Modelling reindeer rut activity using on-animal acoustic recorders and machine learning

Alexander J. Boucher

A Thesis
in
The Department
of
Biology

Presented in Partial Fulfillment of the Requirements
For the Degree of
Master of Science (Biology)

At Concordia University
Montréal, Québec, Canada

July 2023

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Alexander J. Boucher

Entitled: Modelling reindeer rut activity using on-animal acoustic recorders and
                        machine learning

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Biology)

and complies with the regulations of the University and meets the accepted standards with respect
to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. Grant E. Brown

_____ External Examiner
Dr. James W. A. Grant

_____ Examiner
Dr. Grant E. Brown

_____ Examiner
Dr. Eric J. Pedersen

_____ Thesis Supervisor
Dr. Robert B. Weladji

Approved by

_____
Dr. Robert B. Weladji, Graduate Program Director

_____, 2023   _____
                      Dr. Pascale Sicotte, Dean of Arts and Science

Abstract

Modelling reindeer rut activity using on-animal acoustic recorders and machine learning

Alexander J. Boucher

Researchers have been using sound to study the biology of wildlife to understand their ecology and behaviour for decades. By gathering audio from free-ranging species using on-animal recorders, their vocalizations can be used to describe their behaviour and ecology through signal processing. Unfortunately, processing hours of recordings is incredibly time-consuming. By applying machine learning to audio recordings, researchers have used neural networks to decrease the processing time of acoustic data. However, until now, most of this research has focused on analyzing the data of stationary recorders. To show the utility of on-animal recorders in combination with machine learning, we recorded the vocalizations of reindeer (*Rangifer tarandus*) during their rut at the Kutuharju research station in Kaamanen, Finland. We used vocalizations as an activity index to describe the rut activity of male reindeer. In 2019 and 2020, we placed recorders around the necks of seven reindeer during their rut. We trained convolutional neural networks to identify reindeer grunts, which were then used to classify their vocalizations. Of the networks' vocalization classifications, around 95% of them were correct. With such high metrics, we could reliably explore the males' activity patterns using a neural network. We then analyzed the reindeers' vocalization using generalized additive models. The patterns suggested heavier, older males vocalized more than lighter, younger males and, overall, were more active during the day than night. Overall, on-animal acoustic recorders, in tandem with machine learning, proved to be effective tools, and with more attention, they could prove valuable tools for other researchers.

Acknowledgments

Contribution of Authors

Alexander J. Boucher[1], Robert B. Weladji[1], Øystein Holand[2] and Jouko Kumpula[3]

[1] Department of Biology, Concordia University, 7141 Sherbrooke Street West, Montreal, QC, H4B 1R6, Canada

[2] Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Ås, PO Box 5003, Norway

[3] Natural Resources Institute of Finland (Luke), Reindeer Research Station, 99910, Finland

I would like to acknowledge the contribution of the following collaborators on the project:

- Conception and design of the study: A and R
- Data Collection: A, R and O
- Analysis and interpretation of data: A and R
- Drafting of manuscript: A
- Critical revision: A and R

Table of Contents

List of Figures

Figure 1. The on-animal acoustic recorder used to record the rutting vocalizations of the reindeer. The orange case is the 3D-printed weatherproof housing containing the Soroka-15E recorder. The four panels represent the (**A**) interior, (**B**) top, (**C**) side, and (**D**) front. The ruler denotes units in centimetres.

Figure 2. Various spectrograms of noises captured by the on-animal acoustic recorder. The different spectrograms represent (**A**) a series of reindeer grunts, (**B**) a single reindeer grunt, (**C**) the sound captured by the recorder as it scratched against some branches and (**D**) the sound of the wind as it blew against the recorder. The y-axis represents frequency (Hz), and each spectrogram is exactly four seconds long.

Figure 3. A binary data set was created in the above process. For each file, the data set was made by reading a given audio file and applying the above pre-processing steps. Segment duration and spectrogram parameters are detailed in **Table 2** (Dufourq et al., 2022).

Figure 4 A-G. Estimated smooth curves of the 24-hour rut patterns of seven reindeer according to their acoustic recordings represented as grunting behaviour taken as a proportion of time for each hour. (**A**) represents the estimated smooth curve of a 5.5-year-old and its complete data set to serve as an example for the data gathered. (**B-G**) represent six smooth curves of six individuals ranging from 1.5-years-old to 4.5-years-old. Black lines represent the mean, and shaded areas represent the 95% confidence interval. Models were fit using ten basis functions, and edf represents effective degrees of freedom. Days of observation represent the number of days the smooth curves were estimated. 2019 and 2020 represent the year the data were gathered. 1.5-5.5 represent the age of the reindeer, and 1-3 represent different individuals of the same age class.

Figure 5 A-G. . Estimated smooth curves of the daily patterns of seven reindeer according to their acoustic recordings represented as grunting behaviour taken as a proportion of time for each hour. (**A**) represents the estimated smooth curve of a 5.5-year-old and its complete data set to serve as an example for the data gathered (note that the scale for this figure differs from the other six). (**B-G**) represent six smooth curves of six individuals ranging from 1.5-years-old to 4.5-years-old. Black lines represent the mean, shaded areas represent the 95% confidence interval, and the dotted vertical line represents the peak rutting day. Models were fit using 22 basis functions, and edf represents effective degrees of freedom. Days in relation to peak rut refer to the estimated peak rut date of that year's cohort. The peak ruts for the 2019 and 2020 males were estimated to be October 5th and October 6th, respectively. 2019 and 2020 represent the year the data were gathered. 1.5-5.5 represent the age of the reindeer, and 1-3 represent different individuals of the same age class.

Figure 6. Comparison of hourly grunt proportions when an individual is in dominant or subdominant positions. 2019 and 2020 represent the year the data were gathered. 1.5-5.5 represent the age of the reindeer, and 1-3 represent different individuals of the same age class.

Figure 7 A & B. Estimated smooth curves of the 24-hour rut patterns of seven reindeer according to their acoustic recordings represented as grunting behaviour taken as a proportion of time for each hour comparing (**A**) subdominant behaviour to (**B**) dominant behaviour. Black lines represent the mean, and shaded areas represent the 95% confidence interval. Models were fit using ten basis functions, and edf represents effective degrees of freedom**.**

List of Tables

Introduction

In 1859, Charles Darwin stated that sexual selection is a driving force in species evolution. He noted that traits are selected for during an individual's reproduction if they improve fitness, even if they hinder their survival (Darwin, 1859, 1871). Sexual selection acts on many traits, including exotic colouration, acoustic signals, and other sexually dimorphic traits (Andersson, 1994; Baker & Parker, 1979; Bowyer et al., 2020; Cassini, 2020; Charlton et al., 2007; Cooper et al., 2016; Darwin, 1859, 1871; Mukharji, 2021; Zahavi, 1975); however, acoustic signals are one of the more common reproduction traits and are prevalent in anurans, insects, and birds (Andersson, 1994). These signals are used for mate assessments, territory defence, and to ward off competitors (Andersson, 1994; Arak, 1983; Boake & Capranica, 1982; Garcia et al., 2013; Krebs et al., 1978; Mitani, 1985; Vannoni & McElligott, 2008; Wyman et al., 2012).

Individuals use acoustic signals to convey information and regulate interactions (Bradbury & Vehrencamp, 2011). For example, the emitter can convey information for danger avoidance, peer recognition, social learning and mating (Bradbury & Vehrencamp, 2011). However, understanding the contents of these vocalizations has been an ongoing challenge for the field of bioacoustics, though relating our understanding of behavioural ecology to bioacoustics has helped surmount these challenges. Consequently, researchers have been able to explore new aspects of animal behaviour using the sounds they produce. However, the field of bioacoustics is still nascent. Many issues within the literature and emerging technologies still exist, as deriving information from acoustic signals has proven rather difficult; thus, integrating the fields of bioacoustics and behavioural biology has been a slow process (e.g., Charlton et al., 2007; Stowell et al., 2017; Garcia et al., 2013; Studd et al., 2021). Nevertheless, researchers have recently begun using animal

sounds to study their occupancy, behaviour and ecology (Blumstein et al., 2011; Browning et al., 2017; Enari et al., 2017, 2019; Rognan et al., 2012; Stein, 2011).

Broadly, bioacoustics researchers study a range of aspects of animal ecology and behaviour and, across ecosystems, use sound to analyze community richness and demography (Gammon et al., 2005; Habib et al., 2007; Laiolo, 2010; Laiolo & Tella, 2007; McComb et al., 2001; Wimmer et al., 2013). Formerly, in-person observations were necessary to record the sounds emitted by individuals; however, autonomous recording units (ARUs) have begun to emerge (Shonfield & Bayne, 2017; Sousa-Lima et al., 2013; Wimmer et al., 2013). ARUs are favoured over traditional methods because they record continuously, allow researchers to capture a greater degree of a species' behaviour, remove the human influence, and allow for sampling over greater temporal and spatial scales (Darras et al., 2019; Shonfield & Bayne, 2017; Zwart et al., 2014). Moreover, in addition to gleaning biological information, techniques are being developed to document species distribution, abundance, and biodiversity (Blumstein et al., 2011; Enari et al., 2017, 2019; Johnson & Bayne, 2022; Stein, 2011; Zwart et al., 2014). However, ARUs have several drawbacks, as they are stationary and record passively (Fairbrass et al., 2017; Farina et al., 2011; Shonfield & Bayne, 2017). Consequently, individuals may not be near a recorder when they vocalize. Moreover, high environmental noise may mask vocalizations, and the environments housing the recorders may erratically affect their recordings (Fairbrass et al., 2017; Farina et al., 2011; Shonfield & Bayne, 2017). Stationary acoustic monitoring applications often struggle to capture behaviour relevant to the ecology and behaviour of individuals due to their movement. Consequently, researchers have recently started using on-animal acoustic recorders to study their behaviour and demography.

Commonly used in marine environments (Casoli et al., 2022; Holt et al., 2011; Johnson & Tyack, 2003; Silva et al., 2016; Stimpert et al., 2011), recent efforts have been made to use on-

animal acoustic recorders in terrestrial environments as well. Though these recorders have several drawbacks, their primary benefit is that they continuously record an individual's intentional and unintentional sounds (Shonfield & Bayne, 2017). Unfortunately, recorders are often heavy due to power requirements and challenging to retrieve. Additionally, because no commercially available recorders are produced for the terrestrial environment, custom recorders are often required (Cvikel et al., 2015; Ilany et al., 2013; Lynch et al., 2013; Stowell et al., 2017; Studd et al., 2021; Thiebault et al., 2021; Wijers et al., 2018). Moreover, recorders are likely to be damaged by the animals wearing them or by their environment. Consequently, recording quality might change drastically over temporal and spatial scales and between individuals. Although the number of studies to employ on-animal acoustic recorders within the terrestrial environment remains low, preliminary studies have demonstrated promising results. As they also record unintentional sounds, researchers have used on-animal recorders for event/behaviour classification (Stowell et al., 2017; Wijers et al., 2018), to estimate the metabolic cost of sound production (Ilany et al., 2013) and to analyze feeding behaviour (Lynch et al., 2013; Studd et al., 2019, 2021; Thiebault et al., 2021). However, to date, no researchers have used on-animal recorders to study the reproductive activity of terrestrial species.

Because of the amount of data these applications produce, processing and analyzing data remains a significant hurdle for many researchers. In recent years, researchers have begun to collaborate with computer scientists to apply machine learning to the field of bioacoustics (Blumstein et al., 2011; Dufourq et al., 2021; Enari et al., 2019; Mcloughlin et al., 2019; Stowell et al., 2017; Studd et al., 2021; Thiebault et al., 2021; Wijers et al., 2018). These techniques decrease the time required to process and identify features of interest within data sets. Further, tremendous progress has been made using convolutional neural networks (CNNs; Dufourq et al.,

2021, 2022). CNNs are a type of neural network (a type of machine learning that learns to recognize patterns in large amounts of data through the use of layered, interconnected nodes that mimic the structure of the human brain) used for image recognition/classification (refer to O'Shea & Nash, 2015 for more info). The convolutional part of CNN refers to the transformation of images to a grid of mathematical operations/values, which a neural network can interpret (O'Shea & Nash, 2015; Yamashita et al., 2018). The neural network part of CNNs then learns to group similar feature sets from annotated data to recognize and classify similar feature sets in new data using the processed images (Bermant et al., 2019; Zhong et al., 2020).

When used for bioacoustics research, sounds are transformed into spectrograms, which the CNNs then learn to interpret. To date, several commercially available vocalization classification software tools have been produced. However, compared to CNNs, commercially available classifiers classify vocalizations with poorer accuracy (Knight et al., 2017). Yet, creating and training a CNN is a non-trivial task. Researchers need to choose a suitable neural network architecture, a problem, as there are no clear guidelines for applications concerning bioacoustics (Dufourq et al., 2022). Consequently, the development of CNNs lies primarily with researchers who have knowledge of machine learning. Secondly, hyper-parameter (these are parameters chosen by the practitioner and are set before the model training process) tuning is an arduous step wherein significant time is spent developing an optimal model (Dufourq et al., 2022). As a result, these factors make creating an effective CNN complicated. However, within the machine learning field, these issues have been addressed using transfer learning (Pan & Yang, 2010; Weiss et al., 2016).

Transfer learning uses a model trained on one data set to predict a feature of interest on another (Dufourq et al., 2022; Zhong et al., 2020). Because pre-trained models are loaded rather

than built, it simplifies network development and decreases hyper-parameter tuning (Dufourq et al., 2022; Pan & Yang, 2010; Weiss et al., 2016; Zhong et al., 2020). Further, because a pre-trained model is used, training on a user's feature set requires fewer manually annotated vocalizations. Thus, it can accelerate the development of an effective CNN (Dufourq et al., 2022; Zhong et al., 2020). This is because model development without transfer learning typically requires an extensive database of images to avoid overfitting and to ensure sufficient precision and accuracy (Dufourq et al., 2022; Pan & Yang, 2010; Weiss et al., 2016).

In the field of bioacoustics, transfer learning is relatively understudied, though a recent study has already delivered promising results, with high performance metrics for several pre-trained models and data sets using a relatively simple architecture (Dufourq et al., 2022). Consequently, transfer learning will likely simplify the development and training of CNNs for machine learning experts and non-experts. Unfortunately, the aforementioned models were developed for stationary recorders, and few machine learning models have been used to classify the vocalizations of terrestrial species wearing recorders (Casoli et al., 2022; Stowell et al., 2017; Studd et al., 2021; Thiebault et al., 2021; Wijers et al., 2018).

Consequently, several issues arise when pairing machine learning applications with on-animal recorders. Because recorders are easily damaged, recording quality can vary drastically between individuals and environments, over time and with varying weather and because of the increased likelihood of damage, recorders are likely to fail during field deployment. Moreover, because on-animal recorders record the environment of the equipped animal, classifier performance often suffers (Bravo Sanchez et al., 2021; Lostanlen et al., 2019). As a result, copious example noises are required to develop an accurate network. But despite these drawbacks, there are numerous benefits to modelling species' behaviour using machine learning and on-animal

recorders. Recording the intentional and unintentional noises emitted by an animal allows researchers to study a range of topics concerning the behavioural ecology of a target species (Ilany et al., 2013; Lynch et al., 2013; Stowell et al., 2017; Studd et al., 2021; Thiebault et al., 2021; Wijers et al., 2018). Until now, though, no on-animal acoustic recorders have been used to study the mating activity of terrestrial species.

In the context of reproduction, vocalizations play an important role in regulating sexual selection. Among many terrestrial species, these vocalizations are well-developed and aid in guiding interactions (Vannoni et al., 2005). Among Cervidae who organize into herds, social hierarchy is guided by communication (Vannoni et al., 2005; Vannoni & McElligott, 2008). Consequently, in gregarious species like red deer (*Cervus elaphus*; e.g., Charlton et al., 2007; Garcia et al., 2013) and reindeer (*Rangifer tarandus*; e.g., Espmark, 1964; Frey et al., 2007), their vocalizations are highly developed and help regulate sexual selection (Reby & McComb, 2003b). In this context, calls are used for mate selection, male-to-male competition and territory defence (Reby & McComb, 2003b). Among Cervidae, several researchers have studied these calls (Charlton et al., 2007; Charlton & Reby, 2011; Feighny et al., 2006; Liu et al., 2016; Reby & McComb, 2003a; Volodin et al., 2015; Yen et al., 2013); however, there has been little research focusing on the vocalizations of reindeer. Compared to other Cervidae, the acoustic repertoire of reindeer is small (Lent, 1975). Females and offspring use vocalizations for peer recognition, and males use vocalizations during the rut to establish harems (Espmark, 1964, 1971). In harems, males grunt, which are a series of short calls produced by several exhalations, to antagonize other males and court mates (Espmark, 1964; Frey et al., 2007). During the rut, males compete for control over harems, with heavier/dominant males gaining access to more females than smaller/subdominant

males (Espmark, 1964; L'Italien et al., 2012). Thus, when used as a proxy for activity, vocalizations can be used to study the mating activity of male reindeer.

To study their mating activity, on-animal acoustic recorders and machine learning will prove versatile and suitable for the application. Because our study seeks to describe the presence and absence of calls within long, continuous recordings, this study represents an essential preliminary step in demonstrating the utility of custom, on-animal recorders in a field setting in tandem with machine learning and transfer learning. As a result, this study represents a relatively simple application of machine learning, and with the use of transfer learning, it may ease the uptake of machine learning applications by non-machine learning experts. It may also demonstrate the utility of on-animal recorders, possibly aiding in the development of purpose-built tools. Additionally, because the audio recorded from on-animal recorders is inundated with unintentional noises, creating an effective CNN might prove difficult. To date, few studies have attempted to classify on-animal acoustic recordings using CNNs; thus, testing and identifying ongoing issues is essential for developing emerging technologies. Finally, this study will act as an important indicator of the utility of the previously mentioned tools for the application of documenting the behaviour of numerous individuals over a relatively long temporal scale. Consequently, it will demonstrate the utility of emerging technologies for studying existing fields, thus aiding the collaboration of computer scientists and biologists.

Therefore, this study seeks to 1) Evaluate the efficacy of on-animal recorders for studying the rutting activity of male reindeer and identify ongoing issues; 2) Develop and train a convolutional neural network using transfer learning to identify reindeer vocalizations within a series of continuous recordings captured during their rut; with performance metrics of at least 90%; and 3) Describe the activity patterns of male reindeer during their rut to document how their

grunting activity varies by day and hour. Based on past studies, which have noted that older, heavier males are dominant to younger, lighter males, we hypothesize that male reindeer status, age and weight will play an important role in regulating grunting activity. Thus, we predict that older, heavier males will spend more time grunting than younger, lighter males, as they will be subdominant to the older, heavier. Further, we predict that older, heavier males will be dominant for a greater portion of time than younger, lighter males.

As we develop and train our CNNs, we will train two models across our data sets. We will inspect the recordings manually for the first CNN and annotate them accordingly. Then, once we have our first model, we will use our second data set to demonstrate the ease with which a second network can be trained on a similar group once a preliminary network is produced.

Methods

i.   Field site and focal individuals:

The bioacoustics data were captured at the Kutuharju research site in Kaamanen, Finland (69.1° N, 27.2°E). There, a semi-domesticated reindeer herd is kept within a 45 km^2 enclosure, consisting of about 100 individuals (including calves, females, and males). The herd has been studied continuously since 1969. During the rut (mid-September to late October), the herd is brought into a smaller pen (Lauluvaara ~ 13.8 km^2). During this translocation, males were weighed and fitted with collars with VHF locators (allowing for telemetry tracking) and acoustics recorders (used to record our males' intentional and unintentional sounds). Females were fitted with numbered coloured collars, which allowed for individual identification. All animals were of known age and individually recognizable due to long-term bookkeeping. Moreover, parentage analysis was done on offspring so that the reproductive success of each male was known; however, note that for the 2020 sampling period, the offspring born the following spring had not yet had their parentage analyzed.

Bioacoustics data were gathered during the 2019 and 2020 ruts. In 2019, we collected vocalizations from two males; in 2020, we collected vocalizations from six males. Due to equipment issues, however, the sampling time varied from three days to two months. Furthermore, the number of recorders we could gather data from was limited due to technical failures. As a result, we only collected data from seven individuals (two from 2019 and five from 2020). The individual characteristics of each male are listed in **Table 1**. We had males across five different age classes (1.5 – 5.5 years old) with varying weights. Finally, reproductive success notes the number of offspring born to a male the following spring.

9

ii.   Acoustic recorders and acoustic analysis:

During the males' translocation to Lauluvaara, they were outfitted with on-animal acoustic recorders. These recorders contained recording devices (SOROKA-15E, TS-Market Ltd., Zelenograd, Russia; amplitude resolution: 16 bits, sampling rate: 16 kHz; **Figure 1**) to collect the animals' vocalizations. These recorders provided us with continuous recordings throughout the rut. To store the recordings, each recorder was outfitted with a 256-gigabyte microSD card, which was capable of recording over 92 days of audio. Finally, each recorder contained a 9000-milliamp hour 3.6-volt lithium-ion battery, which could power the recorders for over two months, notwithstanding damage.

The housing containing the recorders was 3D printed by TS-Market Ltd. Before putting the recorders on the animals, the batteries were attached to the recorders and tested, and the two halves of the housing were attached with silicon to prevent water from affecting the recorders and screwed together. Then, the recorders were attached to each animal's collar using silicon, metal hose clamps and a rubber sheath for protection.

Each male's status and social rank within the group were documented by observers throughout the ruts. During these sampling events, the size of the male's harem was noted; if he was in a dominant position (i.e., in control of a harem), he was labelled as such, and if he was not, they were labelled subdominant (i.e., not in control of a harem). Throughout each season, the social hierarchy of the males was predicted through observations and agonistic interactions. Further, if there were no observations documenting the status of a male (as sometimes males could not be located for a few days), attempts were made during recording playback to comment on the male's status. For example, if females and calves could be heard in the background of a male's recorder and he was actively grunting, we took that as an indication that a male was dominant. Similarly, if

a competing male could be heard grunting towards a male whose recorder was being played, and he was not responding, then that would be taken as an indication that the male we were listening to was subdominant. Lastly, if a male was being listened to, but no other individuals could be heard, and he was not grunting, then that would be taken as an indication that the male was searching for a herd and was subdominant. However, if the status of a male proved difficult to determine during playback, and no notes were available, then this status was labelled as unknown until such a time as his status could be determined.

We used Sonic Visualiser (Cannam et al., 2010) to analyze the recordings and for annotation. Every recording was manually annotated using the "boxes layer" feature, allowing us to draw bounding boxes around intentional and unintentional noises. Thus, each sound's start and end times were noted within the recordings, along with each call's minimum and maximum frequency. Each bounding box was also given a binary label, presence or absence. To label the presence class, bounding boxes were placed around single and series vocalizations to capture the grunting behaviours of the reindeer (**Figure 2**). The grunts of the animals wearing the recorders were annotated. However, to avoid over-sampling the activity of the focal individuals, the incidental calls of other reindeer were not annotated to not affect the training of the CNN. Bounding boxes of varying lengths were randomly placed throughout each recording to annotate the absence class. To thoroughly train the CNN, we captured a wide range of unintentional sounds (including sounds from biological, natural, and anthropogenic sources; **Figure 2**). As noted by Dufourq et al. (2022), an important preliminary step is to annotate the vocalizations of other species to prevent false positive classifications made by the network later in the process. Consequently, as part of our absence class, we took special care to annotate the vocalizations of any other species whose frequency range overlapped with our males' range. If the frequency of an

unintentional sound was outside of our sampling range, then that sound was not annotated. Sonic Visualiser was used later in the process to confirm the vocalization classifications made by our CNN and to remove false positive or false negative classifications.

iii.   Machine Learning methodology and process:

We trained our CNNs using code and methodology adapted from Dufourq et al. (2022). We thus used a supervised learning method to train our networks. Firstly, we started by annotating 150 audio segments from our 2020 recordings. We annotated 25 recordings from six individuals from the 2020 field season, amounting to 150 recordings. However, due to technical issues with one of the recorders, we did not include the annotations from this individual; when we included the recordings from this individual, the performance of the networks suffered. Due to computer hardware limitations, we used a subset of the remaining 125 annotated audio segments. Recordings were picked to get a range of recordings across individuals from as broad a temporal scale as possible from various environments and weather conditions. As a result, our 2020 training set ended up containing 8605 presence segments (augmented to 14000 during the training process [this was done via time-shifting existing presence annotations]) and 18000 absence segments (of which 14000 were randomly sampled) to create a balanced data set. These two sets of spectrograms were used to train the 10 and 50-epoch 2020 CNNs (an epoch is one complete pass over a set of training files, if a model is being trained over 10 epochs, this means that the model will adjust classification parameters over 10 iterations; **Figure A1 A-D)**.

After annotating the audio recordings, we randomly searched for the hyper-parameters that would allow us to train the most effective CNN. After trying over 40 hyper-parameter combinations, we found that the values in **Table 2** gave us the best-performing network. As CNNs require a fixed input and as the length of multiple grunts can vary drastically, with a series of

grunts lasting from a fraction of a second to over 30, windows longer than a few seconds did not perform any better than those that were only four seconds. As grunts vary little during a series of grunts, window lengths that were more than four seconds increased computation time with no performance gains. The other characteristics we focused on were minimum and maximum call frequency values. As the frequency of reindeer grunts begins below 100 Hz, our minimum frequency was set to zero Hz. For the maximum frequency, we set our frequency to 4000 Hz. While the vocalizations of reindeer are often indiscernible after 2500 Hz, preliminary results indicated that increasing the maximum frequency to 4000 Hz improved network performance. However, setting the maximum frequency above 4000 Hz did not improve network performance and increased computation time.

Four pre-processing steps were conducted to create the inputs for the CNNs, the same as those performed in Dufourq et al. (2022). Firstly, a low pass filter (this acts as a filter that only passes signals below its cut-off frequency while attenuating all signals above it) was applied to each audio file. This reduces aliasing artifacts (these are parts of an image that are incorrectly reconstructed after being downsampled, which may result in parts of an image being distorted, which can affect the interpretation of an image), which can occur during downsampling (this is the process whereby the sampling rate of a recording is decreased; Dufourq et al., 2022). The cut-off rate associated with the filter was selected based on the maximum frequency of the males' vocalizations within our presence class. Within Sonic Visualizer, the grunts of our males were indiscernible after about 1000 Hz, although some of the louder calls were discernable after 1000 Hz, these examples were rare; hence, the low pass filter cut-off was set to 1000 Hz.

Next, each audio file was downsampled to reduce computation times, as higher frequencies were unnecessary. Within our recordings, sampling rates above 4000 Hz did not improve network

performance and only increased computation requirements. Hence, the Nyquist rate (or Nyquist frequency, which is the minimum rate at which a signal needs to be sampled to retain all the acoustic information; Landau, 1967) was set to 4000 Hz and the downsampling rate to twice that of the Nyquist rate (Dufourq et al., 2022). Then, the annotations for both annotation classes were extracted from each. This operates on a sliding window approach. As each segment window is four seconds, a sliding window would begin at a bounding box's initial time. This would then produce a four-second spectrogram. The four-second segment would then slide one second forward in time and extract another four-second segment. This process would repeat until the four-second window overlapped with the end time of the bounding box. The spectrograms extracted would then be labelled with the same binary label as the bounding box (either presence or absence). This process would then be repeated for each bounding box within an entire audio file and across all the audio files. Finally, audio segments were converted into two-dimensional mel-frequency spectrograms (a spectrogram where the frequencies are converted to the mel scale, which is a perceptual scale of pitches judged by listeners to be equal in distance from one another; Stevens et al., 2005) with the associated values found in **Table 2**. **Figure 3** illustrates the pre-processing steps.

During the training process, we tested several of the pre-trained models listed in Dufourq et al. (2022); the ResNet152V2 (He et al., 2016) model had the highest performance metrics. The feature extractor and output layer were fine-tuned to produce as accurate a model as possible. Although it increased the computation time, it increased the networks' classification performance. Like the study conducted by Dufourq et al. (2022) and knowing that the pre-trained models required a three-channel input (often corresponding to the three channels in a colour [RGB] image), we used the exponent method described in the above study. Channel one corresponds to the original spectrogram, whereas channels two and three represent exponential values of the

14

original spectrogram's density values (as transformed onto a normalized scale). This works by applying an exponent to a normalized spectrogram (which has values between 0 and 1); thus, parts of the spectrogram with little noise will have their values decrease, while louder areas will still be visible (Dufourq et al., 2022). While developing our networks, the exponent values that produced the most accurate and precise networks were S^1, S^3 and S^5. This is perhaps due to the high amount of wind within our recordings. Consequently, these exponential values likely produced a series of spectrograms where grunts were more distinguishable from other noises.

For testing, the CNNs predicted two softmax outputs (the final layers of a neural network that convert real number vectors into probability distributions, which then determine an image's final classification class) on each spectrogram within an entire testing file. The final class (presence or absence) was assigned based on the softmax output with a value greater than 0.5. Each file was predicted by using a sliding window approach. The window length was four seconds, which shifted one second at a time until the network predicted an entire recording. Two models were trained on the same 2020 data set, one over ten epochs and another over 50, to see how training over additional epochs might affect the performance of the first CNN. Appendix **Figure A1** demonstrates the difference between the two models. Note that the validation loss values increase over epochs to over one. This is often a sign of overfitting, but even when testing the performance of the networks on individuals it was not trained on, they often had performance metrics of over 90%. This may result from the similarity between grunts or an inherent issue with transfer learning. Whatever the issue, it appeared not to affect the performance of our networks, and to find the cause of this issue, the problem will likely require further attention.

After training the initial networks, we used the 50-epoch network to collect vocalizations from the 2019 recordings. In total, we collected presence annotations from 32 of our 2019

recordings. These annotations were manually verified to ensure accuracy and remove false positive classifications. Absence annotations were also added to provide the network with additional training material. Then, we used a subset of our 2020 annotations and all those from our 2019 recordings to train another CNN. This secondary CNN was trained upon 10778 presence and 11546 absence annotations but, based on the optimization patterns displayed in **Figure A1**, was only trained for 25 epochs.

We used the same 12 audio files to test the networks' performance. These files were not used to train the networks and represented individuals from across the rut and various rutting behaviours. For example, some files contained almost no vocalizations, while others contained over 250. Eight of the twelve recordings originated from the 2020 data, while four originated from the 2019 data. For the networks trained on the 2020 data, the 2020 verification files were labelled as 'trained' under the audio file subset category and the 2019 recordings as 'untrained' (**Table 3**). These untrained files were used to gauge the networks' performance in classifying vocalizations in recordings that did not contain the vocalizations of individuals used to train the networks. All the files were labelled 'trained' for the network trained on both years' recordings under the audio file subset category (**Table 3**).

To measure the performance of the networks, we used recall rate, precision, accuracy and F1 scores (**Equation A 1-4**). Recall rate (sensitivity) indicates how well a model detects sounds of interest (the proportions of true positives to actual positives; Do Nascimento et al., 2021; Mesaros et al., 2016; Navarro et al., 2017). Precision indicates the detector's reliability (what proportion of positive identifications were correct; Do Nascimento et al., 2021; Mesaros et al., 2016; Navarro et al., 2017). Accuracy measures the proportion of classifications a model correctly classifies (Navarro et al., 2017). Finally, F1 score measures the harmonic precision and accuracy mean

(Navarro et al., 2017). Within the equations, classifications are classified into three categories. True positives (TP) occur when a grunt presence is correctly classified, false negatives (FN) occur when an absence of a grunt is classified in a segment that contains a grunt, and false positives (FP) occur when a grunt is classified in a segment when there is no grunt (Nascimento et al., 2021). Of these classifications, we want as many true positive classifications as possible and as few false negatives and false positives. The corresponding performance metrics are listed in **Table 3**.

To train the networks, we used the packages listed in **Table A1**. The script was processed using Python 3, and the CNNs were implemented in Tensorflow 2 (Abadi et al., 2016). Each CNN was trained over several epochs using the Adam optimizer with a batch size of 32 (Kingma & Ba, 2014). Finally, spectrograms were generated using the Librosa library (McFee et al., 2020). Model training and testing were conducted on a 2021 Apple Macbook Pro with an Apple M1 pro processor with 16 GB of unified LPDDR5 RAM running MacOS Ventura 13.1.

iv.   Data analysis:

We conducted statistical analysis in R v. 4.1.3 (R Core Team, 2018). To evaluate the hourly and daily patterns of our male reindeer over the 2019 and 2020 ruts, we used hierarchical generalized additive models (HGAMs, using the 'gam' function in the R package mgcv version 1.8-41; Pedersen et al., 2019; Wood, 2017). GAMs are a nonparametric technique that uses penalized regression splines to fit smooth relationships between independent and dependent variables. Thus, they can describe time series data's nonlinear trends and relationships.

The vocalizations classified by our CNNs were visually and audibly verified, and then the window lengths (seconds) of presence annotations (vocalization detections) were summed by the hour of the day. The hours of the day were based on the recorders' memory and verified using GPS data and audio from when the reindeer were released into the pen. Then, we compared the years

using their estimated peak rut (referred to as peak hereafter). We estimated this date by averaging the births of calves in the spring and backdating using a gestation length of 221 days (Mysterud et al., 2009). The peaks for the 2020 and 2019 cohorts were estimated as October 6th and October 5th, respectively. We also tested for the effect of temperature (degrees Celsius) and weather (sunny, overcast, snow and rain) on the activity of our reindeer. The data were captured by a weather station near Kaamanen, Finland (station code 102047, 69.14° N, 27.27°E). We tested for the interaction of weather and temperature on daily and hourly patterns using HGAMS; however, no discernable patterns were present for either temporal variable that helped to describe their rutting activity. Consequently, temperature and weather did not help to explain grunting patterns and were thus excluded from our models.

Smoother terms were used to compare the patterns of the proportion of time spent grunting across days and hours in terms of individuals (**Equation 1** [i indicates each variable is calculated for each individual]; Zuur & Ieno, 2016).

($\boldsymbol{Equation\ 1}$) $Grunt\ proportion\ g_i(t) \sim Beta(\mu_i(t), \theta)$

$$logit(\mu_i(t)) = f_i(Day) + g_i(Hour) + re(Individual) + Status_i$$

The data were fit using the beta regression family (Wood, 2017) with a logit link function and restricted maximum likelihood, and zero values were offset by 2.2×10^{-14} to avoid having a complete separation with the data (complete separation, or quasi-complete separation, occurs when an outcome variable separates a predictor variable or set of predictor variables completely or nearly completely; Albert & Anderson, 1984). Status was modelled as a categorical variable with two levels: dominant and subdominant (**Equation 1**). The hourly patterns (**Equation 1**) were fit using cyclic cubic regression splines due to the repeating nature of 24-hour days, and to describe the pattern more generally, the splines were fit with ten basis functions (basis functions being the

number of functions each smoother is allowed to use in order to fit a set of data; Wood, 2017). The daily patterns (**Equation 1**) were fit using thin plate regression splines, and to allow for considerable flexibility in describing the pattern, the splines were fit with 22 basis functions. Finally, the smoothers fit by individual (**Equation 1**) were treated as random effects (intercepts) to account for inter-individual variability in the average frequency of vocalizations. Both temporal patterns were fit using individual group-level smoothers with differing wiggliness to describe the pattern of each reindeer (Pedersen et al., 2019). Finally, trends were fit using individual smoothers to describe the behaviour of the reindeer, and 95% confidence intervals were used to describe the uncertainty in the shapes of the estimated functions (Wood, 2017).

To describe the grunting activity of the reindeer, proportions were transformed into time as a proportion of each hour and when values were reported, we did so as a mean and standard error (mean ± SE) and to determine if dominant and subdominant statuses were significantly different, we used a Wilcoxon rank sum test. The HGAM models' performance, fit and autocorrelation were assessed using the 'appraise' function in gratia (Simpson, 2021) and the 'ACF' function in mgcv (Wood, 2017). Finally, models were chosen according to model fit and the Akaike information criterion (Wood, 2017).

Lastly, to document the resting activity of our males, three days for each male were analyzed randomly to document the time males spent resting each day. This could not be analyzed via network classifications as our classifier was trained to analyze for the presence of vocalizations, so their resting activity had to be manually analyzed via recorder playback.

Results

i.   Machine Learning Performance:

Due to the number of spectrograms our 'combined' network was trained upon, it was quicker to train per epoch than the '2020' networks; the output layers took about 425 and 550 seconds to train, respectively, while the feature extraction layers took about 1260 and 1610 seconds. Both networks took about the same time to process a single recording (2.27 hours long), about 225 seconds. However, while all three of our networks performed similarly, they had slightly different classification characteristics. Overall, annotating and training the 2020 data set took about four weeks (150 2.27-hour recordings), while classifying, verifying, and training the 2019 recordings took about three days (32 2.27-hour recordings). Consequently, establishing the initial data set took longer than collecting the annotations for the 2020 network.

When we compare the networks trained on the 2020 data, the 50-epoch network was slightly more accurate and classified fewer false negatives but slightly more false positives on files that contained individuals it was trained on (**Table 3**). When we compare the performance metrics of vocalization classifications on files that contained individuals not included within the 2020 training batch, the network trained over 50 epochs performed better than that trained over ten epochs. Consequently, the 50-epoch network was better at classifying and collecting vocalizations of untrained individuals than the 10-epoch network. Thus, it would be a better tool for researchers wanting to collect additional vocalizations. Overall, the 50-epoch network was more sensitive but more likely to identify vocalizations falsely (**Table 3**). Consequently, this network will require the removal of more false positives but require the addition of fewer false negatives. Both networks performed better when the recordings contained more intentional noises. For example, when we compare recordings containing 200 reindeer grunts versus 20, both had comparable false positive

and false negative counts. However, in the recordings containing 20 vocalizations, these false readings disproportionately affected the performance metrics.

When we compare the 50-epoch model to the 'combined' 25-epoch model, the 25-epoch model's performance metrics were, on average, 4.8 ± 1.4% (mean ± SE) higher than the 50-epoch model (**Table 3**). While the recall rate, precision and F1 scores were only slightly higher for the 25-epoch model, the accuracy was markedly better (**Table 3**). Across the same 12 files, the 25-epoch network had five fewer false positives and 49 fewer false negative classifications. However, both networks had a disproportionate number of false positive and false negative classifications originating from a subset of the files. Moreover, some recorders had higher incidents of false positive and false negative classifications across all their recordings compared to some of the other individuals' audio recordings. Overall, when classifying upon recordings with individuals the networks were trained upon (i.e., 'trained' audio subsets), the network trained across both years' recordings performed better than those trained on the 2020 recordings (**Table 3**). Finally, training models for more than 25 epochs did not appear to have any noticeable effects; it only increased computation time (**Figure A1**).

ii.   Rutting Behaviour:

Between males, rutting activity varied by individual, with the trends suggesting larger males grunted more (**Table 1, Figure 4,** and **Figure 5**). When males were dominant and in control of a harem, they spent, on average, 14.2 ± 31.6 minutes per hour grunting more than subdominant males (HGAM; z-value = -10.68, *p* < 0.001, n=2213; **Figure 6**). The HGAM model accounted for 83.2% of the deviance within the data with an adjusted r-squared of 0.608. Among the males, the younger and older males had very distinct activity patterns, respective to their age groups (**Table

1, **Figure 4,** and **Figure 5**). However, among the 3.5-year-old males, age, weight, and activity patterns differed more by individual (**Figure 4** and **Figure 5**).

The 1.5- and 2.5-year-old males, who weighed 83 and 96 kilograms, respectively, spent almost none of their time in dominant positions, 0% and 3.5%. Conversely, the 4.5- and 5.5-year-old males, who weighed 150 and 140 kilograms, respectively, spent most of their time in dominant positions, 100% and 66.9%. There was more variation between the time spent vocalizing for the 3.5-year-old males and their time spent in dominant positions. Their pre-rut weights were 100, 135 and 155 kilograms. The lightest among them spent nearly no time in a dominant position, 4.0%, while the other two spent about half their time in dominant positions, 51.1% and 53.0%, respectively. However, these data were uncertain due to the recorders' inconsistent recording lengths. Furthermore, there was considerable variety in the grunting patterns of the males, even when dominant. When the activity of each male was divided by status, there was a significant difference between the proportion of time spent grunting in dominant and subdominant positions, 8.85 ± 0.19 minutes versus 1.14 ± 0.05 (Wilcoxon rank sum test, $W = 1137461$, $p < 0.001$; **Figure 6**).

For the hourly data, the males had several consistent patterns. Males were more active during the daylight hours, from about 9:00 to 21:00 (**Figure 4 A, C-G**), and most were less active overnight, with most being the least active just before dawn, at about 06:00 (**Figure 4 A, C-G**). The 1.5-year-old's hourly pattern was the only inconsistency; however, he was subdominant for the entirety of the rut (**Figure 4 B**).

Generally, based on the hourly patterns, males were more active during the day than at night (**Figure 4**). Though the males were more active during the day, there were still many hours when the males displayed next to no vocal activity (e.g., **Figure 4 A**). This led to the day hours

22

having larger 95% confidence intervals than the night hours (**Figure 4**). However, the males were still somewhat active overnight. On average, the males spent about 4.42 ± 1.81 hours at rest (n=7 [three days for each male were analyzed]) each day (**Table 4**), and each time they took a break, they spent about 21.2 ± 19.3 minutes at rest (**Table 4**). When the hourly patterns were subdivided into dominant and subdominant categories, there was only a slight difference between the trends (**Figure 7**). Like the males' hourly trends, the dominant and subdominant trends indicated that the males were more active during the day than at night (**Figure 7**). However, the activity of subdominant males was more consistent, while dominant males had increased activity patterns between the hours of 09:00 and 22:00 (**Figure 7**).

For the daily patterns, there was more variation among males (**Figure 5**). Older males (**Figure 5 A, E-G**) tended to have elevated activity trends compared to younger males (**Figure 5 B-D**). The oldest male (**Figure 5 A**) was dominant most of the rut and was more active before the peak than after. For dominant males, activity typically increased until the peak, slightly decreasing after (**Figure 6**). The second 3.5-year-old was the exception (**Figure 5 E**). He had an uptick in activity about a week before the 2020 peak, and then his activity declined for 2.5 days. The activity of the two remaining males (**Figure 5 F & G**) increased as they approached the peak. However, because their recorders failed prematurely, only so much information could be gleaned from their audio. Furthermore, because there was so much variation in activity from one hour to the next, there was considerable variability in the males' daily activity, especially among dominant males, as they approached the peak (e.g., **Figure 5 A**). Overall, as the peak of the rut approached, dominant males vocalized more than subdominant males (**Figure 5**).

Discussion

The activity patterns extracted from our male reindeer indicated variability between males and between hours and days of the rut. The data suggested a trend by age class, with older, heavier males vocalizing more than younger, lighter males and spending more time in dominant positions. However, this was only the case with some of our males, as displayed with the 3.5-year-old males. Nonetheless, the best predictor of acoustic activity was male status. Dominant males vocalized more than subdominant males. Despite that, there was still much uncertainty within the activity patterns, as all the recorders recorded for varying lengths. Consequently, not all the males' activity patterns may have been fully explored. Nevertheless, our results align with our predictions that older, heavier males will grunt more than younger, light males and spend more of the rut in a dominant position.

Overall, collecting acoustics data may only partially explain reindeer rut behaviour. While their acoustic activities can be taken as an inference of their activity, these activity patterns also need to be taken with some skepticism, as there are likely a host of behaviours that are not covered accurately with sound alone. However, by combining both on-animal acoustic recorders and machine learning applications, researchers can delve into behavioural questions that span greater spatial and temporal scales. With recent developments in machine learning, performance metrics continue to improve, and with the inclusion of methods like transfer learning, methodologies continue to simplify. Conversely, on-animal recorders are the greatest hindrance to performing long-term analysis. They are liable to failure, and their performance over longer time scales is inconsistent. Thus, improving their reliability must be an essential research component moving forward. However, there is still much promise in these emerging technologies; with improvements, they could prove to be incredible tools for researchers.

i.   *On-animal acoustic recorders*

This study, along with several other preliminary studies that used acoustic recorders to record the behaviour of species, have already demonstrated their utility (e.g., Casoli et al., 2022; Insley et al., 2008; Lynch et al., 2013; Stowell et al., 2017; Studd et al., 2019, 2021; Thiebault et al., 2021; Wijers et al., 2018). They remove the human impact on the species and allow researchers to study species over longer temporal scales. Plus, they allow researchers to interpret the activities of species through unintentional sounds. For example, Insley et al. (2008) reported they could differentiate between various behaviours of northern fur seals (*Callorhinus ursinus*) using on-animal recorders. We used a similar approach to document the resting activity of our male reindeer, and we could differentiate males at rest from those walking around or engaged in rut-associated activities.

With the appropriate expertise and observational data, researchers can often differentiate between the audible behaviours of their species. Lynch et al. (2013) listed several behaviours they documented using sounds captured from mule deer (*Odocoileus hemionus*). It should be noted, as mentioned by both Lynch et al. (2013) and Wijers et al. (2018), that while most behaviours are discernable using acoustics, some periods of ambiguous sound signals will likely be recorded, and depending on the objective of the study, may require observational data to confirm. In our case, it would have been difficult to discern the status of males with only sound. In some cases, inferences from other sound sources may make it challenging to determine behaviour (Wijers et al., 2018). These conclusions may be made from vocalizations emitted from the species or other species, anthropogenic sources, and environmental sources (Wijers et al., 2018). Consequently, pairing audio recordings with observations can help researchers confirm behaviour and build appropriate auditory profiles. Additionally, researchers using on-animal recorders should also consider the

behaviour of their species. If a species of interest is not sufficiently audible, it may not be suitable for acoustic analysis.

On their own, the inferences made from acoustic recorders are often limited, but when paired with observations and other tools, the limitations of recorders are often minimized. As context is vital, an essential first step should be to categorize acoustic behaviours using observations or cameras. However, after such a step is complete, it becomes far easier to classify the auditory behaviours of species using sound. When paired with other tools, acoustic analysis can be a powerful tool for researchers. For instance, Wijers et al. (2018) used acoustic recorders and bio-loggers to document the activity patterns of lions (*Panthera leo*), and they noted a high precision of activity classification using the two methods. Similarly, Studd et al. (2019) used accelerometers and audio recorders to document the kill rates and hunting behaviour of Canadian lynx (*Lynx canadensis*). In both cases, two devices allowed the researchers to make better inferences from their data and, in combination, allowed them to use acoustics data for ground-truth calibration. On their own, on-animal acoustic recorders provide researchers with a promising avenue to pursue new research and, for behavioural ecologists, a new lens through which to study their species of interest. However, there are still several issues that plague on-animal acoustic recorders.

As other researchers have reported, the failure rate of on-animal recorders is high, and because no commercial on-animal recorders have been produced, they often require significant troubleshooting, especially during initial deployments (Studd et al., 2021). Also, because the devices can be deployed for a significant amount of time, they can accumulate damage, and if they fail, it can be challenging to discern their cause of failure. Because of this, a recorder's audio quality can change throughout its deployment, and thus, the performance of a network is affected.

Further, the weight of these recorders' power packs often limits their deployment and thus, researchers must consider their use against the species they are studying. Consequently, research questions are relegated to larger or smaller species over shorter times. As such, the number of studies that have employed on-animal recorders is only a few, but interest in them is increasing (Casoli et al., 2022; Lynch et al., 2013; Stowell et al., 2017; Studd et al., 2019, 2021; Thiebault et al., 2021; Wijers et al., 2018).

Because of these issues, few studies have studied terrestrial species over longer temporal scales, but of the few that have, they have only deployed recorders for several days or weeks (Casoli et al., 2022; Lynch et al., 2013; Stowell et al., 2017; Studd et al., 2019, 2021; Thiebault et al., 2021; Wijers et al., 2018). Even our study was limited to only one month, and of our eight recorders, one had a malfunctioning microphone, another failed after three days, and of the eight, only one recorded for longer than 20 days. For our study, the recorders were constantly exposed to varying weather (rain, snow, slush, and wind) and kinetic damage. As such, the units' cause of failure could originate from several sources. However, if the listed issues can be rectified, on-animal recorders could be a valuable tool for researchers.

ii.   Machine learning methodology and transfer learning

As interest in machine learning methods increases, especially within the ecology realm, so does their performance. Likewise, as there continue to be advances in machine learning and speech recognition, there continue to be opportunities for bioacoustics research. However, as Dufourq et al. (2022) outlined, acoustic classifiers are proving to be powerful tools for researchers, but creating and training a model is challenging for machine learning novices.

Creating a CNN from scratch, optimizing it, and fine-tuning hyper-parameters is challenging; thus, they would be adopted by more users if they were easier to build and train

(Dufourq et al., 2022). However, when CNNs are created using transfer learning, hyper-parameter tuning and optimizing are easier, and fewer network design decisions are required (Dufourq et al., 2022). One of the most challenging choices a researcher may face is choosing a suitable pre-trained model. For us, ResNet152V2 (He et al., 2016) was the most appropriate model, but this may only be the case for some practitioners due to the stochastic nature of neural network classifiers. As Dufourq et al. (2022) found, some models perform better than others for some applications; consequently, practitioners must find an appropriate pre-trained model. Overall, when used in combination with recorders and when adequately trained, acoustic classifiers can significantly decrease the processing time of data sets, and when implemented with transfer learning, the process is made simpler.

One benefit of using transfer learning is that it decreases the number of examples needed to train a network. Dufourq et al. (2022) demonstrated that as few as 25 vocalization annotations could achieve an F1 score of 82%. This indicates that even with a few examples, practitioners can begin to automate the process of collecting additional vocalizations. However, with a wide range of individuals and environments, gathering as many annotations as possible is recommended to build a robust classifier (Dufourq et al., 2021, 2022).

This also raises some issues when paired with on-animal acoustic recorders. The quality of our recordings varied considerably between devices and individuals over time. As a result, to maintain the network's performance, we gathered vocalizations throughout the rut. Unfortunately, due to hardware limitations, we were limited with the number of spectrograms our CNN could be trained upon; thus, we could not capture every sound within the environment. Consequently, this led to errors in our classification and generated additional false positive classifications in some of our recordings. Studd et al. (2021) reported similar findings when they used an automated event

classification model to classify the hunting activity of Canadian lynx. Despite this, these issues can be rectified by manually verifying classifications. Accordingly, it is in a practitioner's best interest to collect as many presence and absence annotations as possible to improve the performance of a classifier across individuals, deployments, and recorders. This is especially true when using on-animal recorders, as they record across environments and individuals and are readily damaged.

As these methodologies continue to develop and additional practitioners adopt them, their performance will improve. Recent uses of machine learning for acoustic classification have achieved high performance metrics. For example, Studd et al. (2021) reported F1 scores of 79 to 90% for Canadian lynx feeding event classifications using CNNs. Additionally, using a Random Forest classification method, Wijers et al. (2018) achieved an average per-class behaviour classification rate of 98.5% with a combination of audio, accelerometers and magnetometers. Thus, as these technologies continue to develop and more practitioners adopt them, their performance and ease of use will increase. Altogether, when used for acoustic classification, classifiers can decrease the processing time of acoustic data sets. However, despite these considerable advantages, there are still several drawbacks of which practitioners should be aware.

While utilizing transfer learning can simplify the network fabrication process, having knowledge of a species and machine learning will help optimize networks and hyper-parameters. Creating or adapting a network for a user's application still requires knowledge of the two to create a methodology and select appropriate hyper-parameters. While transfer learning simplifies the process, pre-trained models may only be suitable for some applications, and practitioners may need to develop a model suitable for their data set. There will always be limitations when we marry numerous technologies, especially when considering on-animal recorders and machine learning.

The issues associated with using on-animal recorders are again brought up while considering machine learning. As recorders are liable to damage throughout a deployment, the performance of a classifier may be impacted the longer a recorder is deployed, especially over numerous deployments. Furthermore, the classification performance between individuals/recorders may differ as units are damaged or because individuals may have slight differences in acoustic characteristics. It is in a practitioner's best interest to add a collection of new annotations every deployment for new individuals and environments, as differences in a recorder's audioscape and individuals can affect the performance of a classifier. Even if a preliminary model is built, collecting additional vocalizations is far easier using a prior network trained for a similar application, as demonstrated by the time difference required to build our two networks. Our preliminary network took about four weeks to train, while it only took about three days to collect the vocalizations and train the second network. Once a classifier is built, updating a classifier over time becomes relatively easy. The most challenging part of using a classifier is acquiring the knowledge to build one suitable for a particular application.

Finally, while using a classifier can expedite the data processing phase, the process is yet to be reliable enough to be automated. As a result, classifications still require validation, and practitioners thus need to inspect their classifier's classifications to verify its performance. While machine learning accelerates the analysis of bioacoustics data, the process is not yet automated nor reliable enough to use without human validation. Despite these drawbacks, however, the advantages of using machine learning are numerous.

iii.   *Rut activity*

The activity patterns described herein illustrate the hourly and daily patterns of several age and weight classes of semi-domesticated reindeer. While our results do not disagree with previous

research, it describes their activity in better detail. Moreover, it is one of the first studies to try and illustrate the hourly activity pattern of reindeer. As described by past research (Body et al., 2014; Espmark, 1964; Mossing & Damber, 1981; Tennenhouse et al., 2012), the age and weight of male reindeer are the best predictors of their activity. These studies noted that the activity of older males is higher than younger males (Kojola, 1991; Mysterud et al., 2004; Skogland, 1989; Tennenhouse et al., 2012). However, as described by Tennenhouse et al. (2012), the activity of older males declines during the late rut compared to the early and peak. This corroborates this study's findings, as the activity of our males increased towards the peak and then declined following it. This is generally described as males having used up most of their energy reserves during the early and mid-rut (Leader-Williams, 1980; Tennenhouse et al., 2012). By the final weeks of the rut, older males have used up most of their energy reserves and thus cannot compete with other males for access to females (Bobek et al., 1990; Tennenhouse et al., 2012). This describes the decrease in daily activity seen in our 5.5-year-old male past the peak. This is also reflected in the pattern of his calves born the following spring. Of the 14 offspring he sired, 13 were conceived in the ten days preceding the 2019 peak. However, given our small sample size, especially with this fitness data, I would take this comparison with some skepticism.

Likewise, such patterns help describe the smaller males' patterns. The patterns of younger males are described as opportunistic, wherein they gain access to females whenever possible. Older males are more polygynous, have a better fighting ability, begin the rut earlier than younger males, and are more active during the peak (Espmark, 1964; Mysterud et al., 2008; Tennenhouse et al., 2012). In comparison, younger males are more active during pre-rut than peak-rut weeks (Tennenhouse et al., 2012). There are a couple of possible reasons for this activity pattern. First, during the peak, the reproductive output of older males is highest, so young males may avoid

pursuing females during this time to avoid the aggression of older males. Second, Singer and Zeigenfuss (2002) found that young Dall sheep (*Ovis dalli*) increase their mating activity when older males are absent or scarce. Thus, younger males may increase their activity when competition from older males is absent. Another explanation may be that younger males cannot time their reproductive output like older males (Mysterud et al., 2008; Tennenhouse et al., 2012). Tennenhouse et al. (2012) postulated that this mating naivety in young males might manifest as underdeveloped social rutting behaviour. Perhaps the timing of reproductive effort is a learned skill that improves with age (Tennenhouse et al., 2012). Whether it be learned or situational, the timing of these increased reproduction periods can be seen in the activity of our males as described by their vocal activity.

When males were dominant, they vocalized more often than when they were subdominant. Conversely, while young males were occasionally dominant, they spent more time in subdominant positions than heavier males. However, occasionally, smaller males became dominant and larger males subdominant. For instance, the 2.5-year-old became dominant six days before the peak, and one of the 3.5-year-olds became subdominant on the peak. In both cases, the activity of the two males significantly increased or decreased during these switching events. The 2.5-year-old took control of a small harem for a couple of days, and the 3.5-year-old was displaced by a larger male for four days. In each case, the males' status was a significant determinant of activity, with dominant males vocalizing more than subdominant males, and overall, dominant states coincided more with heavier, older males than younger, lighter males. Whatever the reason, the activity patterns of younger males were less than those of older males, and the exact mechanisms of this difference are still not completely understood.

However, in our acoustic data, this manifested differently in the two age groups. Younger males spent less time at rest than older males. While not engaged in rut-associated activities, they spent considerable time on the move, constantly walking, potentially searching for free-roaming females. Additionally, during audio playback, it was noted that younger males spent much time as satellite males, hanging around the periphery of larger harems, as the vocalizations of other males could be heard in the background of younger males' recorders. This could be seen as a learning mechanism for the young males so they could learn to be better competitors in the future, or it could be seen as opportunistic behaviour. Perhaps the younger males were not interested in confronting older males but instead were on the periphery in the event of herd fission. Furthermore, in addition to older males spending more time at rest, when they did rest, their rests were longer. For all age groups, males spent the most time at rest between 22:00 and 06:00, but there was no consistent time the males were at rest during these hours. While Espmark (1964) noted that most rut activities occurred at dusk and dawn, which disagrees with our findings, he also noted that the males partook in little to no rut activities overnight. Interestingly, while this part of the day was associated with decreased activity, our data suggests that males still participated in rut activities overnight.

While the males' hourly patterns were all somewhat similar, there were still slight differences between them. The early morning hours, from 00:00 to 06:00, were the males' least active, with activity increasing during the morning and early afternoon, then decreasing again during the later afternoon, 15:00 to 18:00, then increasing again towards dusk, 19:00 to 21:00, followed by decreased activity. The exception was with our 1.5-year-old but given that we only have 3.4 days of data for him during the early rut, this individual's pattern should be taken with some skepticism. Unlike the older, heavier males, the smaller males appeared more active towards

dusk during the latter part of the day. In contrast, our heavier males were more active throughout the day. Where Espmark (1964) noted that most copulations occur at dawn and dusk, especially the former, our study did not document much rutting behaviour during the early morning hours. However, our study did document the uptick in activity at dusk. Further, their activity decreased from 12:00 to 14:00, aligning with similar research on other Cervidae. Zeng et al. (2011) noted that male Eld's deer (*Rucervus eldii*) activity peaked during the day. They noted that males were willing to mate at any point of the day but that fewer copulation attempts were observed before 09:00 and between 12:00 and 14:00 (Zeng et al., 2011). They postulated that increased temperatures from 12:00 to 14:00 reduced the mating activity of their deer (Zeng et al., 2011). Similarly, Mellado et al. (2000) noted that buck goats (family: *Bovidae*) were most sexually active at milder temperatures—consequently, male reindeer may display similar activity patterns for similar reasons. Further, male Eld's deer mating attempts displayed an activity pattern similar to the vocal activity patterns of our males (Zeng et al., 2011). So, the activity of reindeer may be like that of other Cervidae and goats.

The hourly activity of younger males again took on an opportunistic approach, with their activity increasing later in the day, perhaps after older males had already expended much of their energy. However, this pattern differed among males; only the uptick of activity later in the day, compared to older, heavier males, was consistent. Likewise, the overnight activity of all our males differed by the individual. Unlike Espmark's (1964) remarks, wherein he said activity is comparatively rare during the night and reindeer rest during most of this time, we found that males were still somewhat active. While their activity was diminished, dominant males were still alert and ready to ward off potential competitors overnight. Similarly, subdominant males were also quite active overnight. While this was not reflected in their vocal activity, it was determined

34

through recorder playback that they often spent much of the night walking around. While we cannot know why they do this, it can be postulated that they may be looking for lone females or smaller harems that a less dominant male controls. However, there is much variability between days and individuals, so the smoothed terms may only partially depict differences between individuals and times of the rut. Due to the limitations of our data and the issues caused by our failed recorders, a complete depiction of the males' activity over the entirety of the rut could not be represented. Consequently, the patterns represented herein should be scrutinized and not taken as a complete depiction of each age or weight class. Future research should focus on depicting a more complete picture of the activity of each age class.

Considering the above, the activity patterns of reindeer depicted through on-animal acoustic recorders should be taken with some skepticism, as the vocal activity of reindeer may not entirely depict their rutting behaviour (Lynch et al., 2013; Wijers et al., 2018). It is often in a researcher's best interest to pair their vocal recordings with in-person observations and other means of data collection. Further, practitioners can only deploy a few recorders due to data collection and hardware limitations. As a result, they may only capture a partial image of their species' behaviour, especially across sexes, ages, and weights. Consequently, depicting a complete image of a species' behaviour may require more than audio. For our application, using only the vocalizations of male reindeer was enough, though our inferences are also somewhat limited. However, even with just audio, researchers can use their species' intentional and unintentional sounds to explore many questions concerning them. Nevertheless, as is the problem with audio, classifying and processing audio takes time and effort.

Also, due to the variability of unintentional sounds, using machine learning networks to recognize the unintentional sounds emitted by a species is difficult. For example, in our case, if

we had used a network to classify the sounds of two males fighting, the sound would have likely been confused with the sound of branches rattling against a pair of antlers, for instance, as they walked under a tree. Therefore, using networks to categorize unintentional sounds may be difficult. However, as technology advances, these issues may be rectified. For now, practitioners must manually verify and listen to their recordings to answer more complex questions. However, as demonstrated, machine learning and on-animal acoustic recorders are suitable technologies for simple applications, such as describing the vocal activity of male reindeer.

Conclusion

Based on our study's findings, our research suggests that reindeer vocalizations play an important role in regulating their rut behaviour. The trends suggest that males who grunt more are often older and heavier. Further, based on our limited reproductive data, those males who grunt more are in control of a harem and, thus, likely have more chances to copulate. Consequently, reindeer vocalizations likely play an important role in regulating sexual selection, and males use these vocalizations for a few reasons, from organizing their harem to antagonizing potential competitors. Subdominant males cannot fully compete with dominant males, which is reflected in their vocal activity, resulting in them vocalizing less than dominant males. This is also reflected in the general hierarchy of the males and the size of the males' harems, with larger males controlling more females than smaller males. Thus, because much of their rut activity is regulated through vocalizations, using on-animal acoustic recorders and machine learning to describe their vocal activity was a good application of the technology. However, there are still some limitations with the two technologies.
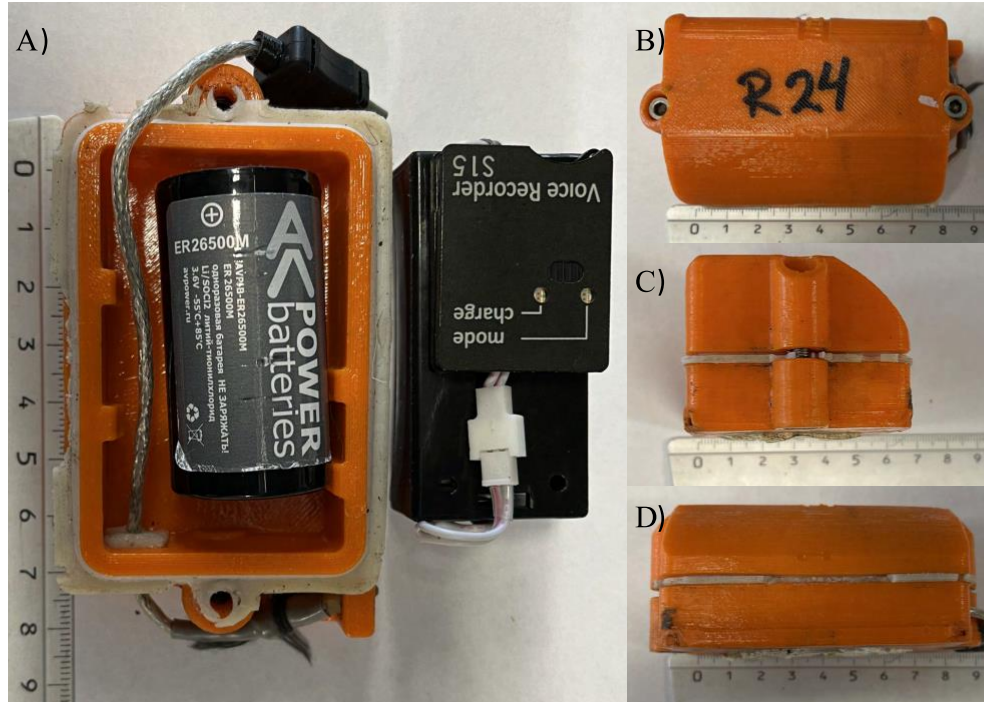
With machine learning and on-animal recorders, most issues are related to using on-animal recorders. This arises from their high rate of failure and reliability issues. Thus, future endeavours should strive to rectify their failure rates. A prefabricated, commercially produced on-animal recorder may be the best solution. Deploying commercial units would hopefully remove many of the reliability issues and increase the use of these technologies by additional practitioners, enabling researchers to have more trust in deploying recorders for longer. Moreover, creating a suitable commercial recorder might bring the added benefit of creating some ubiquity among researchers. If most of the recorders deployed by researchers are the same, with similar recording characteristics, it will help to create similar databases with comparable qualities. This could create

further collaboration between researchers and allow for the increased uptake of pre-trained networks for similar applications and similar species or for the sharing of databases to explore different research questions.

After collecting the recordings, neural networks are reliable enough to process large data sets. Further, when researchers use transfer learning to build their networks, the process is simplified, and the ease of training preliminary models is higher as they require fewer annotations. Thus, researchers with little machine learning knowledge can more efficiently train models; this enabled us to train a CNN to describe the reindeer activity with a high degree of accuracy.
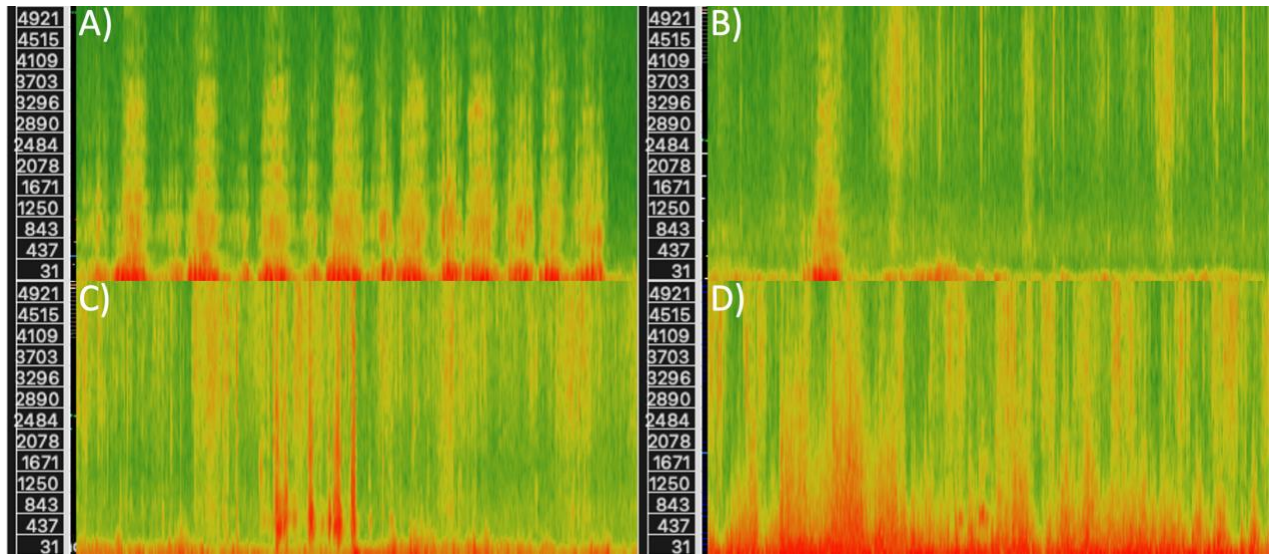
While researchers can trust machine learning for future studies, precautions must be taken when using on-animal acoustic recorders. Nevertheless, the two technologies hold much promise, and the breadth of applications they can be applied to increases as the technologies develop. For future studies, researchers should focus further on exploring the rutting behaviour of reindeer with larger sample sizes. Further, researchers using machine learning should focus on exploring the performance of models for classifying the unintentional noises of animals. As with such sound profiles, researchers can likely illustrate a broader image of a species' behaviour. However, even now, the utility of these technologies is expansive, and practitioners can use them to explore species through a new lens.
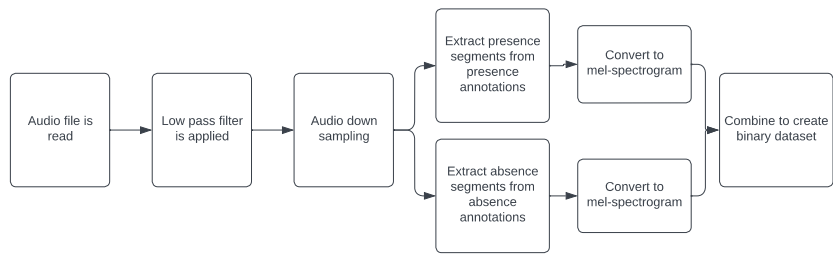
Figure 1. The on-animal acoustic recorder used to record the rutting vocalizations of the reindeer. The orange case is the 3D-printed weatherproof housing containing the Soroka-15E recorder. The four panels represent the (**A**) interior, (**B**) top, (**C**) side, and (**D**) front. The ruler denotes units in centimetres.
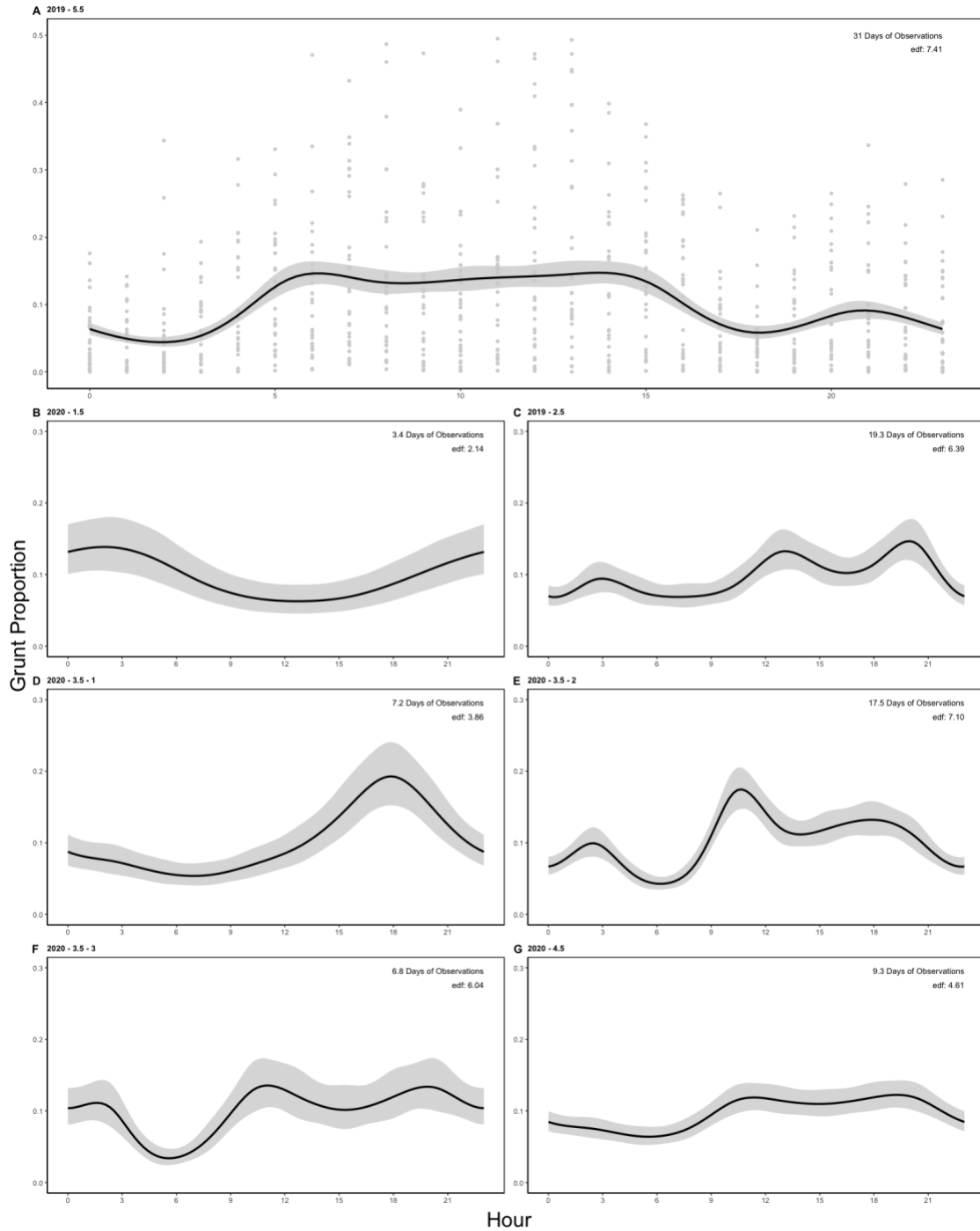
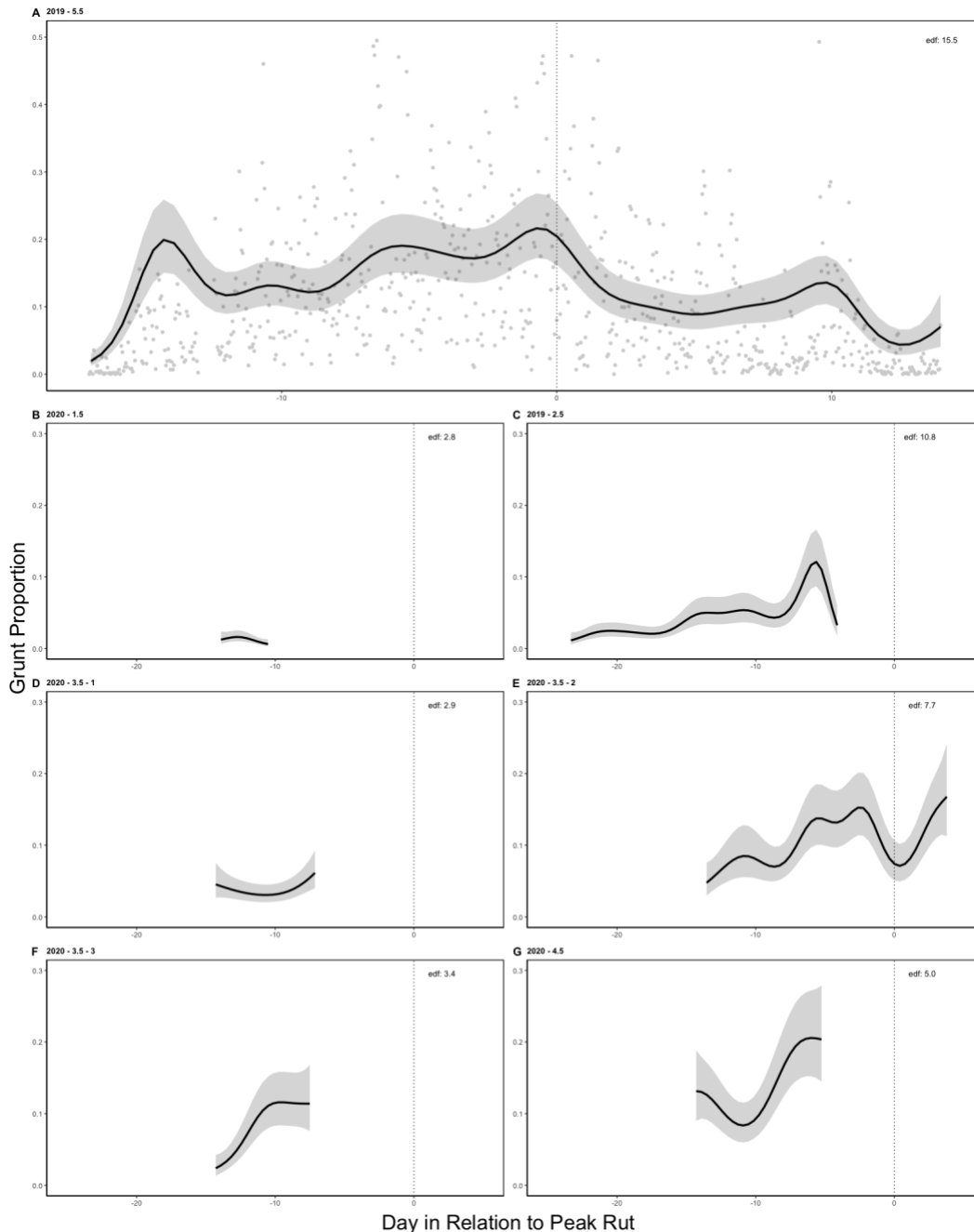**Figure 2.** Various spectrograms of noises captured by the on-animal acoustic recorder. The different spectrograms represent (**A**) a series of reindeer grunts, (**B**) a single reindeer grunt, (**C**) the sound captured by the recorder as it scratched against some branches and (**D**) the sound of the wind as it blew against the recorder. The y-axis represents frequency (Hz), and each spectrogram is exactly four seconds long.
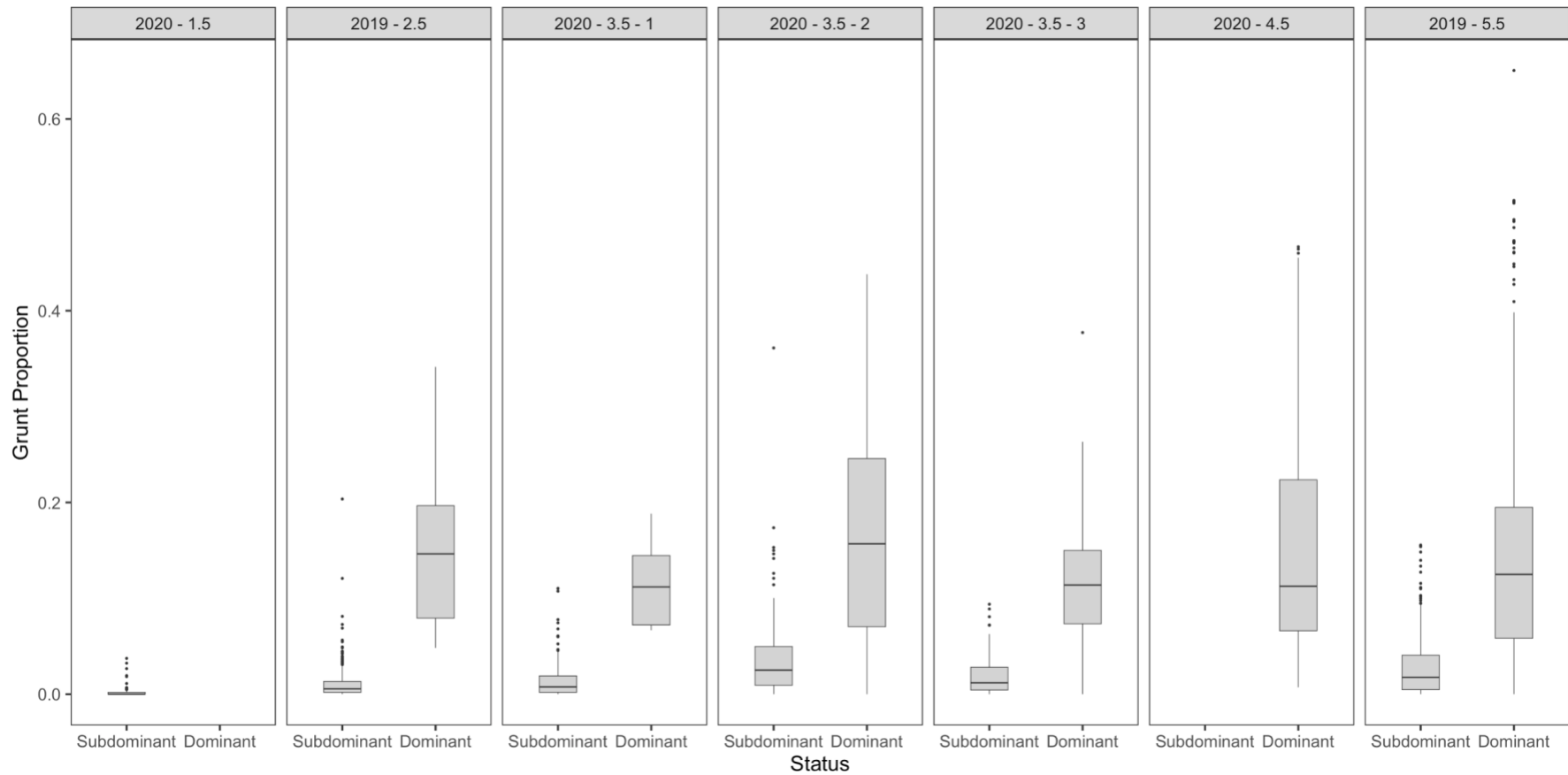
Figure 3. A binary data set was created in the above process. For each file, the data set was made by reading a given audio file and applying the above pre-processing steps. Segment duration and spectrogram parameters are detailed in **Table 2** (Dufourq et al., 2022).
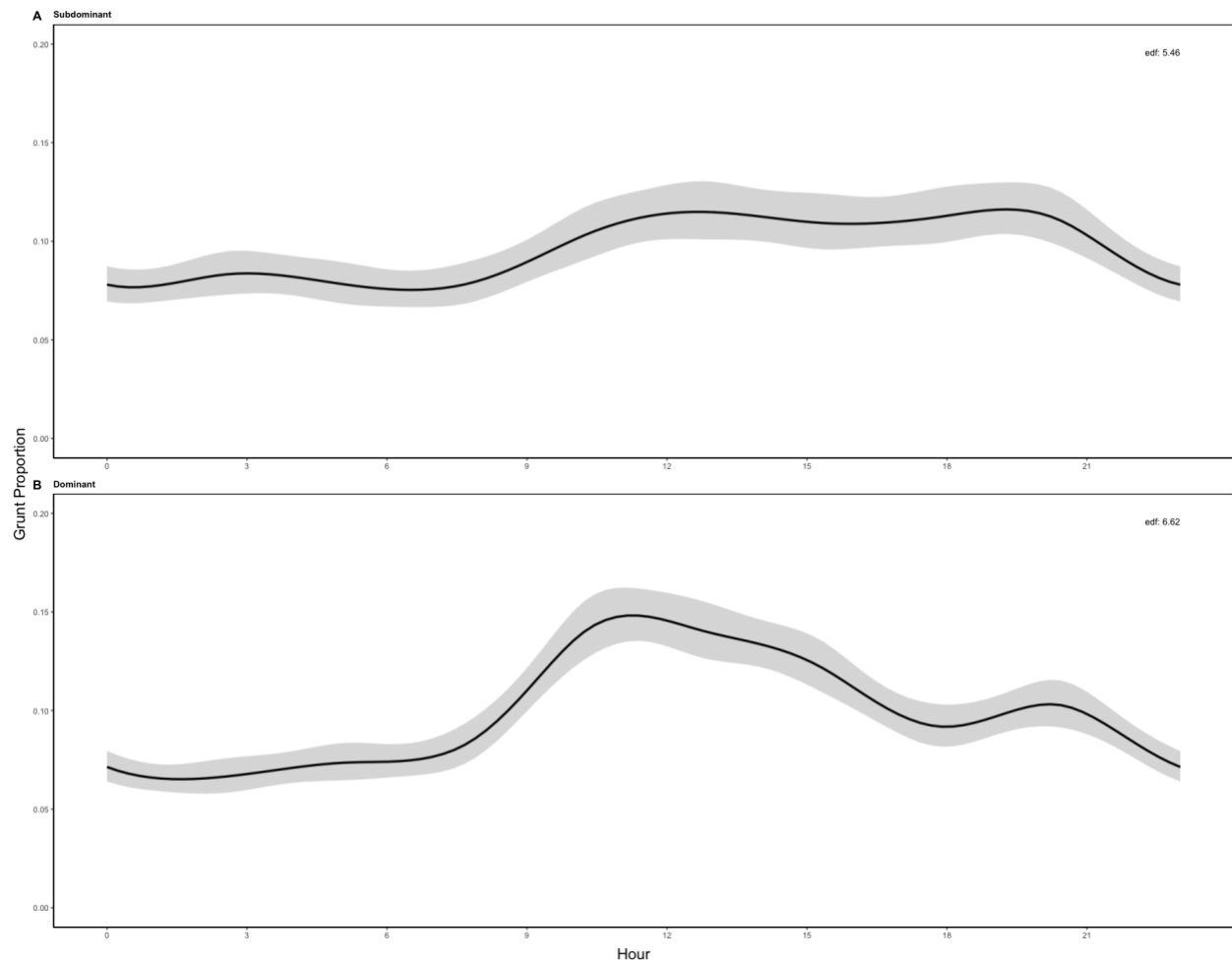
Figure 4 A-G. Estimated smooth curves of the 24-hour rut patterns of seven reindeer according to their acoustic recordings represented as grunting behaviour taken as a proportion of time for each hour. (**A**) represents the estimated smooth curve of a 5.5-year-old and its complete data set to serve as an example for the data gathered. (**B-G**) represent six smooth curves of six individuals ranging from 1.5-years-old to 4.5-years-old. Black lines represent the mean, and shaded areas represent the 95% confidence interval. Models were fit using ten basis functions, and edf represents effective degrees of freedom. Days of observation represent the number of days the smooth curves were estimated. 2019 and 2020 represent the year the data were gathered. 1.5-5.5 represent the age of the reindeer, and 1-3 represent different individuals of the same age class.

Figure 5 A-G. Estimated smooth curves of the daily patterns of seven reindeer according to their acoustic recordings represented as grunting behaviour taken as a proportion of time for each hour. (**A**) represents the estimated smooth curve of a 5.5-year-old and its complete data set to serve as an example for the data gathered (note that the scale for this figure differs from the other six). (**B-G**) represent six smooth curves of six individuals ranging from 1.5-years-old to 4.5-years-old. Black lines represent the mean, shaded areas represent the 95% confidence interval, and the dotted vertical line represents the peak rutting day. Models were fit using 22 basis functions, and edf represents effective degrees of freedom. Days in relation to peak rut refer to the estimated peak rut date of that year's cohort. The peak ruts for the 2019 and 2020 males were estimated to be October 5th and October 6th, respectively. 2019 and 2020 represent the year the data were gathered. 1.5-5.5 represent the age of the reindeer, and 1-3 represent different individuals of the same age class.

43

Figure 6 Comparison of hourly grunt proportions when an individual is in dominant or subdominant positions. 2019 and 2020 represent the year the data were gathered. 1.5-5.5 represent the age of the reindeer, and 1-3 represent different individuals of the same age class.

Figure 7 A & B. Estimated smooth curves of the 24-hour rut patterns of seven reindeer according to their acoustic recordings represented as grunting behaviour taken as a proportion of time for each hour comparing (**A**) subdominant behaviour to (**B)** dominant behaviour. Black lines represent the mean, and shaded areas represent the 95% confidence interval. Models were fit using ten basis functions, and edf represents effective degrees of freedom**.**

Table 1. Individual reindeer sampling details, physical rut characteristics and grunting activity patterns (mean ± SE). The 2019 male pre-rut weights were measured on September 11th, 2019, and their post-rut weights were measured on November 26th, 2019. The 2020 male pre-rut weights were measured on September 21st, 2020, and their post-rut weights were measured on November 9th, 2020. Reproductive success denotes the number of offspring born to a male the following spring.

Individual	Age (Years)	Year sampled	Recording date start	Recording date end	Number of hourly observations (n)	Starting rut weight (kg)	Weight loss at end of rut (kg)	Reproductive success	Percent of each hour spent grunting
1.5	1.5	2020	Sept 22	Sept 25	82	83	-5	NA	0.38 ± 0.47
2.5	2.5	2019	Sept 11	Sept 30	464	96	5	0	1.59 ± 0.07
3.5 - 1	3.5	2020	Sept 21	Sept 28	173	100	1	NA	1.80 ± 0.22
3.5 - 2	3.5	2020	Sept 22	Oct 9	419	135	30	NA	9.89 ± 0.47
3.5 - 3	3.5	2020	Sept 21	Sept 28	164	155	35	NA	7.06 ± 0.53
4.5	4.5	2020	Sept 21	Sept 30	224	150	39	NA	15.01 ± 0.76
5.5	5.5	2019	Sept 18	Oct 18	744	140	28	14	10.64 ± 0.40

Table 2. Pre-processing hyper-parameters used while training and analyzing the rut vocalizations of reindeer from the Kutuharju field research station and the number of testing files used.

Parameters	Units
Low pass filter cut-off (Hz)	1000
Downsampling rate (Hz)	8000
Nyquist rate (Hz)	4000
Segment duration (seconds)	4
Hann window length (samples)	1024
Spectrogram hop size (samples)	256
Number of spectrogram mel frequency bins	128
Spectrogram minimum frequency (Hz)	0
Spectrogram maximum frequency (Hz)	4000
Number of testing files (n)	12
Testing time (minutes)	1638.4

Table 3. The recall rate, precision, accuracy and F1 average values of three convolutional neural networks trained on a series of sound files containing reindeer vocalizations and unintentional sounds.

Model	Audio file subset	Recall rate	Precision	Accuracy	F1
2020 - 10 Epoch Model	Trained (n=8)	91.9	94.6	79.0	93.2
	Untrained (n=4)	81.2	96.7	87.5	88.1
	Total (n=12)	88.3	95.3	84.6	91.5
2020 - 50 Epoch Model	Trained (n=8)	95.5	92.8	82.6	94.1
	Untrained (n=4)	84.2	98.1	89.1	90.4
	Total (n=12)	91.8	94.6	86.9	92.9
Combined - 25 Epoch Model	Total (n=12)	97.8	95.8	93.6	96.7

*The trained subset refers to eight testing files from individuals from the 2020 cohort, while the untrained subset refers to four testing files from the 2019 cohort (from two individuals, those audio files were not included in the training audio file batch). The total subset refers to the average values from a combination of the trained and untrained subsets (in the case of the combined model, the total subset represents the average of 12 testing files that came from individuals from both the 2019 and 2020 sampling periods, both of which were used to train this model). The 2020 - 10 Epoch Model and 2020 - 50 Epoch Model were trained on 37 2.27-hour audio files from five individuals from the 2020 sample period. These two models were trained on 28000 presence and absence segments. The combined - 25 Epoch Model was trained on 55, 2.27-hour audio files from seven individuals from the 2019 and 2020 sample periods. This model was trained on 10778 presence segments and 11546 absence segments. Testing files are audio samples that were not used during the training process.

* Recall rate indicates how well a model detects sounds of interest (the proportions of true positives to actual positives). Precision indicates the detector's reliability (what proportion of positive identifications were correct). Accuracy measures the proportion of classifications a model correctly classifies. Finally, F1 score measures the harmonic precision and accuracy mean. Refer to **equations A1-4** for a formula of each metric.

Table 4. The average rest per day (mean ± SE), the average number of breaks per day (mean ± SE), and the average time spent at rest per break (mean ± SE) across our seven individuals. Three days per individual were analyzed via audio playback across the range of the days we had recorded for each male reindeer.

| Individual	Average rest per day (hh:mm)	Number of rests per day (n)	Average time spent at rest per break (mm:ss)
1.5	3:28 ± 00:44	6.0 ± 0.9	34:44 ± 06:23
2.5	1:50 ± 00:52	11.3 ± 0.2	09:27 ± 02:36
3.5 - 1	4:39 ± 01:02	14.0 ± 0.7	19:57 ± 02:57
3.5 - 2	6:32 ± 00:49	16.7 ± 0.4	23:33 ± 02:51
3.5 - 3	4:00 ± 00:17	9.7 ± 0.9	25:45 ± 02:21
4.5	5:01 ± 00:44	15.7 ± 2.0	19:14 ± 01:09
5.5	5:26 ± 00:40	14.0 ± 0.9	23:18 ± 03:27

Appendix

Table A1. The following Python packages were used to train the convolutional neural networks and classify the rutting audio of the male reindeer. The following files were run on Python 3.10.7 on a 2021 Apple Macbook Pro with an Apple M1 pro processor with 16 GB of unified LPDDR5 RAM running MacOS Ventura 13.1.

Python Package	Version
Jupyter	1.0.0
Keras	2.10.0
Librosa	0.9.2
Matplotlib	3.6.0
Numpy	1.23.3
Pandas	1.5.0
Pickle	0.7.5
Scikit-learn	1.1.2
Scipy	1.9.1
Soundfile	0.10.3.post1
Tensorflow 2	2.7.0
Yattag	1.14.0
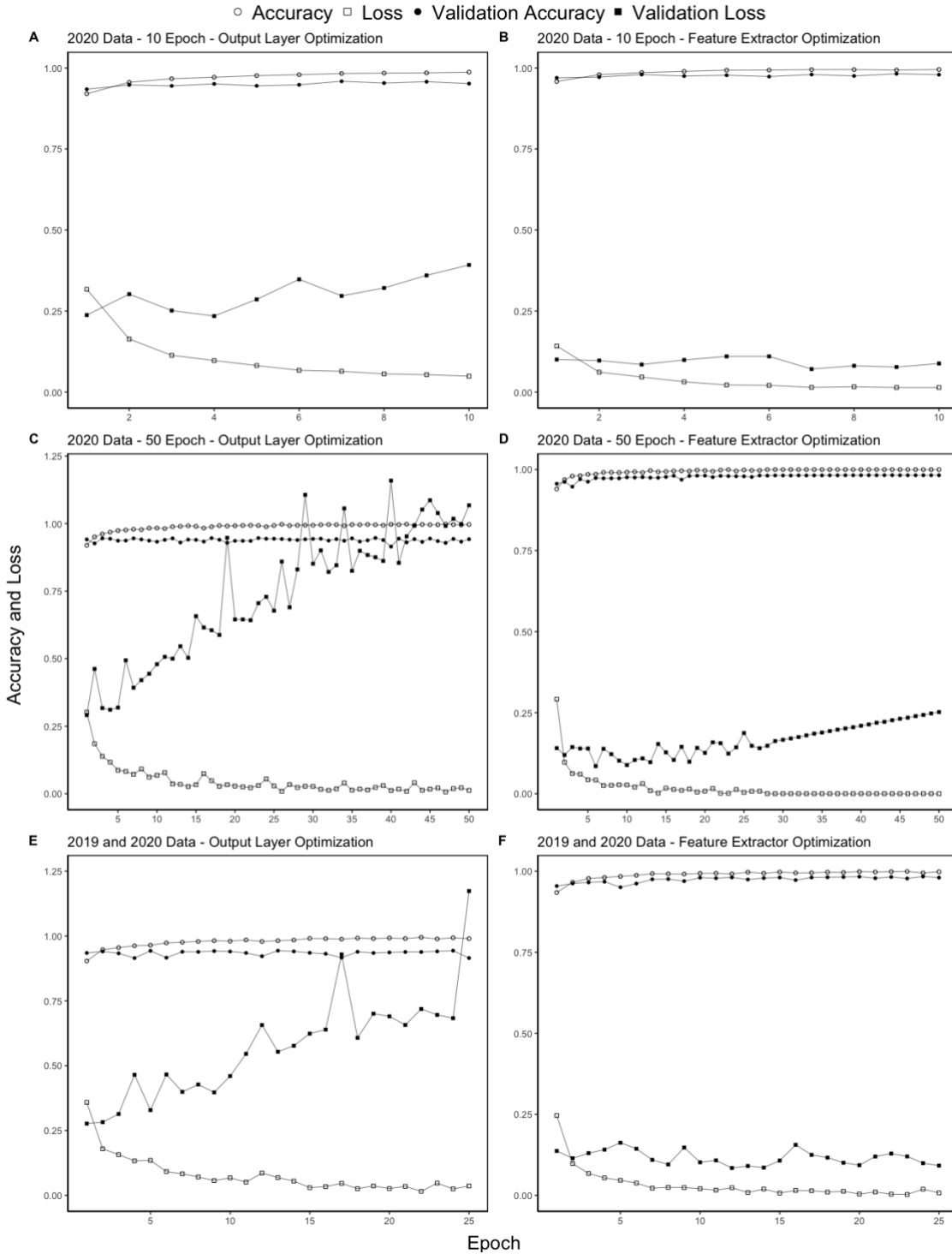
Figure A1 A-F. Figures demonstrating the three networks' accuracy, loss, validation, and validation loss over numerous epochs during the training process. (**A, C & E**) relate to the networks during the output-layer optimization, and (**B, D & F**) relate to the networks during the feature extractor optimization. (**A & B**) relate to the network trained over ten epochs on the 2020 data, (**C & D**) relate to the network trained over 50 epochs on the 2020 data, and (**E & F**) relate to the network trained over 25 epochs on both the 2019 and 2020 data.

Equations A1-4. The following four equations were used to measure our three networks' performance based on their classifications. True positives (TP) occur when a grunt presence is correctly classified, false negatives (FN) occur when an absence of a grunt is classified in a segment that contains a grunt, and false positives (FP) occur when a grunt is classified in a segment when there is no grunt (Do Nascimento et al., 2021; Mesaros et al., 2016).

$$(A.1) \quad Recall\ rate\ or\ sensivity = \frac{TP}{TP + FN} \quad (A.2) \quad Precision = \frac{TP}{TP + FP}$$

$$(A.3) \quad Accuracy = \frac{TP}{TP + FP + FN} \qquad (A.4) \quad F_1 score = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, 265–283. https://doi.org/10.48550/arXiv.1605.08695

Albert, A., & Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, *71*(1), 1–10. JSTOR. https://doi.org/10.2307/2336390

Andersson, M. (1994). *Sexual selection*. Princeton University Press.

Arak, A. (1983). Sexual selection by male–male competition in natterjack toad choruses. *Nature*, *306*(5940), 261–262. https://doi.org/10.1038/306261a0

Baker, R. R., & Parker, G. A. (1979). The evolution of bird coloration. *Philosophical Transactions of the Royal Society B Biological Sciences*, *287*(1018), 63–130. JSTOR. https://doi.org/10.1098/rstb.1979.0053

Bermant, P. C., Bronstein, M. M., Wood, R. J., Gero, S., & Gruber, D. F. (2019). Deep machine learning techniques for the detection and classification of sperm whale bioacoustics. *Scientific Reports*, *9*(1), 12588. https://doi.org/10.1038/s41598-019-48909-4

Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., Deppe, J. L., Krakauer, A. H., Clark, C., Cortopassi, K. A., Hanser, S. F., McCowan, B., Ali, A. M., & Kirschel, A. N. G. (2011). Acoustic monitoring in terrestrial environments using microphone

arrays: Applications, technological considerations and prospectus. *Journal of Applied Ecology*, *48*(3), 758–767. https://doi.org/10.1111/j.1365-2664.2011.01993.x

Boake, C. R. B., & Capranica, R. R. (1982). Aggressive signal in "courtship" chirps of a gregarious cricket. *Science*, *218*(4572), 580–582. https://doi.org/10.1126/science.218.4572.580

Bobek, B., Perzanowski, K., & Weiner, J. (1990). Energy expenditure for reproduction in male red deer. *Journal of Mammalogy*, *71*(2), 230–232. https://doi.org/10.2307/1382171

Body, G., Weladji, R. B., Holand, Ø., & Nieminen, M. (2014). Highly competitive reindeer males control female behavior during the rut. *PLOS ONE*, *9*(4), e95618. https://doi.org/10.1371/journal.pone.0095618

Bowyer, R. T., McCullough, D. R., Rachlow, J. L., Ciuti, S., & Whiting, J. C. (2020). Evolution of ungulate mating systems: Integrating social and environmental factors. *Ecology and Evolution*, *10*(11), 5160–5178. https://doi.org/10.1002/ece3.6246

Bradbury, J. W., & Vehrencamp, S. L. (2011). *Principles of animal communication* (2nd ed.). Sinauer Associates.

Bravo Sanchez, F. J., Hossain, M. R., English, N. B., & Moore, S. T. (2021). Bioacoustic classification of avian calls from raw sound waveforms with an open-source deep learning architecture. *Scientific Reports*, *11*(1), 15733. https://doi.org/10.1038/s41598-021-95076-6

Browning, E., Gibb, R., Glover-Kapfer, P., & Jones, K. E. (2017). Passive acoustic monitoring in ecology and conservation. *WWF Conversation Technology Series*, *1*(2), 1–75. https://doi.org/10.25607/OBP-876

Cannam, C., Landone, C., & Sandler, M. (2010). Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. *ACM Multimedia 2010 International Conference*, 1467–1468. https://doi.org/10.1145/1873951.1874248

Casoli, M., Johnson, M., McHugh, K. A., Wells, R. S., & Tyack, P. L. (2022). Parameterizing animal sounds and motion with animal-attached tags to study acoustic communication. *Behavioral Ecology and Sociobiology*, *76*, 59. https://doi.org/10.1007/s00265-022-03154-0

Cassini, M. H. (2020). Sexual size dimorphism and sexual selection in primates. *Mammal Review*, *50*(3), 231–239. https://doi.org/10.1111/mam.12191

Charlton, B. D., & Reby, D. (2011). Context-related acoustic variation in male fallow deer (*Dama dama*) groans. *PLOS ONE*, *6*(6), e21066. https://doi.org/10.1371/journal.pone.0021066

Charlton, B. D., Reby, D., & McComb, K. (2007). Female red deer prefer the roars of larger males. *Biology Letters*, *3*(4), 382–385. https://doi.org/10.1098/rsbl.2007.0244

Cooper, I. A., Brown, J. M., & Getty, T. (2016). A role for ecology in the evolution of colour variation and sexual dimorphism in Hawaiian damselflies. *Journal of Evolutionary Biology*, *29*(2), 418–427. https://doi.org/10.1111/jeb.12796

Cvikel, N., Levin, E., Hurme, E., Borissov, I., Boonman, A., Amichai, E., & Yovel, Y. (2015). On-board recordings reveal no jamming avoidance in wild bats. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1798), 20142274. https://doi.org/10.1098/rspb.2014.2274

Darras, K., Batary, P., Furnas, B., Grass, I., Mulyani, Y., & Tscharntke, T. (2019). Autonomous sound recording outperforms human observation for sampling birds: A systematic map and user guide. *Ecological Applications*, *29*(6), e01954. https://doi.org/10.1002/eap.01954

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. John Murray.

Darwin, C. (1871). *The descent of man and selection in relation to sex*. John Murray.

Do Nascimento, L. A., Pérez-Granados, C., & Beard, K. H. (2021). Passive acoustic monitoring and automatic detection of diel patterns and acoustic structure of howler monkey roars. *Diversity*, *13*(11). https://doi.org/10.3390/d13110566

Dufourq, E., Batist, C., Foquet, R., & Durbach, I. (2022). Passive acoustic monitoring of animal populations with transfer learning. *Ecological Informatics*, *70*, 101688. https://doi.org/10.1016/j.ecoinf.2022.101688

Dufourq, E., Durbach, I., Hansford, J. P., Hoepfner, A., Ma, H., Bryant, J. V., Stender, C. S., Li, W., Liu, Z., Chen, Q., Zhou, Z., & Turvey, S. T. (2021). Automated detection of Hainan gibbon calls for passive acoustic monitoring. *Remote Sensing in Ecology and Conservation*, *7*(3), 475–487. https://doi.org/10.1002/rse2.201

Enari, H., Enari, H., Maruyama, T., Okuda, K., & Okuda, K. (2019). An evaluation of the efficiency of passive acoustic monitoring in detecting deer and primates in comparison with camera traps. *Ecological Indicators*, *98*, 753–762. https://doi.org/10.1016/j.ecolind.2018.11.062

Enari, H., Enari, H., Okuda, K., Yoshita, M., Kuno, T., & Okuda, K. (2017). Feasibility assessment of active and passive acoustic monitoring of sika deer populations. *Ecological Indicators*, *79*, 155–162. https://doi.org/10.1016/j.ecolind.2017.04.004

Espmark, Y. O. (1964). Rutting behaviour in reindeer (*Rangifer tarandus L.*). *Animal Behaviour*, *12*(1), 159–163. https://doi.org/10.1016/0003-3472(64)90117-4

Espmark, Y. O. (1971). Individual recognition by voice in reindeer mother-young relationship. Field observations and playback experiments. *Behaviour*, *40*(3–4), 295–301. https://doi.org/10.1163/156853971X00438

Fairbrass, A. J., Rennert, P., Williams, C., Titheridge, H., & Jones, K. E. (2017). Biases of acoustic indices measuring biodiversity in urban areas. *Ecological Indicators*, *83*, 169–177. https://doi.org/10.1016/j.ecolind.2017.07.064

Farina, A., Pieretti, N., & Piccioli, L. (2011). The soundscape methodology for long-term bird monitoring: A Mediterranean Europe case-study. *Ecological Informatics*, *6*(6), 354–363. https://doi.org/10.1016/j.ecoinf.2011.07.004

Feighny, J. A., Williamson, K. E., & Clarke, J. A. (2006). North American elk bugle vocalizations: Male and female bugle call structure and context. *Journal of Mammalogy*, *87*(6), 1072–1077. https://doi.org/10.1644/06-MAMM-A-079R2.1

Frey, R., Gebler, A., Fritsch, G., Nygrén, K., & Weissengruber, G. E. (2007). Nordic rattle: The hoarse vocalization and the inflatable laryngeal air sac of reindeer (*Rangifer tarandus*). *Journal of Anatomy*, *210*(2), 131–159. https://doi.org/10.1111/j.1469-7580.2006.00684.x

Gammon, D. E., Baker, M. C., & Tipton, J. R. (2005). Cultural divergence within novel song in the black-capped chickadee (*Poecile atricapillus*). *The Auk*, *122*(3), 853–871. https://doi.org/10.1093/auk/122.3.853

Garcia, M., Charlton, B. D., Wyman, M. T., Fitch, W. T., & Reby, D. (2013). Do red deer stags (*Cervus elaphus*) use roar fundamental frequency (F0) to assess rivals? *PLoS ONE*, *8*(12), e83946. https://doi.org/10.1371/journal.pone.0083946

Gottlander, K. (1987). Variation in the song rate of the male pied flycatcher *Ficedula hypoleuca*: Causes and consequences. *Animal Behaviour*, *35*(4), 1037–1043. https://doi.org/10.1016/S0003-3472(87)80160-4

Habib, L., Bayne, E. M., & Boutin, S. (2007). Chronic industrial noise affects pairing success and age structure of ovenbirds *Seiurus aurocapilla*. *Journal of Applied Ecology*, *44*(1), 176–184. https://doi.org/10.1111/j.1365-2664.2006.01234.x

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Holt, M. M., Hanson, M. B., Emmons, C. K., Baird, R. W., Hogan, J., Foster, J., Giles, D., & Balcomb, K. C. (2011). Investigating acoustics, behavior and vessel noise exposure in endangered killer whales (*Orcinus orca*) using digital acoustic recording tags. *The Journal of the Acoustical Society of America*, *129*(4), 2606–2606. https://doi.org/10.1121/1.3588644

Ilany, A., Barocas, A., Kam, M., Ilany, T., & Geffen, E. (2013). The energy cost of singing in wild rock hyrax males: Evidence for an index signal. *Special Issue: Behavioural Plasticity and Evolution*, *85*(5), 995–1001. https://doi.org/10.1016/j.anbehav.2013.02.023

Insley, S. J., Robson, B. W., Yack, T., Ream, R. R., & Burgess, W. C. (2008). Acoustic determination of activity and flipper stroke rate in foraging northern fur seal females. *Endangered Species Research*, *4*(1), 147–155. https://doi.org/10.3354/esr00050

Johnson, J., & Bayne, E. (2022). Bioacoustically derived migration arrival times in boreal birds: Implications for assessing habitat quality. *Avian Conservation and Ecology*, *17*(2), 13. https://doi.org/10.5751/ACE-02224-170213

Johnson, M. P., & Tyack, P. L. (2003). A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE Journal of Oceanic Engineering*, *28*(1), 3–12. https://doi.org/10.1109/JOE.2002.808212

Kingma, D. P., & Ba, J. (2014). *Adam: A Method for stochastic optimization*. http://arxiv.org/abs/1412.6980

Knight, E. C., Hannah, K. C., Foley, G. J., Scott, C. D., Brigham, R. M., & Bayne, E. (2017). Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. *Avian Conservation and Ecology*, *12*(2), 14. https://doi.org/10.5751/ACE-01114-120214

Kojola, I. (1991). Influence of age on the reproductive effort of male reindeer. *Journal of Mammalogy*, *72*(1), 208–210. https://doi.org/10.2307/1382001

Krebs, J., Ashcroft, R., & Webber, M. (1978). Song repertoires and territory defence in the great tit. *Nature, 271*(5645), 539–542. https://doi.org/10.1038/271539a0

Laiolo, P. (2010). The emerging significance of bioacoustics in animal species conservation. *Conservation Planning within Emerging Global Climate and Economic Realities*, *143*(7), 1635–1645. https://doi.org/10.1016/j.biocon.2010.03.025

Laiolo, P., & Tella, J. L. (2007). Erosion of animal cultures in fragmented landscapes. *Frontiers in Ecology and the Environment*, *5*(2), 68–72. https://doi.org/10.1890/1540-9295(2007)5[68:EOACIF]2.0.CO;2

Landau, H. L. (1967). Sampling, data transmission, and the Nyquist rate. *Proceedings of the IEEE*, *55*(10), 1701–1706. https://doi.org/10.1109/PROC.1967.5962

Leader-Williams, N. (1980). Population dynamics and mortality of reindeer introduced into South

    Georgia. *The Journal of Wildlife Management*, *44*(3), 640–657. JSTOR.

    https://doi.org/10.2307/3808011

Lent, P. C. (1975). A review of acoustic communication in *Rangifer tarandus*. In Luick, J.R, Lent,

    P.C, Klein, D.R, & White, R.G (Eds.), *Proceedings of the First International Reindeer and*

    Caribou Symposium (pp. 398–408).

L'Italien, L., Weladji, R. B., Holand, Ø., Røed, K. H., Nieminen, M., & Côté, S. D. (2012). Mating

    group size and stability in reindeer *Rangifer tarandus*: The effects of male characteristics, sex

    ratio and male age structure. *Ethology*, *118*(8), 783–792. https://doi.org/10.1111/j.1439-

    0310.2012.02073.x

Liu, N., Jiang, Z., Zhang, L., Zhong, Z., Ping, X., Xu, H., & Li, C. (2016). Bioacoustic cues and

    their relations to dominance rank in Père David's deer stags. *Animal Production Science*, *56*(6),

    971–977. https://doi.org/10.1071/AN15048

Lostanlen, V., Salamon, J., Farnsworth, A., Kelling, S., & Bello, J. P. (2019). Robust sound event

    detection in bioacoustic sensor networks. *PLOS ONE*, *14*(10), e0214168.

    https://doi.org/10.1371/journal.pone.0214168

Lynch, E., Angeloni, L., Fristrup, K., Joyce, D., & Wittemyer, G. (2013). The use of on-animal

    acoustical recording devices for studying animal behavior. *Ecology and Evolution*, *3*(7), 2030–

    2037. https://doi.org/10.1002/ece3.608

McComb, K., Moss, C., Durant, S. M., Baker, L., & Sayialel, S. (2001). Matriarchs as repositories

    of social knowledge in African elephants. *Science*, *292*(5516), 491–494.

    https://doi.org/10.1126/science.1057895

McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O. (2020). *Librosa: Audio and music signal analysis in python* (0.8.1). https://doi.org/10.5281/zenodo.4792298

Mcloughlin, M. P., Stewart, R., & McElligott, A. G. (2019). Automated bioacoustics: Methods in ecology and conservation and their potential for animal welfare monitoring. *Journal of The Royal Society Interface*, *16*(155), 20190225. https://doi.org/10.1098/rsif.2019.0225

Mellado, M., Cárdenas, C., & Ruíz, F. (2000). Mating behavior of bucks and does in goat operations under range conditions. *Applied Animal Behaviour Science*, *67*, 89–96. https://doi.org/10.1016/S0168-1591(99)00109-4

Mesaros, A., Heittola, T., & Virtanen, T. (2016). Metrics for polyphonic sound event detection. *Applied Sciences*, *6*(6). https://doi.org/10.3390/app6060162

Mitani, J. C. (1985). Sexual selection and adult male orangutan long calls. *Animal Behaviour*, *33*(1), 272–283. https://doi.org/10.1016/S0003-3472(85)80141-X

Mossing, T., & Damber, J-E. (1981). Rutting behavior and androgen variation in reindeer (*Rangifer tarandus L.*). *Journal of Chemical Ecology*, *7*, 377–389. https://doi.org/10.1007/BF00995760

Mukharji, P. B. (2021). Darwin's bulbuls: South Asian cultures of bird fighting and Darwin's theory of sexual selection. *BJHS Themes*, *6*, 1–17. https://doi.org/10.1017/bjt.2021.3

Mysterud, A., Bonenfant, C., Loe, L. E., Langvatn, R., Yoccoz, N., & Stenseth, N. C. (2008). The timing of male reproductive effort relative to female ovulation in a capital breeder. *The Journal of Animal Ecology*, *77*(3), 469–477. https://doi.org/10.1111/j.1365-2656.2008.01365.x

Mysterud, A., Langvatn, R., & Stenseth, N. C. (2004). Patterns of reproductive effort in male ungulates. *J Zool*, *264*(2), 209–215. https://doi.org/10.1017/S0952836904005618

Mysterud, A., Røed, K. H., Holand, Ø., Yoccoz, N. G., & Nieminen, M. (2009). Age-related gestation length adjustment in a large iteroparous mammal at northern latitude. *Journal of Animal Ecology*, *78*(5), 1002–1006. https://doi.org/10.1111/j.1365-2656.2009.01553.x

Navarro, P. J., Fernández, C., Borraz, R., & Alonso, D. (2017). A machine learning approach to pedestrian detection for autonomous vehicles using high-definition 3D range data. *Sensors*, *17*(1). https://doi.org/10.3390/s17010018

O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *ArXiv e-prints,* 1511.08458, https://doi.org/10.48550/arXiv.1511.08458

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ*, *7*, e6876. https://doi.org/10.7717/peerj.6876

R Core Team. (2018). *R: A language and environment for statistical computing (4.1.3)*. R Foundation for Statistical Computing. https://www.r-project.org/

Reby, D., & McComb, K. (2003a). Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour*, *65*(3), 519–530. https://doi.org/10.1006/anbe.2003.2078

Reby, D., & McComb, K. (2003b). Vocal communication and reproduction in deer. *Advances in The Study of Behavior*, *33*, 231–264. https://doi.org/10.1016/S0065-3454(03)33005-0

Robertson, J. G. M. (1986). Female choice, male strategies and the role of vocalizations in the Australian frog *Uperoleia rugosa*. *Animal Behaviour*, *34*(3), 773–784. https://doi.org/10.1016/S0003-3472(86)80061-6

Rognan, C., Szewczak, J., & Morrison, M. (2012). Autonomous recording of great gray owls in the Sierra Nevada. *Northwestern Naturalist*, *93*, 138–144. https://doi.org/10.2307/23259690

Shonfield, J., & Bayne, E. M. (2017). Autonomous recording units in avian ecological research: Current use and future applications. *Avian Conservation and Ecology*, *12*(1), 14. https://doi.org/10.5751/ACE-00974-120114

Silva, T. L., Mooney, T. A., Sayigh, L. S., Tyack, P. L., Baird, R. W., & Oswald, J. N. (2016). Whistle characteristics and daytime dive behavior in pantropical spotted dolphins (*Stenella attenuata*) in Hawai'i measured using digital acoustic recording tags (DTAGs). *The Journal of the Acoustical Society of America*, *140*(1), 421–429. https://doi.org/10.1121/1.4955081

Simpson, G. L. (2021). Gratia: Graceful 'ggplot'-based graphics and other functions for GAMs fitted using 'mgcv.' *R Package Version 0.3. 0*.

Singer, F. J., & Zeigenfuss, L. C. (2002). Influence of trophy hunting and horn size on mating behavior and survivorship of mountain sheep. *Journal of Mammalogy*, *83*(3), 682–698. https://doi.org/10.1644/1545-1542(2002)083<0682:IOTHAH>2.0.CO;2

Skogland, T. (1989). *Comparative social organization of wild reindeer in relation to food, mates and predator avoidance*. Parey. http://lib.ugent.be/catalog/rug01:000184320

Sousa-Lima, R. S., Norris, T. F., Oswald, J. N., & Fernandes, D. P. (2013). A review and inventory of fixed autonomous recorders for passive acoustic monitoring of marine mammals. *Aquatic Mammals*, *39*(1), 23–53. https://doi.org/10.1109/RIOAcoustics.2013.6683984

Stein, P. J. (2011). Active acoustic monitoring systems for detecting, localizing, tracking, and classifying marine mammals and fish. *The Journal of the Acoustical Society of America*, *129*(4), 2369. https://doi.org/10.1121/1.3587669

Stevens, S. S., Volkmann, J., & Newman, E. B. (2005). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, *8*(3), 185–190. https://doi.org/10.1121/1.1915893

Stimpert, A. K., Au, W. W. L., Parks, S. E., Hurst, T., & Wiley, D. N. (2011). Common humpback whale (*Megaptera novaeangliae*) sound types for passive acoustic monitoring. *The Journal of the Acoustical Society of America*, *129*(1), 476–482. https://doi.org/10.1121/1.3504708

Stowell, D., Benetos, E., & Gill, L. F. (2017). On-bird sound recordings: Automatic acoustic recognition of activities and contexts. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *25*(6), 1193–1206. https://doi.org/10.1109/TASLP.2017.2690565

Studd, E. K., Boudreau, M. R., Majchrzak, Y. N., Menzies, A. K., Peers, M. J. L., Seguin, J. L., Lavergne, S. G., Boonstra, R., Murray, D. L., Boutin, S., & Humphries, M. M. (2019). Use of acceleration and acoustics to classify behavior, generate time budgets, and evaluate responses to moonlight in free-ranging snowshoe hares. *Frontiers in Ecology and Evolution*, *7*, 154. https://doi.org/10.3389/fevo.2019.00154

Studd, E. K., Derbyshire, R. E., Menzies, A. K., Simms, J. F., Humphries, M. M., Murray, D. L., & Boutin, S. (2021). The purr-fect catch: Using accelerometers and audio recorders to document kill rates and hunting behaviour of a small prey specialist. *Methods in Ecology and Evolution*, *12*(7), 1277–1287. https://doi.org/10.1111/2041-210X.13605

Tennenhouse, E. M., Weladji, R. B., Holand, Ø., & Nieminen, M. (2012). Timing of reproductive effort differs between young and old dominant male reindeer. *Annales Zoologici Fennici*, *49*(3), 152–160. JSTOR. https://doi.org/10.5735/086.049.0303

Thiebault, A., Huetz, C., Pistorius, P., Aubin, T., & Charrier, I. (2021). Animal-borne acoustic data alone can provide high accuracy classification of activity budgets. *Animal Biotelemetry*, *9*(1), 28. https://doi.org/10.1186/s40317-021-00251-1

Vannoni, E., & McElligott, A. G. (2008). Low frequency groans indicate larger and more dominant fallow deer (*Dama dama*) males. *PLOS ONE*, *3*(9), e3113. https://doi.org/10.1371/journal.pone.0003113

Vannoni, E., Torriani, M. V. G., & McElliogott, A. G. (2005). Acoustic signaling in cervids: A methodological approach for measuring vocal communication in fallow deer. *Cognition, Brain, Behavior*, *9*(3), 551–566.

Volodin, I. A., Matrosova, V. A., Volodina, E. V., Garcia, A. J., Gallego, L., Márquez, R., Llusia, D., Beltrán, J. F., & Landete-Castillejos, T. (2015). Sex and age-class differences in calls of Iberian red deer during the rut: Reversed sex dimorphism of pitch and contrasting roars from farmed and wild stags. *Acta Ethologica*, *18*(1), 19–29. https://doi.org/10.1007/s10211-013-0179-8

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*(1), 1–40. https://doi.org/10.1186/s40537-016-0043-6

Wijers, M., Trethowan, P., Markham, A., du Preez, B., Chamaillé-Jammes, S., Loveridge, A., & Macdonald, D. (2018). Listening to lions: Animal-borne acoustic sensors improve bio-logger

calibration and behaviour classification performance. *Frontiers in Ecology and Evolution*, *6*, 171. https://doi.org/10.3389/fevo.2018.00171

Wimmer, J., Towsey, M., Roe, P., & Williamson, I. (2013). Sampling environmental acoustic recordings to determine bird species richness. *Ecological Applications*, *23*(6), 1419–1428. https://doi.org/10.1890/12-2088.1

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Chapman & Hall/CRC Press.

Wyman, M. T., Mooring, M. S., McCowan, B., Penedo, M. C. T., Reby, D., & Hart, L. A. (2012). Acoustic cues to size and quality in the vocalizations of male North American bison, *Bison bison*. *Animal Behaviour*, *84*(6), 1381–1391. https://doi.org/10.1016/j.anbehav.2012.08.037

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights into Imaging*, *9*(4), 611–629. https://doi.org/10.1007/s13244-018-0639-9

Yen, S. C., Shieh, B. S., Wang, Y. T., & Wang. Y. (2013). Rutting vocalizations of Formosan sika deer *Cervus nippon taiouanus*—Acoustic structure, seasonal and diurnal variations, and individuality. *Zoological Science*, *30*(12), 1025–1031. https://doi.org/10.2108/zsj.30.1025

Zahavi, A. (1975). Mate selection—A selection for a handicap. *Journal of Theoretical Biology*, *53*(1), 205–214. https://doi.org/10.1016/0022-5193(75)90111-3

Zeng, Z., Song, Y.L., & Zhang, Q. (2011). Copulatory pattern and behavior in a semi-captive population of Eld's deer. *Current Zoology*, *57*(3), 284–292. https://doi.org/10.1093/czoolo/57.3.284

Zhong, M., LeBien, J., Campos-Cerqueira, M., Dodhia, R., Lavista Ferres, J., Velev, J. P., & Aide, T. M. (2020). Multispecies bioacoustic classification using transfer learning of deep convolutional neural networks with pseudo-labeling. *Applied Acoustics*, *166*, 107375. https://doi.org/10.1016/j.apacoust.2020.107375

Zuur, A. F., & Ieno, E. N. (2016). A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution*, *7*(6), 636–645. https://doi.org/10.1111/2041-210X.12577

Zwart, M. C., Baker, A., McGowan, P. J. K., & Whittingham, M. J. (2014). The use of automated bioacoustic recorders to replace human wildlife surveys: An example using nightjars. *PLOS ONE*, *9*(7), e102770. https://doi.org/10.1371/journal.pone.0102770