Analysis of cattle movement patterns in the Vierkaser-Alm

Using machine learning to determine the influence of environmental conditions on movement

Master Thesis

for the attainment of the Master`s degree “Master of Science”,
abbreviated “MSc”

submitted within the University Master Program for Further Education
“Geographical Information Science & Systems – (UNIGIS MSc)”
at the Department of Geoinformatics - Z_GIS,
Faculty of Digital and Analytical Sciences,
University of Salzburg

submitted by

Cristo Leonard van der Werff

Supervisor:
Assoc.-Prof. Dr. Gudrun Wallentin

Zoetermeer, 31 March 2026

Acknowledgements

Gratitude is given to Prof. Dr. Gudrun Wallentin for acting as supervisor and giving help and guidance through the making of this thesis.

Gratitude is given to Sebastian Feldbacher for granting permission to conduct research on the farm and the cattle in Vierkaser.

Abstract

Alpine livestock grazing is a traditional way of farming found in Austria that is increasingly under pressure. Therefore, it is important to apply proper land management practices, especially with limited time and resources. Cattle tracking is a modern way to gather and analyze cow movement data to improve the choice of land management practices. This study analyzed cow movement locations to find the most suitable modelling method, quantify the impact of environmental factors and evaluate the effects of shrub clearing in the pasture.

In the Vierkaser pasture, located in the Salzburg province of Austria, cows were attached a GPS to track their locations for the past five years. Each year, a certain area was cleared from shrub growth. These locations and corresponding environmental conditions were analyzed using generalized linear models. The models were trained on spatial conditions for multiple selected temporal weather conditions.

The model that came out best was a negative binominal regression model, consisting of linear relationships with slope, tree density and grassland cover for nighttime models. Daytime additionally has distance to water source as significant environmental predictor. Elevation, excess greenness and other land cover classes resulted as insignificant for the model predictions.

During both day and night, cows are strongly affected by the amount of grassland cover. In natural environment, cows prefer to be close to pasture that has dense tall trees nearby to act as shade and shelter. Being on a gentle slope is important while resting during nighttime but barely during daytime. Being close to a water source is important during the day, especially during high temperatures, but not so much when it rains. Clearing shrubs in a specific area of a pasture makes cows more attracted to that area for the first year. The long term effects of shrub clearing vary and are unclear.

This study can be used to understand cow movement during the alpine summer and used as a guideline for deciding environmental land management practices in the future.

Contents

List of abbreviations and translations

Below is the list of abbreviations used in this paper:

AIC	Akaike Information Criterion
CHM	Canopy Height Model
DSM	Digital Surface Model
DTM	Digital Terrain Model
GLM	Generalized Linear Model
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
NB	Negative Binominal
RSF	Resource Selection Function
RSM	Resource selection model
RSPF	Resource Selection Probability Function
SAA	Sun Altitude Angle
SQL	Structured Query Language
r_s	Spearman's rank correlation coefficient

1 Introduction

1.1 Relevance

1.1.1 Cattle farming in Austria

Livestock farms make up a large part of the Austrian agriculture industry, as nearly 60% of Austria's agricultural land is grassland (BMLUK, 2024). Alpine farming has been practiced since the Bronze Age. It is often described as livestock farms that are mainly settled in mountain valleys for the winter, but temporary relocated to alpine pastures during summer. This specific period is also known as the 'Almsommer'. This temporary relocation is done to use the arable land in the valleys for crop cultivation, while the not so arable alpine grassland is sufficient for grazing (Gilck, Poschlod, 2019).

These farms provide cultural and ecological value to the landscape of the Austrian Alps. Despite economically relying on subsidies, the scenery they create are of great importance for the tourism industry. For the mountainous Alps, alpine pastures are as equally important as its remote wilderness for biodiversity (Wrbka et al., 2004).

The future of alpine farms is under pressure due to a very wide range of factors. The main contributing factors to this pressure are expected increases in costs for energy and animal feed concentrates, along with a more liberal agriculture market (Buchgraber et al., 2011). Too much pressure can result in land abandonment. While many types of farming are affected by this, it is estimated that farms in mountain areas are at three times higher risk than non-mountain areas (Dax et al., 2021). It is estimated that around 12% of alpine farms have been abandoned in Austria between 1980 and 2000. Many farms were forced to change from full-time to part-time farms, with only 35% of alpine farms remaining full-time during that period (Steinwidder et al., 2011).

Farmers have many ways to manage the alpine pasture for their cows, such as mowing, cutting or planting trees, using fertilizer, etc. The most important qualifications for pastures are considered: A high diversity of species, good grass composition, few weeds, dense grass cover (Wezel et al., 2021). As most farmers work part-time, it is important for them to make worthy and time efficient decisions in their pasture management.

1.1.2 Cattle movement patterns

To get a better understanding of alpine farm management, it is important to understand cow grazing and movement patterns of cattle.

Cattle movement is a complex process determined by a large variety of factors. There are multiple ways to categorize this large collection of factors. These can be categorized in three patterns: Environmental, biological and behavioral factors (Creamer and Horback, 2024). Other papers categorize them by internal and external factors, in which internal factors consist mostly of biological and some behavioral factors while external factors consist on environmental and some behavioral factors (Rivero et al., 2021).

Biological factors consist of internal characteristics of the cattle, such as breed, age, weight pregnancy status. The effect of these biological factors on movement have been known to be complex and unclear (Walburger et al., 2009, Wyffels et al., 2020b).

Behavioral factors consist of internal phenotype, such as memory of the pasture, cognition, perception, emotional state and external phenotype of the behavior and interaction of other cows (Creamer and Horback, 2024). Within the same cow herd, differences between both individuals and subgroups of cows can be found. A dominant cow that plays a leading role will be less affected by non-dominant cows and vice versa (Rivero et al., 2021).

Environmental factors consist of external spatial and temporal factors. Spatial factors include land cover, topography, climate, supplemental feed, vegetation whereas temporal factors include weather, season, time of the day (Rivero et al., 2021).

Another, more general, factor is the stocking rate of a pasture. This is the ratio between the amount of animals and the pasture size over a period of time. This can affect to what extent cattle will utilize relatively less preferred grazing area (Probo et al., 2014).

A common and effective way to analyze cow movement is with the usage of GNSS tracking. This is often practiced in ecological research for both animal movement and behavior. Beside location, trackers can also track other environmental factors such as temperature, jaw movement, angle, etc. (Tomkiewicz et al., 2010).

1.1.3 Goal and objectives

The goal of this thesis is to understand why cows choose to move to specific locations during the summer in alpine pasture. In order to reach this goal, a handful of objectives are established for this research.

The study will train a model on cow locations and environment data. There is a wide variety of models that can be used in the animal movement research field. The first objective is to determine the best model and inputs for cow movement in this study.

Once the best model has been found, analysis will be done to determine the influence of each factor. This will be done for a range of conditions. The second objective is to quantify the influence of each environmental factor on cow movement for specific conditions.

The cow data in this research is available for multiple years. Various land management engineering has been done over the past years, resulting in a slightly different environment for the area each year. The third objective is to analyse how land management in the past has affected cow movement over the years.

1.2 Study area

1.2.1 Location

The study area is located at the Untersberg, Salzburg province, Austria, next to the border with Germany (Figure 1). The area lies at the foot of the Northern Limestone Alps, specifically in the Berchtesgaden Alps. The study area is 444,570 m^2 (44.46 ha). The altitude of the lowest and highest point sits between 1553 and 1760 m.

Figure 1: Overview of the study area location, using the BEV Orthographic photo as background.

The area consists mostly of limestones shaped as plateaus and ridges. It is dominantly covered by grasslands, pine forests, rocks and scree (Kraller et al., 2012). The weather during the Almsommer is relatively temperate and humid. It has an average temperature of 21.1 °C at day and 11.2 °C at night. The precipitation is around 168 mm per month (Mithell, Jones, 2005).

The Alm itself is known as the 'Vierkaseralm am Untersberg'. It originally consist of two active pastures: the Vierkaser meadow and Hirschanger (Figure 3). The Vierkaser meadow contains a small unused building and remnants of a few old buildings. The rest of the study area is natural physical and biological environment, although sometimes tempered by humans. The pastures are partially enclosed by both natural environment and placed fencing.

Before 2021, the farms cattle stayed in other places throughout the years. The Almsommer of 2021 is the first time cows have went used this pasture after 60 years. Every year has a different date for when the cattle are taken up the mountain (Almauftrieb) and taken back to the valleys (Almabtrieb) and different total count of cows (Table 1).

Year	Cow count	Almauftrieb	Almabtrieb
2021	5	June 26	September 10
2022	12	June 4	September 14
2023	10	June 10	September 17
2024	11	June 8	September 11
2025	7	June 7	September 14

Table 1: Overview of the Almsommers during the study period.

1.2.2 Land management

Below is a summary of the land management that has been done in the past years. This land management mainly consists of clearing of pine scrub, making room for grass growth then later for cows to feed on. This will later be referred to as 'shrub clearing'. It is conducted by the 'Almparternschaf', a group that consists of the farmer himself, members of the 'Alpenverein Salzburg' and other volunteers.

Figure 2: Photo of shrub clearings in 2022, vierkaseralm_am_untersberg Instagram.

Every year, other parts of the study area have been cleared. This was done during various moments, usually in August during the Almsommer (Table 2). In 2020, during early October and late November, went to location to clear shrubs an area connecting Meadow with Hirschanger. Additionally, also an area next to Hirschanger was cleared up for better access. In August 2021, areas were cleared leading to an expansion in size for the Meadow and Hirschanger pastures. In August 2022, similarly, more of the area was cleared next to the existing pastures. In August 2023, the largest shrub clearing was conducted. For a total of 500 work hours, very large parts of the area were cleared, giving better access to another fairly large unnamed pasture. In August 2024, another connecting area between the Meadow and Hirschanger was cleared for cows to have better access. Cows now have three decently accessible ways to access Hirschanger from the meadow. Beside that area, some for parts adjacent to Hirschanger were cleared as well. Finally, a dam was constructed in Hirschanger to improve rain containment to serve as drinking water from cows. In June 2025, more areas mostly in the south, adjacent to Hirschanger, were cleared. Waterpipes were placed.

Date	**Area cleared**	**Labor**	**Additional work**
2020-10-04	6,496 m^2 (0.65 ha)	9 people, 150 hours	Road construction
2021-08-22	3,659 m^2 (0.37 ha)	12 people, 300 hours	
2022-08-21	7,395 m^2 (0.74 ha)	13 people, 350 hours	
2023-08-27	19,287 m^2 (1.93 ha)	23 people, 500 hours	
2024-08-18	5,206 m^2 (0.52 ha)	unknown	Dam construction
2025-06-22	4,039 m^2 (0.40 ha)	unknown	Waterpipe placement

Table 2: Overview of the land management clearings in the recent years.

The areas that have underwent shrub clearing have been digitized by Dr. Prof. Wallentin (Figure 3). A few other available patches of predominantly grass that have not underwent shrub clearing were also digitized.

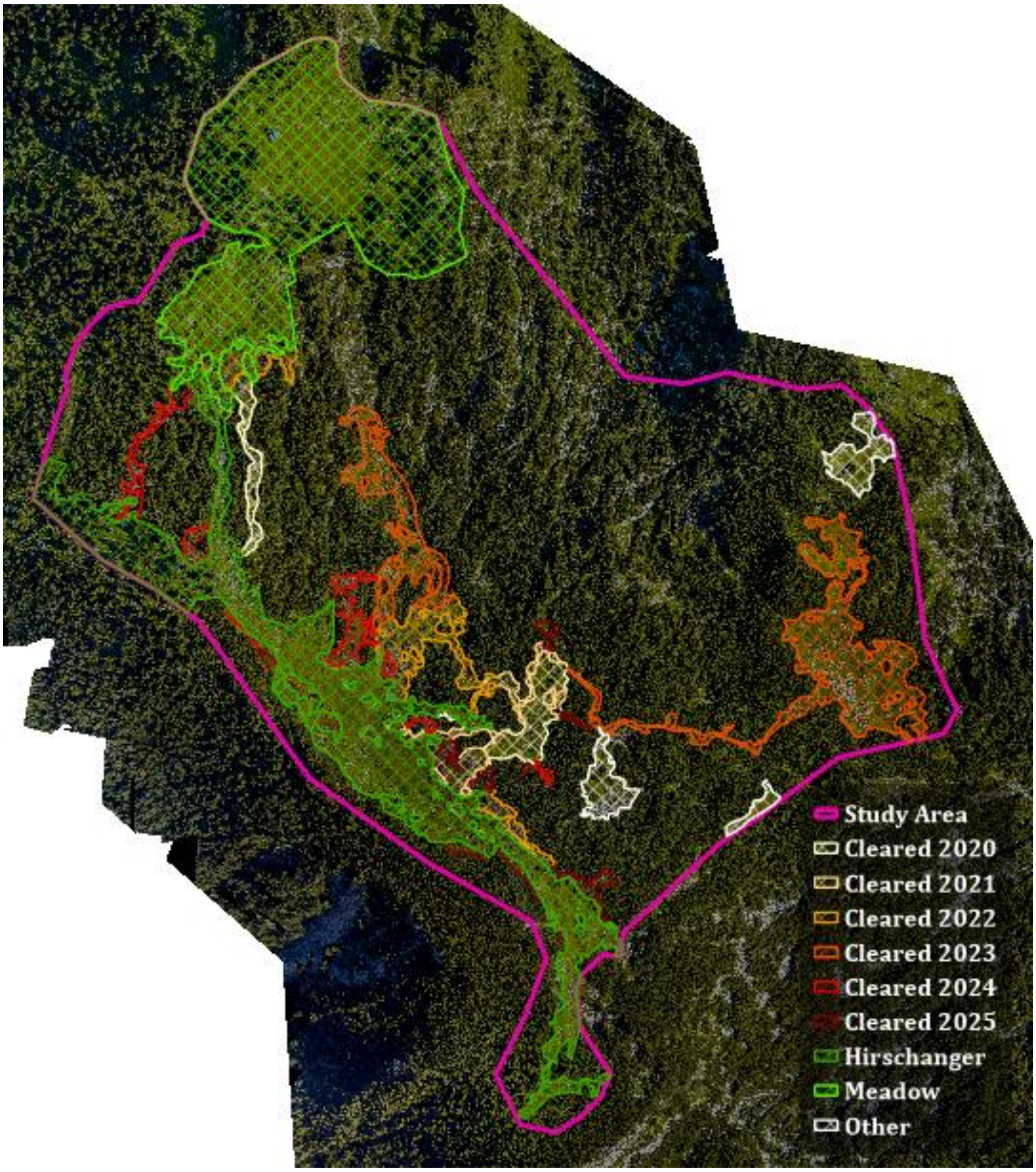

Figure 3: Overview of the pastures inside the study area.

1.3 Environmental factors

The selected environmental factors for this study are based on existing literature and data availability. A study reviewing GPS-tracked cattle grazing shows that environmental factors in cow movement patterns can be categorized in five categories: Water location, shade and shelter in relation to climate and weather, supplemental feed type and location, vegetation characteristics, topography (Rivero et al., 2021). Supplemental feed it not used in the farm for this study and is thus not included as an environmental factor.

An description of each selection is shown below. The categories for factors are slightly altered to fit the structure of this study. Environmental factors include water location, shade and shelter, vegetation, topography and weather factors.

1.3.1 Water location

Water source locations have shown to affect the chosen grazing locations of cattle (Probo et al., 2014). In this study area, there is one water trough location present and a few natural ponds where cows can drink from (Figure 4). Not every pond in the area is practically used by cows. Some are unused due to not being permanently available or significantly dirty.

Figure 4: Picture of cows drinking from one of the ponds found in the study area, photo by Berg Natur Gruppe, Salzburg Alpenverein, 2025.

1.3.2 Shade and shelter

In farms where there is a building available for cows to sleep in, the location of the building would be included for a model like this (Pittarello et al., 2021). In the farm in this study area, the cows do not sleep inside a building. Instead, cows use the natural environment as shade and shelter. Here, the location, density and height of trees can be used to compute a factor that will work as 'shade and shelter' (Rivero et al., 2021).

1.3.3 Vegetation

Vegetation preference plays an important role for a cows movement. Vegetation factors consist of the vegetation type, quality and mass in a pasture. In alpine environments, fertile pasture are most preferred whereas dwarf shrub least. It is common for studies to distinguish vegetation between forest, grassland and shrubland (Raniolo, 2022; Raniolo, 2023). These classifications are also applicable to this study area.

Cows generally prefer to graze in dense patches of forage. The NDVI is often used as indicator for vegetation mass and quality in a pasture, where cattle show a strong preference for vegetation with a high greenness (Rivero et al., 2021).

1.3.4 Topography

Alpine regions are mountainous and often have a strongly varying terrain. Altitude is shown to have a positive effect for herbage quality. While lower altitudes generally enable a wider variety of plant species to grow, these also include plants that are not preferred by cows. Cows can both have a preferences for pastures either in higher altitude (van Dorland et al., 2007) or lower altitude (Falú et al., 2014), depending on the altitudes effect on the local vegetation profile.

While precise vegetation profiles would be ideal, using altitude as environmental factor is a more practical approach in this study.

Slope is known to affect grazing patterns for cow movement. Cows generally prefer to rest on more gentle slopes, leading to relatively higher defecation rates, which results in better nutrition for vegetation on these gentle slopes (Pittarello et al., 2021).

Terrain ruggedness, while occasionally included in some studies, is shown to often not have any relationship to cattle movement patterns (Steyaert et al., 2011; Raynor et al., 2021).

1.3.5 Weather

In cow movement modelling, weather factors often include temperature, rainfall and sometimes wind. Temperature is considered the strongest weather related factor. It can indicate how far a cow will graze from their usual resting spot and water location. High amount of rainfall can also relate to the movement activity of cattle (Rivero et al., 2021). Studies on wind direction and wind speed often have mixed results when it comes to finding a correlation with movement patterns (Halász et al., 2016; Eikelboom et al., 2020).

1.4 Machine learning model

There is a wide range of choices for machine learning models that can be used for animal movement. This chapter gives a brief overview of the model used for this research.

1.4.1 Resource selection models

A common way to correlate animal movement with environment is the use of resource selection modelling. This is often either classification or regression.

The classification involves a model classifying either the behavior or presence of a specific cow. Movement classification involves generating a bunch of pseudo-absent cow locations, where cows could have went. The Random Forest model often comes out as the best choice of model for predicting animal presence classification (Carvalho et al., 2018; O'Malley et al., 2024; Shoemaker et al., 2018).

The regression involves a model predicting the spatial use intensity by cows of a specific area. Commonly, generalized linear models (GLM) are used to . While they are not by definition the most accurate predictors, they can give concrete well interpretable results that can be used to create resource selection functions (RSF), where a single value is be calculated from spatial factors to indicate spatial use probability.

For the goal and objectives this study, it is important to provide a well interpretable result for understanding cow movement and land management. Therefore, this study use generalized linear models.

1.4.2 Generalized linear models

Generalized linear models in animal movement can be fit to predict spatial use intensity. Spatial use intensity is based on how much time, or how often a cow is in a certain area. A generated spatial grid of equal sized shapes is often use to define these areas. This category of GLM can also be defined as a ‘count of occurrences’ use. Traditional GLMs for this use includes models such as linear regression, Poisson regression, binominal regression.

These GLM differ based on their handling of dispersions in the dataset. Linear regression does not handle dispersion, Poisson regression assumes the variance is equal to the mean value, (negative) binominal regression uses an extra parameter to handle dispersion. The most appropriate model should be chosen based on the datasets that are put into the model. (Acciaro et al., 2024)

2 Method

2.1 Animal GPS data

This chapter describes the collection, cleaning and processing of the cow data used in this study.

2.1.1 Data collection

The cow GPS records play a major role in the model. This dataset was collected by UNIGIS, Salzburg University over the past years between 2021 and 2025. This was done through attaching GPS loggers to the collars of a varying number of cows. These were attached in the valleys, before the march of the cows to the alpine study area. After the return of the cows they were removed and the raw data was saved to computer storage.

Multiple types of GPS trackers were used for data collection. An initial goal was to track cow locations every 5 minutes for the entire Almsommer. The available commercial trackers, named 'Q2', were limited in either battery life, irregular time interval of 1-2 hours and occasional device failure. Therefore a custom GPS tracker was created, named 'Bene'. Both of these GPS trackers were used to track the cattle during the Almsommers (Bonath, 2023). Due do this, each cow trajectory had a tracker with varying interval time duration and irregular lifespan. This resulted in a rather skewed dataset with varying total records often for only a part of the entire Almsommer.

Not all cows resided in within the study area. Some of the cows, including Gamsei, Sternei, Karli and Nanni resided on the pasture of a different farmer, outside of the study area. This resulted in some of the trajectories unused for this study (Table 3).

2.1.2 Data cleaning

The raw data of cow GPS records had to be put together into a single database. This was done by writing a python script that generates two SQL query file for each raw data file (Appendix Script F). The first query is used to import the raw .csv files into the Postgres database named 'raw'. The .gpkg and .gpx files had to be manually imported in the Postgres database using QGIS. The second query is used to converting and translating the 'raw' database into a 'processed' database. These two queries were done separately because of the varying methods required to properly import the data into a Postgres database. The individual query content is based on a given SQL template by Dr. Prof. Gudrun.

Then, using QGIS, the dataset was then clipped to the Vierkaser-Alm boundary, limiting it to the study area. This removes irrelevant cow locations, often being in the valleys or road towards the alpine pasture.

Finally, records that lacked a datetime value were removed from the dataset, as these had illogical location and it was unclear if these records were correctly registered by the device.

2.1.3 Data weighting

The goal for this dataset is to turn it into relative spatial use maps. The dataset has varying GPS trackers, counts and timestep differences for each trajectory. Therefore, a weighting is required to be applied. First, a weighting is applied so that each trajectory is weighted equally, despite the total record count or timestep between records. Then, each trajectory is weighted by the amount of tracked cows, so that each year is treated equally. By combining these weighting, this results in a dataset where each individual record is specifically weighted so that calculating an average spatial use intensity map is not over-represented by a specific tracker or year (Table 3).

2.1.4 Temporal conditions

In order to filter the spatial use intensity on time and weather specific cow locations, the temporal datasets need to be retrieved for each GPS record.

The starting time of the day and night differ throughout the Almsommer. Instead of directly using the datetime field of a GPS record, instead, the sun altitude angle (SAA) is calculated. This value describes the height of the sun, with positive values occurring during day and negative values during night. A python script is created to calculate this value for each datetime value in the cow dataset, using the ‘pandas’ and ‘SunPositionCalculator’ libraries (Appendix Script B).

The weather data used in this study comes from OpenWeatherMap. A historical bulk weather dataset of the centroid of the study area was used. This dataset consists of a table that contains a broad range of weather variables for each exact clock-hour. From this dataset, only the temperature (°C) and rainfall in the past hour (mm) will be used for the model. These are joined to each cow location by using the nearest weather record. The data preparation and joining is done using the ‘pandas’ library in python (Appendix Script E).

Year	Tracker	Cowname	Format	Records	Weighting	Usability
2021	bene	Pia	gpkg	10096	0.000017	used
2021	q2	Geri	csv	253	0.000659	used
2021	q2	Gretel	csv	3563	0.000047	used
2021	q2	Pia	csv	4200	0.000040	used
2021	q2	Haensel	gpx	7464	0.000022	used
2021	q2	Susi	csv	1092	0.000153	used
2022	bene	Unknown-1	csv	6	0.015152	used
2022	bene	Unknown-2	csv	2635	0.000035	used
2022	bene	Unknown-3	csv	2668	0.000034	used
2022	bene	Unknown-4	csv	2703	0.000034	used
2022	bene	Unknown-5	csv	4363	0.000021	used
2022	bene	Unknown-6	csv	7	0.012987	used
2022	bene	Unknown-7	csv	1467	0.000062	used
2022	bene	Unknown-8	csv	1036	0.000088	used
2022	q2	Unknown-0	gpx	2596	0.000035	used
2022	q2	Julia	gpx	816	0.000111	used
2022	q2	Pia	gpx	136	0.000668	used
2023	q2	Gretel	csv	1183	0.000141	used
2023	q2	Unknown	csv	1068	0.000156	used
2023	q2	Julia	csv	902	0.000185	used
2023	q2	Haensel	csv	301	0.000554	used
2023	q2	Susi	csv	1471	0.000113	used
2023	q2	Pia	csv	1406	0.000119	used
2024	q2	Gamsei	gpx	0	-	unused
2024	q2	Sternei	gpx	0	-	unused
2024	q2	Julia	gpx	1054	0.000316	used
2024	q2	Haensel	gpx	73	0.004566	used
2024	q2	Karli	gpx	0	-	unused
2024	q2	Pia	gpx	1418	0.000235	used
2025	q2	Haensel	gpx	828	0.000403	used
2025	q2	Nanni	gpx	0	-	unused
2025	q2	Sternei	gpx	0	-	unused
2025	q2	Gamsei	gpx	0	-	unused
2025	q2	Susi	gpx	407	0.000819	used
2025	q2	Pia	gpx	94	0.003546	used

Table 3: Overview of the processed cow GPS records dataset tracks within the study area.

2.1.5 Spatial use intensity

To calculate the spatial use intensity, a spatial grid first has to be defined. An empty grid is created using equal size 50 x 50 m square shaped grid cells. This size was chosen based on the year with the lowest average GPS records, where cells frequently have multiple GPS records inside them without becoming too dispersed (Acciaro et al., 2024). Then, the area of grid cells that overlap with the boundary was calculated Grid cells that cover the study area for 30% or lower, were removed.

Next, the relative spatial use intensity is joined by calculating the sum of the weighted cow points within each grid cell, using the 'Join by location (summary)' tool. This is done using various combinations of filters for the cow datasets. The filters are chosen based on literature and the dataset size. Choosing a too specific time and weather condition results in a low total GPS record count and an unreliable spatial use intensity grid to model a resource selection function on. Therefore, thresholds are chosen within the first and third quartiles (Raniolo et al., 2023), while still holding significant ecological meaning.

First, the datasets are filtered based on time (Figure 4). While some cow behavior can be distinguished for mornings or evening periods, the major movement patterns are best analyzed using day and night periods (Rivero et al., 2021). Creating filters for dawn and dusk also does not uphold the earlier stated quartile condition.

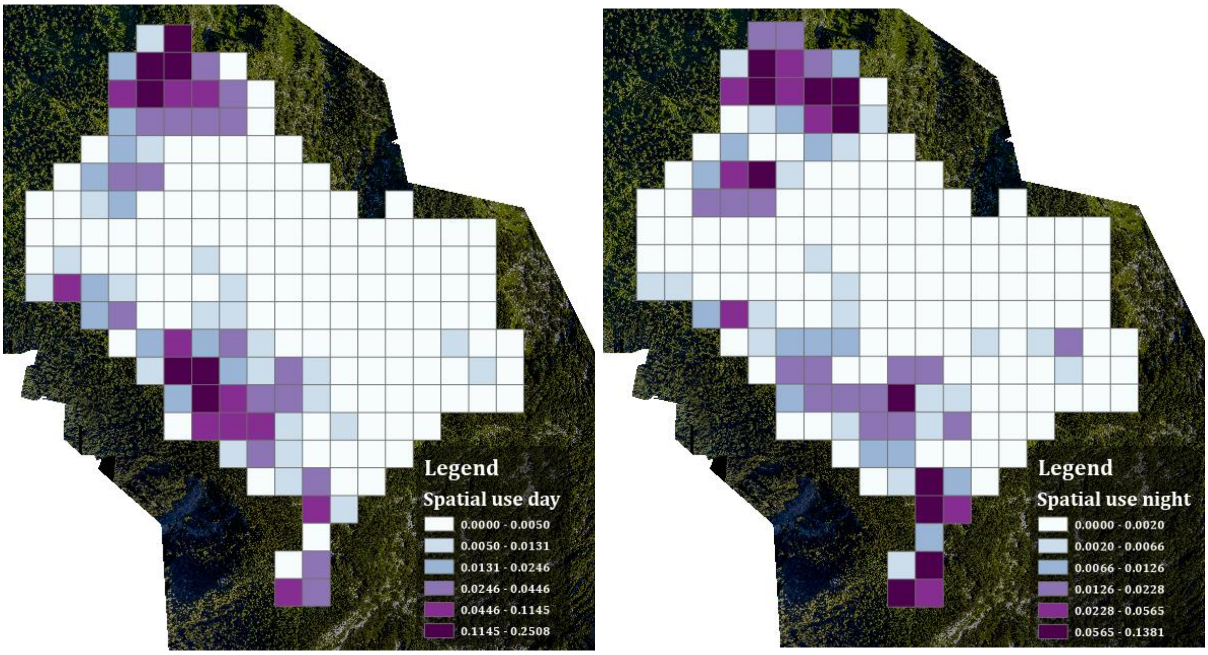

Figure 4: Relative spatial use intensities for day and night.

Datasets are analyzed for specific weather conditions. Cows are relatively resistant to cold temperature, but can show different behavior due to heat stress at temperatures starting at 20 °C (Zhou et al., 2022). At night, while temperatures stayed relatively within the comfort zone of 0-20 °C, a threshold of 12 °C is chosen as, some cow species can show slightly altered walking choices (Hut et al., 2022). Rainy days can cause cows to seek shelter among trees (Rivero et al., 2021). Therefore, for both day and night, the weather condition thresholds that are analyzed are for: Temperatures above and below 20 °C during the day, temperatures above and below 12 °C during the night, rainy and clear days and rainy and clear nights, for a total of 10 conditions (Table 4).

Condition	Time of day	Temperature	Rain
Day	SAA > 0		
Warm day	SAA > 0	TEMP > 20	
Cold day	SAA > 0	TEMP < 20	
Rainy day	SAA > 0		RAIN > 0
Clear day	SAA > 0		RAIN = 0
Night	SAA < 0		
Warm night	SAA < 0	TEMP > 15	
Cold night	SAA < 0	TEMP < 15	
Rainy night	SAA < 0		RAIN > 0
Clear night	SAA < 0		RAIN = 0

Table 4: Overview of each condition for which a spatial use intensity map is created.

2.2 Environmental predictors

The chapters below describe which environmental data sources are used and how these are processed into dataset suitable for direct use in a model. For each predictor, the mean value is calculated for each grid cell using the 'Zonal Statistics' tool in QGIS.

2.2.1 Water source distance

The locations of the water sources for the cows were manually digitized in QGIS using the field knowledge from Prof. Dr. Gudrun Wallentin of the farm and the aerial photo (Figure 5). Then, this point layer was converted into a raster using the Rasterize tool. Finally, the Proximity tool is used to create a new raster with band values containing the distance to the nearest water source.

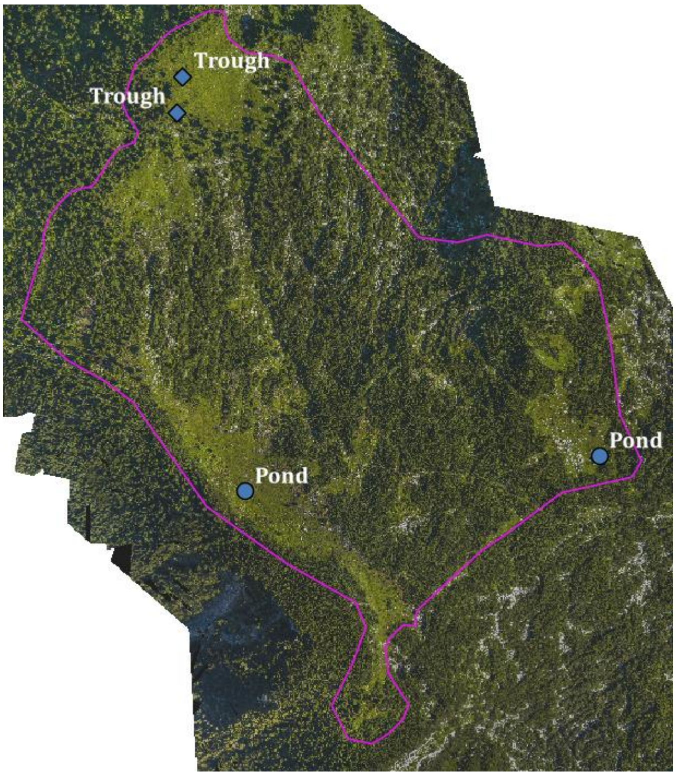

Figure 5: Locations of the water sources found in the study area.

2.2.2 Topography

The most accurate topographic dataset was one that was manually collected. A drone to collect both height data and a RGB imagery in July 2025. This was then processed into .tif files: An aerial mosaic, Digital Terrain Model (DTM) and a Digital Surface Model (DSM). These have a cell size of 0.046 m and a spatial extent covering the entirety of the study area and more.

Next, QGIS was used to create the required topographic datasets for the model. The DTM is directly used as DEM to represent elevation. The same DTM is also used to generate a Slope map, using the 'Slope' tool.

2.2.3 Tree density

The tree density data is generated using QGIS and the SAGA NextGen plugin: First, a Canopy Height Model (CHM) is created by subtracting the DTM from the DSM. While a CHM can be used directly as a covariate for shade, it is not as workable due to the abundant amount of thick shrubs in the area that are not always accessible or practical for cows to hide under. This makes CHM not the most ideal measurement to serve as 'shade and shelter' for cows.

For that reason, a dataset of all trees in the area is required as input. Additionally, a filtered CHM was created, limited by values over 5.5 m. Through manual visual review, it is determined that shrub growth in the area rarely grows above 5.5 m. Using a visualized overlay setup of the aerial photo, CHM, and 6m-filtered CHM, all trees in the study area were digitized as a point feature class. Then, for each tree point feature, the tree height was sampled from the CHM as an attribute. Using this tree feature class, a heatmap is created, using 20m radius Kernel Density Estimation with tree height as weighting, as higher tree can provide more shade. This resulted in a map that provides relative shade intensities from trees (Figure 6).

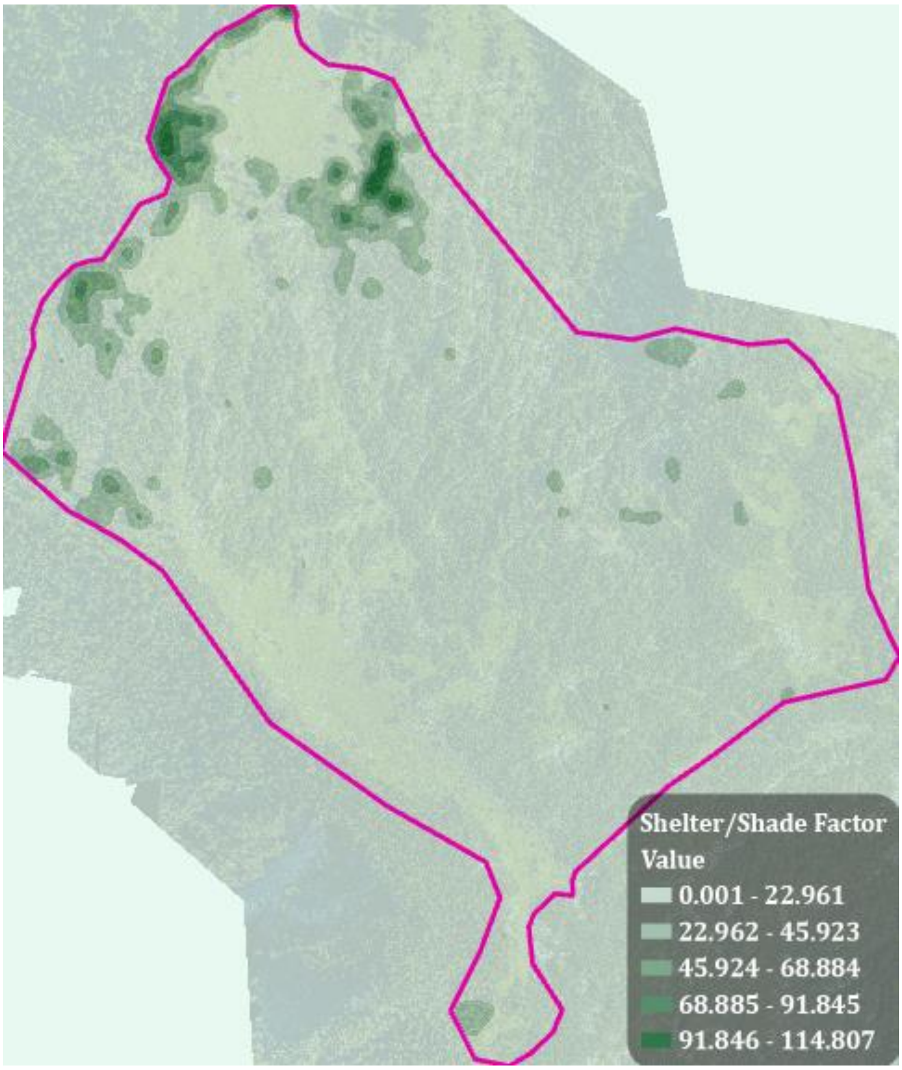

Figure 6: Generated shade/shelter intensity map for the study area, with greener values indicating a higher shade intensity.

2.2.4 Excess Green Index

For vegetation quality, the Normalized Difference Vegetation Index (NDVI) can be used as an indicator for vegetation quality (Raniolo 2022). To compute NDVI, the red and near-infrared (NIR) bands are required. For this study area, the accessible imagery which contain NIR bands did not have the aimed resolution for this study area. An alternative to NDVI is the Excess Green Index (EGI). This index value can be calculated with only RGB bands, as follows:

$$EGI = 2G - R - B$$

In here, G is the green band ratio, R the red band ratio and B the blue band ratio. The ratio is used instead of raw values to account for correct calculations where shadows are present.

This is chosen as to metric to indicate vegetation quality (Figure 7). While this does not represent vegetation quality directly, it can be used as indicator for NDVI and in turn vegetation quality (Xu et al., 2022). This value is calculated by using a created python script with the 'rasterio' and 'numpy' library (Appendix Script C).

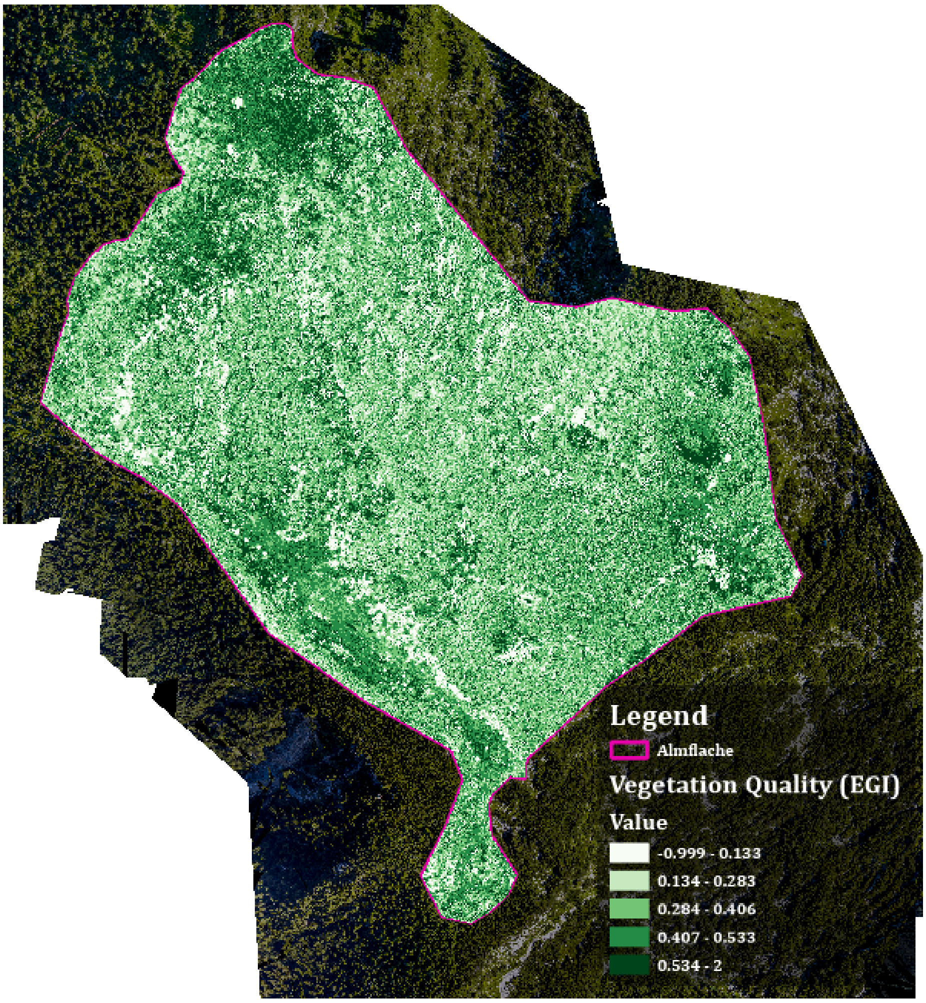

Figure 7: Map of the Excess Green Index for the study area.

2.2.5 Land cover

For vegetation type, a map is created that covers various land cover categories. The most relevant and practically distinguished categories for the model found in the study area are grass, shrub, forest and bare land (Raniolo 2023). In this study area, the class bare land will mostly consists of bare soil, roots, scree and an occasional building.

These categories are calculated using band calculations with the CHM and aerial photo (Figure 8). The Excess Green Index (EGI) is used to distinguish vegetation from non-vegetation, with values over 0.05 consisting of vegetation. Then, for land cover type, the CHM is used with values under 0.16 m representing grassland, values between 0.16 and 5.5 m representing shrubs and values over 5.5 m representing forest (Table 5). These thresholds were determined through a series of visual inspections of specific land types and their corresponding EGI and CHM values.

Land cover type	**True condition for raster calculation**
Bare land	EGI < 0.05
Grass	EGI > 0.05 AND CHM < 0.16
Shrub	EGI > 0.05 AND 0.16 < CHM < 5.5
Forest	EGI > 0.05 AND CHM > 5.5

Table 5: Overview of the conditions for the automatic classification of land cover.

Using the threshold and previously generated EGI, an additional python function was created and added to the original EGI script, using the 'rasterio' library to assign the land cover classification for each pixel (Appendix Script C).

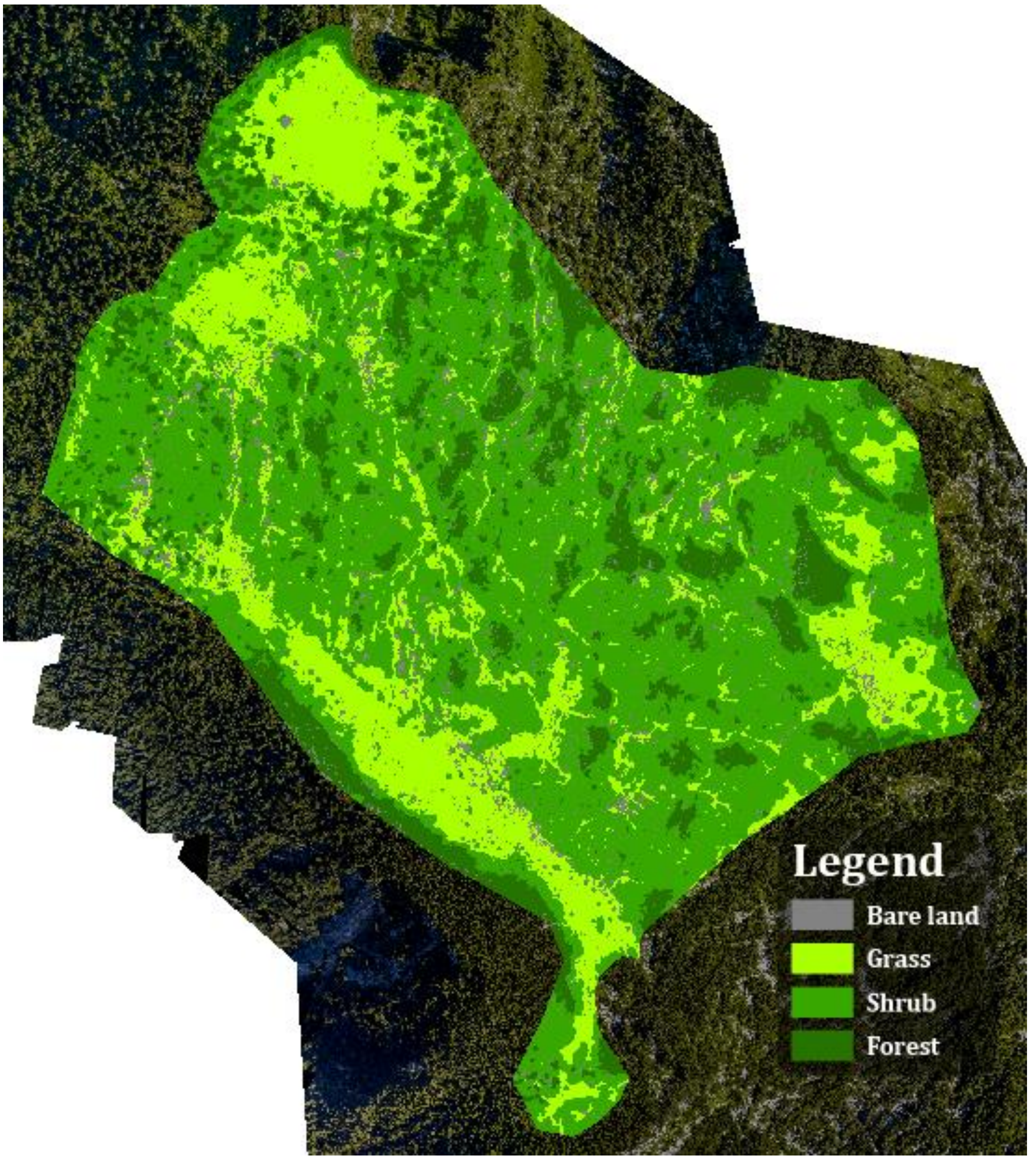

Figure 8: Map of land cover types present in the study area.

2.3 Modelling

The next sections shows how the machine learning models are configured, evaluated and which analysis are performed using the model results.

2.3.1 Model type

There are two types of major machine learning methods that can practically be applied in this study: Classification and regression. In classification, the cow presence is predicted with classes 'present' and 'absent'. While the GPS records act as the present values, the absent values are randomly generated a minimum distance away from the present values in the study area (Appendix Script A). The predictors are then joined to each location. Logistic regression can then be applied to train the model and predict if a cow is likely to be present based on the environmental factors.

In regression, the spatial use intensity is predicted. The study area is split up in grid cells, where the spatial use intensity is calculated from the amount of GPS records present in each grid cells. Then, the mean of the environmental predictors are joined to each grid cell as well. The model will then predict the spatial use intensity based on the environmental factors.

While not part of the main objectives in this study, both classification and regression have been tested using the same datasets. The classification GLM, logistic regression, had a high accuracy of 0.838, but very low recall score of 0.270. The model was unreliable when it came to predicting the present cow locations. The regression models were found to give more consistent results with moderately good Spearman ranking correlation value of 0.68. These model scoring values cannot be directly compared, but are used to give an indication of how well they perform. Due to the relatively poor recall score of the classification model, regression models will be used. Specifically negative binominal regression is chosen as most appropriate GLM for this study, as it takes a datasets dispersion into account when finding the best combination of coefficients for environmental predictors to estimate spatial use intensity.

2.3.2 Model parameterization

The machine learning models are computed using the 'pandas', 'scikit learn' and 'statsmodel' library (Appendix Script D). The final dataset created in chapter 2.2 is read as data frame and filtered to the fields containing cow presence, temperature, wind speed, rainfall past hour, elevation, slope, distance to water source, sun altitude angle, excess green index, tree density and land cover. As land cover is categorical, the 'bareland' land cover field is removed to prevent perfect multicollinearity.

The dataset is separated into a training and test dataset, where 75% of the total data is used for training and 25% for testing. This is done using stratification and the same random seed to establish the same split for each model run.

2.3.3 Resource selection

The results of a fitted NB regression model can be used to compute coefficients for a Resource Selection Function (RSF). This formula consist of an intercept (β_0) and a multitude of coefficients multiplied by the predictor value.

$$w(X) = exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots)$$

These version assumes that all predictors show a linear correlation between predictors and its coefficients. Alternatively, a predictor can also have an quadratic correlation.

$$w(X) = exp(\beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \cdots)$$

Not every environmental factor is by definition impactful on a cow's resource selection. There is a wide variety of predictor combination possibilities. The goal is to seek the combination that fits the model as well as possible. This will be done by testing a broad range of RSF formulas in the NB regression model. This is done by both adding and removing entire predictors from the model, as well as changing the correlations between linear and quadratic.

The model performance can be measured using the Spearman's rank correlation coefficient (r_s). This metric shows how well the ranking of the predicted spatial use intensities correlate to the actual spatial use intensities of the validation dataset. This is relative measure is appropriate for study, as the target value of spatial use is a relative measure. The higher the r_s value, the better the model performance. Another model measure that will be used is Akaike Information Criterion (AIC). This value represents an estimation for prediction error. A high value indicates the model is more prone to predict a wrongful error. Both r_s and AIC will be used to determine which RSF is best to use to fit a model to predict spatial use intensity.

Finally, the Resource Selection Probability Function (RSPF) produces a value that indicates the distribution of resource selection. This is the final value that will be used to create habitat suitability maps that shows which area cows are more likely to use.

$$\tau(X) = \frac{w(X)}{1 + w(X)}$$

2.3.4 Predictor analysis

The derived coefficients cannot be compared to each other directly as the values of these predictors have vast differences in magnitude of order. Therefore, these are standardized to be more easily compared to each other (Silva et al., 2015).

A derived coefficient cannot be compared directly to the same coefficient in another temporal condition, as the *w (x)* for each temporal condition have different scales. Instead, to quantify the impact of a single predictor in multiple conditions, these are plotted using under the same constant mean values of each environmental factor, with only the specific single predictor as variable. The resulting *w(x)* values are then scaled to a probability between 0-1 and plotted for each condition on the same graph. The steepness of the slope indicates the relative effect 'strength' of an environmental factor compared to the same factor in another condition.

2.4 Pasture analysis

This chapter describes how the specific pastures and regions from Chapter 1.2 are analyzed. It includes the computation of spatial use intensity for each area and the before-after analysis of the shrub clearing.

2.4.1 Pasture use intensity

In order to analyze the spatial use for pastures, the relative spatial use intensity is calculated for each pasture and cleared area, for each year. This is done similarly as in chapter 2.1.5, were the sum of weighted cow counts is calculated for each area. Instead of using grid cell as area, the pastures and cleared areas are used.

The habitat suitability for each pasture is also calculated. This is done by calculating the mean predictor values in QGIS using. Then, with the 'Raster Calculation' tool and the derived best fitting resource selection functions from chapter 2.3.3, the habitat suitability of each pasture is calculated.

2.4.2 Pasture impact analysis.

With the spatial use for each year, relative spatial use will be plotted to amount of years after clearing. This is done using Excel, producing various tables and graphs that showcase the differences between cleared areas. This can then be analyzed for potential trends in shrub clearing, spatial use intensity and habitat suitability.

3 Results

3.1 Resource selection functions

This first chapter shows and compares scores of resource selection functions that were trained using the NB model.

3.1.1 Best performing formula

First, the full formula is tested and compared to formulas where a single predictor was removed. Then, if the removal of a specific predictor lead to the best model result, it is further tested by removing an additional predictor. This is done until the removal of a predictor no longer significantly changes either the AIC or r_S. This is done for both day (Table 6) and night (Table 7).

Changes to formula	**Formula for daytime**	**AIC**	**r_s**
All predictors	use_day ~ dtm + slope + wsd + egi + td + lc_grass + lc_shrub + lc_forest	38.63	0.74
Exclude elevation	use_day ~ slope + wsd + egi + td + lc_grass + lc_shrub + lc_forest	36.64	0.73
Exclude excess green index	use_day ~ slope + wsd + td + lc_grass + lc_shrub + lc_forest	34.67	0.73
Exclude shrub and forest land cover	use_day ~ slope + wsd + td + lc_grass	30.69	0.73

Table 6: Overview of changes made to the base daytime model that improved r_S. WSD = Water Source Distance, EGI = Excess Green Index, TD = Tree Density, LC = Land Cover class.

Changes to formula	**Formula for nighttime**	**AIC**	**r_s**
All predictors	use_night ~ dtm + slope + wsd + egi + td + lc_grass + lc_shrub + lc_forest	31.02	0.44
Exclude water source distance	use_night ~ slope + wsd + egi + td + lc_grass + lc_shrub + lc_forest	29.28	0.63
Exclude excess green index	use_night ~ slope + wsd + td + lc_grass + lc_shrub + lc_forest	27.33	0.63
Exclude water source distance	use_night ~ slope + td + lc_grass + lc_shrub + lc_forest	25.34	0.63
Exclude shrub and forest land cover	use_night ~ slope + td + lc_grass	21.42	0.62

Table 7: Overview of changes made to the base nighttime model that improved r_S. WSD = Water Source Distance, EGI = Excess Green Index, TD = Tree Density, LC = Land Cover class.

The addition of an exponent to one of the predictors did not significantly improve the r_s for any of the models for any predictor. The final formulas that predicts the best spatial use for day and night are shown below.

$$w_{day}(X) = exp(\beta_0 + \beta_1 slope + \beta_2 wsd + \beta_3 td + \beta_4 lc\_grass)$$

$$w_{night}(X) = exp(\beta_0 + \beta_1 slope + \beta_2 td + \beta_3 lc\_grass)$$

3.1.2 Conditional functions

These derived RSF are fit to spatial use intensities for each condition specified in 2.1.5 (Table 4). As these models use different input datasets, the AIC value no longer holds any comparative meaning. The formulas are then tested and compared for r_s (Table 8).

Condition	r_s	Intercept	Slope	WSD	TD	GC
Daytime	0.73	-5.138	-0.00372	-0.00376	0.0290	3.745
Daytime above 20 C	0.67	-6.523	0.01905	-0.00843	0.0280	3.152
Daytime below 20 C	0.72	-5.376	-0.00851	-0.00301	0.0290	3.845
Daytime with rain	0.68	-6.768	-0.00246	-0.00071	0.0257	4.062
Daytime without rain	0.71	-0.522	-0.00522	-0.00562	0.0300	3.544
Nighttime	0.62	-4.985	-0.07804		0.0377	3.395
Nighttime above 12 C	0.65	-5.189	-0.08228		0.0385	3.352
Nighttime below 12 C	0.47	-6.643	-0.06324		0.0349	3.487
Nighttime with rain	0.63	-5.370	-0.07792		0.0387	3.524
Nighttime without rain	0.58	-6.122	-0.07684		0.0352	3.006

Table 8: Overview of the RSF formulas performance in spatial use under specific weather conditions. WSD = Water source distance, TD = Tree density, GC = Grassland cover.

3.1.3 Habitat suitability

Using the formulas, habitat suitability maps were created for both daytime (Figure 9) and nighttime (Figure 10). These maps show the RSPF value calculated through the corresponding coefficients used for the RSF.

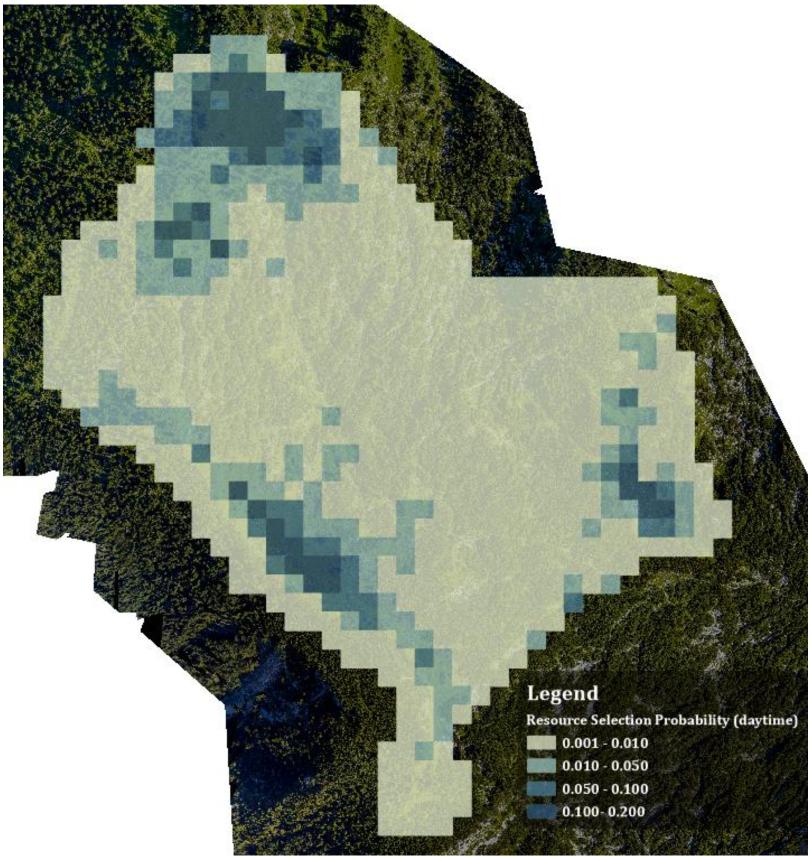

Figure 9: Resource selection probability map for daytime.

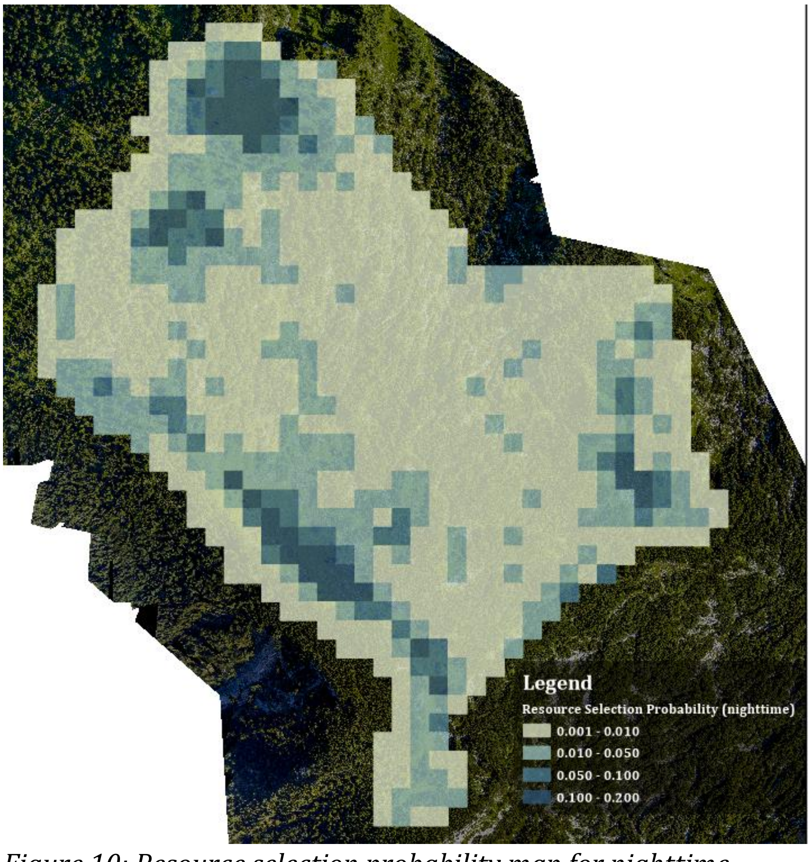

Figure 10: Resource selection probability map for nighttime.

3.2 Effects of predictors

An overview of the mean values, standard deviation and range of the mean predictors present in the dataset grid cells are shown in Table 9.

Predictor	**Unit**	**Mean ± std. dev.**	**Range**
Elevation	m	1665 ± 57	1565 – 1749
Slope	degrees	17.72 ± 5.87	6.95 – 45.75
Water source distance	m	200 ± 93	20 – 397
Excess green index	-	0.36 ± 0.05	0.25 – 0.51
Tree density	-	8.95 ± 11.43	0 – 72.08
Land cover grass	-	0.21 ± 0.24	0 – 1.00
Land cover shrubs	-	0.58 ± 0.23	0 – 0.99
Land cover forest	-	0.16 ± 0.16	0 – 0.94

Table 9: Overview of each environmental predictor used in the model, including their unit, mean and range.

The standardized predictors are calculated and put side by side comparing daytime to nighttime (Figure 11) in addition to comparing each respective time of day for the tested weather conditions (Figure 12 and 13). Positive coefficients imply a higher spatial use when the value increases, negative coefficients imply a lower spatial use when the value increases.

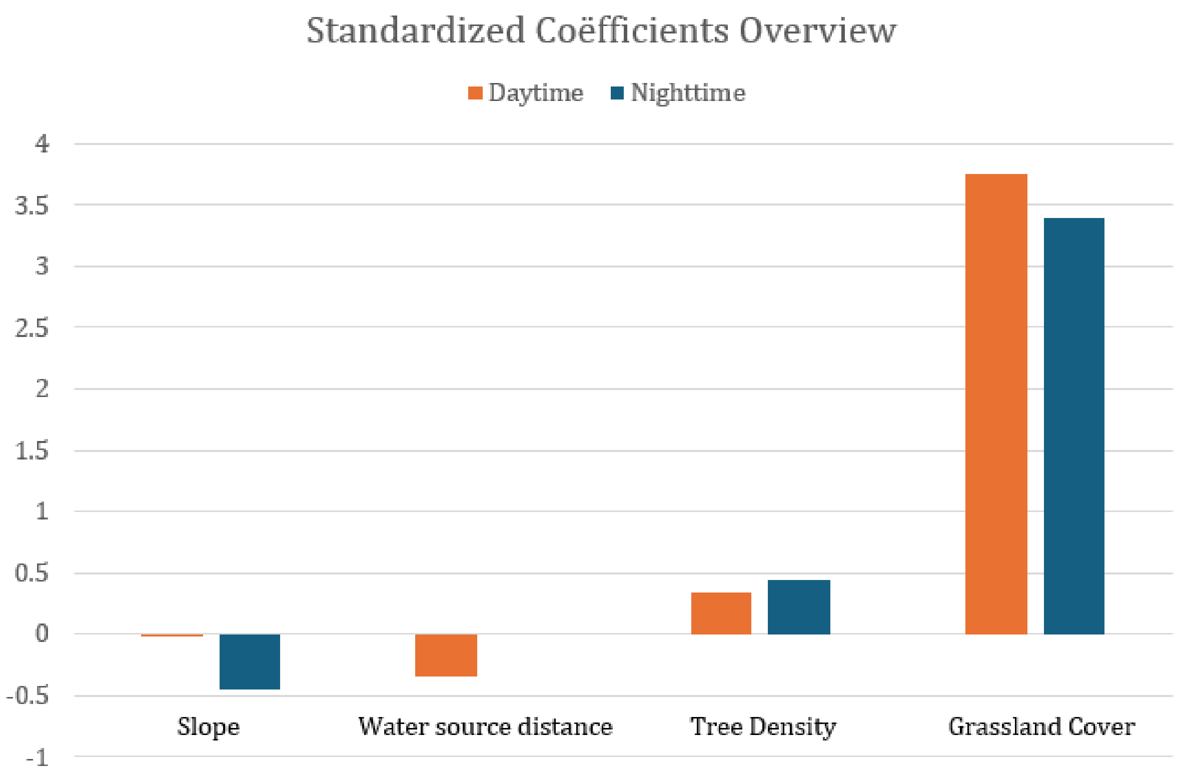

Figure 11: Standardized predictor coefficients for daytime and nighttime.

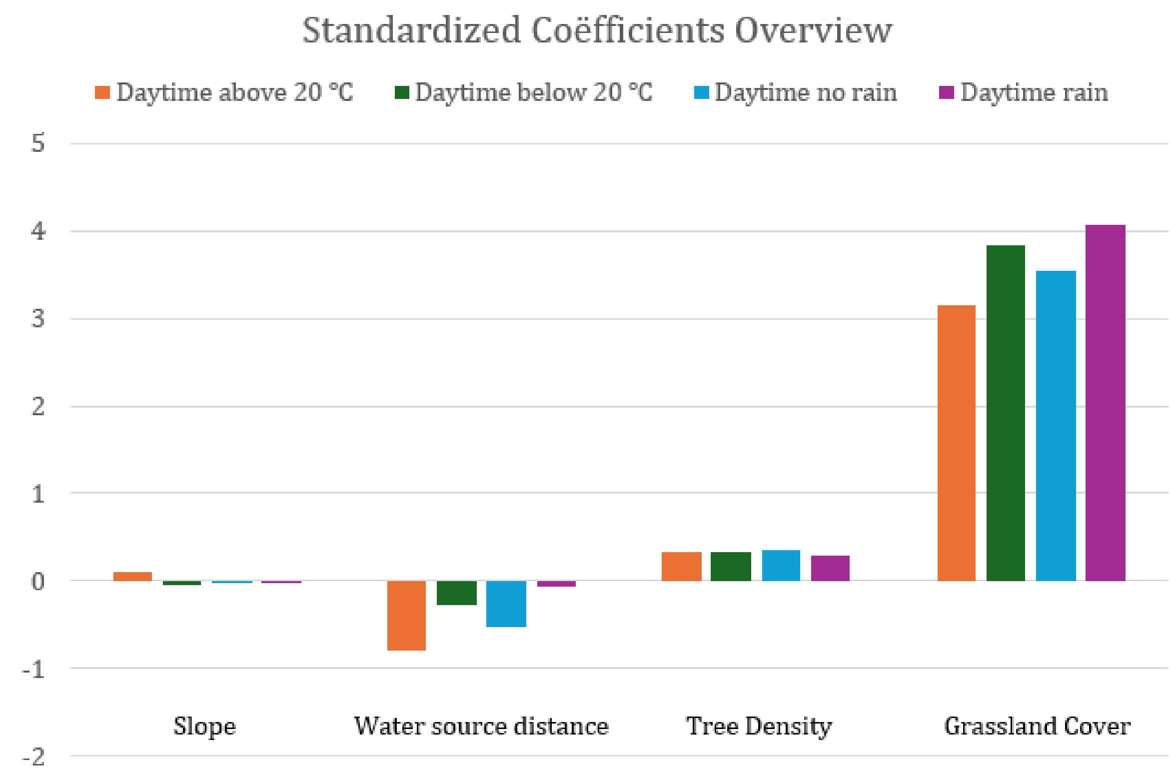

Figure 12: Standardized predictor coefficients for each weather condition during daytime.

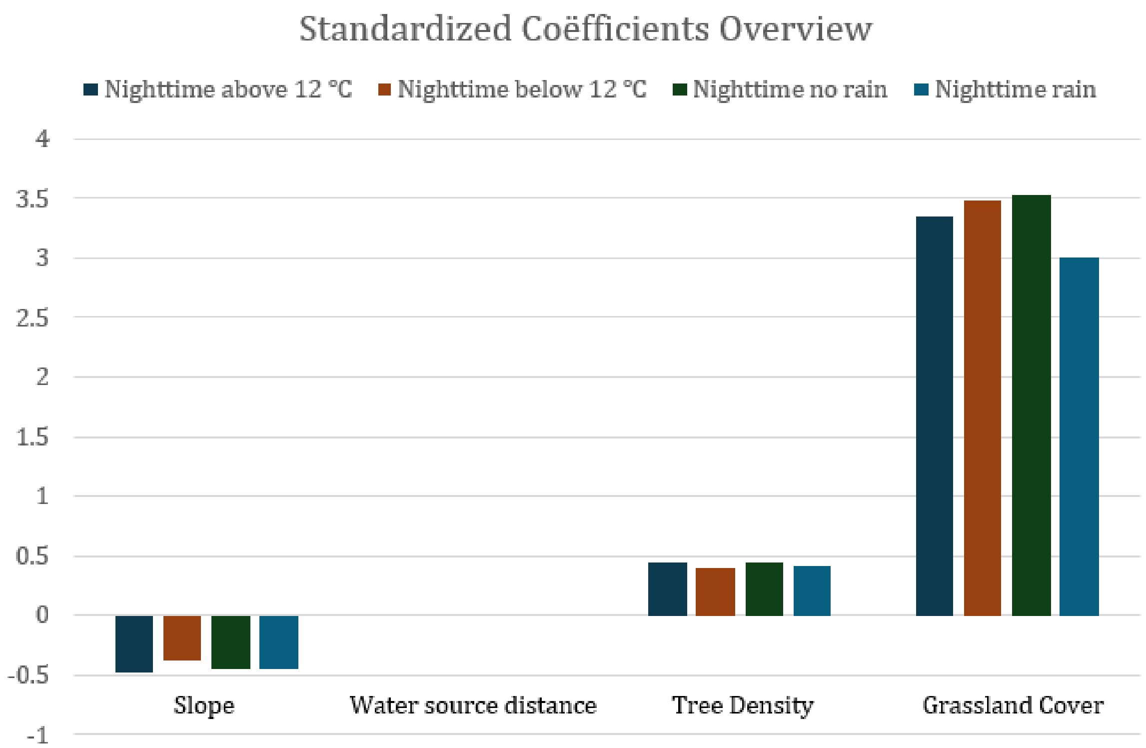

Figure 13: Standardized predictor coefficients for each weather condition during nighttime.

3.2.1 Slope effects

The slope covariate shows a large difference between daytime and nighttime (Figure 14).

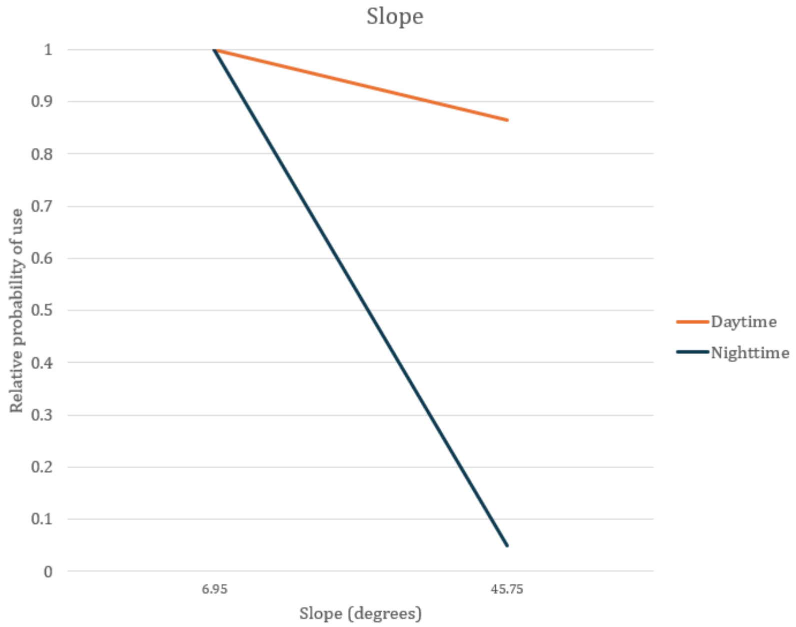

Figure 14: The relative effects of slope on resource selection probability for daytime and nighttime.

Under different weather conditions, slope did not show any differences for cows during nighttime. For daytime, the effect of slope is evidently different, especially for temperatures, even changing between positive and negative correlations (Figure 15).

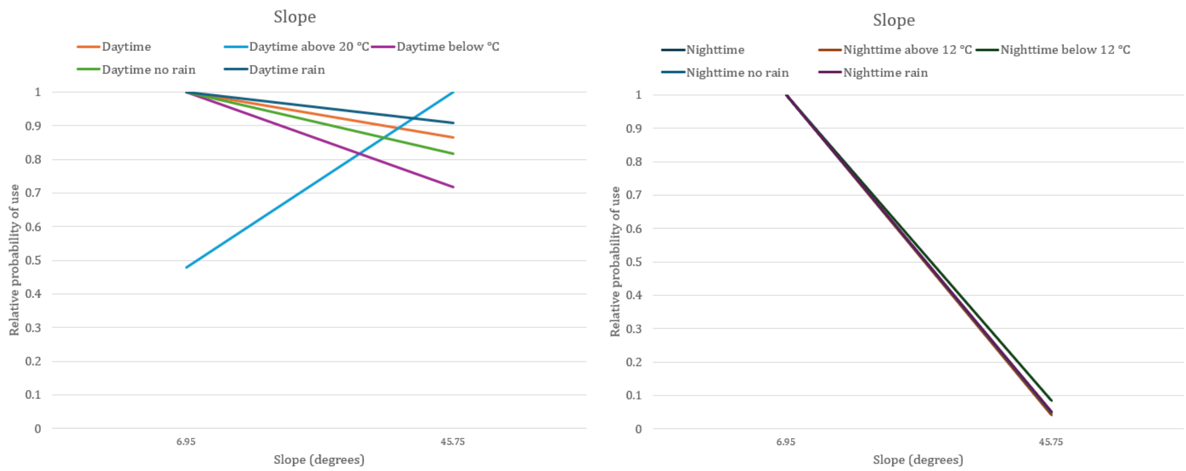

Figure 15: The relative effects of slope on resource selection probability for different weather conditions during daytime and nighttime.

3.2.2 Distance to water source effects

Distance to water source is only a predictor in the RSF for daytime. The distance to water source predictor shows different impacts for each weather condition (Figure 16). Cows are significantly less likely to stay close to water sources during rain than any other condition. Cows are more likely to stay closer to a water source at higher temperatures.

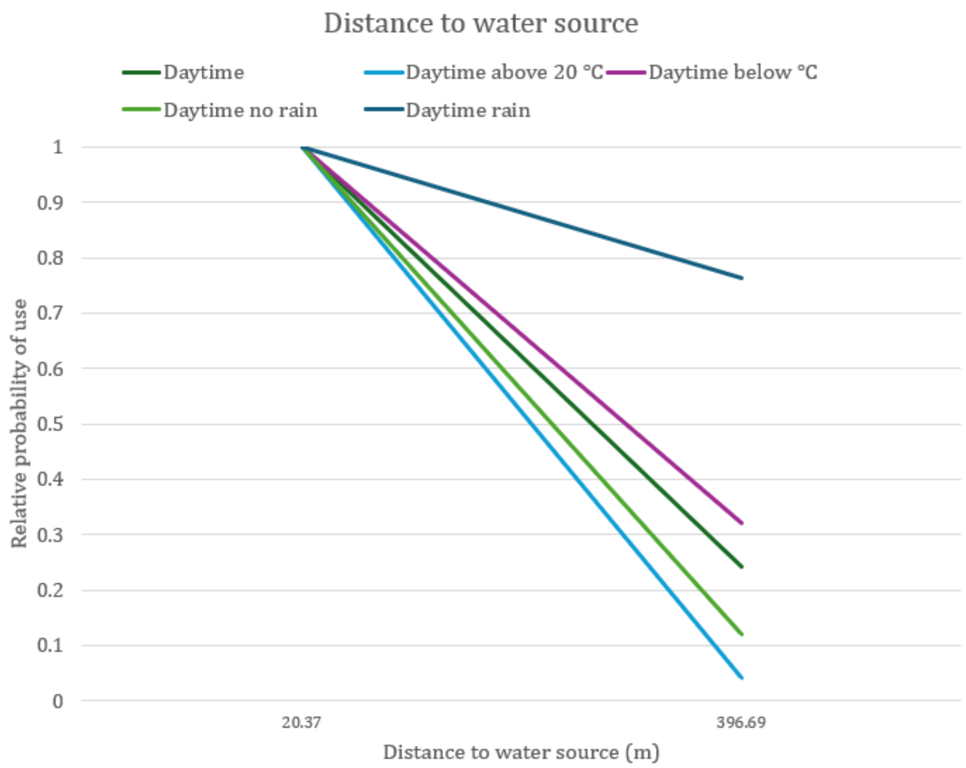

Figure 16: The relative effects of distance to water source on resource selection probability for different weather conditions during daytime.

3.2.3 Tree density effects

Tree density shows similar effects between day and night (Figure 17). There is a stronger preference for locations by higher tree densities during the day.

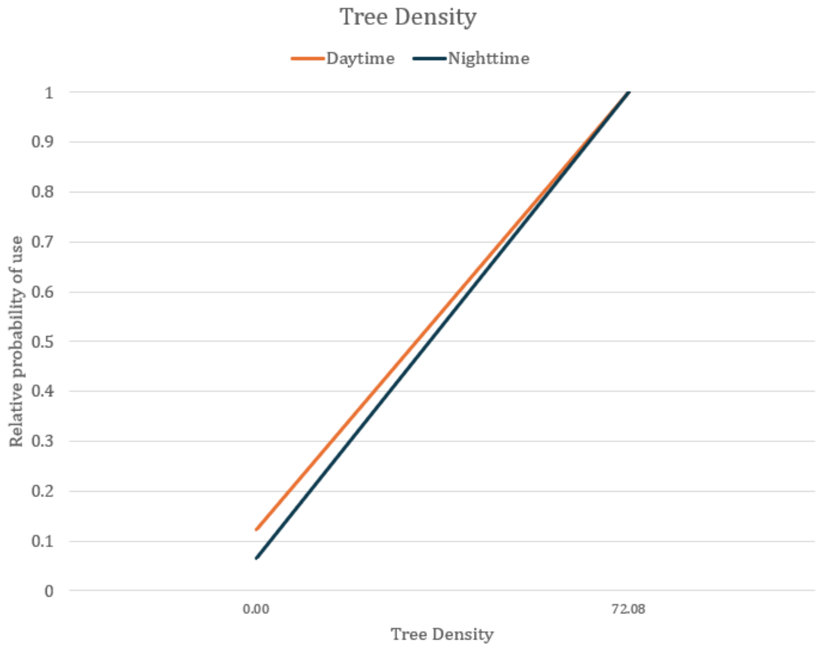

Figure 17: The relative effects of tree density on resource selection probability for daytime and nighttime.

The tree density values show little to no variation in the different weather conditions, for both day and night (Figure 18).

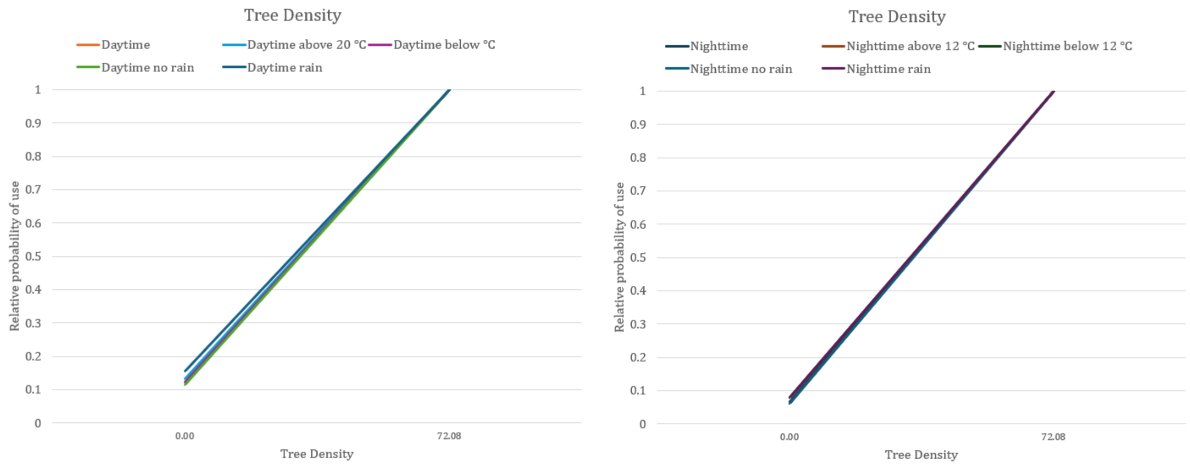

Figure 18: The relative effects of tree density on resource selection probability for different weather conditions during daytime and nighttime.

3.2.4 Grassland cover effects

Grassland cover shows nearly no difference in the resulting resource selection functions between daytime or nighttime (Figure 19).

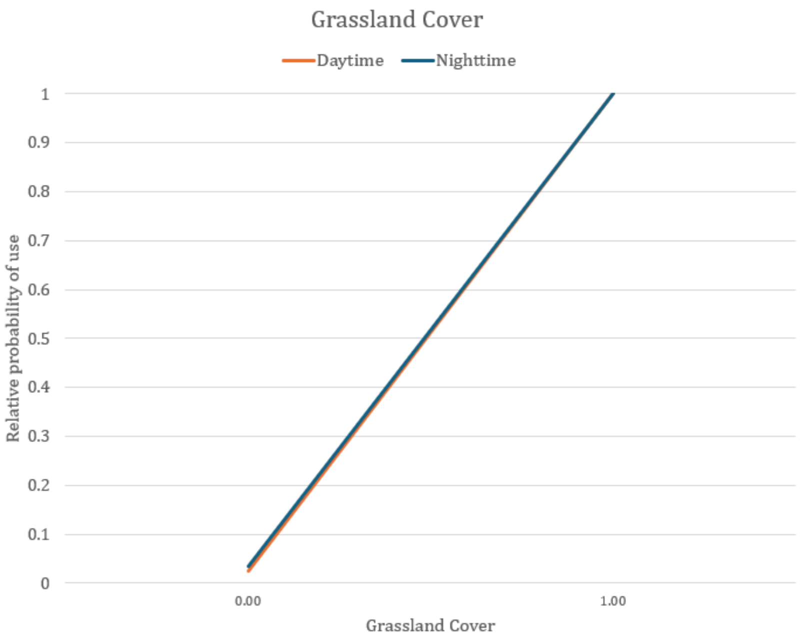

Figure 19: The relative effects of grassland cover on resource selection probability for daytime and nighttime.

Weather conditions barely affect a cow's preference for the amount of grassland cover (Figure 20).

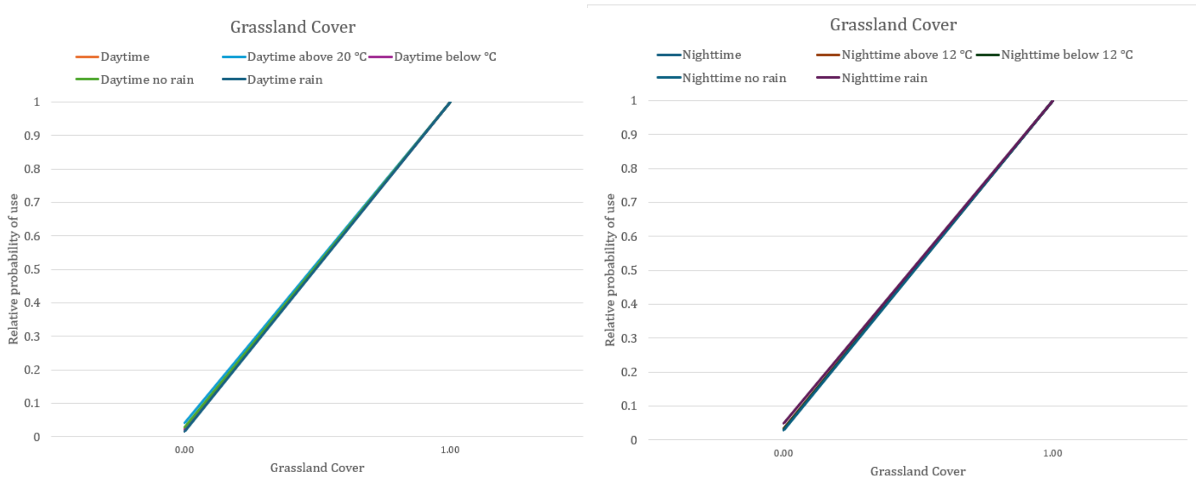

Figure 20: The relative effects of grassland cover on resource selection probability for different weather conditions during daytime and nighttime.

3.3 Pasture preferences

The most used pasture for cows was the Vierkaser meadow, followed by Hirschanger (Figure 21). Both the existing pastures and the cleared areas show different use intensities across the years, with the most drastic changes being in the latter (Figure 22). The spatial use of cleared areas seems to show an increasing trend after clearing, but this was not consistent (Figure 23).

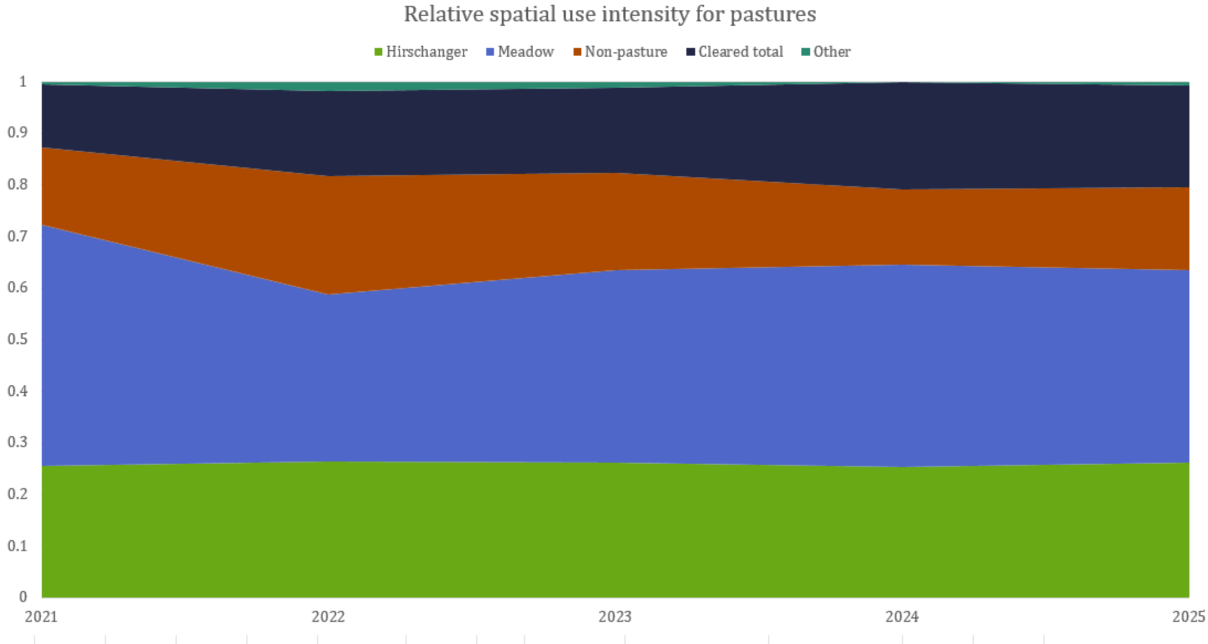

Figure 21: Relative spatial use intensity of all (non-)pastures in the study area.

The size of the pasture is relevant to note, as pastures of the same habitat suitability would naturally have a higher spatial use intensity in the larger one (Table 10).

Pasture	**Area cleared**
Meadow	52,448 m^2 (5.24 ha)
Hirschanger	38,220 m^2 (3.82 ha)
Other	5,167 m^2 (0.52 ha)
2020-10-04	6,496 m^2 (0.65 ha)
2021-08-22	3,659 m^2 (0.37 ha)
2022-08-21	7,395 m^2 (0.74 ha)
2023-08-27	19,287 m^2 (1.93 ha)
2024-08-18	5,206 m^2 (0.52 ha)
2025-06-22	4,039 m^2 (0.40 ha)
Non-pasture	302,859 m^2 (30.28 ha)

Table 10: Size of each analyzed pasture.

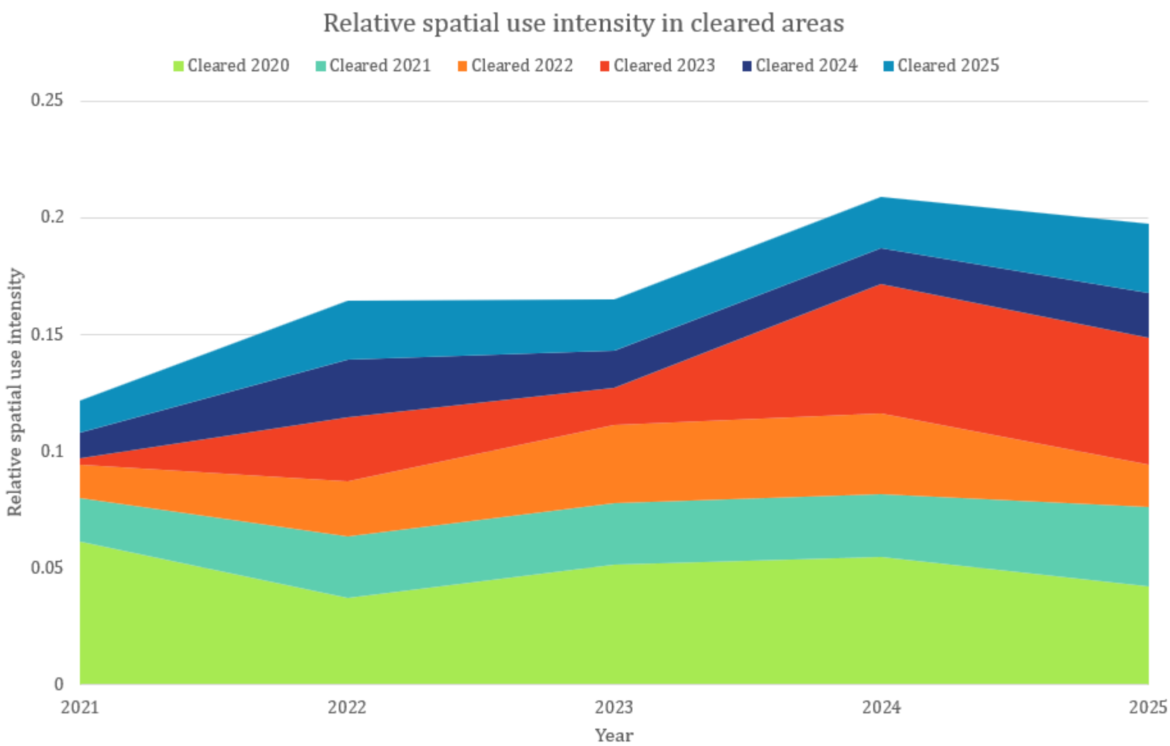

Figure 22: *Relative spatial use intensity of the areas that have not underwent shrub clearing.*

The spatial use changes are also shown with a relative time x-axis that shows spatial use compared to when the area is actually cleared (Figure 23). In this graph, 0 presents the year the shrub clearing happened. This is often in the middle of the Almsommer, around August, but varies per year.

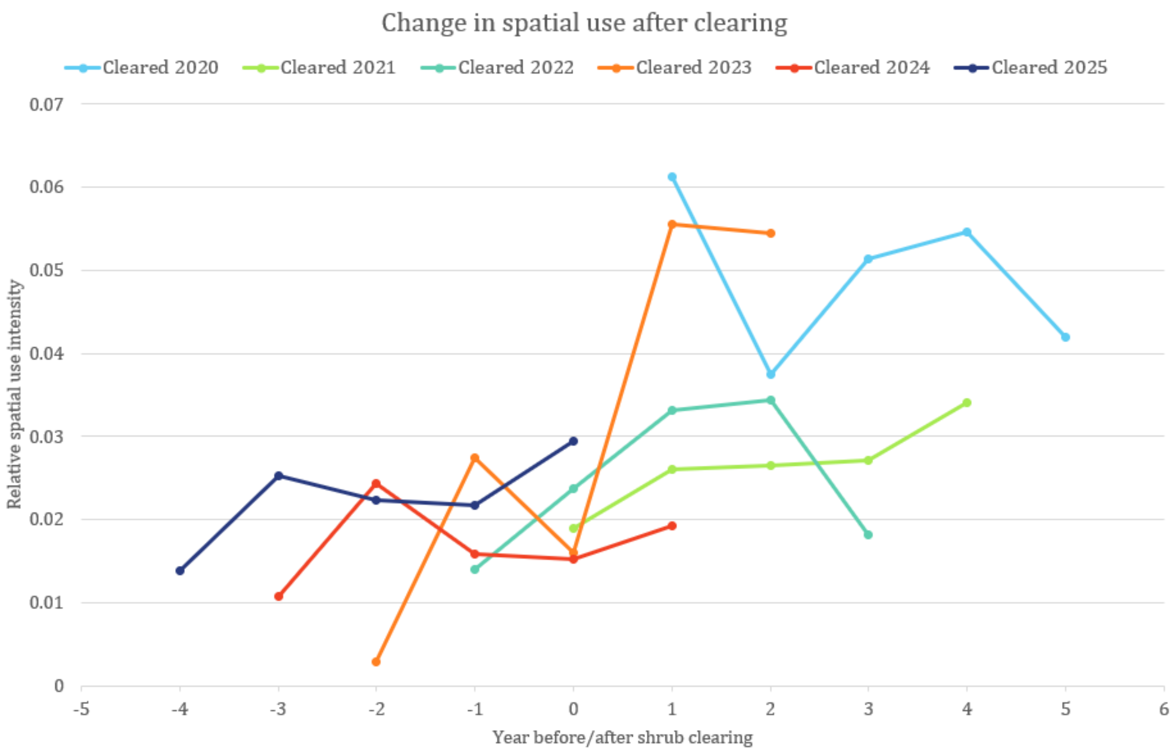

Figure 23: Change in spatial use before and after clearing the given area.

The average modelled habitat suitability was also calculated for each pasture (Figure 24). The Meadow shows to be the best suitable habitat for daytime compared to all other pastures. Especially for daytime, values vary significantly between each pasture whereas most areas are modelled to be fairly suitable for the night.

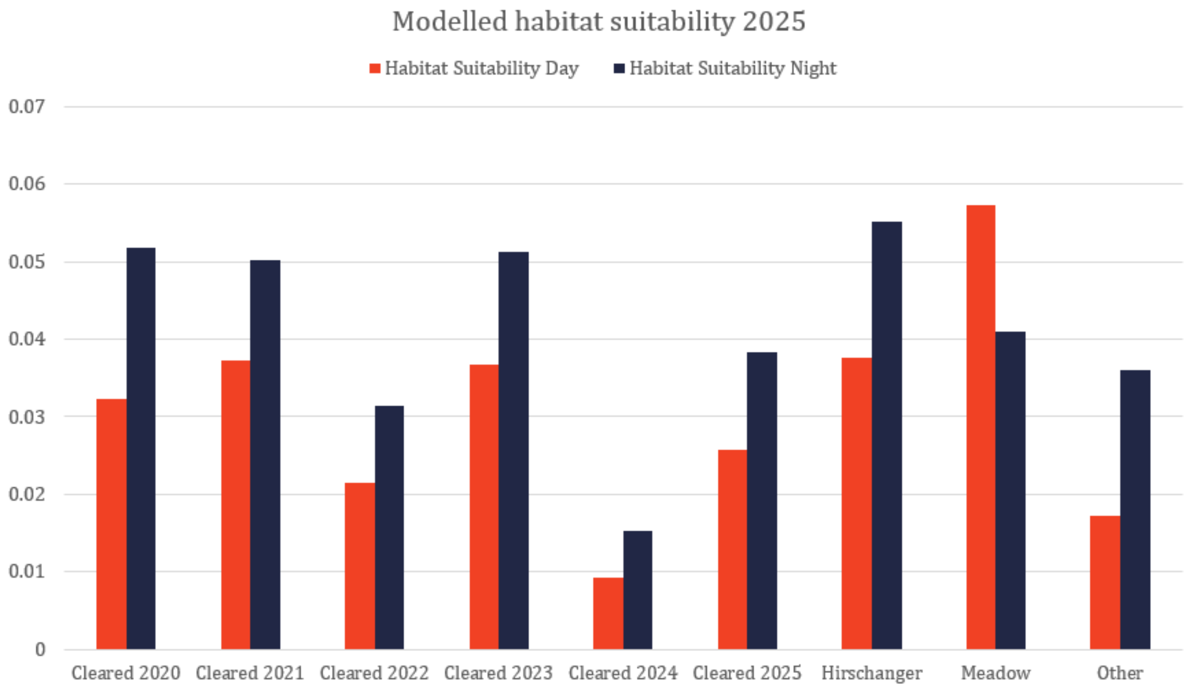

Figure 24: Modelled daytime and nighttime habitat suitability values for each pasture.

4 Discussion

4.1 Resource selection model

This study identifies and quantifies environmental factors to predict and analyze cow movement in resource selection. A GLM was used to model these environmental factors to cattle movement. The following sections discuss the results of the best performing models and the differences of factors between different conditions.

4.1.1 Model accuracy

Models were fit to a wide range of resource selection functions with different combinations of environmental predictions. For daytime, this resulted in the best performing RSF containing slope, water source distance, tree density and grassland cover. For nighttime, this resulted in the best performing RSF containing slope, tree density and grassland cover.

Overall, the r_s for daytime models (0.67-0.73) were significantly better than the ones for nighttime (0.47-0.65) (Table 8). There are multiple ways to interpret and describe the 'correlation strength'. The daytime models can be considered as both a 'moderate-strong' and 'strong' correlation, whereas the nighttime models can be considered as mainly 'moderate' (Akoglu, 2018).

As this study focuses on the environmental factors, excluding biological and social factors, non-perfect scores around 0.7 are fairly common and expected in resource selection (Acciaro, 2024). The relative difference with nighttime models, however, implies that these excluded factors play a larger role in a cow's resource selection compared to just environmental factors. Cows activities, such as lying, eating and walking vary less at increasing temperatures starting from 12 degrees (Hut et al., 2022), which could explain why the model struggles to fit a more consistent RSF with higher r_s value.

4.1.2 Resource selection functions

While the chosen predictors for this study fulfill most of the major environmental factors for cow movement, a few were left out of the models and analysis. One of these is stocking rate, which is known to affect the distances cows are willing to travel from resources such as water supply and shelter, as well as the intensity of grazing less favored areas (Rivero et al., 2021). This was left out of the analysis due to the unclear fencing and natural boundaries of pasture which is required to calculate pasture size. Another

The predictors that came out as having a significant impact on resource selection are slope, water source distance, tree density and grassland cover. Other predictors used in the model, including elevation, excess green index, shrub and forest land cover, were concluded to not have any significant impact on the cows resource selection.

Elevation can be tied with changes in vegetation, temperature or the amount of energy it takes to reach a specific elevation (Rivero et al., 2021). With rather small elevation changes in the study area (Table 9), elevation does not result in very large changes in the mentioned factors.

EGI is an indicator for the herbage mass, however, it does not differentiate between the greenness of edible and inedible vegetation such as shrubland and tree cover. This can cause EGI to be an unstable predictor in a GLM.

Shrub and forest land cover are not grazable. Especially shrubland is considered one of the least attractive types of land cover in most studies (Rivero et al., 2021). Forest can be associated with the Tree Density predictor, but does not take into account the density and height of the trees.

4.2 Movement in weather conditions

This section goes over the differences between the direction and relative strength of a predictor under different temporal conditions (Figure 14-20).

4.2.1 Slope

Slope shows remarkably different effects between day and night (Figure 14). During the day, cows will graze any slope and are barely affected by the slope of the region, nearly negligible. For specific conditions, cows show somewhat similar preferences for slope. Slope has an opposite effect during warm days, where cows instead prefer steeper slopes (Figure 15). Besides complete opposite direction, this effect is also even stronger than the other conditions. While cows vary on how much they are affected by slope, a preference for a steeper slope is more often the result of other spatial relations, rather than the actual driver (Bailey et al., 2021). Some grassland pastures with gentle slopes are relatively far from a water source, another relatively stronger predictor during warm temperatures. However, since slope is barely significant compared to other predictors, slope effects between daytime weather conditions is best left ignored (Figure 11).

During the night, cows seek a much more gentle slope, as they spend most of the night sleeping, which is much more comfortable on a gentle slope (Figure 14). During the nighttime, the weather does not change a cow's preference for slope (Figure 15).

4.2.2 Water source distance

Water source distance only seemed to had a significant effect on the models performance for daytime movement (Figure 16). While cows rest at night, cows do not have the need to drink and be close to a water source.

The effect of distance to water source shows to be in line with how much a cow needs to hydrate (Figure 16). During rain, cows get better hydration and do not need to be close to water sources. During warm temperatures, cows prefer to stay close to water sources. This is due to higher sweating, requiring cows to hydrate more frequently.

4.2.3 Tree density

Tree density shows to be an impactful predictor for both day and night, with a slightly stronger relation during nighttime (Figure 17). For both times of day, there are many reasons to have a preference for areas with a nearby high tree density, which can provide shade and shelter. Seasonal conditions, safety and various weather conditions.

Cows tend to seek for areas that provide shelter and shade during hot or rainy days (Rivero et al., 2021), but the impact of tree density did not vary between different weather conditions that were found in the results (Figure 18).
Areas with a high tree density also imply there is forested land cover, rather than grassland. This steep relationship means that a high tree density can 'compensate' for the loss in grassland cover when considering habitat suitability, whereas sparse or short trees do not.

4.2.4 Land cover

Cows are strongly affected by grassland (Figure 11 and 19). This type of land cover is accessible, a source of food and makes a good resting place due its evenness. It is important to notice that not every area has an equal amount of grass. Some parts of pastures contain sparse shrubs, scree, making them less suitable for cattle.

Between weather conditions, cows preference for grassland does not change (Figure 20). While the preferred vegetation for grazing can differ under different weather conditions, the grassland predictor in this study is too generalized to measure such effects.

4.3 Pasture selection

The spatial use intensity of a pasture is strongly influenced by the size of the pasture. However, for the cleared areas other factors are at play.

4.3.1 Clearing shrubs

While it can vary per year, shrub clearing shows to give an average positive effect on relative spatial use (Figure 22).

The short term effect of clearing shrubs is most predominant. The spatial use increase within the first year of clearing shows to have a substantial impact (Figure 23, increases in values between year 0 and year 1).

The effects of clearing after 2 or more years, shows varying results. There is no guarantee that cows will indefinitely like a specific cleared area over time (Figure 23). This can be due to multiple reasons. Newly cleared pastures are more popular than older cleared pastures. As seen in (Figure 22), the spatial use of cleared pastures overall increases nearly every year. Since the analysis considers relative instead of absolute spatial use, a newly cleared pasture will naturally 'take away' spatial use from another (earlier cleared) pasture.

Another important factor to take into account is stocking rate. The Vierkaser and Hirschanger show to have the highest suitability for day and night respectively. The higher the stocking rate, the higher the likelihood that cows will make use of less suitable areas and have a higher dispersion (Rivero et al., 2021). The stocking rate was significantly lower during 2021 and 2025 (Table 1), with only 5 and 7 cows compared to 10+ present cows in the other years. This can be a reason why the use of the less suitable cleared pastures were much less used than expected. For 2021 to 2022, there was a steep increase in spatial use (Figure 22) despite only a small area being cleared that year (Table 2), including steep increases in use of pastures that had yet to be cleared years later. For 2024 to 2025, a fairly large area was cleared, but only resulted in a slight increase of spatial use in the area.

Not every square meter of cleared are has had to purpose for cows to specifically stay there. Some areas were cleared to grant better moving access between other existing pastures, such as the cleared areas between the Hirschanger and Meadow pastures (Figure 3). Therefore a lower increase in spatial can be expected for such areas.

4.3.2 Future land management

It is suggested to focus on land use by cows during daytime over nighttime, as the preferred area and movement of cows based on environmental predictors is less clear during nighttime (Table 8).

To make an area as appealing as possible during daytime, the key factors to consider are grassland cover, tree density and distance to water source. For daytime, slope can generally be disregarded in land management planning (Figure 11).

The main land management practice that has been applied over the years is shrub clearing. When clearing shrubs, there is more room for grass to grow. Selecting areas with shrubs that are closest to existing water sources or tall tree coverage, have a relative better effect to spatial use by cows.

Alternatively, other methods than clearing shrubs can be considered. There are existing areas, such as the cleared area in 2023 (Figure 3), with an existing high coverage of grassland (Figure 8), that do not have a high spatial use for how well these areas were modelled (Figure 4 and 9). New water sources could be placed in underused areas that already have a high tree density and grassland. Creating better shade and shelter by planting trees is also an option, but can take a long time to grow sufficiently tall. An alternative that works on the short term, would be the construction of artificial shelter.

Many factors come into play when managing an alpine pasture for cattle grazing. The mentioned land management considerations are only based on the results of this study. This is by no means a guide, but can be used as support when creating new land management plans for the future Almsommers.

5 Conclusion

The results of this study gives us a better understanding of the environmental motives for cows to move around the Vierkaser pasture. By using a negative binominal regression model environmental factors can be calculated and modelled as a function to predict relative spatial use intensity. At all times, cows prefer areas that are dominantly covered by grassland and are nearby trees for shade and shelter. During the day, cows prefer to stay in areas that are close to a source of water, especially when temperatures exceed twenty degrees Celsius. However, when it rains, cows no longer intend to stay close to water source. Gentle slopes are important during nighttime but barely relevant during daytime. Clearing shrubs generally results in a higher the use intensity of that area for cows. This is especially so during the first year after clearing shrubs. In the long term, the effect of shrub clearing is less clear. These findings can be considered when deciding the location and size of future shrub clearing and other land management plans.

References

BMLUK - Federal Ministry of Agriculture and Forestry, Climate and Environmental Protection, Regions and Water Management (2024). Statistics, Bovine Livestock.

Buchgraber, K., Schaumberger, A., Pötsch, E. M., Krautzer, B., & Hopkins, A. (2011). Grassland farming in Austria-status quo and future prospective. *Grassland Science in Europe*, *16*(13-24), 95-156.

Gilck, F., & Poschlod, P. (2019). The origin of alpine farming: A review of archaeological, linguistic and archaeobotanical studies in the Alps. *The Holocene*, *29*(9), 1503-1511.

Wrbka, T., Peterseil, J., Schmitzberger, I., & Stocker-Kiss, A. (2004). Alpine farming in Austria, for nature, culture or economic need?. *Frontis*, 165-177.

Dax, T., Schroll, K., Machold, I., Derszniak-Noirjean, M., Schuh, B., & Gaupp-Berghausen, M. (2021). Land abandonment in mountain areas of the EU: An inevitable side effect of farming modernization and neglected threat to sustainable land use. *Land*, *10*(6), 591.

Steinwidder, A., Schneider, M., Wachendorf, M., Starz, W., & Pötsch, E. M. (2011, August). The future of organic grassland farming in mountainous regions of Central Europe. In *Grassland Farming and Land Management Systems in Mountainous Regions. Proceedings of the 16th Symposium of the European Grassland Federation* (pp. 286-296). Organising Committee of the 16th Symposium of the EGF 2011 and Agricultural Research and Education Centre (AREC) Raumberg-Gumpenstein 8952 Irdning, Austria.

Wezel, A., Stöckli, S., Tasser, E., Nitsch, H., & Vincent, A. (2021). Good pastures, good meadows: Mountain farmers' assessment, perceptions on ecosystem services, and proposals for biodiversity management. *Sustainability*, *13*(10), 5609.

Rivero, M. J., Grau-Campanario, P., Mullan, S., Held, S. D., Stokes, J. E., Lee, M. R., & Cardenas, L. M. (2021). Factors affecting site use preference of grazing cattle studied from 2000 to 2020 through GPS tracking: a review. *Sensors*, *21*(8), 2696.

Raniolo, S., Sturaro, E., & Ramanzin, M. (2022). Human choices, slope and vegetation productivity determine patterns of traditional alpine summer grazing. *Italian Journal of Animal Science*, *21*(1), 1126-1139.

Raniolo, S., Sturaro, E., Bovolenta, S., Gianelle, D., & Ramanzin, M. (2023). Activity budget and movement patterns of Brown Swiss and Alpine Grey lactating cows during summer grazing in alpine pastures. *Italian Journal of Animal Science*, *22*(1), 1171-1185.

Probo, M., Lonati, M., Pittarello, M., Bailey, D. W., Garbarino, M., Gorlier, A., & Lombardi, G. (2014). Implementation of a rotational grazing system with large paddocks changes the distribution of grazing cattle in the south-western Italian Alps. *The rangeland journal*, *36*(5), 445-458.

Van Dorland, H. A., Wettstein, H. R., & Kreuzer, M. (2007). Species-rich swards of the Alps: constraints and opportunities for dairy production. *Frontis*, 27-43.

Pittarello, M., Ravetto Enri, S., Lonati, M., & Lombardi, G. (2021). Slope and distance from buildings are easy-to-retrieve proxies for estimating livestock site-use intensity in alpine summer pastures. *PLoS One*, *16*(11), e0259120.

Steyaert, S. M., St∅ en, O. G., Elfström, M., Karlsson, J., Lammeren, R. V., Bokdam, J., ... & Swenson, J. E. (2011). Resource selection by sympatric free-ranging dairy cattle and brown bears Ursus arctos. *Wildlife Biology*, *17*(4), 389-403.

Raynor, E. J., Gersie, S. P., Stephenson, M. B., Clark, P. E., Spiegal, S. A., Boughton, R. K., ... & Augustine, D. J. (2021). Cattle grazing distribution patterns related to topography across diverse rangeland ecosystems of North America. *Rangeland Ecology & Management*, *75*, 91-103.

Falú, E. M. D., Brizuela, M. Á., Cid, M. S., Cibils, A. F., Cendoya, M. G., & Bendersky, D. (2014). Daily feeding site selection of cattle and sheep co-grazing a heterogeneous subtropical grassland. *Livestock Science*, *161*, 147-157.

Shoemaker, K. T., Heffelfinger, L. J., Jackson, N. J., Blum, M. E., Wasley, T., & Stewart, K. M. (2018). A machine-learning approach for extending classical wildlife resource selection analyses. *Ecology and Evolution*, *8*(6), 3556-3569.

O'Malley, W. C., Elbroch, L. M., Zeller, K. A., Beier, P., Beale, M. M., Beausoleil, R. A., ... & Robinson, H. (2024). Machine learning allows for large-scale habitat prediction of a wide-ranging carnivore across diverse ecoregions. Landscape Ecology, 39(5), 106.

Carvalho, J., Santos, J. P. V., Torres, R. T., Santarém, F., & Fonseca, C. (2018). Tree-Based Methods: Concepts, Uses and Limitations under the Framework of Resource Selection Models. *Journal of Environmental Informatics*, *32*(2).

Tomkiewicz, S. M., Fuller, M. R., Kie, J. G., & Bates, K. K. (2010). Global positioning system and associated technologies in animal behaviour and ecological research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1550), 2163-2176.

Halász, A., Nagy, G., Tasi, J. E., Bajnok, M., & Mikó, J. (2016). Weather regulated cattle behaviour on rangeland. *Applied Ecology and Environmental Research*, *14*(4), 149-158.

Eikelboom, J. A., de Knegt, H. J., Klaver, M., van Langevelde, F., van der Wal, T., & Prins, H. H. (2020). Inferring an animal's environment through biologging: Quantifying the environmental influence on animal movement. *Movement Ecology*, *8*(1), 40.

Kraller, G., Warscher, M., Kunstmann, H., Vogl, S., Marke, T., & Strasser, U. (2012). Water balance estimation in high Alpine terrain by combining distributed modeling and a neural network approach (Berchtesgaden Alps, Germany). Hydrology and Earth System Sciences, 16(7), 1969-1990.

Zhou, M., Aarnink, A. J. A., Huynh, T. T. T., Van Dixhoorn, I. D. E., & Koerkamp, P. G. (2022). Effects of increasing air temperature on physiological and productive responses of dairy cows at different relative humidity and air velocity levels. *Journal of dairy science*, *105*(2), 1701-1716.

Acciaro, M., Pittarello, M., Decandia, M., Sitzia, M., Giovanetti, V., Lombardi, G., & Clark, P. E. (2024). Resource selection by Sarda cattle in a Mediterranean silvopastoral system. *Frontiers in Veterinary Science*, *11*, 1348736.

Xu, X., Liu, L., Han, P., Gong, X., & Zhang, Q. (2022). Accuracy of vegetation indices in assessing different grades of grassland desertification from UAV. *International Journal of Environmental Research and Public Health*, *19*(24), 16793.

Hut, P. R., Scheurwater, J., Nielen, M., Van Den Broek, J., & Hostens, M. M. (2022). Heat stress in a temperate climate leads to adapted sensor-based behavioral patterns of dairy cows. *Journal of Dairy Science*, *105*(8), 6909-6922.

Bailey, D. W., Trotter, M. G., Tobin, C., & Thomas, M. G. (2021). Opportunities to apply precision livestock management on rangelands Frontier in Sustainable Food Systems 5.

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish journal of emergency medicine*, *18*(3), 91-93.

Silva, C. P., Garcia, C. E., Estay, S. A., & Barbosa, O. (2015). Bird richness and abundance in response to urban form in a Latin American city: Valdivia, Chile as a case study. *PloS one*, *10*(9), e0138120.

Appendix

Script A: Random cow location generation script

```
#This script will randomly generate 5 fake cow locations for each location

import geopandas as gpd
import pandas as pd

#Input data paths (geopackages paths)
cow_data_path = r""
almflache_path = r""
output_cow_path = r""

#Data reading
cow_data = gpd.read_file(cow_data_path, layer="cows_clipped")
almflache_data = gpd.read_file(almflache_path, layer="Almfläche")
output_cows = cow_data.copy()

#Create new points for each cow
def generate_points_per_cow(cow_data, almflache_data, output_cows):

    #Loop through all cows
    for index, row in cow_data.iterrows():

        #Generate random MULTIPOINT -> POINT
        random_cows_mp = almflache_data.geometry.sample_points(5)
        random_cows_p =
random_cows_mp.explode(index_parts=True).reset_index(drop=True)

        #Join attributes of the currently iterated cow to the geoseries,
except cow_presence
        attributes = row.to_dict()
        df_random_cows = pd.DataFrame([attributes] * len(random_cows_p))
        gdf_random_cows = gpd.GeoDataFrame(df_random_cows,
geometry=random_cows_p)
        gdf_random_cows["cow_presence"] = 0

        #Combine it with the output gdf
        output_cows = pd.concat([gdf_random_cows, output_cows],
ignore_index=True)
        print(f"Currently at: {index}")

    output_cows.to_file(output_cow_path, driver="GPKG",
layer="cows_randomized")
    print("Finished")

generate_points_per_cow(cow_data, almflache_data, output_cows)
```

Script B: Sunlight altitude angle calculator

```
#Calculate the level of darkness / solar altitude angle (SAA) from a location
and datetime
from sun_position_calculator import SunPositionCalculator
import geopandas as gpd
import pandas as pd
import math

#Inputs from geopackage path and layername
cow_data_path = r""
cow_layer = ""

#Reading standard lat/lon since altitude differences are negligible within the
study area
cow_data = gpd.read_file(cow_data_path, layer=cow_layer)
lat = 47.71
lon = 12.95

#Create new field
cow_data["saa"] = None
cow_data["saa"] = cow_data["saa"].astype("float64")

#Calculate SAA in each field from datetime and loc
sunposcalc = SunPositionCalculator()
for index, row in cow_data.iterrows():
    print(f"Calculating row {index}")
    unixtime = pd.Timestamp(row["dt"]).timestamp() * 1000
    position = sunposcalc.pos(unixtime, lat, lon)
    cow_data.loc[index, "saa"] = math.degrees(position.altitude)

cow_data.to_file(cow_data_path, driver="GPKG", layer="cows_w_t_saa_lc_w_chm")
print("SAA Calculations are finished")
```

Script C: Calculate the EGI and Land Cover

```
import rasterio
import numpy

#Paths
mosaic = rasterio.open(r"")
chm = rasterio.open(r"")
output = r""
output_egi = r""

def calculateLandcover():
    #Reading raster bands
    red = mosaic.read(1).astype("float32")
    green = mosaic.read(2).astype("float32")
    blue = mosaic.read(3).astype("float32")
    height = chm.read(1).astype("float32")

    #Calculations for EGI
    rgb_total = red + blue + green
    red_ratio = red / rgb_total
    green_ratio = green / rgb_total
    blue_ratio = blue / rgb_total

    exg = 2 * green_ratio - red_ratio - blue_ratio

    result = numpy.zeros_like(exg, dtype=numpy.uint8)

    #Classification
    result[(exg <= 0.05)] = 1                                         #Bareland
    result[(exg > 0.05) & (height <= 0.16)] = 11                      #Grassland
    result[(exg > 0.05) & (height > 0.16) & (height <= 5)] = 12      #Scrub
    result[(exg > 0.05) & (height > 5)] = 13                          #Forest

    #Save output
    with rasterio.open(output, "w", driver="GTiff", height=result.shape[0],
width=result.shape[1], count=1, dtype=result.dtype,
crs=chm.crs,transform=chm.transform) as outputwriter:
        outputwriter.write(result, 1)

    mosaic.close()
    chm.close()

def calculateEGI():
    #Reading raster bands
    red = mosaic.read(1).astype("float32")
    green = mosaic.read(2).astype("float32")
    blue = mosaic.read(3).astype("float32")
```

```
    #Calculations for EGI
    rgb_total = red + blue + green
    red_ratio = red / rgb_total
    green_ratio = green / rgb_total
    blue_ratio = blue / rgb_total
    exg = 2 * green_ratio - red_ratio - blue_ratio

    #Save output
    with rasterio.open(output_egi, "w", driver="GTiff", height=exg.shape[0],
width=exg.shape[1], count=1, dtype="float32",
crs=chm.crs,transform=chm.transform) as outputwriter:
        outputwriter.write(exg, 1)

    mosaic.close()

calculateLandcover()
calculateEGI()

print("The calculation has finished")
```

Script D: Negative Binominal regression model script

```
import pandas
import statsmodels.api as sm
import statsmodels.formula.api as smf
from scipy.stats import spearmanr
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

#Read input data from path
data_path = r""
data = pandas.read_csv(data_path)

#Split into train and test dataset
train_data, test_data = train_test_split(data, test_size=0.25,
random_state=108384)

#List of environmental factors column names and optional scaling (#comment
this section to remove scaling)
factors = ["dtm", "slope", "wsd", "egi", "td", "lc"]
scaler = StandardScaler()
train_data[factors] = scaler.fit_transform(train_data[factors])
test_data[factors] = scaler.fit_transform(test_data[factors])

#Output path
output_path = r""

#Resource selection formulas to test
rsf_dict = {}
rsf_dict["general_all"] = "use_general ~ dtm + slope + wsd + egi + td +
lc_grass + lc_shrub + lc_forest"
rsf_dict["general"] = "use_general ~ slope + td + lc_grass"
# rsf_dict["day1"] = "use_day ~ dtm + slope + wsd + egi + td + lc_grass +
lc_shrub + lc_forest"
# rsf_dict["day2"] = "use_day ~ slope + wsd + egi + td + lc_grass + lc_shrub +
lc_forest"
# rsf_dict["day3"] = "use_day ~ slope + wsd + td + lc_grass + lc_shrub +
lc_forest"
rsf_dict["day"] = "use_day ~ slope + wsd + td + lc_grass"
# rsf_dict["night1"] = "use_night ~ dtm + slope + wsd + egi + td + lc_grass +
lc_shrub + lc_forest"
# rsf_dict["night2"] = "use_night ~ dtm + slope + egi + td + lc_grass +
lc_shrub + lc_forest"
# rsf_dict["night3"] = "use_night ~ slope + egi + td + lc_grass + lc_shrub +
lc_forest"
# rsf_dict["night4"] = "use_night ~ slope + td + lc_grass + lc_shrub +
lc_forest"
rsf_dict["night"] = "use_night ~ slope + td + lc_grass"
rsf_dict["day_above20"] = "use_day_above20 ~ slope + wsd + td + lc_grass"
```

```
rsf_dict["day_above20_2"] = "use_day_above20 ~ wsd + td + lc_grass"
rsf_dict["day_below20"] = "use_day_below20 ~ slope + wsd + td + lc_grass"
rsf_dict["day_rain"] = "use_day_rain ~ slope + wsd + td + lc_grass"
rsf_dict["day_clear"] = "use_day_clear ~ slope + wsd + td + lc_grass"

rsf_dict["night_above12"] = "use_night_above12 ~ slope + td + lc_grass"
rsf_dict["night_below12"] = "use_night_below12 ~ slope + td + lc_grass"
rsf_dict["night_rain"] = "use_night_rain ~ slope + td + lc_grass"
rsf_dict["night_clear"] = "use_night_clear ~ slope + td + lc_grass"

#Fit the model using a given RSF
def modelData(name, rsf, train_data, test_data):

    #Create copy for future df modifications
    dataframe = train_data.copy()
    test_df = test_data.copy()

    #Create a model using this formula
    model = smf.glm(
        formula=rsf,
        data=dataframe,
        family=sm.families.NegativeBinomial(alpha=1.0)
    )

    #Fit the model to spatial use value
    result = model.fit()

    #Print results
    print(f"{name} using {rsf}")
    summary_table = result.summary2().tables[1]
    summary_table.to_csv(rf"{output_path}\std_summary_{name}.csv")

    #Seperate the spatial use name (for day/night variation purpose)
    use_variable = rsf.split("~")[0].strip()

    #Predict values for the validation set and split them into 10 bins
    test_df["rsf_calc"] = result.predict(test_df)
    rs_value, p_value = spearmanr(test_df[use_variable], test_df["rsf_calc"])

    print(f"AIC = {result.aic} and rs {rs_value} and p_value = {p_value}")

#Test a model for each resource selection formula
for name, rsf in rsf_dict.items():
    modelData(name, train_data, test_data)
```

Script E: Weather data cleaning and joining

```
import pandas as pd
import geopandas as gpd

#This script joins cow timestamps to the nearest weather data record
cow_data_path = r""
weather_data_path = r""
cow_layer = "cows_randomized"
cow_output = "cows_weatherjoined"

#Read files and read datetime as same format
cow_data = gpd.read_file(cow_data_path, layer=cow_layer)
cow_data_sorted = cow_data.sort_values("dt")

weather_data = pd.read_csv(weather_data_path)
weather_data["dt_new"] = pd.to_datetime(weather_data["dt_new"])
weather_data["dt_new"] = weather_data["dt_new"].dt.tz_localize("UTC")
weather_data["dt_new"] = weather_data["dt_new"].astype("datetime64[ms, UTC]")

#Join the weather data to cow data and save
cow_weather_data = pd.merge_asof(cow_data_sorted, weather_data, left_on="dt",
right_on="dt_new", direction="nearest")
cow_weather_geodata = gpd.GeoDataFrame(cow_weather_data, geometry="geometry",
crs=cow_data.crs)
cow_weather_geodata.to_file(cow_data_path, driver="GPKG",
layer="cows_weatherjoined")
```

Script F: Raw cow GPS data SQL generation script

```
###############
# Quick guide #
###############

# 1. Create a new database in your (local) postgresql server
# 2. Enable the PostGIS extension
# 3. Create a 'raw_data' schema and a 'processed' schema
# 4. Import track_points from the .gpx and ,gpkg files into 'raw_data' using
the naming convention below
#       name: [trackertype]_[trackerid]_[year]_[source]_[cow name]
#       primary key: id
#       geometry column: geom
#       source srid: 4326
#       target srid: 4326
#       convert field names to lowercase
# 5. Change the parameters in this .py file below to the correct folder
locations in your pc
# 6. Add/change any data to the data_files dictionary if available
# 7. Run the script
# 8. Import and run the 'all_sql_combined.sql' into your pgadmin

##############
# Parameters #
##############

#Location of your 'raw_data' folder.
#I had to move this inside my PostGres installation folder for it to work
data_folder = r""

#Location of your output folder, in this folder a collection of .sql files
will be saved
output_folder = r""

######################################################
# Dictionary for each cow name and file location     #
# Add available data to this dictionary if available #
# Exact naming convention is VERY IMPROTANT:         #
# '[tracker]_[deviceID]_[year]_[source]_[cowname]'   #
# This is important, because from this name, the     #
# attribute values are read and written              #
######################################################

data_files = {
    "bene_gps01_2021_gpkg_pia":
rf"{data_folder}\2021\2021_Vierkaser_Bene_Pia.gpkg",
    "q2_24403_2021_csv_geri": rf"{data_folder}\2021\Q2 24403 2021-06-26 00 00
00-2021-09-12 00 00 00.csv",
```

```
    "q2_24693_2021_csv_gretel": rf"{data_folder}\2021\Q2 24693 2021-06-26 00 
00 00-2021-09-12 00 00 00.csv",
    "q2_24906_2021_csv_pia": rf"{data_folder}\2021\Q2 24906 - Pia 2021-06-26 
00 00 00-2021-09-12 00 00 00.csv",
    "q2_26845_2021_gpx_haensel": rf"{data_folder}\2021\Q2 26845 2021-06-26 00 
00 00-2021-09-12 00 00 00.gpx",
    "q2_35707_2021_csv_susi": rf"{data_folder}\2021\Q2 35707 2021-06-26 00 00 
00-2021-09-12 00 00 00.csv",

    "bene_gps01_2022_csv_unknown01": rf"{data_folder}\2022\GPS01.csv",
    "bene_gps02_2022_csv_unknown02": rf"{data_folder}\2022\GPS02.csv",
    "bene_gps03_2022_csv_unknown03": rf"{data_folder}\2022\GPS03.csv",
    "bene_gps04_2022_csv_unknown04": rf"{data_folder}\2022\GPS04.csv",
    "bene_gps05_2022_csv_unknown05": rf"{data_folder}\2022\GPS05.csv",
    "bene_gps06_2022_csv_unknown06": rf"{data_folder}\2022\GPS06.csv",
    "bene_gps07_2022_csv_unknown07": rf"{data_folder}\2022\GPS07.csv",
    "bene_gps08_2022_csv_unknown08": rf"{data_folder}\2022\GPS08.csv",
    "q2_24693_2022_gpx_unknown": rf"{data_folder}\2022\Q2 24693 2022-06-03 00 
00 00-2022-10-14 00 00 00.gpx",
    "q2_24906_2022_gpx_julia": rf"{data_folder}\2022\Q2 24906 - JULIA 3 2022-
06-03 00 00 00-2022-10-14 00 00 00.gpx",
    "q2_35811_2022_gpx_pia": rf"{data_folder}\2022\Q2 35811 Pia 1 2022-06-03 
00 00 00-2022-10-14 00 00 00.gpx",

    "q2_24403_2023_csv_gretel": rf"{data_folder}\2023\Q2 24403 GRETEL 2 2023-
06-10 00 00 00-2023-09-19 00 00 00.csv",
    "q2_24693_2023_csv_unknown": rf"{data_folder}\2023\Q2 24693 2023-06-10 00 
00 00-2023-09-19 00 00 00.csv",
    "q2_24906_2023_csv_julia": rf"{data_folder}\2023\Q2 24906 - JULIA 3 2023-
06-10 00 00 00-2023-09-19 00 00 00.csv",
    "q2_26845_2023_csv_haensel": rf"{data_folder}\2023\Q2 26845 Haensel 4 
2023-06-10 00 00 00-2023-09-19 00 00 00.csv",
    "q2_35707_2023_csv_susi": rf"{data_folder}\2023\Q2 35707 SUSI 5 2023-06-10 
00 00 00-2023-09-19 00 00 00.csv",
    "q2_35811_2023_csv_pia": rf"{data_folder}\2023\Q2 35811 Pia 1 2023-06-10 
00 00 00-2023-09-19 00 00 00.csv",

    "q2_24403_2024_gpx_gamsei": rf"{data_folder}\2024\Q2 24403 Gamsei 2024-06-
06 00_00_00-2024-09-12 00_00_00.gpx",
    "q2_24693_2024_gpx_sternei": rf"{data_folder}\2024\Q2 24693 Sternei 2024-
06-06 00_00_00-2024-09-12 00_00_00.gpx",
    "q2_24906_2024_gpx_julia": rf"{data_folder}\2024\Q2 24906 - JULIA 3 2024-
06-06 00_00_00-2024-09-12 00_00_00.gpx",
    "q2_26845_2024_gpx_haensel": rf"{data_folder}\2024\Q2 26845 H・sel 4 2024-
06-06 00_00_00-2024-09-12 00_00_00.gpx",
    "q2_35707_2024_gpx_karli": rf"{data_folder}\2024\Q2 35707 KARLI 5 2024-06-
06 00_00_00-2024-09-12 00_00_00.gpx",
```

```
    "q2_35811_2024_gpx_pia": rf"{data_folder}\2024\Q2 35811 Pia 1 2024-06-06
00_00_00-2024-09-12 00_00_00.gpx",

    "q2_24403_2025_gpx_gamsei": rf"{data_folder}\2025\Q2 24403 Gamsei 2025-06-
06 00_00_00-2025-09-15 00_00_00",
    "q2_24693_2025_gpx_sternei": rf"{data_folder}\2025\Q2 24693 Sternei 2025-
06-06 00_00_00-2025-09-15 00_00_00",
    "q2_24906_2025_gpx_nanni": rf"{data_folder}\2025\Q2 24906 - Nanni 2025-06-
06 00_00_00-2025-09-15 00_00_00",
    "q2_26845_2025_gpx_haensel": rf"{data_folder}\2025\Q2 26845 Hänsel 4 2025-
06-06 00_00_00-2025-09-15 00_00_00",
    "q2_35707_2025_gpx_susi": rf"{data_folder}\2025\Q2 35707 SUSI 5 2025-06-06
00_00_00-2025-09-15 00_00_00",
    "q2_35811_2025_gpx_pia": rf"{data_folder}\2025\Q2 35811 Pia 1 2025-06-06
00_00_00-2025-09-15 00_00_00"
}

all_sql_content = []

#This is the main function
def main_script():

    #First create a table
    create_combined_table()

    #Create a sql file for each file, with a different workflow for each file
+ tracker type.
    #These are all based on the given 'kuehe_to_databse.sql' template
    for name, filepath in data_files.items():

        if "csv" in name and "q2" in name:
            process_q2_csv(name, filepath)

        elif "csv" in name and "bene" in name:
            process_bene_csv(name, filepath)

        elif "gpx" in name:
            process_gpx(name)

        elif "gpkg" in name:
            process_gpkg(name)

        else:
         return

    #Combines all recreated sql files into a single one
    combine_sql(all_sql_content)
```

```python
#Creates an empty table 'vierkaser_daten_gesammelt' for the final processed
table of each cow combined
def create_combined_table():
    sqlstring = """
    create table processed.vierkaser_daten_gesammelt (
        id SERIAL,
        dt timestamp,
        geom geometry,
        alt double precision,
        device_id varchar(20),
        cowname varchar(10),
        source varchar(10),
        tracker varchar(10),
        sat int,
        hdop double precision,
        fix varchar(10),
        age int,
        gpslev int,
        gsmlev int,
        mcc varchar(10),
        mnc int,
        course double precision,
        speed double precision,
        card varchar(10),
        batl int,
        odo int
    );"""
    save_sql("create_combined_table.sql",sqlstring)
    all_sql_content.append(sqlstring)

#Function for processing .csv files
def process_q2_csv(name, filepath):
    tablename = name
    datasource = filepath
    attributes = extract_attributes(name)
    sqlstring = f"""
    create table raw_data.{tablename} (
        dt timestamp,
        lat double precision,
        lng double precision,
        altitude int,
        angle int,
        speed int,
        params varchar(500)
    );

    COPY raw_data.{tablename} (dt, lat, lng, altitude, angle, speed, params)
    FROM '{datasource}'
```

```
    DELIMITER ','
    CSV HEADER;

    create table processed.{tablename} (
        dt timestamp,
        geom geometry(Point, 31258),
        altitude int,
        angle int,
        speed int,
        acc int,
        arm int,
        batl int,
        cellid int,
        gpslev int,
        gsmlev int,
        lac int,
        mcc varchar(20),
        mnc int,
        odo int,
        pwrcut int,
        shock int,
        device_id varchar(20),
        cowname varchar(10),
        source varchar(10),
        tracker varchar(10)
    );

    INSERT INTO processed.{tablename} (dt, geom, altitude, angle, speed, acc,
arm, batl, cellid, gpslev, gsmlev, lac, mcc, mnc, odo, pwrcut, shock)
    SELECT
        dt,
        ST_Transform(ST_SetSRID(ST_MakePoint(lng, lat),4326), 31258) as geom,
        altitude,
        angle,
        speed,
        CAST(split_part(split_part(params, ',',1), '=', 2) AS INTEGER) as acc,
        CAST(split_part(split_part(params, ',',2), '=', 2) AS INTEGER) as arm,
        CAST(split_part(split_part(params, ',',3), '=', 2) AS INTEGER) as
batl,
        CAST(split_part(split_part(params, ',',4), '=', 2) AS INTEGER) as
cellid,
        CAST(split_part(split_part(params, ',',5), '=', 2) AS INTEGER) as
gpslev,
        CAST(split_part(split_part(params, ',',6), '=', 2) AS INTEGER) as
gsmlev,
        CAST(split_part(split_part(params, ',',7), '=', 2) AS INTEGER) as lac,
        CAST(split_part(split_part(params, ',',8), '=', 2) AS VARCHAR(50)) as
mcc,
```

```
        CAST(split_part(split_part(params, ',',9), '=', 2) AS INTEGER) as mnc,
        CAST(split_part(split_part(params, ',',10), '=', 2) AS INTEGER) as
odo,
        CAST(split_part(split_part(params, ',',11), '=', 2) AS INTEGER) as
pwrcut,
        CAST(split_part(split_part(params, ',',12), '=', 2) AS INTEGER) as
shock
    FROM raw_data.{tablename};

    UPDATE processed.{tablename}
    SET
        device_id='{attributes["device_id"]}',
        cowname='{attributes["cowname"]}',
        source='{attributes["source"]}',
        tracker='{attributes["tracker"]}';

    INSERT INTO processed.vierkaser_daten_gesammelt (dt, geom, alt, gpslev,
gsmlev, mcc, mnc, course, speed, batl, odo, device_id, cowname, source,
tracker)
    SELECT
        dt,
        geom,
        altitude,
        gpslev,
        gsmlev,
        mcc,
        mnc,
        angle,
        speed,
        batl,
        odo,
        device_id,
        cowname,
        source,
        tracker
    FROM processed.{tablename};
    """
    save_sql(f"{tablename}.sql", sqlstring)
    all_sql_content.append(sqlstring)

def process_bene_csv(name, filepath):
    tablename = name
    datasource = filepath
    attributes = extract_attributes(name)
    sqlstring = f"""
    create table raw_data.{tablename} (
        sat int,
        hdop double precision,
```

```
        lat double precision,
        lon double precision,
        fix int,
        date date,
        time time,
        age int,
        course double precision,
        speed double precision,
        card varchar(10),
        unknown double precision
        );

    COPY raw_data.{tablename} (sat, hdop, lat, lon, fix, date, time, age,
course, speed, card, unknown)
    FROM '{datasource}'
    DELIMITER ';'
    CSV HEADER;

    create table processed.{tablename} (
        sat int,
        hdop double precision,
        geom geometry(Point, 31258),
        fix int,
        dt timestamp,
        age int,
        course double precision,
        speed double precision,
        card varchar(10),
        device_id varchar(20),
        cowname varchar(10),
        source varchar(10),
        tracker varchar(10)
        );

    INSERT INTO processed.{tablename} (sat, hdop, geom, fix, dt, age, course,
speed, card)
        SELECT
        sat,
        hdop,
        ST_Transform(ST_SetSRID(ST_MakePoint(lon, lat),4326), 31258) as geom,
        fix,
        (date + time)::timestamp,
        age,
        course,
        speed,
        card
        FROM raw_data.{tablename};
```

```
    UPDATE processed.{tablename}
    SET
        device_id='{attributes["device_id"]}',
        cowname='{attributes["cowname"]}',
        source='{attributes["source"]}',
        tracker='{attributes["tracker"]}';

    INSERT INTO processed.vierkaser_daten_gesammelt (dt, geom, course, speed,
fix, hdop, age, device_id, cowname, source, tracker)
    SELECT
        dt,
        geom,
        course,
        speed,
        fix,
        hdop,
        age,
        device_id,
        cowname,
        source,
        tracker
    FROM processed.{tablename};
    """
    save_sql(f"{tablename}.sql", sqlstring)
    all_sql_content.append(sqlstring)

#Function for processing .gpx files
def process_gpx(name):
    tablename = name
    attributes = extract_attributes(name)
    sqlstring = f"""
    create table processed.{tablename} (
        dt timestamp,
        geom geometry(PointZ, 31258),
        course double precision,
        speed double precision,
        fix varchar(50),
        hdop double precision,
        vdop double precision,
        pdop double precision,
        age double precision,
        device_id varchar(20),
        cowname varchar(10),
        source varchar(10),
        tracker varchar(10)
    );
```

```
    INSERT INTO processed.{tablename} (dt, geom, course, speed, fix, hdop,
vdop, pdop, age)
        SELECT
        time,
        ST_Transform(ST_SetSRID(geom,4326), 31258) as geom,
        course,
        speed,
        fix,
        hdop,
        vdop,
        pdop,
        ageofdgpsdata
    FROM raw_data.{tablename};

    UPDATE processed.{tablename}
    SET
        device_id='{attributes["device_id"]}',
        cowname='{attributes["cowname"]}',
        source='{attributes["source"]}',
        tracker='{attributes["tracker"]}';

    INSERT INTO processed.vierkaser_daten_gesammelt (dt, geom, course, speed,
fix, hdop, age, device_id, cowname, source, tracker)
    SELECT
        dt,
        geom,
        course,
        speed,
        fix,
        hdop,
        age,
        device_id,
        cowname,
        source,
        tracker
    FROM processed.{tablename};
    """
    save_sql(f"{tablename}.sql", sqlstring)
    all_sql_content.append(sqlstring)

#Function for processing .gpkg files
def process_gpkg(name):
    tablename = name
    attributes = extract_attributes(name)
    sqlstring = f"""
    ALTER TABLE raw_data.{tablename}
    ALTER COLUMN time TYPE time
    USING time::time;
```

```
    create table processed.{tablename} (
        sat int,
        hdop double precision,
        geom geometry(Point, 31258),
        fix int,
        dt timestamp,
        age int,
        alt double precision,
        course double precision,
        speed double precision,
        card varchar(10),
        device_id varchar(20),
        cowname varchar(10),
        source varchar(10),
        tracker varchar(10)
    );

    INSERT INTO processed.{tablename} (sat, hdop, geom, fix, dt, age, alt,
course, speed, card)
        SELECT
        sat,
        hdop,
        ST_Transform(ST_SetSRID(ST_MakePoint(lon, lat),4326), 31258) as geom,
        fix,
        (date + time)::timestamp,
        age,
        alt,
        course,
        speed,
        card
        FROM raw_data.{tablename};

    UPDATE processed.{tablename}
    SET
        device_id='{attributes["device_id"]}',
        cowname='{attributes["cowname"]}',
        source='{attributes["source"]}',
        tracker='{attributes["tracker"]}';

    INSERT INTO processed.vierkaser_daten_gesammelt (dt, geom, hdop, fix, age,
course, speed, card, sat, device_id, cowname, source, tracker)
    SELECT
        dt,
        geom,
        hdop,
        fix,
        age,
```

```python
        course,
        speed,
        card,
        sat,
        device_id,
        cowname,
        source,
        tracker
    FROM processed.{tablename};
    """
    save_sql(f"{tablename}.sql", sqlstring)
    all_sql_content.append(sqlstring)

#Combines all written .sql files into one big .sql file to execute for your
postgres database
def combine_sql(allcontent):
    all_sql = "\n\n".join(allcontent)
    sql_filename = rf"{output_folder}\all_sql_combined.sql"
    with open(sql_filename, "w", encoding="utf-8") as sql_file:
        sql_file.write(all_sql)

#Re-usable function for saving sql files
def save_sql(sql_name, sql_content):
    sql_filename = rf"{output_folder}\{sql_name}"
    with open(sql_filename, "w", encoding="utf-8") as sql_file:
        sql_file.write(sql_content)

#Re-usable function for extracting attributes from name string
def extract_attributes(name):
    parts = name.split("_")
    return {
        "tracker": parts[0],
        "device_id": parts[1],
        "year": parts[2],
        "source": parts[3],
        "cowname": parts[4]
    }

main_script()
```