NTNU

Department of Computer Science
Department of Physics

Specialization Project

Statistical Analysis Of The Movement Pattern Of Sheep And The Occurrence Of Predators

Author:
Nina Salvesen

Supervisors:
Svein-Olaf Hvasshovd
Jon Andreas Støvneng

Norwegian University of Science and Technology
December 2021

Abstract

This project investigated the movement pattern of sheep in relation to predators on outfield pastures in Norway, by researching relevant literature and statistics on the field. It is a pre-study for a master thesis that will employ machine learning to test if it is possible to classify patterns of predatory behavior in the herd, by usage of data from electronic transmitters worn by the sheep relaying their position in terms of geographical coordinates. The computational analysis will be done with unsupervised and semi-supervised machine learning models to categorize the movement. Important parameters that will be included in the analysis are altitude, activity levels, weather and temperature. The project found that breed of sheep, especially in terms of older versus newer breeds, will be highly important to their reaction to predators. Sheep are habitual animals, and both in terms of hierarchy within a herd, and seasonal and daily routines, there will be fixed patterns for each individual. An hypothesis is therefore that deviance from this will be because of proximity of predators. For further research in the future it is recommended that electronic transmitters are put on the same sheep in a herd for several years, to be able to compare their habitual patterns to find anomalies. It will also be beneficial for the data if herds located in more predator-prone areas are utilized.

Sammendrag

Dette prosjektet studerte bevegelsesmønsteret til sau i sammenheng med rovdyr på utmarksbeiter i Norge, ved å undersøke relevant litteratur og statistikk på emnet. Det er et forstudium for en masteroppgave som skal ta i bruk maskinlæring for å teste om det er mulig å detektere oppførselsesmønster i en flokk på grunn av rovdyr. Dette vil gjøres med bruk av data fra elektroniske e-bjeller sauen har gått med, som logger deres geografiske posisjon i form av koordinater. Beregningene for å kategorisere sauens bevegelser vil bli gjort med uovervåket og semi-uovervåket maskinlæringsmodeller. Parametere som vil være viktige for analysen er høydemeter, aktivitetsnivåer, vær og temperatur. Prosjektet fant at rasen til sauen, spesielt gjeldende for eldre kontra nyere raser, vil ha mye å si for deres reaksjoner i møte med rovdyr. Sau er vanedyr, og hvert individ vil ha faste mønster både med tanke på hierarkiske ordninger i flokken, og deres daglige og sesongbaserte rutiner. En hypoteste er derfor at avvik fra dette vil være grunnet nærhet av rovdyr. For fremtidig forskning er det anbefalt at samme sau i en flokk har på e-bjelle over flere år så det bli enklere å sammenlikne deres bevegelsesmønster og dermed finne uregelmessigheter. Det vil videre være en fordel for datagrunnlaget om flokker i mer rovdyrutsatte områder blir brukt.

Preface

This report is the specialization project of Nina Salvesen for the study program Master of Science in Applied Physics and Mathematics, from the Norwegian University of Science and Technology. The master's thesis is written for the Department of Computer Science under the supervision of Svein-Olaf Hvasshovd, and will study the movement pattern of sheep on outfield pastures in Norway in relation to normal behavior and in close proximity to predators. The analysis will be completed by the combined use of machine learning, ethological theory and statistical analysis. The findings in this report will prepare the research in the master's thesis.

I would like to thank my supervisor Svein-Olaf Hvasshovd, for believing in me, encouraging me and being in general steady and dependable. It has been educational and fun working with this project.
I would also like to thank everyone at NIBIO that has contributed to this thesis and to Jon-Andreas Støvneng, my supervisor at the Intitute of Physics, for all the help.

Contents

List of Figures

List of Tables

Abbreviations

List of all abbreviations in alphabetical order:

- **API** application programming interface
- **CRISP-DM** Cross-Industry Standard Process for Data Mining
- **GIS** Geographic Information System
- **GNNS** Global Navigation Satellite System
- **NIBIO** Norwegian Institute of Bioeconomy
- **NKS** Norsk Kvit Sau (Norwegian White Sheep)
- **NTNU** Norwegian University of Science and Technology
- **SNO** Statens Naturoppsyn (State Nature Control)

1 Introduction

1.1 Motivation

Each year around two million sheep in Norway are let out on outfield pastures to graze during the summer months [1]. Here they will stay mostly unsupervised, before being collected back to the farm in the fall. Each year around 100.000 sheep are lost on outfield pastures, because of e.g. predators, accidents, diseases, poisonous plants, or insects and parasites [2]. The Norwegian Environment Agency reports that in 2020 about 14.700 sheep were confirmed killed and compensated because of predators [3], but the real number is most likely higher. In order to get compensation the farmer has to document or prove to a preponderance of probability that predators were at fault, which sometimes proves problematic since not all carcasses are found, or they may be spoiled to a point where it is hard to determine the cause of death. The authorities will also need to ensure that the sheep's welfare was preserved, and that the death of a sheep is not because of neglect of duty from the farmer [4]. A farmer is entitled to compensation when livestock is killed or harmed due to protected predatory game, and the total cost of payouts from the government in 2020 were 37.7 million NOK [5]. This number has been consistently decreasing these last years, as seen in figure 1.1. Targeted measures by the authorities in collaboration with the farmers are most likely the reason why the number of lost sheep are falling, but it is still a major problem especially in certain areas [5]. Measures taken by the authorities includes designating selected areas for livestock with predatory extraction, and giving out licenses to hunt harmful predators. A graph of compensated sheep by county in 2020 is shown in figure 1.2, where Trøndelag was the county with the highest confirmed sheep loss.

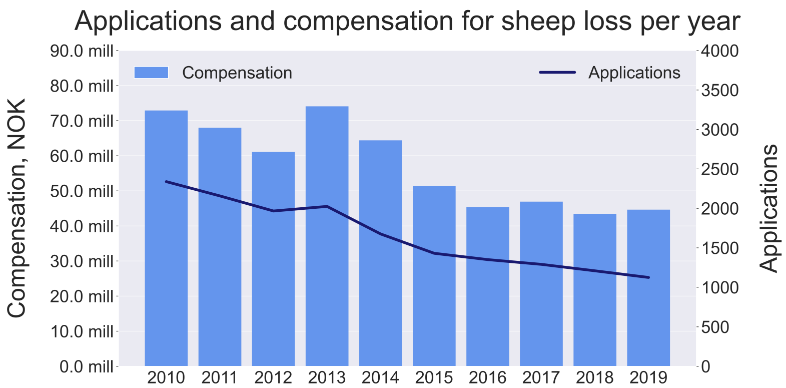


Figure 1.1: The yearly total cost of payouts from the government in NOK due to loss of sheep because of predators, and the amount of applications received for compensation [6].

The predators in focus for this project are those who are protected by government, which are the ones that give reason for compensation to a farmer. Protected predators are in Norway defined as lynx, wolverine, bear, wolf and golden eagle [4]. The parliament sets population targets for these, in order to ensure a sustainable population of predators without it getting too high and thus endangering the livestock industry. There is also a factor of ensuring animal welfare to the livestock, which is in a greater danger of suffering from injury or experiencing a brutal death should the predatory population be too big. The lynx, wolverine and wolf populations in Norway have been well above the targeted level in recent years as shown in table 1.1. This leads to more deaths of livestock than necessary, a higher cost to society because of payouts from the government in compensation, and frustrated farmers.

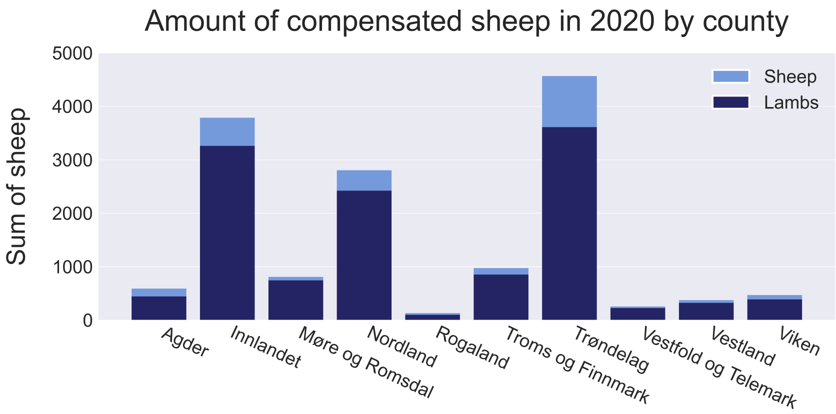


Figure 1.2: The figure shows the distribution of compensation given by the government to each county in 2020 [5].

Predator	Population	Target population	Unit
Wolverine	60	39	Yearly litters
Golden eagle	914-1145	850-1200	Nesting couples
Lynx	67	65	Family groups
Brown bear	8.5	13	Yearly litters
Wolf	8.5	4-6	Yearly litters

Table 1.1: The actual populations and population targets of protected predators in Norway, with numbers from 2021 [7]. The units does not equal the total population, but are estimates based on traces made by the predators. One litter may be of several pups, a family group will consist of multiple animals, nesting couples means a double population of individuals and some eagles will be solitary as well.

It is therefore in everybody's interest to have better systems in place in order to minimize the loss percentage of sheep, to have better control of the herd while it is out on outfield pastures and to better the technology to monitor predators in relation to the livestock.

1.2 Project description

This specialization project is preparational work towards the final master's thesis in spring 2022. It was given in collaboration between the Norwegian Univerity of Science and Tehcnology (NTNU) and the Norwegian Institute of Bioeconomy (NIBIO), with data sets from ewes in Møre and Romsdal and Trøndelag. Usually there are only ewes and their lambs on outfield pastures. The goal of this project is to get a deeper understanding of the possibilities on what to examine further, find out the framework in which to work in, and get the best possible foundation of knowledge for the final thesis based on this. The approach chosen was to research thoroughly sheep behavior and movement patterns, different breeds and their distinctions, sheep reactions when close by predators and predator behavior, and to do an initial analysis of the data sets. The method of the approach was done with the guidance, experience and recommendations from both NIBIO and supervisors. The main objective for the master's thesis is to implement a machine learning program that may be able to determine if there has been predators close by a herd while out on the outfield pastures. To succeed in this, one needs to know what to look for, and what information to feed the program.

As of today, there is not a lot of available data of sheep using electronic transmitters while confirmed being attacked, killed or chased by predators. It is therefore not possible to confidently verify the results with target situations. This implies that further research on more data is needed before any definite conclusions can be drawn. The analysis from this project nevertheless aims towards creating a base on which new research may build on, and point towards what results may look like in cases with predators.

A long term goal for the project of examining the behavior of sheep on outfield pastures by electronic transmitters is to develop a more secure detection of digital stress factors in the sheep and predatory identifiers. If this can be implemented directly into the transmitters and notify the farmer in real time, unnecessary sheep loss may be prevented.

1.3 Stakeholders

A. Farmers: Farmers lose sheep every year, and both from an ethical and an economical perspective reducing the number of lost sheep due to predators will be a major objective. It demands many resources from the farmer each year to ensure the welfare of the herd, track down lost individuals and report killed sheep, all of which a digital solution that gives more control might drastically simplify.

B. The government: There are several departments from the state that handles sheep loss from predators and livestock welfare in general, including the Norwegian Environment Agency, the Norwegian Food Safety Authority, the County Governor and the State Nature Control (SNO). The government also pays out many millions each year in compensation. Streamlining this work will reduce costs and make the process smoother for the government employees and farmers both.

C. NIBIO: NIBIO has been doing research on digital solutions and data analysis for sheep and predators on outfield pastures over several years. As a collaborator to this project, getting specialized computational knowledge and skills from students to help their research might prove very valuable to their purpose of sustainable resource management.

D. Producers of electronic transmitters for sheep: On the marked in Norway there are several providers of electronic transmitters for livestock on pastures. Their solutions will sometimes include a parameter for stress detection from excessive movement or death notifications from lack of movement, but will not discriminate on play or fear. It will be in their interest to further the technology on what types of data analysis are possible to incorporate in their products and thus improve what services they can provide.

2 Method and data

The method of how to approach the data will affect the results greatly, so it is important to choose a strategy that fits to the objectives to be solved. One of the most used strategies in data analytics is the Cross-Industry Standard Process for Data Mining, henceforth called CRISP-DM [8]. This methos was chosen as it is an appropriate fit for the project, and is well documented and tested.

2.1 CRISP-DM

The CRISP-DM method consists of six data project phases in a cycle. The process will be agile and iterative, and as more knowledge is gained from the analysis the steps may need to be adjusted and repeated.

1. Business understanding: The first phase focuses on thoroughly understanding the business, its environment, and the business goals and objectives. Here the business refers to the industry of sheep farming in Norway on outfield pastures. The phase also includes deepening the domain knowledge on the data to be analyzed, which here implies researching the behavioral traits in sheep and predators. The first phase should be used to understand the problem, and how may that be answered by data science. It also includes concretizing the issue to be solved and to reduce the scope of the initial idea, if necessary.

2. Data understanding: The major part of any data science project consists of data preparation [9]. This involves collecting the data, describing it, exploring and visualizing it, and verifying its quality.

3. Data preparation: This involves selecting the relevant data to be used, cleaning the data for errors and checking the anomalies, constructing new attributes and features (feature engineering), integrating all data sources and updating to a convenient format.

4. Modeling: In this phase the model design is chosen, built and implemented, and assessed. The steps up until this phase is revised and iterated until the model is performing to a satisfactory level based on prior domain knowledge.

5. Evaluation: The overall evaluation focuses on results, the process and further work. Are the results presented clearly, are they applicable to the chosen objectives, and what are additional questions that arise? There should also be a review of the process, and what could have been done better. The next possible actions should be considered, and if the business understanding or objectives need to be refined the adaption of the process for the next iteration should be discussed.

6. Deployment: The last phase consists of summarizing the results and findings, plan for maintenance or expiration if necessary, document the project, and conduct an overall review of what went well, what to improve, and what were the takeaways.

The specialization project will mainly answer to the first two steps, while beginning the data cleaning for the next phase. Data cleaning is a time consuming process, and to prepare all data sets readily for usage, and updating as new discoveries are made, will continue throughout the thesis.

2.2 Machine learning

Machine learning is a category of artificial intelligence which applies statistical analysis and mathematical models on data, typically in large amounts, to make a computer detect trends or make predictions [10]. It is said the computer learns itself the patterns in the data, instead of it being explicitly programmed. Tom Mitchell defined it formally as "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E." [11]. The data will be split into a training set and a test set, where the computer will tailor the machine learning model with the gained experience during the training phase, and thus optimize it's performance on the test data. The training process involves tuning the parameters used in the model that will later make the predictions, while the test phase will verify the accuracy and present the results of the program made. The different individual variables or measurables in the data are called features, and it will be critical to emphasize in the analysis independent features that are the most important for the outcome. There are also hyperparameters that control the training procedure in itself, working as configuration variables on how the model will run to train. These are usually constant throughout and given beforehand.

Supervised learning is a category of machine learning where the training data also contains the desired output, called labels, that the model later will try to predict on new data. The model parameters will get feedback on the accuracy from comparison with the labels during training, and may therefore adjust their value accordingly.

In unsupervised learning there are only input data, and the computer must discover itself similarities and patterns in the given material. Important unsupervised tasks include association rule-mining, clustering and anomaly detection. Discovering probabilities of dependencies and co-occurrences of features, and finding their relationships, is called finding association rules. Clustering groups instances into different categories called clusters where the components within are more similar to each other than to those in another cluster. Anomaly detection will identify instances of rare events in data sets with a large amount of normal observations, if it significantly differs from most [12]. In semi-supervised learning, there is only a small amount of labeled data and a large amount of unlabeled data. The computer will first train with the small amount of labeled data, before applying what it learned on the unlabeled data and try to extrapolate. This might still provide valuable results and accurate predictions [13].

A key principle of machine learning is feature engineering, which involves adjusting and designing the input of the model to improve accuracy. Methods of feature engineering might be feature creation of new features by the use of domain knowledge and the raw data, feature extraction of the most important features into the model, removal or imputation in cases of outliers or missing data, or normalization and standardization of values [14]. It is important that when building a machine learning model and applying feature engineering that the model is neither under- nor overfitted. Underfitting a model means it does not have enough valuable information to find a pattern, while in cases of overfitting the model is too complex and it starts learning patterns also with noise and inaccuracies. Finding the right balance between complexity and impartiality will make a model more precise.

2.3 Tools

The programming language Python was used for the data exploration. The library *Pandas* was used for handling the data, which is a software library especially written for Python for data manipulation and analysis. It has a DataFrame object for easy data manipulation which includes tools for reading and writing data from .csv files, which are used in this project. All the plots included in this report were made with *Matplotlib* and *Seaborn*, which are data visualization libraries for Python. Inkscape was used for the illustration. All code files are collected at a Github repository, which can be found by appendix A.

2.4 Data sets

2.4.1 Tingvoll

The main data from NIBIO came from sheep in Tingvoll, Møre og Romsdal, on two farms in Koksvik and Torjul. The herds consist of Grey Trønder sheep at Koksvik, and Spæl and Norwegian White Sheep at Torjul. The data had a total of 89 separate data sets, where two were empty, and there were around 450.000 lines of data points in total. The data spans over five years from 2012 to 2016. Eleven sheep repeated wearing a transmitter for two years, and one wore it for three years.

Checking for repeated patterns of movement through the season for the same sheep will be important in order to look for routines and tendencies. None of the sheep died while out on summer pastures. There are 21 columns of information in the data, where the most important to be used are:

- **Date:** Date of point generation
- **Time:** Time of point generation
- **Lat:** Recorded latitude of point
- **Long:** Recorded longitude of point
- **X and Y:** The magnitude of changes in acceleration due to applied force from movement, given by an accelerometer. The higher the value given, the more activity by the sheep. The values are calculated during the time it takes to fix a point.

2.4.2 Fosen

The data from Fosen, Trøndelag, includes information from three herds with a total of 192 separate sets. The first herd consists of Norwegian White sheep, the second of Norwegian White and Old Norwegian sheep, and the third of White Spæl. The data spans three years from 2018 to 2020, and none of the sheep repeated wearing a transmitter during this time. Thirteen sheep died while on the pastures, where four causes of death were assessed as due to predators, the rest as diseases or unknown. There are a total of about 580.000 lines of data and 10 columns of information, where the most important to be used are:

- **Datetime:** Date and time of point generation
- **Lat:** Recorded latitude of point
- **Lon:** Recorded longitude of point

2.5 Latitude and longitude

Latitude and longitude are geographical coordinates, and in Geographic Information Systems (GIS) they are often given in decimal degrees, as with the data from Tingvoll and Fosen. Latitude is the degrees in the north-south orientation, and thus represented by horizontal lines called parallels, and goes from 0° at the equator to ±90° at the poles. Longitude is the east-west orientation, and thus represented by vertical lines called meridians, and goes from 0° at the prime meridian in Greenwich to ±180° east/west [15]. Latitude and longitude are illustrated in figure 2.1. When moving along latitudinal lines with an unchanged longitude, the trajectory will slice the earth through the center along the meridian, but when moving along longitudinal lines the trajectory will depend on the latitudinal value.

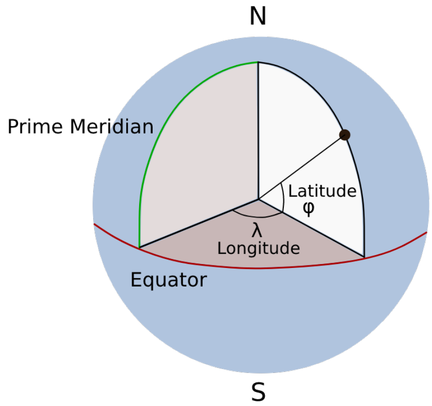


Figure 2.1: Illustration of the earth, and how latitudes and longitudes are calculated with respect to the equator and prime meridian.

The earth in itself is approximately spherical, but because the earth is spinning about a fixed axis though it's center, the fictitious inertial centrifugal force compresses the middle of the earth outwards, and deforms the earth into an obloid spheroid sometimes called an ellipsoid of rotation [16]. This deformity affects how latitudes and longitudes are computed in a Global Navigation Satellite System (GNNS) and a GIS, which need a reference in approximately the same shape of the earth to calculate values. This reference ellipsoid for a given coordinate system should be specified as in accordance with the ISO 19111:2019 standard in order to be sufficiently accurate [17]. Most world wide services like Google Earth and OpenStreetMap, which are 3D geographical coordinate systems that can chart coordinates, uses the WGS84 geodetic reference frame (datum) to define the ellipsoid reference of earth [18, 19]. The identifier used to chart coordinate values on the WGS84 ellipsoid is EPSG 4326 [20]. Note that this is in relation to 3D-mapping of coordinates to locations, when applied to the flat 2D plane on projected coordinate systems like maps, the identifier EPSG 3857 for coordinate transformations is used for WGS84 [21]. In Norway and other European countries, the EUREF89 reference frame is also used, which is based on WGS84 and Europe's tectonic position in 1989. Because of continental drift, Europe and America moves a couple of centimeters apart each year, increasing the difference between EUREF89 and WGS84 [22]. For the data from Tingvoll and Fosen, the exact GNNS and reference frame are unknown, which may affect results in especially altitude calculation. For this project it will be assumed that GPS and WGS84 are used, based on probability, but the difference from other standards are not deemed significant. Further on, the major part of this project will depend on positional differences and qualitative analysis, not exact location, and therefore be independent of reference frame as long as the calculations are consistent.

$$\begin{aligned} a &= \sin^2(\Delta\phi/2) + \cos(\phi_1) \cdot \cos(\phi_2) \cdot \sin^2(\Delta\lambda/2) \\ d &= 2R \cdot \arcsin(\sqrt{a}) \end{aligned} \tag{1}$$

If the distance is less than about 200 km, the earth may be assumed spherical when calculating the length between points [23]. The Haversine formula given in equation (1) below is a method used to calculate the distance traveled by coordinate values and taking into account the curvature of a spherical surface [24]. The Haversine is defined as $hav(\theta) = \sin^2(\theta/2)$, R is the sphere radius and $\Delta\phi$ and $\Delta\lambda$ are the differences in latitude and longitude, respectively. This formula is an approximation, because of the obloid curvature of the earth the Haversine formula may still have an error up to 0.5 % [25]. Vincenty's formula will take into account the reference frame, most commonly WGS84, but it is computationally heavy, and for a qualitative analysis on relatively small distances not necessary [26].

Some error in GNNS values must be assumed, and could be caused by for example signal noise, clock offsets between satellites, atmospheric conditions, or attitude in orbit inclination. But for e.g. the American system GPS, this fault will only account for a couple of meters [27].

2.6 Data cleaning

The first part of cleaning data sets is to visually inspect and get an intuition of what the data looks like. Using a csv-reader one may look for obvious errors such as error messages instead of generated points by filtering on letters. A common occurrence for point generation errors is that latitude and/or longitude are set to zero or to a far off point from the rest. The extreme values of each set must be checked and evaluated. The distance between two consecutive coordinates must be checked for the plausibility of the sheep traveling that far. Incorrect values in the data may either be deleted or imputed with a new value. To ensure consistency of the data so that the analysis is comparable between different sets, a start and end date must be set if not all span the same time frame. Errors in date and time generation must be checked. In addition the size of the data sets should be evaluated. After a fixed time frame is appointed the sizes of all sets should be in the same order of magnitude with each other. A visual inspection will give an overview and an insight to whether the data looks correct and reasonable.

2.7 Interviews and feedback from farmers

Personal experience from farmers who have had sheep on outfield pastures and themselves been exposed to predatory attack on their herd may provide useful insight. Through contact on sheep farmer network groups on Facebook and acquaintances of the partners at NIBIO, farmers were directly contacted and interviewed. This was done in two iterations: first, a nation wide network group were asked open questions to answer at will, and next a group of sheep farmers from the same area as the data sets were interviewed individually.

The open questions asked to the nation wide network were:

- Do you notice any specific behavior from your sheep when you know there might be predators nearby?
- Do you notice differences on this when it comes to breed, with or without lambs, the type of pasture etc.?

- How will the sheep behave normally, compared to when it is sick, is with lamb(s), or other?
- Anything else you can tell me about the behavior and movement pattern of sheep?

The feedback given can be found in its entirety in appendix B.

The questions for the in-depth interview with local farmers were:

- Where are your farm and outfield pasture located?
- Which breed of sheep do you have?
- How will you (shortly) describe the normal behavior of your race of sheep? [If more than one breed: are there any differences?]
- Which predators have your herd encountered, to the best of your knowledge?
- How did the sheep react in the situations where you know predators were nearby?
- Do you notice any difference from what type of predator that is nearby?
- How will the sheep react to human strangers, and/or dogs?
- How often do you inspect your sheep on the outfield pastures?

The feedback given can be found summarized in it entirety in appendix C.

3 Movement patterns of sheep

There are many things that may influence the behavior of how a sheep acts naturally when on outfield pastures, without the security of human protection. A special field of interest in this project is how sheep react when in contact with predators, compared to their normal, relaxed behavior. Since the end goal is to differentiate these two, and hopefully be able to see tendencies of this in the data, it is very important to understand what to look for, and what affects it. Parameters that may influence the nature of sheep behavior include type of breed, hierarchy order, environment and weather, age and gender, resources available and individual personality. The findings on factors influencing the temperament and movement pattern of sheep are presented below.

3.1 Breed of sheep and it's impact

It has become apparent that the breed of sheep will influence their behavior, especially on outfield pastures. Sheep has been domesticated and bred by humans for at least 7 thousand years [28], and trait selection has long been a part of breeding the animal. Especially in more modern times, when sheep farming has gradually evolved from being a livelihood within a family, to be a nation wide economical and sometimes industrial occupation. Traits that maximize profits and streamlines operations are often more wanted and bred upon. This includes traits like flocking less, growing better quality wool, more muscle for more meat per sheep, and more friendly and social towards humans [29]. This will in turn affect their survival instincts on outfield pastures, and make them more dependent on humans as a species. Research indicates that the more bred upon a sheep breed is, the more tame and heavier it will be, and thus more weakened instincts to help guard themselves against predators [30]. However, breed differences and their antipredatorial behaviors are generally poorly documented.

3.1.1 The main Norwegian breeds

Spæl: Originating from one of the oldest breeds in Norway, the Spæl sheep is still considered to have a close resemblance in manner and physical features to the earliest domesticated Norwegian sheep. It has been somewhat crossbred during the first half of the 20^{th} century creating different kinds of Spæl. The main types are Old Norwegian Spæl, White Spæl and Colored Spæl. It has less meat content and fat than most other breeds, but has stronger mother instincts and milks very well. Despite it's slightly smaller and fragile build, it is light on it's feet and robust, more vigilant and runs fast. While on pasture they gather in a herd [31, 32].

Old Norwegian: Also called Wild sheep or Stone age sheep, like Spæl it stems from the old original Norwegian breed. It is small but highly cautious and sturdy, with strong mother- and herd instincts.

Norwegian Fur: Relatively new breed from the 1960s, and slightly larger and heavier than Spæl. It has good mother instincts, and will roam widely on pastures.

Dala: One of the largest breeds, and has a calm, tame demeanor. It has some difficulties with breeding, and will often scatter more on pastures.

Norwegian White Sheep (NKS): This is a crossbred sheep consistent of different both Norwegian and foreign types. Lately some of the other Norwegian breeds like Dala, Rygja and Steigar are sometimes being classified as NKS instead due to crossbreeding. It is a fast growing and heavy breed, very fertile and social towards humans, making it the most populous breed in Norway. It will often walk separate on pastures and not flock together.

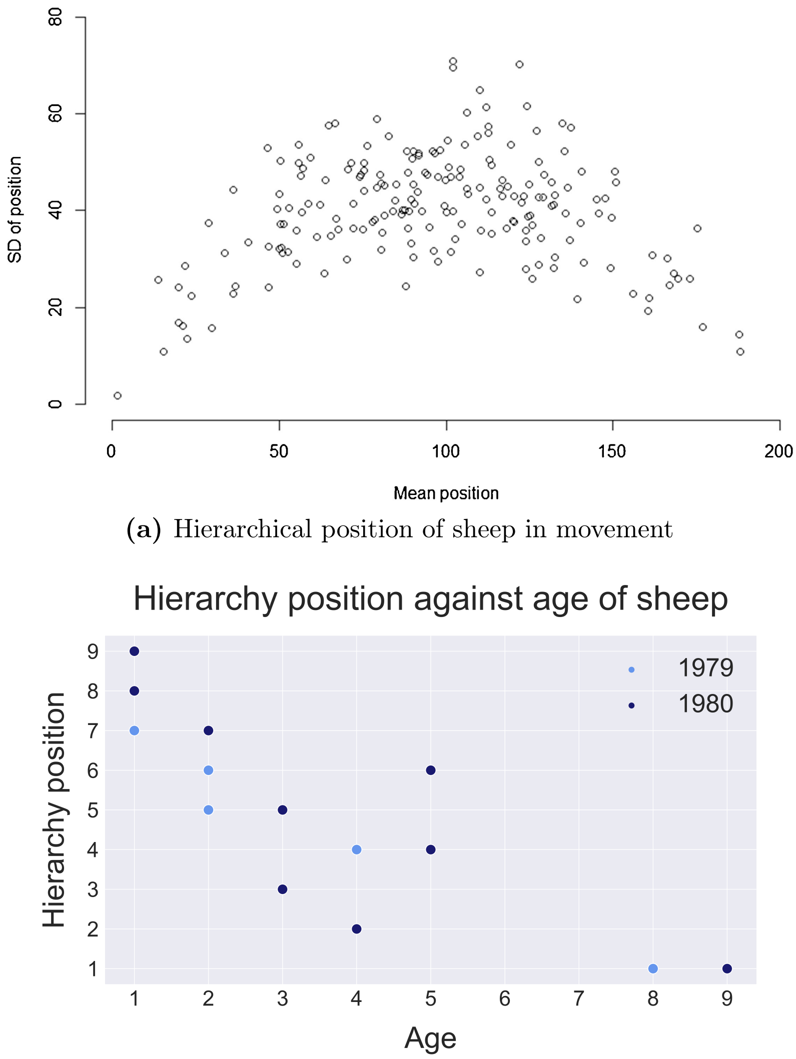
Grey Trønder: A cross between Old Norwegian and Tauter from Trøndelag, with a medium weight. It's alert, frugile and has good herd instincts.

3.1.2 Reaction to predators by breed

Research on the antipredatorial reaction in ewes of several typical Norwegian breeds were done in Trøndelag [30]. The breeds were Old Norwegian Sheep, Spæl, Norwegian Fur Sheep, Suffolk, Steigar Sheep, and Dala sheep. Their reaction was recorded against the threats stuffed wolverines, lynx, and bears, humans and dogs, and big unfamiliar objects. This study confirms that lighter breeds have stronger instincts and reactions against predators than heavier breeds. The breed that had the longest de-reaction time, longest flee distance, most defensive behavior, and flocked the most together was the Old Norwegian Sheep. After that came Spæl and then Fur sheep, and the Fur Sheep was also the most offensive towards the predator types. The rest of the breeds had no to little differences in reactions between them. Steigar Sheep flocked the least together when threatened. The predator figures that initiated the longest de-reaction times were in decreasing order dog, lynx, wolverine and bear. The ones that stimulated the longest flee distance were in decreasing order wolverine, bear, lynx, and then dog. Human and large objects fell significantly below the other threats in reactions. There was also stronger reactions during the fall experiment than the one in spring time.

3.2 Hierarchy and herd behavior

For simplicity when referring to a herd it is the whole collection of sheep belonging to a farmer, while when referring to a flock it will mean a smaller portion of the herd like a family group. When on outfield pastures, the adult sheep are usually only ewes. In a study conducted at the University of New England, Australia, herd position during movement and hierarchical and temperamental differences were examined in adult Merino ewes [33]. The movement was enforced by high value food motivation along a track. Every sheep was given a lameness score each test day, to map their positional variance against their health status and mood. The research show strong correlation in especially preceding runs, and hints towards strong front- and rear hierarchical order. In figure 3.1a, the mean position of each sheep relative each other and the standard deviation of their position are given. The curve plainly shows that the further to the front or back in the herd an ewe is located, the more consistent their relative position is. For the sheep in the middle of the herd, the standard deviation is both higher and varies more, which may indicate that hierarchy is mostly important for sheep with either strong leadership or follower traits. The study also found that for days when sheep showed lameness, their positional mean were 20.5 ± 5 % further back than on days they were not deemed lame. Tømmerberg (1985) followed a herd of Dala sheep over two years on outfield pastures, and recorded their rank and compared it against their age, as seen in figure 3.1b [34]. Here there is a correlation between older age giving higher rank in the herd. He also found that once hierarchy was established in spring, this stayed constant for the whole season out on the pastures.



(a) Hierarchical position of sheep in movement

(b) Hierarchical position by age of sheep

Figure 3.1: Figure (a) shows individual mean position of each sheep (n=196) for all runs and their standard deviation (SD) (figure taken from [33]), and (b) shows hierarchy position of one herd over two years compared to age, where a low position means higher rank [34].

Similar results for Border Leicester sheep and other Merinos have also been found, but more extensive research is needed in order to conclude generally about the spatial leadership for sheep [35, 36].

3.3 Habits of sheep

3.3.1 Home range

Sheep are habitual animals, and when roaming on outfield pastures they will mostly keep to a fixed area each year called home range [31]. William Burt (1943) defined it as "...the area, usually around a home site, over which the animal normally travels in search of food. Territory is the protected part of the home range, be it the entire home range or only the nest." [37]. The extent of the home range will vary with grazing qualities, topography, predators, and quantity of other grazing animals. Tømmerberg (1985) found that the home range increased later in the season, which may be explained by that the lambs are older and need more nutrition, and vegetation grows more slowly towards fall [34]. Each individual sheep or small family group will have their own home range, and will also follow more or less the same route every year. This route will pass down from ewe to daughter, and stay approximately consistent throughout the generations [34]. Trying to change pasture may prove difficult, as the sheep often will return to the area they are familiar with and grew up with. If home range is changed it is most likely due to disturbances, like predators or human intervention [38].

3.3.2 Weather

The movement pattern of sheep will be affected by the weather. On sunny days they are more likely to be in higher altitudes, probably to seek more wind and cooling temperatures. Likewise, they keep to lower grounds if it's cold or rough weather. They roam the most on dry, cloudy days, while staying more put if it's either very hot or stormy. Their behavior is not influenced by a normal amount of rain, but should it rain a lot they will usually find shelter in the form of vegetation or natural formations. When there are a lot of insects, typically when sunny, hot and calm winds, they will move onto higher grounds to try to find more wind to get relief from the nuisance. One behavior to note here is that when bothered by insects sheep will regularly shake their heads and fidget more, which may affect the GPS positioning or activity level. At very hot days, grazing at night will increase while the sheep instead will relax more during the day. Scott and Sutherland (1981) found for Merino sheep that grazing were at a maximum around 10-15°C, and relaxing and shade group formations increased noticeably at temperatures above this [39].

3.3.3 Daily routines

The major part of the day is spent alternating between grazing and chewing cud. Grazing will occupy up to 7-11 hours each day, and cud chewing about 5-9 hours [40, 31]. This depends on the quality of the pasture relative the amount of sheep and time of the season. The most active grazing period will be at dawn and dusk, while the sheep will rest during the early afternoon when temperatures and sun are peaking [34, 39, 41]. Towards night the herd will usually seek higher up in the terrain, and come down again in the morning.

3.3.4 Movement patterns and seasonal habits

Typically the sheep will be released on outfield pastures around the end of May/beginning of June, and be gathered home again in the middle of September. Depending on the breed, a herd will sometimes split into smaller flocks of family groups out on the pastures. As discussed in 3.1.1 the larger and more docile breeds like NKS will split more than e.g. Spæl. When a flock is grazing together they will often move in a wide formation forwards, but if a flock is migrating to another area they will walk in a row behind each other instead. They seek secure places for when resting or chewing cud, either higher up or around vegetation. Their route throughout the season will often stop by mineral blocks if laid out. Migration of a flock is mostly motivated by better vegetation and grazing opportunities elsewhere. At the end of the season a herd will often return homewards by itself, from where it had been grazing. This will also often happen if it's scared by predators [42, 34]. Activity during the night in the late season significantly decreased, which may be explained by darker and longer nights. This is in accordance with the home range also increasing later in the season. Around 90 % of this active time were used for grazing, the rest for walking, running, nursing etc.

3.4 Diseases and behavior when sick

When sheep get sick while on outfield pastures, they may fall behind the flock, become lethargic, have less energy and generally move more slowly than when healthy. This could be because of several diseases, like scrapie, hypokalemia, coccidiosis, or anaplasmosis [31]. Alveld is a disease that arise by consumption of the plant bog asphodel, and will make the sheep sensitive to light and uneasy, and they will often seek shade. Lambs are typically more susceptible to get sick.

3.5 Herd reaction to predators

Tømmerberg observed the herd running upwards in the terrain when frightened by something, and hypothesized that the shift to higher altitudes at night is a measure to better be able to detect approaching predators when dark [34]. This argument is further supported by that this behavior largely followed the length of the day, where the move upwards started earlier in late summer when the sun set earlier. The given feedback from farmers stated that sheep will return either home to the farm if close enough, go to roads or cabins, or seek towards water when predators are nearby their pasture. They will be noticeably stressed and restless. One report from the Norwegian Institute for Nature Research in 2016 found an indication of correlation between heightened deaths because of golden eagles and deaths because of diseases and accidents on outfield pastures, even though the data did not provide a basis to conclude [43]. It is worth noting that a prevalence of more diseases in a herd might cause easier prey and hence more predatory deaths, or that the presence of predators might increase the chances of diseases.

4 Predatory behavior

A short description of the hunting behavior of different predators present in Western Norway and Trøndelag is given below. The total loss of sheep in Norway due to the different predators is given in figure 4.1.

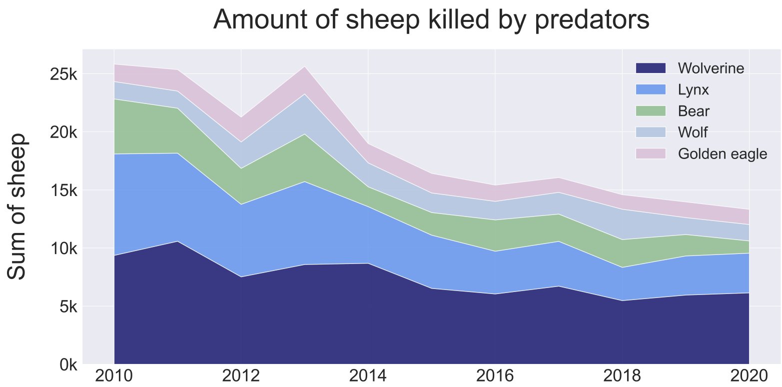


Figure 4.1: The figure shows the amount of sheep lost and compensated due to the different predators since 2010 [3].

4.1 Wolverine

The wolverine is one of the largest threats to sheep on outfield pastures, as shown by figure 4.1. They will kill both lambs and adult sheep, and also eat carrion. They are known to hoard their kills to save for later, and might move them many kilometers from the attack site [44]. The Wolverine will kill the most in the fall [3, 31].

4.2 Eagle

The golden eagle is the main eagle predator for sheep in Norway. The nesting season starts around the beginning of April, and the eaglets leave the nest in July. They mainly hunt lambs and not adult sheep, so the danger of eagles for a herd is mostly present during the first half of the season when the lambs are still quite small. They may also prioritize sick or injured individuals, or eat carrion. They watch their prey either by flying high or low depending on topography and weather, or from perched on a branch. It then dives with great speed and momentum, landing on the back or neck of the lamb, puncturing the victim's prey with their talons. It is not certain the herd is always aware an eagle is nearby, before it has made an attack [45, 46].

4.3 Lynx

The lynx will hunt by sneaking up on their prey from behind, and attacking and killing in a swift movement. Unlike the wolverine, a lynx will not hide or hoard their killings or move it over great distances [47]. Research on Eurasian lynx showed that their roam range increased by more than 10 % during the night, and that they chose more open habitats with higher occurrence of prey available compared to daytime [48]. This implies that lynx might hunt more often at night. Like the eagle, the lynx will hunt the most in the early season [31].

4.4 Bear

The diet of a bear will mostly depend on berries and plants during the fall to increase the fat reserves for the winter, and prey in the spring and early summer to build muscle mass after hibernation [49]. They will hunt by chasing and swatting prey, either while initially hidden or visible out in the open. There will often be several chases before a successful kill, where the bear uses the first chases to look for either young or infirm sheep to target. Bears will sometimes kill excessively, and leave their catch uneaten [50].

4.5 Other

Farmers have also reported loss of sheep due to unleashed dogs or strays [51]. When sheep are chased by dogs they may die of exhaustion and panic, as also several farmers interviewed claimed to be a problem especially during the hunting season. Humans might steal sheep as well, but this is not deemed a common occurrence.

5 Data visualisation and statistics

As per the CRISP-DM strategy it is important to understand the data before starting the analysis, and get an intuition about what it means and looks like. Visualization of the relevant features is a useful tool to be aware of what initial information the data contains. First a plot of what the home range for the different herds looks like was made. The pattern will look different for different sheep in the same herd, as they have their own individual regular routes. An example of the home ranges is shown for Fosen, Koksvik and Torjul in figure 5.1. The map tiles were collected from OpenStreetMap [52].

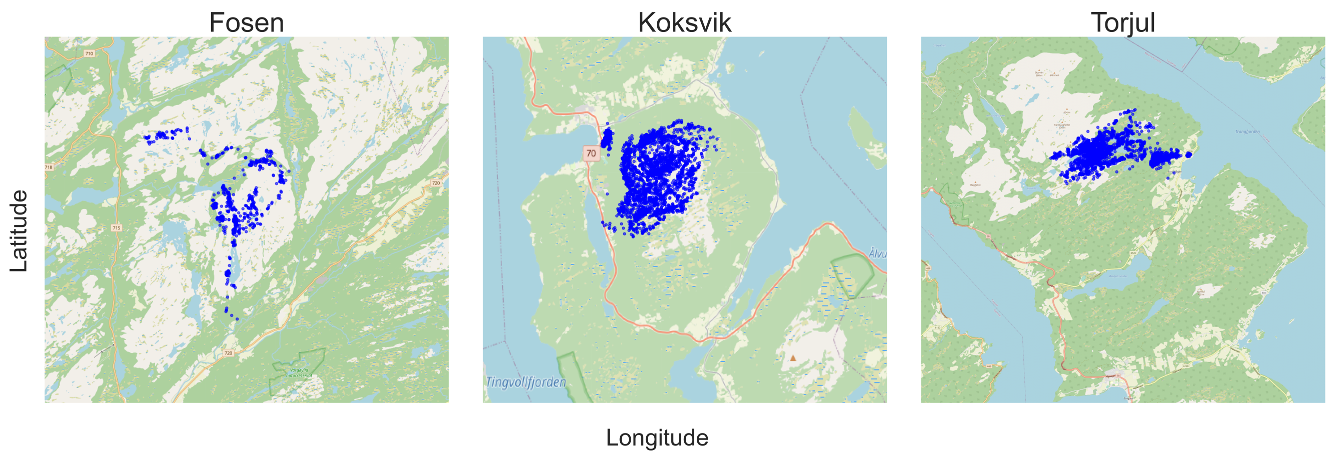


Figure 5.1: Examples of home ranges one season for three sheep in Fosen, Koksvik and Torjul, respectively.

The information from the accelerometer in the electronic transmitters on the sheep in Tingvoll will give important insight to what activity levels are normal and excessive for sheep. A plot of the values of the accelerometer for one sheep in 2012 is given in figure 5.2, and a table of statistics for all years given in table 5.1. Std is the standard deviation, and the percentages represent the lower (0.25) and upper (0.75) percentiles and the median (0.50). A quarter of all values fall below the lower percentile, and three quarters fall below the upper percentile. The total is calculated by The Pythagorean equation as X and Y are orthogonal acceleration vectors [53]. Note that there are no negative values, so the feature only gives magnitude and not direction of the activity change. The histogram plots in figure 5.3 are given for the total, and the nonzero total which removed the rows where the total was zero and thus had at least one value of activity in either X or Y.

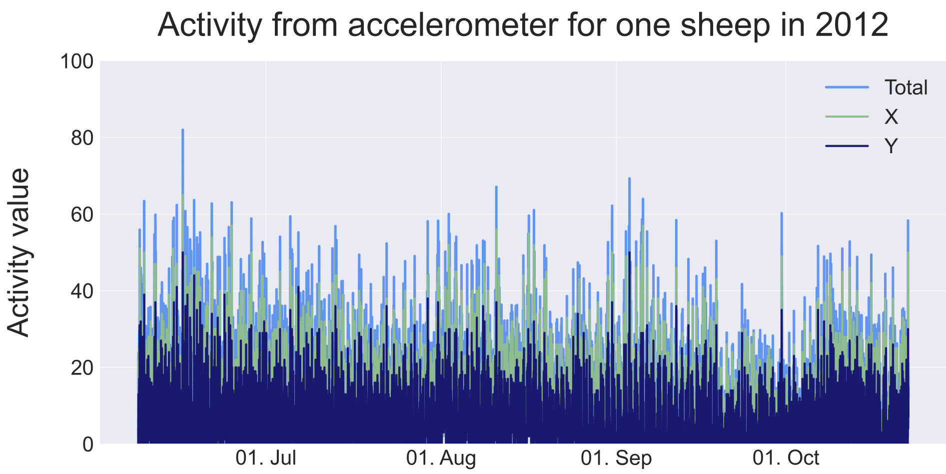


Figure 5.2: The figure shows the measure of activity from the accelerometer in X, Y and the total, from one sheep grazing at Koksvik, Tingvoll, in 2012.

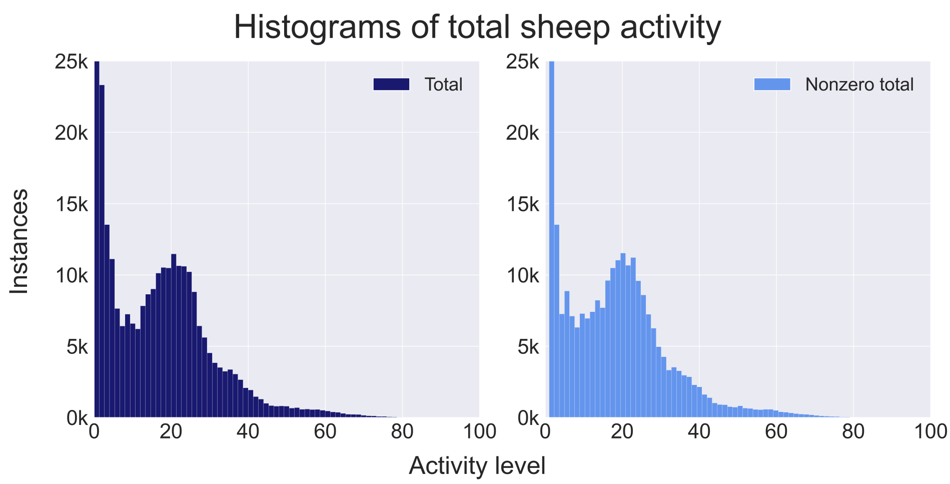


Figure 5.3: The total and nonzero total of the activity at Tingvoll for all years. The zero value for the plot on the left goes up to about 160k, while the value of 1 on the plot to the right goes up to about 55k.

The data from Tingvoll had a measure of the temperature taken by the electronic transmitter, but this will be influenced by sun exposure and radiation of heat from

Statistic	X	Y	Total	$\text{Total}_{\neq 0}$
Mean	9.224	6.484	11.501	16.900
Std	11.297	8.333	13.853	13.811
Min	0.000	0.000	0.000	1.000
Max	100.000	88.000	133.207	133.207
25 %	0.000	0.000	0.000	4.123
50 %	3.000	2.000	4.472	16.125
75 %	17.000	11.000	20.616	24.698

Table 5.1: Statistics for the activity registered from the accelerometer by all sheep at both farms in Tingvoll, for all years. The parameters that are measured are X, Y, the total and the nonzero total.

the body of the sheep. Instead temperature data was collected from Norsk Klimaservicesenter [54]. The closest weather station to Koksvik and Torjul is "Sunndalsøra III", which is located about 30 km further east into Tingvollfjorden from Tingvoll, six metres above mean sea level. Koksvik and Torjul are closer to the open sea, and higher above sea level, but the approximate weather should be enough for a qualitative analysis.

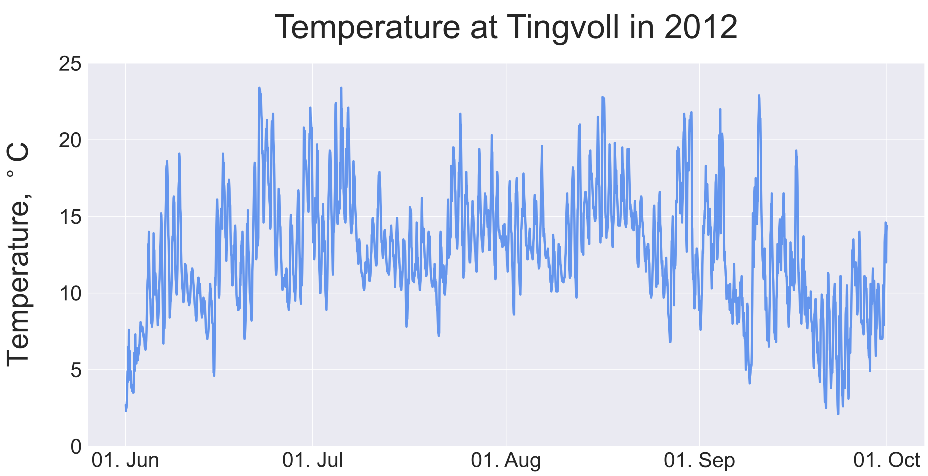


Figure 5.4: The given temperatures at the Tingvoll weather station "Sunndalsøra III" in 2012 [54].

There was a measure of altitude for each point in the data from Tingvoll, but also this had a bad quality and had to be collected from an outside source instead. There was no altitude feature in the data from Fosen. An Application Programming Interface (API) call to Kartverket, which is an open source distributor of geographical data, fetched accurate altitudes based on the coordinates [55]. With this API it is possible to request what reference frame and identifier to use to calculate the altitude, which here was set to WGS84 and EPSG 4326. A plot of one sheep's altitude over one season in 2013 is shown in figure 5.5.

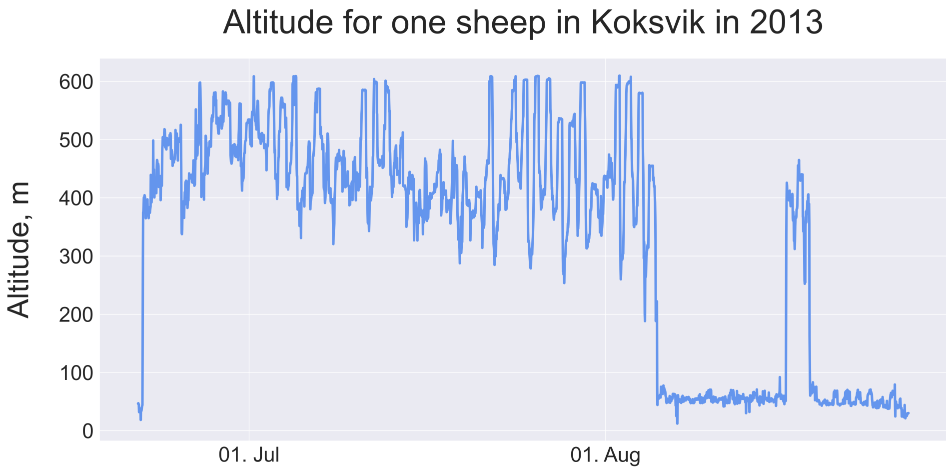


Figure 5.5: The measure of altitude in meter above mean sea level for one sheep in 2013 [55].

To get intuition about the order of magnitude of the size of the data, the number of points in each data set was plotted as shown in figure 5.6. A threshold value for when a set has too few data points to be included in the analysis is proposed to be about 10 % of the average. The average size of a set in Tingvoll is 4644 points, so the threshold was set at 500. Note that this is before the data cleaning is completely done, so the final average and cut-off value will most likely be lower.

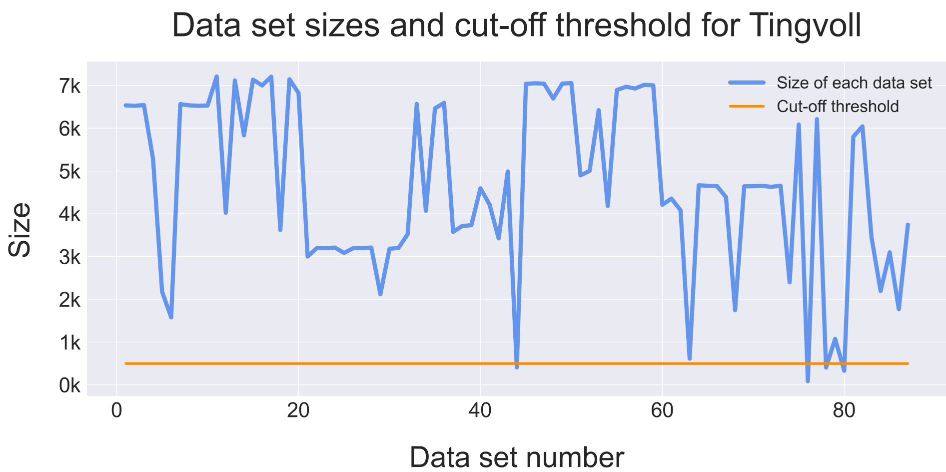


Figure 5.6: The number of points in each data set in the data from Tingvoll, compared to a proposed cut-off threshold value where the sets are too small to be included in the analysis.

6 Discussion

Using electronic transmitters for predator analysis around sheep herds is a relatively new concept not yet thoroughly researched or documented. As a consequence of this, there is an insufficient amount of available data from transmitters on sheep with confirmed death, injury or chase from a predator. The analysis done in this thesis will be based in great majority on data on sheep that lived through the season, without knowing when a predator has been close by. This means unsupervised machine learning must be used. There are however a few valuable cases of confirmed deaths by a predator, which might make it possible to also try a semi-supervised model. The next semester there is a chance of getting access to more data with labeled cases. The goal of the project is to try to separate the sheep into cases of normal behavior, and in closeness to predators. Without more true labels, a machine learning model must work through the data and mostly try to find the behavior patterns of sheep itself. In later analysis the results from this may be compared to the literature study done in sections 3 and 4 to see if there are any clear cases of predators during the time the sheep spent on outfield pastures. An advantage of acquiring more labeled data, and strengthening the supervised part of the model, is that it will most likely be more certain of the patterns related to predatory behavior. With unsupervised learning there will be more guessing and interpreting of the results since there are no guarantees that there for a fact were a predator nearby at that time. There should be tried different approaches to separate the observations, like association rules, clustering, anomaly detection, semi-supervision and combinations of these. Regardless of model chosen the results presented need to be interpreted in light of acknowledged theory and the domain knowledge acquired.

Behavior and instincts for different breed of sheep will have an impact on their reactions, and it is expected that the more cautious and vigilant breeds of Grey Trønder, Spæl and Old Norwegian sheep will have a higher activity level and be more sensitive to changes in their surroundings compared to Norwegian White. This should be something the machine learning model differentiate on. If all data sets are treated equally regardless of breed, the program might report there has never been predators close by NKS sheep from a relative lack of reaction, or report predators close by the lighter breeds more often than true. Initially the model should focus on the older, lighter breeds, as it is expected more prominent and noticeable reactions to predators. If the model is able to identify a behavioral pattern, the same may be tried with the heavier breeds like NKS.

Both the study in section 3.1.2 and feedback from the interviews with farmers mentioned dogs as slightly problematic. It is probable that sheep will react strongly when meeting dogs on outfield pastures, which is not necessarily an uncommon occurrence as people hike in the mountains during summer. Loose dogs were reported to be the most problematic, as they might chase the sheep until it dies of exhaustion. The study also found that the reaction of the herd were stronger in the experiments in the fall than during spring, which may be explained by the herds recent experiences of encounters with predators and/or dogs on the outfield pastures. Regardless, it is important to keep in mind that not all reactions the sheep get need to be because of predators or that they are in danger.

The consistency of the home range for individual sheep might be very valuable in an analysis of predatory proximity. If the route of a sheep deviates greatly from their usual home range, there is a good chance it could be because of predators in the immediate area. This requires that there are enough data of the same individual over several years in their usual pasture. The data given so far does not have many repetitions of the same sheep, but the few in Tingvoll should be investigated closely.

Tømmerberg's hypothesis that the herd flee upwards in the terrain or home towards the farm when scared implies that the altitude level of the sheep might be an important feature for the machine learning model to take into account. The data from Tingvoll had a measure for altitude, but it was declared that these values were not of high quality. Accurate values for the altitude for each pair of coordinates was therefore collected. From figure 5.5 there is a clear repeating pattern of up and down through June and July, and in the beginning of August the sheep came down towards the same height as the farm for some days. This is before the grazing period is over, and might be because of predators as literature suggests. Weather was also stated to have an implication both on the sheep's altitude location and their general activity level. Open source statistics on weather are available from Norsk Klimaservicesenter, and measurables such as temperature can be made into features for the machine learning model. Activity levels from the accelerometer will also have a high probability of being an important feature. Unusual activity levels might be a sign of the sheep being alert or scared. From table 5.1 the mean total activity lies around a value of 11.5, but the median has a value of about 4.5 and the lower percentile zero. This implies that there are probably a lot of either zeros or otherwise low values, which implicate that for a majority of the time the sheep will relax or stay calm. Higher values in activity therefore seems particularly significant.

From the nonzero total, it is possible to measure normal activity levels when the sheep are not relaxing but are in some type of activity, by removing the effect of the shifted distribution from the large number of zero-values. The comparable mean of 16.9 and median of 16.125 signify that the nonzero distribution is not skewed but quite symmetric. The mode of the histogram curves in figure 5.3 centers around 20 for both plots. Sheep spend around 90 % of their active time grazing, the rest for walking, running and nursing. The 90^{th} percentile values are 30 and 35, respectively for the total and nonzero total. This might indicate that an activity level exceeding 35 should be deemed deviant from normal, calm behavior. The sensitivity of the accelerometer is however not known, and the values are calculated for instants during the time it takes to fix a point and not continuously, so the effect of this should be examined further before any conclusions are made. The data from Fosen did not have any measurable activity level, but an option might be to calculate for example the velocity of the movement or distance travelled since the last point generation. There might also be more activity at night in the early season, which the model should distinguish on in the analysis.

If a sheep becomes sick while wearing a transmitter, it is possible the model will notice a difference in behavior and label it as predatory proximity. The effect will however most likely be opposite of the target behavior, by becoming lethargic, less active and slower than usual.

6.1 Cleaning the data sets

The process of cleaning the data sets were started this fall, and will continue into the next semester. There were several problems in the data set from Tingvoll that had to be addressed. There were just under 30.000 lines were the GPS tracker did not manage to generate a point. These faulty points had to be deleted from the set. There were also many lines without a generated point due to inactivity. This shows up as "Mortality" points in the data, but since all sheep started moving again, and none of the sheep from the data set were confirmed dead, this is assumed most likely due to the sheep sleeping or relaxing. The data from Fosen mostly had some missing values, and latitudes and longitudes generated to be zero.

When handling outliers and incorrectly generated points, a threshold value need to be set in order to distinguish between improbable distances traveled by the sheep and cases of the sheep running. With points with a difference in distance calculated by the Haversine formula that is too big to be plausible, another value should be imputed by the use of mean from the two closest feasible points. The threshold value is set to 15 km per hour, as it is assumed sheep will not have the speed nor the stamina to surpass the endurance of a fast human over the course of an hour. Imputation should be done with caution, because it reduces variance in the data set and could be removing important instances where a target situation happened. The rather high value is thus set to give some lenience, as 15 km is considered a long distance traveled hourly for sheep.

6.2 Further work and plan for master thesis

The next semester will be dedicated to preparing, building and evaluating a machine learning model based in the findings in this specialization project. It is still required to clean and prepare the data further before any computational analysis may begin. The extreme coordinate values and distances traveled using the Haversine formula must be checked, a fixed time frame for the analysis must be set, consistency in date and time generation ensured and the data set sizes compared. New features such as temperature, altitude and activity measures are to be implemented into all the data sets in a sensible and uniform format. The data will have to be transformed into a convenient format to easily run the analysis across individuals, years and area.

The bulk of the work will then go into building and tuning a machine learning program, by testing different models and methods, and feature selections. There is a possibility that there may be access to more data sets in the spring, with hopefully more instances of predatory killings. These will also have to be inspected, cleaned and prepared. During this fall there has been some contact with the producers of the electronic transmitters FindMy, where there might be a meeting in early spring to get insight on their technological solutions and access to the data they posses. After a model has been built, the results must be evaluated and interpreted in light of the underlying theory. An assessment should be made on whether the results are applicable, and if they succeeded or not in the overall goal of the project. A Gantt chart of the tentative project timeline is shown in figure 6.1 [56]. The project development will follow sprints with iterations of data preparation, model implementation and evaluation, as in accordance with CRISP-DM.

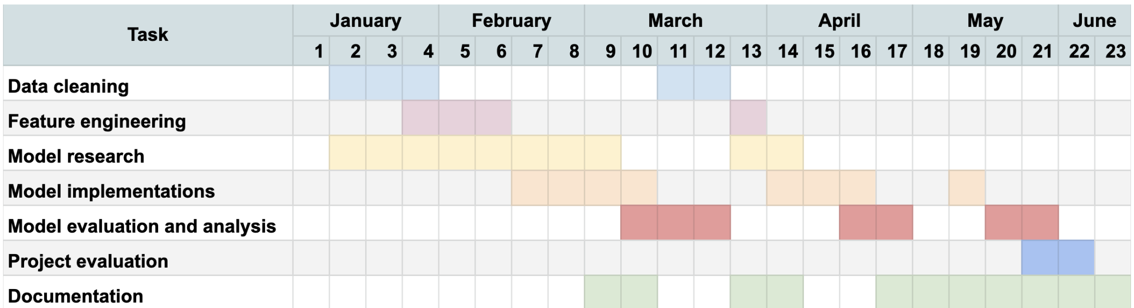


Figure 6.1: Tentative plan for the tasks and their timeline over the next semester in a Gantt chart.

6.3 Recommendations

Predators are more likely to be able to catch slow or injured individuals, or prey that fall behind the safety of the herd. It might be expected that the most vulnerable groups are young lambs, old sheep or sheep without an immediate family group, and therefore these will have more close experiences with predators. The lead ewes might also be in the front and in close proximity when a predator approaches the herd, but as lead sheep they will according to Tømmerberg act more confident and offensive instead of nervous and restless as the lower ranks. If there is an established hierarchical order in a herd, it will be valuable for further research that the low rank sheep in areas with high predatory concentration are tagged with transmitters. This should be done in collaboration with the oldest, young lambs, and lonely grazers, if not the whole herd can be tagged. More data of predatory instances is needed in order to conclude with certainty the behavior of the sheep. In addition, it will be advantageous if the same sheep over multiple consecutive years wear a transmitter in order to compare the divergence of their habitual routes once established.

7 Conclusions

Cleaning the data from outliers and incorrectly generated values should be done at the beginning of next semester, and a threshold value of 15 km traveled per hour calculated by the Haversine formula will catalyse imputation of new values for the coordinates. The project will then follow iterations of data preparation, model implementation and evaluation until satisfactory results are achieved.

Breed of sheep will have a large impact on their behavior and reactions to predators. The lighter and older breeds will have stronger instincts and react more prominently, than the heavier and newer breeds. Therefore will the analysis initially focus on the lighter breeds to more easily see if it is possible to classify predatory behavior in the sheep. Further, the differences of the effect from the model by splitting the data on early and late season, compared to not splitting, will also be inspected.

The computational analysis that will be done next semester will be based on unsupervised and semi-supervised machine learning models, with classification of behavioral pattern by use of methods like association rules, clustering and anomaly detection. Important features to include will be altitude, weather and activity level. The goal is to be able to detect when a predator might have been in close proximity to the herd. This may be used further by implementing the findings into the electronic transmitters worn by sheep on outfield pastures to notify farmers when their herd possibly is in danger.

References

[1] SSB. *12660: Husdyr på utmarksbeite (k) 1995-2020*. 2020. URL: https://www.ssb.no/statbank/table/12660/.

[2] Dyrebeskyttelsen Norge. *Tap av sau på utmarksbeite*. 2020. URL: https://www.dyrebeskyttelsen.no/tap-sau-pa-beite/.

[3] Miljødirektoratet. *Tap av husdyr og tamrein*. Jan. 2021. URL: https://www.miljodirektoratet.no/ansvarsomrader/arter-naturtyper/vilt/rovvilt/husdyr-tap/.

[4] Lovdata. *Forskrift om erstatning når husdyr blir drept eller skadet av rovvilt*. Oct. 2021. URL: https://lovdata.no/dokument/SF/forskrift/2014-05-30-677.

[5] Miljødirektoratet. *Rovvilt tok færre sauer i 2020*. Jan. 2021. URL: https://www.miljodirektoratet.no/aktuelt/nyheter/2021/januar-2021/rovvilt-tok-farre-sauer-i-2020/.

[6] Rovbase. *Erstatning for sau*. 2020. URL: https://www.rovbase.no/erstatning/sau.

[7] Rovdata. *Bestandstatus rovvilt*. Oct. 2021. URL: https://rovdata.no/#LiveContent[bestandsstatus].

[8] Data Science Process Alliance. *What is CRISP-DM*. URL: https://www.datascience-pm.com/crisp-dm-2/.

[9] G. Press. *Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says*. 2016. URL: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/.

[10] A. Tidemann and A. C. Elster. *Maskinlæring*. June 2019. URL: https://snl.no/maskinl%C3%A6ring.

[11] T. M. Mitchell. *Machine Learning*. McGraw Hill, Mar. 1997, p. 2.

[12] D. Sarkar, R. Bali, and T. Sharma. *Practical Machine Learning with Python*. Apress, 2018.

[13] O. Chapelle, B. Schölkopf, and A. Zien. *Semi-Supervised Learning*. MIT Press, 2006.

[14] H. Patel. *What is Feature Engineering - Importance, Tools and Techniques for Machine Learning*. Aug. 2021. URL: https://towardsdatascience.com/what-is-feature-engineering-importance-tools-and-techniques-for-machine-learning-2080b0269f10.

[15] J. K. Rød. *Geografiske koordinater*. Jan. 2021. URL: https://snl.no/geografiske_koordinater.

[16] L. D. Talley et al. *Descriptive Physical Oceanography (ed. 6). Academic Press.* 2011. DOI: https://doi.org/10.1016/C2009-0-24322-4.

[17] ISO. *Geographic information — Referencing by coordinates*. 2019. URL: https://www.standard.no/en/PDF/FileDownload/?redir=true&filetype=Pdf&preview=true&item=1022546&category=5.

[18] Ø. B. Dick, J. K. Rød, and L. Mæhlum. *Geodetisk datum*. Jan. 2021. URL: https://snl.no/geodetisk_datum.

[19] Ø. B. Dick and J. K. Rød. *WGS84*. Feb. 2021. URL: https://snl.no/WGS84.

[20] European Petroleum Survey Group. *About the EPSG Dataset*. URL: https://epsg.org/home.html.

[21] Spatial Reference.org. *EPSG Projection - WGS 84*. URL: https://spatialreference.org/ref/epsg/wgs-84/.

[22] J. K. Rød and L. Mæhlum. *EUREF89*. Mar. 2021. URL: https://snl.no/EUREF89.

[23] J. K. Rød. *Innføring i GIS og Statistikk*. 2nd ed. Fagbokforlaget, 2017, p. 113.

[24] N. R. Chopde and K. N. Mangesh. "*Landmark Based Shortest Path Detection by Using A* and Haversine Formula*". In: *International Journal of Innovative Research in Computer and Communication Engineering* 1 (2 Apr. 2013).

[25] V. Agafonkin. "*Fast geodesic approximations with Cheap Ruler*". In: (May 2016). URL: https://blog.mapbox.com/fast-geodesic-approximations-with-cheap-ruler-106f229ad016.

[26] T. Vincenty. "*Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations*". In: 23 (Apr. 1975).

[27] GPS.gov. *GPS Accuracy*. URL: https://www.gps.gov/systems/gps/performance/accuracy/.

[28] Encyclopaedia Britannica. *Sheep*. May 2020. URL: https://www.britannica.com/animal/domesticated-sheep.

[29] J. R. E. Johanssen and K. M Sørheim. *Atferd og velferd hos sau*. May 2021. URL: https://www.agropub.no/fagartikler/atferd-og-velferd-hos-sau.

[30] I. Hansen, H. S. Hansen, and F. Christiansen. "*Kartlegging av antipredatoratferd hos ulike saueraser*". In: (1998).

[31] T. H. Garmo et al. *Saueboka*. 2nd ed. Landbruksforlaget, 2006.

[32] Norsk Sau og Geit. *Sauerasene i Norge*. URL: https://www.nsg.no/sau/saueraser/.

[33] A. K. Doughty et al. "*The influence of lameness and individuality on movement patterns in sheep*". In: *Elsevier* (Mar. 2018). URL: https://www.sciencedirect.com/science/article/pii/S0376635717302449.

[34] W. O. Tømmerberg. "*Atferd hos frittlevende domestiserte sauer på fjellbeite*". In: (1985).

[35] L. A. Syme. "*Social disruption and forced movement orders in sheep*". In: *Elsevier* (1981). URL: https://www.sciencedirect.com/science/article/pii/S0003347281801765.

[36] V. R. Squires and G. T. Daws. "*Leadership and dominance relationships in Merino and Border Leicester sheep*". In: *Elsevier* (1975). URL: https://www.sciencedirect.com/science/article/pii/030437627590019X.

[37] W. H. Burt. "*Territoriality and home range concepts as applied to mammals*". In: *Journal of Mammalogy* (Aug. 1943). DOI: https://doi.org/10.2307/1374834.

[38] P. A. Jewell. "*The concept of home range in mammals*". In: *Symposium of the Zoological Society of London* 18 (1966), pp. 85–109.

[39] D. Scott and B. L. Sutherland. "*Grazing behaviour of merinos on an undeveloped semi-arid tussock grassland block*". In: *New Zealand Journal of Experimental Agriculture* 9 (Jan. 1981). DOI: https://doi.org/10.1080/03015521.1981.10427794.

[40] G. P. Hughes and D. Reid. "*Studies on the behaviour of cattle and sheep in relation to the utilization of grass*". In: *Journal of Agricultural Science* 41 (4 Oct. 1951), pp. 350–366. DOI: https://doi.org/10.1017/S0021859600049534.

[41] D. E. Tribe. "*Some seasonal observations on the grazing habits of sheep*". In: *Empire Journal of Experimental Agriculture* 27 (1949), pp. 105–115.

[42] A. Murie and W. B. Davis. "*The Wolves of Mount McKinley. Fauna of the National Parks of the United States*". In: *Journal of Mammalogy* 26 (1 Feb. 1945), pp. 100–101. DOI: https://doi.org/10.2307/1375039.

[43] A. Stien et al. "*Kongeørn som tapsårsak for sau og lam. Tapsstudier i Rødsjø beiteområde 2014-2015.*" In: (Aug. 2016). URL: https://brage.nina.no/nina-xmlui/handle/11250/2402971.

[44] E. Østbye. *Jerv*. Nov. 2021. URL: https://snl.no/jerv.

[45] T. E. Katzner et al. *Golden Eagle (Aquila chrysaetos)*. Sept. 2020. DOI: https://doi.org/10.2173/bow.goleag.02.

[46] E. K. Barth. *Kongeørn*. May 2020. URL: https://snl.no/konge%C3%B8rn.

[47] E. K. Rueness. *Gaupe*. Aug. 2021. URL: https://snl.no/gaupe.

[48] M. Filla et al. "*Habitat selection by Eurasian lynx (Lynx lynx) is primarily driven by avoidance of human activity during day and prey availability during night*". In: *Ecology and Evolution* 7.16 (2017), pp. 6367–6381. DOI: https://doi.org/10.1002/ece3.3204.

[49] E. K. Rueness and E. Østbye. *Bjørn*. Nov. 2021. URL: https://snl.no/bj%C3%B8rn.

[50] S. P. French and M. G. French. "*Predatory Behavior of Grizzly Bears Feeding on Elk Calves in Yellowstone National Park, 1986-88*". In: *Bears: Their Biology and Management* 8 (1990), pp. 335–341. DOI: https://doi.org/10.2307/3872937.

[51] E. Østbye. *Sauer løp til de ble «sprengt»: – Skjerp dere folkens*. July 2020. URL: https://www.nrk.no/sorlandet/fant-flere-dode-sauer-_-ber-hundeeiere-skjerpe-seg-1.15095137.

[52] OpenStreetMap contributors. *Open Street Map*. 2021. URL: %5Curl%7B%20https://www.openstreetmap.org%20%7D.

[53] J. F. Aarnes. *Pytagoras' setning*. Oct. 2020. URL: https://snl.no/Pytagoras'_setning.

[54] Norsk Klimaservicesenter. *Observasjoner og værstatistikk*. URL: https://seklima.met.no/observations/.

[55] Kartverket. *Brukarrettleiing for høgdeprofil-API*. Sept. 2021. URL: https://www.kartverket.no/api-og-data/friluftsliv/hoydeprofil.

[56] W. Clark. "*The Gantt Chart: A Working Tool of Management*". In: (1922).

Appendix

A - Code

All code and Python files used in this project are included in a Github repository linked below. Explanations are given in the readme-file. The files are:

- ActivityXY.py
- Altitude.py
- MapPlotFosen.py
- MapPlotTingvoll.py
- MapPlotTogether.py
- SizeCheck.py
- Statistics.py
- README.md

Github Repository Link

`https://github.com/ninasalvesen/Prosjektoppgave`

B - Feedback from Facebook

The responses given to the post on the nation wide Facebook group for sheep farmers are given below. It was answered at will to general questions about sheep behavior.

General feedback on behavior:

- Their movement pattern are usually motivated by grazing opportunities and weather
- The sheep will follow the same fixed routes every year
- Their behavior will also be largely affected by their personality, they will not all act completely the same every time
- They are herd animals with a hierarchy, and an ewe with very low rank may be cast out but will still follow the herd's general movement
- If the herd meets a challenge, the lead ewe will be in the front, the lowest rank at the back and the bulk in the middle. This will happen if the challenge is a predator nearby, unknown humans hiking, dogs, or something inanimate that scares them
- If the above lineup does not scare the challenge away, the lead ewe might try to offensively approach while the rest run

Feedback on behavior with predators nearby:

- Wolverines and eagles cause the most trouble, but the eagles will mostly only trouble the herd when the lambs are small
- With eagles nearby they will often move down below the tree line
- Generally there is a difference in behavior depending on breed of sheep and predator
- They will know when a predator is nearby, even if they not necessarily see it. They become stressed, nervous and restless
- The herd sometimes will spread when in proximity with a predator
- The herd will come home to the farm earlier than usual if it meets predators
- If they meet predators early in the season when they still are quite close to the farm they will come home again
- The ewe will be braver if it meets predators when she has young lambs
- If a sheep dies the herd will often change pastures

Feedback on behavior when sick:

- They will often seek roads, houses, waters or people in general if this is nearby
- Some will hide if they are really sick
- They will move noticeably less in a day

C - Interviews of sheep farmers

Where are your farm and outfield pasture located?
1: Nordre Vang
2: Dovre
3: Dokkfjell
4: Nordre Land
Which breed of sheep do you have?
1: NKS, and some Spæl
2, 4: NKS
3: NKS and Norwegian Fur
How would you describe their normal behavior?
1: Walk very separated and afar from each other, which makes them more vulnerable. They often walk towards where there are roads and people. They will return home at the end of the season themselves.
2: Individual differences if they walk alone or in a small flock. They are more alert the first few days on outfield pastures. Quite bound by their routines, but will deviate from this if it is very hot and dry. Then they will roam more and move further.
3: Bound by routine, and very little herd instincts. Friendly towards strangers.
4: Calm and bounded by routines.
Which predators have your sheep encountered?
1: Lynx and fox
2: Wolverine and eagle
3: Eagle and lynx
4: Lynx
How did the sheep react in situations with predators nearby?
1, 4: Have not observed this.
2: They are very nervous. Solitary sheep more vulnerable.
3: They will not always understand or notice when a predator is nearby. They will be noticeably stressed.
Do you notice any difference from what type of predator that is nearby?
1-4: No
How will the sheep react to human stranger and/or dogs?
1, 3, 4: The same as with predators.
2: My herd is used to dogs, and will not react much.
How often do you inspect the sheep on the outfield pastures?
1: Three times a week
2: Two-three times a week
3: Daily
4: Two times a week