

Norges miljø- og  
biovitenskapelige  
universitet

Masteroppgave 2021 30 stp  
Realtek

# **GPS-basert sammenstilling av saueflokks oppførsel på inn- og utmark ved bruk av geografiske informasjonssystemer og maskinlæring**

GPS-Based Comparison of Free-Range and  
Pasture-Raised Sheep Herd Behaviour Through  
Use of GIS and Machine Learning

Erik Sand  
Geomatikk





---

## Sammendrag

Å ha sau på beite er en tradisjonsrik driftsform av sauehold i Norge. Driften har vært observert og forsket på i mange år, og etter hvert som GPS-teknologien har utviklet seg, har man begynt å bruke denne for å tolke sauens romlige og temporale oppførsel på beite. Metodene som benyttes i dag, brukes i hovedsak for å se på aktiv og inaktiv tid, eller for å se på beitebruk. Dette kan man kombinere, og man kan da se på hvilke typer beite som brukes ved inaktiv tid, og hvilke som brukes ved aktiv tid. Formålet med denne oppgaven er nettopp dette, men samtidig vurdere om man ser den samme oppførselen på inn- og utmark, og på tvers av beiteområder.

Sporingsdata fra sau er hentet fra tre ulike områder, ett innmarksbeite i Hol, og to utmarksbeiter i Nord-Østerdalen. De to beiten i Nord-Østerdalen har store forskjeller i beitekvalitet og tettheten av sau som beiter der. Datasettene er visualisert ved bruk av QGIS, og den temporale fordelingen er visualisert ved ArcGIS Pro. Rådatabehandling og bevegelsesparametere er beregnet ved bruk av Python.

Tre teknikker for å finne en tegn til normaloppførsel er presentert. To av disse er tetthetsbasert og en er basert på observert temporal fordeling på beiteområder. De to tetthetsbaserte metodene ser på heholdsvis avstand mellom målinger, og tettheten hos målingene. Den temporale fordelingen er så predikert med fem forskjellige vinklinger ved bruk av maskinlæringsalgoritmen `RandomForestRegressor`.

Hovedfunn ved tetthetsbasert analyse er at hvileområder hos saueflokker har flere likhetstrekk, både på inn- og utmark. Her kan nevnes områdets helning og at beitekvaliteten er generelt lavere. I tillegg kan man også se at avstanden en sau beveger seg vekk fra tidligere målte posisjoner kan brukes til å finne unormal oppførsel. Temporal fordeling viser at man kan forklare opptil 75% av beitetypfordelingen ved bruk av oppmålt beitegrunnlag. Ved sammenligning fordelingen i andre beiteområder kan man forklare oppimot 16% mer av fordelingen. `RandomForestRegressor` forklarte altså 91% av variasjonen i fordeling på beite.

---

## Abstract

Sheep husbandry in rangeland is an ancient brand of animal husbandry in Norway. Sheep grazing has been observed and researched for years, but as GPS technology has evolved, this is a more frequent way of interpreting the sheep's spatial and temporal behavior when grazing. The methods used today are mainly used for finding active and inactive time, or to visualize grazing locations. These methods could be combined. Thus visualizing which kind of pastures are used for inactive time, and which kinds are used for active time. The purpose of this thesis is just this, and at the same time consider whether the experienced behaviour is observed by both free-range and pasture-raised sheep.

Tracing data from sheep has been obtained from three different areas, one of these are measured on pasture-raised sheep in Hol, and the two others are measured by free-range sheep in Nord-Østerdalen. The two areas in Nord-Østerdalen are vastly different by looking at grazing quality, and the number of sheep in the area. The datasets are visualized using QGIS, and the temporal distribution is visualized by ArcGIS Pro. Raw-data processing and calculation of motion parameters are done using Python.

Three techniques for finding signs of normal behavior are presented. Two of these are density-based and one is based on observed temporal distribution by grazing areas. The density-based techniques uses distance between measured points, and the observed point density respectively. The temporal distribution is predicted through use of five different methods by the machinelearning algorithm RandomForestRegressor.

The main findings of the density based methods is that resting areas in the sheep herd contains several similarities, both free-range and pasture wise. Both grazing quality is observed to be lower, and the slope appears to be the same. In addition the distance a sheep moves away from previously measured positions can be used to derive abnormal behaviour. Temporal distribution explains that graze-type distribution correlate as much as 75% when regarding sheep distribution. When comparing other grazing areas, this increases by as much as 16%. RandomForestRegressor thus explains 91% of the variation in grazing distribution.

---

## Forord

Masteroppgaven markerer slutten på mine fem år ved NMBU. Det har vært en fantastisk tid, preget av mye hardt arbeid, mye moro, og opplevelser jeg vil huske for livet. Vil med dette rette en stor takk til Mannskoret Over Rævne, Trøndernes Fagforening og Samfunnet i Ås for å ha BEsørget en så rik studiehverdag.

Muligheten for å knytte geomatikk og landbruk sammen på masternivå fanget min interesse, og oppgaven ble fort spikret etter en introduksjon til dette emnet. Utfordringene med å finne nye metoder for analyse har vært krevende, men jeg har også lært enormt mye.

En stor takk rettes til veileder Seyed Hossein Chavoshi for gode innspill, råd og veiledning, og for å ha hjulpet meg med anskaffing av data, og anbefalinger av artikler.

Jeg vil også takke Øystein Holand og Nicolai Hermann Jørgensen for datasett og innspill rundt disse. En ekstra takk må rettes til Nicolai for korrekturlesing og konstruktiv diskusjon rundt programmering.

Ønsker samtidig å rette en takk til kollektivet ved Ulrik, Johan, Ane og Simon, for kaffe på kanna og fellesmiddag gjennom hele våren. Det har vært til stor glede i en ellers lite sosial hverdag.

Vil også rette en stor takk til Ingrid Tveite for å ha stilt opp, og sørget for at masteren ble ferdigstilt og korrekturlest.

Til slutt vil jeg rette en stor takk til hovedveileder Håvard Tveite, som har stått på, og bistått med masteroppgaven frem til siste slutt. Jeg er ufattelig imponert og ydmyk over den innsatsen og det engasjementet du la ned i en så vanskelig tid. Jeg vil takke deg for alle innspill, korrekturlesinger og hyggelige telefonsamtaler vi har hatt i samarbeidet.

*Norges miljø- og biovitenskaplige universitet*

*Ås, 31.05.2021*

*Erik Sand*

# Innhold

<b>Figurer</b>	<b>vii</b>
<b>Tabeller</b>	<b>x</b>
<b>1 Innledning</b>	<b>1</b>
1.1 Bakgrunn . . . . .	1
1.2 Problemstilling . . . . .	2
1.3 Aktualitet . . . . .	2
1.3.1 Sau . . . . .	2
1.3.2 Tidligere studier . . . . .	3
1.3.3 Områdetyper . . . . .	4
1.4 Oppsett og struktur . . . . .	4
<b>2 Teori</b>	<b>5</b>
2.1 GNSS-systemene . . . . .	5
2.1.1 Introduksjon til GNSS . . . . .	5
2.1.2 Komponenter i GPS-systemet . . . . .	5
2.1.3 Referansesystem og posisjonfastsetting . . . . .	6
2.2 Sporingssystemer . . . . .	7
2.2.1 Followit . . . . .	8
2.3 Tetthetsbasert mål av punktfordeling . . . . .	8
2.3.1 Punkttetthetsanalyse . . . . .	8
2.4 Avstandsbasert mål av punktfordeling . . . . .	9
2.4.1 G-funksjon . . . . .	9
2.4.2 F-funksjon . . . . .	9
2.4.3 K-funksjon . . . . .	10
2.5 Regresjonstrær . . . . .	10
2.5.1 Over- og undertilpasning . . . . .	11
2.5.2 Modellens godhet . . . . .	12

---

2.6	ModelBuilder . . . . .	12
<b>3</b>	<b>Data og Analyse</b>	<b>13</b>
3.1	Testområder . . . . .	13
3.1.1	Hol . . . . .	14
3.1.2	Nord-Østerdalen . . . . .	15
3.1.3	Forskjell på områder . . . . .	16
3.2	Preprossesering av datasett . . . . .	16
3.2.1	Programvare . . . . .	16
3.2.2	GIS-verktøy . . . . .	17
3.2.3	GPS-data . . . . .	17
3.2.4	Innlesing . . . . .	17
3.2.5	Feilmålinger . . . . .	18
3.2.6	Grunnlagsverdier . . . . .	18
3.3	Romlig analyse . . . . .	21
3.3.1	Sporing . . . . .	21
3.3.2	Flokksporing . . . . .	22
3.3.3	Avstandsbasert mål av punktfordeling . . . . .	22
3.3.4	Punkttetthetsanalyse . . . . .	22
3.4	Grunnlagsanalyse . . . . .	23
3.4.1	Sortering av grunnlag . . . . .	23
3.4.2	Randomforest . . . . .	24
<b>4</b>	<b>Resultater</b>	<b>25</b>
4.1	Observasjoner . . . . .	25
4.1.1	Romlig fordeling . . . . .	25
4.1.2	Ukentlig fordeling . . . . .	27
4.1.3	Visualisering av fordeling . . . . .	28
4.2	Avstandsbasert mål av punktfordeling . . . . .	31
4.2.1	F-funksjon . . . . .	31
4.2.2	G- og K-funksjon . . . . .	32
4.2.3	G-funksjon enkeltsauer . . . . .	33
4.3	Hotspot . . . . .	34
4.3.1	Spekedal . . . . .	34
4.3.2	Bratthøa . . . . .	38
4.3.3	Hol . . . . .	39

---

---

4.4	Randomforest . . . . .	40
<b>5</b>	<b>Diskusjon</b>	<b>43</b>
5.1	Observasjoner . . . . .	43
5.2	Avstandsbasert mål av punktfordeling . . . . .	43
5.3	Hotspot- og grunnlagsanalyse . . . . .	44
5.4	Randomforest . . . . .	44
5.5	Datagrunnlag . . . . .	45
5.5.1	Problematikk datainnlesning . . . . .	45
5.5.2	Målemetode . . . . .	46
<b>6</b>	<b>Konklusjoner og anbefalinger</b>	<b>49</b>
6.1	Konklusjon . . . . .	49
6.2	Anbefalinger . . . . .	50
	<b>Bibliografi</b>	<b>52</b>
	Appendix . . . . .	55

# Figurer

1.1	Dyr drept av rovdyr[4] . . . . .	1
2.1	Satellittens referansesystem[9] . . . . .	6
2.2	Overgang fra geosentrisk til geodetisk referansesystem[10] . . . . .	7
2.3	Radiobjella Tellus fra Followit[21] . . . . .	8
2.4	Sammenstilling av G- og F-funksjonene[13] . . . . .	9
2.5	K-funksjon for to typer mønstre [14] . . . . .	10
2.6	Regresjonstre [53] . . . . .	11
2.7	Undertilpassning, godt kompromi og overtilpassning[40] . . . . .	11
2.8	Eksempel på model fra ArcGIS ModelBuilder . . . . .	12
3.1	Kart over Norge og hvor studieområdene befinner seg. . . . .	13
3.2	Hol beiteområde (60°68'N, 8°01'E). . . . .	14
3.3	Studieområdene Spekedalen og Bratthøa i Nord-Østerdalen . . . . .	15
3.4	Spekedal beiteområde . . . . .	18
3.5	Bratthøa beiteområde . . . . .	19
3.6	Hol beiteområde . . . . .	20
3.7	“Join attributes by location” i QGIS . . . . .	20
3.8	Parametere for beskrivelse av bevegelse[8] . . . . .	21
3.9	Meny for plotting av linjediagram i ArcGIS . . . . .	23
4.1	Sauenes fordeling i Hol beiteområde . . . . .	25
4.2	Sauenes fordeling i Spekedal beiteområde . . . . .	26
4.3	Sauenes fordeling i Bratthøa beiteområde . . . . .	26
4.4	Antall punkter per beitetype Hol 2012 . . . . .	27
4.5	Antall punkter per beitetype i Bratthøa . . . . .	27
4.6	Antall punkter per beitetype i Spekedalen . . . . .	28
4.7	Fordeling Hol . . . . .	29
4.8	Fordeling Bratthøa . . . . .	29

---

4.9	Fordeling Spekedal . . . . .	29
4.10	Fordeling Hol . . . . .	30
4.11	Fordeling Bratthøa . . . . .	30
4.12	Fordeling Spekedal . . . . .	30
4.13	F-funksjon Bratthøa 2014 . . . . .	31
4.14	F-funksjon Bratthøa 2013 . . . . .	31
4.15	F-funksjon Spekedal 2013 . . . . .	31
4.16	F-funksjon Hol 2012 . . . . .	31
4.17	G-funksjon Spekedal 2013 . . . . .	32
4.18	K-funksjon Spekedal 2013 . . . . .	32
4.19	G-funksjon Hol 2012 . . . . .	32
4.20	K-funksjon Hol 2012 . . . . .	32
4.21	G-funksjon på sau 142 000 Spekedal . . . . .	33
4.22	G-funksjon på sau 142 06 Bratthøa . . . . .	33
4.23	G-funksjon på sau 13 Hol i tett innhegning . . . . .	33
4.24	G-funksjon på sau 4 Hol i spredt innhegning . . . . .	33
4.25	Hotspot-analyse Spekedal . . . . .	34
4.26	Inspeksjon av Gloføken . . . . .	35
4.27	Inspeksjon av Spekedalseter . . . . .	36
4.28	Inspeksjon av Storlægda . . . . .	37
4.29	Hotspot-analyse Bratthøa . . . . .	38
4.30	Hotspot-analyse Hol 2014 og 2015 . . . . .	39
4.31	Hotspot-analyse Hol 2012 og 2013 . . . . .	39
4.32	Inspeksjon tetthet Hol . . . . .	40
4.33	Flyfoto av Hol[3] . . . . .	40
4.34	Residualplott metode en . . . . .	40
4.35	Residualplott med innmark som testdatasett . . . . .	41
4.36	Residualplott med utmark som testdatasett . . . . .	41
4.37	Residualplott estimering av Bratthøa 2014 ved bruk av Bratthøa 2013 . . . . .	42
4.38	Residualplott estimering av Bratthøa 2014 ved bruk av Spekedalen, Hol og Bratthøa 2013 . . . . .	42
5.1	Endring i samplingfrekvens hos GPS-mottakeren . . . . .	46
5.2	Punktmønsteret fra målingene hos de 18 sauene på Hol i 2015 . . . . .	48
5.3	Inspeksjon av vandrerute på Einsenuten . . . . .	48



---

Alle kart er laget med bakgrunnskartet Topografisk Norgeskart fra: <http://openwms.statkart.no/skwms1/wms.toporaster4?version=1.3.0&>

# Tabeller

3.1	Samletabell studieområder . . . . .	16
3.2	Oppsett ukentlig sortering . . . . .	24
5.1	Resultater Randomforest . . . . .	45
5.2	Estimert batterikapasitet ved ulike frekvenser . . . . .	46

# 1. Innledning

## 1.1 Bakgrunn

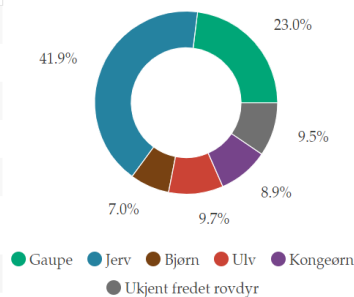
I Norge har man drevet sauedrift i flere tusen år, og det er fortsatt en av våre største primærnæringer. Sau (*Ovis Aries*) er en drøvtygger, og døgnsyklusen på beite består i hovedsak av beiting, drøvtygging og hvile. Den vanligste årlige syklusen for sauehold er at sauene går inne i fjøset om vinteren, og på våren og sommeren er de på henholdsvis inn- og utmarksbeite. Innmarksbeite er en type beite som bærer preg av at det er i nærheten av gården, og at det er inngjerdet. Utmarksbeite derimot ligger lengre unna, og er beite i vill vegetasjon, altså skog og fjellterreng som ikke blir kultivert [27].

Antallet sauer på utmarksbeite i Norge har de siste 20 årene vært ca. to millioner [51]. Dette tallet er mye høyere enn det som kalles vinterfora sau, altså de sauene man har om vinteren, fordi lamminga planlegges til like før beitesesongen. Når de slippes ut på beite, er det lite overvåking, spesielt på utmarksbeite, og derfor har man de siste årene utstyrt søyer med GPS-mottakere, for å finne de ved innsanking til høsten.

Et av de større problemene ved sauedrift er rovdyr. Mange lam blir tatt av kongeørn (*Aquila Chrysaetos*), men det er de fire store rovdyrene som nå er i tilbakevekst i Norge som tar flest. Disse er brunbjørn (*Ursus Arctos*), ulv (*Canis Lupus*), gaupe (*Lynx Lynx*) og jerv (*Gulo Gulo*) [26]. Økningen av disse artene påvirker omgivelsene de nå har funnet sitt habitat i. Den har også fått sitt inntog i media, med den mye omdiskuterte ulvesonen.

Antall dyr erstattet og skadevolder

Art	Sau	Lam	Totalt erstattet	Andel av totalt tap
Gaupe	509	2 847	3 356	23.0 %
Jerv	705	5 410	6 115	41.9 %
Bjørn	534	483	1 017	7.0 %
Ulv	323	1 086	1 409	9.7 %
Kongeørn	6	1 297	1 303	8.9 %
Ukjent fredet rovdyr	268	1 118	1 386	9.5 %
Totalt	2 345	12 241	14 586	100.0 %



Figur 1.1: Dyr drept av rovdyr[4]

Av grafen over ser vi at jerven er den klart største trusselen mot sau. Gaupe utgjør den nest største trusselen, og ulv, kongeørn og bjørn er ganske jevne når det kommer til antall. Bjørn er en større trussel for voksne sauer, og kongeørn for lam.

---

## 1.2 Problemstilling

I denne oppgaven ønsker jeg å besvare følgende problemstilling:

*Kan man finne en standardisert modell for saueflokkers bevegelsesmønstre ved bruk av maskinlæring og geografiske informasjonssystemer?*

For å besvare problemstillingen har jeg også knyttet tre spørsmål til dette:

*Hvilke indre parametere hos en saueflokk viser tydelige tegn til en normalttilstand?*

*Hvilke ytre tegn observerer man hos en saueflokk under normale omstendigheter?*

*Er oppførselen synlig både på innmark og utmark?*

Med indre parametere menes forholdet innad i saueflokken, som avstand til hverandre og hvordan flokken forflytter seg. De indre parametere beregnes ved analyse i QGIS og ArcGIS, og ved bruk av maskinlæring kan jeg finne parametrene med høyest og lavest signifikansnivå. Jeg ønsker også å se på hvilke variabler som har minst variasjon, altså de som er mest stabile, og vurdere hvorvidt disse vil endre seg ved forstyrrelser (for eksempel rovdyrangrep).

Ytre tegn er blant annet beitetype og helning. Ved inntasting av områdedata skal man kunne si noe om hvordan sauene beiter under normale forhold. Beveger sauene seg på unormale områder, kan man ved en varslingsfunksjon, si at noe avviker fra normalen. Slike situasjoner kan være ulveangrep, skade og sykdom.

I oppgaven benyttes datasett fra både inn- og utmark. Observasjon av lik atferd på begge typer beite vil da høyst sannsynlig kunne være beskrivende for normal oppførsel hos sauer på beite. Er oppførselen tydelig, men ikke lik på beiteslagene, viser det tegn til at man må behandle inn- og utmarksbeite etter dette. Altså lage to modeller.

## 1.3 Aktualitet

### 1.3.1 Sau

Sauedrift som driftsform bærer typiske preg av å være uovervåket. Det vil si at sauene slippes ut på beite, og har tilsyn noen få ganger, om noen i det hele tatt i løpet av beitesesongen. Derfor vil teknologi som sier hvordan de kontinuerlig oppfører seg bidra mye til nyttig lærdom innenfor denne gamle og tradisjonsrike formen for dyrehold. Radiobjeller er nettopp dette, en godt etablert teknologi som kan benyttes til sporing av dyr på beite. Selv om sporing av et dyrs GPS-spor kan skje ganske så i sanntid, er det fortsatt en teknikk som ser på data etter at det har skjedd. Sauebønder har i lang tid kunnet forutse noe av flokkens bevegelse og posisjonering gjennom erfaringer fra tidligere år. I hovedsak av direkte observasjon, men også ved å se på spor i terrenget etter årevis med saueutrakk og beiting. Ved å samle sammen denne informasjonen fra forskjellige beiteområder, kan man med dagens maskinlæring lage en generell modell som beskriver saueflokkes oppførsel på tvers av beiter. Viser det seg at oppførselen er signifikant forskjellig kan man lage flere spesifikke modeller for hvert beite.

---

Animalia[5], et norsk fag- og utviklingsorganisasjon for kjøtt- og eggproduksjon, beskriver noe av sauers forventede atferd:

Sauer...

- motiveres til å løpe når de ser andre sauer som løper
- vil stoppe opp når de ser sauer som beveger seg motsatt retning
- foretrekker å gå fra mørke til lyse områder
- velger heller å bevege seg oppover enn nedover
- vil bevege seg mer villig mot åpent område enn mot en “blind ende”

Dette er faktorer som gir rom til å tro at bevegelsesmønsteret i hele eller deler av flokken endres ved forstyrrelser. Samtidig ser man at sauene har tydelige tendenser i sitt normale bevegelsesmønster.

### 1.3.2 Tidligere studier

Man har studert dyr på beite i flere tiår og mange metoder er forsøkt benyttet for å beskrive oppførselen deres. Den gode gamle metoden består av direkte observasjon, og her har man for eksempel observert at en sauflokk med større tetthet vil ha et høyere aktivitetsnivå per sau enn flokker med lavere tetthet. Dette fordi sauene forstyrrer hverandre, og konkurrerer om matfatet [33]. Studien gjort av Mobæk et al. (2012) bruker GPS-data ved at observatøren først noterte oppførselen, for så å bevege seg til punktet og måle med en håndholdt GPS. Dagens metodikk baserer seg på å feste en GPS-mottaker, et akselerometer, gravimeter eller lignende på dyret for å måle data direkte. En kombinasjon av GPS-data og hodets vinkel i forhold til kroppen er brukt for å kategorisere kyr. Ved bruk av K-means algoritmen, en såkalt ustyrt læring, ble bevegelsesmønsteret deres klassifisert i to grupper. Algoritmen fant at de to letteste klassene å skille på var aktiv og inaktiv tid [49].

Man har også forsøkt å klassifisere oppførsel ved hjelp av akselerometer. Her har man forsøkt å finne ut av positurene til dyr, og ut ifra det tolket hva slags aktivitetsmønster de har. Barwick et al. (2020)[7] monterte akselerometer på bein, ører og nakke, og kunne klassifisere beiting, ståing, gåing og ligging med en nøyaktighet på ca. 48%-94% for benmåleren, 67%-88% for nakkemåleren, og 86%-95% for øremåleren. Riaboff et al. (2020)[47] testet forskjellige algoritmer på sine akselerometerdata. Ved bruk av eXtreme Gradient Boost (XGB) kunne oppførsel forklares med en nøyaktighet på 98%, og ved bruk av Random Forest kunne 97% forklares. Denne forskningen ble spilt videre på, og kombinert med GPS-data. Dette ble så benyttet til å modellere hvor mye tid kyr brukte i de forskjellige beitetypene, og hvordan de brukte tiden sin der (beiting, hvile osv.)[48]. Denne typen analyse retter seg dog mer mot innmarksbeite, for å effektivisere bruken av beitet og beitetypene det har.

Man finner også studier som kun benytter GPS-data. Heriblant har man belyst enkeltsauers bevegelsesmønstre. Ved analyse av sju GPS-dager med korresponderende observert bevegelsesdata, fikk de en gjennomsnittlig nøyaktighet på  $80.75\% \pm 5.9\%$  (SE), dette ved bruk av høyfrekvente data[56]. Lavfrekvente data er også benyttet, og en studie gjort av Ungar et al. (2005)[57] gjør nettopp dette. Modellen forklarer 85% av variasjonen i datasettet, og dette viser at muligheten for tolkning av oppførsel ved lav sampling rate er mulig. De positive sidene ved dette er at batteriene i GPSen får lengre levetid, og kan gi verdifull informasjon gjennom hele beitetiden. Tar man disse to studiene i betraktning, ser det ut som om lavfrekvente data er bedre på å klassifisere bevegelse. Faktum er at den høyfrekvente modellen klassifiserer ca. 80% riktig av 39 123 punkter mot den lavfrekvente modellen som forklarer 85% av 195 målinger.

Metodikk for å beregne rovdryfare har også blitt presentert. Her ved at man har sett på antallet GPS-punkter fra kyr observert på samme område, og brukt ulveobservasjoner i nærheten. Studiet er gjort av Clark et al. (2020)[12], og metoden de presenterer for å beregne rovdryfare baserer seg på en modell presentert av Ausband et al. (2010)[6]. Denne modellen sier noe om ulvers bevegelsesmønster, og man kan avlede sannsynligheten for hvor ulven beveger seg i forhold

---

til observasjonene. Clark et al. bruker disse sannsynlighetene og sammenligner med tettheten og posisjonen til kyrene. Formålet med studiet er å beregne sannsynlighetsområder for at ulv og storfe møtes, men sier ikke noe om angrepssjansen ved møte.

Man kan også se på posisjonen til sauene i forhold til hverandre, og dermed gjøre en flokkanalyse. Dette kan være nyttig for blant annet å finne de mest brukte beiteområdene til saueflokker [52]. Bevegelsesmønstre i forhold til simulert rovdyrangrep er også gjort før. Manning et al. (2014)[32] har brukt hund til å jage sau på samme måte som en ulv ville gjort, og fant at romlig fordeling ikke så ut til å endre seg. Studien fant på den annen side at farten var signifikant høyere under angrepet enn før og etter.

### 1.3.3 Områdetyper

Når man ser på posisjonsdata, er det ikke bare dyrenes posisjon man kan se på. Man kan også se på områdedataene, og dyrenes forhold til disse. Riaboff et al.(2020)[48] forsøkte en tilnærming til dette problemet ved å dele inn beiteområdet i ruter på 8x8 meter for enkel definering av områder. Faktorer studien bruker er helning, jordfuktighet og botaniske karaktertrekk, og strukturelle karaktertrekk. Studien varte over fem dager og fant at kyrne beiter nært trær og drikkestasjoner om morgenen, og om kvelden beitet de nært melkestasjonen. Anbefalinger til videre arbeid er å bruke GPS med bedre batterikapasitet, og akselerometer og GPS-frekvenser som er lavere, for å få mer data, og for kontinuerlig overvåkning.

Det å forstå hvordan dyr beveger seg er mye utforsket. Større universiteter har derfor gått sammen om en felles offentlig nettbasert løsning, hvor man kan dele data og studier. Denne løsningen heter Movebank, og brukes av tusenvis av forskere og vilt-forvaltere [58]. Her kan man kombinere sporingsdata fra dyr med blant annet værddata innhentet gjennom samarbeid med NASA.

Studier ved NMBU har også sett på effekten av valg av beitekvalitet. Mobæk et al. (2009)[34] fant at sauens valg av habitat avhenger av årstid og sesong. En konklusjon fra studiet er at variasjonen i valg av habitat kan avhenge av plantenes utvikling i løpet av sommeren.

## 1.4 Oppsett og struktur

Denne oppgaven er delt inn i seks kapitler, hvor Innledningen gir en introduksjon til emnet, presenterer problemstillingen og presenterer tidligere forskning. Kapittel to presenterer teori bak GPS-system, algoritmer for romlig analyse og maskinlæringsalgoritmer, og beskriver også utstyret som er benyttet. Kapittelet Data og Analyse er et todelt kapittel, hvor kapittel 3.1 og 3.2 forklarer metodikken bak preprossesering av datasett og filformater, og beskriver første observasjoner. Kapittel 3.3 og 3.4 beskriver hvilke programvarer og metoder som er benyttet for besvaring av problemstillingen. Kapittel fire presenterer resultatene fra analysen, og kapittel fem diskuterer disse. Til slutt følger Konklusjon, som oppsummerer resultat og diskusjon, og besvarer problemstillingene. Her presenteres også forslag til videre forskning.

## 2. Teori

### 2.1 GNSS-systemene

#### 2.1.1 Introduksjon til GNSS

GNSS-teknologien (Global Navigation Satellite Systems) har siden 1970-tallet vokst frem til å bli en ledende navigasjonsteknologi. Opphavet kommer fra det amerikanske militæret, som med sin GPS (Global Positioning System) lagde det første fullstendige operative systemet. GPS har utviklet seg til å bli en sivil teknologi, noe som har medført et større bruksområde. Dette har økt etterspørselen etter bedre satellitteknologi, noe som i senere tid har gjort det svært lønnsomt å videreutvikle samt vedlikeholde dagens system. Et resultat av dette er at flere systemer enn det militære GPS-systemet har sett dagens lys. GLONASS er det russiske militærets svar på GPS, dette systemet var først operativt den 14. desember 1995 [23]. Kina har også utviklet et eget system, BeiDou. Dette systemet har to versjoner, en åpen sivil versjon, og en for autoriserte brukere. Et regionalt system var ferdig 2013, og hele systemet i 2020 [30].

Et system som fortsatt er under utvikling heter Galileo og er etablert av den europeiske romorganisasjonen (ESA) og EU-kommisjonen. Dette er det siste av systemene, og startet med navigasjonstjenester i desember 2016. Per nå fungerer det som et supplement til GPS, da det ikke er fullstendig operativt selv før 2022 [24]. Flere land har de senere årene jobbet for sine sivile program som GLONASS, BeiDou og Galileo [11]. Tilgangen på denne teknologien har dermed ført til en enorm utvikling av nøyaktighet og størrelse på mottakeren, og igjen ført til en enorm produksjon av forskjellige typer mottakere med forskjellig bruksmål. Her kan blant annet treningsklokker og radiobjeller nevnes som eksempel. Ved at slik teknologi har blitt allemannseie, har det også ført til enorme mengder datasett med muligheter for analyse.

#### 2.1.2 Komponenter i GPS-systemet

GPS-systemet består i hovedsak av tre elementer: romsegmentet, bakkeselementet og brukersegmentet [11]. Disse tre segmentene beskriver hvilke typer teknologi og utstyr som benyttes i målingene.

Romsegmentet består av 24 satellitter i seks forskjellige omløpsbaner rundt jorda. Denne fordelingen er beregnet for å kunne gi 24-timers dekning på jorda fra  $80^{\circ}\text{N}$  til  $80^{\circ}\text{S}$ . Radius på omløpsbanen er på ca. 27 000 km og en rundetid er på en time og 58 minutter. For å skille på satellittene har de en egen ID, kalt PRN (PseudoRandom Noise), noe man finner igjen i rådataene hos en GPS-mottaker. Grunnet for satellittenes nøyaktighet er at de har fire nøyaktige klokker, av typen atomur[25].

Bakkeselementet, eller kontrollsegmentet består av overvåkningsstasjoner, som følger kontinuerlig med på satellittene og deres posisjoner. Dette er fastmonterte basestasjoner, med nøyke innmålte koordinater og avstander mellom hverandre. Den første av disse basene ble bygd i Colorado Springs og dette er fortsatt hovedkontrollstasjonen. GPS-systemet har stasjoner over hele jorda, og det finnes nå 20 bakkestasjoner. Stasjonene kan deles inn i monitorstasjoner, og antennestasjoner.

Monitorstasjonene overvåker satellittene, og rapporterer til kontrollstasjonen i Colorado. Denne sender så ut data til antennestasjonene, som laster opp forbedrede bane- og klokke-data minst en gang i døgnet per satellitt[25, 22].

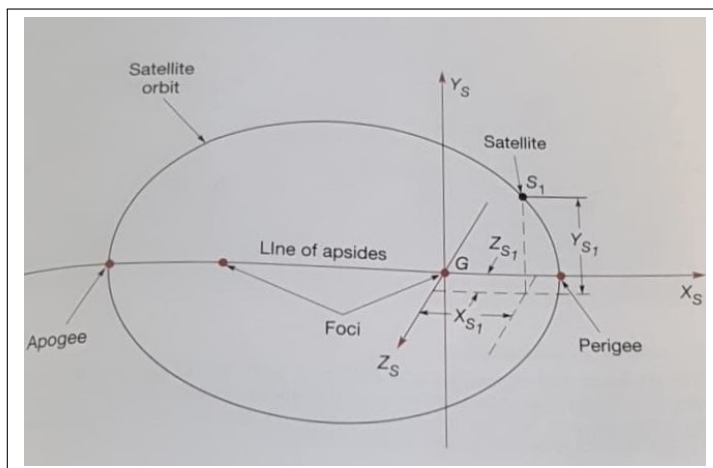
I motsetning til romsegmentet som kun består av satellitter, og kontrollsegmentet som kun består av kontrollstasjoner, så består brukerssegmentet av to kategorier. Disse kalles Standard Position Service (SPS) og Precise Positioning Service (PPS). SPS er den gratis og offentlige versjonen, og har en presisjon på 100 meter i horisontalt plan, og 156 meter i vertikalt plan med en 95% standardfeil. PPS har en presisjon på 18 meter horisontalt og 28 meter vertikalt. Forskjellen på disse to er at SPS bruker L1-signaler i hovedsak på frekvens 1575,42 MHz, men har i senere tid også tatt i bruk L2-signaler på 1227,60 MHz. PPS benytter begge frekvensene[11].

Når man bestemmer posisjonen trenger man signal fra fler satellitter enn en. Minstekravet er på tre satellitter, og da har man det man kaller en fix. Dette gir kun 2D-koordinater, altså uten høyde. For at man skal ha med høyde kreves kontakt med fire satellitter. Mottar man signal fra flere satellitter øker nøyaktigheten på posisjonsfastsettingen [16].

### 2.1.3 Referansesystem og posisjonsfastsetting

For å kunne bestemme brukerens posisjon, må man også kunne bestemme satellittens posisjon. Men for å posisjonsfeste noe må man feste det i forhold til noe, og har derav referansesystemer. Satellittens referansesystem er rom-relatert, og definerer et tredimensjonalt rektangulært system ut fra satellittens omløpsbaner. Dette transformeres så til et jordsentrisk koordinatsystem, som er relatert til jordens massesenter. Dette er ikke leselige koordinater for mennesker, derfor transformeres disse til vanlige geodetiske koordinater som lengde- og breddegrad og høyde.

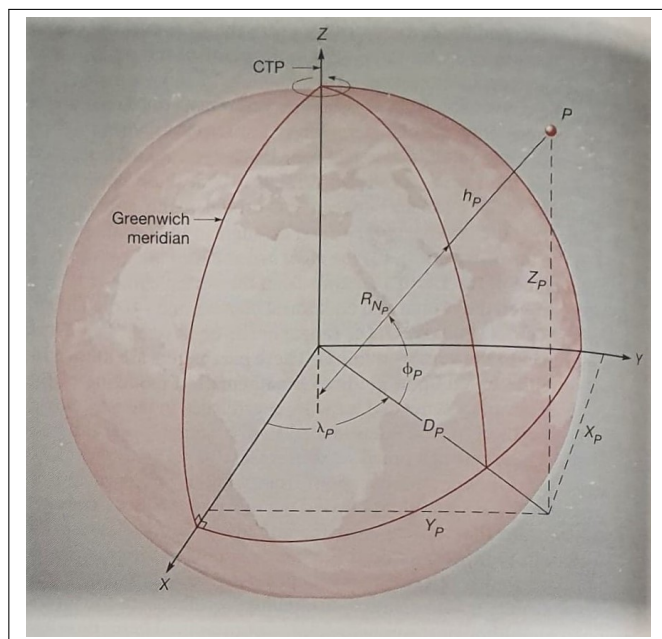
Satellittens referansesystem bygges på det at satellitten først er i omløp, så defineres dens bane stort sett av jordens gravitasjonskraft. Andre faktorer er sola og månens gravitasjonskraft, og solstråling. Disse andre faktorene varierer en del, og gjør at banen ikke blir helt ideel. Figur 2.1 viser den ideelle banen som kun tar hensyn til jordens gravitasjon, modellert med to fokuspunkter, nemlig jordens massesenter (G)[11]. Disse to punktene er bundet sammen av en linje som kalles apselinja. Man ser også at  $X_s$  og  $Y_s$  forklarer henholdsvis posisjonen på ellipsen, og  $Z_s$  beskriver avviket fra det gjennomsnittlige planet. Dette systemet er vanskelig å bruke for å forklare noe på jordas overflate, men ved at man vet satellittens omløpsparametere, kan man transformere til koordinatsystemer som gir mening for vår bruk.



Figur 2.1: Satellittens referansesystem[9]

Jordsentrisk koordinatsystem er noe lettere å forholde seg til, da dette bruker kun et punkt som jordens massesenter. Koordinatene her betegnes også av  $X_e$ ,  $Y_e$  og  $Z_e$ , og beskriver avstanden fra jordens sentrum i hver sin retning. XY-planet ligger på ekvator, hvor X-aksen går gjennom Greenwich-meridianen. Z-aksen ligger på polaksen eller rotasjonsaksen, og Y-aksen dannes av normalen til XZ-planet. For overgangen fra satellittens referansesystem til jordsentrisk, så trengs det fire rotasjonsparametere fra satellitten: (1) vinkelen mellom omløpsbaneplanet og jordens ekvatorielle plan, (2) vinkelen i omløpsbanen fra ekvator til apselinja, (3) vinkelen i ekvatorplanet fra punktet vernal equinox til linja mellom omløpsbanen og ekvatorplanet, og (4) vinkelen i ekvatorplanet fra Greenwich-meridianen til vernal equinox.





Figur 2.2: Overgang fra geosentrisk til geodetisk referansesystem[10]

Alle disse parameterne finnes for satellittene ved bruk av matematiske modeller for omløpsbanen deres. Fortsatt er koordinatene vanskelige å forholde seg til, fordi avstanden til jordsentrum gir høye verdier, og blir for teoretisk. Derfor må man transformere nok en gang, og denne gangen til et system som forholder seg til noe synlig, nemlig til jordoverflaten.

Overgangen fra jordsentrisk eller geosentrisk referansesystem til geodetisk referansesystem vises tydeligst av at man ikke lengre snakker om X-, Y- og Z-koordinater. Man snakker nå om lengde- og breddegrad, og høyde. Lengdegrader sier noe om hvor man befinner seg i forhold til Greenwichmeridianen, og breddegradene sier noe om hvor man er i forhold til ekvator. Man får altså forklart himmelretningene av to koordinater. I tillegg har man også høyden, som beregnes til høyde over havet. Denne typen re-

feransesystem har mange standarder, som bruker forskjellige formler for høydemodellering og ellipsoidmodellering. Men den mest fremtredene er WGS84, som også brukes i denne oppgaven.

## 2.2 Sporingssystemer

Metodikken for å følge dyrs oppførsel på beite er en teknologi som har vært under utvikling lenge. Per nå (2021) er det i hovedsak tre ledende teknikker. Akselerometer, GPS og direkte observasjon.

Akselerometer ser på endringene i fart, og gir detaljerte beskrivelser av bevegelsesmønstre. Fordelen er at man kan lagre dette i større datasett, og ha ganske hyppige målinger. Ulempen er at desto hyppigere, desto lavere batterikapasitet. I tillegg sier akselerometere ingenting om hvor et dyr befinner seg.

Direkte observasjon gir presise beskrivelser av oppførselen, men her begrenses kapasiteten når antallet sauer øker. Ved bruk av kamera og videoopptak kan man øke kapasiteten, men prosesseringen tar lang tid, da denne foregår manuelt.

GPS-sporing gir posisjon, som man kan avlede parametere som fart og retning av. Lagringskapasiteten er høy, og prosesseringen er rask. GPSer krever dog mer batterikapasitet enn akselerometer, noe som går på bekostning av frekvensen av målingene.

---

### 2.2.1 Followit

Alle data i denne studien er samlet inn ved bruk av GPSer fra Followit.

Followit AB er en ledende bedrift innenfor mobil og satellittpoisjonering av dyr, biler, transport, folk og eiendom. Løsningene brukes hovedsaklig av viltforskere, transport- og sikkerhetsindustrien, for husdyr, av jaktsamfunnet og av private til overvåkning av verdifull eiendom[20].

Followit har to typer radiobjeller til salgs. Pellego som er den billigste, og Tellus som er den dyrere versjonen. På Hol og i Nord-Østerdalen har søyene vært utstyrt med Tellus-modellen. Tellus veier 600 eller 800 gram avhengig av batteritype. Den har også en lagringskapasitet på opptil 80 000 posisjoner, og gir også informasjon som tid, høyde, temperatur, aktivitet og dødelighet. Man kan følge med på bevegelsesmønsteret på smarttelefon eller datamaskin, og der kan man også endre på instillingene når det gjelder frekvens. Batterikapasiteten er estimert til to år ved tolv posisjonsmålinger om dagen[21].



Figur 2.3: Radiobjella Tellus fra Followit[21]

Followit gir med andre ord mye data som også kan brukes i modelleringen av miljøet. Mottakerne i Hol ble satt til å finne posisjon hvert 30. minutt, og de i Nord-Østerdalen hvert 60. minutt.

## 2.3 Tetthetsbasert mål av punktfordeling

Punktfordeling kan måles ved første og andre grads egenskaper. Der andregrads egenskaper i punktfordelingen forteller noe om avstanden mellom punkter, så sier førstegrads egenskaper noe om tettheten. Funksjonene for mål av tetthet sier altså noe om hvor mange ganger man finner et tilfelle i studieområdet. Den enkleste formelen for dette er gitt ved:

$$\hat{\lambda} = \frac{n}{a} = \frac{(\#S \in A)}{a} \quad (2.1)$$

$\hat{\lambda}$  beskriver den estimerte intensiteten i området  $a$ , gitt ved antallet punkter  $n$ . Et av problemene ved denne typen estimering er at tettheten avhenger svært mye av hvordan man definerer studieområdet[15].

### 2.3.1 Punkttetthetsanalyse

Punkttetthetsanalyse studerer hvor mange punkt som havner innenfor en bufferavstand til et annet punkt. Bufferavstanden definerer en flate med radius lik bufferavstanden, og har en overflateverdi beregnet utifra nærhet til punktet. Er avstanden større enn bufferavstanden, blir overflateverdien lik null.

Slike overflateverdier beregnes for alle punkter, før disse legges sammen med verdiene for overflatene som overlapper hverandre. Dette sørger for en kontinuerlig overflate mellom nære punkter, og desto fler punkter i en klynge, desto høyere verdi får overflaten[15].

I denne typen analyse er det viktig å finne en representativ bufferavstand, da med for høy verdi vil tettheten bli for generell i hele studieområdet. Er den for liten, vil de minste grupperingene få sterkt fokus, og enkeltpunkter av interesse kan forsvinne.

## 2.4 Avstandsbasert mål av punktfordeling

Avstandsbasert måling ser på avstanden fra et fenomen til et annet, og i dette tilfellet fra en sau til en annen. Det finnes flere metoder for å finne statistiske mål for dette. De presentert i denne oppgaven heter G-, F- og K-funksjon.

### 2.4.1 G-funksjon

G-funksjon viser fordelingen av avstand til nærmeste nabo for hvert punkt [15]. Dette betyr at funksjonen vil gi tydelige utslag om dyrene beveger seg i par/grupper, selv om de er fordelt på et større område. Man vil også kunne oppdage om dyrene holder seg på en jevn avstand fra hverandre. Da vil grafen stige brått rundt denne avstanden. Altså hvis mange sauer har lik avstand til nærmeste nabo, vil dette være svært tydelig i grafen.

Funksjon:

$$G(d) = \frac{\#(d_{\min}(s_i) < d)}{n} \quad (2.2)$$

Funksjonen ser på om  $d_{\min}$  (nærmeste naboavstand) for punktet  $s_i$  er kortere enn  $d$ . Antallet punkter med mindre avstand fordeles på det totale antall punkter  $n$ .

### 2.4.2 F-funksjon

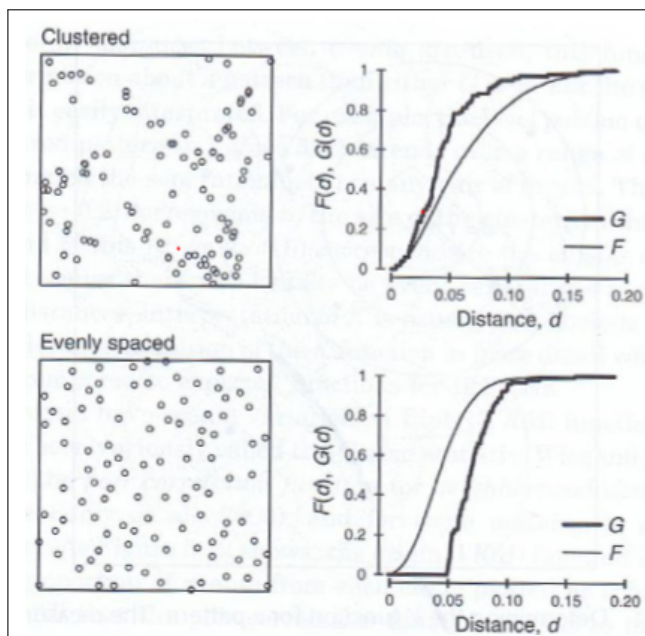
F-funksjonen er nært beslektet G-funksjonen, men baserer seg på å plassere ut en mengde tilfeldige punkt i studieområdet, og beregner de tilfeldige punktenes nærmeste naboavstand til fenomenet. [15] Dette fører til flere målte avstander og gir med det en “glattere” kurve. F-funksjonen vil i dette tilfellet si mer om områder uten sauer, da de tilfeldige punktene plassert her vil da gi lange avstander, som igjen vil vises i grafen.

Funksjon:

$$F(d) = \frac{\#(d_{\min}(p_i, S) < d)}{m} \quad (2.3)$$

F-funksjonen ser på om  $d_{\min}$  (nærmeste naboavstand) fra punktet  $p_i$  til punktene i punktmønsteret  $S$  er kortere enn  $d$ . Fordelingen skjer på antallet tilfeldige punkter  $m$ .

Av figur 2.4 ser man at G-funksjonen blir ekstremt bratt når punktene er jevnt spredd. F-funksjonen derimot vil få en mindre bratt, men tidlig start på kurven, da de tilfeldige punktene som er satt ut vil være under halvparten av avstanden til den nærmeste naboen. Når punktsettet er klyngete vil G-funksjonen stige raskest, siden alle punkter har korte avstander til hverandre. F-funksjonen vil være jevn og lang, da mange av de tilfeldige punktene blir lagt til områder som ikke har klustere, og de blir da langt unna.



Figur 2.4: Sammenstilling av G- og F-funksjonene[13]

### 2.4.3 K-funksjon

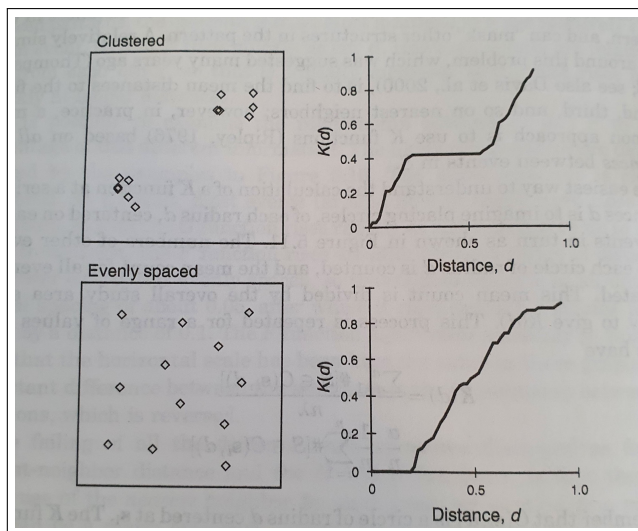
Problemet med G- og F-funksjonene er at de kun ser på de nærmeste punktene, og hvis alt er fordelt i små grupper, vil ikke grafene si noe om avstand mellom grupper. En løsning på dette er K-funksjonen, som sier noe om hvor mange punkter som ligger nærmere enn avstanden  $d$ . Denne funksjonen sørger for at alle avstander til alle sauer vises[15].

Funksjon:

$$K(d) = \frac{\sum_{i=1}^n \#[S \in C(s_i, d)]}{n\lambda} \quad (2.4)$$

$C(s_i, d)$  beskriver sirkelen  $C$  plassert i punktet  $s_i$  med radius  $d$ .  $S$  beskriver området man finner punkter i,  $n$  er det totale antall punkt, og  $\lambda$  er tettheten.

Man ser tydelig av det øverste plottet i figur 2.5 at det er to klynger, da man har et tydelig sprang rundt kort avstand, og nok et sprang når avstanden når neste klynge. I plott to ser man at det ikke er noen særlige hopp grunnet jevn avstand.



Figur 2.5: K-funksjon for to typer mønstre [14]

## 2.5 Regresjonstrær

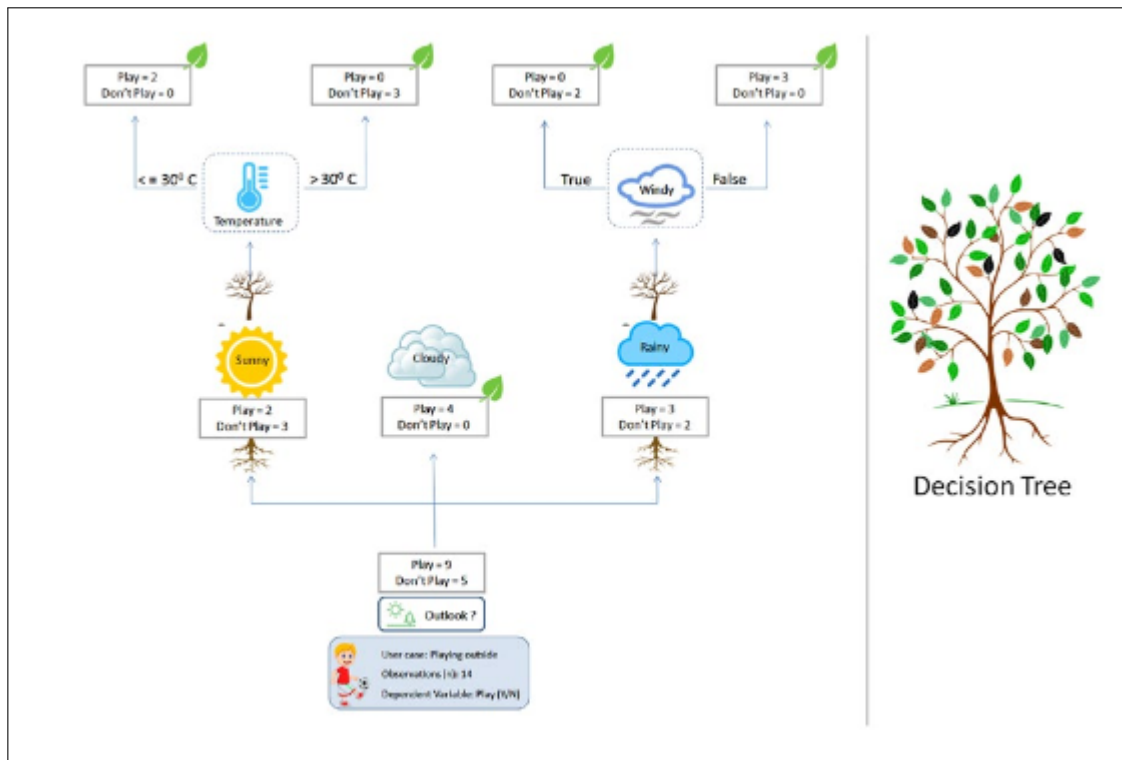
Regresjonstrær (også kalt beslutningstrær) forklarer sin oppbygging gjennom navnet. Dette fordi de er bygd opp i en tre-lignende struktur bestående av noder og grener som knytter disse sammen (se figur 2.6). Hver node representerer en test av en attributt, og hver gren er resultatet av testen. Løvnodene representerer klassifiseringen. Veien fra rota til løvnoden representerer klassifiseringsreglene. For å bygge et regresjonstre, benytter man treningsdata, som bestemmer hvilke variabler som deles ved hvilke noder, og ved hvilken verdi. Den store fordelene med regresjonstrær er at man slipper å transformere variabler, da variablene behandles enkeltvis, og man ser ikke på kombinasjoner av disse[54]. Denne fordelene gjør at man kan bruke læring direkte på datasettet, og ingen preprosessering er nødvendig. Dette kan være svært nyttig i en modell som kontinuerlig oppdateres, da punktinformasjonen kan lastes direkte inn i modellen.

Fra treningsdataene genererer algoritmen spøringer, som skiller datasettet. Denne skillingen skjer på en måte som gjør at informasjonsevinsten (IG) blir størst mulig.

IG beregnes ved formelen:

$$IG(D_p, x_i) = I(D_p) - \frac{N_{left}}{N_p} I(D_{left}) - \frac{N_{right}}{N_p} I(D_{right}) \quad (2.5)$$

$x_i$  beskriver hvilken variabel som splittes,  $N_p$  sier antallet treningsdata i foreldrenoden,  $D_p$  er et utvalg av treningsdatene i foreldrenoden, og  $D_{left}$  og  $D_{right}$  er treningsdataene i løvnodene etter splittingen.  $I$  er funksjonen for beregning av treffsikkerhet, og ved regressjon bruker man Mean squared error (MSE), altså det kvadrerte gjennomsnittet av residualene[41].



Figur 2.6: Regresjonstre [53]

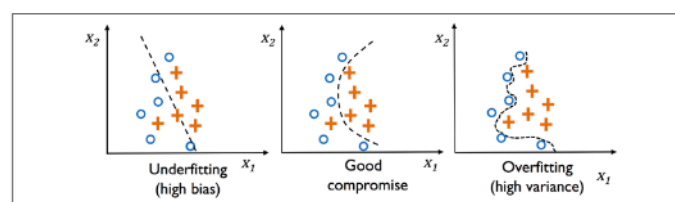
En videreutvikling av denne algoritmen er Randomforest. Randomforest er en ansamlingsteknikk som kombinerer flere mindre regresjonstrær. Her bruker man mindre og randomiserte utvalg av observasjonene og variablene, for å lage flere forskjellige regresjonstrær. Trærne blir da ukorrelerte, og kan finne frem til forskjellige resultaer. Man samler så disse resultatene og finner et uvektet gjennomsnitt, som er prediksjonen[53]. Styrken til Randomforest er at på grunn av tilfeldigheten som benyttes, blir modellen bedre generalisert.

Regresjonstrær bygges altså opp av mange spøringer, og beregner resultatet ut i fra dette. Randomforest kan man se på som en skog av enklere regresjonstrær, og predikerer ved å sammenstille resultat fra disse.

### 2.5.1 Over- og undertilpasning

Over- og undertilpasning er vanlige problem ved maskinlæring [42]. Overtilpassning betyr at modellen har høy varianse, som er for tilpasset treningsdataene. Denne vil derfor prestere dårlig på testdata. En undertilpasset modell blir for generell til å fange mønsteret i treningsdataene, og har dermed ikke lært nok. En god modell har derfor en tilpassning mellom disse to problemene. Figur 2.7 viser hvordan de forskjellige problemene tilpasser seg datasettet.

Siden Randomforest er en generaliserende versjon av regresjonstrær, reduserer modellen sjansen for overtilpassning, og ved bruk av nok treningsdata sørger man for å unngå undertilpassning[?].



Figur 2.7: Undertilpassning, godt kompromi og overtilpassning[40]

## 2.5.2 Modellens godhet

For å finne ut hvor god en modell er, bruker man noe som heter  $R^2$ . Regresjonsmodellen beregner en prediksjon for hvert punkt, og man bruker avviket fra sannheten til prediksjonen for å beregne godhet. Formelen er gitt ved:

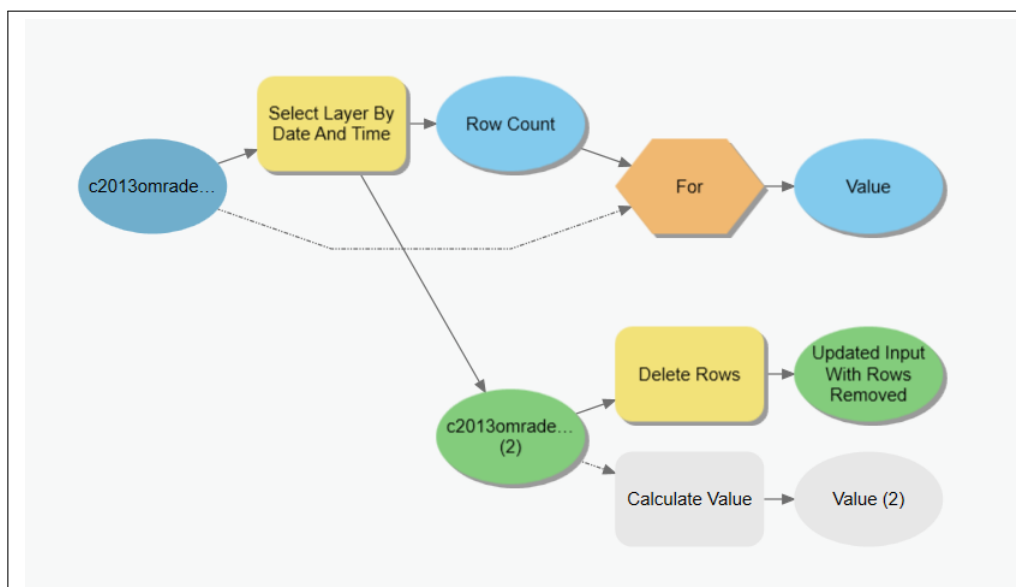
$$r^2 = \frac{SS_R}{SS_T} \quad (2.6)$$

$SS_R$  er summen av den kvadrerte forskjellen mellom beregning og sannhet, og  $SS_T$  er den kvadrerte forskjellen mellom gjennomsnittsverdien for sannheten og dette avviket til sannheten for hvert punkt[31].  $SS_R$  er altså feilen modellen får av variasjonen mellom prediksjon og sannhet, og  $SS_T$  er variasjonen som finnes fra før. Når  $R^2$  er høyere enn 0, viser dette at modellen er bedre enn å predikere ved bruk av gjennomsnittet. Den maksimale scoren man kan få ved  $R^2$ -score er 1. Da forklarer modellen alt. Hvis modellen får f.eks. en  $R^2$  på 0,74, betyr dette at modellen forklarer 74% av variansen i datasettet.

## 2.6 ModelBuilder

ArcGIS har en funksjon som heter ModelBuilder. Dette er et visuelt programmeringsspråk for å behandle geografiske data. Modellene man lager i ModelBuilder er representert ved et diagram som lenker sammen sekvenser av prosesser og geoproseseringsverktøy. En av grunnene til at man bruker ModelBuilder er at outputen fra en prosess kan brukes direkte som input i en annen prosess. Dette gjør at man slipper mellomlagring av midlertidige filer[19].

ModelBuilder har et enkelt visuelt uttrykk som vist i figur 2.8. Blå betyr inputdata, altså ubehandlet data, grønn betyr behandlet data, gul er verktøy og oransje er iterator. Ved kjøring av modellen brukes også fargen rød, for å vise hvilken prosess som kjøres, og denne vises også når modellen avbrytes, for å vise hvor feilen ligger. Hvis noen av de nødvendige kriteriene ikke er oppfylt, vil boksen ha fargen grå. Eksempler på dette kan være at noe ikke er koblet til, eller at inputdata er på feil format. Eksport av ModelBuilder er også svært effektivt, da man kan eksportere både bildefil for å forklare modellen, og man kan eksportere Python-script. I Python kan man modifisere scriptet til å gjøre enda mer spesifikke oppgaver, og rammen for prosessen kan gjøres raskt og effektivt ved “klikk og lim” i ModelBuilder.



Figur 2.8: Eksempel på model fra ArcGIS ModelBuilder

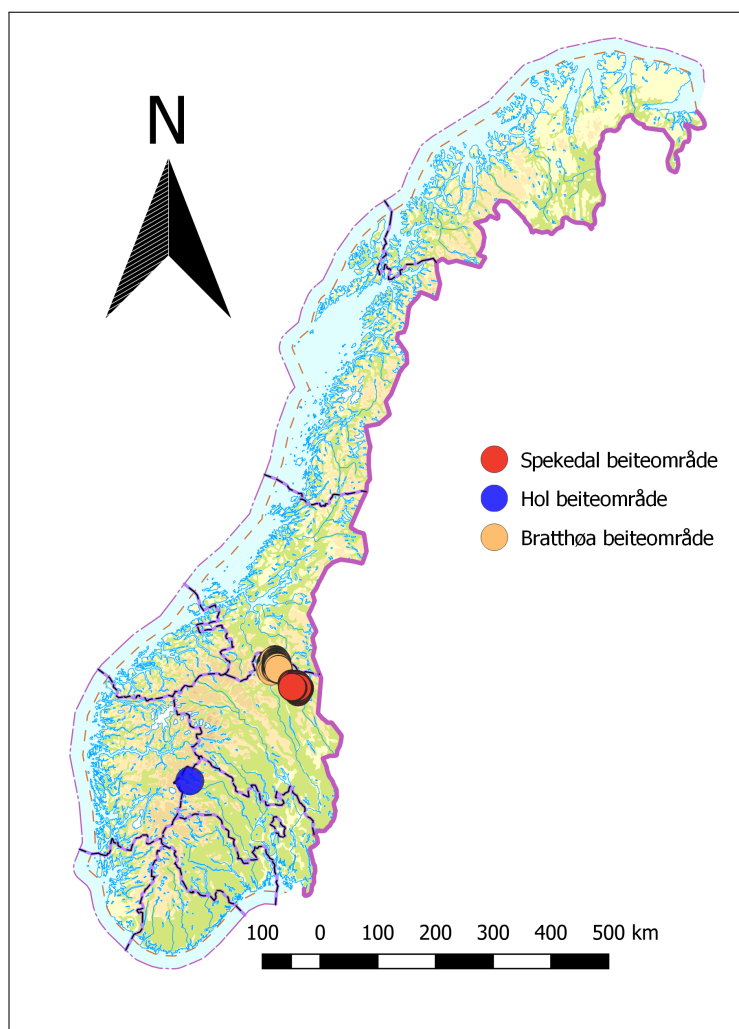


## 3. Data og Analyse

Kapittel 3.1 og 3.2 beskriver preprosesseringen av rå-dataene brukt i analysen. 3.3 og 3.4 beskriver fremgangsmåten for å besvare problemstillingene. Her er det tre forskjellige metoder presentert.

### 3.1 Testområder

Datsettene i denne oppgaven er hentet fra to steder: Hol kommune i Viken fylke og fra Nord-Østerdalen i Innlandet fylke. Hol ligger sørvest på Østlandet, helt mot grensen til Vestlandet, og Nord-Østerdalen ligger Nordøst på Østlandet.

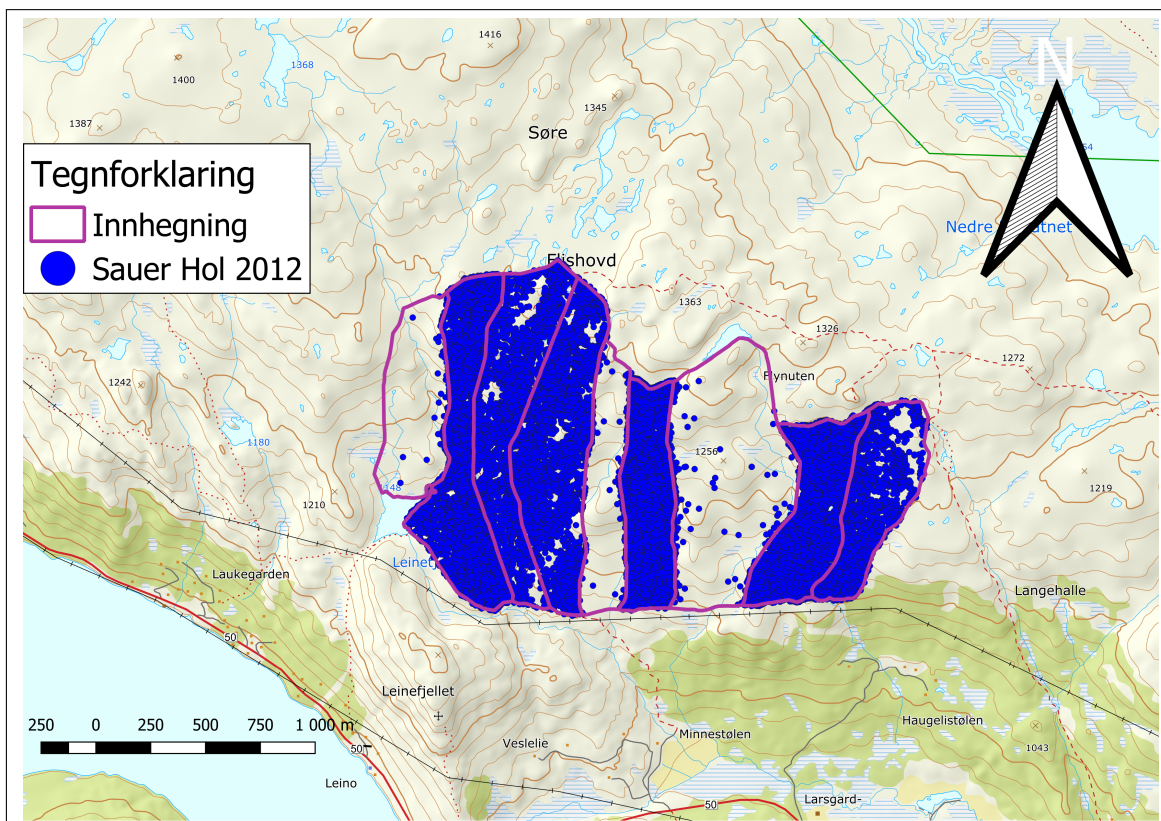


Figur 3.1: Kart over Norge og hvor studieområdene befinner seg.

### 3.1.1 Hol

I Hol er det 23 til 26 søyer og 44 til 49 lam fra samme flokk, som slippes ut i eksperimentelle innhegninger [34]. Dataene blir innsamlet ved bruk av Teullus radiobjeller fra Followit.

Studieområdet Hol ligger i Hol kommune i vestre del av Viken fylke, tidligere Buskerud fylke ( $60^{\circ}40'N$ ,  $7^{\circ}55'E$ ). Dette er et studieområde for studering av sau på beite, og ble opprettet i 2001. Det ble første gang tatt i bruk i 2002. Det har en utstrekning på ca.  $2,7 \text{ km}^2$  og har en høydevariasjon fra 1050 m.o.h. til 1300 m.o.h. [34]. Området er delt opp i flere mindre innhegninger på ca.  $0,3 \text{ km}^2$  for forskning på områder med lik naturtype. Metodikken for forskning her baserer seg på tetthetsanalyse, hvor man ser på forskjeller på beitebruk i sauflokken ved forskjellige tettheter. Dette ved å ha én innhegning med lav tetthet, og én med høy tetthet (25 og 80 sau per  $\text{km}^2$ ). Vegetasjonskartleggingen før etableringen av forsøksområdet er gjort av Yngve Rekdal [44], og denne utredningen sier at 9% klassifiseres som svært godt beite og 79% som nyttbart beite.



Figur 3.2: Hol beiteområde ( $60^{\circ}68'N$ ,  $8^{\circ}01'E$ ).





---

### 3.1.3 Forskjell på områder

Forskjellen på disse to studieområdene gjør at modellen favner et bredere spekter av beitedrift. Den ene studien baserer seg på innmarksbeite (Hol) og den andre på utmark. I tillegg er det to saueraser i drift i Nord-Østerdalen.

Tabell 3.1: Samletabell studieområder

Studieområde	Beitetype	N	E	Tetthet	Nyttbart beite
Hol (spredt)	Innmark	60°40′	7°55′	25 per km <sup>2</sup>	79%
Hol (tett)	Innmark	60°40′	7°55′	80 per km <sup>2</sup>	79%
Spekedalen	Utmark	62°40′	11°21′	0,238 per km <sup>2</sup>	26,2%
Bratthøa	Utmark	62°52′	10°77′	0,935 per km <sup>2</sup>	60,1%

## 3.2 Preprossesering av datasett

Datasettene er gitt av beitegruppa ved Senter for Husdyrforsk ved NMBU (Nicolai Hermann Jørgensen og Øystein Holand), hvor GPS-data er levert som kommaseparerte filer (.csv), og område-data som ESRI shapefiler.

### 3.2.1 Programvare

#### Python

Python er et åpent, interaktivt og objekt-orientert programmeringsspråk. Det at Python er åpent, gjør at det benyttes mye til utvikling i andre programmer, men det er kombinasjonen av regnekraft og klar syntaks som gjør dette til et populært programmeringsspråk. I Python finner man flere praktiske moduler som er nedlastbare og enkle å implementere i koden sin[1].

#### Pandas

Pandas er en offentlig modul, og den viktigste egenskapen Pandas har i denne oppgaven er datastrukturen dataframe. Dette er en representasjon av datatabeller som kan direkte lese .csv og .xlsx filer. Dette gjør konvertering svært praktisk og prosesseringen blir effektiv. Etter prosessering av dataframes, kan den enkelt lagres på disk som både .csv og .xlsx[37]. All opplastning og lagring av filer i Python har brukt Pandas-modulen.

#### Matplotlib

Matplotlib er en åpen modul som brukes mye til visualisering av datasett i Python. Histogram og spredningsplott er noen eksempler på visualiseringsteknikker man finner i Matplotlib[28]. Matplotlib er brukt til visualisering av residualplot, og for å visualisere data i analysen.

#### Scikit-learn

Scikit-learn er også en åpen og gratis modul som brukes til maskinlæring i Python [38]. Her finnes verktøy for analyse, preprossesering og modell-kalibrering. Det er i hovedsak estimatoren Randomforest som er benyttet fra Scikit-learn, men også SelectFromModel som anbefaler den beste modellen fra valgt estimator.

---

## Google Colaboratory

Google Colaboratory (Colab) er en åpen programvare eid av Google. Dette er et Python-miljø på nett, som man kan åpne i en internettleser. Colab kjører i skyen, slik at man ikke er avhengig av å ha den dyreste datamaskinen med den høyeste prosesseringskraften. En annen fordel er at Colab har forhåndsinstallert mange av modulene som nevnt tidligere, slik at de er klare til bruk[50].

### 3.2.2 GIS-verktøy

I oppgaven er to GIS-verktøy benyttet, ArcGIS Pro og QGIS. Grunnen til dette er at ArcGIS pro er bedre på håndtering av temporale data, med tanke på visualisering. I tillegg har ArcGIS pro modelleringsapplikasjonen ModelBuilder, som tidligere beskrevet.

#### QGIS

QGIS (Quantum GIS) er en gratis og åpen GIS-programvare, som brukes til analyse, redigering og visualisering av geografiske data[39]. I denne oppgaven er QGIS versjon 3.18.2 benyttet til å visualisere data og for å produsere heatmaps. QGIS er også benyttet til romlig kobling av datasett, og beregning av G-, F- og K-funksjoner.

#### ArcGIS Pro

ArcGIS Pro er privat eid GIS-programvare, eid av ESRI, og lisens er gitt av NMBU. ArcGIS kan brukes til visualisering, prosessering og analyse av 2D- og 3D-data[18]. I oppgaven har ArcGIS Pro blitt brukt til å visualisere fordeling av temporale data, bygge buffermodell i ModelBuilder og fjerning av støy.

### 3.2.3 GPS-data

GPS-dataene fra Hol består av kolonnene: Dato, Tid, Time to Fix, Latitude, Longitude, 2D3D, Altitude, Dilution of Precision, SVs, FOM, X-akselerasjon, Y-akselerasjon, Temperatur, Bat, Status, SCap GPS og GSM. For å få sortert bort feilmålinger, så fjernes alle rader som har en forringelse av presisjon (DOP) over 5, og ugyldige tidsformat. Dette er beskrevet i vedlegg A.

Data fra Nord-Østerdal består av: ID, breddegrad, lengdegrad, tid, vertikal DOP, sauerase, gård og studieområde.

Siden datasettene er bygd opp på litt forskjellige måter, så må de standardiseres, og attributter som begge har, eller har mulighet til å finne, settes opp. Datasettene består da av: koordinater, ID, tid, DOP og rase. Behandlingen ble gjort i Python, og ferdige filer ble eksportert som csv-filer.

### 3.2.4 Innlesing

CSV-filene ble lagt inn i QGIS ved "Create layer from a Delimited Text File". Deretter ble de lagret som på ESRI shapefil-formatet med nytt koordinatsystem, nemlig WGS84/UTM zone 32N (EPSG:32632). Dette fordi avstandsberegning er lettere med koordinater på Easting-/ Northing-format. Data fra Nord-Østerdal var samlet i en fil, så her ble de delt opp i fire filer, oppdelt etter årstall (2013 eller 2014) og etter område (Bratthøa eller Spekedalen).

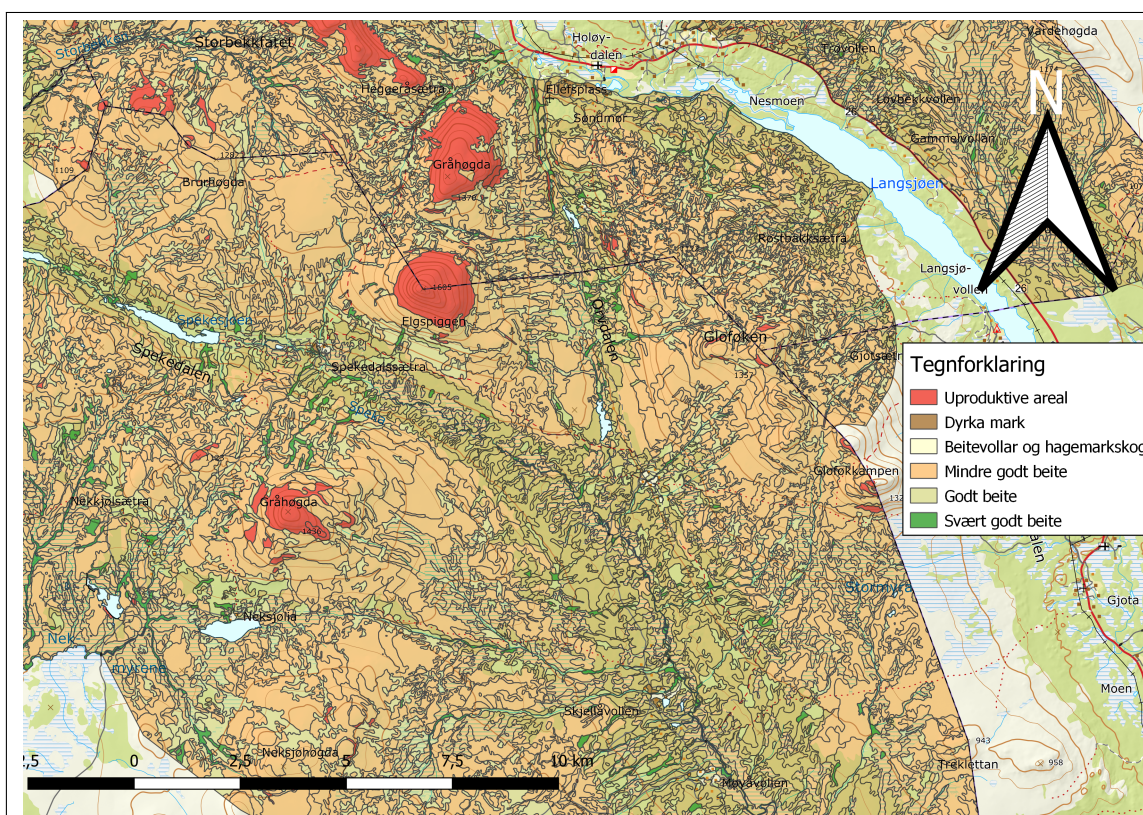


### 3.2.5 Feilmålinger

Feilmålinger kan bestå av så mangt, små avvik til store feil. Derfor er det viktig å identifisere de målingene som er umulige for en sau. Dette gjelder da fart som er altfor høy eller punkter langt utenfor studieområdet. Målinger uten høyde, altså klassifisert som 2D-koordinater kan også vurderes til å bli tatt ut av datasettet. D'Eon et al. (2002) sier at man kan beholde 2D-koordinatet om ønskelig, men dette kan gå ut over den statistiske kvaliteten ved beregning av nøyaktighet[16]. Ved innlesning av datasettet brukes DOP til å skille dårlige og gode målinger, slik at dette filtrerer ut dårlig målte punkter, i tillegg bruker ingen av analysene presentert målingsnøyaktigheten som faktor. Derfor er 2D-koordinater beholdt.

### 3.2.6 Grunnlagsverdier

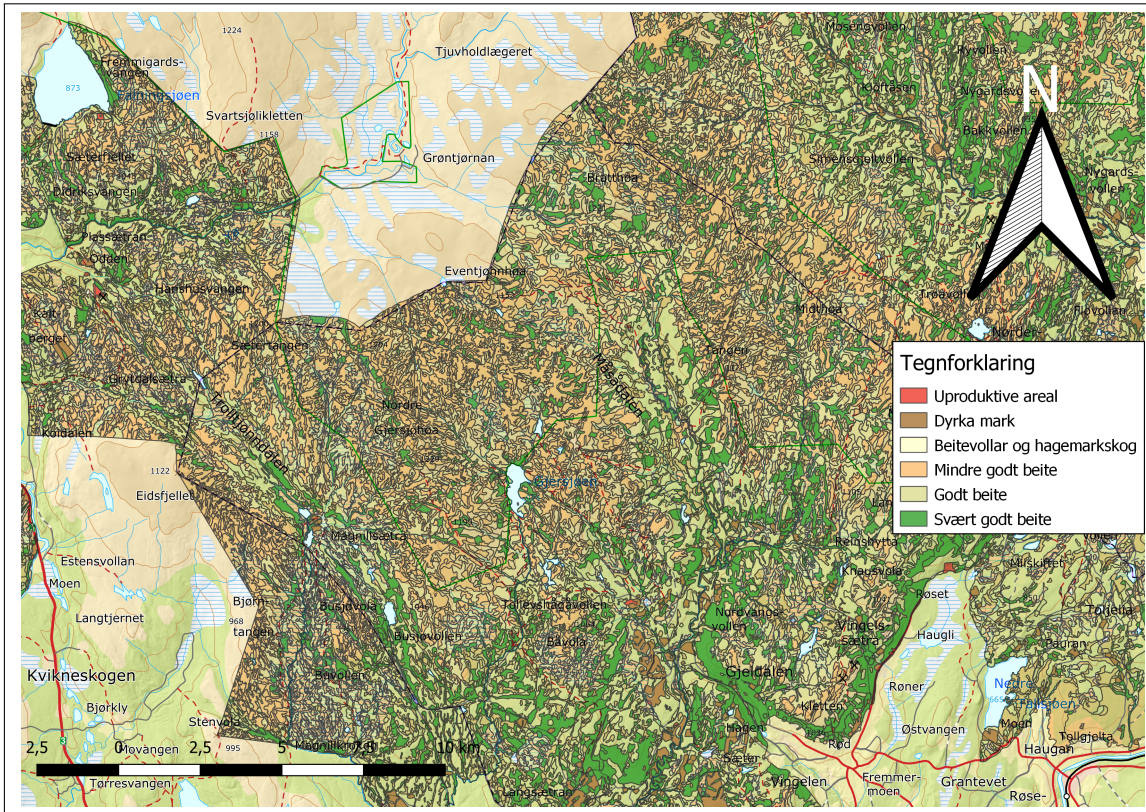
Ved å lese inn grunnlagsfilene i QGIS og kategorisere symbiologien etter beiteverdier, får man raskt og enkelt sett hvordan et beiteområde ser ut. Man får også en rask visuell oppfatning av hvor godt beitet er.



Figur 3.4: Spekedal beiteområde

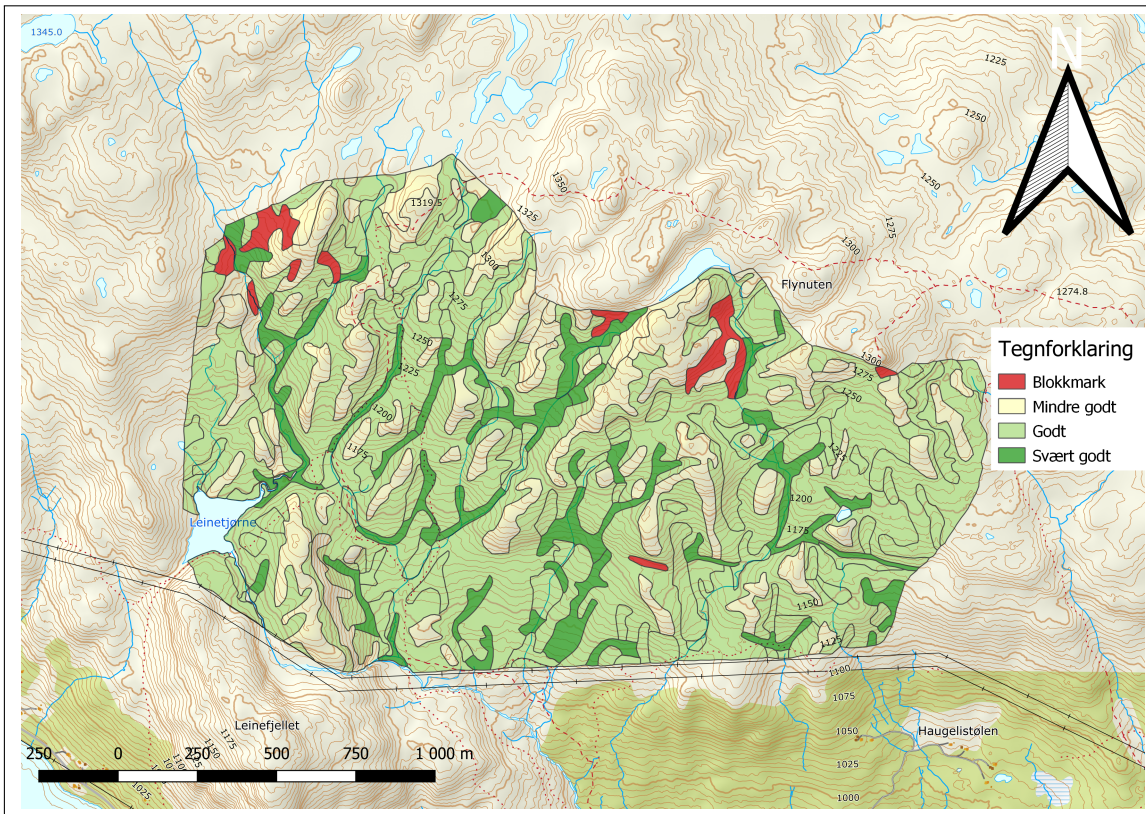
Grunnlagsverdier for Bratthøa og Spekedalen er hentet fra NIBIOs Kilden[36]. Kilden er en nettside hvor offentlig kartlagt informasjon ligger gratis, klare for nedlastning. Datasettet består av 22 kolonner, med informasjon om blant annet: areal, registreringsår, beitekvalitet for sau og storfe (både numerisk og kategorisk), områdebeskrivelse, naturtype m.m. Dataene har registreringsårstall fra og med 1974 frem til og med 2019, noe som er et tegn på at mye av beitekvaliteten er uendret siden 1974, og dermed lik i 2013 og 2014. Filen ble bestilt via NIBIOs løsning for nedlastning, hvor den så ble motatt på mail og lastet ned på ESRI shapefil-formatet. Beiteverdiene er som følger: Mindre godt beite, Godt beite, Svært godt beite, Beitevollar og hagemarkskog, Dyrka mark og Uproduktive areal.





Figur 3.5: Bratthøa beiteområde

Man ser av figur 3.4 at beiteområdet i Spekedalen er preget av høvfjellsområder (rødt) og tyntvoksende områder som er markert brun-oransje. De beste områdene er lavtliggende områder i dalførene, men selv her er det ikke store områdene med svært godt beite. I figur 3.5 ser man at Bratthøa har en helt annen signatur og oppfattelse ved første øyekast. Dette bærer preg av å være mye grønnere, og langt større områder som er svært gode beiteområder. Dette skaper variasjon i datasettene, noe som er bra for generalisering av modeller. Dette gir også et godt grunnlag for å si om oppførsel er lik på ulike beiter.

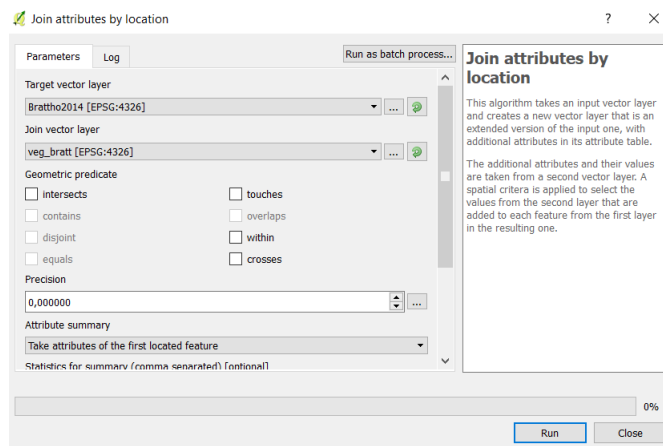


Figur 3.6: Hol beiteområde

Grunnlagsverdiene for Hol beiteområde ble kartlagt i 1997, og fulgte med punktdatasettene for GPS-registreringer fra beitegruppa ved NMBU. Filen er på ESRI shapefile-format og inneholder informasjon om: region, posisjon, areal, omkrets, og består av verdiene: areal, dato, målemetode m.m.

Beiteverdier er som følger: Blokkmark, Mindre godt, Godt og Svært godt. I figur 3.6 ser man Hol studieområde. Her er det hovedsaklig godt beite, med ispett av svært godt og mindre godt beite. Fordelingen av beitetypene er ganske jevn, da dette forsøksområdet ble opprettet for å se på sauens oppførsel ved forskjellige tettheter på tilsvarende beiteområde.

For å kombinere grunnlagsdatasettene med punktdatasettene brukes funksjonen “Join attributes by location” i QGIS. Denne legger til beiteverdi for hver posisjon sauene har befunnet seg. Siden datasettene også har flere kolonner med informasjon som ikke er relevant for denne analysen, så ble ikke disse tatt med i koblingen mellom lagene. Join attributes by location har en løsning for å velge ut de kolonner man vil koble mot datasettet.



Figur 3.7: “Join attributes by location” i QGIS

---

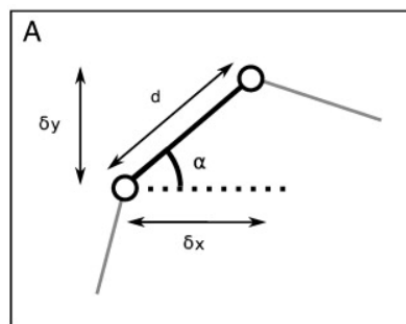
### 3.3 Romlig analyse

Romlig analyse presenterer metodene brukt for å finne bevegelsesparametre og for å analysere punktdatasettens romlige oppførsel.

#### 3.3.1 Sporing

Teimouri et al. (2018)[56] viser til hva slags parametre og beregninger som bør gjøres for lett å kunne tolke GPS-sporene. Ved å beregne seks parametre av posisjonsendring, kan man dele datasettet inn i fire forskjellige klassifiseringer:

- (I) hvile (lav gjennomsnittlig rotasjonsvinkel og lav gjennomsnittshastighet)
- (II) gange (lav gjennomsnittlig rotasjonsvinkel og høy gjennomsnittshastighet)
- (III) beitende (høy gjennomsnittlig rotasjonsvinkel og lav gjennomsnittshastighet)
- (IV) udefinert bevegelse (høy rotasjonsvinkel og høy gjennomsnittshastighet)



Figur 3.8: Parametre for beskrivelse av bevegelse[8]

Dette gjør at rotasjonsvinkel og hastighet må beregnes for hvert punkt, noe som gjøres ved å se på forrige punkt i dyrets spor, og finne endringen som har skjedd.

De tilleggsparemetere som er beregnet er avstand mellom posisjoner ( $d$ ), tidsforskjell, øst-vest-forflytning ( $\delta X$ ), nord-sør-forflytning ( $\delta Y$ ) og absolutt-vinkel ( $a$ ), altså vinkel mellom X-aksen og forflyttningsretningen. Dette er vist ved illustrasjon i figur 3.8.

$$\delta X_t = X_t - X_{t-1} \quad (3.1)$$

$$\delta Y_t = Y_t - Y_{t-1} \quad (3.2)$$

$$d_t = \sqrt{(\delta X)^2 + (\delta Y)^2} \quad (3.3)$$

$$a_t = \tan^{-1}\left(\frac{\delta Y}{\delta X}\right) \quad (3.4)$$

$$v_t = \frac{d_t * 3.6}{T_t - T_{t-1}} \quad (3.5)$$

Ved å følge sporene for hver enkelt sau, kunne man lett finne tid og posisjonsendring mellom målinger, og med dette avlede fart og vinkel på bevegelsen (i forhold til øst). For å finne endringen i rotasjon, kan man bruke funksjonen:

$$\delta a_t = a_t - a_{t-1} \quad (3.6)$$

Beregningene ble gjort i Python (se vedlegg B).



---

Dobos et al. (2015)[17] har forhåndsklassifisert  $> 0,072 \frac{km}{t}$  som stasjonær,  $\geq 0,072 \frac{km}{t}$  og  $\leq 0,118 \frac{km}{t}$  som beiting, og  $< 0,118 \frac{km}{t}$  som vandring for sauer på beite. Ved å ta utgangspunkt i disse, får man et overblikk om hva farten tilsier av oppførsel. Disse fartsverdiene baseres på GPS-målinger for hvert 10. minutt, noe som er tre og seks ganger så hyppig som datasettene presentert i denne oppgaven. Siden farten vil reduseres av at bevegelseslinja er mer generalisert, senkes terskelen til at under  $0,05 \frac{km}{t}$  beskriver hvile, og gange er over  $0,1 \frac{km}{t}$ . Fartsmålingene som havner innimellom her vil da bestå som beitefart. Disse grensene brukes ikke til beregning, men som et bakteppe ved inspeksjon av oppførsel.

### 3.3.2 Flokksporing

Sauenes oppførsel er forsket på i mange år, og i 2009 publiserte Mobæk et al.[34] en artikkel om tetthet og tidens variabilitet med tanke på sauers beiteseleksjon. Den fant at tettheten kan si noe om beitemønsteret, for ved høy tetthet valgte et mindre antall av sauene et høyproduktivt område, og ved lav tetthet kunne de fleste sauene trekke til disse. Dette gjaldt imidlertid for beiting. Når det gjaldt hvile, så var det en liten trend i at de trakk mot lavproduktive områder. Etersom beitingen hadde pågått en stund, kunne man se en reduksjon i ferdsel til de beste områdene, dette på grunn av at de beste plantene begynte å bli spist. I oppgaven blir derfor den oppgitte tettheten av sau (sau per km<sup>2</sup>) og uker siden start av beiting benyttet som variabler i analysen.

Fra en forskningsartikkel senere i samme prosjekt fant man at sauer i de tette innhegningene var mer aktive enn de i mindre tette (hadde mer tid på beina).  $66.2\% \pm 4.6\%$  for tett saueflokk, og  $60.8\% \pm 3.2\%$  for mindre tett, målt i prosent aktiv på dagtid mellom 09:00 og 22:00. Så man på alle 24 timer i dagen, fant man lignende data, men mindre forskjeller. I tillegg fant studien at sauer reduserte sin gjennomsnittlige bevegelse etterhvert i løpet av sesongen, med 28 cm hver 14. dag [33]. Dette gjør variablene: tid siden beitetiden startet, dato, aktiv/inaktiv og tetthet interessante.

### 3.3.3 Avstandsbasert mål av punktfordeling

Det å se på avstand mellom punkter er svært enkelt ved bruk av QGIS. I QGIS 3.18 følger statistikkprogrammet R med, og her finner man flere statistiske mål. G-, F- og K-funksjon er nettopp dette, statistiske mål på punktmønster. Ved å bruke disse funksjonene i QGIS får man generert plot av funksjonen, og man kan fort analysere resultatet. I denne sammenheng kan man bruke det til å se etter tegn på grupperinger (klynger), eller å se på om sauer vandrer alene. Sammenligningen av funksjonene vil si noe om utnyttelse av hele området, da G-funksjonen forklarer hvor langt enkelt-sauer beveger seg, og F-funksjonen sier noe om utstrekningen i hele beiteområdet. K-funksjonen igjen vil på sin side vise om det er tydelige grupperinger, og hvor stor avstanden er mellom disse.

Den første metoden bruker avstandsbasert mål av punktfordeling og forsøker å besvare *Hvilke indre parametere hos en saueflokk viser tydelige tegn til en normaltilstand?* og *Er oppførselen synlig på både innmark og utmark?*

### 3.3.4 Punkttetthetsanalyse

Punkttetthetsanalysen er en effektiv metode for å se på besøkte områder. Disse vil være av stor interesse, da man burde finne ut hva som gjør at sauene trekker dit, og om dette er en fellesnevner i de forskjellige datasettene. Ved bruk av QGIS sin funksjon “Heatmap” for å visualisere kartlaget, får man til nettopp dette. Denne funksjonen har også en innlagt forstørrelsesfunksjon, som gjør at når man forstørret et område, får man et nytt “Heatmap” av det mindre området. Dette gjør at man kan gå inn i dybden av kartlaget, og finne akkurat de stedene som er hyppigst besøkt.



Ved å bevare punktkartlagene slik de er kan man velge å se på de punktene som finnes i det tette området. Ved sortering av valgte punkter kan man sortere etter dato og med det se sammenhengen i når målingene er registret. Man kan også sortere på andre variabler, f.eks. fart og beitegrunnlag. Man kan også se på bakgrunnskartet for å se etter topografi i terrenget, som høydekurver eller farge på kartlaget.

Noe av ulempen med denne analysen er at den ikke tar hensyn til at sauene går innom punktet flere ganger. Det blir derimot de lengre oppholdene som ved hvile om natten som blir trekt frem.

Metode to benytter et tetthetsbasert mål av punktfordeling for å finne interesseområder, og ser på grunnlaget i disse områdene. Metode to ser på følgende problemstillinger: *Hvilke ytre tegn observerer man hos en saueflokk under normale omstendigheter?* og *Er oppførselen synlig både på innmark og utmark?*

## 3.4 Grunnlagsanalyse

Den tredje metoden baserer seg på maskinlæring og prøver å finne en modell for flokkens fordeling basert på grunnlagsdata. Metode to og tre ser på følgende problemstillinger: *Hvilke ytre tegn observerer man hos en saueflokk under normale omstendigheter?* og *Er oppførselen synlig både på innmark og utmark?*

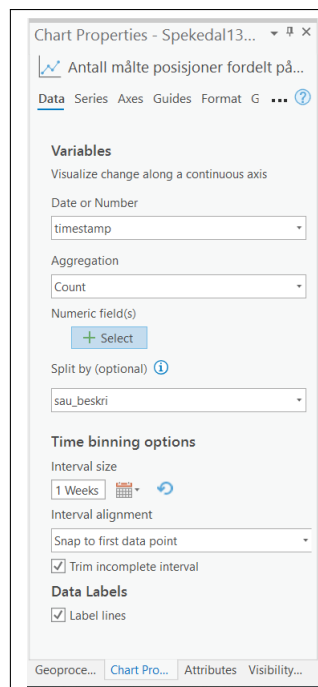
### 3.4.1 Sortering av grunnlag

Raynor et al. (2020)[43] presenterer en metode for å estimere hvor kyr befinner seg, ved å se på topografiske variasjoner. Faktorer tatt i betraktning er beitekvalitet, vanntilgang, flokketetthet og bratthet. Dette er gjort på forskjellige beiteområder i USA, og målingene er gjort på kyr.

Valg av beite forandres i løpet av beitesesongen [34]. Dette kan man analysere over tid, ved å koble datasettene med grunnlagsverdiene. For å sammenstille disse dataene er ArcGIS sine funksjoner for visualisering av data effektive. Ved bruk av “Create chart → Line chart” får man plottet en presentasjon av datasettet, og sortering etter tidskolonna “timestamp” og summering av antall målinger per beitegrunnlag (“sau\_beskri”) gir en god visualisering. “Trim incomplete intervall” gjør at datasettets tid starter når antallet målinger er på et stabilt nivå, man unngår med dette utslipp og sanking av sauene.

Ved manuell inntasting av verdiene kan sorteringen noteres, og antallet målinger registert på de ulike beitelegges i en tabell. Tilleggsverdier, som ukenummer, tetthet og innmark/utmark legges også til som forklarende variabler (se tabell 3.2).

“Antall” beskriver antallet målte punkter oppgitt “Beitetype” inneværende uke. “Fordeling” er den oppgitte fordelingen av beitetypene i studieområdet. “Tetthet” er den oppgitte tettheten av sau fordelt på studieområdet. “Totalt” sier hvor mange punkter som er målt totalt inneværende uke. Ved å kombinere alle de åtte filene for hvert studieår og område i Python ble det 318 rader med informasjon som kan brukes til treningsdata og valideringsdata.



Figur 3.9: Meny for plotting av linjediagram i ArcGIS

Tabell 3.2: Oppsett ukentlig sortering

Antall	Beitetype	Fordeling	Ukenr	Tetthet	Inn- /utmark	Totalt
49	Uegnet	0.019	2	52.5	Innmark	6936
1193	Mindre Godt	0.215	2	52.5	Innmark	6936
4286	Godt	0.662	2	52.5	Innmark	6936
1408	Svaert Godt	0.104	2	52.5	Innmark	6936
72	Uegnet	0.019	3	52.5	Innmark	7140

Visualisering av datasettet gjøres både i Python og ArcGIS. ArcGIS viser “Antall” fordelt per uke, og Python viser den gjennomsnittlige prosentvise fordelingen for hvert beiteområde. To visualiseringer benyttes siden “Antall” er variabelt på tvers av beiteområdene. Man kan beregne den prosentvise fordelingen av en uke, ved å dele “Antall” på “Totalt” i Python. Dette viser hvor mange prosent av sauene som er på hvilket beite.

### 3.4.2 Randomforest

Regresjonstrær er tidligere brukt til klassifisering av beitedyrs oppførsel. Et tidligere studie delte inn data i beiting, gange, ståing og ligging ved direkte observasjon av kyr. De direkte observasjonene ble brukt som fasitdata, og ved bruk av regresjonstre-modell kunne 85% av variasjonen forklares ved bruk av GPS- og akselerometermålinger hvert 20. minutt[57].

Modellering av oppførsel er altså mulig, men for å kunne si noe om den forventede oppførselen, må man først finne trendene. Ved å presentere fem forskjellige metoder for å estimere sauers bruk av beiteområde, skal de ulike trendene og sammenhengene på tvers av områder, beitetypen og år belyses.

Første metode: Lære ved bruk av både utmark og innmark, og estimere oppførsel på ett nytt område. Her kommer alle kolonner til nytte, og man kan se om modellen finner trekk i to av områder, som kan brukes til å forklare et nytt område. Denne metoden vil vise om det er sammenheng på forskjellige beiter og forskjellige saueflokker.

Andre metode: Lære ved bruk av utmark, og se om dette kan brukes til å estimere oppførsel på innmark. Siden to utmarksområder med ulik tetthet av sau og beitetypfordeling benyttes, kan mange av disse forskjellene brukes til å finne felles trender. Metoden vil vise om det finnes en sammenheng mellom inn- og utmark.

Tredje metode: Lære ved bruk av innmark, og se om dette kan estimere utmark. Her er det kun ett område, med lik tetthet og lik beitetypfordeling. Dette gjør at modellen ikke får utnyttet variasjonen her til læring. De andre verdiene som ukenr. og beitetype vil få større vekt, og man får vist om dette også har en trend på tvers av områder.

Fjerde metode: Lære av samme område og estimere neste år. Vil vise om det er sammenheng i oppførselen år etter år.

Femte metode: Lære av to områder og legge til forrige år for testområdet. Denne metoden vil vise om det er bedre å bruke spesialiserte modeller per område, eller om en generell modell er bedre.

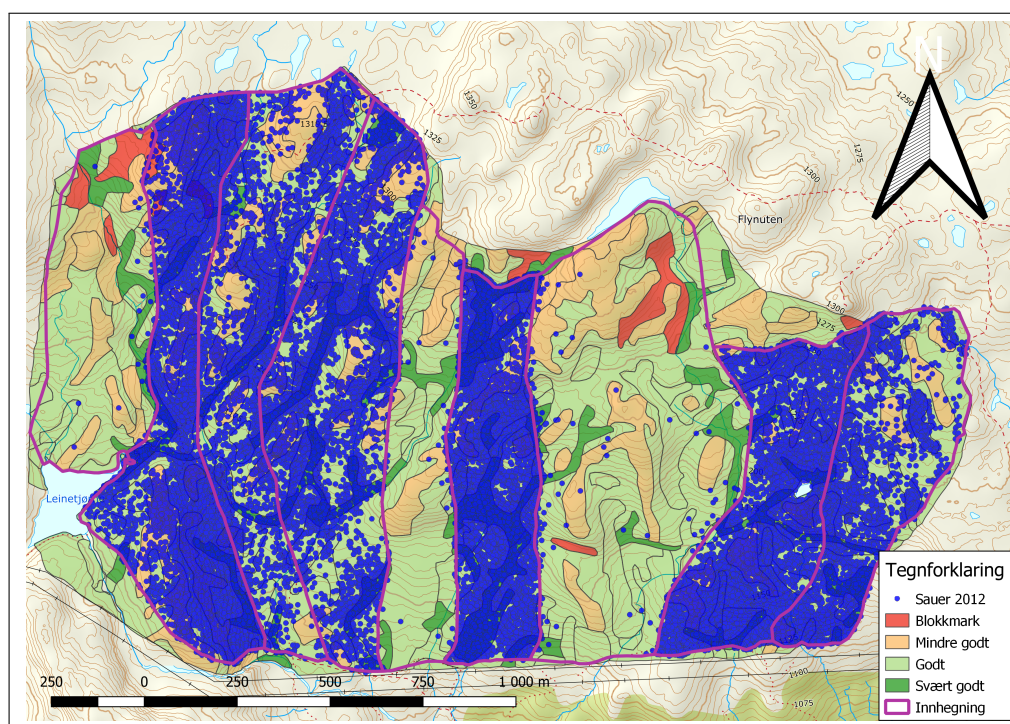
Med disse metodene får man vist hva modellen gjør det godt og dårlig på, og forskjellene kan si noe om den observerte trenden på tvers av datasett, områdetyper og år. Metode fire er en spesialisert modell, og metode fem er en generalisert modell.

## 4. Resultater

### 4.1 Observasjoner

#### 4.1.1 Romlig fordeling

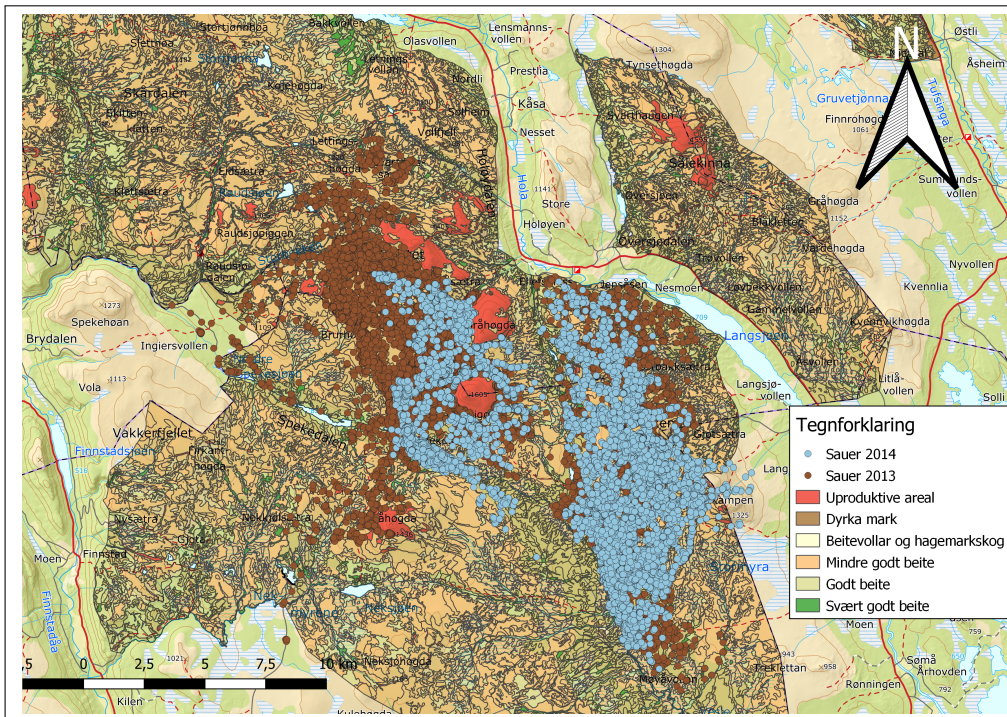
Den enkleste formen for visualisering, er å laste punktdatasettene inn i QGIS. Dette er nyttig for finne potensielle forventninger til resultatene.



Figur 4.1: Sauenes fordeling i Hol beiteområde

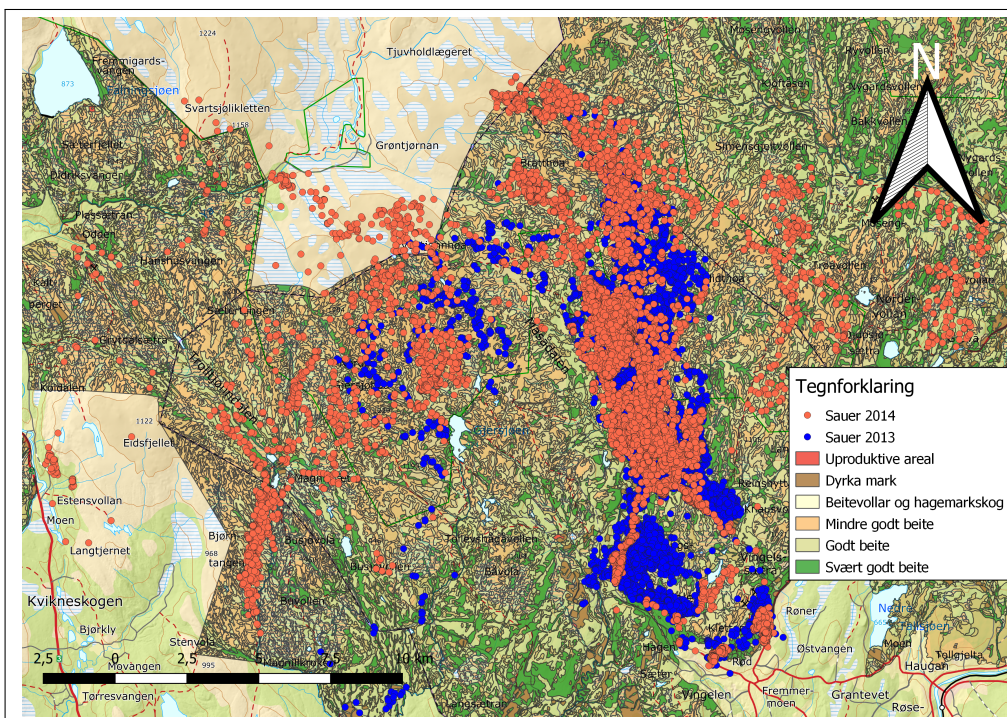
Man ser tydelig at det er tre av beitenene som ikke er i bruk, og seks som er i bruk. Noe vandring i de tomme områdene kan også ses. Man ser også at innhegningene overlapper punkter. Dette kan komme av både GPS-unøyaktighet for punkter og gjerdeinntegningen, eller at sauene faktisk har kommet seg gjennom. Det er også faktisk mulig å se hvilke områder som har en tett bestanddel av sau og de som ikke har. Område to, seks og åtte har mange tette punkter, som kommer av at det er mye vandring, og flere målinger. Område tre, fire og ni har større glipper. De områdene som vises i glippene er av typen mindre godt beite, men også noe godt beite. Signaturen til svært godt beite slipper ikke igjennom hos noen av innhegningene, noe som tyder på at sauene har vært der.





Figur 4.2: Sauenes fordeling i Spekedal beiteområde

Figur 4.2 viser tydelig at sauene unngår “Uproduktive områder”, som i dette tilfellet er høyfjell. Det er også noen “glipper”, hvor mindre godt beite er unngått. Av datasettet, ser man at beiteflokken også er todelt, dette av en sørovergående dal.

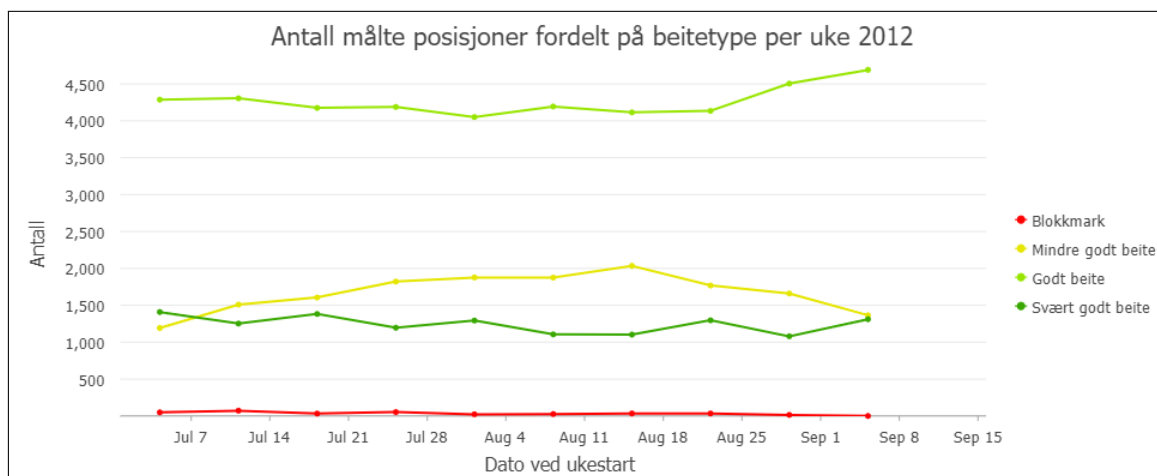


Figur 4.3: Sauenes fordeling i Bratthøa beiteområde

I figur 4.3 ser man at sauene i hovedsak beiter i godt område. Her er det vanskelig å se tydelige områder som unngås. Spredningen er stor, spesielt i 2014.

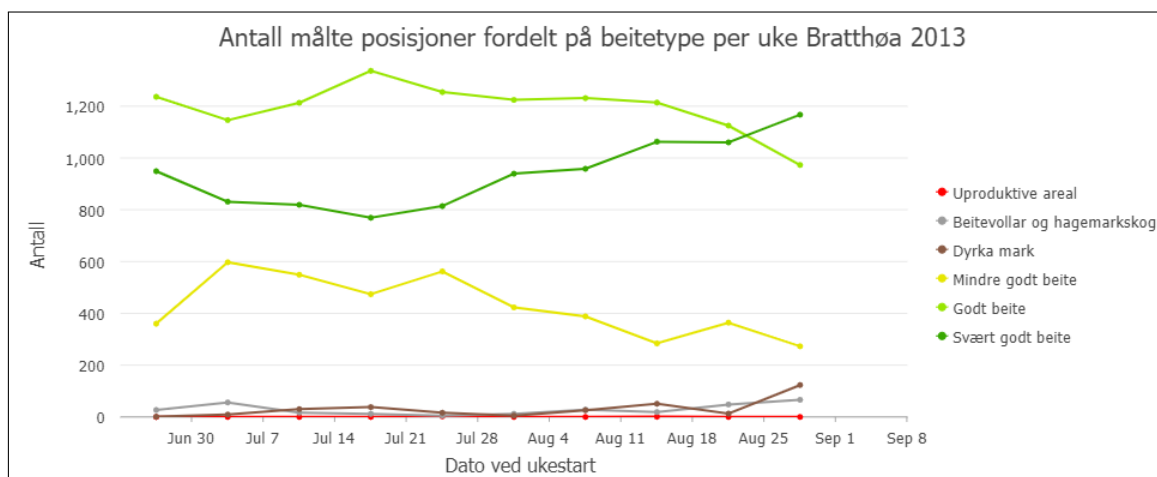
### 4.1.2 Ukentlig fordeling

Ved bruk av ArcGIS sine funksjoner for visualisering av data, ser man fordelingen på beitetyper gjennom beitesesongen.



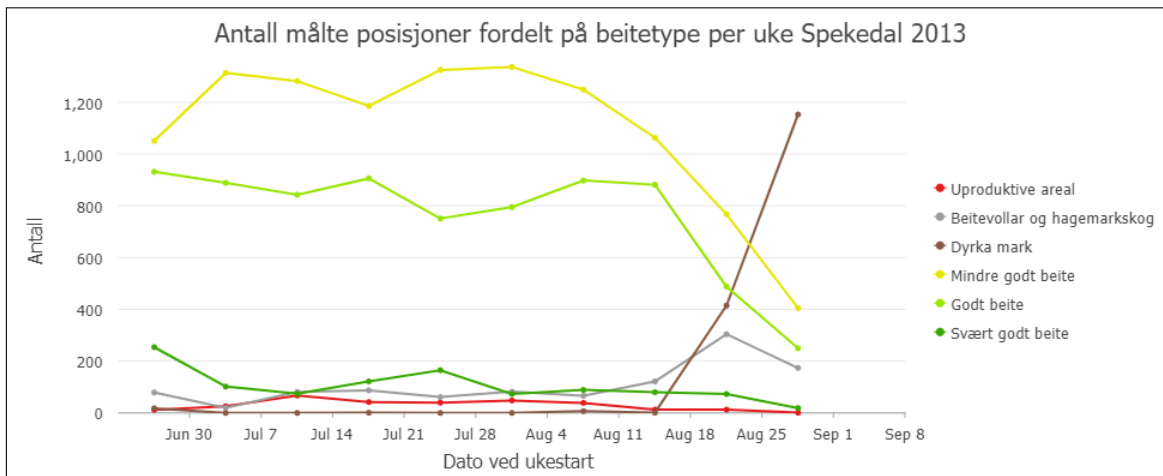
Figur 4.4: Antall punkter per beitetype Hol 2012

Man kan se at det er en trend av høyere besøk på typen “Svært godt beite” i begynnelsen av beitesesongen, og at denne er synkende gjennom sesongen. Samtidig kan man også se at typen “Mindre godt beite” er generelt økende gjennom sesongen. “Godt beite” og “Blokkmark” holder seg relativt stabile, men “Godt beite” har en stigning mot slutten av beitesesongen.



Figur 4.5: Antall punkter per beitetype i Bratthøa

I Bratthøa beiteområde, ser man at “Godt beite” og “Svært godt beite” er de mest benyttede områdene. Man ser at “Godt beite” er jevnt brukt, men noe synkende de siste to ukene. “Svært godt beite” er jevnt stigende fra uke fire, og “Mindre godt beite” er synkende fra uke fem ca.



Figur 4.6: Antall punkter per beitetype i Spekedalen

I Spekedal er det en annen fordeling. Her ser vi at “Mindre godt beite” holder seg stabilt høyt frem til uke sju, hvor “Godt beite” blir mer brukt. Både “Godt beite” og “Mindre godt beite” synker kraftig på slutten, og det ser ut som om sauene kommer hjem, da de mot slutten av beitesesongen er på “Dyrka mark”.

#### 4.1.3 Visualisering av fordeling

Ved å sammenstille plot av den prosentvise fordelingen på de ulike typene beiter, er det lette å oppdage likheter og ulikheter. Figur 4.7 til 4.12 viser dette. Y-aksen viser hvor mange prosent av sauene som befinner seg på beitet, og x-aksen viser ukenummeret. Resultatplottene er en midling av målingene for alle årene på beitet. Vedlegg E viser script for plotting.

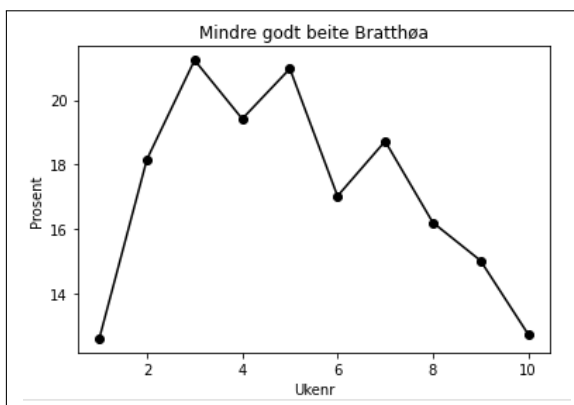
Godt beite og uproduktive beiter er ikke visualisert, da disse har ganske stabile fordelinger, sett bort fra Spekedal 2013, hvor alle sauene beveger seg ned mot dyrka mark.

---

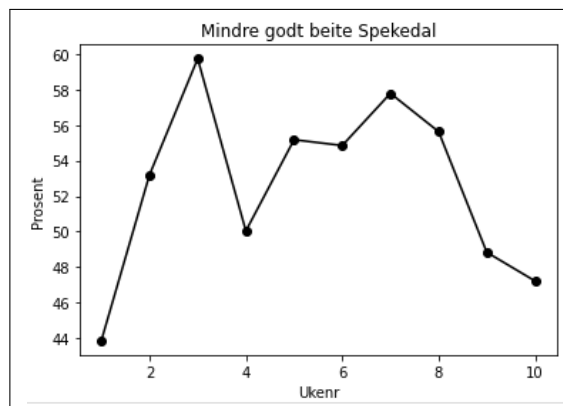
## Mindre godt beite



Figur 4.7: Fordeling Hol



Figur 4.8: Fordeling Bratthøa

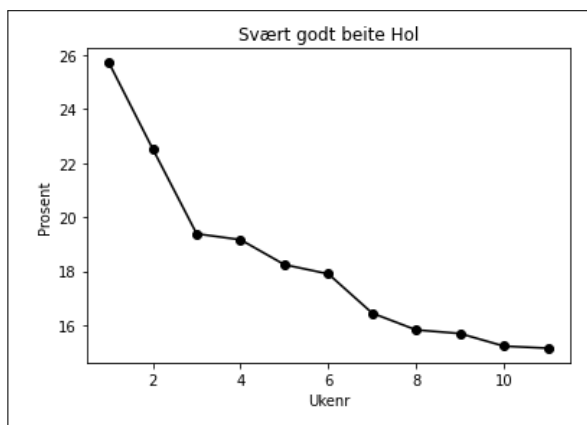


Figur 4.9: Fordeling Spekedal

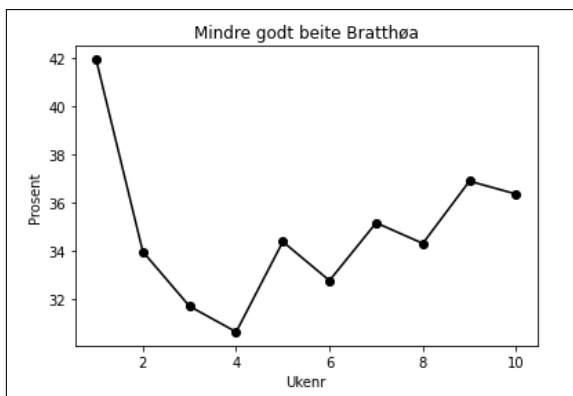
Vi ser at bruken av dårlig beite på innmark er stigende gjennom beitesesongen, og på utmark er den noe synkende. Man ser også en viss likhet mellom Spekedal og Bratthøa, brå stigning i starten, så en gradvis nedtrapping, men på Bratthøa brukes det generelt mindre av dårligere beite. At det er avvik her har nok med å gjøre at Bratthøa består av bedre beite enn Spekedal.

---

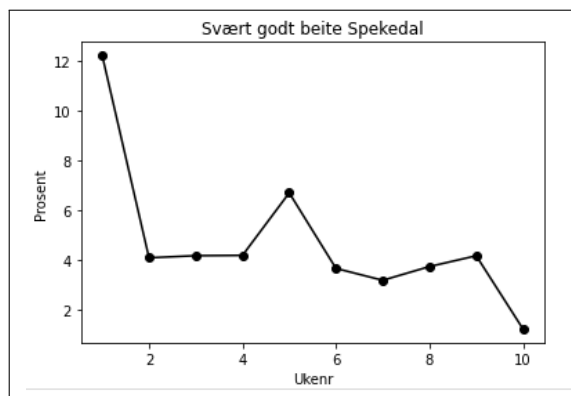
## Svært godt beite



Figur 4.10: Fordeling Hol



Figur 4.11: Fordeling Bratthøa



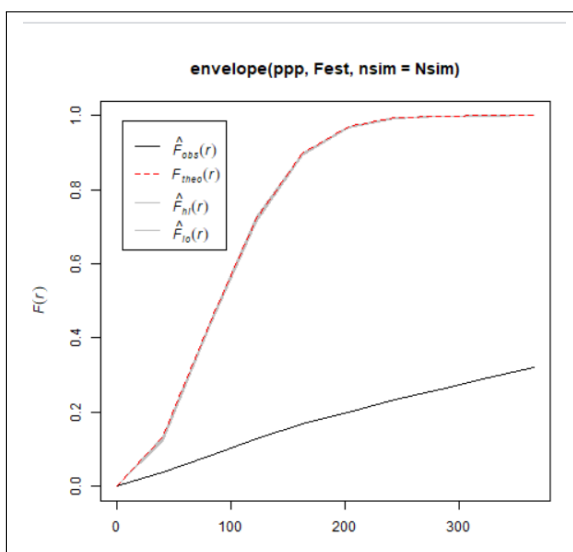
Figur 4.12: Fordeling Spekedal

Man ser at observasjoner på svært godt beite er synkende fra start for alle datasett, men spesielt Spekedal og Hol synker mye. Bratthøa har en stigning etter den første måneden. Det kan se ut som om det har vært en kamp om de beste beiteene i starten av sesongen, og denne trenden finnes hos alle områdene.

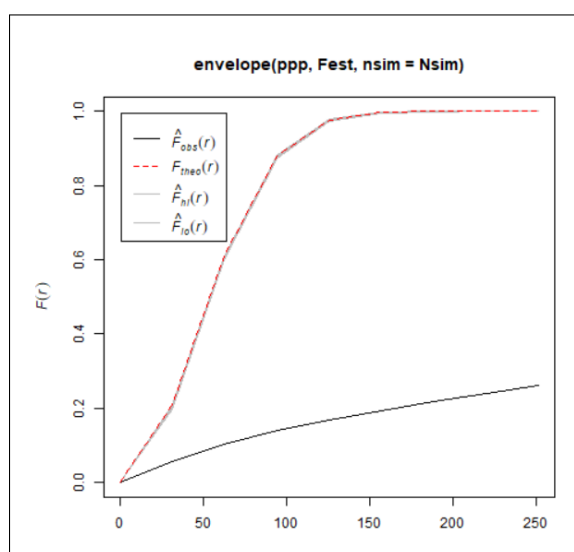


## 4.2 Avstasjonsbasert mål av punktfordeling

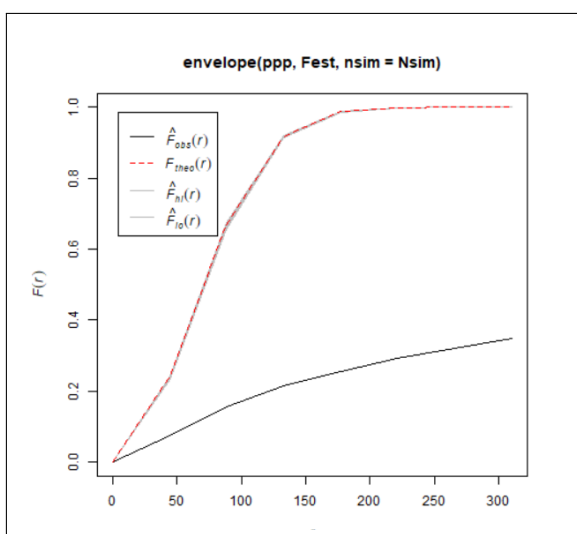
### 4.2.1 F-funksjon



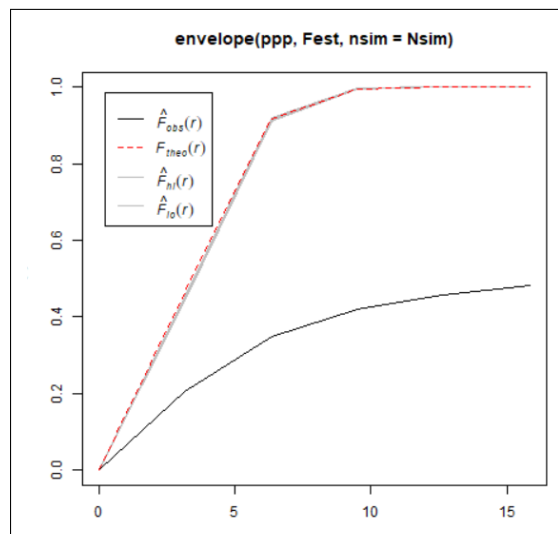
Figur 4.13: F-funksjon Bratthøa 2014



Figur 4.14: F-funksjon Bratthøa 2013



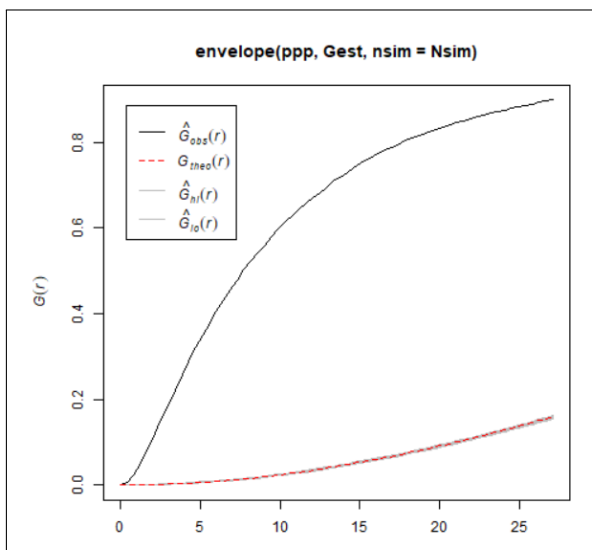
Figur 4.15: F-funksjon Spekedal 2013



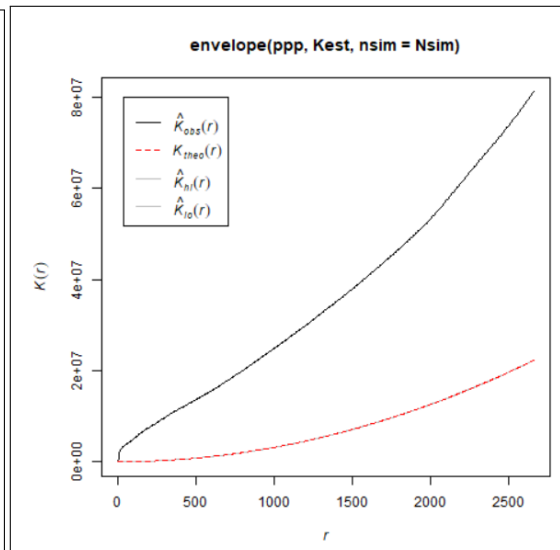
Figur 4.16: F-funksjon Hol 2012

Av figur 4.13 - 4.16 ser man at f-funksjonen for fordelingen i beiteområdet er ganske like, spesielt de to årene på Bratthøa og Spekedalen i 2013. Som vist i figurene er den sorte streken (observert) under den røde (teoretiske), noe som betyr at utnyttelse av det totale området er lite i forhold til den forventning ved tilfeldig punktplassering.

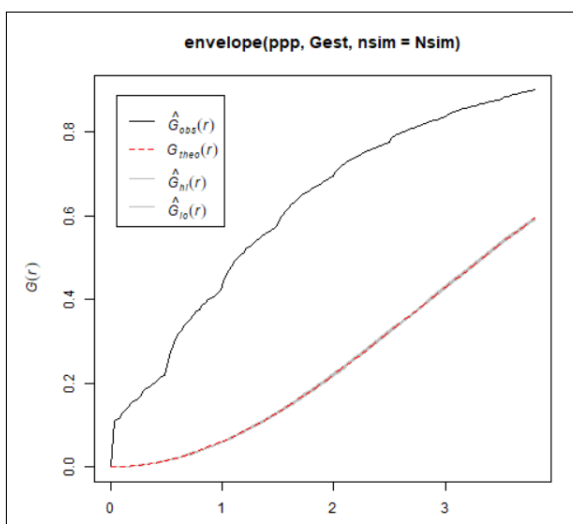
## 4.2.2 G- og K-funksjon



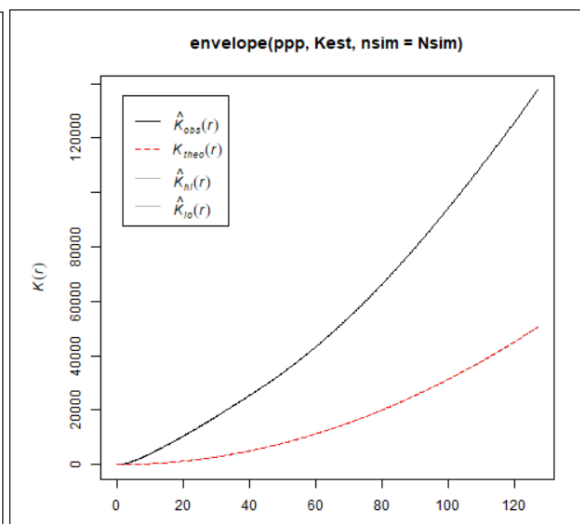
Figur 4.17: G-funksjon Spekedal 2013



Figur 4.18: K-funksjon Spekedal 2013



Figur 4.19: G-funksjon Hol 2012



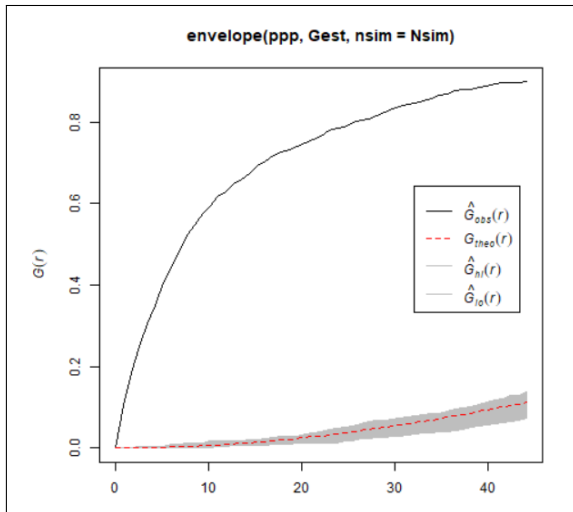
Figur 4.20: K-funksjon Hol 2012

G-funksjonene viser tydelig at avstandene mellom punkter er mye lavere enn forventet ved tilfeldig spredning. Man ser også at 80% av sauene i utmark har en nærmeste nabo på under ca. 200 meter, og på innmark er denne på ca. 30 meter.

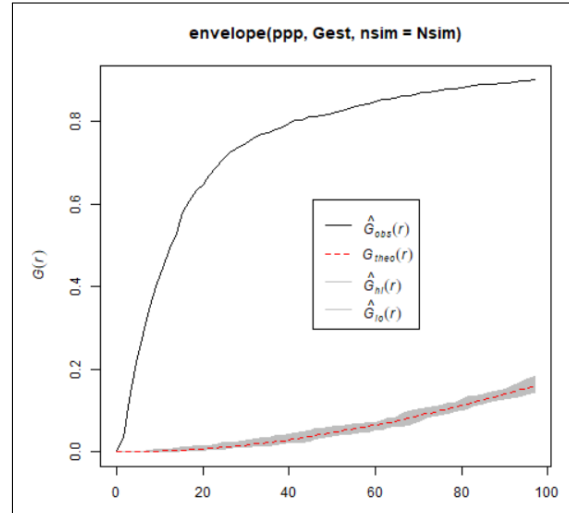
K-funksjonene viser at de fleste naboavstandene mellom punkter blir svært mange mot slutten. Det er ingen tydelige hopp, slik at fordelingen ser ut til å være jevn, avstandsmessig. Det blir også for mange innmålte punkter til å avlede noen god informasjon.

### 4.2.3 G-funksjon enkeltsauer

En representativ g-funksjon for en sau på utmarksbeite er presentert i figur 4.21. Denne viser at 60% av punktene er mindre enn 100 meter fra et annet punkt, og 80% av punktene er mindre enn 400 meter unna andre punkter. Figur 4.22 viser en sau på Bratthøa, hvor grafen skilte seg betydelig fra de andre utmarkssauene. Her finner man avstander opp til 1 kilometer, men man ser fortsatt at over 60% av punktene ligger under 300 meter fra andre punkter.

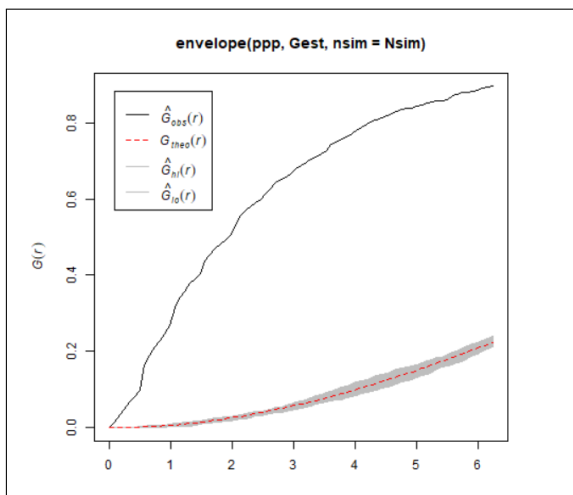


Figur 4.21: G-funksjon på sau 142 000 Spekedal

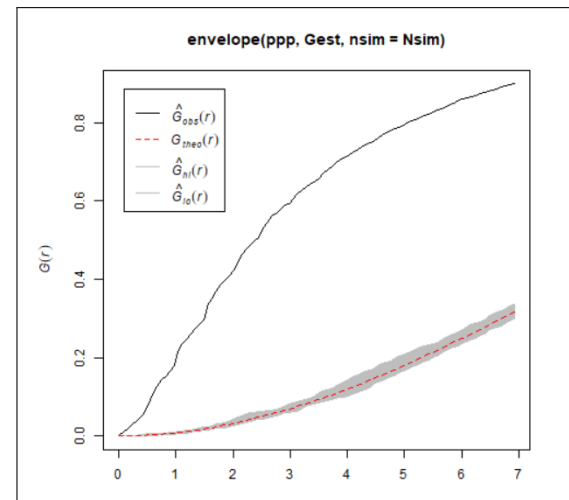


Figur 4.22: G-funksjon på sau 142 06 Bratthøa

På innmarksbeitet i Hol er det klart mindre avstander mellom punktene, og det viser seg å ikke være store forskjeller mellom de tette saueflokkene og de spredte. På innmark er 80% av punktene mindre enn 40 meter unna andre punkt.



Figur 4.23: G-funksjon på sau 13 Hol i tett innhegning

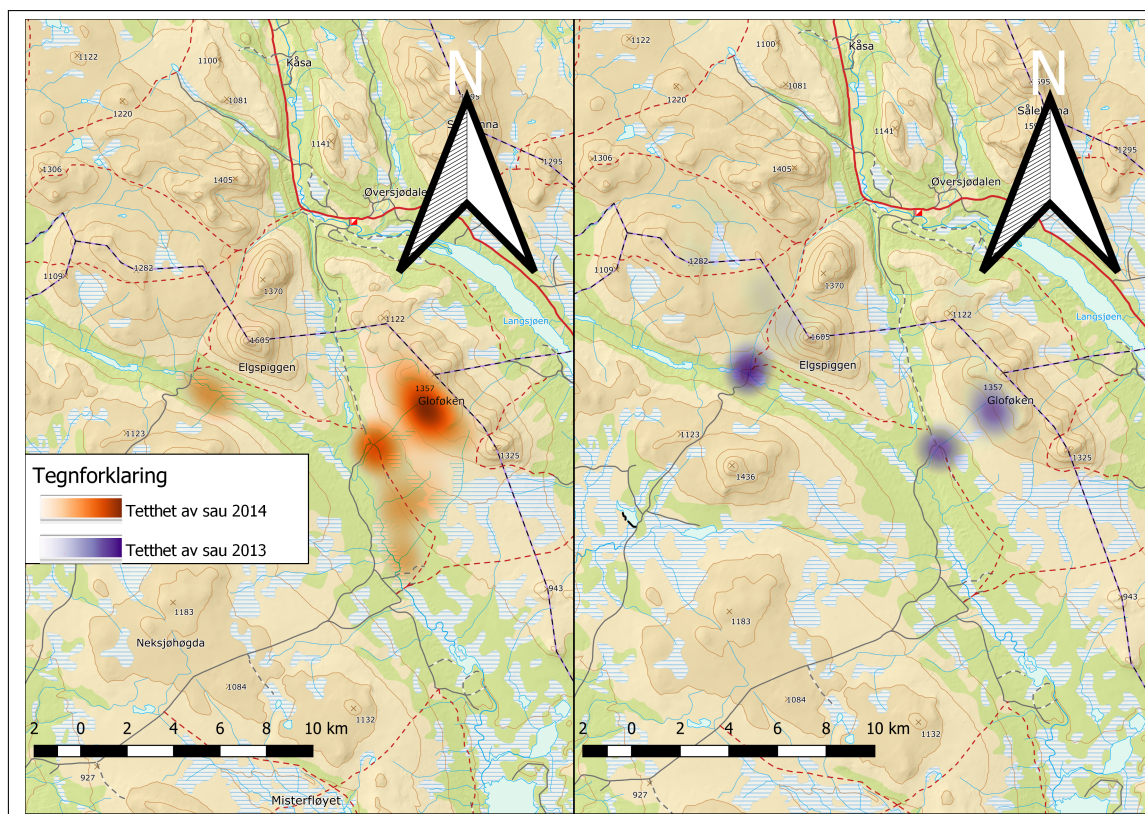


Figur 4.24: G-funksjon på sau 4 Hol i spredt innhegning

## 4.3 Hotspot

Alle de åtte kartlagene er visualisert som “Heatmap”, og de tettste områdene er nærmere undersøkt. Faktorer av interesse er når området er i bruk, hva som kjennetegner bakgrunnskartet og grunnlagskartet, og om det er likhet mellom årene. Ved å se på fart, får man også en oversikt over hva slags oppførsel sauen har hatt.

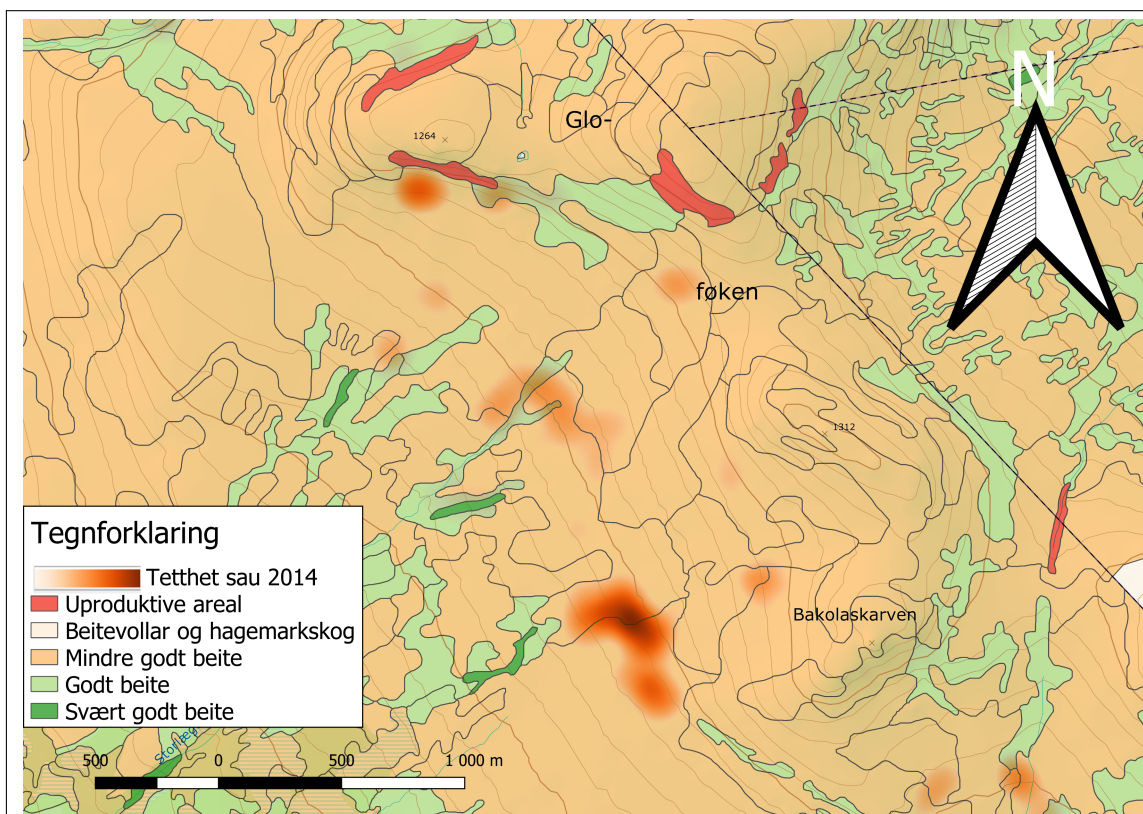
### 4.3.1 Spekedal



Figur 4.25: Hotspot-analyse Spekedal

I figur 4.24 ser man tre tydelige hovedklynger som går igjen begge år, to nede i dalen, og en oppe på fjellet. Området i vest er en seter kalt Spekedalseter, området i midten er en grend som heter Storlægda, og mot øst har vi fjellområdet Gloføken. Alle disse er av interesse og vil bli videre inspisert.

## Gloføyen



Figur 4.26: Inspeksjon av Gloføyen

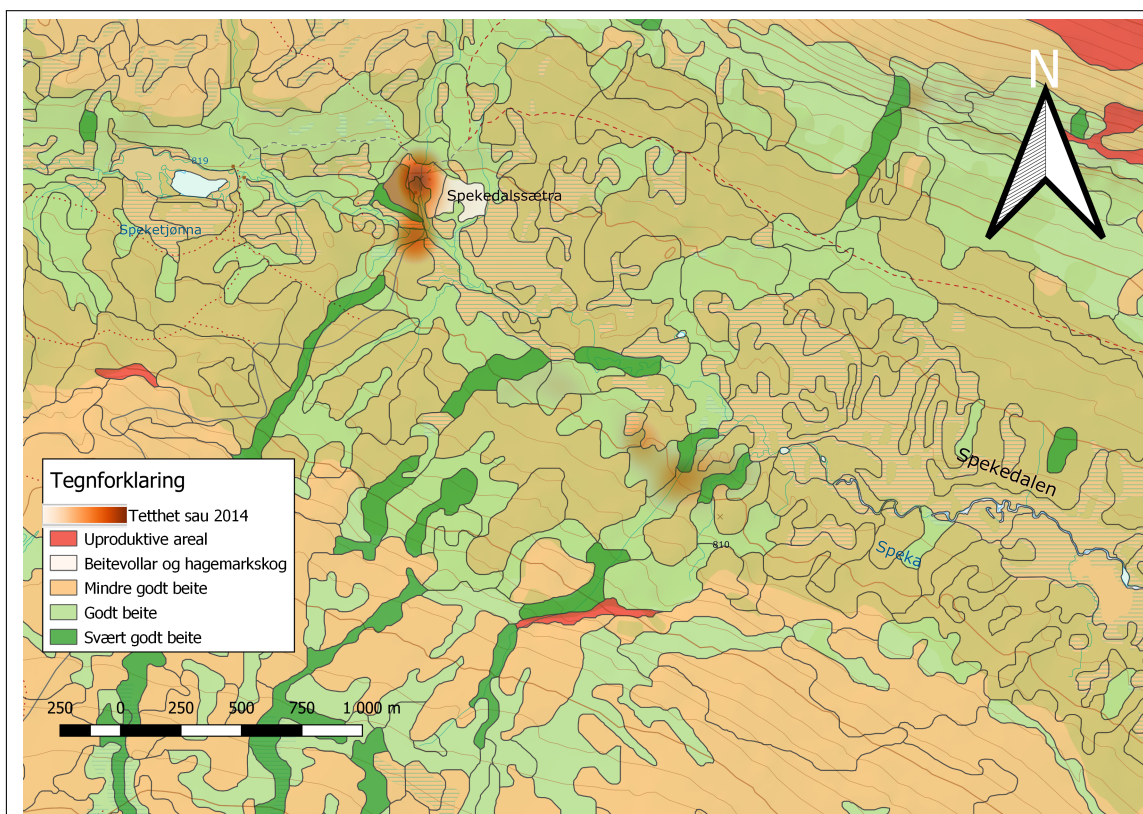
Av posisjonene logget på Gloføyen er ca. 11% godt beite, ca. 2% er svært godt, og 87% er mindre godt beite. Ser man på de tilhørende målene for farten til sauene i området, ser man at 43% er under  $0,05 \frac{km}{t}$ . Ved å sjekke fordelingen gjennom døgnet, ser man at området er mye brukt på natta og morgenen, både av NKS og Spælsau. Området benyttes altså mye som hvilested, spesielt i juli og august.

Den store trenden begge år er at området brukes hele tiden i starten og mot slutten av beitetiden. Det er mange “ankomster” om kvelden og “avreiser” på morgenen.

Ser man på høydekurvene på bakgrunnskartet er dette et høytliggende og sør-vestvendt område. Dette medfører at solen står på midt på dagen og litt utover kvelden. Det kan se ut som om sauene kommer til steinete områder hvor solen har varmet disse opp for å hvile.



## Spekedalseter



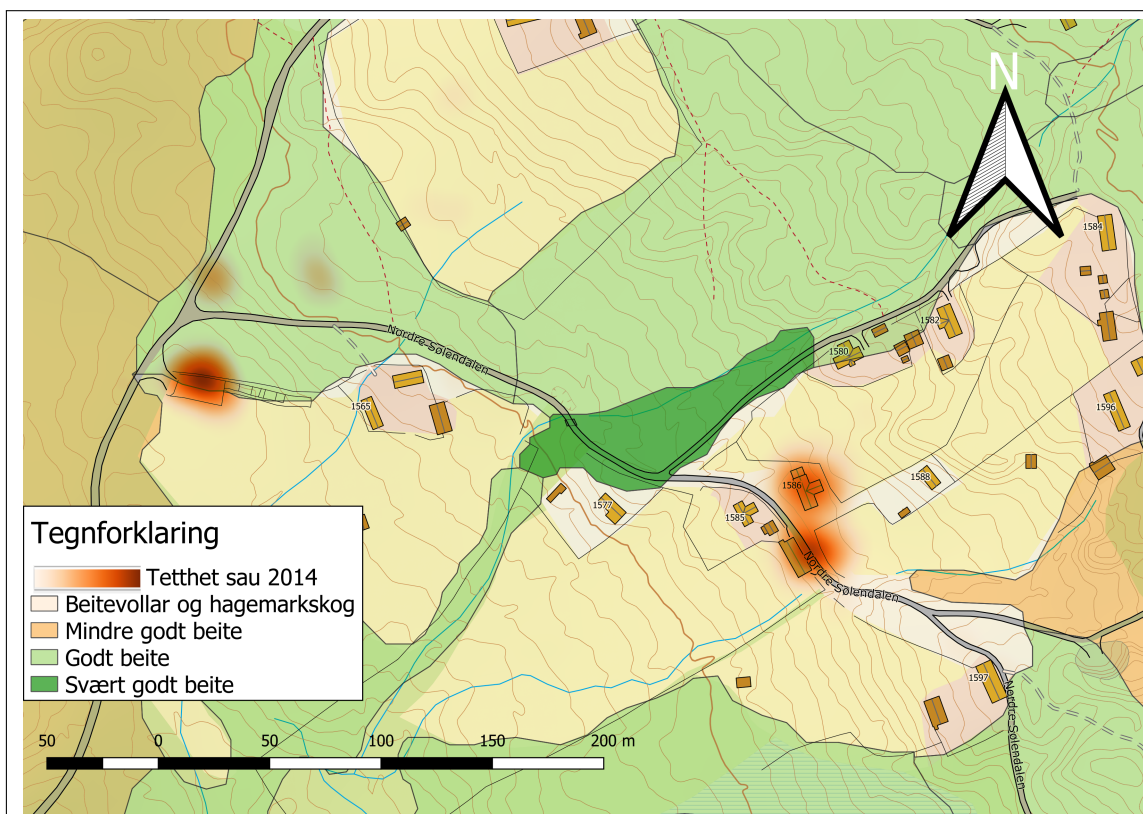
Figur 4.27: Inspeksjon av Spekedalseter

Spekedalseter er et frodig område med gode beiter. Her renner elven Speka gjennom, som også sørger for drikke til sauene. At de velger å holde til i dette området er ikke så rart.

I 2013 ser det ut som om sanking ble gjort rundt 21.-23. august, og sauene ble holdt her til 02. september med GPS-sporing på. Dette bidrar til den høye andelen av dyrka mark på den temporale utforskingen.

Ser man på tiden før sauene blir holdt igjen, indikerer det at området brukes noe på morgenen og mye til overnatting. Overnattingen er noe sporadisk, men merkbart høyere ved månedsskiftet juli-august. Denne trenden finner man begge år. Når sauene her ikke befinner seg på dyrkamark rundt setra, så er de på godt beite nede i dalen (til høyre i figur 4.26). Dette dalføret bærer altså preg av å være godt beite, men det er ikke så mye brukt til overnatting. Altså sover sauene på de åpne beitemarkene rundt setra, og beiter i skogen på morgenen.

## Storlægda

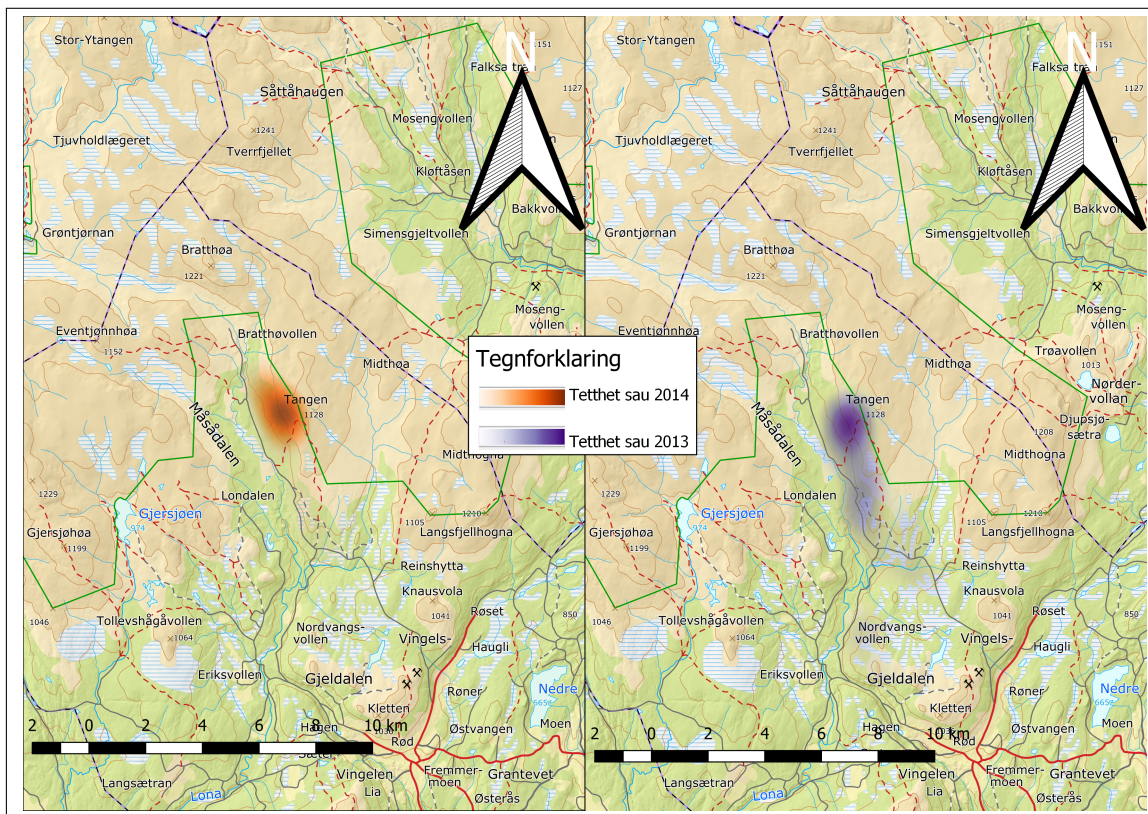


Figur 4.28: Inspeksjon av Storlægda

Ved inspeksjon av de tetteste områdene i Storlægda, finner man at sauene i dette området bruker det både natt, morgen og kveld. Det er ikke så mange lange besøk, men hyppige vandringer inn like før midnatt, og vandring ut om morgenen. Bevegelseshastighetene i disse tre områdene viser at 55% er under  $0,05 \frac{km}{t}$ , 35% er over  $0,1 \frac{km}{t}$ . Dette er et tydelig tegn på at området brukes mye til hvile.

Dette området har høy tetthet både i 2013 og 2014, og grunnen til dette er ganske klar. Det er innmark. Dette vises også av å se på grunnlagsverdiene, da 58% av punktene i området er målt på Beitevoll og hagemarkskog.

### 4.3.2 Bratthøa



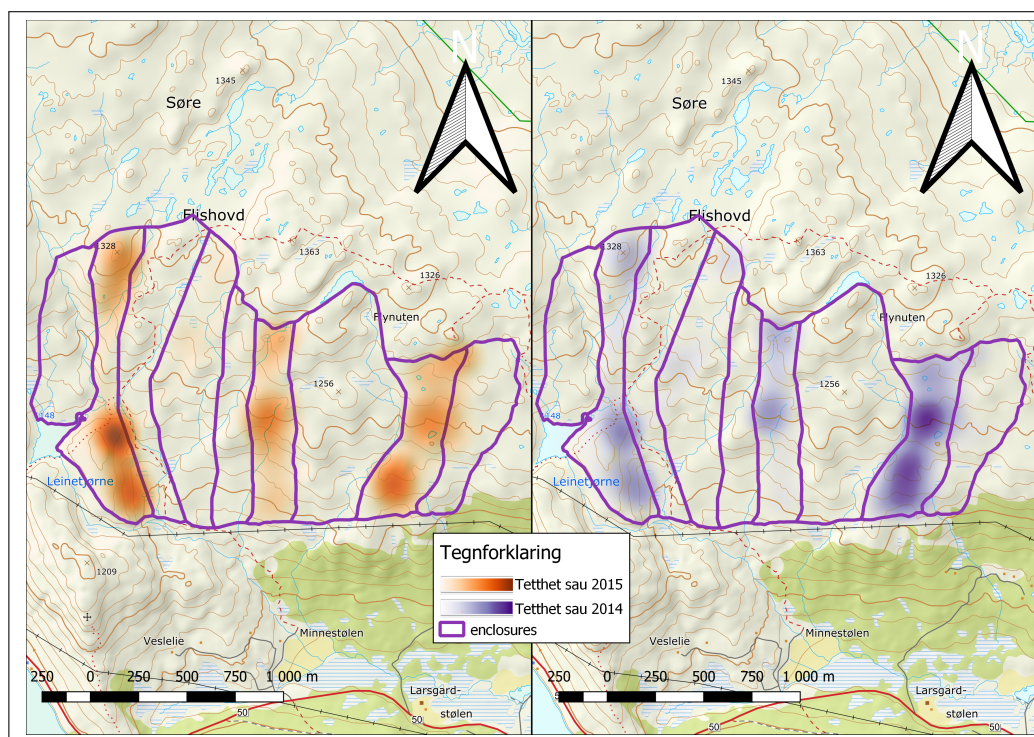
Figur 4.29: Hotspot-analyse Bratthøa

Her er det et område som skiller seg klart ut, og grunnen er at mange av sauene bruker dette om natta. Typiske trekk er at det går en dag mellom hvert besøk, men i noen tilfeller kommer de tilbake på kvelden samme dag. Flere av sauene har lengre opphold, gjerne opptil en måned. Dette er ganske tydelig en soveplass, da ca. 50% av målingene viser en fart på under  $0,05 \frac{km}{t}$ . Til motsetning er 25% over  $0,1 \frac{km}{t}$ , som kan forklares ved inn- og utvandring av hvileområdet.

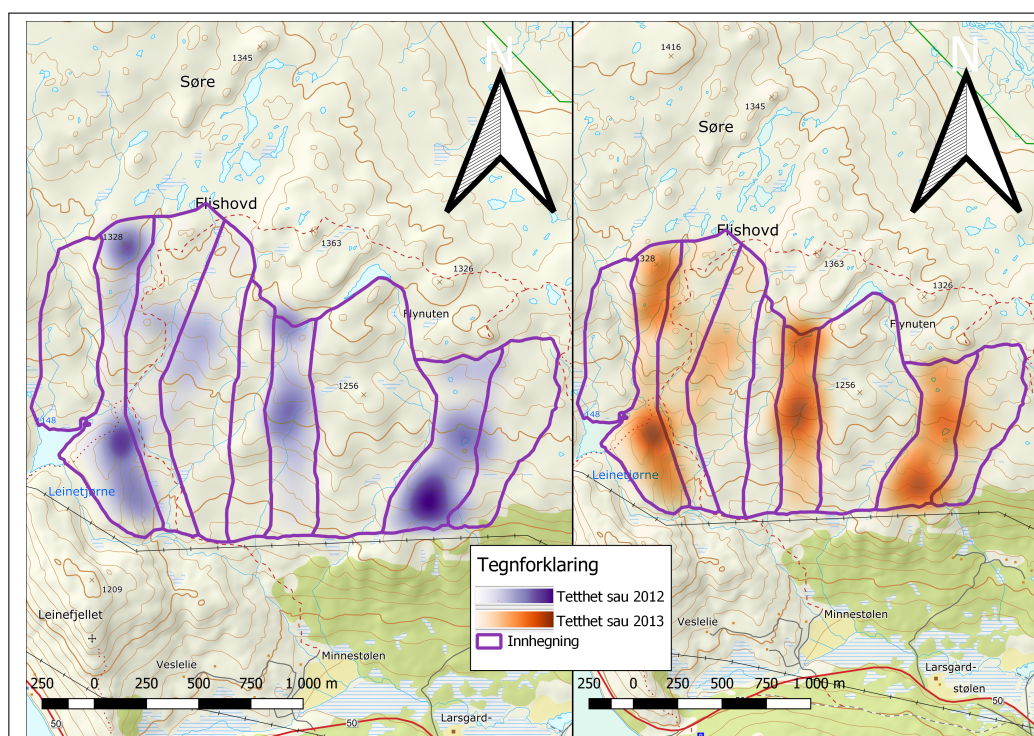
Bakgrunnskartet viser at området er hellende mot vest, og av flyfoto ser man at det er i overgangen skog til høyfjell sauene befinner seg. Ser man på beitefordelingen er 34,6% Godt beite, 46,8% Svært godt beite og 18% mindre godt beite.



### 4.3.3 Hol



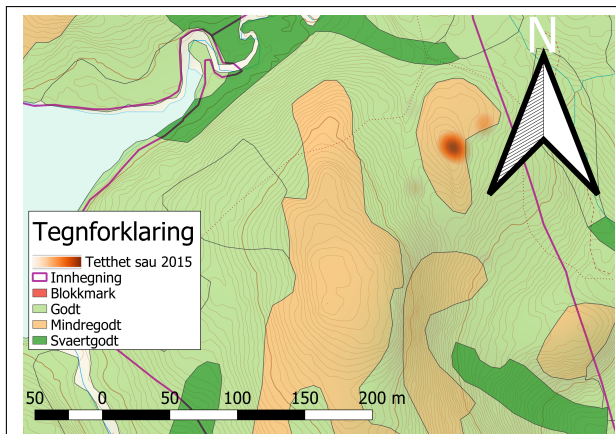
Figur 4.30: Hotspot-analyse Hol 2014 og 2015



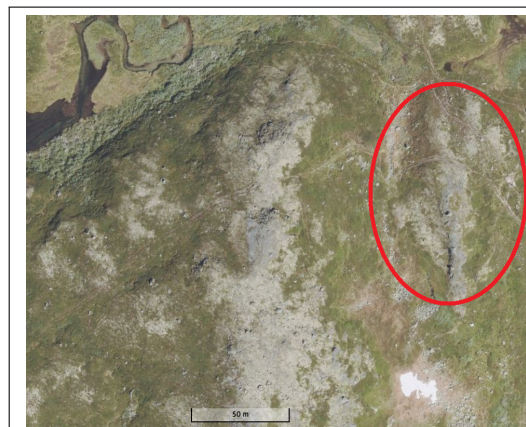
Figur 4.31: Hotspot-analyse Hol 2012 og 2013

Man ser fort av analysene at det er åtte områder av interesse. Tre i innhegning nummer to fra venstre, tre i nummer åtte fra venstre og to i nummer seks fra venstre.

Nærmere inspeksjon av disse områdene og når de brukes viser at det er overnattingssteder, da de fleste punkter er registrert på nattestid. Dette bekreftes også ved fartskolonna, hvor ca. 60% av målingene har en endring på mindre enn 0,05 km/t. Ser man på grunnlagsverdiene, så er denne i hovedsak “Mindre godt”, og noen steder “Godt”. Sammenligner man områdene med flybilder (se figur 4.31 og 4.32), ser man en gjenganger, steinur og berg. Ved analyse av bakgrunnskartet ser man helningen og flatheten, og områdene er generelt sørvendte eller ligger på en høyde.



Figur 4.32: Inspeksjon tetthet Hol



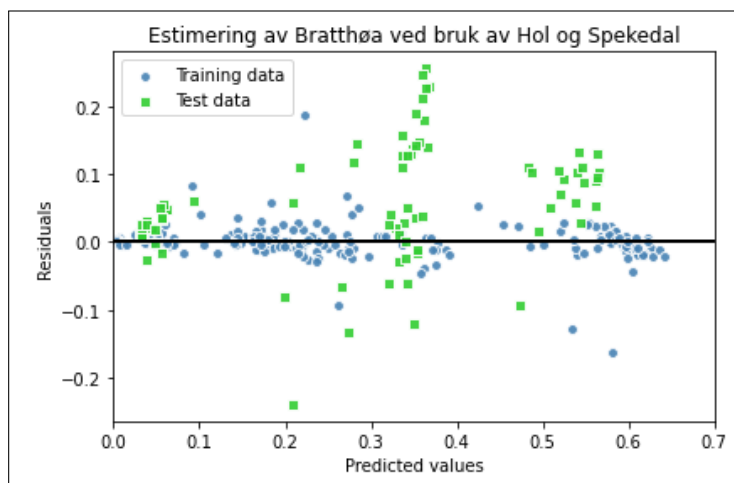
Figur 4.33: Flyfoto av Hol[3]

## 4.4 Randomforest

### Metode en

Ved prediksjon som presentert i første metode, brukes Spekedal og Hol som treningssett og Bratthøa som test. Dette gir en  $R^2$  på 0,625, altså 62,5% av variansen på Bratthøa kan forklares ved variansen i Spekedal og Hol.

Av residualplottet i figur 4.34 ser man at det er noe systematisk feilberegning rundt 0,33 og 0,51 for testdatasettet. Treningsdatasettet følger linja ganske bra.



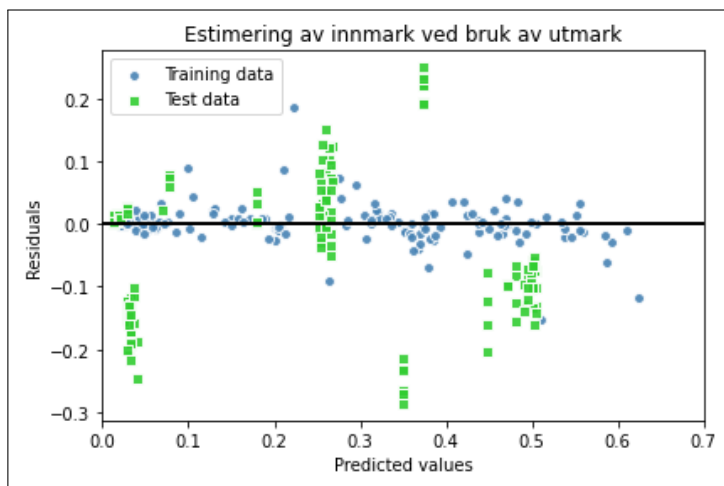
Figur 4.34: Residualplott metode en

---

## Metode to

For å sjekke om man kan finne en sammenheng mellom utmark og innmark, kan man se om innmark lar seg modellere ved hjelp av utmark. Ved å bruke Hol som testsett og Bratthøa og Spekedal som treningssett får Randomforest en  $R^2$  på 0,717.

I figur 4.35 ser man at det er mange systematiske avvik ved predikering av testdatasettet, og disse feilene ligger på rundt 20%, noe som er ganske høyt. Treningsdatasettet er ganske forklart, med bare noen mindre enkelte avvik.

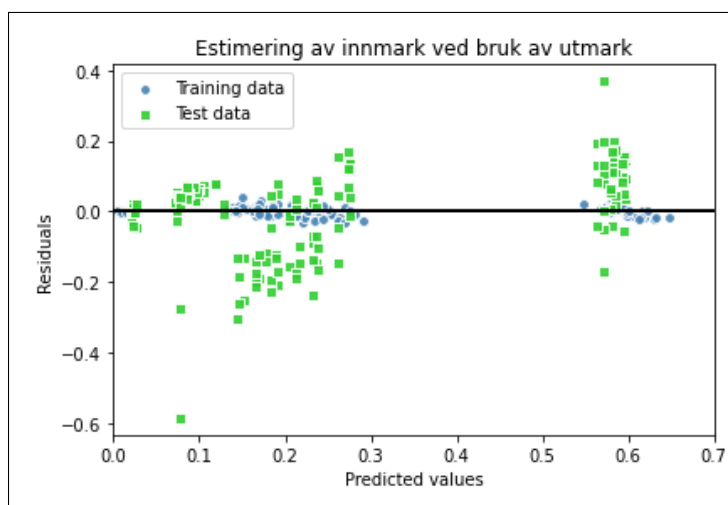


Figur 4.35: Residualplott med innmark som testdatasett

## Metode tre

Metode tre er nok en sjekk for om man kan observere sammenheng mellom utmark og innmark, men her benyttes ett innmarksområde for å modellere to utmarksområder. Her får Randomforest en  $R^2$  på 0,610. Her vil læring på tetthet være meningsløst, da det bare er en verdi for dette. Dette kan forklare at nøyaktigheten synker.

Her ser man at testdatasettet har mindre tegn til systematisk avvik enn for metode to, men avvikene man ser er høyere. Treningsdatasettet er nesten helt korrekt forklart. Dette er et tydelig tegn på overtilpassning av modellen.



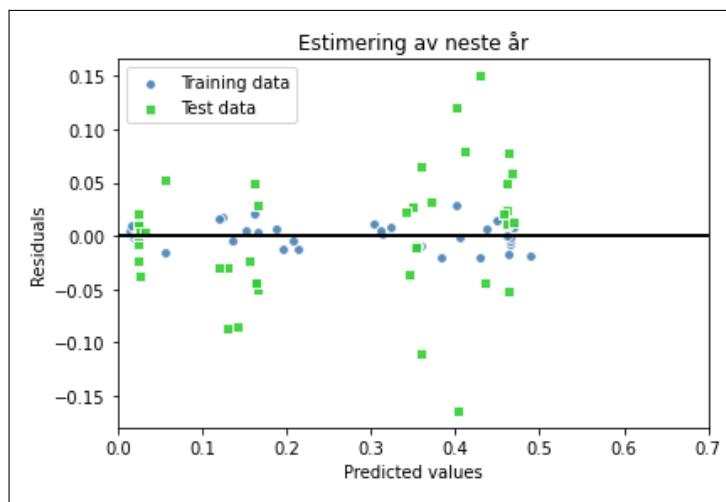
Figur 4.36: Residualplott med utmark som testdatasett

---

## Metode fire

Ved å bruke datasettet fra Bratthøa i 2013 til å modellere Bratthøa i 2014, får Randomforest en  $R^2$  på 0,875.

Her ser man spredte avvik for testdatasettet, men disse er til gjengjeld små, og på det meste 15% feil. Her ser man også at treningsdatasettet ligger tett inntil linja, som er tegn på overtilpassning.

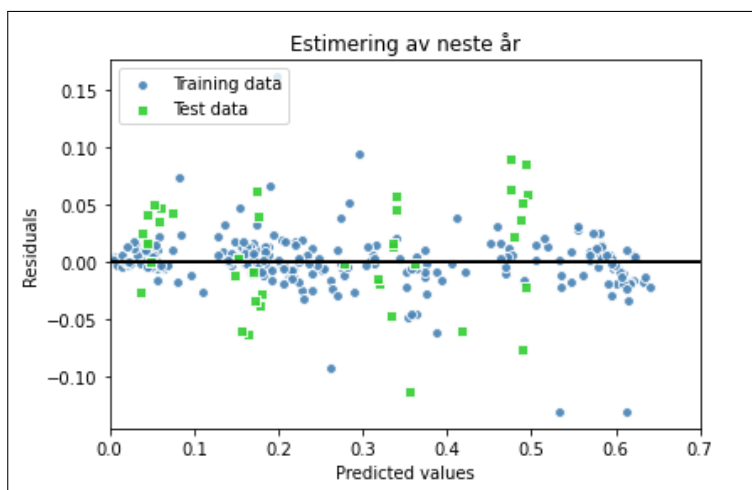


Figur 4.37: Residualplott estimering av Bratthøa 2014 ved bruk av Bratthøa 2013

## Metode fem

Ved å bruke alle treningsdataene fra Spekedal og Hol, samt 2013 fra Bratthøa, så estimerer modellen oppførsel i 2014 på Bratthøa med en  $R^2$  på 0,915. Det at resultatet øker ved bruk av andre områder, sier at det her finnes sammenheng, og tendenser som hjelper på tvers av beiteområder.

I metode fem ser man at variansen er enda mindre enn for metode fire, og man ser også at treningsdatasettet har større utslag enn testdatasettet. Dette er tegn på at modellen er god, da både test- og treningsdata har like avvik, og forklaringsgraden er høy.



Figur 4.38: Residualplott estimering av Bratthøa 2014 ved bruk av Spekedalen, Hol og Bratthøa 2013

## 5. Diskusjon

### 5.1 Observasjoner

#### Romlig fordeling

Av den romlige fordelingen ser man spesielt for Spekedal og Høl at det er områder som ikke har hatt besøk av sau. Dette er i hovedsak områder av typen uegnet og mindre godt. Mindre gode beiter består av forskjellige naturtyper. Disse kan finnes ved å utforske NIBIOs Kilden[36]. Om det skulle vise seg at sauene unngår en bestemt naturtype, vil dette si mye om hvordan eksterne faktorer påvirker sau. Med tanke på en varslingsfunksjon for unormal oppførsel, kunne varselet kommet om en sau beveger seg inn på et slik område, eller har oppholdt seg her lenge.

Av den temporale fordelingen ser man i hovedtrekk at mindre godt beite brukes mer mot slutten enn mot starten av beitetiden for innmarksbeite. En studie utført av Mobæk et al. (2009) på samme beiteområde konkluderer blant annet med at forskjellige planter blomstrer på forskjellig tid, og dette fører til den endrede bruken av beiteområde[34]. I tillegg kan man se at bruken av svært godt beite på utmark holder seg stabilt eller stiger, noe som kan tyde på at sauene finner nye områder.

### 5.2 Avstandsbasert mål av punktfordeling

#### F-funksjon

F-funksjonen blir i dette tilfellet beskrivende for hvor mye av beitet som benyttes. Denne målingen blir nok noe feil, da f-funksjonen i QGIS deler området inn i et omsluttende rektangel, og ikke etter beiteområdets utstrekning, noe som gjør at mye ekstra tomrom blir tatt med. Dette gjør at man i fordelingen ikke ser mer enn ca. 30% på utmarksområdet, og at man ser ca. 50% på innmarksområdet. F-funksjonen må nok beskrives som en for stor og omfattende funksjon til å se på sauers fordeling i beitet. Dette fordi mange tilfeldige punkter genereres i et for stort område i forhold til beitebruken.

#### K-funksjon

K-funksjonen har ingen tydelige preg av klustering, da grafen ikke har noen tydelige utslag. Området blir nok for stort, og det er for mange punkter til å kunne avlede videre informasjon fra denne. Målet var å få delt inn data i mindre grupper, og få sett på K-funksjonen per time og sammenlignet disse, men grunnet at datasettene består av ca. 60 dager à 24 timer, gir det 1440 bilder per datasett for utmark, og det dobbelte på innmark. Forsøk på å en slik iterering ble gjort med maskinlæring uten suksess.

---

## G-funksjon

G-funksjonen viser noen forskjeller på inn- og utmark. Mye av forklaringen kan ligge i at sauene på utmark har målt hver time, og kan da bevege seg lengre mellom hver måling, enn sauene på innmark som har målt hvert 30. minutt. Mange punkter vil nok ha forrige måling som nærmeste nabo, noe som betyr at de er avhengige av hverandre, og denne avhengigheten vil vises tydeligere ved hyppigere målinger.

Det g-funksjonen sier er at for sauer på utmark, er svært få målinger lengre unna andre punkter enn 300 meter. For innmark gjelder dette for 40 meter. Observerte avstander lengre enn dette vil altså være av mulig interesse. Man bør kanskje også se på g-funksjonen midt i beitesesongen eller på et mindre tidsrom, da dette kan bestå av et mer utforskende mønster, med mulige lengre avstander.

## Enkeltsauer

Bruk av g-funksjon på en sau viser seg å forklare mye. Man ser at de aller fleste punkter har korte nabo-avstander. Disse er ikke av særlig interesse, der er derimot de lengste avstandene av stor interesse. Disse kan si noe om det lengste en sau har beveget seg unna sin "komfortsone". Ved inspeksjon av de høye verdiene hos sau 142 06 i figur 4.21 kunne man se at sauen har en forflytting på 31 km på fem timer. Disse punktene ligger langs vei, altså kan man anta at sauen er fraktet med bil. Dette er varsler som en bonde burde fått hvis han ikke visste om forflyttingen.

## Temporal inndeling

Ett av formålene med disse analysene var å sammenligne forskjellige uker, dager og timer med hverandre, og sett på grafen for disse mindre tidsrommene. Kunne man sett at funksjonene fordelte seg ganske likt over tid, kunne man sagt at det finnes en normal. Håpet var å få iterert seg gjennom dette i ArcGIS ModelBuilder, men dette har ikke latt seg gjøre.

## 5.3 Hotspot- og grunnlagsanalyse

Hotspot-analysen viser tydelig interessante områder. De stedene hvor sauene befinner seg mye, sier mye om oppførselen. Fordelen ved denne analysen er at man får funnet de områdene med lite bevegelse, altså hvileområder da disse vil ha svært mange punkter. Mobæk et al.[34] viser at inaktiv tid finnes mest på mindre godt beite, og aktiv tid er mest på svært godt beite. Det at sauene hviler på de dårligere beitetypene stemmer godt overens med resultatene i denne analysen. Mobæk sin studie er fra samme område på Hol, og her beskrives det mindre gode området som områder med mose, snø, lavhei, myr og steinete landskap. Av flyfoto ser det ut som om hvileområder ligger på steinete landskap, og dette ville det vært interessant å sett videre på.

En ulempe ved analysen er at fart, tidspunkt, grunnlag m.m., må sjekkes manuelt, noe som er svært tidkrevende. Skulle mange flere områder blitt analysert, burde en god produksjonsløype for å automatisere dette bli satt opp.

## 5.4 Randomforest

Ved test av fordelingen av områdetyper mot sauenes fordeling får man forklart 75% av variasjonen. Dette er den viktigste faktoren for hvordan sauene fordeler seg i beitet. Denne verdien sier også at om modellen skal være god, må den altså predikere bedre enn dette. Det er det kun metode fire og fem som gjør, og metode en, to og tre gjør det dårligere enn ved å kun bruke grunnlagsfordelingen. Dette viser at å lære på en type område, for så å predikere en ny ikke er en god fremgangsmåte.



---

Man ser at det er en avvikende trend mellom inn- og utmark, da man forklarer ca. 4% mindre av innmark ved bruk av utmark, og 14% mindre av utmark ved bruk av innmark.

Metode fire og fem viser at man kan forklare mye ved bruk av et tidligere datasett fra beiteområdet. Her forklares opptil 12% mer av variansen. Dette kan også være et tegn på overtilpassning, altså at modellen estimerer lik oppførsel år etter år, for når man legger til data fra andre beiter, blir resultatet faktisk bedre. Dette kan forklares av at modellen generaliseres, og når den generaliserte modellen er bedre enn den spesialiserte modellen, viser det at flokkoppførselen til sauer faktisk kan predikerbar ved hjelp av andre områder.

Negative sider ved modellen er at den kun benytter tre forskjellige områder i læringen. Dette kan gjøre at modellen blir overtilpasset ved at man bruker samme område til læring og testing. Det ville fortsatt vært nyttig å se på flere datasett, for å se om modellen blir bedre av variasjon i datasett.

Modellen ser på et begrenset antall rader, og man kunne valgt å sortere etter mindre tid enn uker, f.eks. dager, 12 timer for natt og dag, eller 8 timer, hvor dagene er delt i natt, formiddag og ettermiddag.

## Hovedfunn ved Randomforest

Tabell 5.1: Resultater Randomforest

Metode nr.	Andel testdata	Andel treningsdata	$R^2$	Antall variabler benyttet
1	25,3%	74,7%	0,625	6
2	49,4%	50,6%	0,717	5
3	50,6%	49,4%	0,610	4
4	50%	50%	0,875	4
5	87,3%	12,7%	0,915	6

Modellen viser at sammenhengen i sauenes fordeling på inn- og utmark er lavere enn sammenhengen gitt av beitetypfordelingen. Den viser også at tidligere års oppførsel er den beste metoden å predikere oppførselen på, men ved å lære av andre områder tilføres modellen variasjon, slik at den predikerer bedre.

## 5.5 Datagrunnlag

### 5.5.1 Problematikk datainnlesning

Innlesningen av filene gav noen feilmeldinger, hvor avvik ble oppdaget. Flere rader var ufullstendige, da gjerne i starten av måleperioden. Disse og unøyaktige målinger (høy DOP) ble tatt bort. Slike målinger vil man komme borti ved senere innhenting av datasett. Dette er både en fordel og en ulempe, da de tidligste målingene ofte vil være støygivende i en modell fordi de beskriver utslipp. Atferd ved utslipp er svært annerledes enn resten av beitetiden. Men med en god modell, kunne man kanskje også beregnet denne oppførselen også.

Det var flere feil som ble oppdaget i datasettet, blant annet at posisjonen til en sau var registrert på forskjellige steder på nøyaktig samme tidspunkt. Feilen her ligger nok i at et forsøk fra 2015 ville se på høyfrekvent sporing av sau, slik at posisjon ble registrert flere ganger samme minutt (se figur 5.1). Når slike feil dukket opp, ble det siste av de to punktene slettet.

98621	29721,313733830116689;1233;1,200000000000000;0;0;5;30.08.2015	03.06.00;400;1;1
98622	29721,761815349571407;1234;1,200000000000000;0;0;5;30.08.2015	03.06.00;400;1;1
98623	29721,306795329786837;1232;1,200000000000000;0;0;5;30.08.2015	03.07.00;400;1;1
98624	29721,778294559568167;1237;1,200000000000000;0;0;5;30.08.2015	03.08.00;400;1;1
98625	29720,638575970195234;1230;1,200000000000000;0;0;5;30.08.2015	03.09.00;400;1;1
98626	29718,392097479663789;1230;1,200000000000000;0;0;5;30.08.2015	03.10.00;400;1;1
98627	29718,711463320069015;1226;1,100000000000000;0;0;5;30.08.2015	03.11.00;400;1;1
98628	29720,626433709636331;1235;1,100000000000000;0;0;5;30.08.2015	03.12.00;400;1;1
98629	29711,666540290229023;1207;0,900000000000000;4;6;5;30.08.2015	03.13.00;400;1;1
98630	29719,518805880099535;1232;0,900000000000000;2;0;5;30.08.2015	03.14.00;400;1;1
98631	29720,639443280175328;1229;0,900000000000000;0;0;5;30.08.2015	03.14.00;400;1;1
98632	29720,068716949783266;1232;0,800000000000000;0;0;5;30.08.2015	03.15.00;400;1;1
98633	29720,086063059978187;1233;0,900000000000000;0;0;5;30.08.2015	03.16.00;400;1;1
98634	29721,739265349693596;1236;0,900000000000000;0;0;5;30.08.2015	03.17.00;400;1;1
98635	29721,287714719772339;1232;0,900000000000000;0;0;5;30.08.2015	03.18.00;400;1;1
98636	29717,948352489620447;1227;1,000000000000000;1;1;5;30.08.2015	03.19.00;400;1;1

Figur 5.1: Endring i samplingfrekvens hos GPS-mottakeren

Siden Hol er innmarksbeite, med inngjerda område, blir analysen av flokkoppførselen vanskelig når de bryter ut. Dette fordi at oppførselen definert som innmarksoppførsel blir for lik utmarksoppførsel. Dette gjør det vanskelig å finne ut om det er forskjeller i oppførselen på de to beitetypene. En annen fremgang enn å slette punktene, kan være å sette punkter utenfor det gitte innmarksområdet til utmarkspunkter, og dermed vil oppførselen bli tolket likt. Man kan også se på om det er mulig å benytte et virtuelt gjerde, som gir varsel til sauene eller bonden når sauene går utenfor det bestemte området.

## 5.5.2 Målemetode

### Followit

Teimori (2018)[55] fant at så lenge steglengden mellom hvert punkt er separerbare fra feilmålinger, er segmentering av sauens spor mulig. Teimouri sier også at ved høy frekvens på målingene vil farten bli overestimert på grunn av målefeil, og ved lav frekvens vil farten bli underestimert på grunn av at sporet blir generalisert. Det kommer altså an på hvilken type studie man ønsker, når man velger målefrekvens på GPSen. Ved høye frekvenser får man tydelig speart klassene beiting, hvile og vandring, og ved lav frekvens får man muligens klassifisert aktiv og inaktiv tid.

Followit oppgir en måle- og batterikapasitet på tolv målinger per dag i to år[21]. I tabell 5.2 er hyppigere frekvensers batterikapasitet estimert.

Tabell 5.2: Estimert batterikapasitet ved ulike frekvenser

Frekvens	Batteritid
Annenhver time	2 år
Hver time	1 år
Hvert 30. min	6 mnd
Hvert 15. min	3 mnd
Hvert 10. min	2 mnd
Hvert 5. min	1 mnd

Man ser av tabell 5.2 at målingene på Hol kan vare i ca. to måneder etter innsanking, og målingene fra Nord-Østerdalen kunne vart i opptil åtte måneder til. Siden man får avledet mest informasjon fra flest målinger, ville jeg anbefalt å bruke en frekvens på hvert 30. minutt. Dette fordi at man da har ca. to måneder med ekstra for å være sikker på at batteriet holder. På innmark er det mulig å samle sauer oftere, så her kunne man forsøkt med hyppigere målinger, f.eks. hvert 10. minutt, og fått forskjellene fra måned til måned.

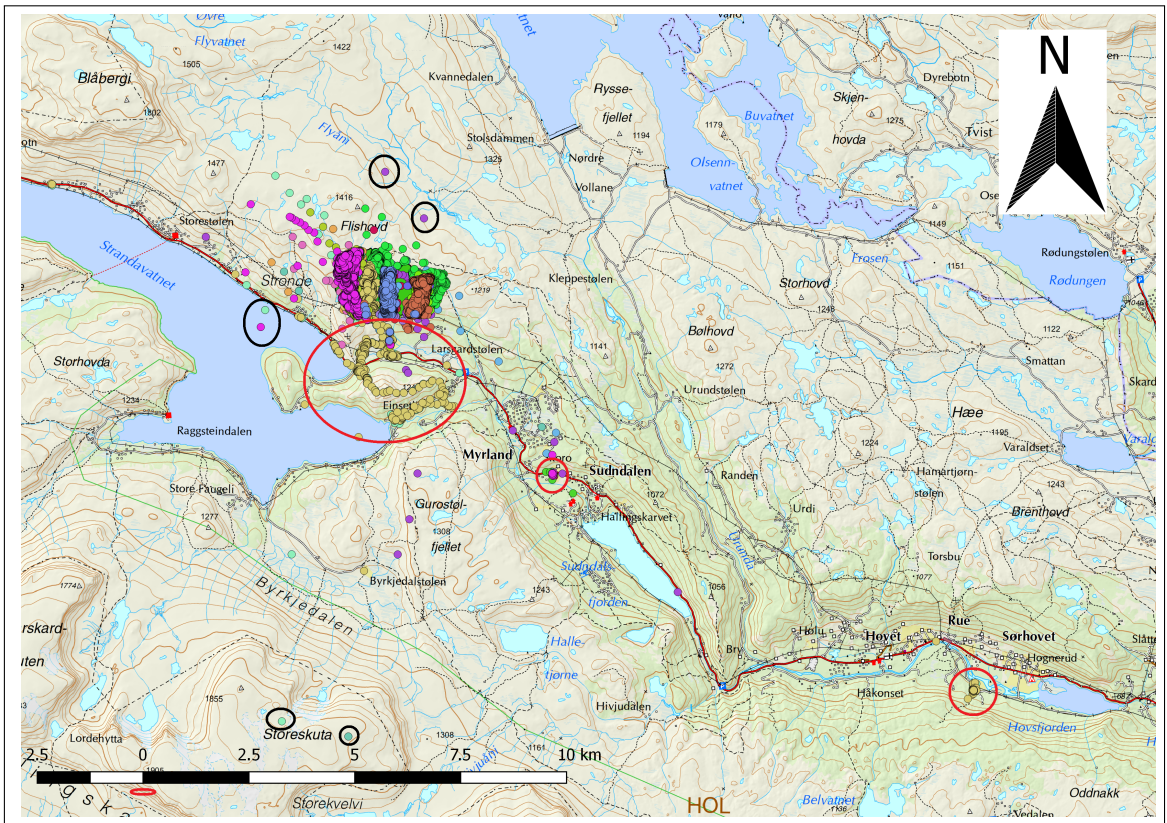


---

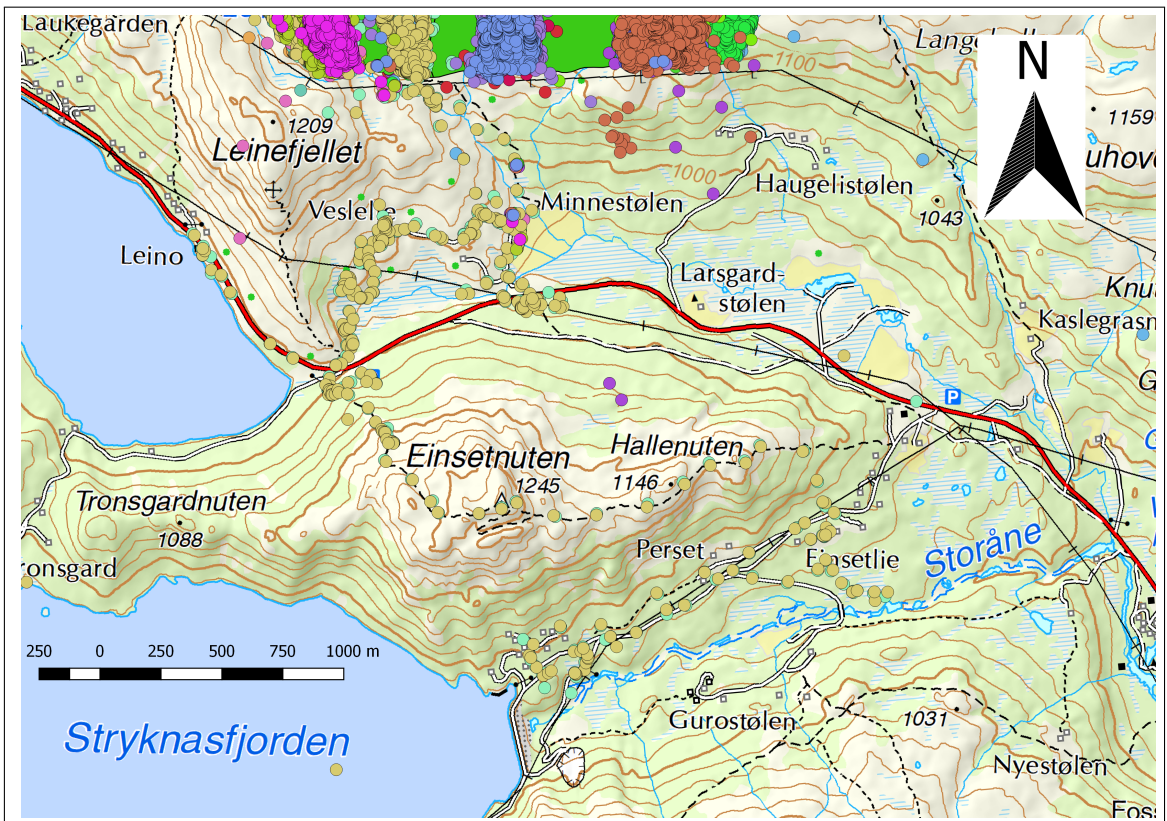
## Unøyaktighet

Unøyaktighet kan komme av flere ting, og noe av dette avhenger av målingene. En del punkter er tatt ut på grunn av for høy avstand fra studieområdet (Hol) eller for høy DOP.

Selv om en god del feilkilder ble sortert ut, ser man i figur 5.2 flere punkter utenfor studieområdet på Hol. Noen av disse er markert med sort og rødt. De sorte ringene markerer punkter som er tydelige feilmålinger, da de ligger langt unna det originale sporet eller befinner seg på steder som det ikke er mulig å befinne seg (ref. punktet på vannet). De røde ringene viser områder hvor det ser ut som om sauene har brutt ut av innhegningen, og vandret fritt avsted. Forstørrer man det største området, kan man se hva som har foregått. Området er et fjell kalt Einsetnuten, som vist i figur 5.3. Her ser man et tydelig spor av to sauer som har fulgt stien over fjellet. Ser man på den røde ringen nede til høyre i figur 5.2 er tilsynelatende samme sauer hentet hjem igjen.



Figur 5.2: Punktmønsteret fra målingene hos de 18 sauene på Hol i 2015



Figur 5.3: Inspeksjon av vandrerute på Einsenuten

## 6. Konklusjoner og anbefalinger

### 6.1 Konklusjon

Fra resultatene kommer det frem at man kan finne tendenser for normalsituasjon hos en saueflokk. De tre metodene presentert for å besvare problemstillingene, tilsier det finnes både indre parametre og ytre tegn som peker på at man kan se en normaltilstand hos saueflokker. Noe av oppførselen kan sies å være lik på inn- og utmark, og noe er ulikt. De ytre tegn er i hovedsak like for begge typer beite. Dette styrker påstanden om at man kan lage en standardisert modell for saueflokkes bevegelsesmønstre.

#### Avstandsbasert mål av punktfordeling

Avstandsbasert mål av punktfordeling har vist seg å være en dårlig fremgangsmetode, i allefall ved bruk på hele datasettet. F-funksjonen viser at store deler av områdene ikke blir tatt med i beregningen (under 70% for utmark og under 50% for innmark). Man kan til gjengjeld se at utstrekningen på inn- og utmark ikke er lik. G-funksjonen viser at innmarksfordelingen blir for tett til å kunne sammenlignes med utmark, men at innmark og utmark er like hver for seg. Nærmeste naboavstand er altså ikke lik for begge typer beite, men hver beitetype har tydelige tegn på en forventning. Avstand mellom målinger i en saueflokk er altså en tydelig indre parameter som viser tegn til likhet, men forventningen er altså forskjellig på inn- og utmark.

G-funksjonen på enkeltsauer viser hvor langt en sau er villig til å bevege seg ut av sin “komfortsone”. Analysen viser hvilke avstander som tolkes som unaturlig lange, men dette gjelder i hovedsak utmark. Denne testen styrker at avstand mellom målingene har en normal forventning, og man kan også se at det er lett å finne avvik fra normalen.

#### Hovedfunn ved hotspot- og grunnlagsanalyse

Med hotspot-analysen er de mest besøkte områdene tydelig visualisert. Fellestrekk for disse stedene over de forskjellige årene er funnet, og hovedfunnene er som følger:

- Populære hvilesteder for sau er typisk i sør/sør-vestvendte helninger, både for inn- og utmark
- Sauer hviler i hovedsak på beitetypen klassifisert som “Mindre Godt”
- Av flyfoto kan man se at hvilesteder har mye stein og berg
- Mange sauer trekker tilbake til gård og innmark, da dette er åpne områder med godt beite
- Hvilesteder i utmark ligger ofte over tregrensa, eller steder uten trær
- Man kan se tendenser til lik oppførsel i valg av hvileområde hos både Spælsau og NKS
- Man kan se at saueflokker av forskjellig tetthet velger samme type hvileområder
- Man kan se samme trend i valg av hvileområde hos sauer på inn- og utmark

---

Hotspot- og grunnlagsanalysen viser sterke tendenser til lik oppførsel på inn- og utmark. Her ved at tydelige ytre tegn når sauene velger hvilested. Disse tegnene er dårligere beiter, helst sørvest-vendt og med steinete underlag.

## Randomforest

Randomforest-analysen viser at det å bruke flere typer beite for læring, bidrar til å generalisere modellen, som igjen sørger for et mer nøyaktig resultat. Ved at modellen faktisk har mulighet til å lære av forskjellige områder, kan man se at det finnes en sammenheng. Denne gjelder både for inn- og utmark, og modellen bruker kun ytre tegn, noe som styrker at dette har en normal tendens.

## Konklusjonen

Konklusjonen er at det finnes både indre parametere som avstand innad i flokken, og ytre tegn som hvile på steinete områder, som viser seg å ha en forventet normal oppførsel for sauer på beite. Disse tendensene kan man altså måle seg frem til ved blant annet bruk av g-funksjon eller “heatmap” i QGIS, eller ved bruk av RandomForestRegressor i maskinlæringen. Oppførselen er synlig både på inn- og utmark, selv om forventningen er forskjellig.

## 6.2 Anbefalinger

### Avstandsbasert mål av punktfordeling

Sauene har en maksimal avstand de kan gå mellom målinger, uten at denne avstanden blir observert. Denne avstanden gis ved g-funksjonen, da denne sier de lengste avstandene som er målt, og hva sannsynligheten for å se slike avstander er. Dette kan man bruke med forskjellige sikkerhetsmarginer, som g-funksjonens verdi ved 99% eller 95% som eksempel. Ved å legge på denne avstanden som buffer rundt punktdatasettet, får man beiteområdets utstrekning, og man kan også måle hvor mange punkter som har godt eller svært godt beite innenfor denne avstanden.

Man kan også se på resultatet fra g-funksjonen for hvert enkelt dyr, og sammenstille disse. Her kan man få ut forventning ved forskjellige sannsynligheter, og ved bruk av variasjonen i observasjonene, kan man sette opp et konfidensintervall. Er de observerte avstandene for en sau mye mindre eller større enn dette intervallet, kan man anta at noe unormalt, som rovdyrangrep eller sykdom har intruffet.

### Tetthetsbasert mål

En videreutvikling av tetthetsbasert mål av punktfordelingen er å dele datasettet ved en minimumsfart, og/eller dag-/nattestid, slik at man får skilt ut hvileområder. Ved hotspot-analyse på dette avledete datasettet kan man da oppdage mye besøkte beiteområder.

### Grunnlagsanalyse

Grunnlagsanalysen har vist at maskinlæring for å finne sammenheng i beitebruk er nyttig. Dette kan man jobbe videre på med å se på koblingen mellom tid på døgnet og fordelingen av beitetype. Dette for å finne ut om sauer har en tendens til å sove i mindre godt beite, og beite i svært godt beite for eksempel. Med satellittbilder kan man finne signaturer i de mest populære områdene, etablere om det er en trend, og lage en estimering ut fra hvor mye slike områder brukes. I tillegg har NIBIOs Kilden[36] offentlige grunnlagsdatasett, hvor de beskriver naturtyper, slik at man kan se på om eksempelvis heisamfunn eller lauvskog brukes til noe spesielt. Dette er en større oppdeling

---

av beiteslag enn den presentert i oppgaven, og kan gi verdifull informasjon hvis tendenser kan observeres.

En annen faktor som er observert er helning. Denne kan man beregne ved bruk av Norges digitale høydemodell [2]. Ved å se på punktfordelingen i de forskjellige helningene, får man en sannsynlighet for hvor mye disse områdene benyttes. Denne sannsynligheten kan legges som et kartlag over et nytt område, og viser muligens hvor sannsynlige hvileområder for sau befinner seg.

Det er mange ytre faktorer å se på. Her tilbyr Movebank [35] åpne løsninger for topografi, atmosfære, hydrologi, barimetri, regnnivå og landdekke for å nevne noe. Her er det mye data som kan kobles opp mot oppførselen til sauene. Et forslag til videre forskning er å bruke Prinsipalkomponentanalyse, en maskinlæringsalgoritme som finner de verdiene med størst variasjon i datasettet. De verdiene som varierer minst kan tenkes å være de som er stabile ved en normalsituasjon, og ustabile ved en unormal situasjon.

# Bibliografi

- [1] General python faq. <https://docs.python.org/3/faq/general.html#what-is-python> (lest 19.05.2021).
- [2] Hoydedata. <https://hoydedata.no/LaserInnsyn> (lest 25.05.2021).
- [3] Norgeskart. <https://www.norgeskart.no> (hentet 14.05.2021).
- [4] Rovbase: Erstatning for sau. <https://www.rovbase.no/erstatning/sau/> (lest 28.01.2021).
- [5] Animalia. Sauens atferd. <https://www.animalia.no/no/Dyr/sau/velferd-hos-sau/sauens-atferd/> (lest: 28.01.2021).
- [6] David E. Ausband, Michael S. Mitchell, Kevin Doherty, Peter Zager, Curt M. Mack, and Jim Holyan. Surveying Predicted Rendezvous Sites to Monitor Gray Wolf Populations. *Journal of Wildlife Management*, 74(5):1043 – 1049, 2010.
- [7] Jamie Barwick, David William Lamb, Robin Dobos, Mitchell Welch, Derek Schneider, and Mark Trotter. Identifying sheep activity from tri-axial acceleration signals using a moving window classification model. *Remote Sensing*, 12(4), 2020.
- [8] Clément Calenge, Stéphane Dray, and Manuela Royer-Carenzi. The concept of animals' trajectories from a data analysis perspective. *Ecological Informatics*, 4(1):34–41, 2009.
- [9] Paul R. Wolf Charles D. Ghilani. *Elementary Surveying: An Introduction to Geomatics*, chapter 13.4 Reference Coordinate Systems, page 349. Pearson, 2016.
- [10] Paul R. Wolf Charles D. Ghilani. *Elementary Surveying: An Introduction to Geomatics*, chapter 13.4 Reference Coordinate Systems, page 352. Pearson, 2016.
- [11] Paul R. Wolf Charles D. Ghilani. *Elementary Surveying: An Introduction to Geomatics*. Pearson, 2016.
- [12] Patrick E. Clark, Joe Chigbrow, Douglas E. Johnson, Larry L. Larson, Ryan M. Nielson, Mounir Louhaichi, Tyanne Roland, and John Williams. Predicting spatial risk of wolf-cattle encounters and depredation. *Rangeland Ecology and Management*, 73(1):30 – 52, 2020.
- [13] David J. Unwin David O'Sullivan. *Geographic Information Analysis*, chapter 5 Point pattern analysis, page 135. John Wiley and Sons, 2010.
- [14] David J. Unwin David O'Sullivan. *Geographic Information Analysis*, chapter 5 Point pattern analysis, page 136. John Wiley and Sons, 2010.
- [15] David J. Unwin David O'Sullivan. *Geographic Information Analysis*. John Wiley and Sons, 2010.
- [16] Robert G. D'Eon, Robert Serrouya, Graham Smith, and Christopher O. Kochanny. Gps radiotelemetry error and bias in mountainous terrain. *Wildlife Society Bulletin (1973-2006)*, 30(2):430–439, 2002.
- [17] R.C. Dobos, D.B. Taylor, M.G. Trotter, B.E. McCorkell, D.A. Schneider, and G.N. Hinch. Characterising activities of free-ranging merino ewes before, during and after lambing from guss data. *Small Ruminant Research*, 131:12–16, 2015.

- 
- [18] ESRI. About arcgis pro. Tilgjengelig fra: <https://pro.arcgis.com/en/pro-app/latest/get-started/get-started.htm> (lest 19.05.2021).
- [19] ESRI. What is modelbuilder. <https://pro.arcgis.com/en/pro-app/latest/help/analysis/geoprocessing/modelbuilder/what-is-modelbuilder-.htm> (lest 06.05.2021).
- [20] Followit. Followit ab. <https://www.followit.se/> (lest 12.02.2021).
- [21] Followit. Tellus livestock - robust bestseller for tough conditions. <https://www.followit.se/livestock/> (lest 12.02.2021).
- [22] U.S. Space Force. Gps overview. <https://www.gps.gov/systems/gps/> (lest 24.05.2021).
- [23] Børje Forssell. Glonass i store norske leksikon på snl.no. <https://snl.no/GLONASS> (lest 01.03.2021).
- [24] Norvald Forssell, Børje; Kjerstad. Galileo i store norske leksikon på snl.no. <https://snl.no/Galileo> (lest 01.03.2021).
- [25] Norvald Forssell, Børje; Kjerstad. Gps i store norske leksikon på snl.no. <https://snl.no/GPS> (lest 26.02.2021).
- [26] Strand GH, de Boon A. et al. Hansen, I. Impacts of norwegian large carnivore management strategy on national grazing sector. 16:2470–2483, 2019.
- [27] Odd Magne Harstad. Beite i store norske leksikon. <https://snl.no/beite> (lest 10.03.2021).
- [28] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.
- [29] N.H. Jørgensen, G. Steinheim, and Ø. Holand. Does scale matter? variation in area use across spatiotemporal scales of two sheep breeds in two contrasting alpine environments. *Rangeland Ecology and Management*, 71(2):189–195, 2018.
- [30] Børje Kjerstad, Norvald; Forssell. Beidou i store norske leksikon på snl.no. <https://snl.no/BeiDou> (lest 01.03.2021).
- [31] Gunnar Loevaas. *Statistikk for universiteter og høyskoler*. Universitetsforlaget, 2013.
- [32] Jaime K. Manning, Eloise S. Fogarty, Mark G. Trotter, Derek A. Schneider, Peter C. Thomson, Russell D. Bush, and Greg M. Cronin. A pilot study into the use of global navigation satellite system technology to quantify the behavioural responses of sheep during simulated dog predation events. *Animal Production Science*, 54(10):1676–1681, 2014.
- [33] Ragnhild Mobaek, Atle Mysterud, Leif Egil Loe, Øystein Holand, and Gunnar Austrheim. Experimental evidence of density dependent activity pattern of a large herbivore in an alpine ecosystem. *Oikos*, in press, 09 2012.
- [34] Ragnhild Mobaek, Atle Mysterud, Leif Egil Loe, Øystein Holand, and Gunnar Austrheim. Density dependent and temporal variability in habitat selection by a large herbivore; an experimental approach. *Oikos*, 118(2):209–218, 2009.
- [35] Movebank. Enviromental data products accessed by the env-data system. <https://www.movebank.org/cms/movebank-content/envdata-products> (lest 25.05.2021).
- [36] NIBIO. Kilden. <https://kilden.nibio.no>.
- [37] The pandas development team. *pandas-dev/pandas: Pandas*, February 2020.
- [38] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [39] QGIS Development Team. Qgis geographic information system. <http://qgis.osgeo.org>, 2009.
-



- 
- [40] S. Raschka and V. Mirjalili. *Python Machine Learning*, chapter 3, page 76. Packt Publishing Ltd., 2019.
- [41] S. Raschka and V. Mirjalili. *Python Machine Learning*. Packt Publishing Ltd., 2019.
- [42] S. Raschka and V. Mirjalili. *Python Machine Learning*, chapter 3, page 75. Packt Publishing Ltd., 2019.
- [43] E.J. Raynor, S.P. Gersie, M.B. Stephenson, P.E. Clark, S.A. Spiegel, R.K. Boughton, D.W. Bailey, A. Cibils, B.W. Smith, J.D. Derner, R.E. Estell, R.M. Nielson, and D.J. Augustine. Cattle grazing distribution patterns related to topography across diverse rangeland ecosystems of north america. *Rangeland Ecology and Management*, 75:91–103, 2021.
- [44] Yngve Rekdal. Vegetasjon og beite ved minnestølen. 01 2001.
- [45] Yngve Rekdal. Vegetasjon og beite i rendalen Østfjell. spekedalen, sølendalen og joten. 2007.
- [46] Yngve Rekdal. Vegetasjon og utmarksbeite i vingelen. rapport frå vegetasjonskartlegging i tolga kommune. 2009.
- [47] L. Riaboff, S. Poggi, A. Madouasse, S. Couvreur, S. Aubin, N. Bédère, E. Goumand, A. Chauvin, and G. Plantier. Development of a methodological framework for a robust prediction of the main behaviours of dairy cows using a combination of machine learning algorithms on accelerometer data. *Computers and Electronics in Agriculture*, 169:105179, 2020.
- [48] Lucile Riaboff, Sébastien Couvreur, Aurélien Madouasse, Marie Roig-Pons, Sébastien Aubin, Patrick Massabie, Alain Chauvin, Nicolas Bédère, and Guy Plantier. Use of predicted behavior from accelerometer data combined with gps data to explore the relationship between dairy cow behavior and pasture characteristics. *Sensors*, 20(17), 2020.
- [49] Mac Schwager, Dean M. Anderson, Zack Butler, and Daniela Rus. Robust classification of animal tracking data. *Computers and Electronics in Agriculture*, 56(1):46–59, 2007.
- [50] A. So, T.V. Joseph, R.T. John, A. Worsley, and S. Asare. *The The Data Science Workshop: Learn how you can build machine learning models and create your own real-world data science projects, 2nd Edition*. Packt Publishing, 2020.
- [51] Geir-Harald Strand. The combined effects of centralization and carnivore management on sheep farmers and sheep farming in norway. *Human Dimensions of Wildlife*, 0(0):1–16, 2020.
- [52] Simen M. Svenkerud. Kartlegging av sauers bruk av beite ved bruk av gps-klaver og geografiske analyseverktøy. 2015.
- [53] Manohar Swamynathan. *Mastering Machine Learning with Python in Six Steps*, chapter 3 Step 3 Fundamentals of Machine Learning, page 221. Apress, 2019.
- [54] Manohar Swamynathan. *Mastering Machine Learning with Python in Six Steps*. Apress, 2019.
- [55] Maryam Teimouri. Extracting movement patterns from geographical trajectory data. 2018.
- [56] Maryam Teimouri, Ulf Geir Indahl, Hanne Sickel, and Håvard Tveite. Deriving animal movement behaviors using movement parameters extracted from location data. *ISPRS International Journal of Geo-Information*, 7(2), 2018.
- [57] Eugene D. Ungar, Zalmen Henkin, Mario Gutman, Amit Dolev, Avraham Genizi, and David Ganskopp. Inference of Animal Activity From GPS Collar Data on Free-Ranging Cattle. *Rangeland Ecology and Management*, 58(3):256 – 266, 2005.
- [58] Kays R. Wikelski M, Davidson SC. Movebank: archive, analysis and sharing of animal movement data. [www.movebank.org](http://www.movebank.org), (lest 20.05.2021).

---

# Vedlegg

## Vedlegg A: Pythonfil for innlesning av data

```
1 # -*- coding: utf-8 -*-
2 """
3 Created on Wed Feb 10 10:11:54 2021
4
5 @author: eriks
6 """
7
8 """
9 Importerer nødvendige pakker
10 """
11
12 import pandas as pd
13 import numpy as np
14 import matplotlib.pyplot as plt
15 import datetime
16 import math as m
17
18 """
19 Innlesning av data
20 """
21 head15 = ["YYYY", "MM", "DD", "Time", "TTF", "Lat", "Long", "2D3D", "Alt", "DOP", "
22           SVs", "FOM", "X", "Y", "Temp", "Bat", "Status", "SCap", "GPS", "GSM"]
23 head14 = ["YYYY", "MM", "DD", "Time", "TTF", "North.", "East.", "Zone", "2D3D", "
24           Alt", "DOP", "SVs", "FOM", "X", "Y", "Temp", "Bat", "Status", "SCap", "GPS", "GSM"]
25 dataset1 = pd.read_csv("OneDrive_2021-02-10/Tellus_data_2012/RAW/142-47.csv", names
26                       =head15, usecols=head15, skiprows=[0,1], index_col=False)
27
28
29 dataset1 = dataset1.dropna()
30 dataset1.reset_index(drop=True, inplace=True)
31
32 """
33 Tar bort alle rader med DOP over 5 og umaalt.
34 """
35 dataset1['DOP'] = pd.to_numeric(dataset1['DOP'], downcast="integer")
36
37
38 droplist = []
39 for i, j in enumerate(dataset1['DOP']):
40     if j is None or j > 5 or j == 0:
41         droplist.append(i)
42
43 dataset1.drop(index=droplist, inplace=True)
44 dataset1.reset_index(drop=True, inplace=True)
45
46 """
47 Konverterer Lat og Long til integer
48 """
49 dataset1['Long'] = pd.to_numeric(dataset1['Long'], downcast="float")
50 dataset1['Lat'] = pd.to_numeric(dataset1['Lat'], downcast="float")
51
52 """
53 Sletter kolonner fra datasettet som ikke er nødvendig i naavaerende analyse
54 """
55 dataset1.drop(columns=['GSM', 'GPS', 'SCap', 'Status', 'Bat', 'FOM', 'SVs',
56                       '2D3D', 'TTF'], inplace=True)
57
58 """
59 Enkel display av trajectory
60 """
61 plt.figure()
62 plt.plot(dataset1['Long'], dataset1['Lat'], 'o-')
63 plt.ylabel("Northing")
64 plt.xlabel("Easting")
65 plt.show()
```

---

```

66 """
67 Konverterer dato og tid til datetime. Try/Except fjerner data som kommer paa
68 feil format. (Eks. 25:00:00)
69 """
70
71 addlist = []
72 droplist = []
73 for i, row in dataset1.iterrows():
74     try:
75         date_str = "{0}-{1}-{2} {3}".format(row['YYYY'],row['MM'],row['DD'],row['
Time'])
76         time = datetime.datetime.strptime(date_str, "%Y-%m-%d %H:%M:%S")
77         addlist.append(time)
78     except:
79         droplist.append(i)
80         print(i)
81
82 dataset1.drop(droplist, inplace=True)
83 dataset1['Datetime'] = addlist
84 dataset1.drop(['YYYY','MM','DD','Time'], axis=1, inplace=True)
85
86 """
87 Definerer sauens id, sauerase (NKS=1, Spael=2), beitetype (Innmark=1, Utmark=2)
88 og tetthet
89 """
90 print(dataset1.head())
91
92 dataset1['ID'] = int(47)
93 dataset1['Rase'] = int(1)
94 dataset1['Beitetype'] = int(1)
95
96 dataset1.to_csv("Behandlet/Hol/2012/47.csv")

```

---

## Vedlegg B: Pythonfil for beregning av fart, avstand og retning

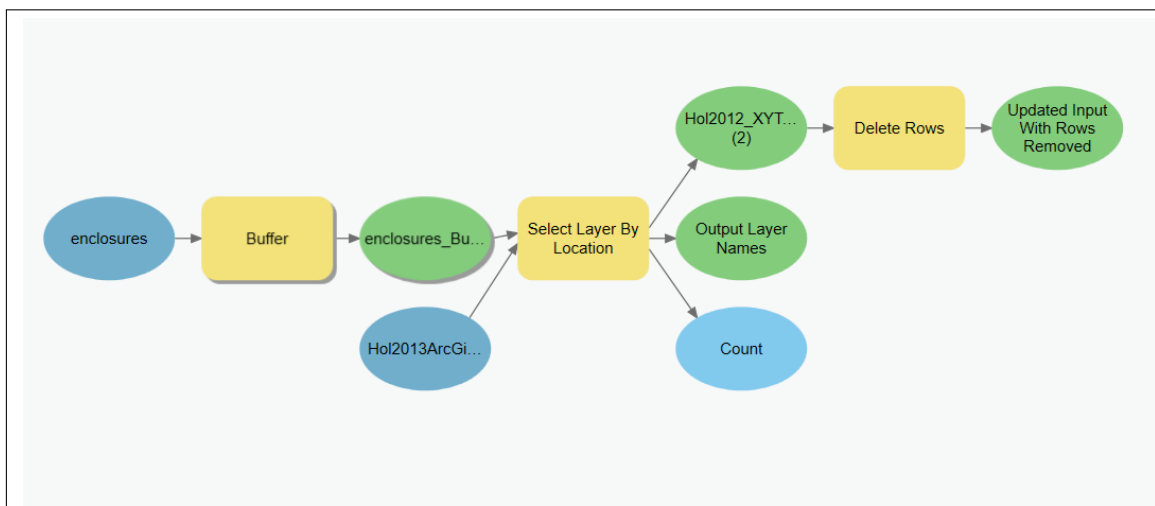
```
1 # -*- coding: utf-8 -*-
2 """
3 Created on Mon Mar 15 15:44:39 2021
4
5 @author: eriks
6 """
7
8
9 import pandas as pd
10 import numpy as np
11 import matplotlib.pyplot as plt
12 import datetime
13 import math as m
14
15 """
16 Innlesing av data
17 """
18 dataset1 = pd.read_csv("Behandlet/Med omragedata(4)/bratthoi3rett2.csv", sep=';',
19                        decimal=',')
20
21 """
22 Enkel utskrift for aa sjekke at datasettet ser greit ut
23 """
24 print(dataset1.head())
25
26 """
27 Sorterer data etter individuelle id-nummer og etablerer et tomt datasett
28 """
29 dfs = dict(tuple(dataset1.groupby('ID')))
30 dataset = []
31
32 """
33 Itererer gjennom alle enkeltsauer
34 """
35 for i, name in enumerate(dfs):
36     frame = dfs.get(name)
37
38     """
39     "Tvinger" posisjonskolonnene til formatet float
40     """
41     frame['X'] = pd.to_numeric(frame['X'], downcast="float")
42     frame['Y'] = pd.to_numeric(frame['Y'], downcast="float")
43
44     """
45     Setter opp tomme lister for de parametere som skal legges til datasettet
46     """
47     distance_list = []
48     angle_list = []
49     fart_list = []
50     delta_x = []
51     delta_y = []
52
53     """
54     Resetter start- og sluttverdier
55     """
56     start_x = 0
57     start_y = 0
58     end_x = 0
59     end_y = 0
60
61     """
62     Itererer gjennom posisjoner for hver enkelt sau
63     """
64     for index, row in frame.iterrows():
65
66         """
67         Sjekker om dette er startposisjon, hvis saa er tilfelle blir alle
68         nye parametere satt til 0.
69         """
70         if start_x == 0:
71             start_x = row['X']
72             start_y = row['Y']
```

```

71     distance_list.append(0)
72     angle_list.append(0)
73     fart_list.append(0)
74     delta_x.append(0)
75     delta_y.append(0)
76     time = datetime.datetime.strptime(row['timestamp'],
77                                       "%d.%m.%Y %H.%M.%S")
78     """
79     Hvis det ikke er startposisjon, saa beregnes parametere
80     """
81     else:
82         end_x = row['X']
83         end_y = row['Y']
84
85         """ Beregner x- og y-forskjellen fra forrige punkt """
86         x_forskjell = end_x - start_x
87         y_forskjell = end_y - start_y
88         distance = m.sqrt((x_forskjell**2)+(y_forskjell**2))
89         angle = m.atan2(y_forskjell,x_forskjell)
90
91         diff = datetime.datetime.strptime(row['timestamp'], "%d.%m.%Y %H.%M.%S")
92     - time
93         fart = distance*3.6/diff.total_seconds()
94
95         distance_list.append(distance)
96         angle_list.append(angle)
97         fart_list.append(fart)
98         delta_x.append(x_forskjell)
99         delta_y.append(y_forskjell)
100
101         start_x = end_x
102         start_y = end_y
103         time = datetime.datetime.strptime(row['timestamp'], "%d.%m.%Y %H.%M.%S")
104
105     """ Setter inn nye parametere i datasettet """
106     frame['Distance'] = distance_list
107     frame['Vinkel'] = angle_list
108     frame['Fart'] = fart_list
109     frame['Delta_x'] = delta_x
110     frame['Delta_y'] = delta_y
111
112     dataset.append(frame)
113
114     """
115     Skriver nytt datasett til fil
116     """
117     combined_csv = pd.concat([f for f in dataset])
118
119     combined_csv.to_csv("Behandlet/Med Omragedata(4)/Ferdig med rett fart/Bratthoa13.
120                          csv")

```

Vedlegg C: Buffermodell fra ArcGIS ModelBuilder for å fjerne punkter utenfor studieområdet



```

1 # -*- coding: utf-8 -*-
2 """
3 Generated by ArcGIS ModelBuilder on : 2021-05-24 13:40:51
4 """
5 import arcpy
6
7 def Model(): # Model
8
9     # To allow overwriting outputs change overwriteOutput option to True.
10    arcpy.env.overwriteOutput = False
11
12    Hol2013ArcGis_XYTableToPoint = "Hol2013ArcGis_XYTableToPoint"
13    enclosures = "enclosures"
14
15    # Process: Buffer (Buffer) (analysis)
16    enclosures_Buffer = "C:\\Users\\eriks\\OneDrive\\Dokumenter\\ArcGIS\\Projects\\
17    Sau spor\\Sau spor.gdb\\enclosures_Buffer"
18    arcpy.analysis.Buffer(in_features=enclosures, out_feature_class=
19    enclosures_Buffer, buffer_distance_or_field="5 Meters", line_side="FULL",
20    line_end_type="ROUND", dissolve_option="NONE", dissolve_field=[], method="
21    PLANAR")
22
23    # Process: Select Layer By Location (Select Layer By Location) (management)
24    Hol2012_XYTableToPoint2_2, Output_Layer_Names, Count = arcpy.management.
25    SelectLayerByLocation(in_layer=[Hol2013ArcGis_XYTableToPoint], overlap_type="
26    COMPLETELY_WITHIN", select_features=enclosures_Buffer, search_distance="0
27    Meters", selection_type="NEW_SELECTION", invert_spatial_relationship="INVERT")
28
29    # Process: Delete Rows (Delete Rows) (management)
30    Updated_Input_With_Rows_Removed = arcpy.management.DeleteRows(in_rows=
31    Hol2012_XYTableToPoint2_2)[0]
32
33    if __name__ == '__main__':
34        # Global Environment settings
35        with arcpy.EnvManager(scratchWorkspace=r"C:\Users\eriks\OneDrive\Dokumenter\
36        ArcGIS\Projects\Sau spor\Sau spor.gdb", workspace=r"C:\Users\eriks\OneDrive\
37        Dokumenter\ArcGIS\Projects\Sau spor\Sau spor.gdb"):
38            Model()

```

---

## Vedlegg D: Python notebook-fil for Randomforest-estimering

```
1 # -*- coding: utf-8 -*-
2 """Master random forest.ipynb
3
4 Automatically generated by Colaboratory.
5
6 Original file is located at
7     https://colab.research.google.com/drive/1f_Qi2MnVOT7BC2RS3UPAoa0P5FWu1I6t
8
9 # Masteroppgave
10 ## Adferdsanalyse sau
11
12 ### Erik Sand
13
14 ## Importerer nødvendige pakker
15 """
16
17 import pandas as pd
18 import numpy as np
19 import matplotlib.pyplot as plt
20 import io
21
22 from sklearn.preprocessing import StandardScaler
23
24 from sklearn.ensemble import RandomForestRegressor
25 from sklearn import preprocessing
26
27 from sklearn.model_selection import train_test_split
28 from sklearn.feature_selection import SelectFromModel
29 from sklearn.feature_selection import SelectKBest
30 from sklearn.model_selection import cross_val_score
31
32 """# Opplasting av filer og transformering til DataFrame"""
33
34 from google.colab import files
35
36 x_train = files.upload()
37
38 """Transformerer fra csv til dataframes."""
39
40 train_df = pd.read_csv(io.BytesIO(x_train['Total.csv']))
41
42 """# Forberedelse av data
43
44 """
45
46 # Beregner prosentvis fordeling
47 train_df['Prosent fordeling'] = train_df['Count']/train_df['Antall']
48
49 x = train_df[['Beiteverdi', 'Fordeling', 'Ukenr', 'Tetthet', 'Antall', 'Beitetype.1']]
50
51 # Bruker LabelEncoder for aa transformere tekstkolonna 'Beitetype.1' til
52 # numeriske verdier
53 le = preprocessing.LabelEncoder()
54 x['Beitetype.1'] = le.fit_transform(x['Beitetype.1'])
55
56 y = train_df[['Prosent fordeling']]
57
58 """Deler inn i trenings- og testdata."""
59
60 # Kodecelle for aa dele opp datasettene tilfeldig
61
62 X_train, X_test, Y_train, Y_test = train_test_split(x,
63                                                    y,
64                                                    test_size = 0.4,
65                                                    random_state = 1000)
66 Y_train = np.ravel(Y_train)
67 Y_test = np.ravel(Y_test)
68
69 # Metode en
70 X_train = x.iloc[:236]
71 X_test = x.iloc[236:]
```



```

72
73 Y_train = np.ravel(y.iloc[:236])
74 Y_test = np.ravel(y.iloc[236:])
75
76 # Metode to
77 X_train = x.iloc[156:]
78 X_test = x.iloc[:156]
79
80 Y_train = np.ravel(y.iloc[156:])
81 Y_test = np.ravel(y.iloc[:156])
82
83 # Metode tre
84 X_train = x.iloc[:156]
85 X_test = x.iloc[156:]
86
87 Y_train = np.ravel(y.iloc[:156])
88 Y_test = np.ravel(y.iloc[156:])
89
90 # Metode fire
91 X_train = x.iloc[236:276]
92 X_test = x.iloc[276:]
93
94 Y_train = np.ravel(y.iloc[236:276])
95 Y_test = np.ravel(y.iloc[276:])
96
97 # Metode fem
98 X_train = x.iloc[:276]
99 X_test = x.iloc[276:]
100
101 Y_train = np.ravel(y.iloc[:276])
102 Y_test = np.ravel(y.iloc[276:])
103
104 """Bruker feature selection for aa finne de beste parameterne, og bruker disse i
    random forest regressor.
105
106 """
107
108 sel = SelectFromModel(RandomForestRegressor())
109
110 sel.fit(X_train, Y_train)
111
112 """Kopierer utskriften av beste modell, og bruker denne.
113
114 """
115
116 clf = RandomForestRegressor(bootstrap=True, ccp_alpha=0.0,
117                             criterion='mse', max_depth=None,
118                             max_features='auto',
119                             max_leaf_nodes=None,
120                             max_samples=None,
121                             min_impurity_decrease=0.0,
122                             min_impurity_split=None,
123                             min_samples_leaf=1,
124                             min_samples_split=2,
125                             min_weight_fraction_leaf=0.0,
126                             n_estimators=100, n_jobs=None,
127                             oob_score=False,
128                             random_state=None, verbose=0,
129                             warm_start=False)
130
131 clf.fit(X_train, Y_train)
132
133 # Prediksjoner
134 Y_pred_train = clf.predict(X_train)
135 Y_pred_test = clf.predict(X_test)
136
137 # Score
138 train_score = clf.score(X_train, Y_train)
139
140 test_score = clf.score(X_test, Y_test)
141
142 print('Train score (mse): %.3f' % train_score)
143

```

```

144 print('Test score (mse): %.3f' % test_score)
145
146 # Faa numerisk feature importance
147 importances = list(clf.feature_importances_)
148 # Lister opp tuples med variabler og viktighet
149 feature_list = ['Beiteverdi', 'Fordeling', 'Ukenr', 'Tetthet', 'Antall', 'Beitetype.1']
150 feature_importances = [(feature, round(importance, 2)) for feature, importance in
151                        zip(feature_list, importances)]
152 # Sorterer feature importance til viktigste foerst
153 feature_importances = sorted(feature_importances, key = lambda x: x[1], reverse =
154                             True)
155 # Utskrift
156 [print('Variabel: {:20} viktighet: {}'.format(*pair)) for pair in
157     feature_importances];
158
159 # Plot for aa visualisere residualer
160 plt.scatter(Y_pred_train,
161            Y_pred_train - Y_train,
162            c='steelblue',
163            edgecolor='white',
164            marker='o',
165            s=35,
166            alpha=0.9,
167            label='Training data')
168 plt.scatter(Y_pred_test,
169            Y_pred_test - Y_test,
170            c='limegreen',
171            edgecolor='white',
172            marker='s',
173            s=35,
174            alpha=0.9,
175            label='Test data')
176 plt.title('Estimering av neste aar')
177 plt.xlabel('Predicted values')
178 plt.ylabel('Residuals')
179 plt.legend(loc='upper left')
180 plt.hlines(y=0, xmin=-10, xmax=50, lw=2, color='black')
181 plt.xlim([0, 0.7])
182 plt.tight_layout()
183 plt.show()

```

## Vedlegg E: Python notebook-fil for plotting av prosentvis fordeling i Google Colab

```
1 # -*- coding: utf-8 -*-
2 """Master linear.ipynb
3
4 Automatically generated by Colaboratory.
5
6 Original file is located at
7     https://colab.research.google.com/drive/1coDX0k5BvZ3kiSmIMpjE_STPTng_H58-
8
9 # Masteroppgave
10 ## Adferdsanalyse sau
11
12 ### Erik Sand
13
14 ## Importerer nødvendige pakker
15 """
16
17 import pandas as pd
18 import numpy as np
19 import matplotlib.pyplot as plt
20 import io
21
22 """# Opplasting av filer og transformering til DataFrame"""
23
24 from google.colab import files
25
26 x_train = files.upload()
27
28 """Transformerer fra csv til dataframes."""
29
30 train_df = pd.read_csv(io.BytesIO(x_train['Total.csv']))
31
32 """# Forberedelse av data
33
34 """
35
36 # Beregner prosentvis fordeling
37 train_df['Prosent fordeling'] = train_df['Count']/train_df['Antall']
38
39 # Velger hvilke verdier som skal plottes
40 y_sub = train_df[train_df['Beiteverdi']==3]
41 y_sub = y_sub[train_df['Beitetype.1']=='Utmark']
42 y_sub = y_sub[train_df['Fordeling']>=0.1]
43
44 # Sorterer per uke og beregner gjennomsnittet
45 x_sub = y_sub.groupby("Ukenr").mean()
46
47 x = x_sub.index                # Ukenr
48 y = x_sub['Prosent fordeling']*100    # Prosent
49
50
51 # Plottefunksjon
52 plt.title('Mindre godt beite Bratthoa')
53 plt.ylabel('Prosent')
54 plt.xlabel('Ukenr')
55 plt.plot(x,y,'ok-')
```



**Norges miljø- og biovitenskapelige universitet**  
Noregs miljø- og biovitenskapelige universitet  
Norwegian University of Life Sciences

Postboks 5003  
NO-1432 Ås  
Norway