Heide-Jørgensen *et al.* 2020). From December until June the sea ice remains, but during winter at the Scoresby Sound opening, there is polynya (Heide-Jørgensen *et al.* 2020).

2.2 Acoustic data

2.2.1 Tag data collection

The fieldwork was based at the Hjørnedal field camp, at the southwest of Milne Land (Figure 1.3) (Blackwell *et al.* 2018). Orientation and acoustic tags (Acousonde™– <http://www.acousonde.com/>) were attached with suction cups to the back of eight narwhals, two males and six females during the August months of the years 2013 to 2016, and in August 2019 (Table 1.1) (Blackwell *et al.* 2018). Tagging was done near shore by live-capturing the whales with set nets and the help of local Inuit hunters (Blackwell *et al.* 2018). Attached tags were held with magnesium corrodible links guaranteeing their release after three to eight days, while their position was accessible by Argos transmitter (Blackwell *et al.* 2018).

The two channels of the Acousonde that record low and high-frequency sounds (sampling rates 25,811 Hz and 154,868 Hz, respectively) were alternately used in Freya, Thora, Mára, and Eistla (Blackwell *et al.* 2018). For Mutti, Balder, Jonas and Frida only low-frequency sampling was used (Blackwell *et al.* 2018). These choices were made due to data storage and battery considerations, which last longer with lower frequencies, and consequently, will allow longer recordings hence facilitating an understanding of narwhals' behavior over longer time frames (Blackwell *et al.* 2018).

Table 1.1: Characteristics of the individual narwhals considered in this study

Whale	Sex	Length	Year	Samples	Total Record
Thora	Female	341	2014	150	3.83
Eistla	Female	360	2016	150	3.46
Frida	Female	380	2015	149	2.32
Mára	Female	390	2014	10	0.04
Freya	Female	420	2013	200	1.31
Mutti	Female	465	2019	150	3.03
Balder	Male	372	2016	200	4.35
Jonas	Male	510	2019	200	6.10

Whale – name; **Sex** – sex; **Length** – length in centimeters excluding the tusk; **Year** – the year of tagging; **Samples** – number of samples selected (see section 2.2.4); **Total Record** – Number of days of each record. Days where acoustic information was not available or processed and the initial silence after narwhals capture and tagging were discarded from the tag duration.

2.2.2 Description of tag data

As part of other studies (Blackwell *et al.* (2018) and ongoing projects), each record was annotated for *clicking* and *non-clicking periods*. *Clicking periods* were considered continuous when silent breaks lasting no more than 10 seconds were present (Blackwell *et al.* 2018). The key issue is therefore that while we can safely assume that no clicks were produced during *non-clicking periods*, during *clicking periods* the animal might produce multiple clicks per second or no clicks at all in each second. To calculate a click production rate we therefore need to know the mean number of clicks produced during a second belonging to a *clicking period*.

The data set provided by SBB consisted of time-depth profiles at one-second resolution. For each second, information was provided on whether the whale was in a *clicking period* or not at that corresponding depth and time.

2.2.3 Estimation of density via cue counting

Consider a passive acoustic density estimation exercise. Assuming we have a random sample of sensors over a given area one wants to make inferences for the number of sounds counted over a given time can be converted into an animal density for said time and space by the following formula (Marques *et al.* 2013),

$$\hat{D} = \frac{n(1 - \hat{f})}{\hat{p}a\hat{r}}$$

Equation 1.1: General formula to estimate animal density

where \hat{D} is the estimated animal density, n is the number of sounds detected, \hat{f} is the false positive rate (sounds from other sources, not the sound of interest), \hat{p} is the probability of sounds being detected by the sensors, a is the area covered by the sensors and \hat{r} are multipliers (including naturally sound production rate) responsible for turning sound density to an animal density (Marques *et al.* 2013).

In particular, for a cue counting approach, the following formula is used,

$$\hat{D} = \frac{n(1 - \hat{f})}{\hat{p}aT\hat{r}}$$

Equation 1.2: Formula to estimate animal density from a cue counting approach

where, n is the clicks that were detected during T time units, \hat{f} is the false positives proportion, \hat{p} is the probability of a click being detected in the area size where the sounds occurred (a), and \hat{r} is the cue rate (mean number of clicks per whale per time unit) estimated (Marques *et al.* 2013). In this thesis we use seconds as time unit.

After estimating the click rates, the most important missing variable that would be required to calculate the density is the detection probability in a given area around a sensor. Therefore, in this thesis, only the mean cue rate is calculated, as a stepping stone for obtaining density estimates.

2.2.4 Sample selection

It would be too time-consuming to count the number of clicks present in each second for all the data provided because each whale had more than 1500 seconds belonging to *clicking periods*. Nonetheless, we need to be able to estimate the mean number of clicks produced per second during these *clicking periods*, which is a fundamental piece of information to estimate the cue rate. Therefore, a random sample of seconds was selected.

First, the seconds present in the *non-clicking periods* were deleted, leaving only data from *clicking periods*. The first and last seconds of each *clicking period* were also deleted, since these seconds were those during which clicking started or stopped, and were therefore incomplete seconds of clicking information.

To select the samples for analysis, we simply divided the number of available seconds of data from *clicking periods* (for example, 2500) by the number of samples desired (for example, 150). In this case, $2500 / 150 = 16.7$, rounded down to 16, so we selected every 16th second from *clicking periods* for more detailed analyses (see below), until the desired number of samples was obtained. The sample size selected for each record was chosen according to each record's length.

For each sample selected as described above, we counted the number of clicks produced by the whale (using software MT Viewer) and used these counts to estimate how the average number of clicks, their variances, and their coefficients of variations change with different sample sizes (Figure 1.7, 1.8 and 1.9). Given her short record duration, we did not consider Mára in this analysis.

To evaluate the potential variability associated with a given random start we sampled Freya's dataset with a second random sample of 100 seconds, with a different random start. In this way, two different sampling schemes were implemented (Freya1 and Freya2). To evaluate their sensitivity, we investigated if there was a big difference in the number of echolocation clicks, with a high coefficient of variation, between each sampling, and comparing each sampling per se with the two samplings together (with a total of 200 samples). In addition, it was verified if the variance in the average for Freya1 and Freya2 datasets would explain the differences between each mean value obtained for each sampling (Figure 1.10).

2.2.5 Click counting method

Echolocation clicks in each sample selected were counted with the help of MT Viewer, a custom-written program for analysis of Acousonde data (W.C. Burgess, pers. comm.) that displays the data in

the time domain (time in seconds on the x-axis and the sound pressure in Pascal on the y-axis). To allow a clearer view of the clicks and decrease the amount of flow noise, the data collected during low-frequency sampling were high-pass filtered at 1.5 kHz before click counting (Blackwell *et al.* 2018).

Each second evaluated for clicks is referred to as a sample. Samples with buzzes (more than 20 clicks per second) were recorded as zero clicks per second because buzz click amplitudes are much smaller than regular clicks, making them essentially undetected on PAM recorders that are not animal-borne (Blackwell *et al.* 2018). An indicator variable was used to distinguish a true zero (no clicks) from a zero given buzz clicks were present.

New samples, adjacent to those initially selected, were evaluated if there was a noisy background in the original sample and it was impossible to count the clicks. The same happened if the selected sample occurred when Acousonde was switching its sampling between the low and high-frequency hydrophones. We assume these are periods missing at random with respect to click rates and hence do not have an impact on the reported cue production rates. For some surface samples (at depths less than 10 meters), it became difficult to distinguish whether the sounds were made by the tagged whale or by other whales around, because near the surface there is more communication between whales and pressure release effects at the surface lead to poor signal-to-noise ratios (SNR) when tags are outside of the water (Blackwell *et al.* 2018). Therefore, it was taken note of the uncertain samples, and a new sample was not chosen, in this case, since there were only a hand full of uncertain surface samples. Hence, these few uncertain samples will not have a significant impact on the mean number of clicks, and if surface samples would be excluded the results would not show the real behavior of the whale. Only when it was really impossible to know how many clicks there were, the surface sample was classified as unavailable. This happened to one sample from Frida's record, and hence that sample had to be deleted for the analysis. As a result, Frida's sampling continued with 149 samples instead of 150.

The raw data for further analysis was the number of clicks counted per sample in each of the tags.

2.2.6 Estimation of echolocation click rate

With tag data from this thesis we estimate the echolocation click rate (\hat{r}) from equation 1.2 (Marques *et al.* 2013). For tag i , we estimated the corresponding cue production rate (r_i) by,

$$\begin{aligned} \text{Average } r_i &= \frac{r_i \text{ during clicking} \times \text{clicking duration} + r_i \text{ during silence} \times \text{silence duration}}{\text{Total recording time}} \\ &= \frac{r_i \text{ during clicking} \times \text{clicking duration} + 0 \times \text{silence duration}}{\text{Total recording time}} \\ &= \frac{r_i \text{ during clicking} \times \text{clicking duration}}{\text{Total recording time}} \end{aligned}$$

Leading to,

$$r_i = \frac{n_i \times c_i}{T_i}$$

Equation 1.3: Formula to estimate echolocation click rate

where n_i is the number of seconds initially recorded as *clicking periods* in tag i , c_i is the mean number of clicks per second in sampling data i and T_i is the total number of seconds of acoustic recording considered in tag i . Note that c_i is the sum of the c_{ij} , where c_{ij} is the number of clicks in second j sampled from whale i , as represented by the following equation,

$$c_i = \frac{\sum_j(c_{ij})}{J_i}$$

Equation 1.4: Formula to estimate the mean number of clicks per second in tag i

where J_i is the number of samples selected for tag i and c_{ij} is the number of clicks in sample j for tag i . For a better understanding, Figure 1.4 shows a simplified example of how the cue rates can be calculated.

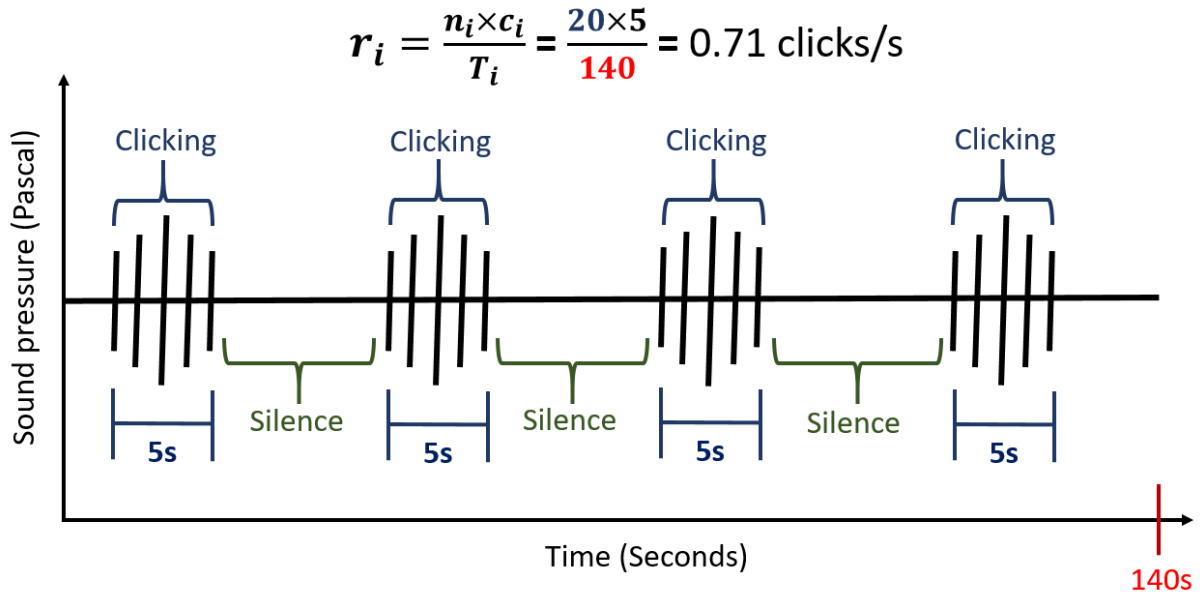


Figure 1.4: Simplified schematic figure that represents how the clicks and silence are present in MT Viewer and how a cue rate can be calculated.

Furthermore, we also estimated the cue rate variance, $var(r_i)$, for each click rate using the delta method (Powell 2007) and taking into account the variability from the sampling process, meaning that a finite sampling correction was incorporated, as represented by the following formula,

$$var(r_i) = \left(\frac{n_i}{T_i}\right)^2 \times \frac{1}{(J_i)^2} \times var(c_i) \times \frac{T_i - J_i}{T_i - 1}$$

Equation 1.5: Formula to estimate variance of echolocation click rate

where, n_i is the number of seconds recorded as clicking in tag i , c_i is the mean number of clicks per second in sampling data i , T_i is the total number of seconds in tag i , and J_i is the sample size used from tag i . For an easier interpretation, it was estimated the cue rate coefficient of variation in percentage, $cv(r_i)$, by (Abdi 2010),

$$cv(r_i) = \frac{\sqrt{var(ri)}}{ri} \times 100$$

Equation 1.6: Formula to estimate coefficient of variation of echolocation click rate in percentage

where, $\sqrt{var(ri)}$ is the standard deviation of ri .

Generally, the variability in a population cue rate depends only on the variability across tags. In this case, the mean cue rate depends on the variability across tags and induced from the sampling procedure.

It was estimated a simple average cue rate (unweighted average) for the subpopulation by,

$$r(\text{unweighted}) = \frac{\sum(r_i)}{N}$$

Equation 1.7: Formula to estimate a simple average cue rate (unweighted average)

where, N is the total number of tags which, in this case, were 8 tags.

We could consider that tags with longer durations would lead to a cue rate with higher accuracy (Marques *et al.* 2009). Therefore, another estimate was produced for the subpopulation, considering a weighted average, where the weights are the different tags durations, as represented by the following equation (Marques *et al.* 2009),

$$r(\text{weighted}) = \frac{\sum(n_i \times c_i)}{\sum(T_i)}$$

Equation 1.8: Formula to estimate a weighted average cue rate, where the weights are the different tags durations

where $\sum(n_i \times c_i)$ is the total number of clicks counted in all tags, and $\sum(T_i)$ is the total number of seconds present in all tags.

A table with the values obtained for each sampling data was created (Table 1.2). Since Freya's dataset had two different samplings, we considered the combined sampling scheme for further analysis (Freya1+2).

To estimate the variance of the unweighted and weighted average cue rates, a bootstrap resampling approach was considered. The values present in table 1.2 were resampled 999 times by bootstrapping the tags and the variance and standard deviation were computed. Then, the coefficient of variation was calculated by equation 1.6.

2.3 Statistical analysis

All data analysis were implemented in R (R Core Team 2020).

2.3.1 Analysis of presence and absence of sounds production

All tag data provided were modeled to analyze the probability of foraging sounds being produced as a function of depth. The *non-clicking periods* represented the absence of foraging sounds, and the *clicking periods* represented the presence of foraging sounds.

Narwhals were typically initially silent because of their capture and tagging. To avoid artificially lowering presumed natural cue rates given this effect of tagging, we removed those silent periods (Blackwell *et al.* 2018). To do so, plots for each narwhal were created to select the seconds where the foraging behavior started without those unusual gaps of silence. We considered that normal acoustic behavior was initiated from the first dive, where sounds were being produced regularly over time. To distinguish between the absence and presence of sounds over time and depth, each point was colored differently depending on clicking state (Figure 1.5).

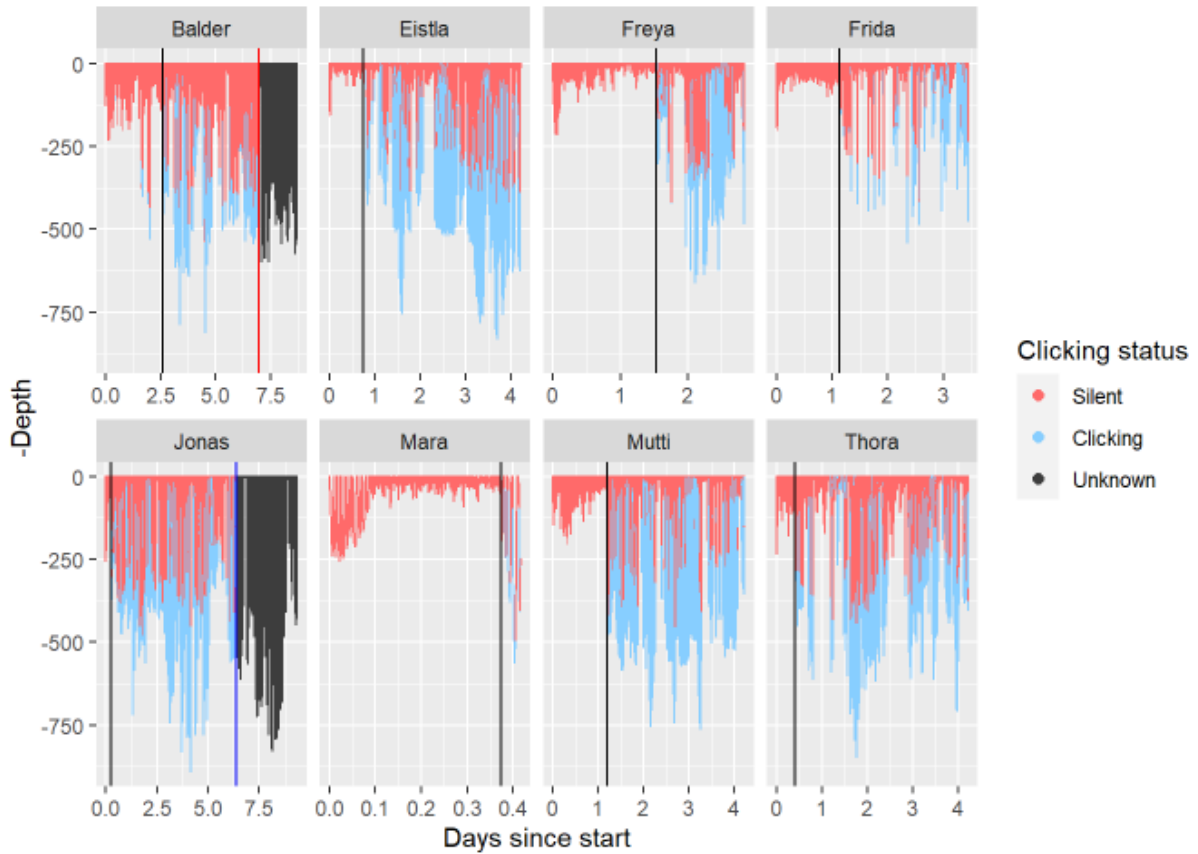


Figure 1.5: Presence (in color blue), absence (in color pink) and unknown (in color black) of echolocation behavior by each narwhal as a function of depth and days since capture and tagging. Black lines represent the days from which narwhals start their foraging behavior. After the red line, in Balder’s dataset, acoustic information was not available, and in the second part of Jonas’ dataset (after the dark blue line) sound was available but not processed.

After knowing the time at which the foraging behavior started for each whale, the samples between long periods of silence were deleted (Figure 1.6). The firsts 23 s of sounds produced in Balder’s record after about 12 h of silence were ignored and removed since Balder then remained silent for nearly 24 h more with unusually short clicking times framed by big gaps of silence (Blackwell *et al.* 2018). In addition, the samples after about seven days until the end of Balder's record had to be removed since his acoustic information was no longer available because the acoustic channel stopped recording when the Acousonde's battery got depleted (after the red line in Figure 1.5). For the second part of Jonas’ dataset (after the dark blue line in Figure 1.5), which periods correspond to *clicking* and *non-clicking periods* were not available, and therefore this period was ignored for sampling (Figure 1.6).

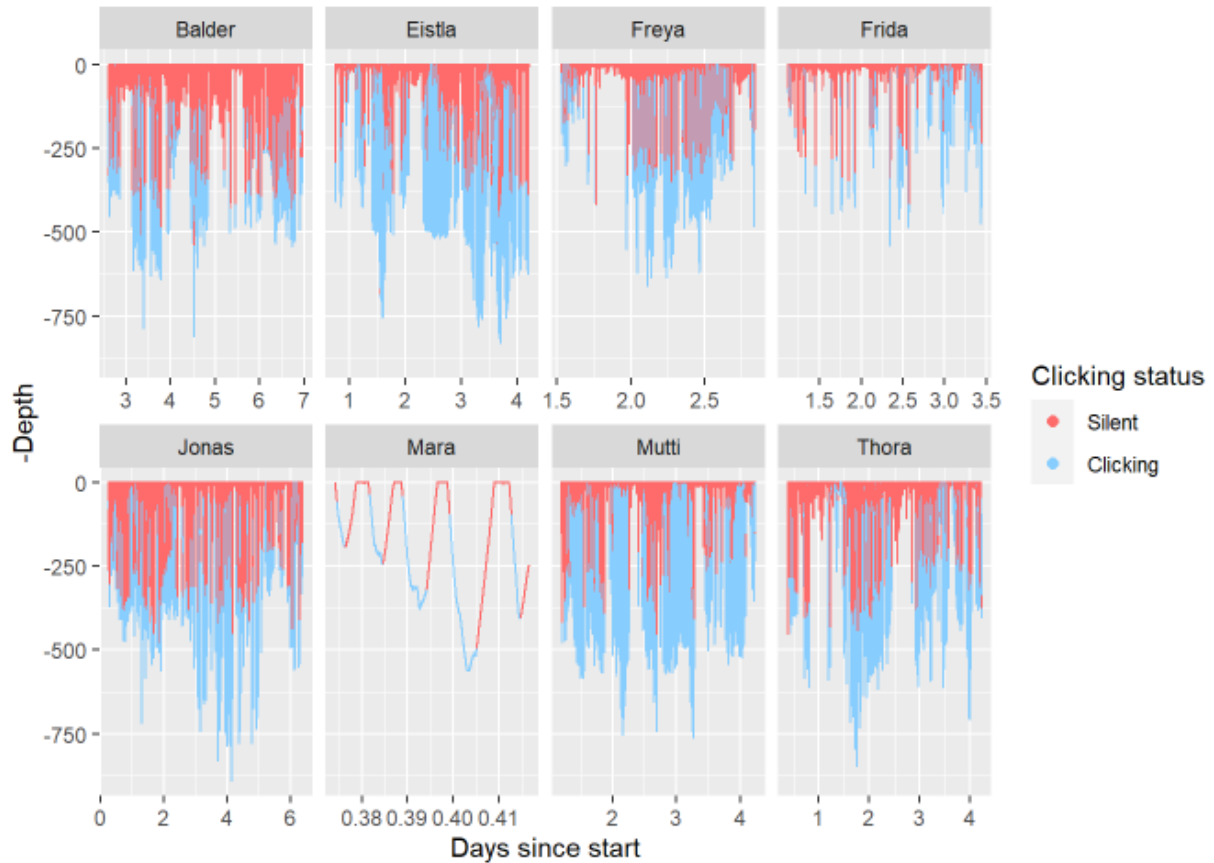


Figure 1.6: Presence (in color blue) and absence (in color pink) of echolocation behavior by each narwhal as function of depth and days since capture and tagging. These data were corrected by deleting long periods of silence as well as periods in which acoustic information was not available or processed.

We then modelled the probability of a given second being in a *clicking period* (response variable) as a function of depth (explanatory variable), using a binomial generalized additive model (GAM). Narwhal ID was considered a fixed factor in the model. In this way, the probability of foraging sounds being produced was estimated as a function of depth, per whale, by making predictions from the GAM (Figure 1.11). For computational efficiency, the data were subsampled considering a systematic sampling with random start, selecting only one out of every 10 samples in the original data.

2.3.2 Trends of the number of clicks over time and depth

To verify if there was a trend in the number of clicks over time and depth during the *clicking periods*, sampling data were modeled. The selection of the best model was made through observation and interpretation of the residuals, and the Akaike information criterion (AIC) values. The number of clicks counted in the samples selected was used as the response variable, and the explanatory variable was the time differences in minutes to see the trend over time, and the depth in meters to see the trend over depth. It was considered time since the first sample selected started by making the differences in minutes between the first selected sample and each other samples selected. The selection of the best model was made through interpretation of the AIC values of generalized linear models (GLMs) and GAMs of the negative binomial and Poisson families. As a result, the model selected to observe each trend, over time and depth, was a GLM of the negative binomial family with all the sampling data together and with narwhal ID as a fixed factor (Figure 1.12 and 1.13).

2.3.3 Test for the influence of narwhals sex and length in echolocation click rates

With the cue rates estimates obtained for each whale, we tested whether there was an influence of narwhals' sex and length on echolocation clicking rates. A boxplot of the click rates for females and males was done for data visualization (Figure 1.15). Click rate estimates of the eight whales were used as the response variable, and the length in centimeters and sex for each whale were used as the explanatory variables. The selection of the best model was made through observation and interpretation of the residuals, and the AIC values of GLMs and GAMs of the Gaussian and Gamma families. Models with the Gamma and Gaussian families were done because our outcome (click rates) is a non-negative continuous variable. As a result, a GLM of the Gaussian family with a log link was implemented since it had a lower AIC value (Figure 1.14).

2.3.4 Estimation of the cue rate in the second part of Jonas' dataset

To illustrate how one might be able to estimate acoustic cue rates based on tags without acoustic data we estimated the click rate for the last part of Jonas' tag, since after about six days until the end of Jonas' record the data from the acoustic tag was not originally processed. We used the above GAM model to predict the number of clicks Jonas would produce in each second as a function of depth. To predict the sound production without acoustics, we multiplied the probability of foraging sounds being produced as a function of depth by the mean number of sounds produced as a function of depth. Summing all these values and dividing by the length of recording without acoustics, we obtained a model-based estimated cue rate.

3 RESULTS

3.1 Effects of tagging

During the analysis of presence and absence of sounds produced, it was verified that narwhals were stressed after their capture and tagging because they were silent for a long period after their release, which is not a natural behavior (Figure 1.5). It was also observed that each individual had a different reaction to the capture and tagging since the time it took them to recover was different (Figure 1.5).

3.2 Sample size effect on cue rate estimates

During the sample selection analysis, it was concluded that 100 samples were enough to obtain a good estimate of cue rates in Freya's record (Figure 1.7 and 1.8). Moreover, there was not a big difference in the mean values with 100 or 200 samples selected (Figure 1.9). In addition, it was verified that the mean number of clicks per second obtained for the Freya1 dataset was within the variance of the mean obtained for the Freya2 dataset and vice-versa (Figure 1.10). The mean values obtained for Freya1, Freya2, and Freya1+2 datasets were 3.96, 4.92, and 4.44, respectively.

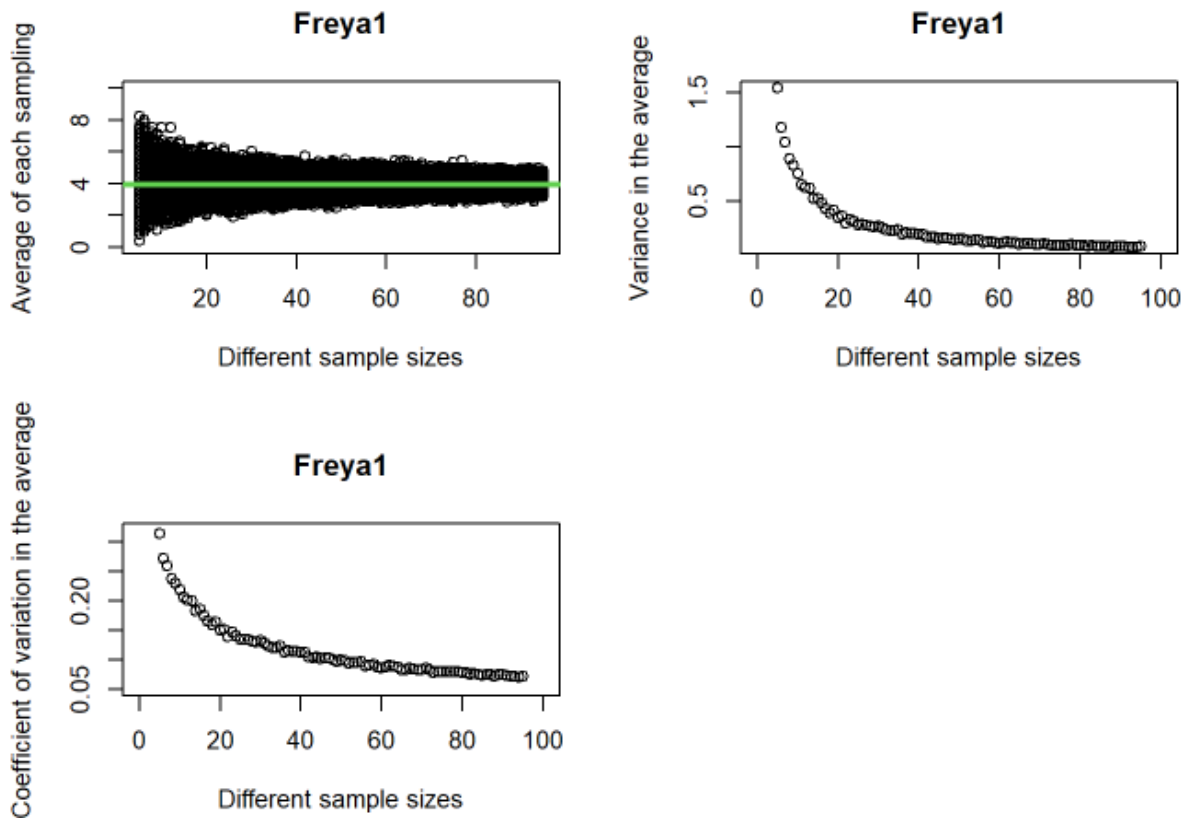


Figure 1.7: Plots showing how the increasing sample sizes change the estimate of the mean obtained and how the variances of the mean and their coefficients of variation change with increasing sample sizes for Freya1 dataset (first sampling made from Freya's record). The green line represents the mean obtained for Freya1 dataset.

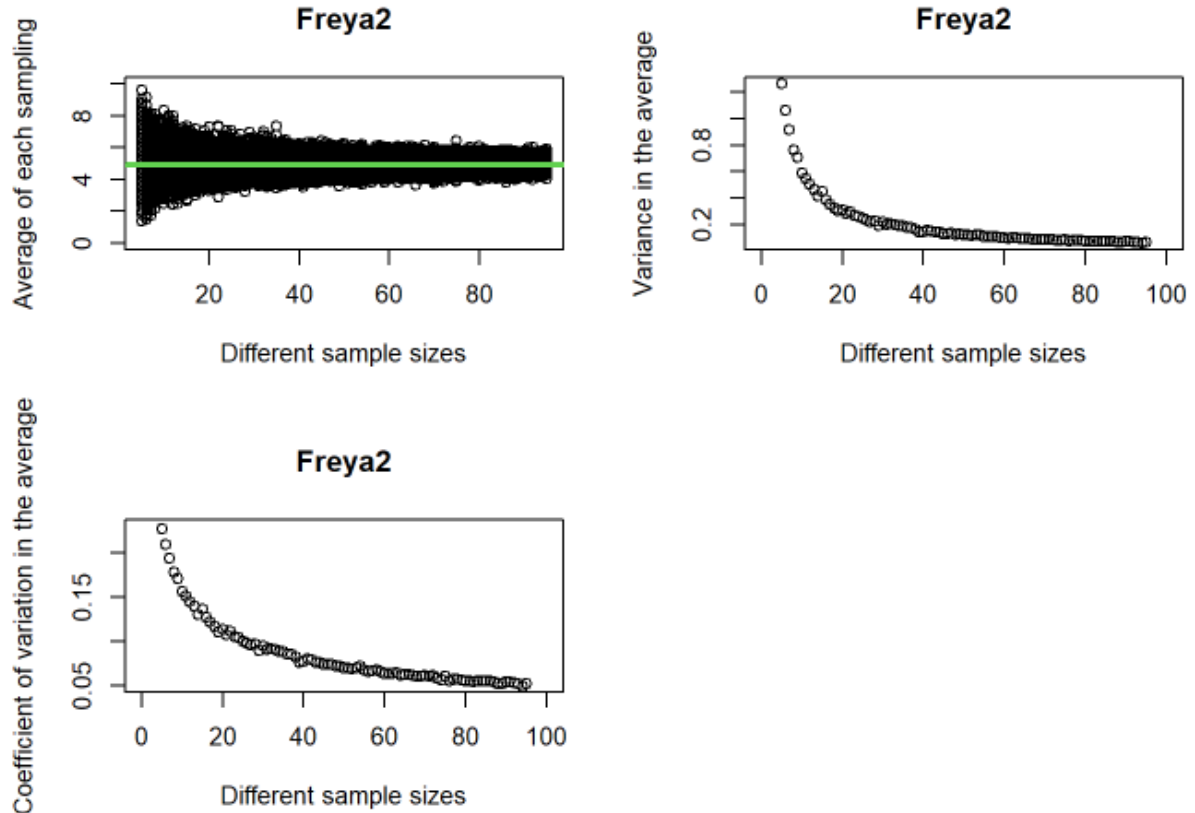


Figure 1.8: Plots showing how the increasing sample sizes change the estimate of the mean obtained and how the variances of the mean and their coefficients of variation change with increasing sample sizes for the Freya2 dataset (second sampling made from Freya's record). The green line represents the mean obtained for Freya2 dataset.

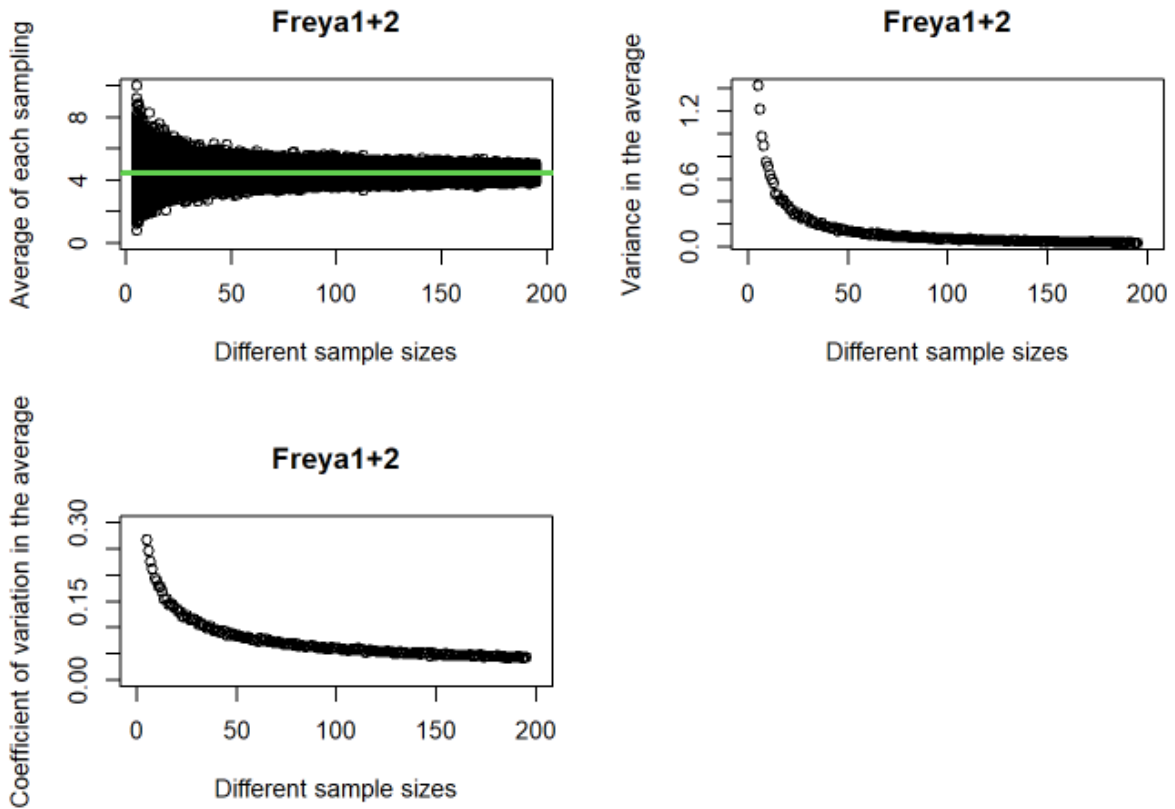


Figure 1.9: Plots showing how the increasing sample sizes change the estimate of the mean obtained and how the variances of the mean and their coefficients of variation change with increasing sample sizes for Freya1+2 dataset (the two samplings made from Freya's record were pooled). The green line represents the mean obtained for Freya1+2 dataset.

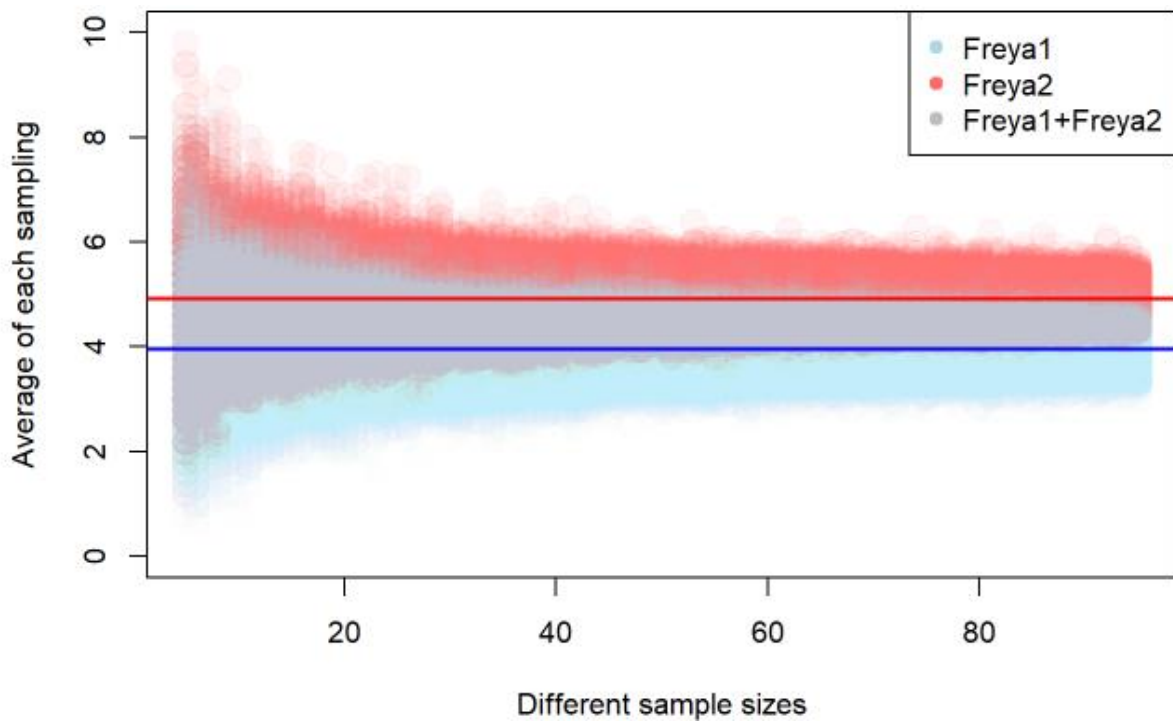


Figure 1.10: Plot showing how the increasing sample sizes change the estimate of the average obtained by a sub-sample of k samples out of 100 that were selected for the Freya1 and Freya2 datasets, the first and second samplings made from Freya's record, respectively. The blue line represents the mean obtained for Freya1 dataset, and the red line represents the mean obtained for Freya2 dataset. The blue and red points are the average values obtained for the Freya1 and Freya2 sub-samples, respectively. The grey points are the average values for the Freya1 and Freya2 sub-samples that were overlapped.

3.3 Relationship between depth and probability of foraging sounds produced

The presence vs absence analysis showed that the probability of clicking and buzzing occurrence increased significantly with increasing depth for each narwhal (Figure 1.11). The whales were buzzing and clicking regularly at depths deeper than 400m (Figure 1.11). The biggest difference in the probability of clicking as a function of depth was between Frida and Balder (Figure 1.11). Mára's dataset had only 1h of recording, so the reliability of her prediction is likely poor. The wiggleness of the estimated function probably reflects that uncertainty (Figure 1.11).

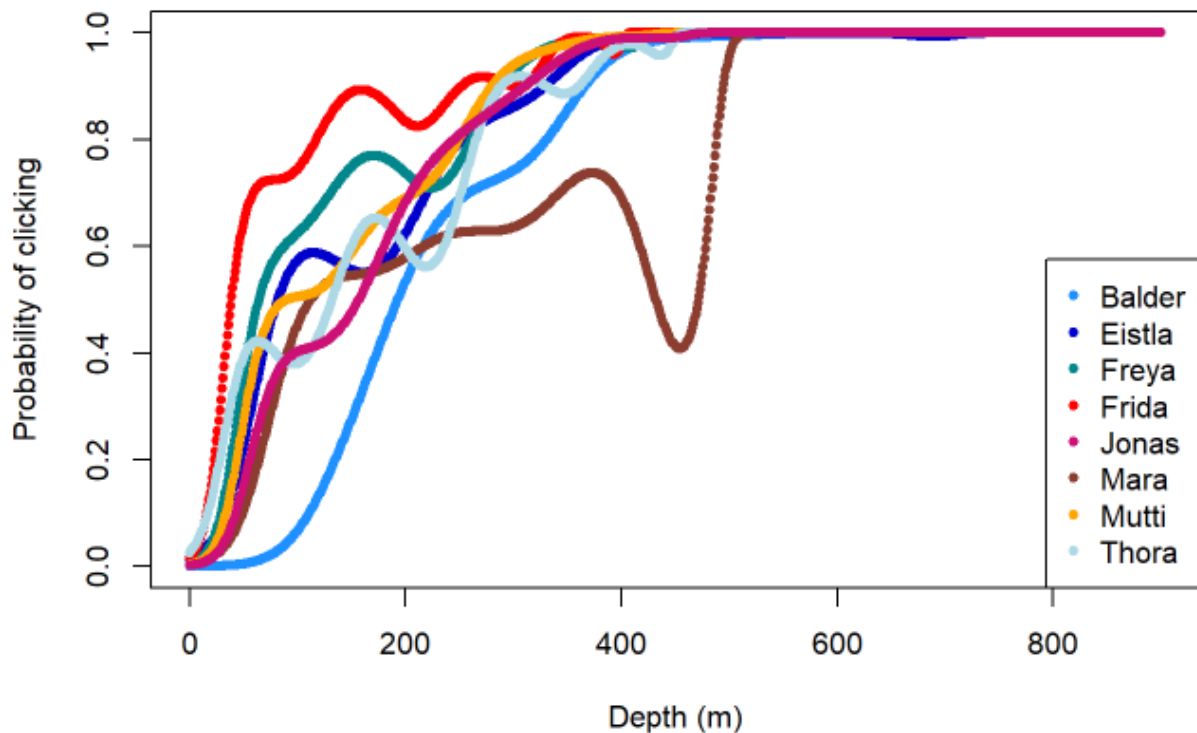


Figure 1.11: Plot showing the probability of echolocation clicks produced by narwhals (“Prob.Clicking”) with whale variable as a factor along with the depth (in meters). Each narwhal is represented by a different color.

3.4 Trends of the number of clicks over time and depth

It was verified that there was not a trend in the number of echolocation clicks per second present in the *clicking period* over time (p -value = 0.12332), but there was a significant tendency with deeper depths during this period (p -value = $6.06e-06$), where the number of clicks increased with depth for the Scoresby Sound's narwhals stock (Figure 1.12 and 1.13).

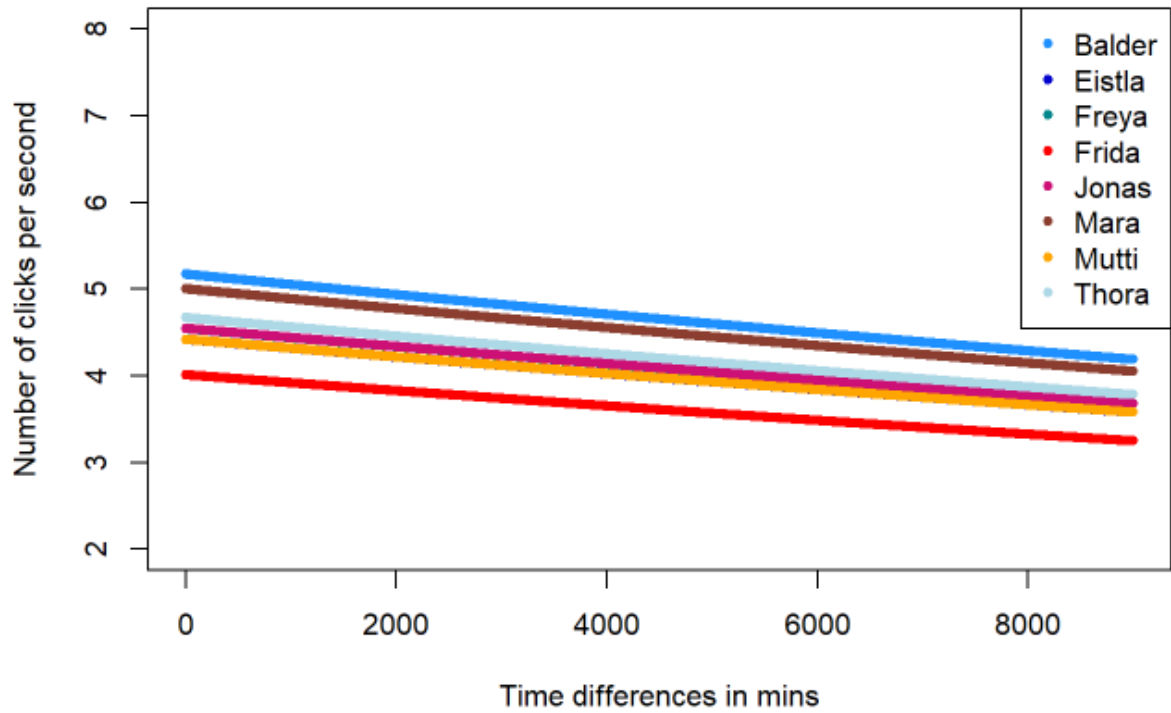


Figure 1.12: Plot showing the trend of the number of clicks during the *clicking period* over time. The time variable is the time differences in minutes between the first sample selected with the other samples selected for each sampling data. Each narwhal is represented by a different color.

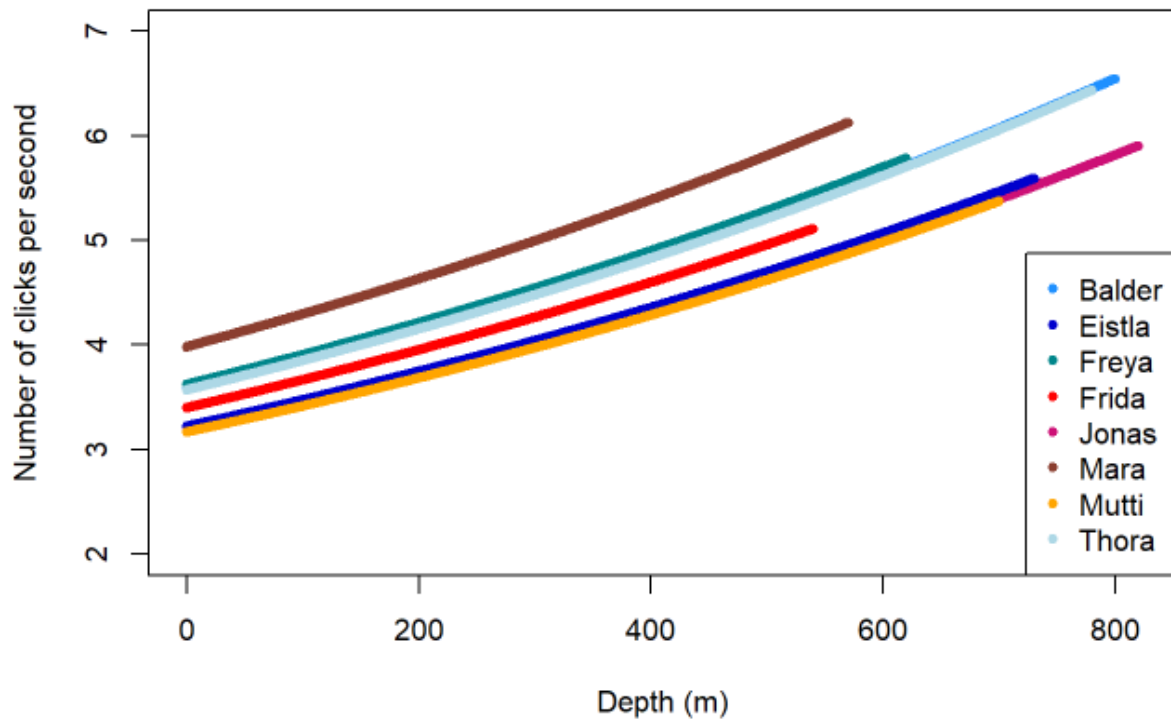


Figure 1.13: Plot showing the trend of the number of clicks during the *clicking period* with deeper depths in meters. Each narwhal is represented by a different color.

3.5 Click rate estimates

The values obtained for each sampling data are shown in Table 1.2, where click rate estimates are present in decreasing order. It can be observed that the click rate coefficients of variation values obtained for each individual were always less than 10% (Table 1.2).

The unweighted and weighted mean click rate obtained from all the samplings was 1.40 clicks/s with a coefficient of variation equal to 11.41% and 1.27 clicks/s with a coefficient of variation equal to 10.20%, respectively.

Table 1.2: Table with respective values for each sampling data

Whale	Silence	Clicking	Mean	cr	crvar	crev
Mára	0.02	0.02	5.00	2.12	2.52E-02	7.47
Mutti	1.60	1.43	4.20	1.98	8.79E-05	0.47
Eistla	2.19	1.27	4.15	1.53	5.51E-05	0.49
Freya	0.88	0.43	4.44	1.46	1.91E-05	0.30
Thora	2.77	1.06	4.33	1.20	5.02E-05	0.59
Jonas	4.37	1.73	4.01	1.16	1.64E-05	0.35
Balder	3.39	0.96	4.70	1.04	2.16E-05	0.45
Frida	1.88	0.44	3.81	0.72	1.75E-05	0.58

Whale – name; **Silence** – number of days spent in silence; **Clicking** - number of days spent clicking; **Mean** - mean number of echolocation clicks of each sampling data; **cr** - click rate estimates; **crvar** – variances of each cr; **crev** - coefficients of variations of each cr in percentage

3.6 Test for the influence of narwhals sex and length in echolocation click rates

There was no evidence to suggest that significant differences in mean click rates between males and females (p -value = 0.2107) might be present, nor that narwhals' length influenced echolocation click rates (p -value = 0.3133) (Figure 1.14). Our low sample size, with only two males, makes it more difficult to obtain significant differences. Therefore, these results should be interpreted carefully (Figure 1.15).

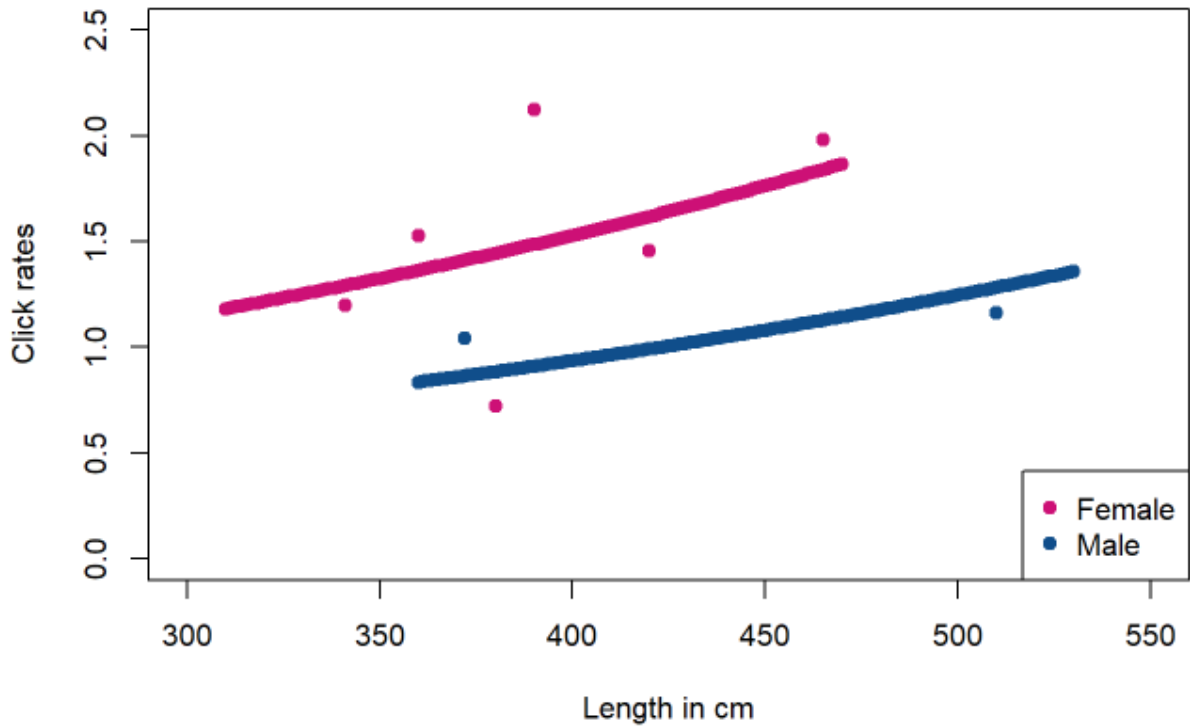


Figure 1.14: Plot showing the relationship between the click rates and narwhals' length, in centimeters, for each sex. Color pink and blue represent the females and males, respectively.

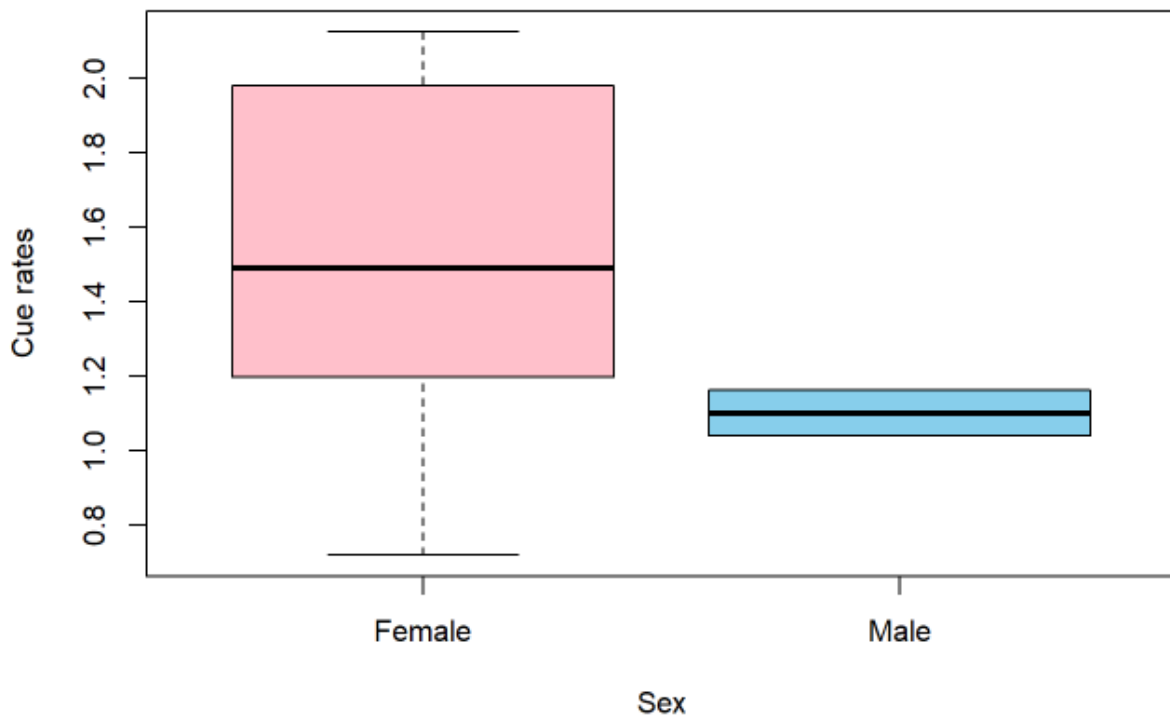


Figure 1.15: Boxplot of the click rates for females and males. The black line (highlighted) represents the click rate medians obtain for females (1.49 clicks/s) and males (1.10 clicks/s). The first quartile is the line after the minimum click rate (present at the lower extreme) and the third quartile is the line before the maximum click rate (present at the upper extreme). In the males' boxplot, the minimum and maximum have the same values as the first and third quartile, respectively, because our data have only two males.

3.7 Click rate for the last part of Jonas' tag

The click rate obtained for the last part of Jonas' tag, without acoustics processed, was 1.43 clicks/s, which was slightly higher than the click rate obtained for the first part (1.16 clicks/s). During the last part, Jonas dove more frequently to deeper depths, with a mean depth greater (138.36 m) than the mean depth obtained for the first part of Jonas' tag (110.79 m).

4 DISCUSSION

We performed a sample selection for each individual record, counted the echolocation clicks in each sample selected, and estimated the mean cue rate of eight narwhals in Scoresby Sound. We evaluated the relationship between depth and the probability of regular clicks being produced, the trend between the number of clicks with depth and over time, and tried to assess the influence of sex and length in echolocation click rates.

In this thesis, it was shown that narwhals from Scoresby Sound produce regular clicks, used during foraging, with a higher probability of occurrence at depths deeper than 400m. We also showed that, during their foraging activity, they echolocate regularly without predictable changes over time, but there was a significant tendency to produce more clicks per second at deeper depths. With these models, we estimated a model-based cue rate for one of our males, where only time-depth profiles were available without the acoustics processed. It was not possible to show whether narwhals click rates were influenced by sex and length. The small sample size would provide very low power to detect any real effects that might exist. Click rate estimates obtained had low coefficient of variations, and therefore, we anticipate these might provide the basis to establish accurate density estimates with PAM in the future (Gomez & Gomez 1984). If that is the case, then we will have achieved our goal which was to provide a stepping stone to promote a way to assess narwhal populations, and hence to help monitor and manage their population.

We did not find differences in cue rates between males and females, but this might be essentially lack of power given a small sample size. Further studies, with more individuals, are needed to confirm, or to disprove these results. It was thought that larger animals might have a higher cue rate, but this was not shown probably for the same reasons. In the future, it might be important to see whether bottom depth or the available prey for narwhals in Scoresby Sound might be responsible for specific cue rates (Tervo *et al.* 2021). If that is the case, then cue rates would be harder to transfer across sites with different prey distributions.

4.1 Effects of tagging

As described by Blackwell *et al.* (2018), capture and tagging caused a strong reaction by narwhals, leading some individuals to be silent for almost two days (Figure 1.5). The mean length of the post-release silence was 23h (Blackwell *et al.* 2018). This initial long period of silence is not considered a normal behavior, but happens to be the same reaction narwhals show in response to noise pollution and their predator, the killer whale (Finley *et al.* 1990; Blackwell *et al.* 2018).

In addition, the length of the post-release silence differed between whales (Figure 1.5). Blackwell *et al.* (2018) reported no correlation between the length of the post-release silence and the duration of capture and tagging, which ranged between 30 and 90 minutes. Hence, Blackwell *et al.* (2018) suggest that the first location where whales went after release, in Føn fjord, may not be an area with narwhals'

preferred prey since buzzing rates for the majority of narwhals were lower in this area. Shuert *et al.* (2021) showed that the handling duration had an impact on narwhals' swimming behavior and activity, which were intensified when more than 40 minutes of handling were required. They reported that narwhals did shallower dives and spent most of their time at shallow depths after their release, but more investigation about the effects of narwhals capturing and tagging is still required (Shuert *et al.* 2021). In addition, the mitigation of these effects should also be studied (Hawkins 2004; Casper 2009). We could have much more data available for this thesis if the responses of narwhals to their tagging and capture were mitigated.

Blackwell *et al.* (2018) had to use long-term deployments on narwhals to satisfy the long-term goals of the project, i.e., to study the effects of airgun pulses from seismic exploration on narwhal behavior (Heide-Jørgensen *et al.* 2021). Given the impossibility of tagging narwhals at sea because of their shyness and skittish nature, capture of the animals was required (Blackwell *et al.* 2018).

4.2 Sample size effect on cue rate estimates

To obtain a reliable cue rate estimate, 100 samples selected from Freya's record were apparently a good compromise since the coefficient of variation of the average number of clicks was less than 20% (Figure 1.7 and 1.8) (Gomez & Gomez 1984). Furthermore, selecting 100 or 200 samples from that dataset resulted in a similar mean number of clicks with very little variation (Figure 1.9). It was observed that the differences in the mean number of clicks between the two samplings done for Freya's dataset were consistent with the variance in the mean values obtained, as well (Figure 1.10). In other words, the mean values with 100 or 200 samples selected were similar.

These initial results provided us with the knowledge of how many samples would be necessary for each of the other whales, where the record duration and sample size selected for Freya's dataset were used as a benchmark.

4.3 Relationship between depth and probability of foraging sounds produced

The probability of making clicks and buzzes was strongly dependent on depth values in Scoresby Sound (Figure 1.11). Narwhals had the highest probability of producing foraging sounds at depths deeper than 400m, which supports one of the hypotheses of this study (Figure 1.11). This result is explained by the fact that at depths shallower than 100m narwhals socialize and hence regular clicks and buzzes used during foraging activity are uncommon to occur (Blackwell *et al.* 2018). On the contrary, where foraging activity mostly occurs, communication between individuals is rare (Blackwell *et al.* 2018). Our results are in accordance with Blackwell *et al.* (2018), which showed that buzzes are mainly made between 350 m and 650 m depth and with other studies that found that narwhals in East Greenland dive most frequently to the mid-surface region, which presents their preferred prey, capelin (*Mallotus villosus*) (Watt *et al.* 2013; Watt *et al.* 2015; Blackwell *et al.* 2018). Blackwell *et al.* (2018) showed that higher buzzing rates occurred in one of Scoresby Sound's preferred summer areas, the Gåsefjord complex. In addition, Heide-Jørgensen *et al.* (2020) showed that during summer and at depths in the range 300-650m, there is a niche selection by narwhals because a temperature lower than 2°C is present. Narwhals may prefer this niche since they would prey more easily at that temperature because of their slow endurance swimming (Williams *et al.* 2011; Heide-Jørgensen *et al.* 2020).

Also, a big difference between Frida and Balder's foraging activity was shown here, which Blackwell *et al.* (2018) also described. Blackwell *et al.* (2018) showed that 55% of buzzing was made

by Frida at depths of less than 150 m, closer to the surface and for a longer time than Balder who had only 1% of his buzzing activity in that depth range. These results may suggest that males and females have different acoustic behaviors, but a greater number of individuals is needed to be certain about these differences. The niche selection and the strong site fidelity that narwhals present show their sensitivity to changes in their habitat, so understanding how they will respond and cope with disturbances in their optimal habitat is crucial (Laidre *et al.* 2008; Heide-Jørgensen *et al.* 2020).

4.4 Trends of the number of clicks over time and depth

The number of echolocation clicks per second present in the *clicking period* did not show a trend over time, but with deeper depths, during this period, narwhals from Scoresby Sound produced more clicks per second (Figure 1.12 and 1.13). Therefore, when narwhals are making echolocation clicks, mainly at depths deeper than 400m, they echolocate regularly regardless of the time present in the moment but tend to echolocate more frequently with increasing depths.

Blackwell *et al.* (2018) analyzed six tags that were also used in this thesis and showed that narwhals in Scoresby Sound presented a short duty cycle of echolocation, where 27% of the time narwhals were clicking, and had a mean clicking bout duration in the range 5.5-10 min. Although this low percentage of time spent clicking makes echolocation clicks difficult to detect on PAM recorders, our results showed that during echolocation narwhals did not show variability in their echolocation activity over time (Blackwell *et al.* 2018). Thus, constant echolocation clicks over time are suitable for density estimates by PAM.

With our models, we were able to obtain a click rate for the second part of Jonas' tag, without having his *clicking periods* processed for the sound data, by combining the probability of foraging sounds as a function of depth with the number of clicks per second, while clicking, as a function of depth since with increasing depths, during clicking time, echolocation is more frequent. This is interesting as it illustrates how one might use historical time-depth data to estimate cue production rates from tags without the acoustic sensor. This may work particularly well in deep-diving species such as narwhals, for which depth is presumably a good predictor of clicking state.

4.5 Click rate estimates

There was not great dispersion around the click rate values for each whale since their coefficients of variation were less than 10%, and therefore click rates were precise (Table 1.2) (Gomez & Gomez 1984). The unweighted and weighted mean click rates obtained from all the samplings had coefficients of variation below 12%, suggesting that this amount of sampling is reliable when used in the future for density estimates with PAM for the Scoresby Sound subpopulation of narwhals (Gomez & Gomez 1984). The estimates made by bootstrap ignore the variance between each tag, but as we saw in our results, the variability of this variance is small. The unweighted and weighted mean click rate values obtained were identical, but the weighted mean click rate may be the best to use for the density estimate with PAM since the variability from the different record durations is taken into account.

However, it should be noted that a far preferable approach that would remove the need to implement the sampling procedure would be to have these tags automatically processed for clicks. That would remove the sampling requirement and would lead to a cue rate without the sampling-induced variability associated with each whale. Of course that ignores the variability over the time the whale was not being sampled, that is, the cue rate is only valid for the time the recording lasted. To account for such additional

variability, one would have to make assumptions about the number of cues produced over time outside the sampling period. For that a model for sound production over time would be required, and that is beyond the scope of our study.

In the future, it should be evaluated whether the cue production rates obtained here could be generalized to other areas since we showed that cue rates depend on depths. The three narwhal populations—in East Greenland, West Greenland and Northern Hudson Bay—present different dive behaviors in distinct seasons, have different diets, and use different depth categories along the water column with diverse purposes, which may result in notorious differences between the sounds produced in distinct clusters or even between individuals (Marcoux *et al.* 2012; Watt *et al.* 2015). It is hard to extrapolate our cue production rates to areas where narwhals have different depth preferences, hence, cue productions rates might have to be estimated separately for each population.

4.6 Test for the influence of narwhals sex and length in echolocation click rates

We could not show that the sex and size (length) of narwhals from Scoresby Sound influence their click rates (Figure 1.14). This is not surprising given our sample size. We only had eight tags to consider in this analysis, making it more difficult to detect differences, in particular to see differences between females and males since our dataset only had two males (Figure 1.15). Thus, to confirm whether acoustic behavior differs between male and female narwhals in the future, we would need data from a greater number of individuals, particularly males.

Tervo *et al.* (2021) showed that there was no influence of narwhals' sex and size on their dive duration and depth, however, the ability to store more oxygen is related to larger body size, allowing longer and deeper dives (Schreer & Kovacs 1997; Kooyman 2009; Tervo *et al.* 2021). In addition, there seems to be a tendency for males in East Greenland to have a larger body size than narwhal females (Garde *et al.* 2015). Moreover, Watt *et al.* (2013) observed that males have a richer diet on benthic prey in comparison to females. Therefore, this may be a sign of a dive depth difference between males and females in East Greenland, especially in winter, when narwhals can dive below 1500m depth (Heide-Jørgensen *et al.* 2015). However, Watt *et al.* (2015) showed that females from East Greenland spend more time at deep depths than males, which may be due to males being less selective in choosing their prey or having a more efficient foraging activity because of their larger size (Beck *et al.* 2003; Dietz *et al.* 2007; Watt *et al.* 2015). If click rates turn out to be different between females and males, and even between different lengths in the future, density estimates for narwhals might not be reliable since a wide variation in cue rates makes it difficult to turn a density of cues into a population density (Marques *et al.* 2013).

Narwhal's dives in Scoresby Sound are mostly V-dives (i.e., shaped as a V), where the duration of time at the bottom is short (Tervo *et al.* 2021). The duration of V-dives is not related to different narwhal sizes, but it is likely that prey distribution has an effect on dive duration since buzzing rates start during the descent phase and remain until the ascent phase (Tervo *et al.* 2021). As a result, the distribution of narwhals prey seems to be scattered along the water column, and also a spatial patchy prey distribution is present (Tervo *et al.* 2021). During transiting foraging, buzzing rates are much lower, and narwhals dive deeper and faster than during stationary foraging (Tervo *et al.* 2021). Additionally, dives with horizontal straight movements are mostly present during transiting foraging, and in contrast, the majority of horizontal movements in stationary foraging are tortuous (Tervo *et al.* 2021). If in the future, prey distribution will have an effect on narwhals click rates, the density estimates will have to be calculated separately for each narwhal population since each population has a different diet (Watt *et al.* 2013; Watt *et al.* 2015).

4.7 Click rate for the last part of Jonas' tag

The click rate obtained for the last part (without acoustics) of Jonas' tag was slightly higher than the click rate obtained for the first part (with acoustics). During the last part, Jonas dove more frequently to deeper depths, with a mean depth greater than the mean depth obtained for the first part, which may indicate that our models made good predictions to estimate the click rate without acoustics. Therefore, the mean cue rate for the subpopulation might be more reliable using the cue rates for the entire record duration of each tag, including the part where the acoustics were not available.

In the future, it should be confirmed whether it is better, i.e. if we could achieve greater precision, by using the click rates for the entire tag duration, including the model information, or by just using the parts where tags provide us the acoustics. This could be evaluated by comparing the real cue rate value, which would be estimated after counting all the clicks in MT Viewer for the full tag, with a model-based estimated cue rate, and with a cue rate estimated using the sampling data, which would require to have a sampling for the last part of Jonas' tag, where the acoustics were available but not processed. Also, the coefficients of variation values should be interpreted for these cue rates.

4.8 Final considerations

The results from this thesis are relevant for the future of narwhals in East Greenland since the cue rates obtained will improve our knowledge about narwhal behavior and thereby help understand the effects of anthropogenic sounds, climate change, and other threats in the near future (Blackwell *et al.* 2018). Moreover, an accurate density estimate based on PAM, needed to carry out narwhals management and conservation is only possible with knowledge of the species' acoustic behavior (Macaulay 2020). The data collected for this thesis happened presumably before anthropogenic sounds in the arctic were a widespread and considerable threat to narwhals (Podolskiy & Sugiyama 2020). With climate change and the increasing threats, the importance of monitoring the state of narwhals from Scoresby Sound over the years should not be understated (Podolskiy & Sugiyama 2020). It is also necessary to have a greater knowledge about the consequences of narwhals capture and tagging as well as an understanding of how to mitigate these effects on narwhals (Hawkins 2004; Casper 2009; Shuert *et al.* 2021).

Narwhals spend most of their time underwater and occur over a wide area at low densities, and traditional visual survey methods (Marques *et al.* 2009) are difficult to carry out in narwhal habitats. As a result, PAM methods sound promising and could be highly effective (Marques *et al.* 2009). PAM has never been used for the estimation of narwhal density, and the mean click rate obtained in this thesis gets us one step closer, allowing a PAM density estimate for the subpopulation of narwhals from Scoresby Sound (Marques *et al.* 2013). The missing steppingstone would be the effective detection area, or equivalently, the probability of detection in a given area around a sensor. This could be obtained in a number of ways, including distance sampling, spatially explicit capture recapture, with trial based methods (Marques *et al.* 2009) or based on acoustic propagation methods (Küsel *et al.* 2011).

Furthermore, it is also important to know whether cue rates are influenced by narwhals' prey distribution, whether they differ between females and males, and have a greater knowledge about prey abundance, how they will be affected by climate change, and confirm whether and to what extent cue production rates obtained in this thesis can be generalized to other areas (Watt *et al.* 2013).

5 REFERENCES (Conservation Biology Journal)

- Abdi H. 2010. Coefficient of variation. *Encyclopedia of Research Design* **1**:169-171.
- Ahonen H, Stafford KM, Lydersen C, de Steur L, Kovacs KM. 2019. A multi-year study of narwhal occurrence in the western Fram Strait—detected via passive acoustic monitoring. *Polar Research* **38**.
- Beck CA, Bowen WD, McMillan JI, Iverson SJ. 2003. Sex differences in the diving behaviour of a size-dimorphic capital breeder: the grey seal. *Animal Behaviour* **66**:777-789.
- Blackwell SB, Tervo OM, Conrad AS, Sinding MHS, Hansen RG, Ditlevsen S, Heide-Jørgensen MP. 2018. Spatial and temporal patterns of sound production in East Greenland narwhals. *PLOS ONE* **13**:e0198295.
- Breed GA, Matthews CJD, Marcoux M, Higdon JW, LeBlanc B, Petersen SD, Orr J, Reinhart NR, Ferguson SH. 2017. Sustained disruption of narwhal habitat use and behavior in the presence of Arctic killer whales. *Proceedings of the National Academy of Sciences* **114**:2628-2633.
- Brown TA, Galicia MP, Thiemann GW, Belt ST, Yurkowski DJ, Dyck MG. 2018. High contributions of sea ice derived carbon in polar bear (*Ursus maritimus*) tissue. *PLOS ONE* **13**:e0191631.
- Burgess WC, Tyack PL, Le Boeuf BJ, Costa DP. 1998. A programmable acoustic recording tag and first results from free-ranging northern elephant seals. *Deep Sea Research Part II: Topical Studies in Oceanography* **45**:1327-1351.
- Casper RM. 2009. Guidelines for the instrumentation of wild birds and mammals. *Animal behaviour* **78**:1477-1483.
- Chambault P, *et al.* 2020. The impact of rising sea temperatures on an Arctic top predator, the narwhal. *Scientific Reports* **10**:18678.
- Cosens SE, Dueck LP. 1993. Icebreaker noise in Lancaster Sound, N.W.T., Canada: Implications for marine mammal behavior. *Marine Mammal Science* **9**:285-300.
- DeRuiter SL, Bahr A, Blanchet M-A, Hansen SF, Kristensen JH, Madsen PT, Tyack PL, Wahlberg M. 2009. Acoustic behaviour of echolocating porpoises during prey capture. *Journal of Experimental Biology* **212**:3100-3107.
- Dietz R, Shapiro AD, Bakhtiari M, Orr J, Tyack PL, Richard P, Eskesen IG, Marshall G. 2007. Upside-down swimming behaviour of free-ranging narwhals. *BMC Ecology* **7**:14.
- Digby PSB. 1953. Plankton Production in Scoresby Sound, East Greenland. *Journal of Animal Ecology* **22**:289-322.
- Finley KJ, Miller GW, Davis RA, Greene CR. 1990. Reactions of belugas, *Delphinapterus leucas*, and narwhals, *Monodon monoceros*, to ice-breaking ships in the Canadian high arctic. *Canadian Bulletin of Fisheries and Aquatic Sciences* **224**:97-117.
- Ford JK, Fisher HD. 1978. Underwater acoustic signals of the narwhal (*Monodon monoceros*). *Canadian Journal of Zoology* **56**:552-560.
- Garde E, Hansen SH, Ditlevsen S, Tvermosegaard KB, Hansen J, Harding KC, Heide-Jørgensen MP. 2015. Life history parameters of narwhals (*Monodon monoceros*) from Greenland. *Journal of Mammalogy* **96**:866-879.
- Garde E, Heide-Jørgensen MP, Hansen SH, Nachman G, Forchhammer MC. 2007. Age-Specific Growth and Remarkable Longevity in Narwhals (*Monodon monoceros*) from West Greenland as Estimated by Aspartic Acid Racemization. *Journal of Mammalogy* **88**:49-58.
- Gomez KA, Gomez AA. 1984. *Statistical Procedures for Agricultural Research*. John Wiley and Sons, Inc, New York.
- Graham ZA, Garde E, Heide-Jørgensen MP, Palaoro AV. 2020. The longer the better: evidence that narwhal tusks are sexually selected. *Biology Letters* **16**:20190950.

- Grebmeier JM. 2012. Shifting Patterns of Life in the Pacific Arctic and Sub-Arctic Seas. *Annual Review of Marine Science* **4**:63-78.
- Hauser DDW, Laidre KL, Stern HL. 2018. Vulnerability of Arctic marine mammals to vessel traffic in the increasingly ice-free Northwest Passage and Northern Sea Route. *Proceedings of the National Academy of Sciences* **115**:7617-7622.
- Hawkins P. 2004. Bio-logging and animal welfare: practical refinements. *Memoirs of National Institute of Polar Research*:58–68.
- Hay KA 1984. The life history of the narwhal (*Monodon monoceros* L.) in the eastern Canadian Arctic. McGill University Montreal, Canada.
- Heide-Jørgensen M, Dietz R, Laidre K, Richard P. 2002. Autumn movements, home ranges, and winter density of narwhals (*Monodon monoceros*) tagged in Tremblay Sound, Baffin Island. *Polar Biology* **25**:331-341.
- Heide-Jørgensen MP. 2018. Narwhal: *Monodon monoceros*. Pages 627-631 in Würsig B, Thewissen JGM, and Kovacs KM, editors. *Encyclopedia of Marine Mammals (Third Edition)*. Academic Press.
- Heide-Jørgensen MP, *et al.* 2021. Behavioral Response Study on Seismic Airgun and Vessel Exposures in Narwhals. *Frontiers in Marine Science* **8**.
- Heide-Jørgensen MP, *et al.* 2020. Some like it cold: Temperature-dependent habitat selection by narwhals. *Ecology and Evolution* **10**:8073-8090.
- Heide-Jørgensen MP, Garde E. 2011. Fetal growth of narwhals (*Monodon monoceros*). *Marine Mammal Science* **27**:659-664.
- Heide-Jørgensen MP, Hammeken N, Dietz R, Orr J, Richard P. 2001. Surfacing Times and Dive Rates for Narwhals (*Monodon monoceros*) and Belugas (*Delphinapterus leucas*). *Arctic* **54**:284-298.
- Heide-Jørgensen MP, Hansen RG, Westdal K, Reeves RR, Mosbech A. 2013a. Narwhals and seismic exploration: Is seismic noise increasing the risk of ice entrapments? *Biological Conservation* **158**:50-54.
- Heide-Jørgensen MP, Laidre KL, Burt ML, Borchers DL, Marques TA, Hansen RG, Rasmussen M, Fossette S. 2010. Abundance of narwhals (*Monodon monoceros*) on the hunting grounds in Greenland *Journal of Mammalogy* **91**:1135-1151.
- Heide-Jørgensen MP, Nielsen NH, Hansen RG, Schmidt HC, Blackwell SB, Jørgensen OA. 2015. The predictable narwhal: satellite tracking shows behavioural similarities between isolated subpopulations. *Journal of Zoology* **297**:54-65.
- Heide-Jørgensen MP, Richard PR, Dietz R, Laidre KL. 2013b. A metapopulation model for Canadian and West Greenland narwhals. *Animal Conservation* **16**:331-343.
- Hiby A, Ward A. 1986. Analysis of cue-counting and blow rate estimation experiments carried out during the 1984/85 IDCR minke whale assessment cruise. Report of the International Whaling Commission **36**:473-476.
- Johnson M, Aguilar de Soto N, Madsen P. 2009. Studying the behaviour and sensory ecology of marine mammals using acoustic recording tags: A review. *Marine Ecology Progress Series* **395**:55-73.
- Johnson M, Madsen PT, Zimmer WMX, Aguilar de Soto N, Tyack PL. 2004. Beaked whales echolocate on prey. *Proceedings of the Royal Society B: Biological Sciences* **271**:S383-S386.
- Johnson MP, Tyack PL. 2003. A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE Journal of Oceanic Engineering* **28**:3-12.
- Koblitz JC, Stilz P, Rasmussen MH, Laidre KL. 2016. Highly Directional Sonar Beam of Narwhals (*Monodon monoceros*) Measured with a Vertical 16 Hydrophone Array. *PLOS ONE* **11**:e0162069.
- Kooyman GL. 2009. Diving Physiology. Pages 327-332 in Perrin WF, Würsig B, and Thewissen JGM, editors. *Encyclopedia of Marine Mammals (Second Edition)*. Academic Press, London.

- Küsel ET, Mellinger DK, Thomas L, Marques TA, Moretti D, Ward J. 2011. Cetacean population density estimation from single fixed sensors using passive acoustics. *Journal of the Acoustical Society of America* **129**:3610-3622.
- Laidre K, Heide-Jørgensen MP. 2005. Winter feeding intensity of narwhals (*Monodon monoceros*). *Marine Mammal Science* **21**:45-57.
- Laidre K, Heide-Jørgensen MP, Dietz R. 2002. Diving behavior of narwhals (*Monodon monoceros*) at two coastal localities in the Canadian High Arctic. *Canadian Journal of Zoology* **80**:624-635.
- Laidre K, Heide-Jørgensen MP, Orr J. 2006. Reactions of narwhals, *Monodon monoceros*, to killer Whale, *Orcinus orca*, attacks in the Eastern Canadian Arctic. *Canadian Field Naturalist* **120**.
- Laidre KL, Heide-Jørgensen MP, Dietz R, Hobbs R, Jørgensen OA. 2003. Deep-diving by narwhals *Monodon monoceros*: Differences in foraging behavior between wintering areas? *Marine Ecology Progress Series* **261**:269-281.
- Laidre KL, Stirling I, Lowry LF, Wiig Ø, Heide-Jørgensen MP, Ferguson SH. 2008. Quantifying the sensitivity of Arctic marine mammals to climate-induced habitat change. *Ecological Applications* **18**:S97-S125.
- Louis M, *et al.* 2020. Influence of past climate change on phylogeography and demographic history of narwhals, *Monodon monoceros*. *Proceedings of the Royal Society B* **287**:20192964.
- Lowry L, Laidre K, Reeves R. 2017. *Monodon monoceros*. The IUCN (International Union for Conservation of Nature) Red List of Threatened Species: e.T13704A50367651.
- Macaulay J. 2020. Passive Acoustic Monitoring of Harbour Porpoise Behaviour, Distribution and Density in Tidal Rapid Habitats. University of St Andrews.
- Madsen P, Kerr I, Payne R. 2004. Echolocation clicks of two free-ranging, oceanic delphinids with different food preferences: false killer whales *Pseudorca crassidens* and Risso's dolphins *Grampus griseus*. *The Journal of experimental biology* **207**:1811-1823.
- Marcoux M, Auger-Méthé M, Humphries M. 2012. Variability and context specificity of narwhal (*Monodon monoceros*) whistles and pulsed calls. *Marine Mammal Science* **28**:649-665.
- Marcoux M, Auger-Méthé M, Humphries MM. 2009. Encounter frequencies and grouping patterns of narwhals in Koluktoo Bay, Baffin Island. *Polar Biology* **32**:1705-1716.
- Marques TA, Thomas L, Martin SW, Mellinger DK, Ward JA, Moretti DJ, Harris D, Tyack PL. 2013. Estimating animal population density using passive acoustics. *Biological Reviews* **88**:287-309.
- Marques TA, Thomas L, Ward J, DiMarzio N, Tyack PL. 2009. Estimating cetacean population density using fixed passive acoustic sensors: An example with Blainville's beaked whales. *The Journal of the Acoustical Society of America* **125**:1982-1994.
- Merchant ND, Fristrup KM, Johnson MP, Tyack PL, Witt MJ, Blondel P, Parks SE. 2015. Measuring acoustic habitats. *Methods in Ecology and Evolution* **6**:257-265.
- Miller PJO, Johnson MP, Tyack PL. 2004. Sperm whale behaviour indicates the use of echolocation click buzzes "creaks" in prey capture. *Proceedings of the Royal Society B: Biological Sciences* **271**:2239-2247.
- Ngô MC, Heide-Jørgensen MP, Ditlevsen S. 2019. Understanding narwhal diving behaviour using Hidden Markov Models with dependent state distributions and long range dependence. *PLOS Computational Biology* **15**:e1006425.
- North Atlantic Marine Mammal Commission (NAMMCO). 2017. Report of the NAMMCO-JCNB Joint Scientific Working Group on Narwhal and Beluga. Copenhagen, Denmark.
- North Atlantic Marine Mammal Commission (NAMMCO). 2019. Report of the Scientific Committee 26th Meeting. Tórshavn, Faroe Islands.
- Nowacek DP, Christiansen F, Bejder L, Goldbogen JA, Friedlaender AS. 2016. Studying cetacean behaviour: new technological approaches and conservation applications. *Animal Behaviour* **120**:235-244.

- Podolskiy EA, Sugiyama S. 2020. Soundscape of a Narwhal Summering Ground in a Glacier Fjord (Inglefield Bredning, Greenland). *Earth and Space Science* **125**:e2020JC016116.
- Porsild MP. 1922. Scattered Observations on Narwhals. *Journal of Mammalogy* **3**:8-13.
- Powell LA. 2007. Approximating variance of demographic parameters using the delta method: a reference for avian biologists. *The Condor* **109**:949-954.
- R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rasmussen M, Koblitz J, Laidre K. 2015. Buzzes and High-Frequency Clicks Recorded from Narwhals (*Monodon monoceros*) at Their Wintering Ground. *Aquatic Mammals* **41**:256-264.
- Reeves RR, *et al.* 2014. Distribution of endemic cetaceans in relation to hydrocarbon development and commercial shipping in a warming Arctic. *Marine Policy* **44**:375-389.
- Scheidat M, Geelhoed SCV, Noort CA 2019. Feasibility of a Passive Acoustic Monitoring (PAM) network for Harbour Porpoises on the Dutch Continental Shelf. Wageningen Marine Research, IJmuiden.
- Schreer J, Kovacs K. 1997. Allometry of diving capacity in air-breathing vertebrates. *Canadian Journal of Zoology* **75**:339-358.
- Shuert CR, Marcoux M, Hussey NE, Watt CA, Auger-Méthé M. 2021. Assessing the post-release effects of capture, handling and placement of satellite telemetry devices on narwhal (*Monodon monoceros*) movement behaviour. *Conservation Physiology* **9**.
- Simmonds MP. 2017. Evaluating the welfare implications of climate change for cetaceans. Pages 125-135. *Marine Mammal Welfare*. Springer.
- Stafford KM, Laidre KL, Heide-Jørgensen MP. 2012. First acoustic recordings of narwhals (*Monodon monoceros*) in winter. *Marine Mammal Science* **28**:E197-E207.
- Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM. 2013. Climate change: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change (IPCC). Cambridge, United Kingdom
- Tervo OM, Ditlevsen S, Ngô MC, Nielsen NH, Blackwell SB, Williams TM, Heide-Jørgensen MP. 2021. Hunting by the Stroke: How Foraging Drives Diving Behavior and Locomotion of East-Greenland Narwhals (*Monodon monoceros*). *Frontiers in Marine Science* **7**.
- Thomas L, Marques TA. 2012. Passive acoustic monitoring for estimating animal density. *Acoustics Today* **8**:35-44.
- Walmsley SF, Rendell L, Hussey NE, Marcoux M. 2020. Vocal sequences in narwhals (*Monodon monoceros*). *Journal of the Acoustical Society of America* **147**:1078-1091.
- Watkins WA, Schevill WE, Ray C. 1971. Underwater Sounds of *Monodon* (Narwhal). *The Journal of the Acoustical Society of America* **49**:595-599.
- Watt C, Orr J, Heide-Jørgensen MP, Nielsen N, Ferguson S. 2015. Differences in dive behavior among the world's three narwhal (*Monodon monoceros*) populations correspond with dietary differences. *Marine Ecology Progress Series* **525**.
- Watt CA, Heide-Jørgensen MP, Ferguson SH. 2013. How adaptable are narwhal? A comparison of foraging patterns among the world's three narwhal populations. *Ecosphere* **4**:71.
- Weilgart L. 2017. Din of the Deep: Noise in the Ocean and Its Impacts on Cetaceans. Pages 111-124 in Butterworth A, editor. *Marine Mammal Welfare: Human Induced Change in the Marine Environment and its Impacts on Marine Mammal Welfare*. Springer International Publishing, Cham.
- Westbury MV, Petersen B, Garde E, Heide-Jørgensen MP, Lorenzen ED. 2019. Narwhal Genome Reveals Long-Term Low Genetic Diversity despite Current Large Abundance Size. *iScience* **15**:592-599.

- Williams TM, Noren SR, Glenn M. 2011. Extreme physiological adaptations as predictors of climate-change sensitivity in the narwhal, *Monodon monoceros*. *Marine Mammal Science* **27**:334-349.
- Wisniewska DM, Johnson M, Nachtigall PE, Madsen PT. 2014. Buzzing during biosonar-based interception of prey in the delphinids *Tursiops truncatus* and *Pseudorca crassidens*. *Journal of Experimental Biology* **217**:4279-4282.

6 APPENDIX

All the code required to implement the analyses in this thesis is presented below.

APPENDIX

Table of Contents

Introduction	34
Reading data	34
Analysis of presence and absence of sounds production.....	36
Calculation of the probability of foraging sounds produced - GAM model	48
Click sampling by whale	50
Freya	53
Seeing how increasing the sample size changes the estimates	53
Seeing if the variance in the average explains the differences between the means obtained for Freya1 and Freya2	57
Data exploration	58
Trends of the number of clicks over time	60
Relationship between the number of clicks with depth	72
Balder	83
Seeing how increasing the sample size changes the estimates	83
Data exploration	84
Trends of the number of clicks over time	85
Relationship between the number of clicks with depth	87
Eistla	89
Seeing how increasing the sample size changes the estimates	89
Data exploration	90
Trends of the number of clicks over time	91
Relationship between the number of clicks with depth	93
Mutti.....	95
Seeing how increasing the sample size changes the estimates	95
Data exploration	96
Trends of the number of clicks over time	97
Relationship between the number of clicks with depth	99
Thora	101
Seeing how increasing the sample size changes the estimates	101
Data exploration	102
Trends of the number of clicks over time	103
Relationship between the number of clicks with depth	105
Frida	107
Seeing how increasing the sample size changes the estimates	107
Data exploration	108
Trends of the number of clicks over time	109
Relationship between the number of clicks with depth	111
Jonas.....	113
Seeing how increasing the sample size changes the estimates	113
Data exploration	114
Trends of the number of clicks over time	115

Relationship between the number of clicks with depth	117
Mara	119
Data exploration	119
Trends of the number of clicks over time	120
Relationship between the number of clicks with depth	122
Models with all the whales.....	124
Trends of the number of clicks over time	124
Relationship between the number of clicks with depth	128
Estimating Click Rates	131
Table with respective values for each whale (clickrates table).....	134
Test for the influence of narwhals sex and length in echolocation clicks rates	139
Estimation of the cue rate in the second part of Jonas data.....	144
References	151

Introduction

In this document, the analysis of the narwhal click rates is described.

In section [Reading Data](#), original tag data and sampling for this tag data will be imported to R. Hereinafter, original tag data will be analyzed in the topic [Analysis of presence and absence of sounds production](#), where will be observed the relationship between depth and probability of foraging sounds produced. Then, in [Click sampling by whale](#) it will be seen for each sampling of tag data how increasing the sample size changes the estimates, will be done the data exploration and will be observed if there is any trend between the number of clicks with depth and over time. In the section [Estimating Click Rates](#) will be created a new table with respective values for each whale to test the influence of sex and length in echolocation clicks rates. In the section [Estimation of the cue rate in the second part of Jonas data](#) will be estimated the cue rate on the part of the Jonas data where acoustic information was available but not processed.

Reading data

It was done two samplings for the Freya data (Freya1 and Freya2) because it will be seen if there is a big difference in the mean of echolocation clicks, with a high coefficient of variation, between each sampling, and comparing each sampling separated with the two samplings together. This will be done to have certainty about the sample size that can be chosen for each narwhal regarding the results that will be obtained for Freya.

```
originaldata <- read.delim("originaldata.txt", sep = "")
sampledata <- read.delim("sampledata.txt", sep = "")

# Freya
FreyaO <- subset(originaldata, whale == "Freya") #Original data

Freya1 <- subset(sampledata[1:100, ], whale == "Freya") #100 samples for Freya1
Freya2 <- subset(sampledata[101:200, ], whale == "Freya") #100 samples for Freya2
```

```
# limit of clicks that are counted in a second
```

```
clim <- 20
```

```
Freya1T <- Freya1
```

```
Freya2T <- Freya2
```

Freya1 has 6 seconds with more than 20 clicks.

```
sum(Freya1T$clicks > clim)
```

```
## [1] 6
```

Freya2 has 3 seconds with more than 20 clicks.

```
sum(Freya2T$clicks > clim)
```

```
## [1] 3
```

More than 20 clicks/s changed to 0 clicks/s because they are not detected on PAM recorders given their low amplitude. This was done in R only for Freya since the counting of clicks in MT Viewer included also the buzzes (more than 20 clicks/s).

```
Freya1T$clicks[Freya1T$clicks > clim] <- 0
```

```
Freya2T$clicks[Freya2T$clicks > clim] <- 0
```

For the other whales, it was not necessary to make this change in R because it was put 0 clicks/s when there were more than 20 clicks/s during the counting of clicks in MT Viewer. The sample size chosen for the other whales was selected according to each record duration and comparing with the sample size and record duration obtained for Freya, so only one sampling was done for each other whale.

```
# Balder
```

```
BalderO <- subset(originaldata, whale == "Balder") #Original data
```

```
Balder1 <- subset(sampledata, whale == "Balder") #200 samples for Balder
```

```
# Eistla
```

```
EistlaO <- subset(originaldata, whale == "Eistla") #Original data
```

```
Eistla1 <- subset(sampledata, whale == "Eistla") #150 samples for Eistla
```

```
# Mutti
```

```
MuttiO <- subset(originaldata, whale == "Mutti") #Original data
```

```
Mutti1 <- subset(sampledata, whale == "Mutti") #150 samples for Mutti
```

```
# Thora
```

```
ThoraO <- subset(originaldata, whale == "Thora") #Original data
```

```

Thora1 <- subset(sampledata, whale == "Thora") #150 samples for Thora

# Frida
FridaO <- subset(originaldata, whale == "Frida") #Original data

Frida150 <- subset(sampledata, whale == "Frida") #150 samples for Frida

Frida149 <- Frida150[-144, ] #149 samples for Frida
## We had to delete the sample 144 because to do the
## analysis I can not have any NA

# Jonas
JonasO <- subset(originaldata, whale == "Jonas") #Original Jonas data

JonasO$row <- (1:nrow(JonasO))

JonasO1 <- JonasO[1:551015, ] #first part of Jonas data processed for sounds

JonasO2 <- JonasO[551016:807030, ] #second part of Jonas data not processed for sounds
Jonas <- subset(sampledata, whale == "Jonas") #200 samples for Jonas

# Mara
MaraO <- subset(originaldata, whale == "Mara") #Original data

Mara1 <- subset(sampledata, whale == "Mara") #10 samples for Mara

```

Analysis of presence and absence of sounds production

Here are plots for each narwhal done in the ggplot tool where can be seen the depths when the sounds are present (in color blue) and the duration in days of each recording. In order to see how long it took the whales to start their foraging behavior, the days that appear on the x values are the days since the capture and tagging. Black lines represent the days from which narwhals start their foraging behavior. After the red line, in Balder data, acoustic information is not available, and in the second part of Jonas data (after the blue line) sound was available but not processed. We can also see that the record's durations were different for each whale. Mára data had a power reset on her record, which led to just one hour of clicking and buzzing (Blackwell *et al.* (2018)). To be able to make the plot, the maximum value of obtainable memory in R was increased to 8000000MB.

```

BalderOW <- BalderO
EistlaOW <- EistlaO
FreyaOW <- FreyaO
FridaOW <- FridaO
MuttiOW <- MuttiO

```

```

ThoraOW <- ThoraO
MaraOW <- MaraO
JonasOW <- JonasO

# Creation of a data frame called whales
FreyaOW$row <- (1:nrow(FreyaOW))
BalderOW$row <- (1:nrow(BalderOW))
EistlaOW$row <- (1:nrow(EistlaOW))
FridaOW$row <- (1:nrow(FridaOW))
MaraOW$row <- (1:nrow(MaraOW))
MuttiOW$row <- (1:nrow(MuttiOW))
ThoraOW$row <- (1:nrow(ThoraOW))
## Adding the number of rows/seconds since the recorded
## started and the names of the whales

Whales <- rbind(BalderOW, EistlaOW, FreyaOW, FridaOW, MuttiOW,
  ThoraOW, MaraOW, JonasOW)

Whales1 <- Whales

Whales1[is.na(Whales1)] <- 0 # transformation of NaN to 0 (absence of sounds)
Whales1[, -c(1, 2, 4, 5, 6)] = (Whales1[, -c(1, 2, 4, 5, 6)] !=
  0) * 1 # values different from zero are equal to 1 (presence of sounds)

Whales1$Clicking <- as.factor(Whales1$Clicking)
### so that clicking is a factor (with only 0 and 1 values)
### and not continuous

memory.limit(size = 8e+06) ##maximum value of obtainable memory in R

## [1] 8e+06

# Observation where the animals were clicking
Whales1$Clicking2 <- Whales1$Clicking
Whales1$Clicking2 <- factor(Whales1$Clicking, levels = 0:2)
Whales1$Clicking2[Whales1$whale == "Jonas" & Whales1$row >= 551016] <- 2
Whales1$Clicking2[Whales1$whale == "Balder" & Whales1$row >=
  601441] <- 2

Whales1$Clicking2 <- factor(Whales1$Clicking2, levels = c("0",
  "1", "2"), labels = c("Silent", "Clicking", "Unknown"))

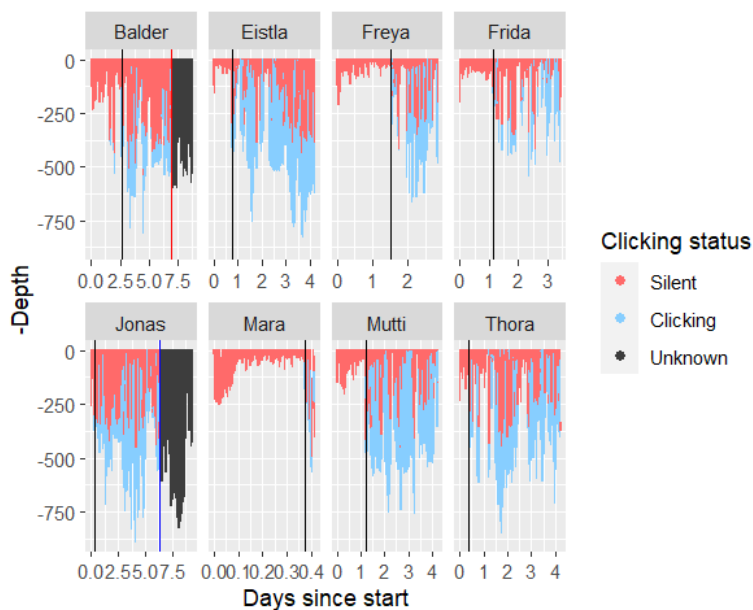
ggplot(data = Whales1, aes(x = (row/(60 * 60 * 24)), y = -Depth,
  colour = Clicking2)) + geom_point(shape = ".") + scale_color_manual(values = c(Silent =
  "indianred1",

```

```

Clicking = "skyblue1", Unknown = "gray24"), name = "Clicking status") +
facet_wrap(~whale, ncol = 4, scales = "free_x") + xlab("Days since start") +
guides(colour = guide_legend(override.aes = list(size = 2,
  shape = 19))) + geom_vline(data = filter(Whales1, whale ==
"Balder"), aes(xintercept = 225750/86400), colour = "black") +
geom_vline(data = filter(Whales1, whale == "Balder"), aes(xintercept = 601441/86400),
  colour = "red") + geom_vline(data = filter(Whales1, whale ==
"Eistla"), aes(xintercept = 64412.5/86400), colour = "black") +
geom_vline(data = filter(Whales1, whale == "Freya"), aes(xintercept = 132410/86400),
  colour = "black") + geom_vline(data = filter(Whales1,
whale == "Frida"), aes(xintercept = 97114/86400), colour = "black") +
geom_vline(data = filter(Whales1, whale == "Mara"), aes(xintercept = 32330/86400),
  colour = "black") + geom_vline(data = filter(Whales1,
whale == "Mutti"), aes(xintercept = 104275/86400), colour = "black") +
geom_vline(data = filter(Whales1, whale == "Thora"), aes(xintercept = 33870/86400),
  colour = "black") + geom_vline(data = filter(Whales1,
whale == "Jonas"), aes(xintercept = 551016/86400), colour = "blue") +
geom_vline(data = filter(Whales1, whale == "Jonas"), aes(xintercept = 23910/86400),
  colour = "black")

```



In Balder data, the firsts 23s of sounds produced after about 12h of silence were ignored and removed since Balder remained silent for nearly 24h between those 23s, which were unusually short clicking times framed

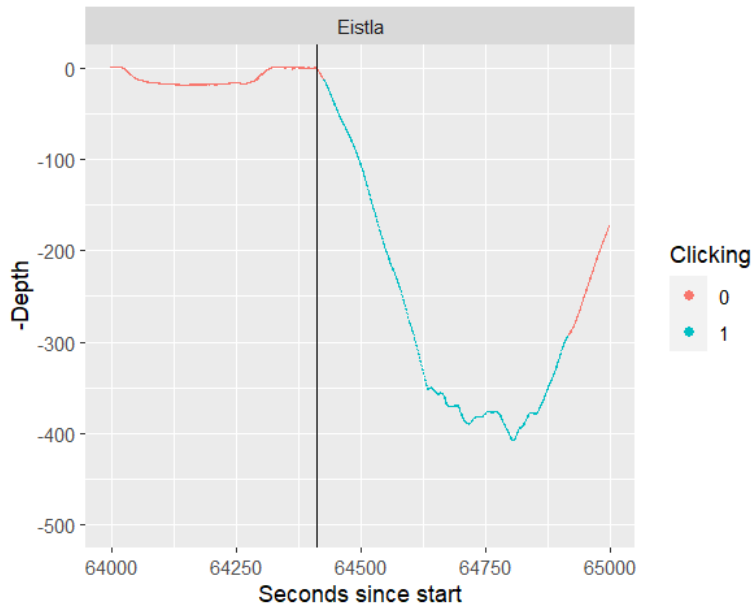
by big gaps of silence (Blackwell *et al.* (2018)). Also, 68s of clicking after those 23s were deleted to have a consistent clicking during the foraging behavior. All other whales were silent for the entire period before the black line. The samples from red and blue lines to the end of the records had to be deleted too.

Before the foraging behavior started, narwhals had long periods of silence because of their capture and tagging, so these periods (before the black line) had to be removed, as they are not representative of narwhals' foraging behavior (Blackwell *et al.* (2018)). To do so, plots for each narwhal were created to select the seconds where the foraging behavior started without those unusual gaps of silence. We considered that the foraging behavior was initiated from the first dive, where sounds were being produced regularly over time. The next plots show the same as above, but for a better visualization, it is only represented the first dive where the sounds were being produced regularly over time for each whale.

First dive considered for Eistla tag

```
ggplot(data = Whales1[Whales1$whale %in% "Eistla", ], aes(x = (row),
  y = -Depth, colour = Clicking)) + geom_point(shape = ".") +
  facet_wrap(~whale, ncol = 4, scales = "free_x") + xlab("Seconds since start") +
  guides(colour = guide_legend(override.aes = list(size = 2,
    shape = 19))) + xlim(64000, 65000) + ylim(-500, 0) +
  geom_vline(data = filter(Whales1, whale == "Eistla"), aes(xintercept = 64412.5),
    colour = "black")
```

Warning: Removed 362479 rows containing missing values (geom_point).



First dive considered for Freya tag

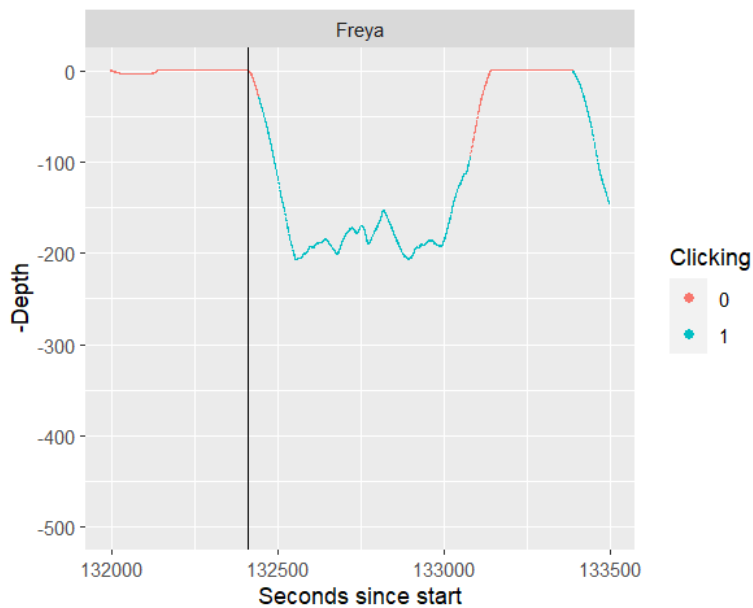
```
ggplot(data = Whales1[Whales1$whale %in% "Freya", ], aes(x = (row),
  y = -Depth, colour = Clicking)) + geom_point(shape = ".") +
  facet_wrap(~whale, ncol = 4, scales = "free_x") + xlab("Seconds since start") +
```

```

guides(colour = guide_legend(override.aes = list(size = 2,
  shape = 19))) + xlim(132000, 133500) + ylim(-500, 0) +
geom_vline(data = filter(Whales1, whale == "Freya"), aes(xintercept = 132410),
  colour = "black")

```

Warning: Removed 244100 rows containing missing values (geom_point).



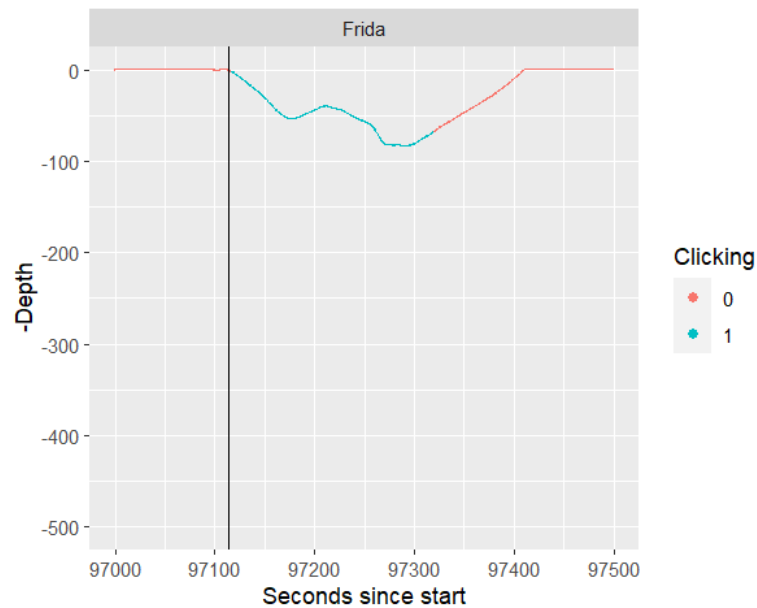
First dive considered for Frida tag

```

ggplot(data = Whales1[Whales1$whale %in% "Frida", ], aes(x = (row),
  y = -Depth, colour = Clicking)) + geom_point(shape = ".") +
facet_wrap(~whale, ncol = 4, scales = "free_x") + xlab("Seconds since start") +
guides(colour = guide_legend(override.aes = list(size = 2,
  shape = 19))) + xlim(97000, 97500) + ylim(-500, 0) +
geom_vline(data = filter(Whales1, whale == "Frida"), aes(xintercept = 97114),
  colour = "black")

```

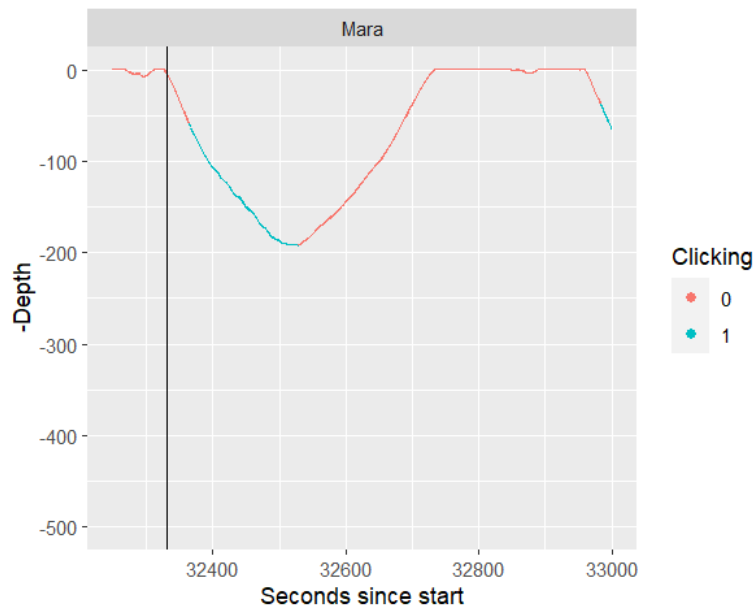
Warning: Removed 297130 rows containing missing values (geom_point).



First dive considered for Mara tag

```
ggplot(data = Whales1[Whales1$whale %in% "Mara", ], aes(x = (row),
  y = -Depth, colour = Clicking)) + geom_point(shape = ".") +
  facet_wrap(~whale, ncol = 4, scales = "free_x") + xlab("Seconds since start") +
  guides(colour = guide_legend(override.aes = list(size = 2,
    shape = 19))) + xlim(32250, 33000) + ylim(-500, 0) +
  geom_vline(data = filter(Whales1, whale == "Mara"), aes(xintercept = 32330),
    colour = "black")
```

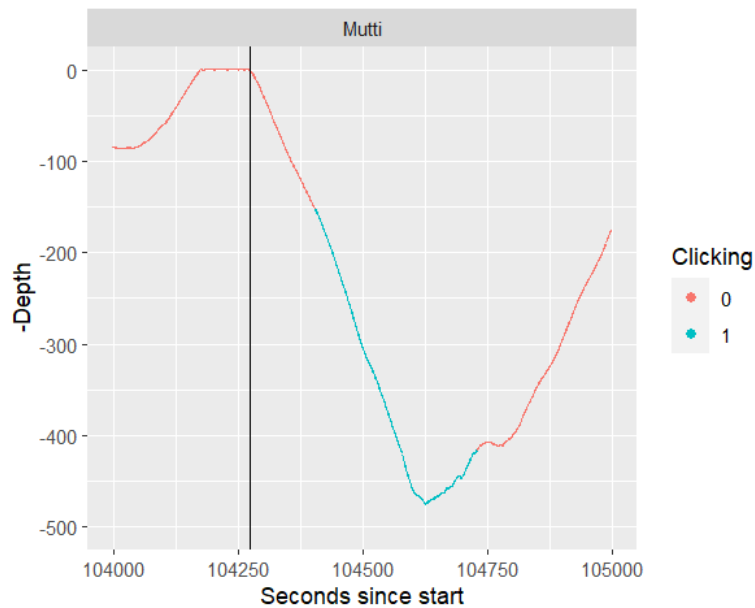
Warning: Removed 35244 rows containing missing values (geom_point).



First dive considered for Mutti tag

```
ggplot(data = Whales1[Whales1$whale %in% "Mutti", ], aes(x = (row),
  y = -Depth, colour = Clicking)) + geom_point(shape = ".") +
  facet_wrap(~whale, ncol = 4, scales = "free_x") + xlab("Seconds since start") +
  guides(colour = guide_legend(override.aes = list(size = 2,
    shape = 19))) + xlim(104000, 105000) + ylim(-500, 0) +
  geom_vline(data = filter(Whales1, whale == "Mutti"), aes(xintercept = 104275),
    colour = "black")
```

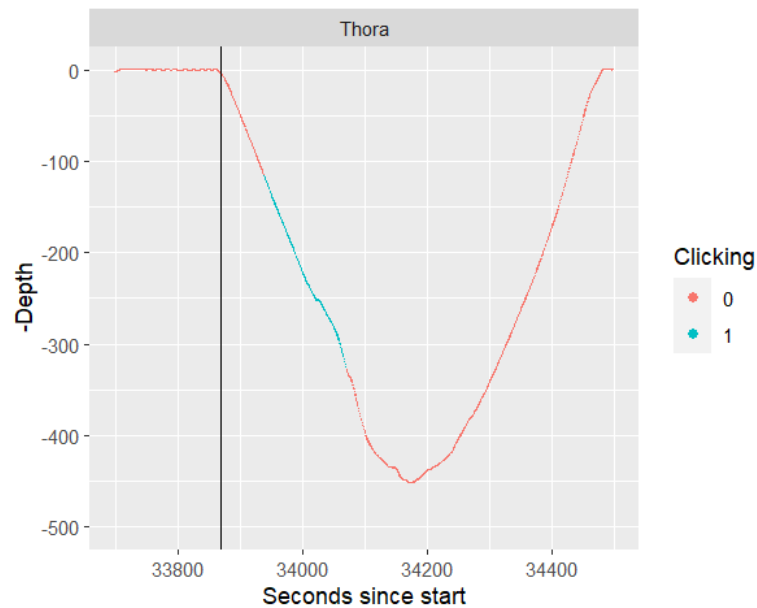
Warning: Removed 364824 rows containing missing values (geom_point).



First dive considered for Thora tag

```
ggplot(data = Whales1[Whales1$whale %in% "Thora", ], aes(x = (row),
  y = -Depth, colour = Clicking)) + geom_point(shape = ".") +
  facet_wrap(~whale, ncol = 4, scales = "free_x") + xlab("Seconds since start") +
  guides(colour = guide_legend(override.aes = list(size = 2,
    shape = 19))) + xlim(33700, 34500) + ylim(-500, 0) +
  geom_vline(data = filter(Whales1, whale == "Thora"), aes(xintercept = 33870),
    colour = "black")
```

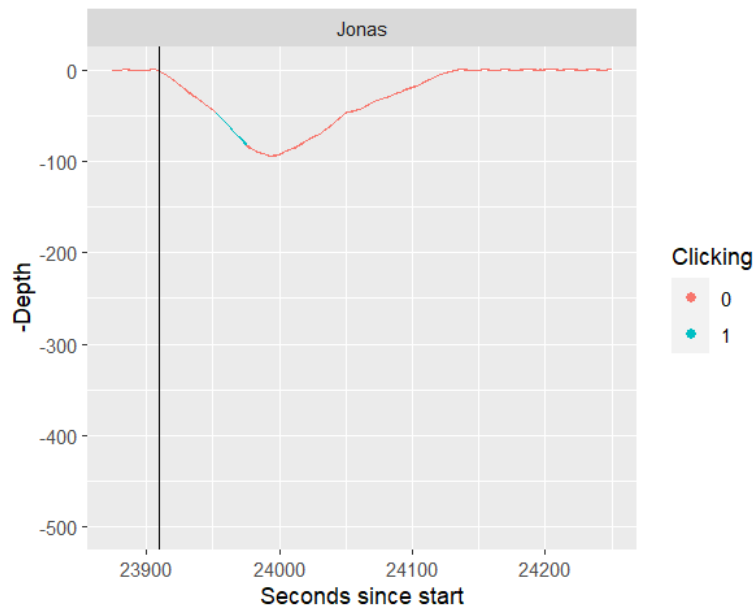
Warning: Removed 364266 rows containing missing values (geom_point).



First dive considered for Jonas tag

```
ggplot(data = Whales1[Whales1$whale %in% "Jonas", ], aes(x = (row),
  y = -Depth, colour = Clicking)) + geom_point(shape = ".") +
  facet_wrap(~whale, ncol = 4, scales = "free_x") + xlab("Seconds since start") +
  guides(colour = guide_legend(override.aes = list(size = 2,
    shape = 19))) + xlim(23875, 24250) + ylim(-500, 0) +
  geom_vline(data = filter(Whales1, whale == "Jonas"), aes(xintercept = 23910),
    colour = "black")
```

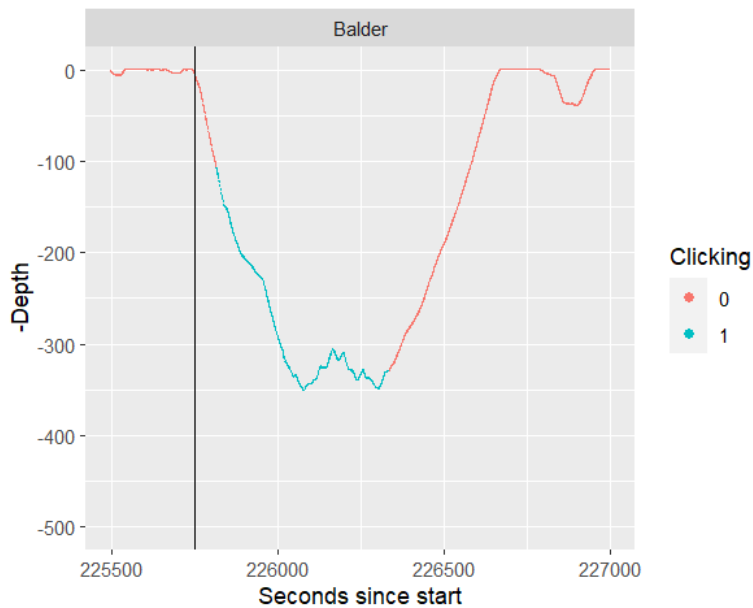
Warning: Removed 806654 rows containing missing values (geom_point).



First dive considered for Balder tag

```
ggplot(data = Whales1[Whales1$whale %in% "Balder", ], aes(x = (row),
  y = -Depth, colour = Clicking)) + geom_point(shape = ".") +
  facet_wrap(~whale, ncol = 4, scales = "free_x") + xlab("Seconds since start") +
  guides(colour = guide_legend(override.aes = list(size = 2,
    shape = 19))) + xlim(225500, 227000) + ylim(-500, 0) +
  geom_vline(data = filter(Whales1, whale == "Balder"), aes(xintercept = 225750),
    colour = "black")
```

Warning: Removed 751201 rows containing missing values (geom_point).



The next plots for each narwhal done in the ggplot tool, where can be seen the depths when the sounds are present (in color blue) and the duration in days of each recording, have the data for each narwhal corrected. As we can see, after taking out the values on each data set, we get a much nicer plot for each whale.

```
# Taking out the rows/seconds where the whales are not in
# their foraging behavior consistently
BalderW <- BalderOW[-c(1:225750, 601441:752702), ]
EistlaW <- EistlaOW[-c(1:64412.5), ]
FreyaW <- FreyaOW[-c(1:132410), ]
FridaW <- FridaOW[-c(1:97114), ]
MaraW <- MaraOW[-c(1:32330), ]
MuttiW <- MuttiOW[-c(1:104275), ]
ThoraW <- ThoraOW[-c(1:33870), ]
JonasW <- JonasOW[-c(1:23910, 551016:807030), ]

# Merging again
whale <- rbind(FreyaW, BalderW, EistlaW, FridaW, MuttiW, ThoraW,
  MaraW, JonasW)

# Again, putting 0 instead of NA, and put 1 where there was
# a sound
whales11 <- whale

whales11[is.na(whales11)] <- 0
whales11[, -c(1, 2, 4, 5, 6)] = (whales11[, -c(1, 2, 4, 5, 6)] !=
  0) * 1
```

```

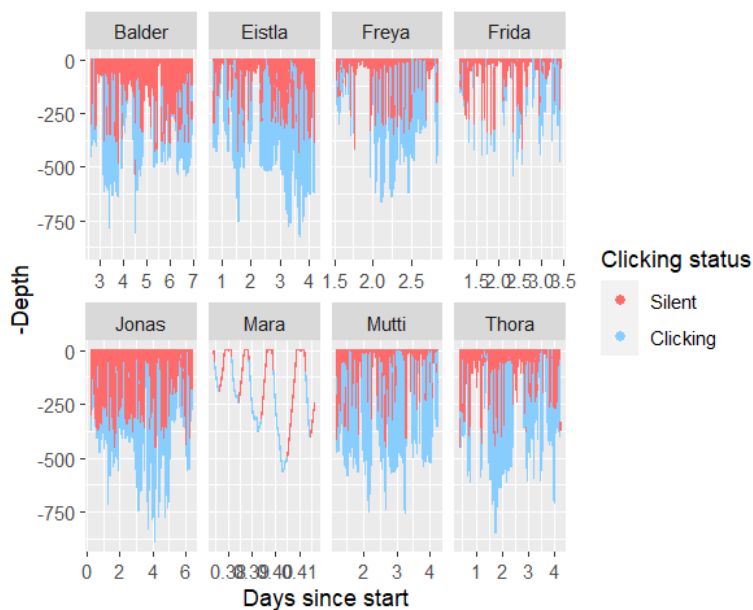
whales11$Clicking <- as.factor(whales11$Clicking)
### so that clicking is a factor (with only 0 and 1 values)
### and not continuous

# Observation where the narwhals were clicking, now with
# the data corrected

whales11$Clicking <- factor(whales11$Clicking, levels = c("0",
  "1"), labels = c("Silent", "Clicking"))

ggplot(data = whales11, aes(x = (row/(60 * 60 * 24)), y = -Depth,
  colour = Clicking)) + geom_point(shape = ".") + scale_color_manual(values = c(Silent =
  "indianred1",
  Clicking = "skyblue1"), name = "Clicking status") + facet_wrap(~whale,
  ncol = 4, scales = "free_x") + xlab("Days since start") +
guides(colour = guide_legend(override.aes = list(size = 2,
  shape = 19)))

```

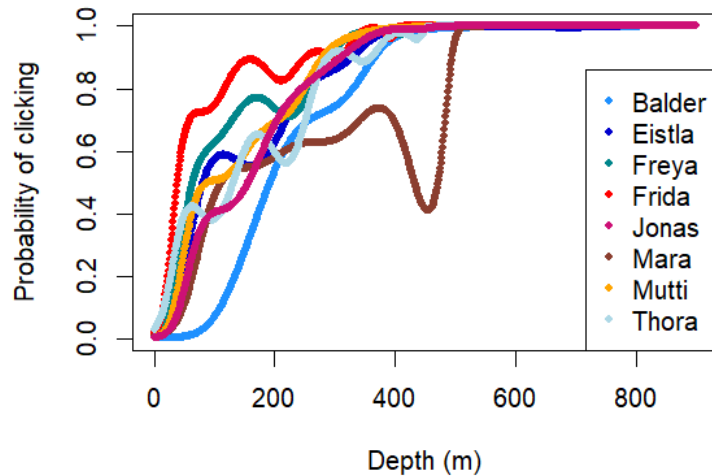


Calculation of the probability of foraging sounds produced - GAM model

Plot representing the relationship between depth and probability of foraging sounds produced for all the whales with whale as a factor (so it is not ignored the behavior differences between each whale). For computational efficiency, the data was subsampled considering a systematic sampling with random start, selecting only one out of every 10 samples in the original data.

```
# Plot for the probability obtained by the function gam()
# for all the whales (with whale as a factor)

gamwhales <- gam(Clicking ~ s(Depth, by = as.factor(whale)) +
  whale, family = "binomial", data = whales11[seq(1, nrow(whales11),
  by = 10), c(2, 3, 5)])
depthall = seq(0, 900, by = 1)
predallBalder <- predict(gamwhales, newdata = data.frame(Depth = depthall,
  whale = "Balder"), type = "response")
plot(depthall, predallBalder, xlab = "Depth (m)", ylab = "Probability of clicking",
  col = "dodgerblue1", pch = 20)
predallEistla <- predict(gamwhales, newdata = data.frame(Depth = depthall,
  whale = "Eistla"), type = "response")
points(depthall, predallEistla, col = "mediumblue", pch = 20)
predallFreya <- predict(gamwhales, newdata = data.frame(Depth = depthall,
  whale = "Freya"), type = "response")
points(depthall, predallFreya, col = "turquoise4", pch = 20)
predallFrida <- predict(gamwhales, newdata = data.frame(Depth = depthall,
  whale = "Frida"), type = "response")
points(depthall, predallFrida, col = "red", pch = 20)
predallMara <- predict(gamwhales, newdata = data.frame(Depth = depthall,
  whale = "Mara"), type = "response")
points(depthall, predallMara, col = "coral4", pch = 20)
predallMutti <- predict(gamwhales, newdata = data.frame(Depth = depthall,
  whale = "Mutti"), type = "response")
points(depthall, predallMutti, col = "orange", pch = 20)
predallThora <- predict(gamwhales, newdata = data.frame(Depth = depthall,
  whale = "Thora"), type = "response")
points(depthall, predallThora, col = "lightblue", pch = 20)
predallJonas <- predict(gamwhales, newdata = data.frame(Depth = depthall,
  whale = "Jonas"), type = "response")
points(depthall, predallJonas, col = "deeppink3", pch = 20)
legend("bottomright", legend = c("Balder", "Eistla", "Freya",
  "Frida", "Jonas", "Mara", "Mutti", "Thora"), col = c("dodgerblue1",
  "mediumblue", "turquoise4", "red", "deeppink3", "coral4",
  "orange", "lightblue"), cex = 1, pch = 20)
```

Summary GAM for all the data. The probability of foraging sounds produced increased significantly (p-value < 0.05) with increasing depth values for each narwhal. Narwhals were buzzing and clicking regularly at depths deeper than 400m. It was noticeable that there was a big foraging activity difference between Frida and Balder. Mara's data (in color brown) had only 1h of recording, leading to a less reliable GAM (data is explained more precisely from models with bigger sample sizes).

```
summary(gamwhales)
```

```
##
## Family: binomial
## Link function: logit
##
## Formula:
## Clicking ~ s(Depth, by = as.factor(whale)) + whale
##
## Parametric coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.9232   0.2646  -18.606 < 2e-16 ***
## whaleEistla  3.3451   0.2708  12.352 < 2e-16 ***
## whaleFreya   356.5524 160.4293  2.222  0.026250 *
## whaleFrida   2503.3397 719.5008  3.479  0.000503 ***
## whaleJonas   13.9137   7.6849   1.811  0.070214 .
## whaleMara     3.1354   1.3673   2.293  0.021840 *
## whaleMutti   91.9415   46.4457   1.980  0.047754 *
## whaleThora   191.1969  27.9163   6.849  7.44e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```

## Approximate significance of smooth terms:
##              edf Ref.df Chi.sq  p-value
## s(Depth):as.factor(whale)Balder 6.735  7.014 3766.29 < 2e-16 ***
## s(Depth):as.factor(whale)Eistla 8.794  8.963 6753.73 < 2e-16 ***
## s(Depth):as.factor(whale)Freya  8.664  8.887 2177.16 < 2e-16 ***
## s(Depth):as.factor(whale)Frida  8.932  8.995 4930.77 < 2e-16 ***
## s(Depth):as.factor(whale)Jonas  8.228  8.501 6923.10 < 2e-16 ***
## s(Depth):as.factor(whale)Mara   6.823  6.992  37.71 6.81e-06 ***
## s(Depth):as.factor(whale)Mutti  8.715  8.920 3712.52 < 2e-16 ***
## s(Depth):as.factor(whale)Thora  8.975  8.999 8054.79 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) = 0.753  Deviance explained = 72.1%
## UBRE = -0.66019  Scale est. = 1      n = 211199

```

Click sampling by whale

To make predictions from the models, it was considered time since the first sample selected started by making the differences between the first selected sample and each other selected samples using `difftime` function with the unit minutes, and, then converted in numeric format for all narwhals data because it was not possible to predict with month-day-year hour:min:sec format.

```

# Freya

# converting the date and time column in format
# month-day-year hour:min:sec using the functions from both
# base R and the lubridate package:

# convert date/time in format month-day-year hour:min:sec
Freya1T$time <- as.POSIXct(Freya1T$time, format = "%m/%d/%y %H:%M:%S",
  tz = "America/Tijuana")
Freya2T$time <- as.POSIXct(Freya2T$time, format = "%m/%d/%y %H:%M:%S",
  tz = "America/Tijuana")

# Considering time since the first sample selected started
Freya1T$mins <- difftime(time1 = Freya1T$time, time2 = Freya1T$time[1],
  units = "mins")
Freya2T$mins <- difftime(time1 = Freya2T$time, time2 = Freya2T$time[1],
  units = "mins")

# Convert time differences (mins) in numeric format
Freya1T$mins = as.numeric(Freya1T$mins)
Freya2T$mins = as.numeric(Freya2T$mins)

```

```

# Balder

# convert date/time in format month-day-year hour:min:sec
Balder1$time <- as.POSIXct(Balder1$time, format = "%m/%d/%y %H:%M:%S",
  tz = "America/Tijuana")

# Considering time since the first sample selected started

Balder1$mins <- difftime(time1 = Balder1$time, time2 = Balder1$time[1],
  units = "mins")

# Convert time differences (mins) in numeric format
Balder1$mins = as.numeric(Balder1$mins)

# Eistla

# convert date/time in format month-day-year hour:min:sec
Eistla1$time <- as.POSIXct(Eistla1$time, format = "%m/%d/%y %H:%M:%S",
  tz = "America/Tijuana")

# Considering time since the first sample selected started
Eistla1$mins <- difftime(time1 = Eistla1$time, time2 = Eistla1$time[1],
  units = "mins")

# Convert time differences (mins) in numeric format
Eistla1$mins = as.numeric(Eistla1$mins)

# Mutti

# convert date/time in format month-day-year hour:min:sec
Mutti1$time <- as.POSIXct(Mutti1$time, format = "%m/%d/%y %H:%M:%S",
  tz = "America/Tijuana")

# Considering time since the first sample selected started
Mutti1$mins <- difftime(time1 = Mutti1$time, time2 = Mutti1$time[1],
  units = "mins")

# Convert time differences (mins) in numeric format
Mutti1$mins = as.numeric(Mutti1$mins)

# Thora

# convert date/time in format month-day-year hour:min:sec

```

```

Thora1$time <- as.POSIXct(Thora1$time, format = "%m/%d/%y %H:%M:%S",
  tz = "America/Tijuana")

# Considering time since the first sample selected started
Thora1$mins <- difftime(time1 = Thora1$time, time2 = Thora1$time[1],
  units = "mins")

# Convert time differences (mins) in numeric format
Thora1$mins = as.numeric(Thora1$mins)

# Frida

# convert date/time in format month-day-year hour:min:sec
Frida149$time <- as.POSIXct(Frida149$time, format = "%m/%d/%y %H:%M:%S",
  tz = "America/Tijuana")

# Considering time since the first sample selected started
Frida149$mins <- difftime(time1 = Frida149$time, time2 = Frida149$time[1],
  units = "mins")

# Convert time differences (mins) in numeric format
Frida149$mins = as.numeric(Frida149$mins)

# Jonas

# convert date/time in format month-day-year hour:min:sec
Jonas$time <- as.POSIXct(Jonas$time, format = "%m/%d/%y %H:%M:%S",
  tz = "America/Tijuana")

# Considering time since the first sample selected started
Jonas$mins <- difftime(time1 = Jonas$time, time2 = Jonas$time[1],
  units = "mins")

# Convert time differences (mins) in numeric format
Jonas$mins = as.numeric(Jonas$mins)

# Mara

# convert date/time in format month-day-year hour:min:sec
Mara1$time <- as.POSIXct(Mara1$time, format = "%m/%d/%y %H:%M:%S",
  tz = "America/Tijuana")

# Considering time since the first sample selected started
Mara1$mins <- difftime(time1 = Mara1$time, time2 = Mara1$time[1],

```

```
units = "mins")
```

```
# Convert time differences (mins) in numeric format
```

```
Mara1$mins = as.numeric(Mara1$mins)
```

Freya

To evaluate sensitivity to the sample selected it was done two samplings only for Freya data. For the other whales, only one sampling was done for each since the chosen sample size and record duration for Freya were used as a comparison.

Seeing how increasing the sample size changes the estimates

The average number of sounds per second for clicks present in the clicking period was calculated for Freya1 (first sampling made for Freya data) data (with 100 samples), and the value obtained was 3.96.

The next plots show for each sample size, a sub-sample of k samples out of 100 that were selected, between sample sizes of x_{min} and x_{max} , and the average was computed. To have a reliable mean of the number of clicks per second, it will be evaluated in the plots below how the increasing sample sizes change the estimate of the mean obtained and how the variances of the mean and their coefficients of variation change with increasing sample sizes. Given the short tag duration, we did not consider Mara in this analysis.

Seeing how increasing the sample size changes the estimates for Freya1. 100 samples seem to be enough to have a good estimate of narwhal cue rates and density since the coefficient of variation of the average number of clicks is less than 20% (Gomez & Gomez (1984)).

```
nobsF1 <- length(Freya1T$clicks)
```

```
# Different sample sizes
```

```
limitF1 <- 5
```

```
nevalF1 <- limitF1:(nobsF1 - limitF1)
```

```
# Number of repetitions for each sample size
```

```
nrepsF1 <- 1000
```

```
# Object to store the averages
```

```
resultadosF1 <- matrix(NA, ncol = length(nevalF1), nrow = nrepsF1)
```

```
# For each sample size
```

```
for (i in 1:length(nevalF1)) {
```

```
  # For each repetition
```

```
  for (j in 1:nrepsF1) {
```

```
    amostraF1 <- sample(Freya1T$clicks, size = nevalF1[i],
```

```
      replace = TRUE)
```

```
    resultadosF1[j, i] <- mean(amostraF1)
```

```
  }
```

```
}
```

```
par(mfrow = c(2, 2))
```

```
plot(x = c(5, 95), y = c(0, 10), xlab = "Different sample sizes",
```

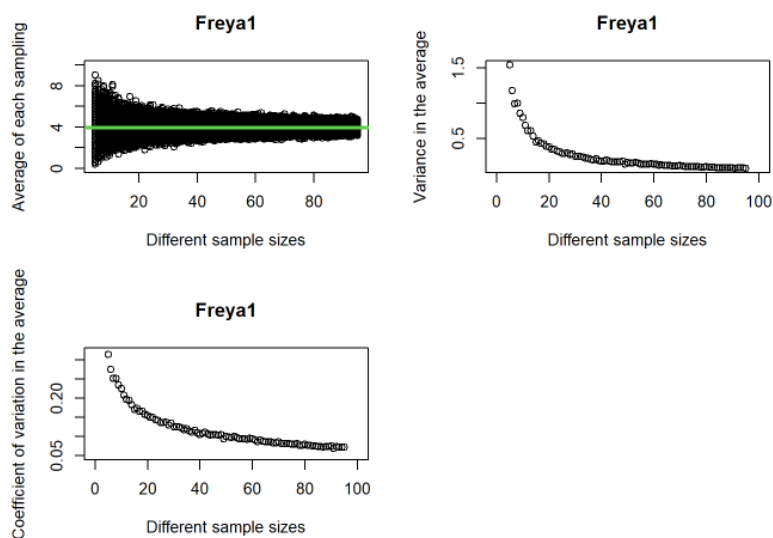
```

ylab = "Average of each sampling", main = list("Freya1"),
type = "n")
for (i in 1:length(nevalF1)) {
  points(rep(nevalF1[i], nrepsF1), resultadosF1[, i])
}
abline(h = mean(Freya1T$clicks), col = 3, lwd = 3)

plot(nevalF1, apply(resultadosF1, 2, var), xlim = c(0, 100),
xlab = "Different sample sizes", ylab = "Variance in the average",
main = list("Freya1"))

# looking at the CV (coefficient of variation)
plot(nevalF1, apply(resultadosF1, 2, sd)/apply(resultadosF1,
2, mean), ylab = "Coefficient of variation in the average",
xlab = "Different sample sizes", ylim = c(0.05, 0.32), xlim = c(0,
100), main = list("Freya1"))

```



The average number of sounds per second for clicks present in the clicking period was calculated for Freya2 (second sampling made for Freya data) data (with 100 samples), and the value obtained was 4.92.

Seeing how increasing the sample size changes the estimates for Freya2. 100 samples seem to be enough to have a good estimate of narwhal cue rates and density since the coefficient of variation of the average number of clicks is less than 20% (Gomez & Gomez (1984)).

```

nobsF2 <- length(Freya2T$clicks)
# Different sample sizes
limitF2 <- 5
nevalF2 <- limitF2:(nobsF2 - limitF2)
# Number of repetitions for each sample size
nrepsF2 <- 1000
# Object to store the averages

```

```

resultadosF2 <- matrix(NA, ncol = length(nevalF2), nrow = nrepsF2)

# For each sample size
for (i in 1:length(nevalF2)) {
  # For each repetition
  for (j in 1:nrepsF2) {
    amostraF2 <- sample(Freya2T$clicks, size = nevalF2[i],
      replace = TRUE)
    resultadosF2[j, i] <- mean(amostraF2)
  }
}
par(mfrow = c(2, 2))

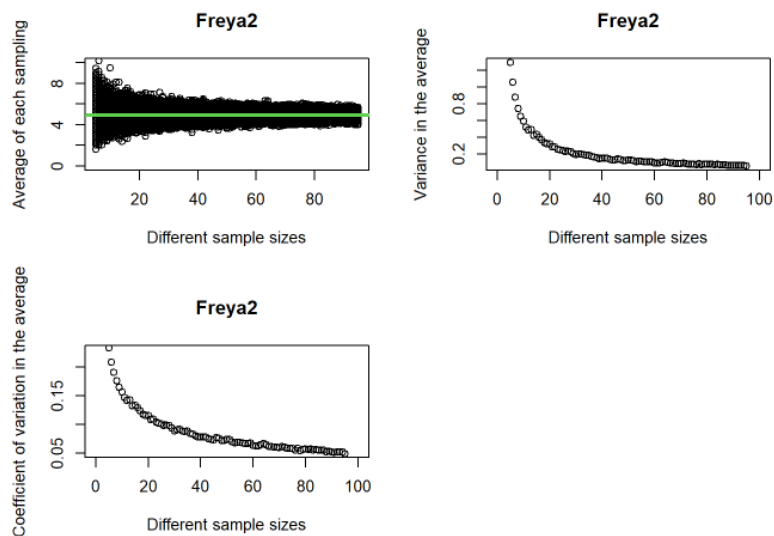
plot(x = c(5, 95), y = c(0, 10), xlab = "Different sample sizes",
  ylab = "Average of each sampling", main = list("Freya2"),
  type = "n")
for (i in 1:length(nevalF2)) {
  points(rep(nevalF2[i], nrepsF2), resultadosF2[, i])
}
abline(h = mean(Freya2T$clicks), col = 3, lwd = 3)

plot(nevalF2, apply(resultadosF2, 2, var), xlim = c(0, 100),
  xlab = "Different sample sizes", ylab = "Variance in the average",
  main = list("Freya2"))

# looking at the CV (coefficient of variation)
plot(nevalF2, apply(resultadosF2, 2, sd)/apply(resultadosF2,
  2, mean), ylab = "Coefficient of variation in the average",
  xlab = "Different sample sizes", ylim = c(0.05, 0.23), xlim = c(0,
  100), main = list("Freya2"))

Freya12 <- rbind(Freya1T, Freya2T)

```



The average number of sounds per second for clicks present in the clicking period was calculated for Freya1 and Freya2 data (with a total of 200 samples), and the value obtained was 4.44.

Seeing how increasing the sample size changes the estimates for Freya1+2 (the two samplings made for Freya data were joined). There is not a big difference in having 100 or 200 samples since with 100 or more samples the mean will be identical and their variation very little.

```
Freya12 <- rbind(Freya1T, Freya2T)
```

```
nobsF12 <- length(Freya12$clicks)
```

```
# Different sample sizes
```

```
limitF12 <- 5
```

```
nevalF12 <- limitF12:(nobsF12 - limitF12)
```

```
# Number of repetitions for each sample size
```

```
nrepsF12 <- 1000
```

```
# Object to store the averages
```

```
resultadosF12 <- matrix(NA, ncol = length(nevalF12), nrow = nrepsF12)
```

```
# For each sample size
```

```
for (i in 1:length(nevalF12)) {
```

```
  # For each repetition
```

```
  for (j in 1:nrepsF12) {
```

```
    amostraF12 <- sample(Freya12$clicks, size = nevalF12[i],  
      replace = TRUE)
```

```
    resultadosF12[j, i] <- mean(amostraF12)
```

```
  }
```

```
}
```

```
par(mfrow = c(2, 2))
```

```
plot(x = c(5, 195), y = c(0, 10), xlab = "Different sample sizes",  
  ylab = "Average of each sampling", main = list("Freya1+2"),
```



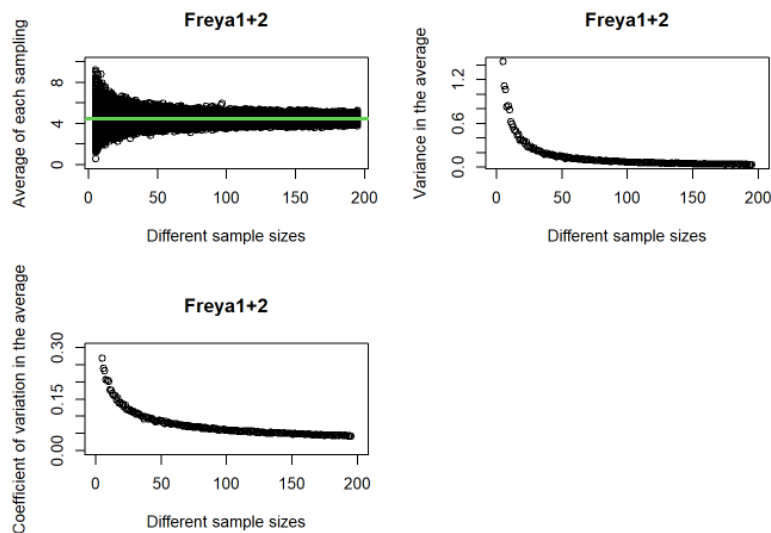
```

type = "n")
for (i in 1:length(nevalF12)) {
  points(rep(nevalF12[i], nrepsF12), resultadosF12[, i])
}
abline(h = mean(Freya12$clicks), col = 3, lwd = 3)

plot(nevalF12, apply(resultadosF12, 2, var), xlim = c(0, 200),
     xlab = "Different sample sizes", ylab = "Variance in the average",
     main = list("Freya1+2"))

# looking at the CV (coefficient of variation)
plot(nevalF12, apply(resultadosF12, 2, sd)/apply(resultadosF12,
2, mean), ylab = "Coefficient of variation in the average",
     xlab = "Different sample sizes", ylim = c(0, 0.3), xlim = c(0,
200), main = list("Freya1+2"))

```



Seeing if the variance in the average explains the differences between the means obtained for Freya1 and Freya2

In the next plot, it can be observed a blue line representing the mean obtained for Freya1 (first sampling made for Freya data), and a red line representing the mean obtained for Freya2 (second sampling made for Freya data). The color points are the average values obtained by a sub-sample of k samples out of 100 that were selected, between sample sizes of x_{min} and x_{max} . The blue and red points are the average values obtained for Freya1 and Freya2 sub-samples, respectively. The grey points are the average values for Freya1 and Freya2 sub-samples that were overlapped.

```

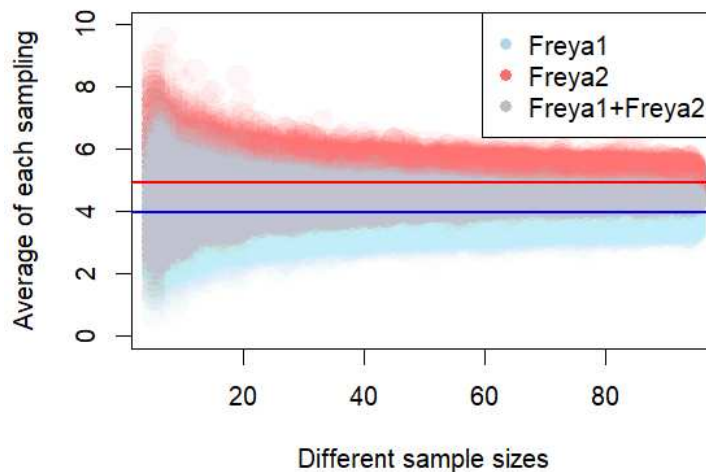
par(mfrow = c(1, 1))
plot(x = c(5, 95), y = c(0, 10), xlab = "Different sample sizes",
     ylab = "Average of each sampling", type = "n")

```

```

for (i in 1:length(nevalF2)) {
  points(rep(nevalF2[i], nrepsF2), resultadosF2[, i], col = alpha("indianred1",
    0.05), pch = 19, cex = 2)
}
for (i in 1:length(nevalF1)) {
  points(rep(nevalF1[i], nrepsF1), resultadosF1[, i], col = alpha("lightblue",
    0.025), pch = 19, cex = 2)
}
abline(h = mean(Freya2T$clicks), col = "red", lwd = 2)
abline(h = mean(Freya1T$clicks), col = "blue", lwd = 2)
legend("topright", legend = c("Freya1", "Freya2", "Freya1+Freya2"),
  col = c("lightblue", "indianred1", "grey"), pch = 19, cex = 1)

```



It can be concluded that the means obtained for Freya1 and Freya2 data (represented by color lines) are within the interval of the average of each sampling from the other data. Hence, the variance in the average explains the differences between the means obtained for Freya1 and Freya2.

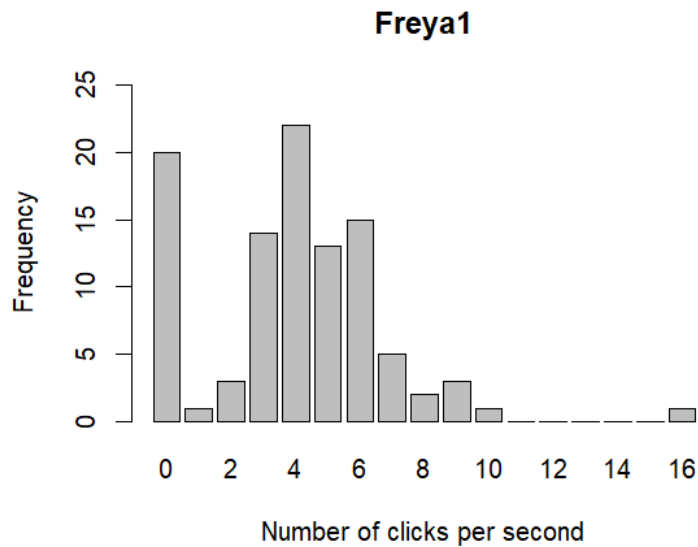
Data exploration

Graph and table of frequencies of the number of clicks counted in Freya1 data.

```

maxnF1 <- max(Freya1T$clicks)
tempF1 <- factor(Freya1T$clicks[Freya1T$clicks <= maxnF1], levels = 0:maxnF1,
  labels = 0:maxnF1)
barplot(table(tempF1), xlab = "Number of clicks per second",
  ylab = "Frequency", main = list("Freya1"), ylim = c(0, 25))

```



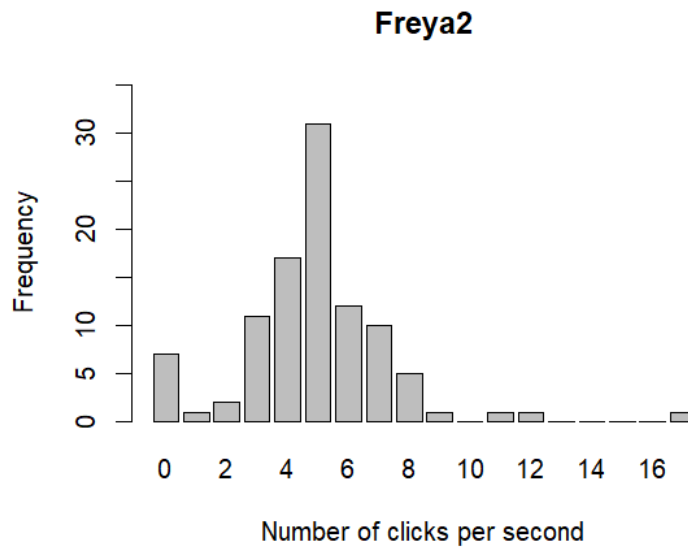
```
table(tempF1)
```

```
## tempF1
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
## 20 1 3 14 22 13 15 5 2 3 1 0 0 0 0 0 1
```

*## In Freya1T, 4 clicks/s were the most frequent count and
0 clicks/s were counted 20 times.*

Graph and table of frequencies of the number of clicks counted in Freya2 data.

```
par(mfrow = c(1, 1))
maxnF2 <- max(Freya2T$clicks)
tempF2 <- factor(Freya2T$clicks[Freya2T$clicks <= maxnF2], levels = 0:maxnF2,
  labels = 0:maxnF2)
barplot(table(tempF2), xlab = "Number of clicks per second",
  ylab = "Frequency", main = list("Freya2"), ylim = c(0, 35))
```



```
table(tempF2)
```

```
## tempF2
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
## 7 1 2 11 17 31 12 10 5 1 0 1 1 0 0 0 0 1
```

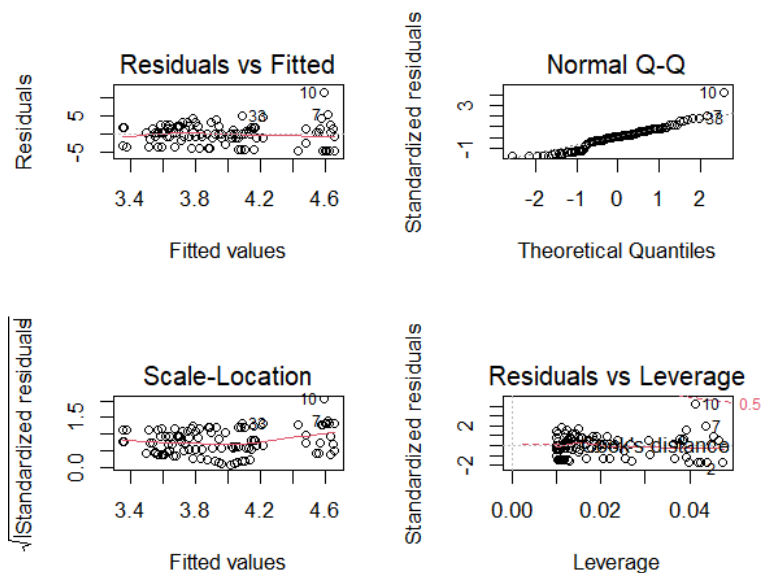
```
## In Freya2T, 5 clicks/s were the most frequent count and
## 0 clicks/s were counted 7 times.
```

Trends of the number of clicks over time

The number of clicks counted in 100 samples was used as the response variable, and the time differences in minutes between samples was the explanatory variable. To know which model would best fit my data, it was seen the residuals, presence of overdispersion, and the AIC values in the LMs (linear models), GLMs (generalized linear models), and GAMs (generalized additive models). It was chosen first the Poisson family since the number of clicks is a count data, but most of the time, overdispersion was present, so to solve this problem, quasipoisson and negative binomial families were used instead. The models with quasipoisson family do not show AIC values, so these models were ignored in the end because the best model was chosen based on the lowest AIC value. In this R document, only Freya1 data shows how the final model was selected. For the other whales, it was only showed the model selected since it was done the same that it is showed for Freya1 data.

The plots below represent the residuals of an LM done for Freya1 data. In general, the residuals of the LM do not seem to be good, because the plot “Scale-Location” suggest that the variance of residuals is not constant, so there is heteroscedasticity, and the plot “Normal Q-Q” shows a non normal residuals distribution. This is expected since the data is a count data. Because the assumptions of Homoscedasticity and normality of the model’s residuals are violated, the LMs will not be used to see the trend of the number of clicks over time.

```
lmFreya1 = lm(Freya1T$clicks ~ Freya1T$mins)
par(mfrow = c(2, 2))
plot(lmFreya1)
```



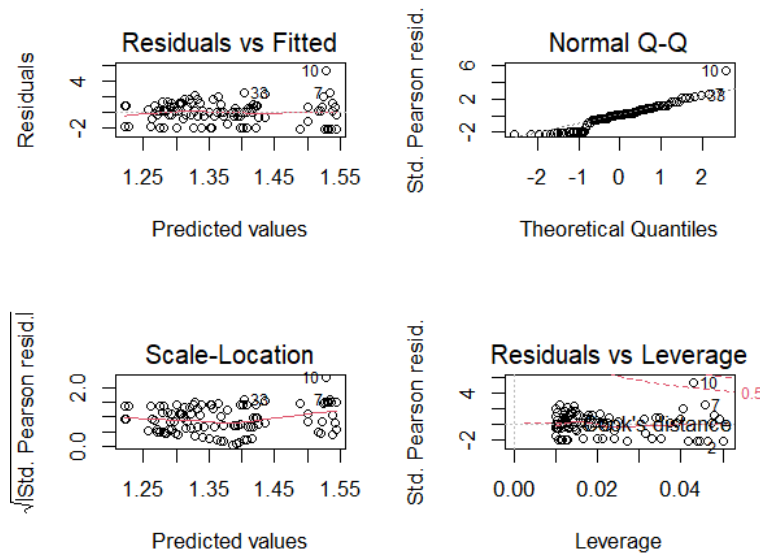
Summary of the LM

```
summary(lmFreya1)
```

```
##
## Call:
## lm(formula = Freya1T$clicks ~ Freya1T$mins)
##
## Residuals:
##   Min     1Q   Median     3Q    Max
## -4.6536 -1.1862  0.1511  1.6416 11.4053
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.6578652  0.6031932   7.722 9.89e-12 ***
## Freya1T$mins -0.0006983  0.0005366  -1.301  0.196
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.761 on 98 degrees of freedom
## Multiple R-squared:  0.01698, Adjusted R-squared:  0.006954
## F-statistic: 1.693 on 1 and 98 DF, p-value: 0.1962
```

Residuals for the Poisson GLM done for Freya1 data. The plot “Residuals vs Leverage” of GLM shows one outlier in Freya1 that will influence the regression results and the residual deviance of the summary in Freya1 is high and not close to the degrees of freedom, so this model is not well-fitted. Also, Freya1 has some presence of patterns and the line that crosses them has a small curvature in the points within residual plots.

```
glmFreya1 = glm(Freya1T$clicks ~ Freya1T$mins, family = poisson(link = "log"))
par(mfrow = c(2, 2))
plot(glmFreya1)
```



Summary of the Poisson GLM

```
summary(glmFreya1)

##
## Call:
## glm(formula = Freya1T$clicks ~ Freya1T$mins, family = poisson(link = "log"))
##
## Deviance Residuals:
##   Min       1Q   Median       3Q      Max
## -3.0599 -0.6039  0.0824  0.8107  4.1255
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.545e+00  1.041e-01  14.843  <2e-16 ***
## Freya1T$mins -1.726e-04  9.566e-05 -1.804  0.0712 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 242.53 on 99 degrees of freedom
## Residual deviance: 239.32 on 98 degrees of freedom
## AIC: 514.85
##
## Number of Fisher Scoring iterations: 5
```

The Pearson estimate should be close to 1. Here we see overdispersion (Pearson estimate > 1) in the Poisson GLM. To solve this problem it will be used a quasipoisson and a negative-binomial regression model.

```
sum(residuals(glmFreya1, type = "pearson")^2)/df.residual(glmFreya1)
## [1] 1.846601
```

Summary of the quasipoisson GLM

```
glmQPFreya1 = glm(Freya1T$clicks ~ Freya1T$mins, family = quasipoisson)
par(mfrow = c(1, 2))
summary(glmQPFreya1)
##
## Call:
## glm(formula = Freya1T$clicks ~ Freya1T$mins, family = quasipoisson)
##
## Deviance Residuals:
##   Min       1Q   Median       3Q      Max
## -3.0599 -0.6039  0.0824  0.8107  4.1255
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.5447013  0.1414227  10.923 <2e-16 ***
## Freya1T$mins -0.0001726  0.0001300  -1.328  0.187
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 1.846602)
##
## Null deviance: 242.53 on 99 degrees of freedom
## Residual deviance: 239.32 on 98 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

Calculation of the theta value required for the negative binomial GLM

```
Freya1.nb <- glm.nb(Freya1T$clicks ~ Freya1T$mins^4, link = log)
c(theta = Freya1.nb$theta, SE = Freya1.nb$SE)
```

```
## theta SE
## 3.051961 0.944656
```

Summary of the negative binomial GLM

```
glmNBFreya1 = glm.nb(Freya1T$clicks ~ Freya1T$mins, link = log,
  init.theta = 3.051961)
summary(glmNBFreya1)
```

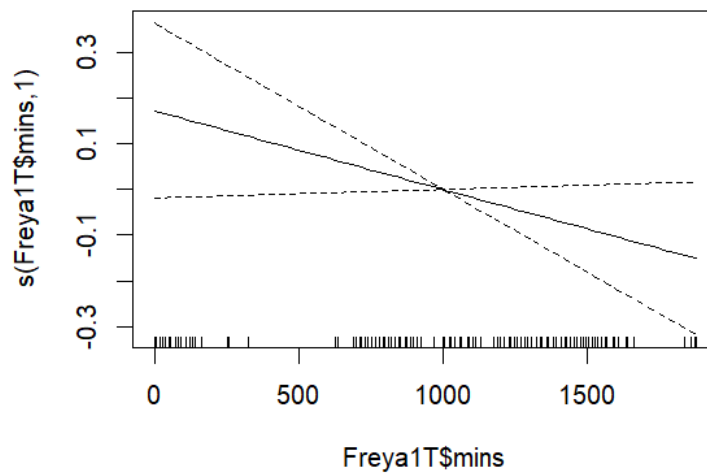
```
##
## Call:
## glm.nb(formula = Freya1T$clicks ~ Freya1T$mins, init.theta = 3.051960514,
## link = log)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -2.38410 -0.40549 0.05522 0.53169 2.25466
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.5470404 0.1627827 9.504 <2e-16 ***
## Freya1T$mins -0.0001750 0.0001469 -1.191 0.234
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(3.052) family taken to be 1)
##
## Null deviance: 133.56 on 99 degrees of freedom
## Residual deviance: 132.15 on 98 degrees of freedom
## AIC: 486.94
##
## Number of Fisher Scoring iterations: 1
##
## Theta: 3.052
## Std. Err.: 0.945
##
## 2 x log-likelihood: -480.939
```

Pearson estimate for the negative binomial GLM

```
sum(residuals(glmNBFreya1, type = "pearson")^2)/df.residual(glmNBFreya1)
## [1] 0.7870001
```

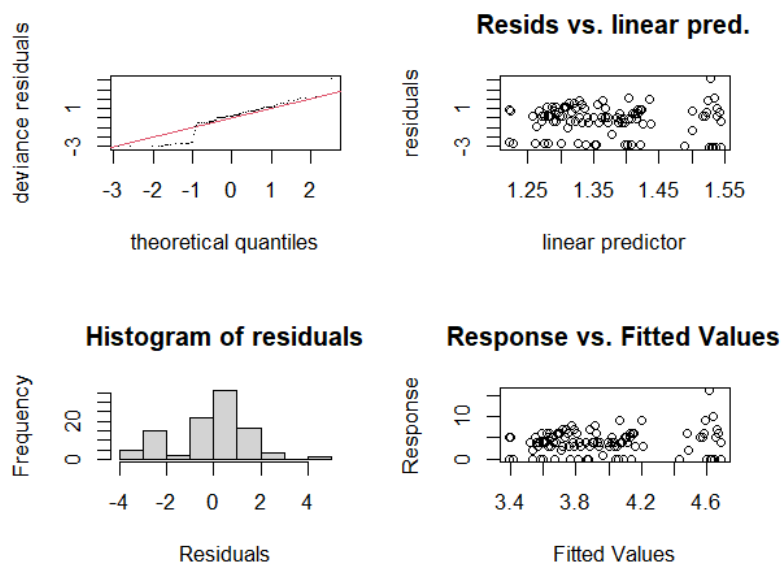

Residuals of the Poisson GAM

```
gamFreya1 = gam(Freya1T$clicks ~ s(Freya1T$mins), family = poisson,  
  link = "log")  
par(mfrow = c(1, 1))  
plot(gamFreya1)
```



Residuals of the Poisson GAM

```
par(mfrow = c(2, 2))  
gam.check(gamFreya1)
```



```
##
## Method: UBRE Optimizer: outer newton
## full convergence after 10 iterations.
## Gradient range [-1.011969e-06,-1.011969e-06]
## (score 1.433166 & scale 1).
## Hessian positive definite, eigenvalue range [1.011816e-06,1.011816e-06].
## Model rank = 10 / 10
##
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##          k' edf k-index p-value
## s(Freya1T$mins) 9  1  1.23  0.99
```

Summary of the Poisson GAM

```
summary(gamFreya1)
```

```
##
## Family: poisson
## Link function: log
##
## Formula:
## Freya1T$clicks ~ s(Freya1T$mins)
##
## Parametric coefficients:
##      Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.37225  0.05045  27.2 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##      edf Ref.df Chi.sq p-value
## s(Freya1T$mins)  1    1  3.254  0.0713 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) = 0.00661  Deviance explained = 1.32%
## UBRE = 1.4332  Scale est. = 1      n = 100
```

Pearson estimate for the Poisson GAM

```
sum(residuals(gamFreya1, type = "pearson")^2)/df.residual(gamFreya1)
```

```
## [1] 1.846602
```

Summary of quasi Poisson GAM

```
gamQPFreya1 = gam(Freya1T$clicks ~ s(Freya1T$mins), family = quasipoisson)
summary(gamQPFreya1)
```

```
##
## Family: quasipoisson
## Link function: log
##
## Formula:
## Freya1T$clicks ~ s(Freya1T$mins)
##
## Parametric coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.37225   0.06856   20.02 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##      edf Ref.df   F p-value
## s(Freya1T$mins)  1    1  1.762  0.187
##
## R-sq.(adj) = 0.00661  Deviance explained = 1.32%
## GCV = 2.4918  Scale est. = 1.8468   n = 100
```

Summary of negative binomial GAM

```
gamNBFreya1 = gam(Freya1T$clicks ~ s(Freya1T$mins), family = nb())
summary(gamNBFreya1)
```

```
##
## Family: Negative Binomial(2.889)
## Link function: log
##
## Formula:
## Freya1T$clicks ~ s(Freya1T$mins)
##
## Parametric coefficients:
##      Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.3722   0.0775   17.7 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##      edf Ref.df Chi.sq p-value
## s(Freya1T$mins)  1    1  1.376  0.241
##
```

```
## R-sq.(adj) = 0.0066 Deviance explained = 1.04%
## -REML = 243.77 Scale est. = 1 n = 100
```

Pearson estimate for the negative binomial GAM

```
sum(residuals(gamNBFreya1, type = "pearson")^2)/df.residual(gamNBFreya1)
```

```
## [1] 0.7624112
```

The glmNBFreya1 (negative binomial GLM) had a smaller AIC value, so it will be chosen this model to see the Trend of the number of clicks over time for Freya1 data.

```
# Best AIC for Freya1
```

```
AIC(lmFreya1, glmFreya1, gamFreya1, glmQPFreya1, glmNBFreya1,
    gamQPFreya1, gamNBFreya1)
```

```
##          df    AIC
## lmFreya1 3.000000 490.8684
## glmFreya1 2.000000 514.8517
## gamFreya1 2.000079 514.8518
## glmQPFreya1 2.000000    NA
## glmNBFreya1 3.000000 486.9386
## gamQPFreya1 3.000049    NA
## gamNBFreya1 3.000325 486.9711
```

Calculation of the theta value required for the negative binomial GLM for Freya1 data.

```
# plot(Freya1T$clicks~Freya1T$mins)
```

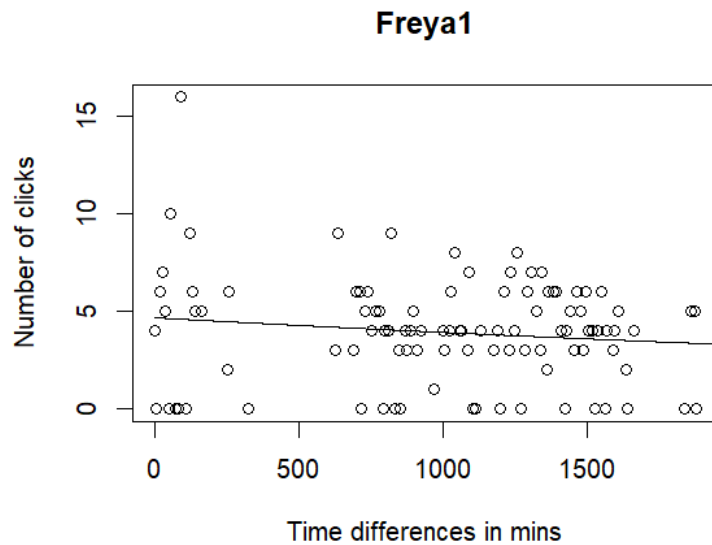
```
Freya1.nb <- glm.nb(Freya1T$clicks ~ Freya1T$mins^4, link = log)
```

```
c(theta = Freya1.nb$theta, SE = Freya1.nb$SE)
```

```
## theta SE
## 3.051961 0.944656
```

The results from the negative binomial GLM show that the number of clicks decreases non significantly over time for Freya1 data. In conclusion, there is not a trend in the number of clicks over time for Freya1 data.

```
glmNBFreya1 = glm.nb(clicks ~ mins, link = log, init.theta = 3.051961,
    data = Freya1T)
DateTime1 = seq(0, 2000, by = 1)
predglmNBFreya1 <- predict(glmNBFreya1, newdata = data.frame(mins = DateTime1),
    type = "response")
plot(Freya1T$mins, Freya1T$clicks, xlab = "Time differences in mins",
    ylab = "Number of clicks", main = list("Freya1"))
lines(DateTime1, predglmNBFreya1)
```



Summary of negative binomial GLM for Freya1

```
summary(glmNBFreya1)
```

```
##
## Call:
## glm.nb(formula = clicks ~ mins, data = Freya1T, init.theta = 3.051960514,
## link = log)
##
## Deviance Residuals:
##   Min       1Q   Median       3Q      Max
## -2.38410 -0.40549  0.05522  0.53169  2.25466
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.5470404  0.1627827   9.504  <2e-16 ***
## mins        -0.0001750  0.0001469  -1.191   0.234
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(3.052) family taken to be 1)
##
## Null deviance: 133.56 on 99 degrees of freedom
## Residual deviance: 132.15 on 98 degrees of freedom
## AIC: 486.94
##
## Number of Fisher Scoring iterations: 1
##
```

```
##
##      Theta: 3.052
##      Std. Err.: 0.945
##
## 2 x log-likelihood: -480.939
```

From here, it will be only showed the models chosen.

The glmNBFreya2 (negative binomial GLM) had a smaller AIC value, so it will be chosen this model to see the Trend of the number of clicks over time for Freya2 data.

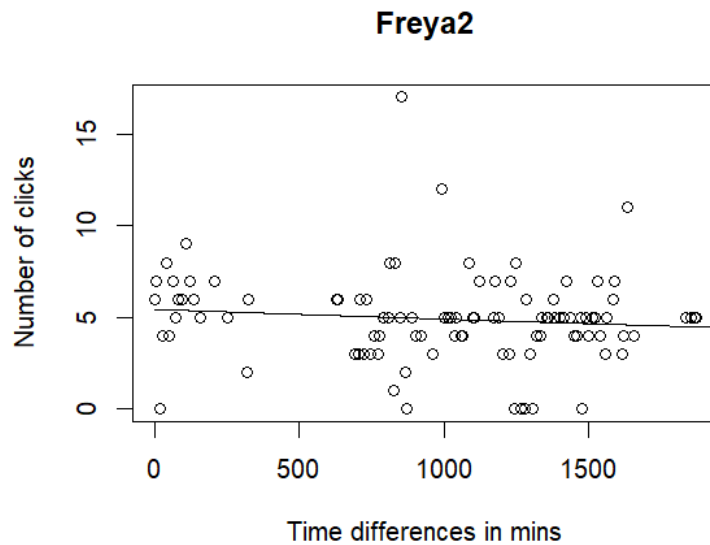
Calculation of the theta value required for the negative binomial GLM for Freya2 data.

```
# plot(Freya2T$clicks~Freya2T$mins)
Freya2.nb <- glm.nb(Freya2T$clicks ~ Freya2T$mins^4, link = log)
c(theta = Freya2.nb$theta, SE = Freya2.nb$SE)

## theta    SE
## 23.51501 19.38791
```

The number of clicks decreases non significantly over time for Freya2 data. In conclusion, there is not any clear pattern over time for Freya2 data.

```
glmNBFreya2 = glm.nb(clicks ~ mins, link = log, init.theta = 23.51501,
  data = Freya2T)
DateTime1 = seq(0, 2000, by = 1)
predglmNBFreya2 <- predict(glmNBFreya2, newdata = data.frame(mins = DateTime1),
  type = "response")
plot(Freya2T$mins, Freya2T$clicks, xlab = "Time differences in mins",
  ylab = "Number of clicks", main = list("Freya2"))
lines(DateTime1, predglmNBFreya2)
```



Summary of negative binomial GLM for Freya2

```
summary(glmNBFreya2)
```

```
##
## Call:
## glm.nb(formula = clicks ~ mins, data = Freya2T, init.theta = 23.51501017,
## link = log)
##
## Deviance Residuals:
##  Min    1Q  Median    3Q   Max
## -3.1286 -0.4155  0.0565  0.3469  3.6314
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.696e+00  1.063e-01  15.953  <2e-16 ***
## mins        -1.035e-04  9.592e-05 -1.079   0.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(23.515) family taken to be 1)
##
## Null deviance: 121.75 on 99 degrees of freedom
## Residual deviance: 120.58 on 98 degrees of freedom
## AIC: 467.86
##
## Number of Fisher Scoring iterations: 1
##
```

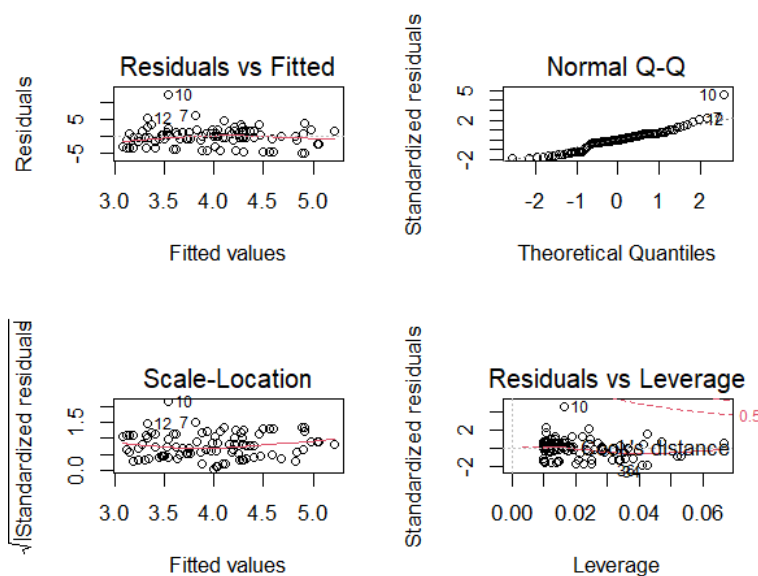
```
##
##      Theta: 23.5
##      Std. Err.: 19.4
##
## 2 x log-likelihood: -461.862
```

Relationship between the number of clicks with depth

The number of clicks counted in 100 samples was used as the response variable, and the depth in meters for each sample was the explanatory variable. To know which model would best fit my data, it was seen the residuals, presence of overdispersion, and the AIC values in the LMs (linear models), GLMs (generalized linear models), and GAMs (generalized additive models). It was chosen first the Poisson family since the number of clicks is a count data, but most of the time, overdispersion was present, so to solve this problem, quasipoisson and negative binomial families were used instead. The models with quasipoisson family do not show AIC values, so these models were ignored in the end because the best model was chosen based on the lowest AIC value. In this R document, only Freya1 data shows how the final model was selected. For the other whales, it was only showed the model selected since it was done the same that it is showed for Freya1 data.

Residuals of the LM. Because the assumptions of Homoscedasticity and normality of the model's residuals are violated, the LMs will not be used to see the relationship between the number of clicks with depth.

```
lmdepthF1 = lm(Freya1T$clicks ~ Freya1T$depth)
par(mfrow = c(2, 2))
plot(lmdepthF1)
```



Summary of the LM

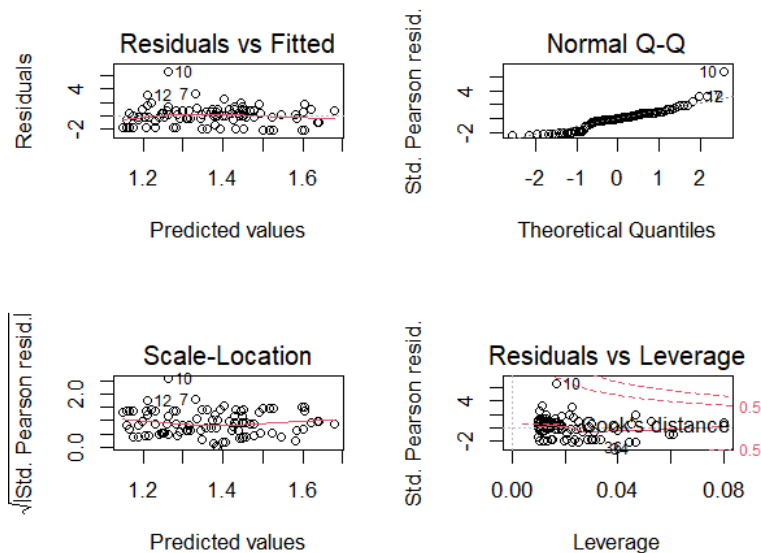
```
summary(lmdepthF1)
```



```
##
## Call:
## lm(formula = Freya1T$clicks ~ Freya1T$depth)
##
## Residuals:
##   Min     1Q   Median     3Q    Max
## -4.9079 -1.1999 -0.0157  1.6444 12.4642
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.041441  0.549163  5.538 2.56e-07 ***
## Freya1T$depth 0.003518  0.001824  1.928  0.0567 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.733 on 98 degrees of freedom
## Multiple R-squared:  0.03656, Adjusted R-squared:  0.02673
## F-statistic: 3.719 on 1 and 98 DF, p-value: 0.05669
```

Residuals of the Poisson GLM

```
glmdepthF1 = glm(Freya1T$clicks ~ Freya1T$depth, family = poisson(link = "log"))
par(mfrow = c(2, 2))
plot(glmdepthF1)
```



Summary of the Poisson GLM. The residual deviance of the summary in Freya1 are high and not close to the degrees of freedom, so this model is not well-fitted.

```
summary(glmdepthF1)
```

```
##  
## Call:  
## glm(formula = Freya1T$clicks ~ Freya1T$depth, family = poisson(link = "log"))  
##  
## Deviance Residuals:  
##   Min     1Q  Median     3Q      Max   
## -3.1526 -0.7271  0.0098  0.7418  4.8376   
##  
## Coefficients:  
##              Estimate Std. Error z value Pr(>|z|)      
## (Intercept)  1.1391250  0.1056574  10.781 < 2e-16 ***  
## Freya1T$depth 0.0008750  0.0003307   2.646  0.00816 **  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for poisson family taken to be 1)  
##  
##   Null deviance: 242.53  on 99  degrees of freedom  
## Residual deviance: 235.59  on 98  degrees of freedom  
## AIC: 511.12  
##  
## Number of Fisher Scoring iterations: 5
```

Pearson estimate for the Poisson GLM. Here we see overdispersion (Pearson estimate>1) in the Poisson GLM. To solve this problem it will be used a quasipoisson and a negative-binomial regression model.

```
sum(residuals(glmdepthF1, type = "pearson")^2)/df.residual(glmdepthF1)
```

```
## [1] 1.948371
```

Summary of the quasipoisson GLM

```
glmQPdepthF1 = glm(Freya1T$clicks ~ Freya1T$depth, family = quasipoisson)
```

```
summary(glmQPdepthF1)
```

```
##  
## Call:  
## glm(formula = Freya1T$clicks ~ Freya1T$depth, family = quasipoisson)  
##  
## Deviance Residuals:  
##   Min     1Q  Median     3Q      Max   
## -3.1526 -0.7271  0.0098  0.7418  4.8376   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 1.1391250 0.1474809 7.724 9.8e-12 ***
## Freya1T$depth 0.0008750 0.0004616 1.895 0.061 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 1.948372)
##
## Null deviance: 242.53 on 99 degrees of freedom
## Residual deviance: 235.59 on 98 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

Calculation of the theta value required for the negative binomial GLM for Freya1 data.

```
depth1.nbF1 <- glm.nb(Freya1T$clicks ~ Freya1T$depth^4, link = log)
c(theta = depth1.nbF1$theta, SE = depth1.nbF1$SE)

## theta SE
## 3.194483 1.010274
```

Summary of the negative binomial GLM

```
glmNBdepthF1 = glm.nb(Freya1T$clicks ~ Freya1T$depth, link = log,
  init.theta = 3.194483)
summary(glmNBdepthF1)

##
## Call:
## glm.nb(formula = Freya1T$clicks ~ Freya1T$depth, init.theta = 3.194483277,
## link = log)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -2.46313 -0.51311 0.00594 0.46509 2.86318
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.1198104 0.1548947 7.229 4.85e-13 ***
## Freya1T$depth 0.0009463 0.0005001 1.892 0.0585 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(3.1945) family taken to be 1)
##
## Null deviance: 135.93 on 99 degrees of freedom
```

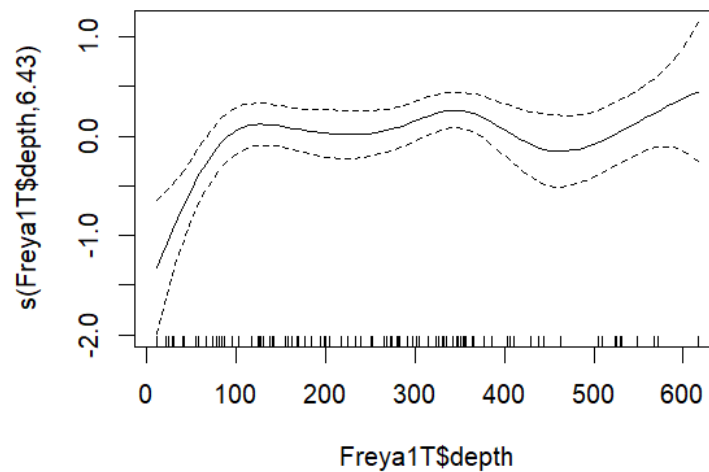
```
## Residual deviance: 132.60 on 98 degrees of freedom
## AIC: 485.07
##
## Number of Fisher Scoring iterations: 1
##
##      Theta: 3.19
##      Std. Err.: 1.01
##
## 2 x log-likelihood: -479.073
```

Pearson estimate for the negative binomial GLM

```
sum(residuals(glmNBdepthF1, type = "pearson")^2)/df.residual(glmNBdepthF1)
## [1] 0.8924479
```

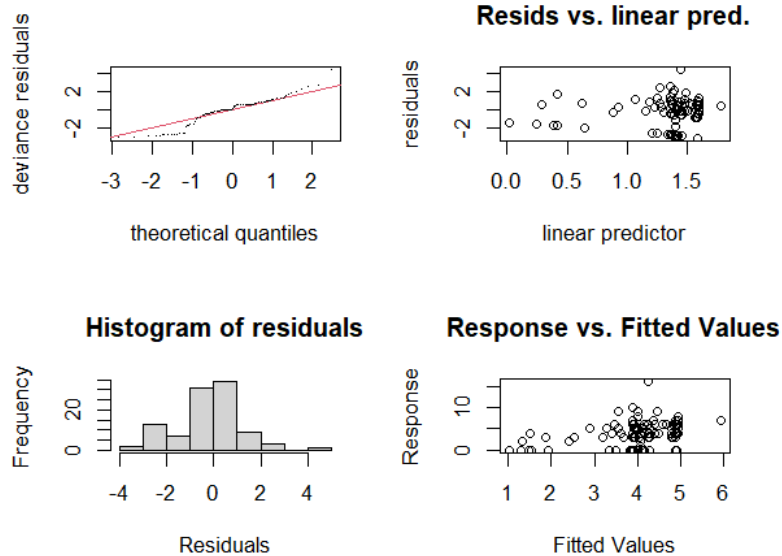
Residuals of the Poisson GAM

```
gamdepthF1 = gam(Freya1T$clicks ~ s(Freya1T$depth), family = poisson,
  link = "log")
par(mfrow = c(1, 1))
plot(gamdepthF1)
```



Residuals of the Poisson GAM

```
par(mfrow = c(2, 2))
gam.check(gamdepthF1)
```



```
##
## Method: UBRE Optimizer: outer newton
## full convergence after 3 iterations.
## Gradient range [1.038807e-07,1.038807e-07]
## (score 1.261061 & scale 1).
## Hessian positive definite, eigenvalue range [0.01133146,0.01133146].
## Model rank = 10 / 10
##
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##           k' edf k-index p-value
## s(Freya1T$depth) 9.00 6.43 0.99 0.52
```

Summary of the Poisson GAM

```
summary(gamdepthF1)
```

```
##
## Family: poisson
## Link function: log
##
## Formula:
## Freya1T$clicks ~ s(Freya1T$depth)
##
## Parametric coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.33633 0.05258 25.41 <2e-16 ***
```

```
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df Chi.sq p-value
## s(Freya1T$depth) 6.434 7.575 22.09 0.00375 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) = 0.0705  Deviance explained = 12.9%
## UBRE = 1.2611  Scale est. = 1      n = 100
```

Pearson estimate for the Poisson GAM

```
sum(residuals(gamdepthF1, type = "pearson")^2)/df.residual(gamdepthF1)
```

```
## [1] 1.826964
```

Summary of the quasipoisson GAM

```
gamQPdepthF1 = gam(Freya1T$clicks ~ s(Freya1T$depth), family = quasipoisson)
summary(gamQPdepthF1)
```

```
##
## Family: quasipoisson
## Link function: log
##
## Formula:
## Freya1T$clicks ~ s(Freya1T$depth)
##
## Parametric coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.36218  0.07075  19.25 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df  F p-value
## s(Freya1T$depth) 2.157 2.695 2.35 0.113
##
## R-sq.(adj) = 0.0457  Deviance explained = 6.1%
## GCV = 2.4281  Scale est. = 1.9308  n = 100
```

Summary of the negative binomial GAM

```
gamNBdepthF1 = gam(Freya1T$clicks ~ s(Freya1T$depth), family = nb())
summary(gamNBdepthF1)
```

```
##
## Family: Negative Binomial(3.265)
## Link function: log
##
## Formula:
## Freya1T$clicks ~ s(Freya1T$depth)
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.3565    0.0754  17.99 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf Ref.df Chi.sq p-value
## s(Freya1T$depth) 2.323  2.9  7.266 0.0781 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) = 0.048  Deviance explained = 5.83%
## -REML = 242.26  Scale est. = 1      n = 100
```

Pearson estimate for the negative binomial GAM

```
sum(residuals(gamNBdepthF1, type = "pearson")^2)/df.residual(gamNBdepthF1)
## [1] 0.8930996
```

For Freya1, it will be chosen gamNBdepthF1 (negative binomial GAM) to see the relationship between the number of clicks with depth because it had the lowest AIC.

Best AIC for Freya1

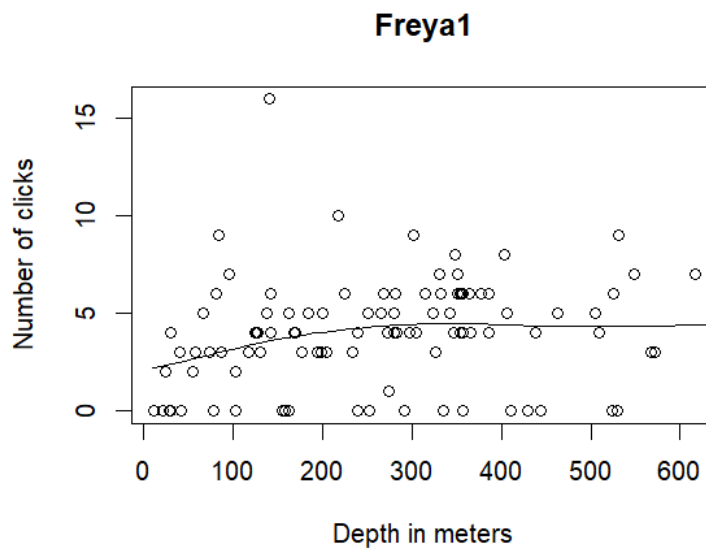
```
AIC(lmdepthF1, glmdepthF1, gamdepthF1, glmQPdepthF1, glmNBdepthF1,
    gamQPdepthF1, gamNBdepthF1)
##             df  AIC
## lmdepthF1  3.000000 488.8569
## glmdepthF1  2.000000 511.1218
## gamdepthF1  7.433657 497.6413
## glmQPdepthF1 2.000000    NA
## glmNBdepthF1 3.000000 485.0732
## gamQPdepthF1 4.156671    NA
## gamNBdepthF1 4.900158 484.2576
```

The results from the negative binomial GAM show that the number of clicks increases non significantly with increasing depth for Freya1 data. In conclusion, there is not a relationship between the number of clicks with depth for Freya1 data.

```

gamNBdepthF1 = gam(clicks ~ s(depth), family = nb(), data = Freya1T)
depthFreya1 = seq(10, 700, by = 10)
predgamNBdepthF1 <- predict(gamNBdepthF1, newdata = data.frame(depth = depthFreya1),
  type = "response")
plot(Freya1T$depth, Freya1T$clicks, xlab = "Depth in meters",
  ylab = "Number of clicks", main = list("Freya1"))
lines(depthFreya1, predgamNBdepthF1)

```



Summary of the negative binomial GAM for Freya1 data

```

summary(gamNBdepthF1)
##
## Family: Negative Binomial(3.265)
## Link function: log
##
## Formula:
## clicks ~ s(depth)
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.3565    0.0754  17.99 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf Ref.df Chi.sq p-value
## s(depth)  2.323   2.9  7.266  0.0781 .

```



```
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) = 0.048  Deviance explained = 5.83%
## -REML = 242.26  Scale est. = 1      n = 100
```

From here, it will be only showed the models chosen.

For Freya2, it will be chosen glmNBdepthF2 (negative binomial GLM) to see the relationship between the number of clicks with depth because it had the lowest AIC.

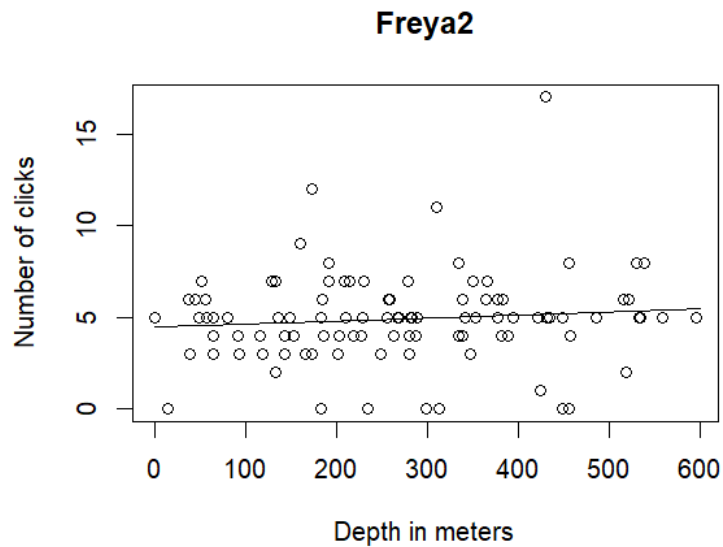
Calculation of the theta value required for the negative binomial GLM for Freya2 data.

```
depth2.nbF2 <- glm.nb(Freya2T$clicks ~ Freya2T$depth^4, link = log)
c(theta = depth2.nbF2$theta, SE = depth2.nbF2$SE)

## theta SE
## 23.18059 18.89555
```

The results from the negative binomial GLM show that the number of clicks increases non significantly with increasing depth for Freya2 data. In conclusion, there is not a relationship between the number of clicks with depth for Freya2 data.

```
glmNBdepthF2 = glm.nb(clicks ~ depth, link = log, init.theta = 23.18059,
  data = Freya2T)
depthFreya2 = seq(0, 600, by = 10)
predglmNBdepthF2 <- predict(glmNBdepthF2, newdata = data.frame(depth = depthFreya2),
  type = "response")
plot(Freya2T$depth, Freya2T$clicks, xlab = "Depth in meters",
  ylab = "Number of clicks", main = list("Freya2"))
lines(depthFreya2, predglmNBdepthF2)
```



Summary of the negative binomial GLM for Freya2 data

```
summary(glmNBdepthF2)
```

```
##
## Call:
## glm.nb(formula = clicks ~ depth, data = Freya2T, init.theta = 23.18059171,
## link = log)
##
## Deviance Residuals:
##   Min       1Q   Median       3Q      Max
## -3.0716 -0.4334 -0.0190  0.3915  3.5186
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.5037724  0.1043124  14.416  <2e-16 ***
## depth        0.0003306  0.0003359   0.984   0.325
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(23.1806) family taken to be 1)
##
## Null deviance: 121.50 on 99 degrees of freedom
## Residual deviance: 120.53 on 98 degrees of freedom
## AIC: 468.06
##
## Number of Fisher Scoring iterations: 1
##
```

```
##
##      Theta: 23.2
##      Std. Err.: 18.9
##
## 2 x log-likelihood: -462.06
```

Balder

Seeing how increasing the sample size changes the estimates

The average number of sounds per second for clicks present in the clicking period was calculated for Balder data (with 200 samples), and the value obtained was 4.7.

The next plots show for each sample size, a sub-sample of k samples out of 200 that were selected, between sample sizes of x_{min} and x_{max} , and the average was computed.

Seeing how increasing the sample size changes the estimates for Balder. 200 samples seem to be enough to have a good estimate of narwhal cue rates and density since the coefficient of variation of the averages number of clicks is less than 20% (Gomez & Gomez (1984)).

```
nobsB <- length(Balder1$clicks)
# Different sample sizes
limitB <- 5
nevalB <- limitB:(nobsB - limitB)
# Number of repetitions for each sample size
nrepsB <- 1000
# Object to store the averages
resultadosB <- matrix(NA, ncol = length(nevalB), nrow = nrepsB)

# For each sample size
for (i in 1:length(nevalB)) {
  # For each repetition
  for (j in 1:nrepsB) {
    amostraB <- sample(Balder1$clicks, size = nevalB[i],
                      replace = TRUE)
    resultadosB[j, i] <- mean(amostraB)
  }
}
par(mfrow = c(2, 2))

plot(nevalB, apply(resultadosB, 2, var), xlim = c(0, 200), ylim = c(0,
  1.5), xlab = "Different sample sizes", ylab = "Variance in the average",
  main = list("Balder"))

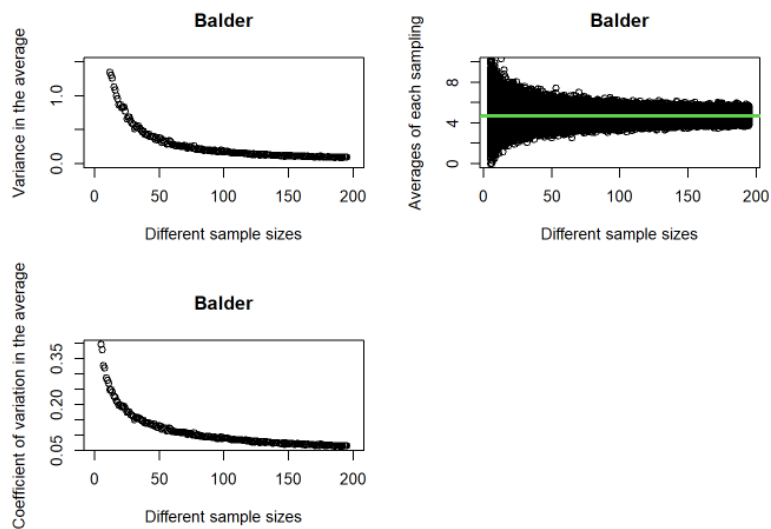
plot(x = c(5, 195), y = c(0, 10), xlab = "Different sample sizes",
  ylab = "Averages of each sampling", main = list("Balder"),
```

```

type = "n")
for (i in 1:length(nevalB)) {
  points(rep(nevalB[i], nrepsB), resultadosB[, i])
}
abline(h = mean(Balder1$clicks), col = 3, lwd = 3)

# looking at the CV (coefficient of variation)
plot(nevalB, apply(resultadosB, 2, sd)/apply(resultadosB, 2,
mean), ylab = "Coefficient of variation in the average",
xlab = "Different sample sizes", xlim = c(0, 200), main = list("Balder"))

```



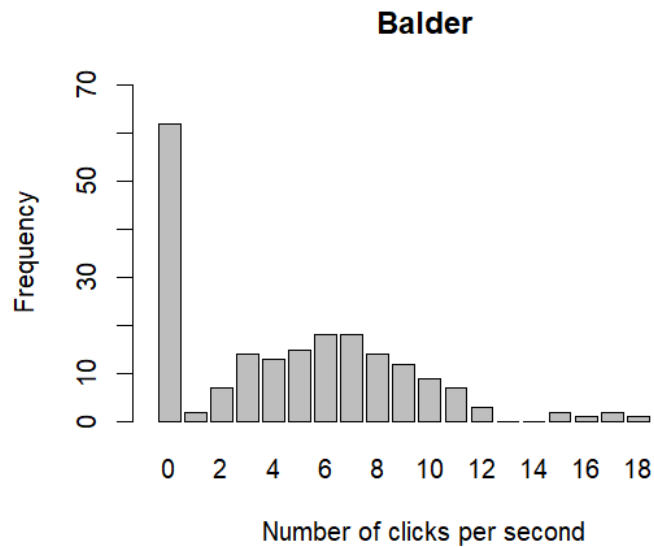
Data exploration

Graph and table of frequencies of the number of clicks counted in Balder data.

```

maxnB1 <- max(Balder1$clicks)
tempB1 <- factor(Balder1$clicks[Balder1$clicks <= maxnB1], levels = 0:maxnB1,
  labels = 0:maxnB1)
barplot(table(tempB1), xlab = "Number of clicks per second",
  ylab = "Frequency", main = list("Balder"), ylim = c(0, 70),
  xlim = c(0, 25))

```



```
table(tempB1)
## tempB1
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
## 62 2 7 14 13 15 18 18 14 12 9 7 3 0 0 2 1 2 1
## In Balder1, 0 clicks/s were the most frequent count (62
## times counted) followed by 6 and 7 clicks/s (18 times
## counted).
```

Clicks counted in 47 seconds were turned into 0 clicks/s because more than 20 Clicks/s are not detected on PAM recorders given its low amplitude.

Trends of the number of clicks over time

The number of clicks counted in 200 samples was used as the response variable, and the time differences in minutes between samples was the explanatory variable. To know which model would best fit my data, it was done the same process that was done and showed previously for Freya1 data. Therefore, it will only be shown the model selected.

The glmNBB1 (negative binomial GLM) had a smaller AIC value, so it will be chosen this model to see the Trend of the number of clicks over time for Balder data.

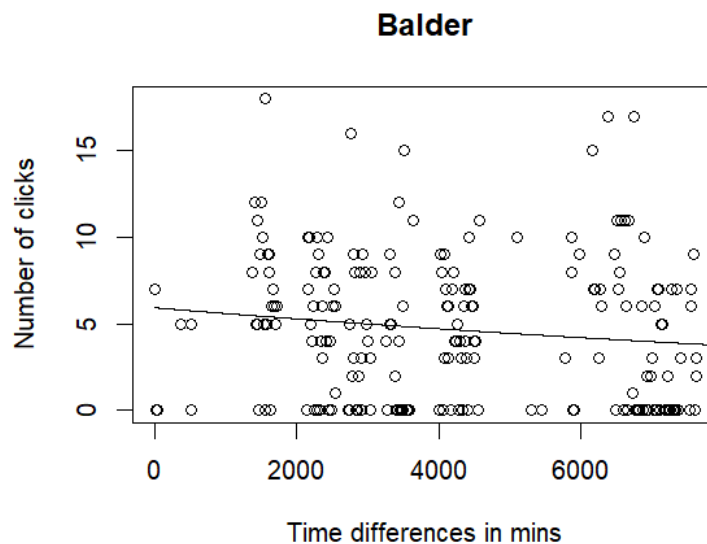
Calculation of the theta value required for the negative binomial GLM for Balder data.

```
# plot(Balder1$clicks~Balder1$mins)
par(mfrow = c(1, 1))
Balder1.nb <- glm.nb(Balder1$clicks ~ Balder1$mins^4, link = log)
c(theta = Balder1.nb$theta, SE = Balder1.nb$SE)
```

```
## theta SE
## 0.8487441 0.1230611
```

The results from the negative binomial GLM show that the number of clicks decreases (negative estimate) non significantly over time for Balder data. In conclusion, there is not any clear pattern over time for Balder data.

```
glmNBB1 = glm.nb(clicks ~ mins, link = log, init.theta = 0.8487441,
  data = Balder1)
DateTimeB1 = seq(0, 10000, by = 1)
predglmNBB1 <- predict(glmNBB1, newdata = data.frame(mins = DateTimeB1),
  type = "response")
plot(Balder1$mins, Balder1$clicks, xlab = "Time differences in mins",
  ylab = "Number of clicks", main = list("Balder"))
lines(DateTimeB1, predglmNBB1)
```



Summary of the negative binomial GLM for Balder data

```
summary(glmNBB1)
##
## Call:
## glm.nb(formula = clicks ~ mins, data = Balder1, init.theta = 0.8487440652,
## link = log)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -1.87853 -1.71726 -0.09517 0.46207 1.61995
##
```

```
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.782e+00 1.875e-01  9.506 <2e-16 ***
## mins      -5.716e-05 4.009e-05 -1.426  0.154
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(0.8487) family taken to be 1)
##
## Null deviance: 239.60 on 199 degrees of freedom
## Residual deviance: 237.65 on 198 degrees of freedom
## AIC: 1061.3
##
## Number of Fisher Scoring iterations: 1
##
##
##           Theta: 0.849
##           Std. Err.: 0.123
##
## 2 x log-likelihood: -1055.340
```

Relationship between the number of clicks with depth

The number of clicks counted in 200 samples was used as the response variable, and the depth in meters for each sample was the explanatory variable. To know which model would best fit my data, it was done the same process that was done and showed previously for Freya1 data. Therefore, it will only be shown the model selected.

The glmNBB2 (negative binomial GLM) had a smaller AIC value, so it will be chosen this model to see the relationship between the number of clicks with depth for Balder data.

Calculation of the theta value required for the negative binomial GLM for Balder data.

```
Balder1.nb1 <- glm.nb(Balder1$clicks ~ Balder1$depth^4, link = log)
c(theta = Balder1.nb1$theta, SE = Balder1.nb1$SE)

## theta    SE
## 0.8365472 0.1206131
```

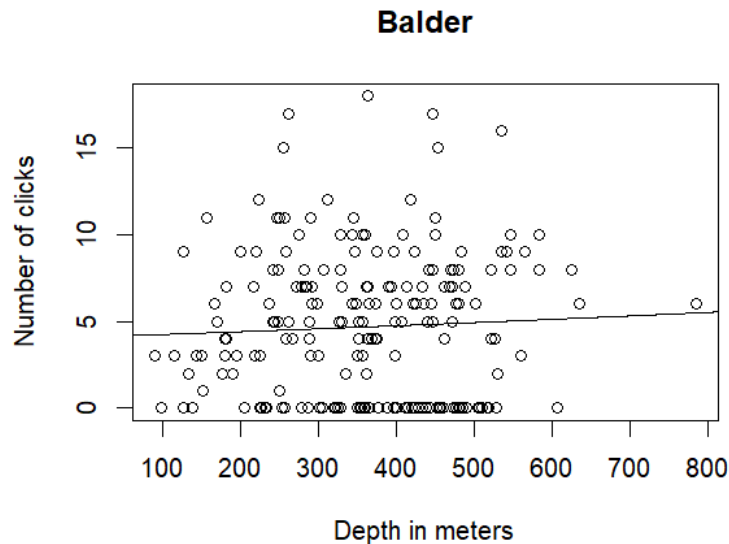
The results from the negative binomial GLM show that the number of clicks increases (positive estimate) non significantly with increasing depth for Balder data. In conclusion, there is not a relationship between the number of clicks with depth for Balder data.

```
glmNBB2 = glm.nb(clicks ~ depth, link = log, init.theta = 0.8365472,
  data = Balder1)
depthBalder = seq(0, 900, by = 1)
predglmNBB2 <- predict(glmNBB2, newdata = data.frame(depth = depthBalder),
```

```

type = "response")
plot(Balder1$depth, Balder1$clicks, xlab = "Depth in meters",
      ylab = "Number of clicks", main = list("Balder"))
lines(depthBalder, predglmNBB2)

```



Summary of the negative binomial GLM for Balder data

```

summary(glmNBB2)

##
## Call:
## glm.nb(formula = clicks ~ depth, data = Balder1, init.theta = 0.8365471475,
## link = log)
##
## Deviance Residuals:
##  Min     1Q   Median     3Q    Max
## -1.81464 -1.76593  0.02672  0.41884  1.48742
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.4117966  0.2651949   5.324 1.02e-07 ***
## depth        0.0003743  0.0006972   0.537  0.591
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(0.8365) family taken to be 1)
##
## Null deviance: 237.51 on 199 degrees of freedom

```



```

## Residual deviance: 237.23 on 198 degrees of freedom
## AIC: 1063
##
## Number of Fisher Scoring iterations: 1
##
##          Theta: 0.837
##        Std. Err.: 0.121
##
## 2 x log-likelihood: -1056.990

```

Eistla

Seeing how increasing the sample size changes the estimates

The average number of sounds per second for clicks present in the clicking period was calculated for Eistla data (with 150 samples), and the value obtained was 4.7.

The next plots show for each sample size, a sub-sample of k samples out of 150 that were selected, between sample sizes of x_{min} and x_{max} , and the average was computed.

Seeing how increasing the sample size changes the estimates for Eistla. 150 samples seem to be enough to have a good estimate of narwhal cue rates and density since the coefficient of variation of the averages number of clicks is less than 20% (Gomez & Gomez (1984)).

```

nobsE <- length(Eistla1$clicks)
# Different sample sizes
limitE <- 5
nevalE <- limitE:(nobsE - limitE)
# Number of repetitions for each sample size
nrepsE <- 1000
# Object to store the averages
resultadosE <- matrix(NA, ncol = length(nevalE), nrow = nrepsE)

# For each sample size
for (i in 1:length(nevalE)) {
  # For each repetition
  for (j in 1:nrepsE) {
    amostraE <- sample(Eistla1$clicks, size = nevalE[i],
                      replace = TRUE)
    resultadosE[j, i] <- mean(amostraE)
  }
}
par(mfrow = c(2, 2))

plot(nevalE, apply(resultadosE, 2, var), xlim = c(0, 150), xlab = "Different sample sizes",

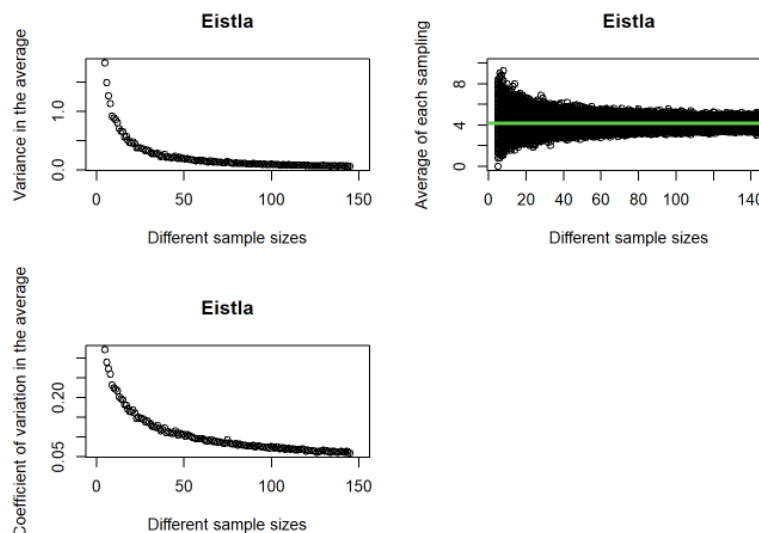
```

```
ylab = "Variance in the average", main = list("Eistla"))
```

```
plot(x = c(5, 145), y = c(0, 10), xlab = "Different sample sizes",
     ylab = "Average of each sampling", main = list("Eistla"),
     type = "n")
for (i in 1:length(nevalE)) {
  points(rep(nevalE[i], nrepsE), resultadosE[, i])
}
abline(h = mean(Eistla1$clicks), col = 3, lwd = 3)
```

looking at the CV (coefficient of variation)

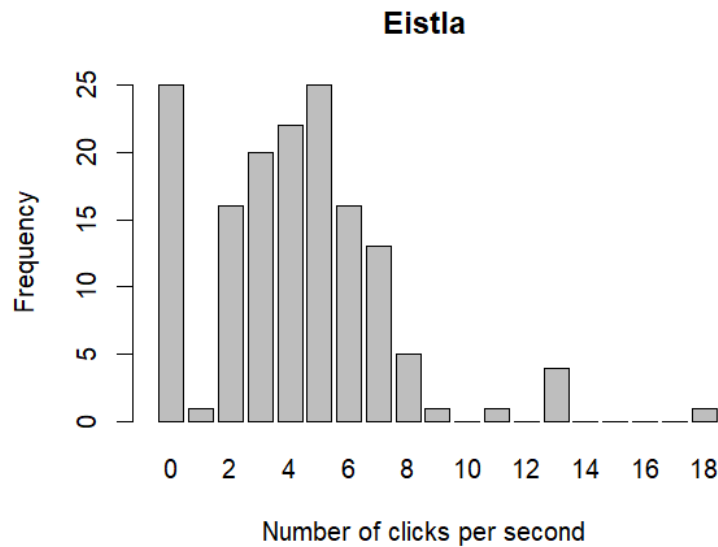
```
plot(nevalE, apply(resultadosE, 2, sd)/apply(resultadosE, 2,
mean), ylab = "Coefficient of variation in the average",
     xlab = "Different sample sizes", xlim = c(0, 150), main = list("Eistla"))
```



Data exploration

Graph and table of frequencies of the number of clicks counted in Eistla data.

```
maxnE1 <- max(Eistla1$clicks)
tempE1 <- factor(Eistla1$clicks[Eistla1$clicks <= maxnE1], levels = 0:maxnE1,
                 labels = 0:maxnE1)
barplot(table(tempE1), xlab = "Number of clicks per second",
          ylab = "Frequency", main = list("Eistla"), ylim = c(0, 25),
          xlim = c(0, 22))
```



```
table(tempE1)

## tempE1
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
## 25 1 16 20 22 25 16 13 5 1 0 1 0 4 0 0 0 0 1
```

*# In Eistla1, 0 and 5 clicks/s were the most frequent count
(25 times counted) followed by 4 clicks/s (22 times
counted).*

Clicks counted in 14 seconds were turned into 0 clicks/s because more than 20 Clicks/s are not detected on PAM recorders given its low amplitude.

Trends of the number of clicks over time

The number of clicks counted in 150 samples was used as the response variable, and the time differences in minutes between samples was the explanatory variable. To know which model would best fit my data, it was done the same process that was done and showed previously for Freya1 data. Therefore, it will only be shown the model selected.

The glmNBE1 (negative binomial GLM) had a smaller AIC value, so it will be chosen this model to see the Trend of the number of clicks over time for Eistla data.

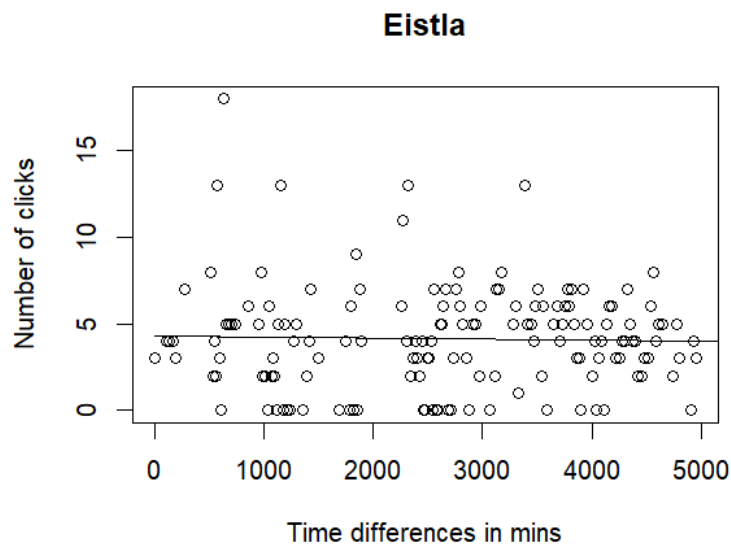
Calculation of the theta value required for the negative binomial GLM for Eistla data.

```
# plot(Eistla1$clicks~Eistla1$mins)
par(mfrow = c(1, 1))
Eistla1.nb <- glm.nb(Eistla1$clicks ~ Eistla1$mins^4, link = log)
c(theta = Eistla1.nb$theta, SE = Eistla1.nb$SE)
```

```
## theta SE
## 2.861314 0.646925
```

The results from the negative binomial GLM show that the number of clicks decreases (negative estimate) non significantly over time for Eistla data. In conclusion, there is not any clear pattern over time for Eistla data.

```
glmNBE1 = glm.nb(clicks ~ mins, link = log, init.theta = 2.861314,
  data = Eistla1)
DateTimeE1 = seq(0, 6000, by = 1)
predglmNBE1 <- predict(glmNBE1, newdata = data.frame(mins = DateTimeE1),
  type = "response")
plot(Eistla1$mins, Eistla1$clicks, xlab = "Time differences in mins",
  ylab = "Number of clicks", main = list("Eistla"))
lines(DateTimeE1, predglmNBE1)
```



Summary of the negative binomial GLM for Eistla data

```
summary(glmNBE1)
##
## Call:
## glm.nb(formula = clicks ~ mins, data = Eistla1, init.theta = 2.861313613,
## link = log)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -2.28679 -0.78546 -0.03119 0.51197 2.64610
##
```

```
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.461e+00 1.368e-01 10.683 <2e-16 ***
## mins      -1.418e-05 4.604e-05 -0.308 0.758
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(2.8613) family taken to be 1)
##
## Null deviance: 188.5 on 149 degrees of freedom
## Residual deviance: 188.4 on 148 degrees of freedom
## AIC: 740
##
## Number of Fisher Scoring iterations: 1
##
##
##           Theta: 2.861
##           Std. Err.: 0.647
##
## 2 x log-likelihood: -733.998
```

Relationship between the number of clicks with depth

The number of clicks counted in 150 samples was used as the response variable, and the depth in meters for each sample was the explanatory variable. To know which model would best fit my data, it was done the same process that was done and showed previously for Freya1 data. Therefore, it will only be shown the model selected.

The glmNBE2 (negative binomial GLM) had a smaller AIC value, so it will be chosen this model to see the relationship between the number of clicks with depth for Eistla data.

Calculation of the theta value required for the negative binomial GLM for Eistla data.

```
# plot(Eistla1$clicks~Eistla1$depth)
Eistla1.nb1 <- glm.nb(Eistla1$clicks ~ Eistla1$depth^4, link = log)
c(theta = Eistla1.nb1$theta, SE = Eistla1.nb1$SE)

## theta SE
## 3.0098693 0.6964728
```

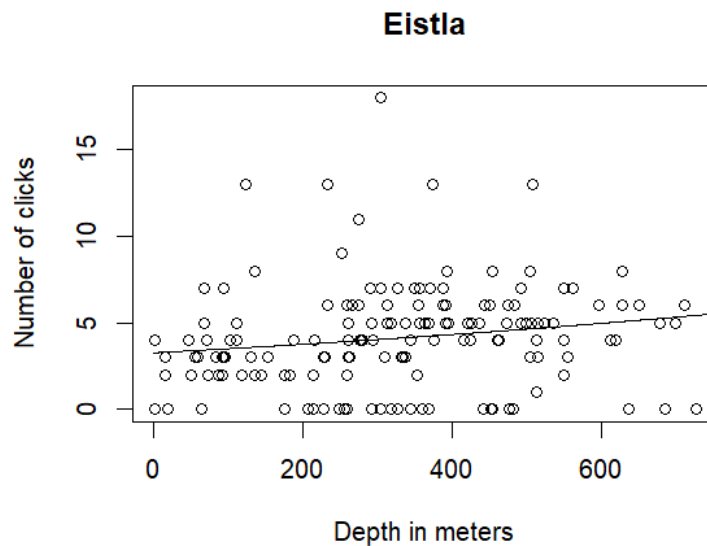
The results from the negative binomial GLM show that the number of clicks increases (positive estimate) significantly (p-value < 0.05) with increasing depth for Eistla data. In conclusion, there is a relationship between the number of clicks with depth for Eistla data.

```
glmNBE2 = glm.nb(clicks ~ depth, link = log, init.theta = 3.0098693,
  data = Eistla1)
depthEistla = seq(0, 900, by = 1)
```

```

predglmNBE2 <- predict(glmNBE2, newdata = data.frame(depth = depthEistla),
  type = "response")
plot(Eistla1$depth, Eistla1$clicks, xlab = "Depth in meters",
  ylab = "Number of clicks", main = list("Eistla"))
lines(depthEistla, predglmNBE2)

```



Summary of the negative binomial GLM for Eistla data

```

summary(glmNBE2)

##
## Call:
## glm.nb(formula = clicks ~ depth, data = Eistla1, init.theta = 3.00986919,
## link = log)
##
## Deviance Residuals:
##   Min       1Q   Median       3Q      Max
## -2.49433 -0.62070 -0.00381  0.38850  2.79987
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.1897491  0.1334152   8.918  <2e-16 ***
## depth        0.0006944  0.0003515   1.976  0.0482 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(3.0099) family taken to be 1)
##

```

```
## Null deviance: 192.61 on 149 degrees of freedom
## Residual deviance: 188.97 on 148 degrees of freedom
## AIC: 736.51
##
## Number of Fisher Scoring iterations: 1
##
##
##      Theta: 3.010
##      Std. Err.: 0.696
##
## 2 x log-likelihood: -730.505
```

Mutti

Seeing how increasing the sample size changes the estimates

The average number of sounds per second for clicks present in the clicking period was calculated for Mutti data (with 150 samples), and the value obtained was 4.2.

The next plots show for each sample size, a sub-sample of k samples out of 150 that were selected, between sample sizes of x_{min} and x_{max} , and the average was computed.

Seeing how increasing the sample size changes the estimates for Mutti. 150 samples seem to be enough to have a good estimate of narwhal cue rates and density since the coefficient of variation of the averages number of clicks is less than 20% (Gomez & Gomez (1984)).

```
nobsM <- length(Mutti1$clicks)
# Different sample sizes
limitM <- 5
nevalM <- limitM:(nobsM - limitM)
# Number of repetitions for each sample size
nrepsM <- 1000
# Object to store the averages
resultadosM <- matrix(NA, ncol = length(nevalM), nrow = nrepsM)

# For each sample size
for (i in 1:length(nevalM)) {
  # For each repetition
  for (j in 1:nrepsM) {
    amostraM <- sample(Mutti1$clicks, size = nevalM[i], replace = TRUE)
    resultadosM[j, i] <- mean(amostraM)
  }
}
par(mfrow = c(2, 2))

plot(nevalM, apply(resultadosM, 2, var), xlim = c(0, 150), xlab = "Different sample sizes",
```

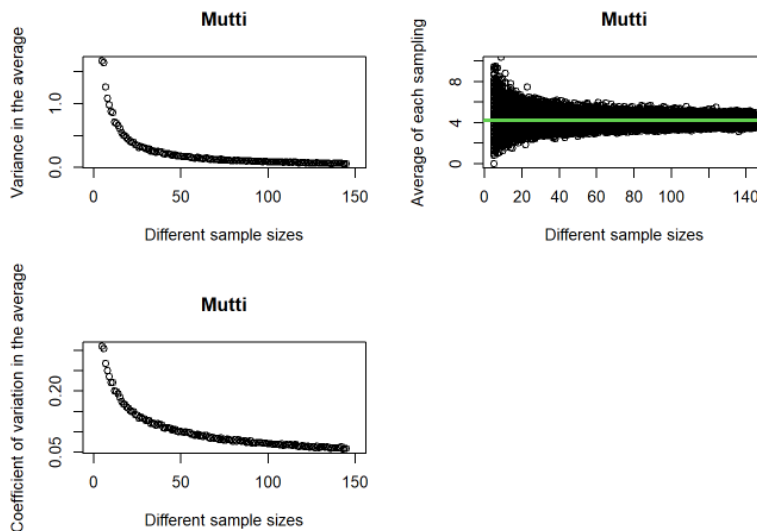
```

ylab = "Variance in the average", main = list("Mutti"))

plot(x = c(5, 145), y = c(0, 10), xlab = "Different sample sizes",
     ylab = "Average of each sampling", main = list("Mutti"),
     type = "n")
for (i in 1:length(nevalM)) {
  points(rep(nevalM[i], nrepsM), resultadosM[, i])
}
abline(h = mean(Mutti1$clicks), col = 3, lwd = 3)

# looking at the CV (coefficient of variation)
plot(nevalM, apply(resultadosM, 2, sd)/apply(resultadosM, 2,
mean), ylab = "Coefficient of variation in the average",
     xlab = "Different sample sizes", xlim = c(0, 150), main = list("Mutti"))

```



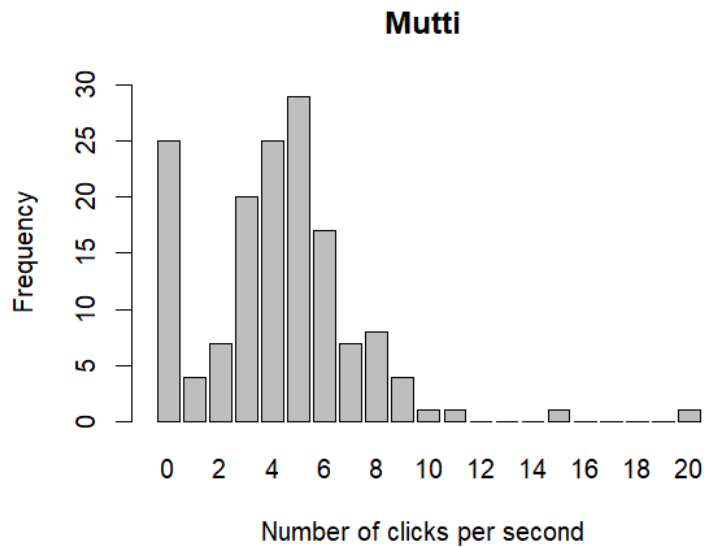
Data exploration

Graph and table of frequencies of the number of clicks counted in Mutti data.

```

maxnM1 <- max(Mutti1$clicks)
tempM1 <- factor(Mutti1$clicks[Mutti1$clicks <= maxnM1], levels = 0:maxnM1,
                labels = 0:maxnM1)
barplot(table(tempM1), xlab = "Number of clicks per second",
         ylab = "Frequency", main = list("Mutti"), ylim = c(0, 30),
         xlim = c(0, 25))

```

```
table(tempM1)
```

```
## tempM1
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
## 25 4 7 20 25 29 17 7 8 4 1 1 0 0 0 1 0 0 0 0 1
```

```
# In Mutti1, 5 clicks/s was the most frequent count (29
# times counted) followed by 0 and 4 clicks/s (both were 25
# times counted).
```

Clicks counted in 14 seconds were turned into 0 clicks/s because more than 20 Clicks/s are not detected on PAM recorders given its low amplitude.

Trends of the number of clicks over time

The number of clicks counted in 150 samples was used as the response variable, and the time differences in minutes between samples was the explanatory variable. To know which model would best fit my data, it was done the same process that was done and showed previously for Freya1 data. Therefore, it will only be shown the model selected.

The glmNBM1 (negative binomial GLM) had a smaller AIC value, so it will be chosen this model to see the Trend of the number of clicks over time for Mutti data.

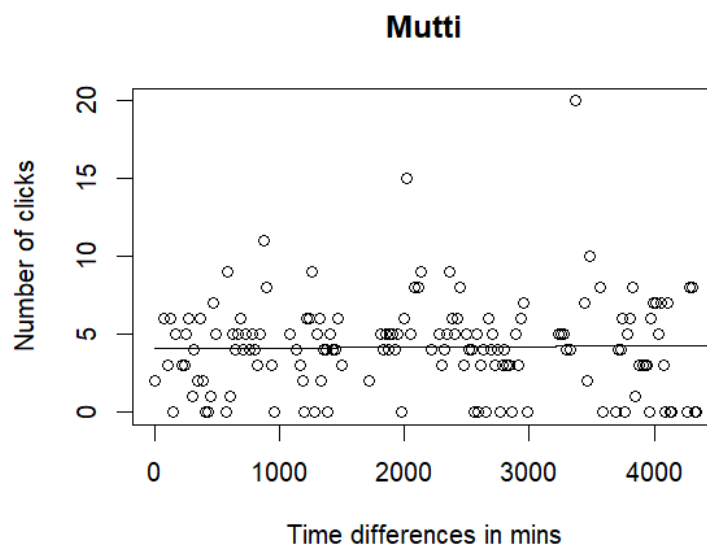
Calculation of the theta value required for the negative binomial GLM for Mutti data.

```
# plot(Mutti1$clicks~Mutti1$mins)
par(mfrow = c(1, 1))
Mutti1.nb <- glm.nb(Mutti1$clicks ~ Mutti1$mins^4, link = log)
c(theta = Mutti1.nb$theta, SE = Mutti1.nb$SE)
```

```
## theta SE
## 3.0709747 0.7228199
```

The results from the negative binomial GLM show that the number of clicks increases (positive estimate) non significantly over time for Mutti data. In conclusion, there is not any clear pattern over time for Mutti data.

```
glmNBM1 = glm.nb(clicks ~ mins, link = log, init.theta = 3.0709747,
  data = Mutti1)
DateTimeM1 = seq(0, 6000, by = 1)
predglmNBM1 <- predict(glmNBM1, newdata = data.frame(mins = DateTimeM1),
  type = "response")
plot(Mutti1$mins, Mutti1$clicks, xlab = "Time differences in mins",
  ylab = "Number of clicks", main = list("Mutti"))
lines(DateTimeM1, predglmNBM1)
```



Summary of the negative binomial GLM for Mutti data

```
summary(glmNBM1)
##
## Call:
## glm.nb(formula = clicks ~ mins, data = Mutti1, init.theta = 3.070974677,
## link = log)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -2.31970 -0.43826 -0.05062 0.49057 2.99439
##
```

```
## Coefficients:
##          Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.411e+00 1.203e-01 11.730 <2e-16 ***
## mins      1.126e-05 4.778e-05 0.236 0.814
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(3.071) family taken to be 1)
##
## Null deviance: 191.44 on 149 degrees of freedom
## Residual deviance: 191.38 on 148 degrees of freedom
## AIC: 740.8
##
## Number of Fisher Scoring iterations: 1
##
##
##          Theta: 3.071
##          Std. Err.: 0.723
##
## 2 x log-likelihood: -734.801
```

Relationship between the number of clicks with depth

The number of clicks counted in 150 samples was used as the response variable, and the depth in meters for each sample was the explanatory variable. To know which model would best fit my data, it was done the same process that was done and showed previously for Freya1 data. Therefore, it will only be shown the model selected.

The glmNBM2 (negative binomial GLM) had a smaller AIC value, so it will be chosen this model to see the relationship between the number of clicks with depth for Mutti data.

Calculation of the theta value required for the negative binomial GLM for Mutti data.

```
# plot(Mutti1$clicks~Mutti1$depth)
Mutti1.nb1 <- glm.nb(Mutti1$clicks ~ Mutti1$depth^4, link = log)
c(theta = Mutti1.nb1$theta, SE = Mutti1.nb1$SE)

##   theta    SE
## 3.6338546 0.9313743
```

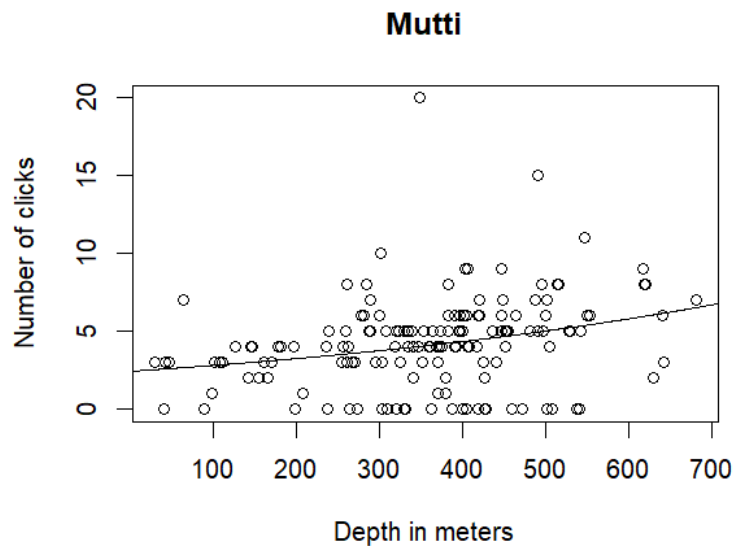
The results from the negative binomial GLM show that the number of clicks increases (positive estimate) significantly (p-value < 0.05) with increasing depth for Mutti data. In conclusion, there is a relationship between the number of clicks with depth for Mutti data.

```
glmNBM2 = glm.nb(clicks ~ depth, link = log, init.theta = 3.6338546,
  data = Mutti1)
depthMutti = seq(0, 900, by = 1)
```

```

predglmNBM2 <- predict(glmNBM2, newdata = data.frame(depth = depthMutti),
  type = "response")
plot(Mutti1$depth, Mutti1$clicks, xlab = "Depth in meters", ylab = "Number of clicks",
  main = list("Mutti"))
lines(depthMutti, predglmNBM2)

```



Summary of the negative binomial GLM for Mutti data

```

summary(glmNBM2)

##
## Call:
## glm.nb(formula = clicks ~ depth, data = Mutti1, init.theta = 3.633854578,
## link = log)
##
## Deviance Residuals:
##  Min    1Q  Median    3Q   Max
## -2.5656 -0.4910  0.0630  0.4134  3.2779
##
## Coefficients:
##      Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.8994850 0.1740695  5.167 2.37e-07 ***
## depth      0.0014399 0.0004412  3.264 0.0011 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(3.6339) family taken to be 1)
##

```

```
## Null deviance: 204.78 on 149 degrees of freedom
## Residual deviance: 194.06 on 148 degrees of freedom
## AIC: 730.63
##
## Number of Fisher Scoring iterations: 1
##
##
##      Theta: 3.634
##      Std. Err.: 0.931
##
## 2 x log-likelihood: -724.625
```

Thora

Seeing how increasing the sample size changes the estimates

The average number of sounds per second for clicks present in the clicking period was calculated for Thora data (with 150 samples), and the value obtained was 4.3333333.

The next plots show for each sample size, a sub-sample of k samples out of 150 that were selected, between sample sizes of x_{min} and x_{max} , and the average was computed.

Seeing how increasing the sample size changes the estimates for Thora. 150 samples seem to be enough to have a good estimate of narwhal cue rates and density since the coefficient of variation of the averages number of clicks is less than 20% (Gomez & Gomez (1984)).

```
nobsT <- length(Thora1$clicks)
# Different sample sizes
limitT <- 5
nevalT <- limitT:(nobsT - limitT)
# Number of repetitions for each sample size
nrepsT <- 1000
# Object to store the averages
resultadosT <- matrix(NA, ncol = length(nevalT), nrow = nrepsT)

# For each sample size
for (i in 1:length(nevalT)) {
  # For each repetition
  for (j in 1:nrepsT) {
    amostraT <- sample(Thora1$clicks, size = nevalT[i], replace = TRUE)
    resultadosT[j, i] <- mean(amostraT)
  }
}
par(mfrow = c(2, 2))

plot(nevalT, apply(resultadosT, 2, var), xlim = c(0, 150), xlab = "Different sample sizes",
```

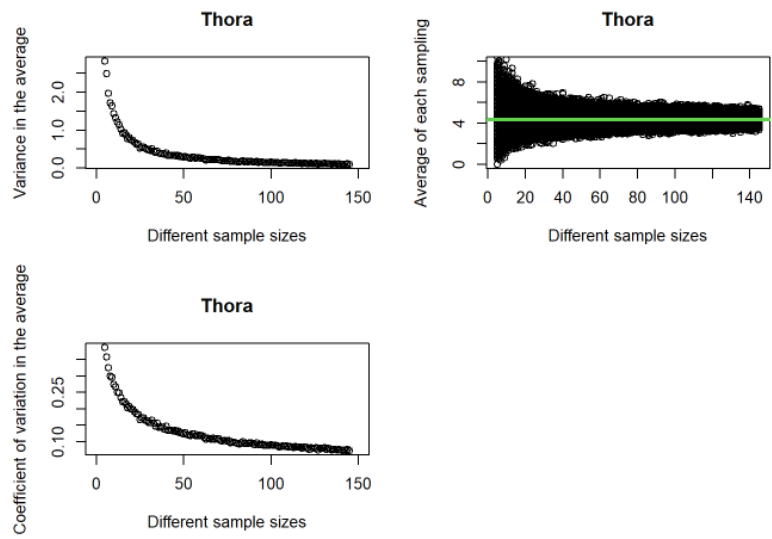
```

ylab = "Variance in the average", main = list("Thora"))

plot(x = c(5, 145), y = c(0, 10), xlab = "Different sample sizes",
     ylab = "Average of each sampling", main = list("Thora"),
     type = "n")
for (i in 1:length(nevalT)) {
  points(rep(nevalT[i], nrepsT), resultadosT[, i])
}
abline(h = mean(Thora1$clicks), col = 3, lwd = 3)

# looking at the CV (coefficient of variation)
plot(nevalT, apply(resultadosT, 2, sd)/apply(resultadosT, 2,
mean), ylab = "Coefficient of variation in the average",
     xlab = "Different sample sizes", xlim = c(0, 150), main = list("Thora"))

```



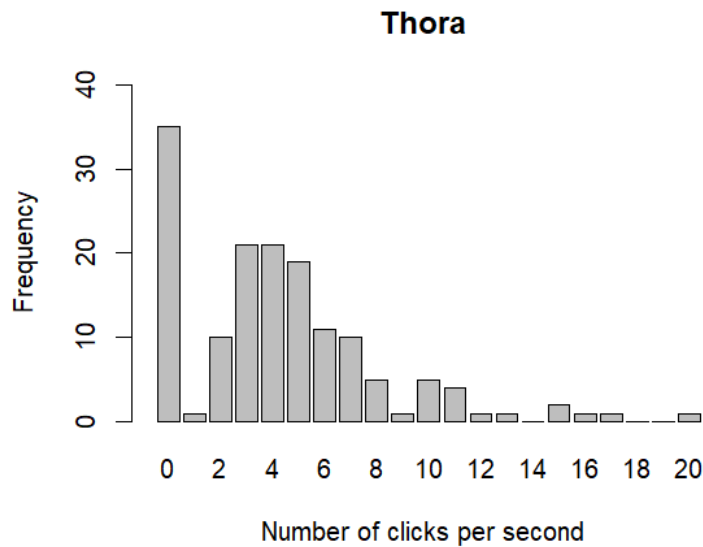
Data exploration

Graph and table of frequencies of the number of clicks counted in Thora data.

```

maxnT1 <- max(Thora1$clicks)
tempT1 <- factor(Thora1$clicks[Thora1$clicks <= maxnT1], levels = 0:maxnT1,
  labels = 0:maxnT1)
barplot(table(tempT1), xlab = "Number of clicks per second",
  ylab = "Frequency", main = list("Thora"), ylim = c(0, 40),
  xlim = c(0, 25))

```



```
table(tempT1)
## tempT1
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
## 35 1 10 21 21 19 11 10 5 1 5 4 1 1 0 2 1 1 0 0 1
# In Thora1, 0 clicks/s was the most frequent count (35
# times counted) followed by 3 and 4 clicks/s (both were 21
# times counted).
```

Clicks counted in 14 seconds were turned into 0 clicks/s because more than 20 Clicks/s are not detected on PAM recorders given its low amplitude.

Trends of the number of clicks over time

The number of clicks counted in 150 samples was used as the response variable, and the time differences in minutes between samples was the explanatory variable. To know which model would best fit my data, it was done the same process that was done and showed previously for Freya1 data. Therefore, it will only be shown the model selected.

The glmNBT1 (negative binomial GLM) had a smaller AIC value, so it will be chosen this model to see the Trend of the number of clicks over time for Thora data.

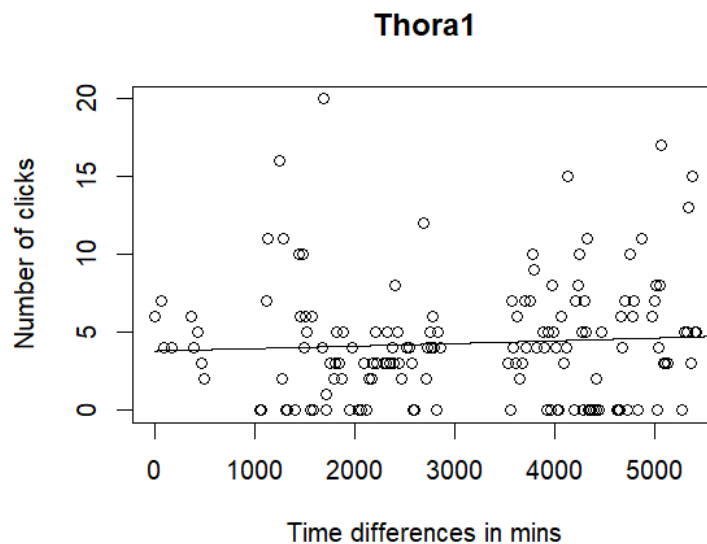
Calculation of the theta value required for the negative binomial GLM for Thora data.

```
# plot(Thora1$clicks~Thora1$mins)
par(mfrow = c(1, 1))
Thora1.nb <- glm.nb(Thora1$clicks ~ Thora1$mins^4, link = log)
c(theta = Thora1.nb$theta, SE = Thora1.nb$SE)
```

```
## theta SE
## 1.361199 0.242646
```

The results from the negative binomial GLM show that the number of clicks increases (positive estimate) non significantly over time for Thora data. In conclusion, there is not any clear pattern over time for Thora data.

```
glmNBT1 = glm.nb(clicks ~ mins, link = log, init.theta = 1.361199,
  data = Thora1)
DateTimeT1 = seq(0, 6000, by = 1)
predglmNBT1 <- predict(glmNBT1, newdata = data.frame(mins = DateTimeT1),
  type = "response")
plot(Thora1$mins, Thora1$clicks, xlab = "Time differences in mins",
  ylab = "Number of clicks", main = list("Thora1"))
lines(DateTimeT1, predglmNBT1)
```



Summary of the negative binomial GLM for Thora data

```
summary(glmNBT1)
##
## Call:
## glm.nb(formula = clicks ~ mins, data = Thora1, init.theta = 1.361198685,
## link = log)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -2.0175 -0.6804 -0.0827 0.3617 2.2663
##
```



```
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.343e+00 1.901e-01 7.065 1.61e-12 ***
## mins      3.919e-05 5.502e-05 0.712 0.476
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(1.3612) family taken to be 1)
##
## Null deviance: 183.59 on 149 degrees of freedom
## Residual deviance: 183.04 on 148 degrees of freedom
## AIC: 774.62
##
## Number of Fisher Scoring iterations: 1
##
##
##           Theta: 1.361
##           Std. Err.: 0.243
##
## 2 x log-likelihood: -768.618
```

Relationship between the number of clicks with depth

The number of clicks counted in 150 samples was used as the response variable, and the depth in meters for each sample was the explanatory variable. To know which model would best fit my data, it was done the same process that was done and showed previously for Freya1 data. Therefore, it will only be shown the model selected.

The glmNBT2 (negative binomial GLM) had a smaller AIC value, so it will be chosen this model to see the relationship between the number of clicks with depth for Thora data.

Calculation of the theta value required for the negative binomial GLM for Thora data.

```
# plot(Thora1$clicks~Thora1$depth)
Thora1.nb1 <- glm.nb(Thora1$clicks ~ Thora1$depth^4, link = log)
c(theta = Thora1.nb1$theta, SE = Thora1.nb1$SE)

## theta    SE
## 1.3578225 0.2417603
```

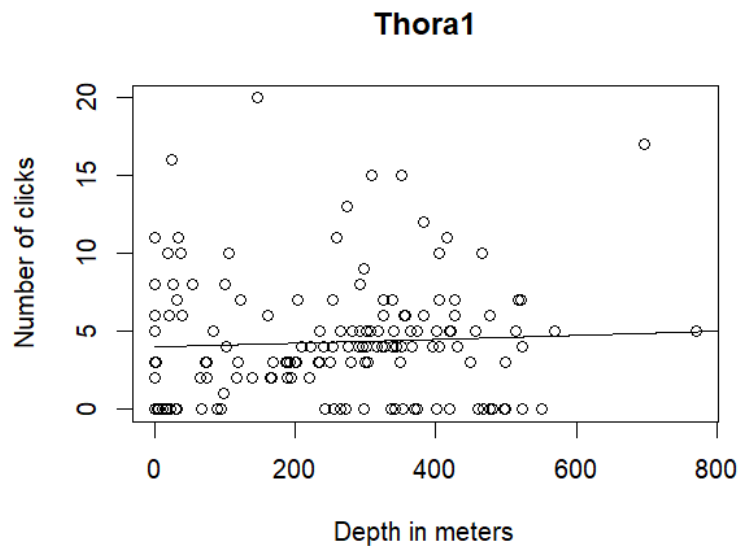
The results from the negative binomial GLM show that the number of clicks increases (positive estimate) non significantly with increasing depth for Thora data. In conclusion, there is not a relationship between the number of clicks with depth for Thora data.

```
glmNBT2 = glm.nb(clicks ~ depth, link = log, init.theta = 1.3578225,
  data = Thora1)
depthThora = seq(0, 900, by = 1)
```

```

predglmNBT2 <- predict(glmNBT2, newdata = data.frame(depth = depthThora),
  type = "response")
plot(Thora1$depth, Thora1$clicks, xlab = "Depth in meters", ylab = "Number of clicks",
  main = list("Thora1"))
lines(depthThora, predglmNBT2)

```



Summary of the negative binomial GLM for Thora data

```

summary(glmNBT2)

##
## Call:
## glm.nb(formula = clicks ~ depth, data = Thora1, init.theta = 1.357822455,
## link = log)
##
## Deviance Residuals:
##  Min    1Q  Median    3Q   Max
## -2.0153 -0.6560 -0.1027  0.3311  2.2184
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.3941534  0.1452765  9.597 <2e-16 ***
## depth      0.0002782  0.0004707  0.591  0.554
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(1.3578) family taken to be 1)
##

```

```
## Null deviance: 183.34 on 149 degrees of freedom
## Residual deviance: 182.98 on 148 degrees of freedom
## AIC: 774.81
##
## Number of Fisher Scoring iterations: 1
##
##
##      Theta: 1.358
##      Std. Err.: 0.242
##
## 2 x log-likelihood: -768.812
```

Frida

Seeing how increasing the sample size changes the estimates

The average number of sounds per second for clicks present in the clicking period was calculated for Frida data (with 149 samples), and the value obtained was 3.8120805.

The next plots show for each sample size, a sub-sample of k samples out of 149 that were selected, between sample sizes of x_{min} and x_{max} , and the average was computed.

Seeing how increasing the sample size changes the estimates for Frida. 149 samples seem to be enough to have a good estimate of narwhal cue rates and density since the coefficient of variation of the averages number of clicks is less than 20% (Gomez & Gomez (1984)).

```
nobsF <- length(Frida149$clicks)
# Different sample sizes
limitF <- 5
nevalF <- limitF:(nobsF - limitF)
# Number of repetitions for each sample size
nrepsF <- 1000
# Object to store the averages
resultadosF <- matrix(NA, ncol = length(nevalF), nrow = nrepsF)

# For each sample size
for (i in 1:length(nevalF)) {
  # For each repetition
  for (j in 1:nrepsF) {
    amostraF <- sample(Frida149$clicks, size = nevalF[i],
                      replace = TRUE)
    resultadosF[j, i] <- mean(amostraF)
  }
}
par(mfrow = c(2, 2))
```

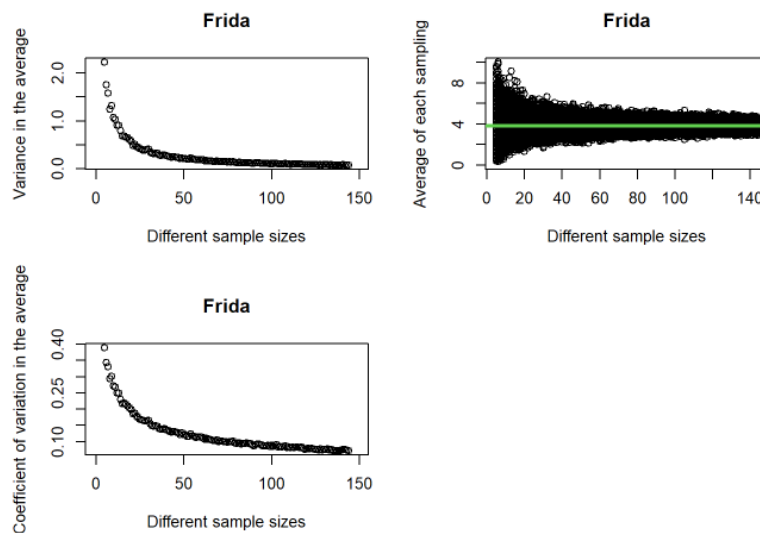
```
plot(nevalF, apply(resultadosF, 2, var), xlim = c(0, 149), xlab = "Different sample sizes",
     ylab = "Variance in the average", main = list("Frida"))
```

```
plot(x = c(5, 144), y = c(0, 10), xlab = "Different sample sizes",
     ylab = "Average of each sampling", main = list("Frida"),
     type = "n")
```

```
for (i in 1:length(nevalF)) {
  points(rep(nevalF[i], nrepsF), resultadosF[, i])
}
abline(h = mean(Frida149$clicks), col = 3, lwd = 3)
```

looking at the CV (coefficient of variation)

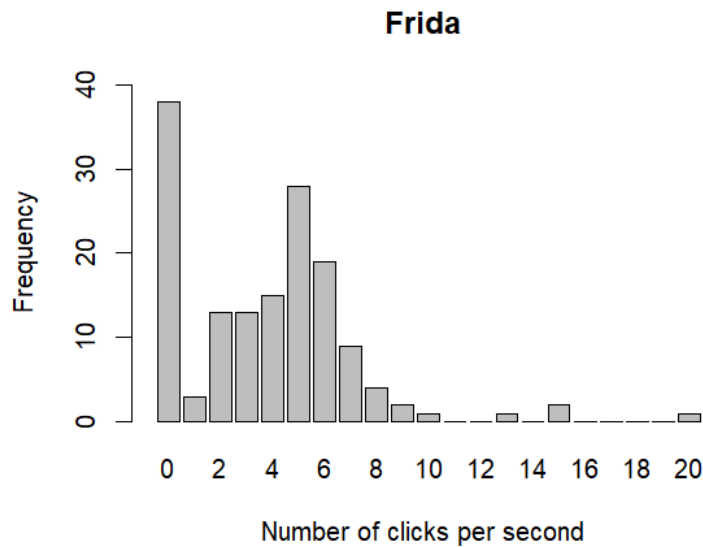
```
plot(nevalF, apply(resultadosF, 2, sd)/apply(resultadosF, 2,
mean), ylab = "Coefficient of variation in the average",
     xlab = "Different sample sizes", xlim = c(0, 149), main = list("Frida"))
```



Data exploration

Graph and table of frequencies of the number of clicks counted in Frida data.

```
maxnF <- max(Frida149$clicks)
tempF <- factor(Frida149$clicks[Frida149$clicks <= maxnF], levels = 0:maxnF,
               labels = 0:maxnF)
barplot(table(tempF), xlab = "Number of clicks per second", ylab = "Frequency",
         main = list("Frida"), ylim = c(0, 40), xlim = c(0, 25))
```



```
table(tempF)
## tempF
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
## 38 3 13 13 15 28 19 9 4 2 1 0 0 1 0 2 0 0 0 0 1
# In Frida, 0 clicks/s was the most frequent count (38
# times counted) followed by 5 clicks/s (28 times counted).
```

Clicks counted in 18 seconds were turned into 0 clicks/s because more than 20 Clicks/s are not detected on PAM recorders given its low amplitude.

Trends of the number of clicks over time

The number of clicks counted in 149 samples was used as the response variable, and the time differences in minutes between samples was the explanatory variable. To know which model would best fit my data, it was done the same process that was done and showed previously for Freya1 data. Therefore, it will only be shown the model selected.

The glmNBF (negative binomial GLM) had a smaller AIC value, so it will be chosen this model to see the Trend of the number of clicks over time for Frida data.

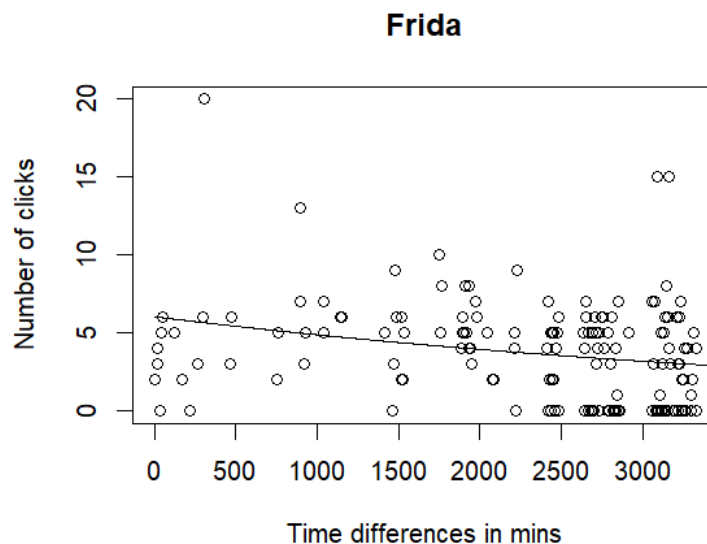
Calculation of the theta value required for the negative binomial GLM for Frida data.

```
# plot(Frida149$clicks~Frida149$mins)
par(mfrow = c(1, 1))
Frida.nb <- glm.nb(Frida149$clicks ~ Frida149$mins^4, link = log)
c(theta = Frida.nb$theta, SE = Frida.nb$SE)
```

```
## theta SE
## 1.5583435 0.3136075
```

The results from the negative binomial GLM show that the number of clicks decreases significantly (p-value < 0.05) over time for Frida data. In conclusion, there seems to be a pattern over time for Frida data.

```
glmNBF = glm.nb(clicks ~ mins, link = log, init.theta = 1.5583435,
  data = Frida149)
DateTimeF = seq(0, 4000, by = 1)
predglmNBF <- predict(glmNBF, newdata = data.frame(mins = DateTimeF),
  type = "response")
plot(Frida149$mins, Frida149$clicks, xlab = "Time differences in mins",
  ylab = "Number of clicks", main = list("Frida"))
lines(DateTimeF, predglmNBF)
```



Summary of the negative binomial GLM for Frida data.

```
summary(glmNBF)
##
## Call:
## glm.nb(formula = clicks ~ mins, data = Frida149, init.theta = 1.558343652,
## link = log)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -2.22018 -1.82816 0.00093 0.42688 2.29891
##
## Coefficients:
```

```

##          Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.802e+00 1.972e-01  9.14 < 2e-16 ***
## mins      -2.115e-04 8.105e-05  -2.61 0.00906 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(1.5583) family taken to be 1)
##
## Null deviance: 191.98 on 148 degrees of freedom
## Residual deviance: 185.49 on 147 degrees of freedom
## AIC: 728.98
##
## Number of Fisher Scoring iterations: 1
##
##
##      Theta: 1.558
##      Std. Err.: 0.314
##
## 2 x log-likelihood: -722.981

```

Relationship between the number of clicks with depth

The number of clicks counted in 149 samples was used as the response variable, and the depth in meters for each sample was the explanatory variable. To know which model would best fit my data, it was done the same process that was done and showed previously for Freya1 data. Therefore, it will only be shown the model selected.

The glmNBFrida (negative binomial GLM) had a smaller AIC value, so it will be chosen this model to see the relationship between the number of clicks with depth for Frida data.

Calculation of the theta value required for the negative binomial GLM for Frida data.

```

# plot(Frida149$clicks~Frida149$depth)
Frida.nb1 <- glm.nb(Frida149$clicks ~ Frida149$depth^4, link = log)
c(theta = Frida.nb1$theta, SE = Frida.nb1$SE)

##  theta    SE
## 1.5688881 0.3172617

```

The results from the negative binomial GLM show that the number of clicks increases (positive estimate) significantly (p-value < 0.05) with increasing depth for Frida data. In conclusion, there is a relationship between the number of clicks with depth for Frida data.

```

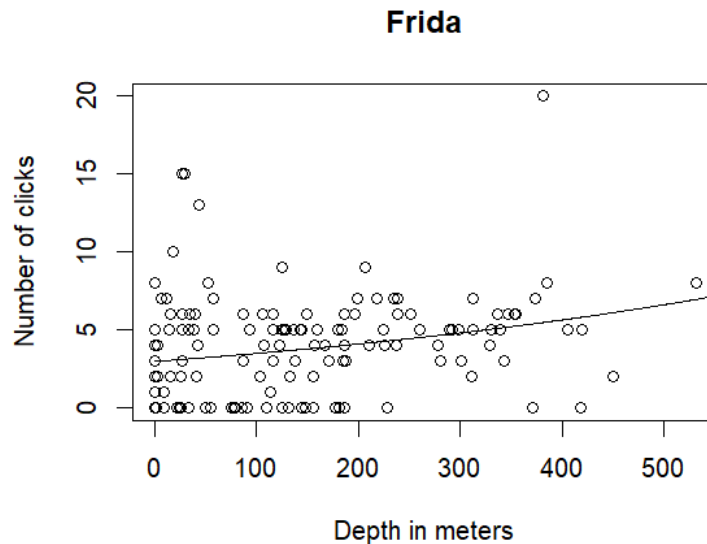
glmNBFrida = glm.nb(clicks ~ depth, link = log, init.theta = 1.5688881,
  data = Frida149)
depthFrida = seq(0, 600, by = 1)
predglmNBFrida <- predict(glmNBFrida, newdata = data.frame(depth = depthFrida),

```

```

type = "response")
plot(Frida149$depth, Frida149$clicks, xlab = "Depth in meters",
     ylab = "Number of clicks", main = list("Frida"))
lines(depthFrida, predglmNBFrida)

```



Summary of the negative binomial GLM for Frida data

```

summary(glmNBFrida)

##
## Call:
## glm.nb(formula = clicks ~ depth, data = Frida149, init.theta = 1.568888077,
## link = log)
##
## Deviance Residuals:
##  Min     1Q   Median     3Q    Max
## -2.20661 -1.83245 -0.00993  0.40684  2.28798
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.1004968  0.1180429   9.323 < 2e-16 ***
## depth        0.0015852  0.0006146   2.579  0.00991 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(1.5689) family taken to be 1)
##
## Null deviance: 192.63 on 148 degrees of freedom

```



```

## Residual deviance: 185.80 on 147 degrees of freedom
## AIC: 728.67
##
## Number of Fisher Scoring iterations: 1
##
##
##      Theta: 1.569
##      Std. Err.: 0.317
##
## 2 x log-likelihood: -722.675

```

Jonas

Seeing how increasing the sample size changes the estimates

The average number of sounds per second for clicks present in the clicking period was calculated for Jonas data (with 200 samples), and the value obtained was 4.095.

The next plots show for each sample size, a sub-sample of k samples out of 200 that were selected, between sample sizes of x_{min} and x_{max} , and the average was computed.

Seeing how increasing the sample size changes the estimates for Jonas. 200 samples seem to be enough to have a good estimate of narwhal cue rates and density since the coefficient of variation of the averages number of clicks is less than 20% (Gomez & Gomez (1984)).

```

nobsJ <- length(Jonas$clicks)
# Different sample sizes
limitJ <- 5
nevalJ <- limitJ:(nobsJ - limitJ)
# Number of repetitions for each sample size
nrepsJ <- 1000
# Object to store the averages
resultadosJ <- matrix(NA, ncol = length(nevalJ), nrow = nrepsJ)

# For each sample size
for (i in 1:length(nevalJ)) {
  # For each repetition
  for (j in 1:nrepsJ) {
    amostraJ <- sample(Jonas$clicks, size = nevalJ[i], replace = TRUE)
    resultadosJ[j, i] <- mean(amostraJ)
  }
}
par(mfrow = c(2, 2))

plot(nevalJ, apply(resultadosJ, 2, var), xlim = c(0, 200), xlab = "Different sample sizes",
      ylab = "Variance in the average", main = list("Jonas"))

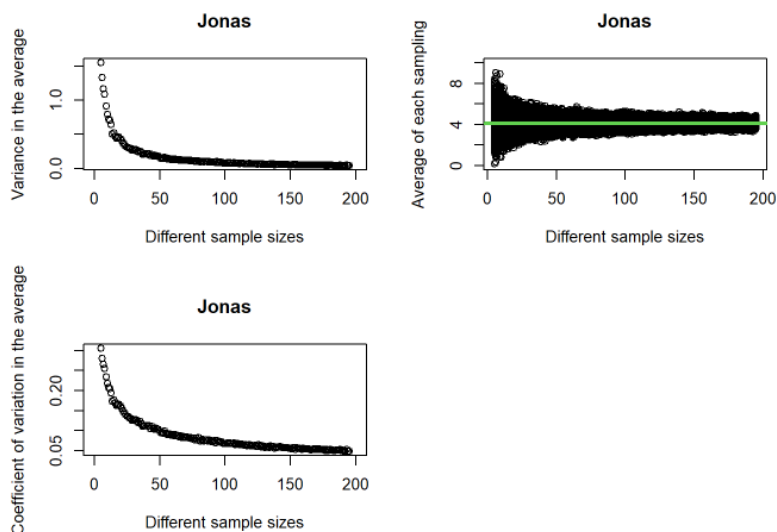
```

```

plot(x = c(5, 195), y = c(0, 10), xlab = "Different sample sizes",
     ylab = "Average of each sampling", main = list("Jonas"),
     type = "n")
for (i in 1:length(nevalJ)) {
  points(rep(nevalJ[i], nrepsJ), resultadosJ[, i])
}
abline(h = mean(Jonas$clicks), col = 3, lwd = 3)

# looking at the CV (coefficient of variation)
plot(nevalJ, apply(resultadosJ, 2, sd)/apply(resultadosJ, 2,
mean), ylab = "Coefficient of variation in the average",
     xlab = "Different sample sizes", xlim = c(0, 200), main = list("Jonas"))

```



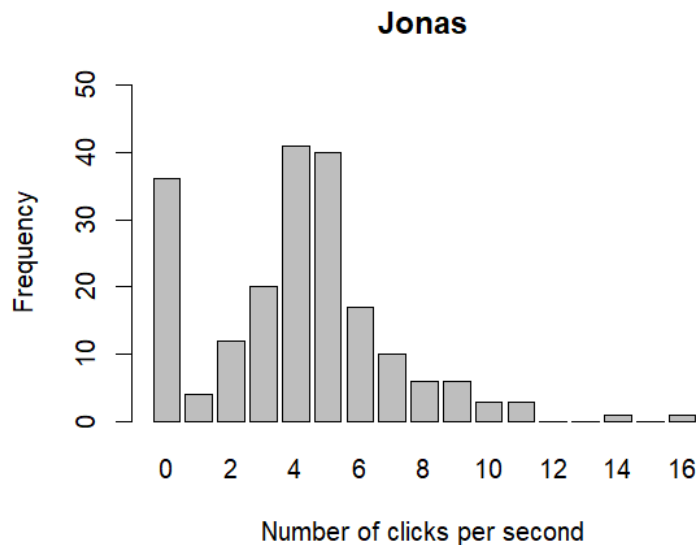
Data exploration

Graph and table of frequencies of the number of clicks counted in Jonas data.

```

maxnJ <- max(Jonas$clicks)
tempJ <- factor(Jonas$clicks[Jonas$clicks <= maxnJ], levels = 0:maxnJ,
               labels = 0:maxnJ)
barplot(table(tempJ), xlab = "Number of clicks per second", ylab = "Frequency",
         main = list("Jonas"), ylim = c(0, 50))

```



```
table(tempJ)
```

```
## tempJ
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
## 36 4 12 20 41 40 17 10 6 6 3 3 0 0 1 0 1
```

```
# In Jonas data, 4 and 5 clicks/s were the most frequent
# count (41 and 40 times counted, respectively) followed by
# 0 clicks/s (36 times counted).
```

Clicks counted in 23 seconds were turned into 0 clicks/s because more than 20 Clicks/s are not detected on PAM recorders given its low amplitude.

Trends of the number of clicks over time

The number of clicks counted in 200 samples was used as the response variable, and the time differences in minutes between samples was the explanatory variable. To know which model would best fit my data, it was done the same process that was done and showed previously for Freya1 data. Therefore, it will only be shown the model selected.

The glmNBJ (negative binomial GLM) had a smaller AIC value, so it will be chosen this model to see the Trend of the number of clicks over time for Jonas data.

Calculation of the theta value required for the negative binomial GLM for Jonas data.

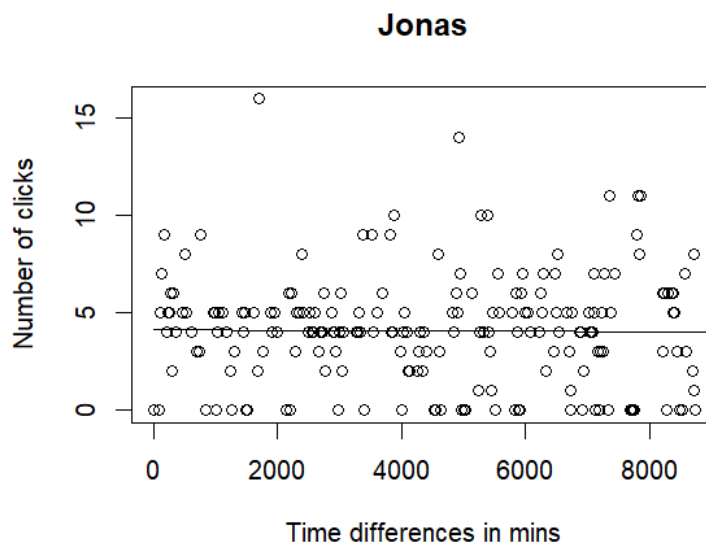
```
Jonas.nb <- glm.nb(Jonas$clicks ~ Jonas$mins^4, link = log)
c(theta = Jonas.nb$theta, SE = Jonas.nb$SE)
```

```
## theta SE
## 2.9701185 0.6159317
```

The results from the negative binomial GLM show that the number of clicks decreases (negative estimate) non significantly over time for Jonas data. In conclusion, there is not any clear pattern over time for Jonas data.

```
# plot(Jonas$clicks~Jonas$mins)
par(mfrow = c(1, 1))

glmNBJ = glm.nb(clicks ~ mins, link = log, init.theta = 2.9701185,
  data = Jonas)
DateTimeJ = seq(0, 9000, by = 1)
predglmNBJ <- predict(glmNBJ, newdata = data.frame(mins = DateTimeJ),
  type = "response")
plot(Jonas$mins, Jonas$clicks, xlab = "Time differences in mins",
  ylab = "Number of clicks", main = list("Jonas"))
lines(DateTimeJ, predglmNBJ)
```



Summary of the negative binomial GLM for Jonas data

```
summary(glmNBJ)

##
## Call:
## glm.nb(formula = clicks ~ mins, data = Jonas, init.theta = 2.970118456,
## link = log)
##
## Deviance Residuals:
##  Min     1Q   Median     3Q    Max
## -2.27874 -0.78587 -0.02994  0.28627  2.46104
```

```
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.423e+00 1.081e-01 13.162 <2e-16 ***
## mins      -2.950e-06 2.118e-05 -0.139  0.889
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(2.9701) family taken to be 1)
##
## Null deviance: 258.19 on 199 degrees of freedom
## Residual deviance: 258.17 on 198 degrees of freedom
## AIC: 980.03
##
## Number of Fisher Scoring iterations: 1
##
##
##           Theta: 2.970
##           Std. Err.: 0.616
##
## 2 x log-likelihood: -974.031
```

Relationship between the number of clicks with depth

The number of clicks counted in 200 samples was used as the response variable, and the depth in meters for each sample was the explanatory variable. To know which model would best fit my data, it was done the same process that was done and showed previously for Freya1 data. Therefore, it will only be shown the model selected.

The glmNBJ1 (negative binomial GLM) had a smaller AIC value, so it will be chosen this model to see the relationship between the number of clicks with depth for Jonas data.

Calculation of the theta value required for the negative binomial GLM for Jonas data.

```
Jonas.nb1 <- glm.nb(Jonas$clicks ~ Jonas$depth^4, link = log)
c(theta = Jonas.nb1$theta, SE = Jonas.nb1$SE)
```

```
## theta SE
## 3.1720641 0.6804949
```

The results from the negative binomial GLM show that the number of clicks increases (positive estimate) significantly (p-value < 0.05) with increasing depth for Jonas data. In conclusion, there is a relationship between the number of clicks with depth for Jonas data.

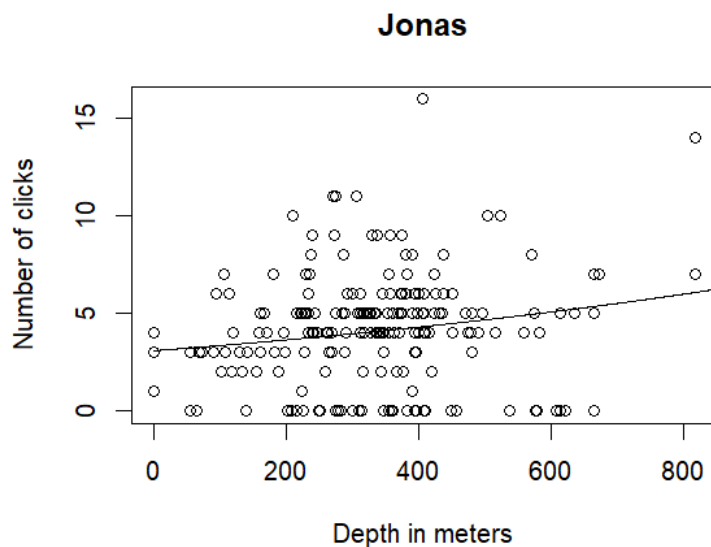
```
# plot(Jonas$clicks~Jonas$depth)
```

```
glmNBJ1 = glm.nb(clicks ~ depth, link = log, init.theta = 3.1720641,
```

```

data = Jonas)
depthJonas = seq(0, 900, by = 1)
predglmNB1 <- predict(glmNB1, newdata = data.frame(depth = depthJonas),
  type = "response")
plot(Jonas$depth, Jonas$clicks, xlab = "Depth in meters", ylab = "Number of clicks",
  main = list("Jonas"))
lines(depthJonas, predglmNB1)

```



Summary of the negative binomial GLM for Jonas data

```

summary(glmNB1)

##
## Call:
## glm.nb(formula = clicks ~ depth, data = Jonas, init.theta = 3.172063833,
## link = log)
##
## Deviance Residuals:
##   Min       1Q   Median       3Q      Max
## -2.50748 -0.60817 -0.00985  0.40327  2.41066
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.1331436  0.1289260  8.789  <2e-16 ***
## depth        0.0008245  0.0003514  2.346  0.019 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##

```

```
## (Dispersion parameter for Negative Binomial(3.1721) family taken to be 1)
##
## Null deviance: 265.14 on 199 degrees of freedom
## Residual deviance: 259.74 on 198 degrees of freedom
## AIC: 974.75
##
## Number of Fisher Scoring iterations: 1
##
##
## Theta: 3.172
## Std. Err.: 0.680
##
## 2 x log-likelihood: -968.751
```

Mara

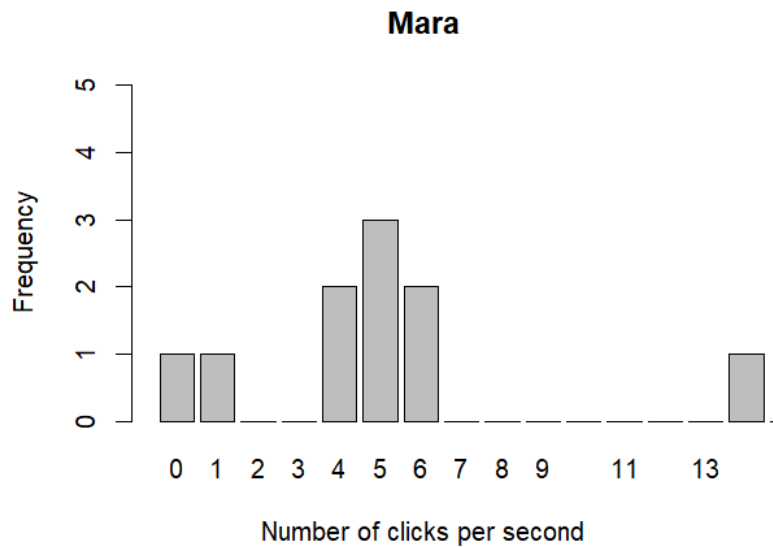
The average number of sounds per second for clicks present in the clicking period was calculated for Mara data (with 10 samples), and the value obtained was 5.

A power reset on Mara record lead to just 1h of clicking and buzzing recorded (Blackwell *et al.* (2018)), so only 10 samples were selected for Mara data. Because of her small sample size, the analysis where it is seen how increasing the sample size changes the estimates were not done.

Data exploration

Graph and table of frequencies of the number of clicks counted in Mara data.

```
maxnMa <- max(Mara1$clicks)
tempMa <- factor(Mara1$clicks[Mara1$clicks <= maxnM1], levels = 0:maxnM1,
  labels = 0:maxnM1)
barplot(table(tempMa), ylim = c(0, 5), xlim = c(0, 16), xlab = "Number of clicks per second",
  ylab = "Frequency", main = list("Mara"))
```



```
table(tempMa)

## tempMa
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
## 1 1 0 0 2 3 2 0 0 0 0 0 0 0 1 0 0 0 0 0 0

# In Mara1, 5 clicks/s was the most frequent count (3 times
# counted) followed by 4 and 6 clicks/s (both were 2 times
# counted).
```

Since Mara did not have samples with more than 20 Clicks/s, it was not necessary to turn any selected seconds into 0 clicks/s.

Trends of the number of clicks over time

The number of clicks counted in 10 samples was used as the response variable, and the time differences in minutes between samples was the explanatory variable. To know which model would best fit my data, it was done the same process that was done and showed previously for Freya1 data. Therefore, it will only be shown the model selected.

The gamMara (poisson GAM) had a smaller AIC value, so it will be chosen this model to see the Trend of the number of clicks over time for Mara data.

The results from the poisson GAM show that the number of clicks increases and then decreases significantly (p-value < 0.05) over time for Mara data. In conclusion, there seems to be a pattern over time for Mara data.

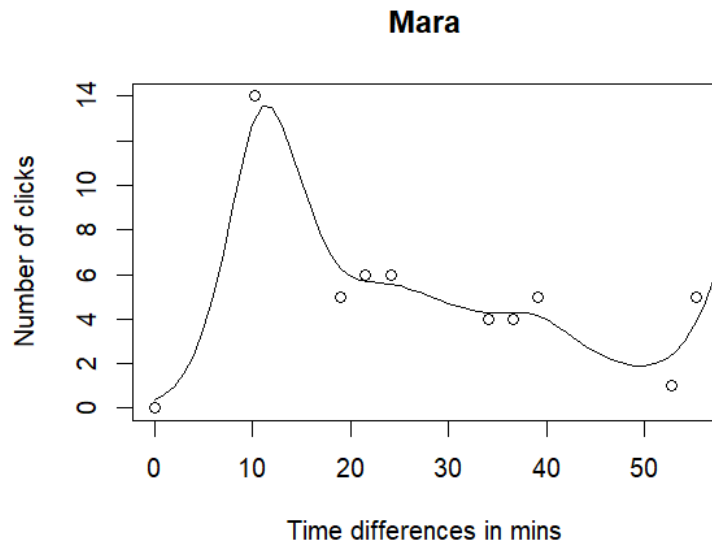
```
gamMara = gam(clicks ~ s(mins), family = poisson, link = "log",
  data = Mara1)
DateTimeMara = seq(0, 60, by = 1)
predgamMara <- predict(gamMara, newdata = data.frame(mins = DateTimeMara),
```



```

type = "response")
plot(Mara1$mins, Mara1$clicks, xlab = "Time differences in mins",
     ylab = "Number of clicks", main = list("Mara"))
lines(DateTimeMara, predgamMara)

```



Summary of the poisson GAM for Mara data

```

summary(gamMara)
##
## Family: poisson
## Link function: log
##
## Formula:
## clicks ~ s(mins)
##
## Parametric coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.3459   0.2041   6.594 4.29e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df Chi.sq p-value
## s(mins) 5.626  6.663 16.87 0.0144 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##

```

```
## R-sq.(adj) = 0.851 Deviance explained = 90%
## UBRE = 0.58986 Scale est. = 1 n = 10
```

Relationship between the number of clicks with depth

The number of clicks counted in 10 samples was used as the response variable, and the depth in meters for each sample was the explanatory variable. To know which model would best fit my data, it was done the same process that was done and showed previously for Freya1 data. Therefore, it will only be shown the model selected.

The glmNBMara (negative binomial GLM) had a smaller AIC value, so it will be chosen this model to see the relationship between the number of clicks with depth for Mara data.

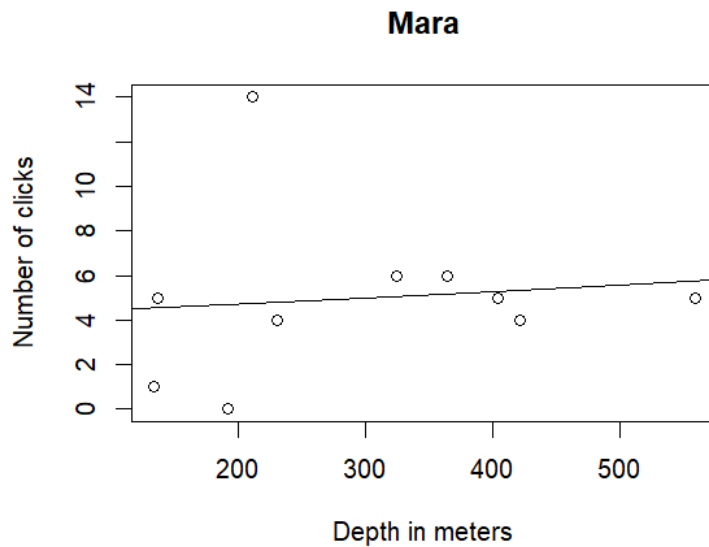
Calculation of the theta value required for the negative binomial GLM for Mara data.

```
Mara.nb1 <- glm.nb(Mara1$clicks ~ Mara1$depth^4, link = log)
c(theta = Mara.nb1$theta, SE = Mara.nb1$SE)

## theta SE
## 3.171334 2.551907
```

The results from the negative binomial GLM show that the number of clicks increases (positive estimate) non significantly with increasing depth for Mara data. In conclusion, there is not a relationship between the number of clicks with depth for Mara data.

```
glmNBMara = glm.nb(clicks ~ depth, link = log, init.theta = 3.171334,
  data = Mara1)
depthMara = seq(0, 600, by = 1)
predglmNBMara <- predict(glmNBMara, newdata = data.frame(depth = depthMara),
  type = "response")
plot(Mara1$depth, Mara1$clicks, xlab = "Depth in meters", ylab = "Number of clicks",
  main = list("Mara"))
lines(depthMara, predglmNBMara)
```



Summary of the negative binomial GLM for Mara data

```
summary(glmNBMara)
```

```
##
## Call:
## glm.nb(formula = clicks ~ depth, data = Mara1, init.theta = 3.171334653,
## link = log)
##
## Deviance Residuals:
##   Min       1Q   Median       3Q      Max
## -2.4017 -0.3473 -0.1379  0.1932  1.9203
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.4421086  0.5609124   2.571  0.0101 *
## depth        0.0005535  0.0017001   0.326  0.7448
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(3.1713) family taken to be 1)
##
## Null deviance: 12.035  on 9  degrees of freedom
## Residual deviance: 11.939  on 8  degrees of freedom
## AIC: 57.552
##
## Number of Fisher Scoring iterations: 1
##
```

```
##
##      Theta: 3.17
##      Std. Err.: 2.55
##
## 2 x log-likelihood: -51.552
```

Models with all the whales

Here it was done the same as above, we observed the trends of the number of clicks over time and depth, but with all the sampling data together with narwhal ID as a fixed factor

```
Freya12 <- rbind(Freya1T, Freya2T)

# Freya12 <- Freya12[, -c(9)]

# convert date/time in format month-day-year hour:min:sec
Freya12$time <- as.POSIXct(Freya12$time, format = "%m/%d/%y %H:%M:%S",
  tz = "America/Tijuana")

# Considering time since the first sample selected started
Freya12$mins <- difftime(time1 = Freya12$time, time2 = Freya12$time[1],
  units = "mins")

# Convert time differences (mins) in numeric format
Freya12$mins = as.numeric(Freya12$mins)

# convert date/time in format month-day-year hour:min:sec
Frida150$time <- as.POSIXct(Frida150$time, format = "%m/%d/%y %H:%M:%S",
  tz = "America/Tijuana")

# Considering time since the first sample selected started
Frida150$mins <- difftime(time1 = Frida150$time, time2 = Frida150$time[1],
  units = "mins")

allwhales <- rbind(Freya12, Balder1, Eistla1, Mutti1, Thora1,
  Frida150, Mara1, Jonas)

allwhales2 <- allwhales[-994, ] #149 samples for Frida (I had to delete the sample 144 for Frida150 data
because to do the analysis I can not have any NA)
```

Trends of the number of clicks over time

The number of clicks counted in sampling data was used as the response variable, and the time differences in minutes between samples was the explanatory variable. To know which model would best fit my data, it was seen the AIC values for GLMs and GAMs with negative binomial and poisson families.

```

glmNBallwhales2 = glm.nb(clicks ~ mins + whale, link = log, data = allwhales2)
gllmallwhalesP2 = glm(clicks ~ mins + whale, family = poisson(link = "log"),
  data = allwhales2)
gamallwhalesP2 = gam(clicks ~ mins + whale, family = poisson,
  link = "log", data = allwhales2)
gamNBallwhales2 = gam(clicks ~ mins + whale, family = nb(), data = allwhales2)

```

The glmNBallwhales2 (negative binomial GLM) had a smaller AIC value, so it will be chosen this model to see the Trend of the number of clicks over time for all the sampling data.

```
AIC(glmNBallwhales2, gllmallwhalesP2, gamallwhalesP2, gamNBallwhales2)
```

```

##           df   AIC
## glmNBallwhales2 10 6107.529
## gllmallwhalesP2  9 6910.081
## gamallwhalesP2  9 6910.081
## gamNBallwhales2 10 6107.577

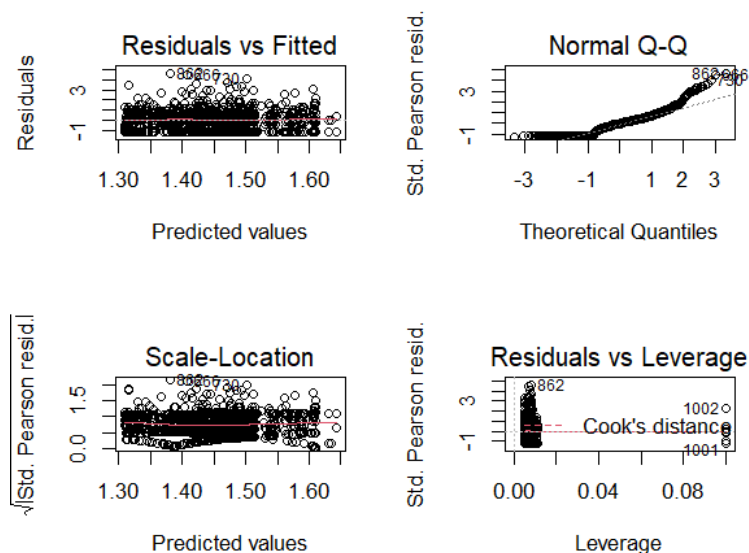
```

Residuals of the negative binomial GLM. In the plot “Residuals vs Leverage” there is some outliers that could influence the regression results

```

par(mfrow = c(2, 2))
plot(glmNBallwhales2)

```



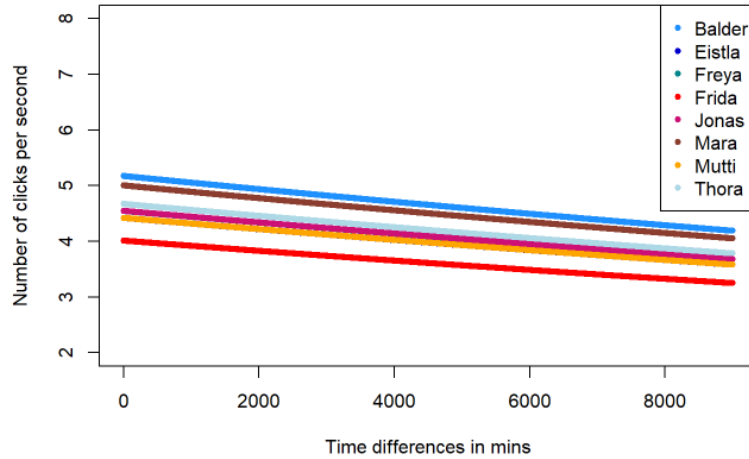
The results from the negative binomial GLM show that the number of clicks decreases non significantly (p-value = 0.12332) over time for all the sampling data. In conclusion, there is not a pattern over time for all the sampling data.

```

par(mfrow = c(1, 1))

glmNBallwhales2 = glm.nb(clicks ~ mins + whale, link = log, data = allwhales2)
DateTimeallwhales = seq(0, 9000, by = 1)
predglmNBallBalder2 <- predict(glmNBallwhales2, newdata = data.frame(mins = DateTimeallwhales,
  whale = "Balder"), type = "response")
plot(DateTimeallwhales, predglmNBallBalder2, xlab = "Time differences in mins",
  ylab = "Number of clicks per second", col = "dodgerblue1", pch = 20,
  ylim = c(2, 8))
predglmNBallEistla2 <- predict(glmNBallwhales2, newdata = data.frame(mins = DateTimeallwhales,
  whale = "Eistla"), type = "response")
points(DateTimeallwhales, predglmNBallEistla2, col = "mediumblue",
  pch = 20)
predglmNBallFreya2 <- predict(glmNBallwhales2, newdata = data.frame(mins = DateTimeallwhales,
  whale = "Freya"), type = "response")
points(DateTimeallwhales, predglmNBallFreya2, col = "turquoise4",
  pch = 20)
predglmNBallFrida2 <- predict(glmNBallwhales2, newdata = data.frame(mins = DateTimeallwhales,
  whale = "Frida"), type = "response")
points(DateTimeallwhales, predglmNBallFrida2, col = "red", pch = 20)
predglmNBallJonas2 <- predict(glmNBallwhales2, newdata = data.frame(mins = DateTimeallwhales,
  whale = "Jonas"), type = "response")
points(DateTimeallwhales, predglmNBallJonas2, col = "deeppink3",
  pch = 20)
predglmNBallMara2 <- predict(glmNBallwhales2, newdata = data.frame(mins = DateTimeallwhales,
  whale = "Mara"), type = "response")
points(DateTimeallwhales, predglmNBallMara2, col = "coral4",
  pch = 20)
predglmNBallMutti2 <- predict(glmNBallwhales2, newdata = data.frame(mins = DateTimeallwhales,
  whale = "Mutti"), type = "response")
points(DateTimeallwhales, predglmNBallMutti2, col = "orange",
  pch = 20)
predglmNBallThora2 <- predict(glmNBallwhales2, newdata = data.frame(mins = DateTimeallwhales,
  whale = "Thora"), type = "response")
points(DateTimeallwhales, predglmNBallThora2, col = "lightblue",
  pch = 20)
legend("topright", legend = c("Balder", "Eistla", "Freya", "Frida",
  "Jonas", "Mara", "Mutti", "Thora"), col = c("dodgerblue1",
  "mediumblue", "turquoise4", "red", "deeppink3", "coral4",
  "orange", "lightblue"), cex = 1, pch = 20)

```



Summary of the negative binomial GLM for all the sampling data

```
summary(glmNBallwhales2)
```

```
##
## Call:
## glm.nb(formula = clicks ~ mins + whale, data = allwhales2, link = log,
##   init.theta = 1.973431811)
##
## Deviance Residuals:
##   Min       1Q   Median       3Q      Max
## -2.25327 -0.74812 -0.03072  0.39261  2.68022
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.643e+00  8.765e-02  18.747 < 2e-16 ***
## mins        -2.346e-05  1.523e-05  -1.541  0.12332
## whaleEistla -1.574e-01  9.565e-02  -1.646  0.09986 .
## whaleFreya  -1.296e-01  9.813e-02  -1.321  0.18644
## whaleFrida  -2.543e-01  9.808e-02  -2.593  0.00952 **
## whaleJonas  -1.292e-01  8.582e-02  -1.505  0.13225
## whaleMara   -3.320e-02  2.798e-01  -0.119  0.90556
## whaleMutti  -1.571e-01  9.768e-02  -1.609  0.10770
## whaleThora  -1.022e-01  9.379e-02  -1.090  0.27571
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(1.9734) family taken to be 1)
##
```

```
## Null deviance: 1534.7 on 1208 degrees of freedom
## Residual deviance: 1525.8 on 1200 degrees of freedom
## AIC: 6107.5
##
## Number of Fisher Scoring iterations: 1
##
##
##      Theta: 1.973
##      Std. Err.: 0.141
##
## 2 x log-likelihood: -6087.529
```

Relationship between the number of clicks with depth

The number of clicks counted in sampling data was used as the response variable, and the depth in meters for each sample was the explanatory variable. To know which model would best fit my data, it was seen the AIC values for GLMs and GAMs with negative binomial and poisson families.

```
glmNballwhales = glm.nb(clicks ~ depth + whale, link = log, data = allwhales2)
glmallwhalesP = glm(clicks ~ depth + whale, family = poisson(link = "log"),
  data = allwhales2)
gamallwhalesP = gam(clicks ~ depth + whale, family = poisson,
  link = "log", data = allwhales2)
gamNballwhales = gam(clicks ~ depth + whale, family = nb(), data = allwhales2)
```

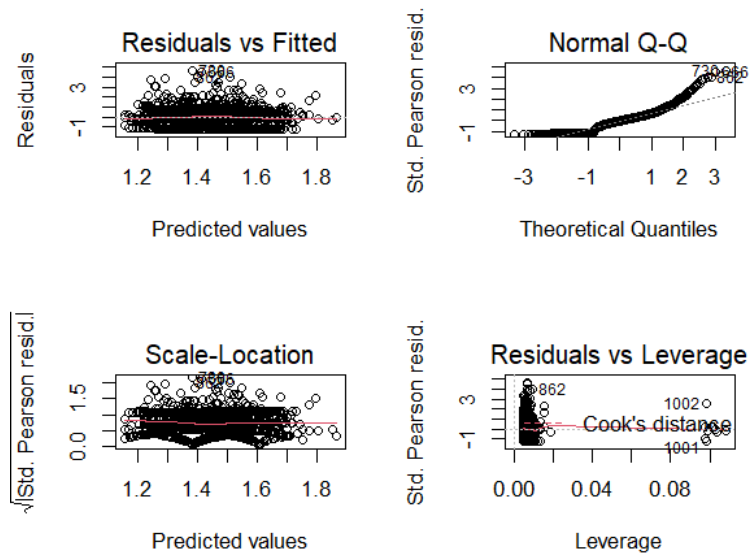
The glmNballwhales (negative binomial GLM) had a smaller AIC value, so it will be chosen this model to see the relationship between the number of clicks with depth for all the sampling data.

```
AIC(glmNballwhales, glmallwhalesP, gamallwhalesP, gamNballwhales)
```

```
##           df    AIC
## glmNballwhales 10 6090.036
## glmallwhalesP  9 6858.487
## gamallwhalesP  9 6858.487
## gamNballwhales 10 6090.085
```

Residuals of the negative binomial GLM. In the plot “Residuals vs Leverage” there is some outliers that could influence the regression results

```
par(mfrow = c(2, 2))
plot(glmNballwhales)
```

The results from the negative binomial GLM show that the number of clicks increases (positive estimate) significantly (p-value = 6.05e-06) with increasing depth for the sampling data. In conclusion, there is a relationship between the number of clicks with depth for all the sampling data.

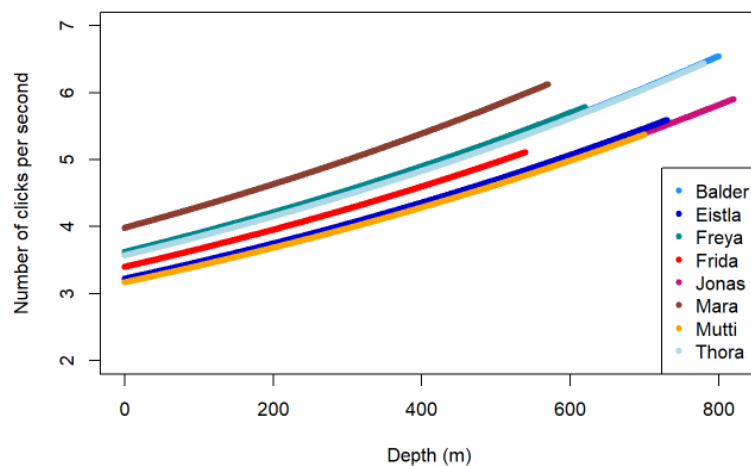
```
par(mfrow = c(1, 1))

glmNBallwhales = glm.nb(clicks ~ depth + whale, link = log, data = allwhales2)
depthallJonas = seq(0, 820, by = 1)
predglmNBallJonas <- predict(glmNBallwhales, newdata = data.frame(depth = depthallJonas,
  whale = "Jonas"), type = "response")
plot(depthallJonas, predglmNBallJonas, xlab = "Depth (m)", ylab = "Number of clicks per second",
  col = "deeppink3", pch = 20, ylim = c(2, 7))
depthallBalder = seq(0, 800, by = 1)
predglmNBallBalder <- predict(glmNBallwhales, newdata = data.frame(depth = depthallBalder,
  whale = "Balder"), type = "response")
points(depthallBalder, predglmNBallBalder, col = "dodgerblue1",
  pch = 20)
depthallEistla = seq(0, 730, by = 1)
predglmNBallEistla <- predict(glmNBallwhales, newdata = data.frame(depth = depthallEistla,
  whale = "Eistla"), type = "response")
points(depthallEistla, predglmNBallEistla, col = "mediumblue",
  pch = 20)
depthallFreya = seq(0, 620, by = 1)
predglmNBallFreya <- predict(glmNBallwhales, newdata = data.frame(depth = depthallFreya,
  whale = "Freya"), type = "response")
points(depthallFreya, predglmNBallFreya, col = "turquoise4",
  pch = 20)
```

```

depthallFrida = seq(0, 540, by = 1)
predglmNballFrida <- predict(glmNballwhales, newdata = data.frame(depth = depthallFrida,
  whale = "Frida"), type = "response")
points(depthallFrida, predglmNballFrida, col = "red", pch = 20)
depthallMara = seq(0, 570, by = 1)
predglmNballMara <- predict(glmNballwhales, newdata = data.frame(depth = depthallMara,
  whale = "Mara"), type = "response")
points(depthallMara, predglmNballMara, col = "coral4", pch = 20)
depthallMutti = seq(0, 700, by = 1)
predglmNballMutti <- predict(glmNballwhales, newdata = data.frame(depth = depthallMutti,
  whale = "Mutti"), type = "response")
points(depthallMutti, predglmNballMutti, col = "orange", pch = 20)
depthallThora = seq(0, 780, by = 1)
predglmNballThora <- predict(glmNballwhales, newdata = data.frame(depth = depthallThora,
  whale = "Thora"), type = "response")
points(depthallThora, predglmNballThora, col = "lightblue", pch = 20)
legend("bottomright", legend = c("Balder", "Eistla", "Freya",
  "Frida", "Jonas", "Mara", "Mutti", "Thora"), col = c("dodgerblue1",
  "mediumblue", "turquoise4", "red", "deeppink3", "coral4",
  "orange", "lightblue"), cex = 1, pch = 20)

```



Summary of the negative binomial GLM for all the sampling data

```
summary(glmNballwhales)
```

```

##
## Call:
## glm.nb(formula = clicks ~ depth + whale, data = allwhales, link = log,
##   init.theta = 2.032733003)
##
## Deviance Residuals:

```

```

##   Min    1Q  Median    3Q    Max
## -2.32599 -0.64534 -0.01847  0.38772  2.70703
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.2744715  0.0848762  15.016 < 2e-16 ***
## depth        0.0007552  0.0001669   4.524 6.06e-06 ***
## whaleEistla -0.1050543  0.0919270  -1.143  0.253
## whaleFreya  0.0121240  0.0858258   0.141  0.888
## whaleFrida  -0.0512702  0.0999345  -0.513  0.608
## whaleJonas  -0.1182877  0.0851233  -1.390  0.165
## whaleMara   0.1071330  0.2700504   0.397  0.692
## whaleMutti  -0.1218287  0.0917302  -1.328  0.184
## whaleThora  -0.0021025  0.0929279  -0.023  0.982
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(2.0327) family taken to be 1)
##
## Null deviance: 1556.9 on 1208 degrees of freedom
## Residual deviance: 1530.1 on 1200 degrees of freedom
## (1 observation deleted due to missingness)
## AIC: 6090
##
## Number of Fisher Scoring iterations: 1
##
##
##           Theta: 2.033
##           Std. Err.: 0.147
##
## 2 x log-likelihood: -6070.036

```

Estimating Click Rates

In this section, it will be estimated the echolocation clicks rate for each tag data, their variance, and the coefficient of variation of these rates. In addition, the number of seconds that were spent clicking and in silence will also be calculated for each tag data.

```

FreyaW <- FreyaW[, -c(2, 4)]
BalderW <- BalderW[, -c(2, 4)]
EistlaW <- EistlaW[, -c(2, 4)]
MuttiW <- MuttiW[, -c(2, 4)]
ThoraW <- ThoraW[, -c(2, 4)]
FridaW <- FridaW[, -c(2, 4)]

```

```

MaraW <- MaraW[, -c(2, 4)]
JonasW <- JonasW[, -c(2, 4)]
## The columns Depth and Buzzing were removed

FreyaO2 <- na.omit(FreyaW)
BalderO2 <- na.omit(BalderW)
EistlaO2 <- na.omit(EistlaW)
MuttiO2 <- na.omit(MuttiW)
ThoraO2 <- na.omit(ThoraW)
FridaO2 <- na.omit(FridaW)
MaraO2 <- na.omit(MaraW)
JonasO112 <- na.omit(JonasW)
## All NaN values were removed from the original data, so
## we can have only the number of seconds spent clicking.

SilenceF1 = nrow(FreyaW) - nrow(FreyaO2)
SilenceF2 = nrow(FreyaW) - nrow(FreyaO2)
SilenceF12 = nrow(FreyaW) - nrow(FreyaO2)
SilenceB = nrow(BalderW) - nrow(BalderO2)
SilenceE = nrow(EistlaW) - nrow(EistlaO2)
SilenceM = nrow(MuttiW) - nrow(MuttiO2)
SilenceT = nrow(ThoraW) - nrow(ThoraO2)
SilenceF = nrow(FridaW) - nrow(FridaO2)
SilenceJ = nrow(JonasW) - nrow(JonasO112)
SilenceMa = nrow(MaraW) - nrow(MaraO2)
## Calculation of the number of seconds spent in silence
## for each tag data.

ClickrateF1 = (nrow(FreyaO2) * mean(Freya1T$clicks))/nrow(FreyaW)
ClickrateF2 = (nrow(FreyaO2) * mean(Freya2T$clicks))/nrow(FreyaW)
ClickrateF12 = (nrow(FreyaO2) * mean(Freya12$clicks))/nrow(FreyaW)
ClickrateB = (nrow(BalderO2) * mean(Balder1$clicks))/nrow(BalderW)
ClickrateE = (nrow(EistlaO2) * mean(Eistla1$clicks))/nrow(EistlaW)
ClickrateM = (nrow(MuttiO2) * mean(Mutti1$clicks))/nrow(MuttiW)
ClickrateT = (nrow(ThoraO2) * mean(Thora1$clicks))/nrow(ThoraW)
ClickrateF = (nrow(FridaO2) * mean(Frida149$clicks))/nrow(FridaW)
ClickrateJ = (nrow(JonasO112) * mean(Jonas$clicks))/nrow(JonasW)
ClickrateMa = (nrow(MaraO2) * mean(Mara1$clicks))/nrow(MaraW)
## Calculation of Echolocation Clicks rate for each tag
## data.

varcrF1 = (((nrow(FreyaO2)/nrow(FreyaW))^2) * (1/((nrow(Freya1T))^2)) *
  var(Freya1T$clicks) * ((nrow(FreyaW) - nrow(Freya1T))/(nrow(FreyaW) -
  1)))

```

```

varcrF2 = (((nrow(FreyaO2)/nrow(FreyaW))^2) * (1/((nrow(Freya2T))^2)) *
  var(Freya2T$clicks) * ((nrow(FreyaW) - nrow(Freya2T))/(nrow(FreyaW) -
  1)))
varcrF12 = (((nrow(FreyaO2)/nrow(FreyaW))^2) * (1/((nrow(Freya12))^2)) *
  var(Freya12$clicks) * ((nrow(FreyaW) - nrow(Freya12))/(nrow(FreyaW) -
  1)))
varcrB = (((nrow(BalderO2)/nrow(BalderW))^2) * (1/((nrow(Balder1))^2)) *
  var(Balder1$clicks) * ((nrow(BalderW) - nrow(Balder1))/(nrow(BalderW) -
  1)))
varcrE = (((nrow(EistlaO2)/nrow(EistlaW))^2) * (1/((nrow(Eistla1))^2)) *
  var(Eistla1$clicks) * ((nrow(EistlaW) - nrow(Eistla1))/(nrow(EistlaW) -
  1)))
varcrM = (((nrow(MuttiO2)/nrow(MuttiW))^2) * (1/((nrow(Mutti1))^2)) *
  var(Mutti1$clicks) * ((nrow(MuttiW) - nrow(Mutti1))/(nrow(MuttiW) -
  1)))
varcrT = (((nrow(ThoraO2)/nrow(ThoraW))^2) * (1/((nrow(Thora1))^2)) *
  var(Thora1$clicks) * ((nrow(ThoraW) - nrow(Thora1))/(nrow(ThoraW) -
  1)))
varcrF = (((nrow(FridaO2)/nrow(FridaW))^2) * (1/((nrow(Frida149))^2)) *
  var(Frida149$clicks) * ((nrow(FridaW) - nrow(Frida149))/(nrow(FridaW) -
  1)))
varcrJ = (((nrow(JonasO112)/nrow(JonasW))^2) * (1/((nrow(Jonas))^2)) *
  var(Jonas$clicks) * ((nrow(JonasW) - nrow(Jonas))/(nrow(JonasW) -
  1)))
varcrMa = (((nrow(MaraO2)/nrow(MaraW))^2) * (1/((nrow(Mara1))^2)) *
  var(Mara1$clicks) * ((nrow(MaraW) - nrow(Mara1))/(nrow(MaraW) -
  1)))
## Calculation of Echolocation Clicks rate variance for
## each tag data.

cvF1 = (sqrt(varcrF1)/ClickrateF1) * 100
cvF2 = (sqrt(varcrF2)/ClickrateF2) * 100
cvF12 = (sqrt(varcrF12)/ClickrateF12) * 100
cvB = (sqrt(varcrB)/ClickrateB) * 100
cvE = (sqrt(varcrE)/ClickrateE) * 100
cvM = (sqrt(varcrM)/ClickrateM) * 100
cvT = (sqrt(varcrT)/ClickrateT) * 100
cvF = (sqrt(varcrF)/ClickrateF) * 100
cvJ = (sqrt(varcrJ)/ClickrateJ) * 100
cvMa = (sqrt(varcrMa)/ClickrateMa) * 100
## Calculation of coefficient of variation for each tag
## data.

```

Table with respective values for each whale (clickrates table)

```
clickrates <- data.frame(whaleID = 1:10, whale = NA, sex = NA,
  length = NA, samples = NA, totalrecord = NA, silence = NA,
  clicking = NA, mean = NA, cr = NA, crvar = NA, crcv = NA)
clickrates$whale[4] <- "Freya1"
clickrates$sex[4] <- "Female"
clickrates$length[4] <- 420
clickrates$samples[4] <- nrow(Freya1T)
clickrates$totalrecord[4] <- nrow(FreyaW)
clickrates$silence[4] <- SilenceF1
clickrates$clicking[4] <- nrow(FreyaO2)
clickrates$mean[4] <- mean(Freya1T$clicks)
clickrates$cr[4] <- ClickrateF1
clickrates$crvar[4] <- varcrF1
clickrates$crcv[4] <- cvF1
clickrates$whale[5] <- "Freya2"
clickrates$sex[5] <- "Female"
clickrates$length[5] <- 420
clickrates$samples[5] <- nrow(Freya2T)
clickrates$totalrecord[5] <- nrow(FreyaW)
clickrates$silence[5] <- SilenceF2
clickrates$clicking[5] <- nrow(FreyaO2)
clickrates$mean[5] <- mean(Freya2T$clicks)
clickrates$cr[5] <- ClickrateF2
clickrates$crvar[5] <- varcrF2
clickrates$crcv[5] <- cvF2
clickrates$whale[6] <- "Freya1+2"
clickrates$sex[6] <- "Female"
clickrates$length[6] <- 420
clickrates$samples[6] <- nrow(Freya12)
clickrates$totalrecord[6] <- nrow(FreyaW)
clickrates$silence[6] <- SilenceF12
clickrates$clicking[6] <- nrow(FreyaO2)
clickrates$mean[6] <- mean(Freya12$clicks)
clickrates$cr[6] <- ClickrateF12
clickrates$crvar[6] <- varcrF12
clickrates$crcv[6] <- cvF12
clickrates$whale[9] <- "Balder"
clickrates$sex[9] <- "Male"
clickrates$length[9] <- 372
clickrates$samples[9] <- nrow(Balder1)
clickrates$totalrecord[9] <- nrow(BalderW)
clickrates$silence[9] <- SilenceB
clickrates$clicking[9] <- nrow(BalderO2)
```

```

clickrates$mean[9] <- mean(Balder1$clicks)
clickrates$scr[9] <- ClickrateB
clickrates$crvar[9] <- varcrB
clickrates$crv[9] <- cvB
clickrates$whale[3] <- "Eistla"
clickrates$sex[3] <- "Female"
clickrates$length[3] <- 360
clickrates$samples[3] <- nrow(Eistla1)
clickrates$totalrecord[3] <- nrow(EistlaW)
clickrates$silence[3] <- SilenceE
clickrates$clicking[3] <- nrow(EistlaO2)
clickrates$mean[3] <- mean(Eistla1$clicks)
clickrates$scr[3] <- ClickrateE
clickrates$crvar[3] <- varcrE
clickrates$crv[3] <- cvE
clickrates$whale[2] <- "Mutti"
clickrates$sex[2] <- "Female"
clickrates$length[2] <- 465
clickrates$totalrecord[2] <- nrow(MuttiW)
clickrates$silence[2] <- SilenceM
clickrates$clicking[2] <- nrow(MuttiO2)
clickrates$mean[2] <- mean(Mutti1$clicks)
clickrates$scr[2] <- ClickrateM
clickrates$crvar[2] <- varcrM
clickrates$crv[2] <- cvM
clickrates$samples[2] <- nrow(Mutti1)
clickrates$whale[7] <- "Thora"
clickrates$sex[7] <- "Female"
clickrates$length[7] <- 341
clickrates$samples[7] <- nrow(Thora1)
clickrates$totalrecord[7] <- nrow(ThoraW)
clickrates$silence[7] <- SilenceT
clickrates$clicking[7] <- nrow(ThoraO2)
clickrates$mean[7] <- mean(Thora1$clicks)
clickrates$scr[7] <- ClickrateT
clickrates$crvar[7] <- varcrT
clickrates$crv[7] <- cvT
clickrates$whale[10] <- "Frida"
clickrates$sex[10] <- "Female"
clickrates$length[10] <- 380
clickrates$samples[10] <- nrow(Frida149)
clickrates$totalrecord[10] <- nrow(FridaW)
clickrates$silence[10] <- SilenceF
clickrates$clicking[10] <- nrow(FridaO2)

```

```

clickrates$mean[10] <- mean(Frida149$clicks)
clickrates$cr[10] <- ClickrateF
clickrates$crvar[10] <- varcrF
clickrates$cr cv[10] <- cvF
clickrates$whale[8] <- "Jonas"
clickrates$sex[8] <- "Male"
clickrates$length[8] <- 510
clickrates$samples[8] <- nrow(Jonas)
clickrates$totalrecord[8] <- nrow(JonasW)
clickrates$silence[8] <- SilenceJ
clickrates$clicking[8] <- nrow(JonasO112)
clickrates$mean[8] <- mean(Jonas$clicks)
clickrates$cr[8] <- ClickrateJ
clickrates$crvar[8] <- varcrJ
clickrates$cr cv[8] <- cvJ
clickrates$whale[1] <- "Mara"
clickrates$sex[1] <- "Female"
clickrates$length[1] <- 390
clickrates$samples[1] <- nrow(Mara1)
clickrates$totalrecord[1] <- nrow(MaraW)
clickrates$silence[1] <- SilenceMa
clickrates$clicking[1] <- nrow(MaraO2)
clickrates$mean[1] <- mean(Mara1$clicks)
clickrates$cr[1] <- ClickrateMa
clickrates$crvar[1] <- varcrMa
clickrates$cr cv[1] <- cvMa

```

View(clickrates)

clickrates

```

## whaleID whale sex length samples totalrecord silence clicking mean
## 1 1 Mara Female 390 10 3665 2109 1556 5.000000
## 2 2 Mutti Female 465 150 261550 138304 123246 4.200000
## 3 3 Eistla Female 360 150 299068 189211 109857 4.153333
## 4 4 Freya1 Female 420 100 113191 76064 37127 3.960000
## 5 5 Freya2 Female 420 100 113191 76064 37127 4.920000
## 6 6 Freya1+2 Female 420 200 113191 76064 37127 4.440000
## 7 7 Thora Female 341 150 331197 239682 91515 4.333333
## 8 8 Jonas Male 510 200 527105 377438 149667 4.095000
## 9 9 Balder Male 372 200 375690 292497 83193 4.700000
## 10 10 Frida Female 380 149 200517 162606 37911 3.812081
## cr crvar cr cv
## 1 2.1227831 2.517278e-02 7.4741184
## 2 1.9790985 8.790590e-05 0.4737417
## 3 1.5256488 5.509142e-05 0.4865051

```



```
## 4 1.2988923 8.250169e-05 0.6992916
## 5 1.6137753 6.638007e-05 0.5048656
## 6 1.4563338 1.912204e-05 0.3002660
## 7 1.1973689 5.015656e-05 0.5914743
## 8 1.1627406 1.641406e-05 0.3484377
## 9 1.0407706 2.158654e-05 0.4464127
## 10 0.7207358 1.748874e-05 0.5802340
```

A new table (clickrates2) was created because Freya data had 2 different samplings, so it was decided to use the values obtained from the two samplings together (Freya1+2). The next table is Without Freya1 and Freya2 data.

```
clickrates2 <- clickrates[-c(4, 5), ]
```

```
View(clickrates2)
```

```
clickrates2
```

```
## whaleID whale sex length samples totalrecord silence clicking mean
## 1 1 Mara Female 390 10 3665 2109 1556 5.000000
## 2 2 Mutti Female 465 150 261550 138304 123246 4.200000
## 3 3 Eistla Female 360 150 299068 189211 109857 4.153333
## 6 6 Freya1+2 Female 420 200 113191 76064 37127 4.440000
## 7 7 Thora Female 341 150 331197 239682 91515 4.333333
## 8 8 Jonas Male 510 200 527105 377438 149667 4.095000
## 9 9 Balder Male 372 200 375690 292497 83193 4.700000
## 10 10 Frida Female 380 149 200517 162606 37911 3.812081
## cr crvar crcv
## 1 2.1227831 2.517278e-02 7.4741184
## 2 1.9790985 8.790590e-05 0.4737417
## 3 1.5256488 5.509142e-05 0.4865051
## 6 1.4563338 1.912204e-05 0.3002660
## 7 1.1973689 5.015656e-05 0.5914743
## 8 1.1627406 1.641406e-05 0.3484377
## 9 1.0407706 2.158654e-05 0.4464127
## 10 0.7207358 1.748874e-05 0.5802340
```

Mean cue rate (unweighted mean): 1.400685 clicks/s for clickrates2 table

```
mean(clickrates2$cr)
```

```
## [1] 1.400685
```

To obtain the variability of the unweighted mean cue rate, a resampling methodology of the values obtained for the 8 whales (in clickrates2 table) was carried out using a bootstrap data frame. The values were resampled 999 times by bootstrapping the tags.

```
df <- clickrates2
```

```
# fix randomization seed, make sample() reproducible
```

```

set.seed(1)
# take a random sample of rows from clickrates2
sampleSize <- nrow(df)

# do many bootstrap replications
bSamples <- 999
# make container for results
bResults <- rep(NA, bSamples)

# loop over bootstraps
for (b in seq_len(bSamples)) {
  # A data.frame of bootstrap
  bData <- df[sample(x = 1:sampleSize, size = sampleSize, replace = TRUE),
  ]
  # compute your statistic of interest (unweighted mean
  # cue rate)
  bValue <- mean(bData[["cr"]])
  # store results in container
  bResults[[b]] <- bValue
}
# show what we computed

# compute the variance of the unweighted mean cue rate
var(bResults)

## [1] 0.02551547

```

Variance of the unweighted mean cue rate: 0.0255155 for clickrates2 table

```

# Calculation of the coefficient of variation for the
# unweighted mean cue rate
cvmean2 = 100 * (sd(bResults)/mean(bResults))

```

Coefficient of variation of the unweighted mean cue rate: 11.4146734% for clickrates2 table

```

crmeanW2 = (sum(clickrates2$clicking * clickrates2$mean)/sum(clickrates2$totalrecord))

```

Mean cue rate (weighted mean): 1.2743985 clicks/s for clickrates2 table

To obtain the variability of the weighted mean cue rate, a resampling methodology of the values obtained for the 8 whales (in clickrates2 table) was carried out using a bootstrap data frame. The values were resampled 999 times by bootstrapping the tags.

```

df <- clickrates2
# fix randomization seed, make sample() reproducible
set.seed(1)
# take a random sample of rows from clickrates2

```

```

sampleSize <- nrow(df)

# do many bootstrap replications
b2Samples <- 999
# make container for results
b2Results <- rep(NA, b2Samples)

# loop over bootstraps
for (b2 in seq_len(b2Samples)) {
  # A data.frame of bootstrap
  b2Data <- df[sample(x = 1:sampleSize, size = sampleSize,
    replace = TRUE), ]
  # compute your statistic of interest (weighted mean cue
  # rate)
  b2Value <- (sum(b2Data[["clicking"]] * b2Data[["mean"]]))/sum(b2Data[["totalrecord"]])
  # store results in container
  b2Results[[b2]] <- b2Value
}
# show what we computed

# compute the variance of the weighted mean cue rate
var(b2Results)

## [1] 0.01704149

```

Variance of the weighted mean cue rate: 0.0170415 for clickrates2 table

```

# Calculation of the coefficient of variation for the
# weighted mean cue rate
cvmean22 = 100 * (sd(b2Results)/mean(b2Results))

```

Coefficient of variation of the weighted mean cue rate: 10.2041764% for clickrates2 table

Test for the influence of narwhals sex and length in echolocation clicks rates

The click rates of the eight whales were used as the response variable, and the length in centimeters and sex for each whale were the explanatory variables. To know which model would fit best my data, it was seen the residuals and the AIC values in the GLMs (generalized linear models), and GAMs (generalized additive models). The AIC values were seen for the models with families Gaussian with log link and Gamma with log, and identity link.

```

clickrates2 <- clickrates[-c(4, 5), ] #Without Freya1 and Freya2 data

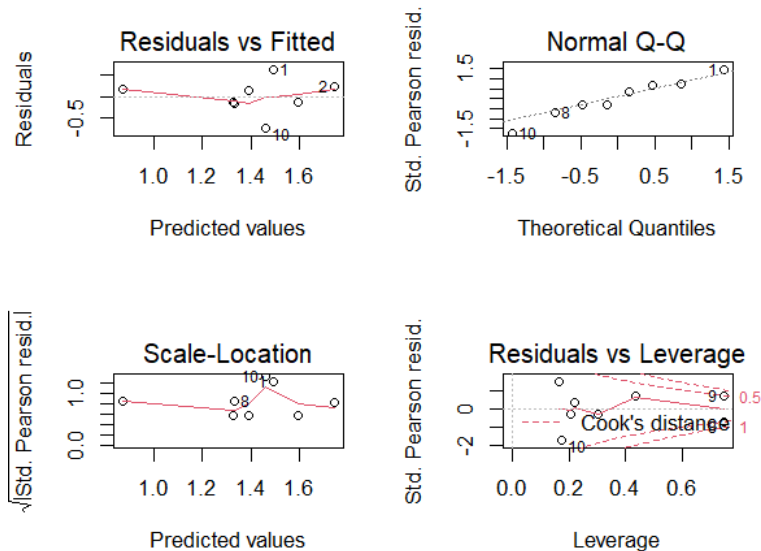
```

Residuals of the Gaussian GLM with identity link. Because the assumptions of Homoscedasticity and normality of the model's residuals are violated, gaussian models with the identity link will not be used.

```

glmGLS2 = glm(cr ~ length + sex, family = gaussian(link = "identity"),
  data = clickrates2)
par(mfrow = c(2, 2))
plot(glmGLS2)

```



It was done GLMs and GAMs with a family that considers the outcome (click rates) a non-negative continuous variable: Gamma family with log, and identity link. It was also done a GAM and a GLM of the Gaussian family with a log link since this way there is no assumption of the normality of residuals, and because this family considers the outcome a continuous decimal data. It will be seen from all these models which one will have the lowest AIC value.

```

glmGamma1LS = glm(cr ~ length + sex, family = Gamma(link = "log"),
  data = clickrates2)
glmGamma2LS = glm(cr ~ length + sex, family = Gamma(link = "identity"),
  data = clickrates2)
glmGLS = glm(cr ~ length + sex, family = gaussian(link = "log"),
  data = clickrates2)
gamGamma1LS = gam(cr ~ s(length, k = 7) + sex, family = Gamma(link = "log"),
  data = clickrates2)
gamGamma2LS = gam(cr ~ s(length, k = 7) + sex, family = Gamma(link = "identity"),
  data = clickrates2)
gamGLS = gam(cr ~ s(length, k = 7) + sex, family = gaussian(link = "log"),
  data = clickrates2)

```

The glmGLS (Gaussian GLM with log link) had the smaller AIC value, so it will be chosen this models to observe if there is an influence of narwhals length and sex in echolocation clicks rates.

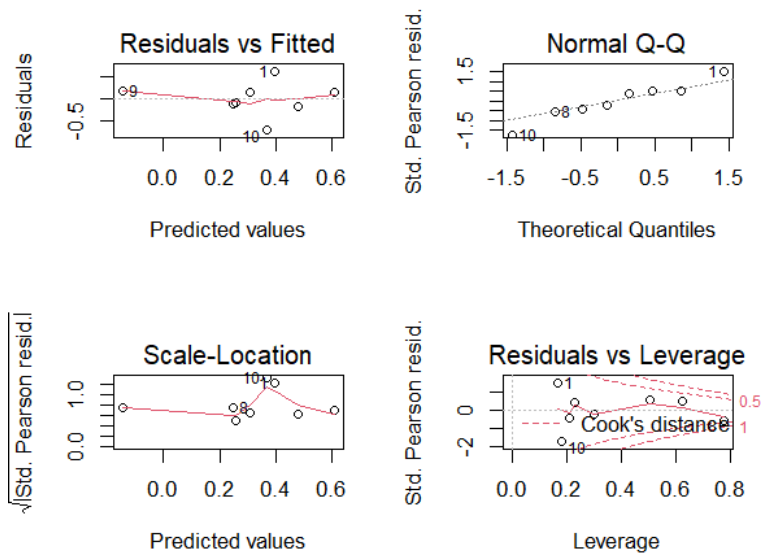
```
# Best AIC
```

```
AIC(glmGLS, gamGLS, glmGamma1LS, glmGamma2LS, gamGamma1LS, gamGamma2LS)
```

```
##           df    AIC
## glmGLS    4.000000 14.49395
## gamGLS    4.000001 14.49396
## glmGamma1LS 4.000000 14.62793
## glmGamma2LS 4.000000 14.89602
## gamGamma1LS 4.000001 14.62793
## gamGamma2LS 4.000002 14.89603
```

Residuals of the Gaussian GLM with log link

```
par(mfrow = c(2, 2))
plot(glmGLS)
```



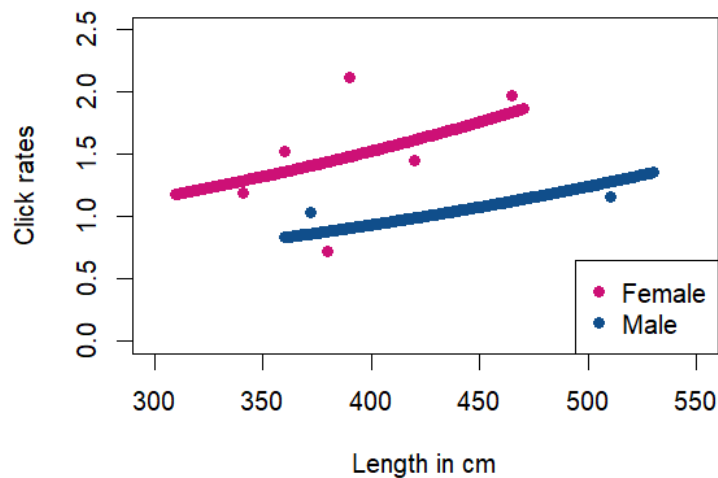
The results from the Gaussian GLM with log link show that the click rates increases non significantly with increasing length. It seems that there is not an influence of narwhals length in echolocation clicks rates.

```
par(mfrow = c(1, 1))
glmGLS = glm(cr ~ length + sex, family = gaussian(link = "log"),
  data = clickrates2)
LSM = seq(360, 530, by = 1)
LSF = seq(310, 470, by = 1)
predglmGLSFemale <- predict(glmGLS, newdata = data.frame(length = LSF,
  sex = "Female"), type = "response")
plot(LSF, predglmGLSFemale, ylim = c(0, 2.5), xlim = c(300, 550),
  xlab = "Length in cm", ylab = "Click rates", col = "deeppink3",
  pch = 19)
```

```

predglmGLSMale <- predict(glmGLS, newdata = data.frame(length = LSM,
sex = "Male"), type = "response")
points(LSM, predglmGLSMale, col = "dodgerblue4", pch = 19)
legend("bottomright", legend = c("Female", "Male"), col = c("deeppink3",
"dodgerblue4"), cex = 1, pch = 19)
points(cr ~ length, data = clickrates2, pch = 19, col = c("deeppink3",
"dodgerblue4")[as.factor(clickrates2$sex)])

```



Summary of the Gaussian GLM with log link. The p-value for the dummy variable (sexMale) is not significant. Therefore, there is not a significant difference in the means cue rates between males and females, but more males are needed to confirm these results in the future!

```
summary(glmGLS)
```

```

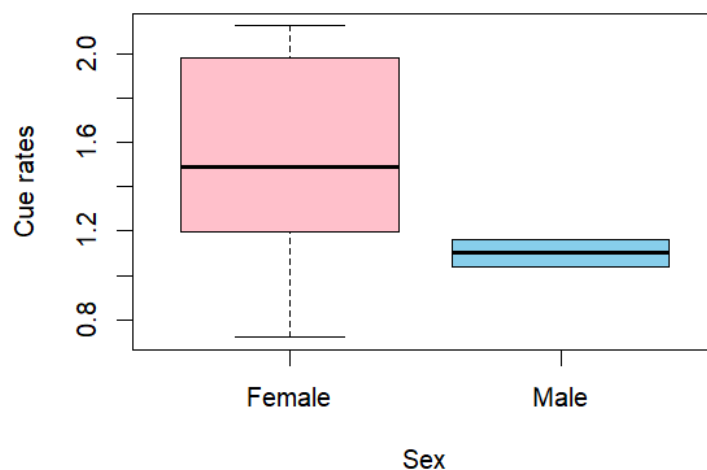
##
## Call:
## glm(formula = cr ~ length + sex, family = gaussian(link = "log"),
## data = clickrates2)
##
## Deviance Residuals:
## 1 2 3 6 7 8 9 10
## 0.63778 0.13920 0.16263 -0.16157 -0.09363 -0.11908 0.17663 -0.72244
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.718903 0.998060 -0.720 0.504
## length 0.002857 0.002460 1.162 0.298
## sexMale -0.490003 0.358511 -1.367 0.230

```

```
##
## (Dispersion parameter for gaussian family taken to be 0.2109525)
##
## Null deviance: 1.5645 on 7 degrees of freedom
## Residual deviance: 1.0548 on 5 degrees of freedom
## AIC: 14.494
##
## Number of Fisher Scoring iterations: 5
```

Boxplot of click rates for females and males. It is necessary to verify if click rates really do not have differences between males and females using a greater number of individuals since our data only had eight individuals with just two males.

```
res <- boxplot(cr ~ sex, data = clickrates2, ylab = "Cue rates",
  xlab = "Sex", col = c("pink", "skyblue"))
```



Quartile summaries of the boxes. Here is present the values for the minimum, first quartile (Q1), median, third quartile (Q3), and maximum click rates for each sex

```
colnames(res$stats) <- res$names
rownames(res$stats) <- c("min", "lower quartile", "median", "upper quartile",
  "max")
res$stats
##           Female  Male
## min          0.7207358 1.040771
## lower quartile 1.1973689 1.040771
## median        1.4909913 1.101756
## upper quartile 1.9790985 1.162741
## max          2.1227831 1.162741
```

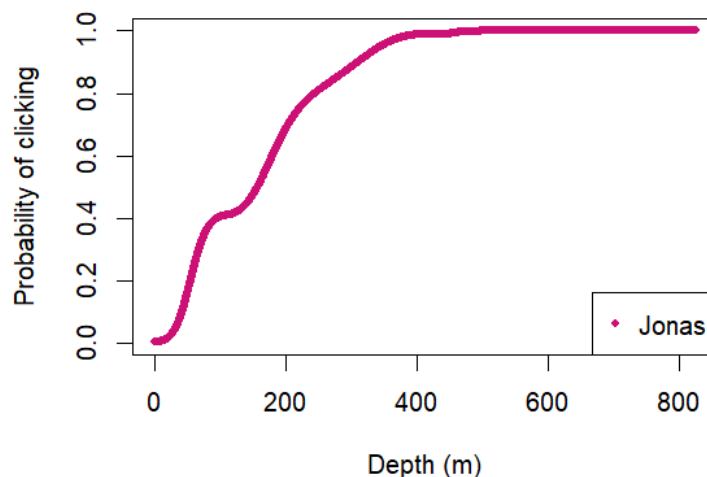
Estimation of the cue rate in the second part of Jonas data

With our results, it is possible to obtain a click rate for the entire duration of the Jonas' tag, without having the second part of his clicking periods processed for the sound data, by combining the probability of foraging sounds as a function of depth with the number of clicks per second, while clicking, as a function of depth since with increasing depths, during clicking time, echolocation is more frequent.

Here we predicted the probability of sounds for the depths presented in the second part of the Jonas data since we want a cue rate where the acoustics were not processed. We used the same binomial GAM where we predicted the probability of sounds as a function of depth for all the whales.

```
JonasO2 <- JonasO[551016:807030, ] #second part of Jonas data not processed for sounds
```

```
gamwhalesJJ <- gam(Clicking ~ s(Depth), family = "binomial",  
  data = whales11[whales11$whale %in% "Jonas", ])  
depthJJ = JonasO2$Depth  
predallJonasJJ <- predict(gamwhalesJJ, newdata = data.frame(Depth = depthJJ,  
  whale = "Jonas"), type = "response")  
plot(depthJJ, predallJonasJJ, xlab = "Depth (m)", ylab = "Probability of clicking",  
  col = "deeppink3", pch = 20)  
legend("bottomright", legend = c("Jonas"), col = c("deeppink3"),  
  cex = 1, pch = 20)
```

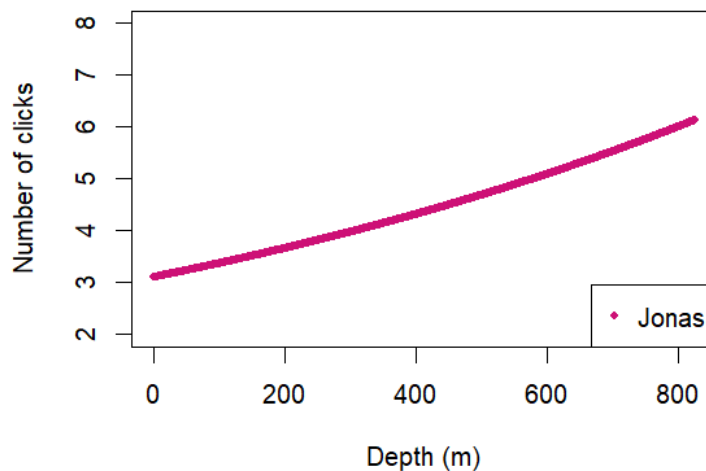


Here we predicted the number of sounds produced for the depths presented in the part of Jonas data without acoustics. We used the same negative binomial GLM where we predicted the number of sounds as a function of depth for all the whales.


```

glmNBallwhalesJJ = glm.nb(clicks ~ depth, link = log, data = allwhales[allwhales$whale %in%
  "Jonas", ])
depthJJ = JonasO2$Depth
predglmNBallJonasJJ <- predict(glmNBallwhalesJJ, newdata = data.frame(depth = depthJJ,
  whale = "Jonas"), type = "response")
plot(depthJJ, predglmNBallJonasJJ, xlab = "Depth (m)", ylab = "Number of clicks",
  col = "deeppink3", pch = 20, ylim = c(2, 8))
legend("bottomright", legend = c("Jonas"), col = c("deeppink3"),
  cex = 1, pch = 20)

```

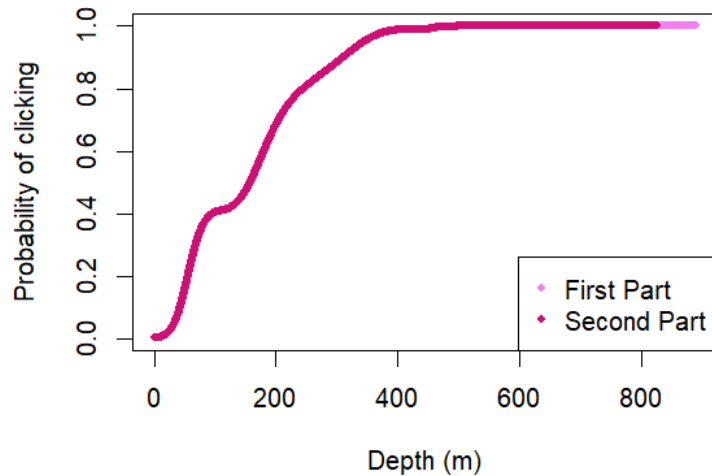


Prediction of the probability of sounds for the depths presented in the first (where the acoustics were processed) vs second part (where the acoustics were not processed - as shown in the above plots) of Jonas data

```

gamJonas12Parts <- gam(Clicking ~ s(Depth), family = "binomial",
  data = whales11[whales11$whale %in% "Jonas", ])
depthJonas1Part = whales11[whales11$whale %in% "Jonas", ]$Depth
predgamJonas1Parts <- predict(gamJonas12Parts, newdata = data.frame(Depth = depthJonas1Part),
  type = "response")
plot(depthJonas1Part, predgamJonas1Parts, xlab = "Depth (m)",
  ylab = "Probability of clicking", col = "violet", pch = 20)
legend("bottomright", legend = c("First Part", "Second Part"),
  col = c("violet", "deeppink3"), cex = 1, pch = 20)
depthJJ = JonasO2$Depth
predgamJonas2Parts <- predict(gamJonas12Parts, newdata = data.frame(Depth = depthJJ),
  type = "response")
points(depthJJ, predgamJonas2Parts, col = "deeppink3", pch = 20)

```

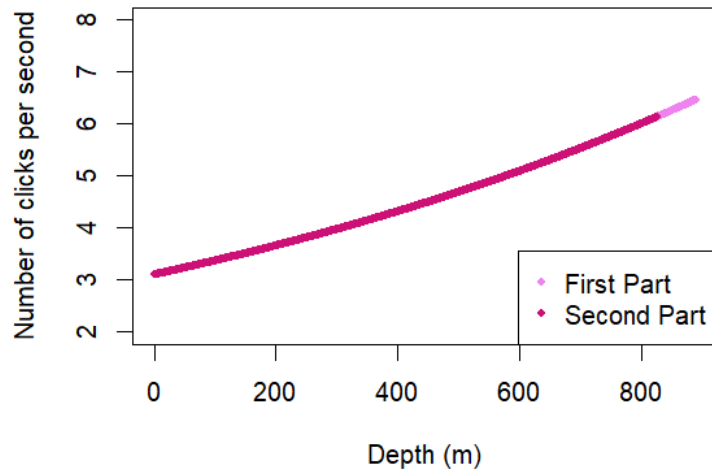


Prediction of the number of sounds produced for the depths presented in the first (where the acoustics were processed) vs second part (where the acoustics were not processed) of Jonas data

```

glmNBJonas12Parts = glm.nb(clicks ~ depth, link = log, data = allwhales2[allwhales2$whale %in%
  "Jonas", ])
depthJonas1Part = whales11[whales11$whale %in% "Jonas", ]$Depth
predglmNBJonas1Parts <- predict(glmNBJonas12Parts, newdata = data.frame(depth = depthJonas1Part,
  whale = "Jonas"), type = "response")
plot(depthJonas1Part, predglmNBJonas1Parts, xlab = "Depth (m)",
  ylab = "Number of clicks per second", col = "violet", pch = 20,
  ylim = c(2, 8))
legend("bottomright", legend = c("First Part", "Second Part"),
  col = c("violet", "deeppink3"), cex = 1, pch = 20)
depthJJ = JonasO2$Depth
predglmNBJonas2Parts <- predict(glmNBJonas12Parts, newdata = data.frame(depth = depthJJ,
  whale = "Jonas"), type = "response")
points(depthJJ, predglmNBJonas2Parts, col = "deeppink3", pch = 20)

```



```
predJonasJJ <- sum(predglmNBallJonasJJ * predallJonasJJ)
```

```
predJonasJJ/nrow(JonasO2)
```

```
## [1] 1.426719
```

Cue rate for the second part of Jonas data: 1.426719 clicks/s

The click rate for the last part of the data was slightly higher than the first part (23%)

```
((predJonasJJ/nrow(JonasO2)) - (ClickrateJ))/(ClickrateJ)
```

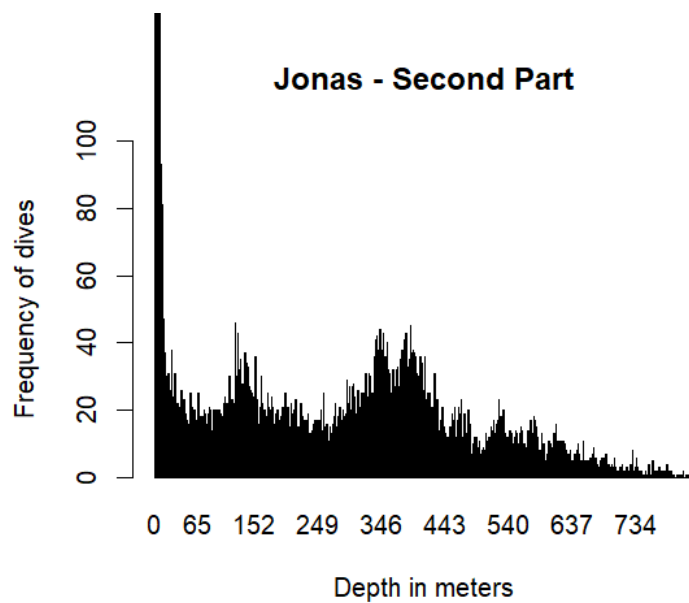
```
## [1] 0.2270312
```

Histogram of depths for the second part of Jonas data with absolute frequencies. Jonas dove more frequently to deeper depths during his second part of the data

```
maxnJ2 <- max(JonasO2$Depth)
```

```
tempJ2 <- factor(JonasO2$Depth[JonasO2$Depth <= maxnJ2], levels = 0:maxnJ2,
  labels = 0:maxnJ2)
```

```
barplot(table(tempJ2), xlab = "Depth in meters", ylab = "Frequency of dives",
  main = list("Jonas - Second Part"), ylim = c(0, 100))
```

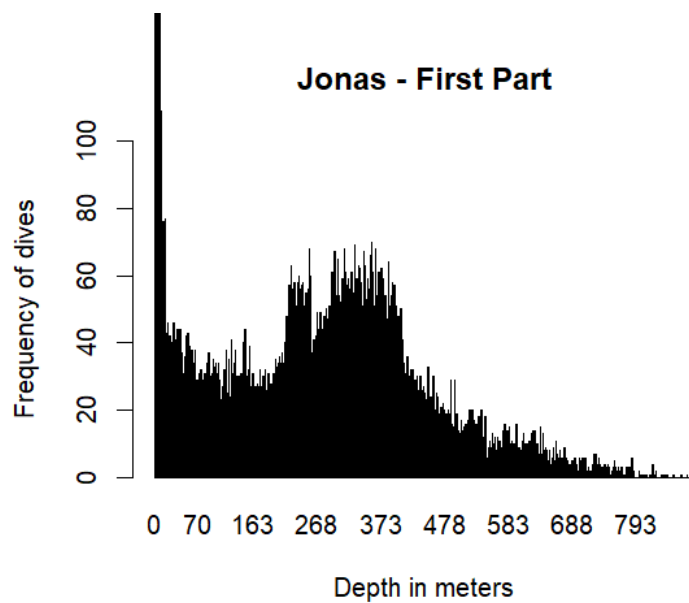


Mean depth where Jonas dove during his second part.

```
mean(JonasO2$Depth)
## [1] 138.3621
```

Histogram of depths for the first part of Jonas data with absolute frequencies

```
maxnJ <- max(whales11[whales11$whale %in% "Jonas", ]$Depth)
tempJ <- factor(whales11[whales11$whale %in% "Jonas", ]$Depth[whales11[whales11$whale %in%
"Jonas", ]$Depth <= maxnJ], levels = 0:maxnJ, labels = 0:maxnJ)
barplot(table(tempJ), xlab = "Depth in meters", ylab = "Frequency of dives",
main = list("Jonas - First Part"), ylim = c(0, 100))
```



Mean depth where Jonas dove during his first part

```
mean(whales11[whales11$whale %in% "Jonas", ]$Depth)
## [1] 110.7887
```

Histogram of depths for the first and second parts of Jonas data. Here the frequencies are in percentage and the red lines represent the mean depth where Jonas dove during the first and second parts of the data

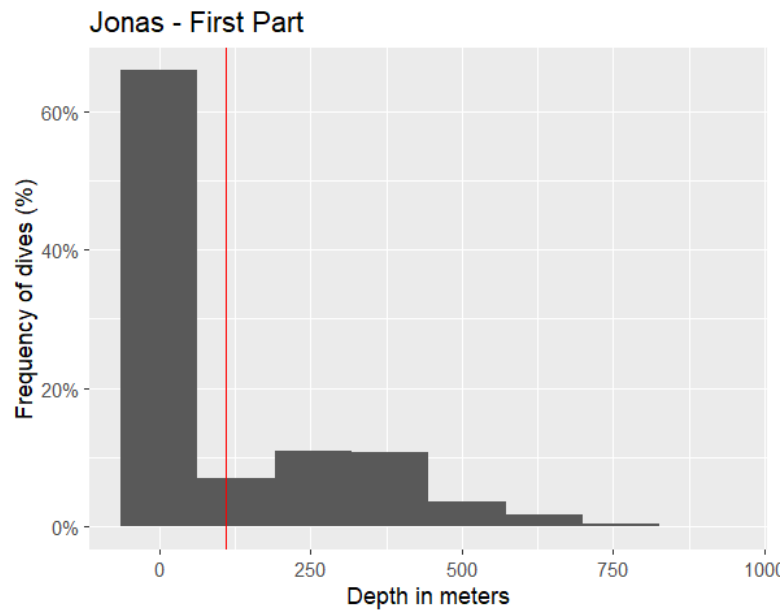
```
var1 <- data.frame(whales11[whales11$whale %in% "Jonas", ]$Depth)
var2 <- data.frame(JonasO2$Depth)

names(var1)[names(var1) == "whales11.whales11.whale.in.Jonas....Depth"] <- "Depth"
names(var2)[names(var2) == "JonasO2.Depth"] <- "Depth"

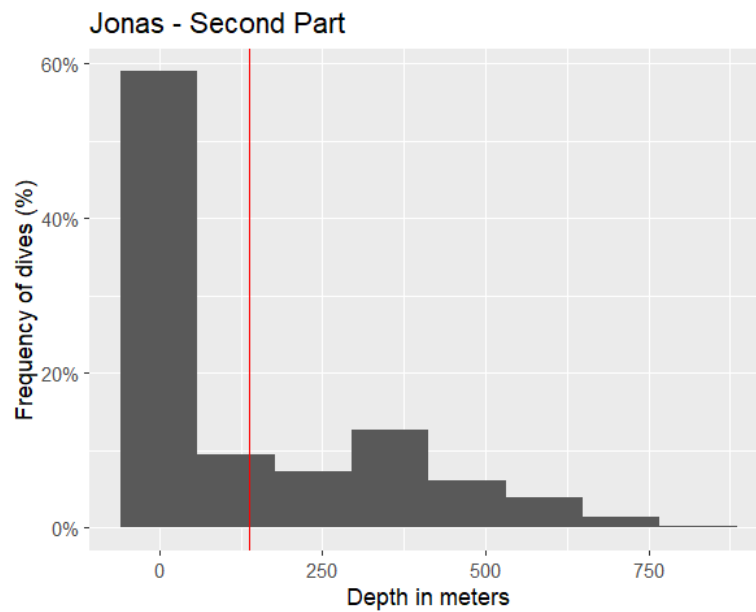
var1$part <- 1
var2$part <- 2

var12 <- rbind(var1, var2)

ggplot(var1, aes(x = Depth)) + geom_histogram(aes(y = stat(count)/sum(count)),
  bins = 8) + scale_y_continuous(labels = scales::percent) +
  geom_vline(data = var1, aes(xintercept = 110.7887), colour = "red") +
  xlab("Depth in meters") + ylab("Frequency of dives (%)") +
  ggtitle("Jonas - First Part")
```



```
ggplot(var2, aes(x = Depth)) + geom_histogram(aes(y = stat(count)/sum(count)),
  bins = 8) + scale_y_continuous(labels = scales::percent) +
  geom_vline(data = var2, aes(xintercept = 138.3621), colour = "red") +
  xlab("Depth in meters") + ylab("Frequency of dives (%)") +
  ggtitle("Jonas - Second Part")
```



References

- Blackwell, S.B., Tervo, O.M., Conrad, A.S., Sinding, M.H.S., Hansen, R.G., Ditlevsen, S. & Heide-Jørgensen, M.P. (2018). Spatial and temporal patterns of sound production in east greenland narwhals. *PLOS ONE*, **13**, e0198295. Retrieved from <https://doi.org/10.1371/journal.pone.0198295>
- Gomez, K.A. & Gomez, A.A. (1984). *Statistical procedures for agricultural research*. John Wiley; Sons, Inc, New York.