



Kunnskap for en bedre verden

## Oppdatering av turstier

*Kan midling av frivillig innsamlede GNSS-data automatiseres for å effektivisere kartlegging av turstier?*

Helene Thomesen Flekke  
Tiril Sølve Halvorsen  
Håkon Hegreberg  
Tom E. Eriksen

Gradering: Åpen

Bachelor i geomatikk

Innlevert: Mai 2021

Veileder: Sverre Stikbakke

Norges teknisk-naturvitenskapelige universitet  
Institutt for vareproduksjon og byggingsteknikk

Oppgavens tittel: Oppdatering av turstier	Dato: 20.05.21		
	Antall sider: 70		
	Masteroppgave:	Bacheloroppgave	X
Navn: Helene Thomesen Flekke, Tiril Sølve Halvorsen, Håkon Hegreberg, Tom E. Eriksen			
Veileder: Sverre Stikbakke			

Det siste tiåret har det vært en økning i bruken av forbrukerteknologi for å samle inn geografiske data. Denne studien vil bruke frivillig innsamlet geografisk informasjon samlet med GNSS for å gjøre kartleggingen av turstier mer effektiv. Den høyeste prioriteten vil være fullstendigheten av turstinettet, mens et høyt nøyaktighetsnivå vil prioriteres i mindre grad.

For å undersøke dette problemet vil to metoder bli utviklet. Begge metodene kartlegger turstier ved å midle GNSS-data. Den første metoden er en manuell prosess som bruker hendelsesintensitet til å finne den mest sannsynlige posisjonen til turstien. Den andre metoden er en enkel automatisert prosess for å midle GNSS-data, men den tar ikke i bruk hendelsesintensitet.

Vi finner at den første metoden er en effektiv metode for å skape et turstinettverk, og metoden kan gjøre mye GNSS-data av usikker kvalitet til brukbare data. Problemet med denne metoden er at den ikke er automatisert.

Den andre metoden er også en effektiv metode for å skape et turstinettverk, og metoden er automatisert. Metoden har svakheter, den kan ikke ta hensyn til hendelsesintensitet.

Mens begge metodene effektivt kan skape et turstinettverk fra GNSS-data, vil det fremdeles være nødvendig med en manuell gjennomgang av turstinettverket for å bekrefte at kvaliteten blir tilstrekkelig.

Turstiene er tilgjengelige her: <https://ntnu-gis.maps.arcgis.com/apps/inspector/index.html?appid=2c4f08d3b39c4cbdbdf83a3c6a7415f2>

Stikkord:

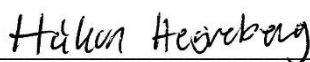
Geografiske informasjonssystemer (GIS)
Geomatikk
ArcGIS
VGI



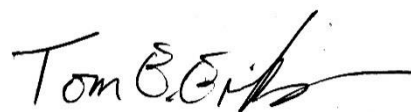
Helene T. Flekke



Tiril S. Halvorsen



Håkon Hegreberg



Tom E. Eriksen

# Abstract

For the past decade there has been an increase in the utilization of consumer technology to gather geographical data. This study will use volunteered geographical information gathered with GNSS to make mapping of hiking trails more efficient. The highest priority will be the completeness of the hiking trail network, while a high level of accuracy will be less prioritised.

To examine this problem, two methods will be developed. Both methods will infer a hiking trail by averaging GNSS data. The first method is a manual process that uses density estimation to find the most likely position of the hiking trail. The second method is a simple automated process for averaging GNSS data, without density estimation.

We find that the first method is an efficient method to create a hiking trail network, and the method can turn a lot of GNSS data of uncertain quality into useable data. The issue with this method is that it is not automated.

The second method is also an efficient method to create a hiking trail network, and the method is automated. The method has weaknesses, it cannot take the density of the data into account.

While both methods will efficiently create a hiking trail network from GNSS data, it is still necessary for a manual review of the datasets to confirm that the quality is satisfactory.

The hiking trails are available here: <https://ntnu-gis.maps.arcgis.com/apps/instant/interactivelegend/index.html?appid=2c4f08d3b39c4cbdbdf83a3c6a7415f2>

# Forord

Tormod Urke er universitetslektor ved institutt for vareproduksjon og byggteknikk (IVB). Han har, sammen med Jørgen Caspersen fra MjøsPlan, spilt inn denne oppgaven som et forslag. Deres forslag var «Oppdatering av turstier».

Bakgrunnen for oppgaven er at det eksisterer mange gamle og nye turstier som ikke er kartlagt. Vi ønsket å se på om det var mulig å bruke moderne metoder for å løse dette godt kjente problemet. Oppgaven vil også gi en utvidet forståelse av to ulike metoder som tar i bruk frivillig innsamlet data.

Vi vil med denne oppgaven avslutte vårt treårige bachelorstudium i geomatikk ved Norges teknisk-naturvitenskapelige universitet. Gjennom disse tre årene har vi tilegnet oss utvidet kunnskap om innsamling, systematisering, forvaltning, analysering og presentasjon av stedfestet informasjon.

Vi ønsker å takke Tormod Urke og Jørgen Caspersen fra MjøsPlan, som foreslo denne oppgaven. Det har vært en interessant problemstilling å løse, og det blir spennende å se på utvikling innen dette feltet fremover.

Vi vil også takke vår veileder, Sverre Stikbakke, og øvrige ansatte ved NTNU som har bistått oss under vårt arbeid. Vår veileder har vært imøtekommende og svært hjelpelig over Microsoft Teams. Mange konsepter og metoder var nye for både Sverre og oss, og vi satt stor pris på at han tok seg tid til å lese seg opp på nye emner.

Sverre har en personlig interesse for å loggføre utendørs aktivitet, da særlig terrengsykling. Han har bidratt med sine personlig innsamlede data, som vi har tatt i bruk i oppgaven vår. Vi har også vært så heldige å få lov til å bruke data samlet inn av Arild Hegreberg, så en stor takk til dem begge.

# Innholdsfortegnelse

Abstract .....	iii
Forord .....	iv
Innholdsfortegnelse .....	v
Figurliste.....	viii
Tabelliste .....	x
Ordforklaringer og forkortelser .....	xi
1 Introduksjon .....	1
1.1 Bakgrunn for oppgaven .....	1
1.2 Problemstilling.....	1
1.3 Oppgavens oppbygning .....	2
2 Teori .....	3
2.1 Volunteered Geographical Information.....	3
2.2 Relevante datasett .....	4
2.3 Vektordata og topologi .....	4
2.4 GNSS .....	5
2.4.1 Feilkilder og følgefeil.....	5
2.4.2 Nøyaktighet .....	6
2.4.3 Kvalitet med eksterne GNSS-enheter .....	7
2.5 GPX-formatet .....	8
2.6 Kernel Density Estimation.....	9
2.7 Tidligere arbeid.....	9
2.7.1 Eksisterende kartkonstruksjonsalgoritmer .....	9
2.7.2 Metoder for kvalitetsvurdering.....	11
3 Metode.....	12
3.1 Midling .....	12
3.2 Datasett .....	12
3.2.1 Innsamling av data .....	12
3.2.2 Datasett og studieområde .....	13
3.3 Programvare.....	13
3.3.1 ArcGIS-programvare.....	13

3.3.2	Jupyter Notebook og ArcPy .....	14
3.4	Verktøy i ArcGIS Pro .....	15
3.5	Metode 1 - KDE .....	16
3.5.1	Del 1 – Preprosessering .....	17
3.5.2	Del 2 – Analyse .....	17
3.5.3	Del 3 – Reklassifisering og senterlinje .....	19
3.6	Metode 2 - Buffer .....	20
3.6.1	Del 1 – Preprosessering .....	20
3.6.2	Del 2 – Analyse .....	20
3.6.3	Del 3 - Publisering .....	21
3.7	Vurdering av kvalitet og nøyaktighet .....	21
4	Resultater .....	22
4.1	Metode 1 - KDE .....	22
4.1.1	Geometri .....	22
4.1.2	Topologi .....	25
4.2	Metode 2 – Buffer .....	26
4.2.1	Geometri .....	27
4.2.2	Topologi .....	29
4.3	Metode 1 versus Metode 2 .....	30
5	Diskusjon .....	31
5.1	Om bruk av VGI .....	31
5.2	Valg av metoder .....	32
5.3	Drøfting av metode 1 – KDE .....	33
5.3.1	Diskusjon av metode .....	33
5.3.2	Diskusjon av resultat .....	35
5.3.3	Diskusjon av nøyaktighet og kvalitet .....	37
5.4	Drøfting av metode 2 – Buffer .....	38
5.4.1	Diskusjon av metode .....	38
5.4.2	Diskusjon av resultat .....	41
5.4.3	Diskusjon av nøyaktighet og kvalitet .....	44
5.5	Metode 1 versus metode 2 .....	46
6	Konklusjon .....	49
7	Videre arbeid .....	50

Litteraturliste .....	51
Vedleggsliste .....	54
Vedlegg I - Resultater metode 1 .....	55
A. Totenvika.....	55
B. Bodø.....	56
Vedlegg II - Resultater metode 2 .....	57
A. Totenvika.....	57
B. Bodø.....	58
Vedlegg III – ModelBuilder .....	59
Metode 1.....	59

# Figurliste

Figur 1: Flytdiagram av metode 1 .....	16
Figur 2: Reklassifisering av KDE for Bodø .....	19
Figur 3: Senterlinje med KDE og GPX-punkter .....	19
Figur 4: Flytdiagram for metode 2 .....	20
Figur 5: Referansedata, Elveg 2.0, med buffersone 10-15 meter .....	21
Figur 6: Eksempel på område hvor senterlinjen har stort avvik .....	23
Figur 7: Vegkryss med god geometri .....	23
Figur 8: Kryss med dårlig geometri .....	23
Figur 9: Vegkryss med dårlig geometri .....	24
Figur 10: Samme kryss som i figur 9, men med flyfoto .....	24
Figur 11: Kurve med flere underliggende målinger .....	24
Figur 12: Kurve med en underliggende måling .....	24
Figur 13: Vegkryss med dårlig topologi .....	25
Figur 14: Vegkryss med feil antall kryss .....	25
Figur 15: Område hvor parallelle veger har blitt slått sammen .....	25
Figur 16: Oversikt over Totenvika .....	26
Figur 17: Oversikt over Bodø .....	26
Figur 18: Åpne områder i Totenvika .....	27
Figur 19: Åpne områder i Totenvika .....	27
Figur 20: Kupert terreng i Totenvika .....	27
Figur 21: Boligfelt Bodø .....	28
Figur 22: Kupert terreng Bodø .....	28
Figur 23: Problematisk vegkryss i Totenvika .....	28
Figur 24: Parallelle stier i Bodø .....	29
Figur 25: Dårlig kryssløsning i Bodø .....	29
Figur 26: Topologisk ukorrekt kryss .....	29
Figur 27: Sammenligning av metodene 1.3 .....	30
Figur 28: Sammenligning av metodene 1.4 .....	30
Figur 29: Sammenligning av metodene 1.1 .....	30
Figur 30: Sammenligning av metodene 1.2 .....	30
Figur 31: Område med dårlig måleserie .....	34



Figur 32: Område med utliggere .....	35
Figur 33: Område med mange punktmålinger .....	36
Figur 34: KDE med 15m søkeradius.....	36
Figur 35: KDE med 10m søkeradius.....	36
Figur 36: Vegkryss med feil antall noder .....	37
Figur 37: Område hvor en dårlig måleserie påvirker buffer-polygonen .....	39
Figur 38: Aggregerte polygoner, merged linjer og senterlinje.....	40
Figur 39: Små vegstubber på under 150 meter .....	41
Figur 40: Turstinettverk Kupert terreng i Bodø .....	42
Figur 41: Ulike kryssløsninger med forskjellige parametere. Her er 5, 10 og 15m synlig.....	42
Figur 42: Sammenligning av parametere 1.1: Senterlinje med 5 meter bufferradius skaper mange parallelle linjer.....	43
Figur 43: Sammenligning av parametere 1.3 .....	43
Figur 44: Sammenligning av parametere 1.2 .....	43
Figur 45: Sving i boligfelt i Bodø 1.2 .....	44
Figur 46: Sving i boligfelt i Bodø 1.1 .....	44
Figur 47: Ny sti .....	47

# Tabelliste

Tabell 1: Oversikt over GPX-datasett .....	13
Tabell 2: Oversikt over brukte verktøy i ArcGIS Pro .....	15
Tabell 3: Parameterinnstillinger brukt i KDE-analysen .....	18

# Ordforklaringer og forkortelser

<b>Cluster</b>	Gruppering av et sett med objekter
<b>DOP</b>	Dilution Of Precision
<b>Elveg 2.0</b>	Produktet Elveg 2.0 - elektronisk vegnett
<b>Feature</b>	Representasjon av et fenomen
<b>Feature Class</b>	Samling av geografiske Features av samme geometriske type
<b>FKB</b>	Felles KartdataBase
<b>GIS</b>	Geografiske informasjonssystemer
<b>GNSS</b>	Global Navigation Satellite System
<b>GPX</b>	GPS Exchange Format
<b>HDOP</b>	Horizontal Dilution Of Precision
<b>KDE</b>	Kernel Density Estimation
<b>Midling</b>	Tilsvarende gjennomsnitt
<b>ModelBuilder</b>	ArcGIS-verktøy for å bygge arbeidsflyt for geoprosessering
<b>OSM</b>	OpenStreetMap
<b>PAM</b>	Partitioning Around Medoids
<b>Python</b>	Objekt-orientert programmeringsspråk
<b>Trackpoints</b>	Sortert liste med relaterte GNSS-punkter som beskriver en sti
<b>TTF</b>	Time To First Fix
<b>VGI</b>	Volunteered Geographic Information
<b>Waypoints</b>	Enkeltstående GNSS-punkter uten relasjon
<b>WGS84</b>	World Geodetic System 1984

# 1 Introduksjon

## 1.1 Bakgrunn for oppgaven

Frivillig innsamling av geografiske data brukes stadig oftere. Dette er en relativt ny metode for datainnsamling, og det er en rekke ulike bruksområder. Kartverket har en kartløsning kalt Rett i kartet, som gir brukerne muligheten til å bidra til å holde kartene oppdatert. Her må hver eneste sti manuelt godkjennes for å sikre dataenes kvalitet. Dette vil gi en kvalitetssikret, men ikke spesielt effektiv prosess for kartlegging.

Det har tidligere blitt forsøkt å skape effektive prosesser for å kartlegge stier og lignende ved å bruke GNSS-spor. TopoFusion er en av få programvarer utviklet for dette formålet, men dette er utviklet til bruk av enkeltpersoner for private formål, og ikke for kartlegging av større områder.

Da vi skulle finne et tema vi kunne skrive om i vår bacheloroppgave var «Oppdatering av turstier» en oppgaven vi kunne velge. Oppgaven var spilt inn av Tormod Urke og Jørgen Caspersen fra MjøsPlan, som kom på dette forslaget under en samtale. Deres bakgrunn for forslaget var manglende kartlegging av stier og turmuligheter.

Ikke bare virket dette som et interessant tema, oppgaven ga oss også muligheten til å bruke og videreutvikle GIS-ferdighetene vi har anskaffet oss i løpet av studietiden. Vi ønsket å løse problemstillingen ved å skape en effektivisert prosess som bruker frivillig innsamlede geografiske data til å kartlegge turstier.

## 1.2 Problemstilling

Innsamling av geografiske data gjort av frivillige, også kalt VGI, er et voksende felt. På bakgrunn av dette ønsket vi å utvikle en metode som benytter seg av VGI. Med brukergenerert informasjon kan en spare seg for den tidskrevende innsamlingsperioden mange kartleggingsprosjekter har. Turstier går gjerne over store landområder, og det ville

vært et svært omfattende og kostbart arbeid for en landmåler å kartlegge på egenhånd. I tillegg ville det vært en kostbar prosess for å oppnå det ferdige produktet.

Vår problemstilling er derfor formulert som følger:

*Kan midling av frivillig innsamlede GNSS-data automatiseres for å effektivisere kartlegging av turstier?*

På bakgrunn av denne problemstillingen ønsker vi å se på muligheten for å kartlegge turstier på en effektiv måte ved å ta i bruk frivillig innsamlet data. Vi ønsker å se på hvilken kvalitet som kan forventes ved midling av GNSS-data, og om det er mulig å automatisere deler av prosessen.

I vårt prosjekt har vi lagt hovedfokuset på midling av GNSS-spor, samlet inn ved bruk av ulike mottakere med ukjent kvalitet. Turstinnettverkets fullstendighet vil være den høyeste prioriteringen, fremfor at det skal ha høy nøyaktighet.

Vi har tatt i bruk data samlet inn av Sverre Stikbakke og Arild Hegreberg. Sverre brukte en sykkelcomputer med GNSS-mottaker og Arild brukte en treningsklokke.

### **1.3 Oppgavens oppbygning**

I kapittel 2 tar vi for oss teori som er relevant for oppgaven. I kapittel 3 presenterer vi to ulike metoder for effektivisering av turstier ved hjelp av frivillig innsamlet data. Metode 1 bruker KDE-analyse, mens metode 2 lager buffer rundt alle linjer. I kapittel 4 presenteres oppgavens resultat, før resultatene drøftes og diskuteres i kapittel 5. Avslutningsvis konkluderer vi basert på resultatene våre.

## 2 Teori

Dette kapittelet vil gi en forklaring på sentrale begreper for oppgaven som er gjennomført. Innledningsvis gis en utvidet forklaring av begrepet volunteered geographical information og en introduksjon til relevante datasett for oppgaven. Vektordata og topologi redegjøres også for. Deretter beskrives GNSS-teknologien med mulige feilkilder og forventet nøyaktighet. Formatet GPX presenteres også.

En sentral del av oppgaven er analysen kernel density estimation. Dette redegjøres det for, samt de relevante datasettene for oppgaven. Avslutningsvis er tidligere kartkonstruksjonsalgoritmer forklart, med metoder for hvordan kvaliteten til algoritmene blir målt.

### 2.1 Volunteered Geographical Information

Begrepet volunteered geographic information (VGI) ble først brukt av M. Goodchild i 2007. Han brukte dette begrepet for å beskrive det nye fenomenet der privatpersoner oppretter, samler inn og deler geografisk informasjon gjennom nettsider som OpenStreetMap. Goodchild forteller om det utbredte engasjementet blant privatpersoner uten formelle kvalifikasjoner, som nå oppretter geografisk informasjon. Dette var arbeid som tidligere stort sett var forbeholdt offentlige etater. Privatpersonene er hovedsakelig utrente, og resultatene har en usikker grad av nøyaktighet, men Goodchild mener at privatpersonene samlet vil ha dyp innvirkning på geografiske informasjonssystemer (Goodchild, 2007).

Det er flere land som nå benytter seg av VGI til ulike formål innenfor nasjonal kartlegging. VGI gir mulighet til å kartlegge nye objekter eller ny informasjon som ikke har blitt samlet inn av offentlige etater, eller som ikke er innenfor etatens prioriteringer eller deres økonomiske rammer (Antoniou *mfl.*, 2017).

Implementering av VGI kan være problematisk. Ved registrering av data pekes det på risiko ved å få inn falsk eller feilaktig data. I tillegg er det bekymringer rundt personvern for brukere. Ved kvalitetskontroll nevnes problemer med at antallet deltakere ikke nødvendigvis

garanterer for kvalitet av innmålt data. Som løsning anbefales det å implementere en metode for kvalitetssikring (Antoniou *mfl.*, 2017).

## 2.2 Relevante datasett

### Databasen Tur- og friluftsruter

Den nasjonale databasen for tur- og friluftsruter forvaltes av Kartverket. Formålet med datasettet er blant annet å gi viktig basisinformasjon i arbeidet med kartlegging og verdsetting av friluftsområder. Datainnsamlingen baserer seg i stor grad på frivillig innrapportering via publikumsløsningen Rett i kartet. Her kan det lastes opp GPX-fil eller tegnes i kartet (Sosi Produktspesifikasjon, 2017).

Produktspesifikasjonen beskriver krav til kvalitet til datasettet. Anbefalt nøyaktighet på data innsamlet med GNSS-mottaker er 10 til 15 meter. Stedfestingsnøyaktighet for innsendte data kontrolleres ved å sammenligne dem med primærdata, som datasettet FKB-TraktorvegSti (Sosi Produktspesifikasjon, 2017).

### Datasettet Elveg 2.0

Elveg 2.0 omfatter alle kjørbare veger lengre enn 50 meter, samt gang- og sykkelveger. Vegene er representert som veglenkegeometri (Kartverket og Statens Vegvesen, 2019).

### Bakgrunnskartet GeocacheBasis

GeocacheBasis er en karttjeneste som fremhever infrastruktur og generell topografisk informasjon (GeodataOnline, ingen dato). GeocacheBasis blir brukt som bakgrunnskart. Kartet er basert på FKB-data.

## 2.3 Vektordata og topologi

Vektordata er koordinatfestede objekter i form av punkter, linjer og polygoner. Topologi beskriver hvordan vektorobjekter blir representert, og sammenkoblingen mellom hvert objekt. Representasjon av geografiske data som har relasjoner, kan gjøres i form av linjetopologi. Linjetopologi bygges opp av kanter og noder som representerer veger og kryss.

Kanter blir knyttet sammen av noder, og det blir dermed formet et sammenhengende nettverk (Lloyd, 2010).

## **2.4 GNSS**

GNSS er en samlebetegnelse for satellittsystemer som blant annet GPS, GLONASS og Galileo. Dette er systemer som tar i bruk data fra satellitter til global navigasjon og stedfestelse. GNSS består av tre deler; en romdel, en kontrolldel og en brukerdel. Romdelen er selve satellitten som går i bane rundt jorden og sender avstandsdata til mottakeren. Kontrollsentraler og bakkestasjoner som følger satellittene, utgjør kontrolldelen. Sammen med data fra selve satellitten, vil brukerdelen motta korreksjonsdata fra kontrolldelen. Brukerdelen er mottakeren som mottar data og bruker dette til å beregne sin posisjon (Skogseth og Norberg, 2014, s.51-56).

Om dagen er smarttefonteknologi svært utbredt. Stadig flere har en smarttelefon selv, eller har tilgang på en. Hver moderne smarttelefon har vanligvis en intern GNSS-mottaker, og det har de siste årene blitt mulig å hente ut GNSS-posisjonsdata fra disse. Dette muliggjør bruken av en slik enhet som en GNSS-mottaker. Smarttelefoner er utstyrt med flere sensorer, som gjør det mulig å navigere utendørs, selv når en ikke har tilgang på gode GNSS-signaler (Dabove, Di Pietra og Piras, 2020).

### **2.4.1 Feilkilder og følgefeil**

Ved nøyaktige GNSS-målinger er det en rekke feilkilder en må ta hensyn til. Dette er blant annet banefeil, klokkefeil, flerveisinterferens, klokkefeil og elektronisk støy. Målinger gjort med telefon og lignende vil ikke gi like gode resultater som ved bruk av profesjonelt GNSS-utstyr. Det er derfor vanskelig å ta hensyn til, og dermed oppheve, disse feilkildene. En feilkilde det ikke tas hensyn til ved måling med telefoner er planlegging av målinger. Dette vil føre til at en ikke oppnår optimal satellittgeometri på alle målinger (Skogseth og Norberg, 2014, s. 63-64).

Det er mulighet for å miste signaler eller at de forstyrres, særlig om en befinner seg under tak eller omgitt av høye bygg. For å oppnå god navigering utendørs, er det sentralt å ha en grunnleggende forståelse for nøyaktigheten og kvaliteten på data som en samler inn med



telefon. Det er mulig å oppnå relativt gode resultater ved bruk av A-GNSS, eller assisted-GNSS, som er utbredt blant GNSS-kompatible mobiltelefoner (Dabove, Di Pietra og Piras, 2020).

Assistert GNSS brukes i hovedsak i mobile enheter, som telefoner og håndholdte GPS-er. Formålet med denne teknologien er å oppnå et bedre og mer nøyaktig resultat, og med dette forbedre mottakernes ytelse. A-GNSS er et system som ofte forbedrer oppstartsytelsen, det vil si TTFF, hos GNSS-mottakere. TTFF står for «Time To First Fix» og er et mål på hvor lang tid GNSS-mottakeren bruker på å oppnå «fix» etter den er skrudd på. Med «fix» menes forbindelse med satellitter som gir navigasjonsdata. TTFF forbedres ved at de nødvendige dataene overføres til enheten via et radionettverk, fremfor satellittkoblingen da satellittene overfører informasjon tregere. Mottakeren vil da «varmes opp» for fix og dermed redusere TTFF (Fernández-Prades *mfl.*, 2012).

## 2.4.2 Nøyaktighet

For krav til posisjonsbestemmelse med satellitter innenfor kart- og oppmålingsarbeider finnes standarden satellittbasert posisjonsbestemmelse. Standarden beskriver blant annet metoder for å kvalitetssikre målearbeidet. Feilkildene for GNSS-målinger, beskrevet i kapittelet over, vil variere over tid. Derfor kan to uavhengige målinger for samme punkt, med 45 minutter mellom hver måling, eller tre målinger med 15 minutter mellom, gjøre påliteligheten til punktmålingen bedre («Satellittbasert posisjonsbestemmelse», 2009).

Når metoden for innsamling av data planlegges, er det viktig å se på hva det endelige produktet skal være og hvilken informasjon det skal inneholde. Kravet til nøyaktighet vil avhenge av hva produktet skal fremstille. Innmåling av fastmerker vil ha et høyere krav til nøyaktighet en for eksempel kartlegging av turstier. Det vil derfor være både tids- og kostnadsbesparende å tilpasse metoden etter hva det endelige produktet skal vise.

Å bruke frivillig innsamlet data er en metode en kan ta i bruk for å spare både tid og penger. Bruken av VGI gjør at en kan unngå planlegging av prosessen i tillegg til innsamlingen og prosessering av data. Dette er ikke en metode som tas i bruk ved innsamling av svært nøyaktige data, men vil til gjengjeld redusere arbeidsmengden for de som samler dataene. Innsamling av data ute i felt tar mye tid, så innhenting av store datamengder vil være ressurskrevende (Skogseth og Norberg, 2014, s. 11-13).

I Kartverkets datasett tur- og friluftsruter ligger kravet til nøyaktighet på 10-15 meter. Det relativt lave kravet skyldes feilkilder når en bruker GNSS-mottaker (Sosi Produktspesifikasjon, 2017).

### 2.4.3 Kvalitet med eksterne GNSS-enheter

I en rapport skrevet av European GNSS Agency i 2020 (GSA, 2020), undersøkte de utviklingen innen GNSS-brukerteknologi og kvaliteten på denne typen data. De har konkludert med at det stadig skjer store fremskritt for denne teknologien og at behovene for kontinuerlig datainnsamling, automatisering og sikker posisjonering.

I desember 2016 åpnet den innledende tjenesten av Galileo. Dette gjorde det mulig for teknologiindustrien å begynne å utvikle mottakere som benytter seg av Galileo. Det skjedde en rask utvikling, og alle nye smarttelefoner begynte å bruke Galileo som standard. Rapporten viser til at rundt 70 prosent av de tilgjengelige GNSS-mottakere på markedet i dag kan ta i bruk Galileo-signaler. Drøye 60 prosent av mottakerne kan også ta i bruk signalene fra samtlige av de andre satellittnavigasjonssystemene, som GPS og GLONASS.

Avansert GNSS-teknologi har i lang tid vært forbeholdt profesjonelt landmålingsutstyr. Utviklingen innen forbrukerteknologien har muliggjort en økning av robusthet i posisjonsbestemmelsen, i tillegg til høyere nøyaktighet (GSA, 2020).

Det er fremdeles rom for forbedring for denne teknologien, da det ikke er mulig å utføre kartleggingsoppgaver med høy presisjon med denne typen mottakere. En rekke feilkilder, da særlig flerveisinterferens, kan ha stor innflytelse på målingene. Brukeren av enheten vil derfor oppnå et resultat med lav nøyaktighet (Yan *mfl.*, 2020).

Goodspeed *mfl.* (2018) har utført en studie hvor de sammenligner kvaliteten mellom GNSS-klokken Garmin Fenix 2, og mobiltelefonen Samsung Galaxy S5. Gjennomførelsen blir gjort av 28 frivillige brukere, som hadde på GNSS-enhetene i minimum åtte timer om dagen, uansett hva de gjorde. Kvaliteten for punktlokasjoner ble sammenlignet med steder som har stedfestet informasjon, som for eksempel skoler, treningsstudio og restauranter. GNSS-sporene utendørs ble sammenlignet med hverandre ved å se på hvor godt de går overens med Google Earths referansekart. Konklusjonen er at en mobil har mye bedre kvalitet innendørs. Dette skyldes trolig at telefonene har funksjoner som kan bruke WiFi og Bluetooth til å

forbedre sin posisjon, i motsetning til treningsklokken. Treningsklokken hadde mer nøyaktige data utendørs. Det gjaldt for både bil- og gåturer (Goodspeed *mfl.*, 2018).

Merry og Bettinger (2019) utførte en studie av GNSS-nøyaktigheten til en smarttelefon under ulike trær ved forskjellige årstider. Dette var for å samle inn data hvor det var ulik mengde løv på trærne. Personen som utførte målingene gikk en runde og målte hvert punkt en gang, to ganger om dagen. De konstaterte at den gjennomsnittlige HDOP-verdien var på rundt 7-13 meter, avhengig av årstidene.

Gilgen-Ammann, Schweizer og Wyss (2020) testet nøyaktigheten til ulike GNSS-klokker fra Garmin, Coros, Apple, Polar og Suunto. I undersøkelsen utførte de tester ved gåing, jogging og sykling i tre ulike geografiske områder; byområder, skog og bane. De sjekket klokkenes evne til å tolke riktig lengde over ulike målte distanser. Over distansen i skogområdene ble distansen alltid undervurdert av klokken, som nok skyldes færre punktmålinger og dermed dårligere geometrisk utforming av linjer. De kom fram til at fart hadde størst påvirkning på resultatet, i form av større avvik fra den virkelige distansen. Jogging var den aktiviteten som hadde størst påvirkning ved økning av hastighet. Dette skyldes trolig mer bevegelse av armen der treningsklokken var plassert, ved høyere fart.

## 2.5 GPX-formatet

GPX-formatet er en de facto standard for utveksling av GNSS-data mellom applikasjoner, men det er ikke en offisiell standard. Formatet er åpent, og uten lisens. GPX-filer er basert på XML-skjema, og data blir lagret i tagger. Filene må ha lokasjonsdata, mens data som tid og høyde er valgfritt. Koordinater blir registrert i WGS84-datamet (Topografix, ingen dato).

GPX-data blir registrert som punkter. Punktene kan registreres som waypoints eller trackpoints. Waypoints er enkeltstående punkter uten relasjon. Trackpoints er en sortert liste med relaterte punkter som beskriver en sti, der punkter blir registrert med et regelmessig mellomrom (Topografix, 2009). For å kunne gjenoppbygge innmålte stier i programvare er det nødvendig at punktene er registrert i sortert rekkefølge, derfor kreves det GPX-filer med punkter registrert som trackpoints til dette prosjektet.

## 2.6 Kernel Density Estimation

Et romlig punktmønster er et sett med koordinater som tilsvarer hendelser. En rekke GPX-trackpoints former dermed et romlig punktmønster. Punktmønstre kan analyseres for å kartlegge intensiteten av hendelser. En punktmønsteranalyse kan antyde om punktmønsteret er romlig strukturert eller ikke. Om punktene er clustret kan det tilsi en høyere sannsynlighet for at flere hendelser vil oppstå i noen områder. Om punktene er spredt er det mer sannsynlig at hendelser oppstår lengre unna hverandre (Lloyd, 2010).

En metode for å kartlegge hendelsesintensitet av punktmønstre, er kernel density estimation (KDE). KDE-analysen kan ha punkter og linjer som input, og output er et rasterlag. Analysen beregner tettheten av objekter i et nabolag, rundt disse objektene. Prosessen i analysen er å konseptuelt bygge en kurvet overflate over hver linje. Overflateverdien er størst ved linjen, og vil avta når en beveger seg fra linjen. Overflateverdien blir null når overflaten når en spesifisert søkeradius fra linjen (pro.arcgis.com, 2021a).

## 2.7 Tidligere arbeid

Over lengre tid har kart blitt konstruert ved bruk av flyfoto og satellittbilder. Metodene som tas i bruk er krevende og komplekse. Dette har ført til at dagens kart og bilder er kostbare å lage og oppdatere. Med stadig bedre tilgjengelighet til GNSS-data gjennom GNSS-mottakere i smarttelefoner og treningsklokker, har kartkonstruksjonsalgoritmer basert på VGI blitt realistiske å ta i bruk. Kartkonstruksjonsalgoritmer bruker GNSS-data i form av punkter til å rekonstruere et nettverk. Dette kan være rekonstruksjon av veger, stier, broer, fortau og tunneller. Selv om nøyaktigheten til GNSS-mottakere i smarttelefoner ikke kan sammenlignes med profesjonelt landmålingsutstyr, vises det at en form for midling av flere GNSS-spor kan brukes til å rekonstruere et nettverk (Zhou *mfl.*, 2020).

### 2.7.1 Eksisterende kartkonstruksjonsalgoritmer

I dette delkapittelet beskrives to eksisterende kartkonstruksjonsalgoritmer og deres metoder for å clustre datasettene. Først blir Kharita-algoritmen beskrevet, denne baserer seg på

clusteralgoritmen k-means. Deretter blir PAM-algoritmen beskrevet, denne baserer seg på clusteralgoritmen k-medoids.

## **Kharita**

Kharita-algoritmen spør om det er mulig å nøyaktig utlede det underliggende vegnettverket kun basert på GNSS-spor. Problemstillingen fokuserer på vegnettverk og -geometri. To algoritmer lages, en offline algoritme, og en online algoritme som kan skape og oppdatere kartet når GNSS-spor blir lastet opp.

Kharita-algoritmen er basert på clusteralgoritmen k-means. Algoritmen bruker først punktenes retning og plassering for å clustre punktene. Retning beregnes fra punkters posisjon i forhold til det påfølgende punktet. Videre plasseres punkter i det geometriske senteret av clusteret det tilhører, og kanter tegnes mellom punktene. Prosessen repeteres for hvert punkt. Siste steg i prosessen er en metode som bruker graph spanners for å fjerne kanter som ikke er korrekte. Graph spanners er en teknikk som kan være viktig ved rekonstruksjon av ulike former for kryss og rundkjøringer for bilveger, der rekonstruksjonen har høye krav til nøyaktighet.

Konklusjonen er at metodene deres betydelig overgår andre eksisterende metoder på grunn av deres utnyttelse av vinkler og hastighetsinformasjon, via clustering, og topologiske struktur, via graph spanners (Stanojevic *mfl.*, 2017).

## **PAM**

PAM-algoritmen undersøker muligheten for å bruke bare GNSS-spor samlet inn gjennom dognadskartlegging, for å automatisk trekke ut gangstier og -nettverk.

En clustermetode ble brukt for å finne signifikante punkter for å representere hver gruppe av punkter, representasjonene kalles medoids. For dette prosjektet ble clustermetoden Partitioning Around Medoids (PAM) valgt på grunn av dens robusthet mot støy og utliggere. Metoden minimerer ulikheten, eller den geometriske avstanden, til punktene i en klynge. PAM har to steg, først blir de initielle medoids valgt, deretter blir de initielle medoids byttet ut med ubenyttede medoids hvis funksjonen kan bli redusert. Prosessen itereres til funksjonen ikke lenger kan bli redusert. Nettverket blir konstruert etter alle medoids er valgt ut. Input er medoids hentet ut i forrige steg, og output er det genererte fotgjengernettet.

Konklusjonen er at algoritmen kan automatisk konstruere vegnettverk for fotgjengere med ulike GNSS-spor fra varierende kilder. De finner en korrelasjon mellom antall GNSS-spor og geometrisk korrekthet. De finner også at kvaliteten på det rekonstruerte nettverket er svært avhengig av kvaliteten til GNSS-sporene, spesielt i områder der det er dårlig sikt til satellitter (Kasemsuppakorn og Karimi, 2013).

## **2.7.2 Metoder for kvalitetsvurdering**

### **Kharita**

Kharita-algoritmen fokuserer på rekonstruksjon av vegnett med komplisert geometri, som rundkjøringer. Evaluering av kvaliteten gjøres med to metoder utviklet av Biagioni og Eriksson (2012) i tidligere forskning på kartkonstruksjonsalgoritmer, kalt GEO-metoden og TOPO-metoden. GEO-metoden fokuserer på geometrien, og måler hvor godt det rekonstruerte kartet stemmer overens med et referansekart av høy kvalitet. TOPO-metoden måler hvor nøyaktig kryss blir konstruert med korrekte koblinger til andre vegsegmenter. Artikkelen forteller at for Kharita-metoden er det TOPO-verdien som er den sanne testen for kartkvaliteten, da denne vil fortelle om metoden faktisk egner seg for rekonstruksjon av kompliserte vegnettverk. Beregningene vil gi en verdi i form av f-score som brukes for vurdering (Stanojevic *mfl.*, 2017).

### **PAM**

Fordi publikumsinnsamlede data ikke inneholder kvalitetsmålinger som HDOP og fordi det ikke eksisterte et referansekart av høy kvalitet for området, ble datasettet for PAM-metoden visuelt inspisert. Algoritmens ytelse ble validert ved å legge det rekonstruerte nettverket over flyfoto av høy kvalitet. Den visuelle inspeksjonen bekreftet at alle genererte vegsegmenter ble plassert korrekt over de faktiske vegene for fotgjengere. De finner også at topologien for det meste er korrekt. Konklusjonen av den visuelle inspeksjonen er at denne algoritmen også egner seg for publikumsinnsamlede data (Kasemsuppakorn og Karimi, 2013).

## 3 Metode

I metodekapittelet blir først midling, datasettet, programvare og verktøy forklart. Deretter blir metodeoppsett utdypet. Sist blir metode for kvalitetskontroll beskrevet.

Beskrivelse av metodeoppsett presenteres i tre deler. I metodene utdypes først preprosessering av GPX-filer for å klargjøre filene til videre behandling. Deretter utdypes metode for å estimere turstiens posisjon, både med tanke på geometri og topologi. I siste del utdypes prosessen for å trekke ut senterlinje fra estimatet.

### 3.1 Midling

Som beskrevet i kapittel 2.4.1 er det flere feilkilder som vil påvirke målinger. I standarden for satellittbasert posisjonsbestemmelse anbefales det å gjøre uavhengige målinger for å redusere muligheten for feil. Gjentatte målinger vil avdekke målinger som skiller seg ut med store avvik («Satellittbasert posisjonsbestemmelse», 2009). Midling av de gjentatte målingene for et punkt vil da kunne gi et mer pålitelig resultat. Midling vil tilsvare gjennomsnitt.

Som Goodchild (2007) forteller har resultatene av VGI en usikker grad av nøyaktighet. Kasemsuppakorn og Karimi (2013) finner at det er en korrelasjon mellom antall GNSS-spor og kvaliteten til et rekonstruert vegnettverk. Metodene som blir presentert her vil se på muligheten for å bruke VGI av ulik grad av nøyaktighet til å lage et pålitelig turstinnettverk.

### 3.2 Datasett

#### 3.2.1 Innsamling av data

Data ble hentet fra Bodø og Totenvika fra bekjente. Datasettene er innsamlet over en 10 årsperiode. Data i Bodø ble innsamlet med en Garmin Forerunner 920XT, dette er en multisportsklokke. Data i Totenvika ble innsamlet med en Garmin Edge 130, dette er en GNSS-sykkelcomputer.

### 3.2.2 Datasett og studieområde

Datasettene brukt i analysen kommer fra Bodø og Totenvika. Det er totalt 64 GPX-filer fordelt over Bodø og Totenvika. De geografiske områdene er varierte; Bodø har majoriteten av GNSS-sporene i skog og mark, mens de resterende GNSS-sporene er i boligområder. Totenvika består av en miks mellom veger og stier langs dyrket jord og stier i skog og mark.

#### GPX-datasett

Tabell 1: Oversikt over GPX-datasett

Datasett	GPX-filer	Tidsperiode	Trackpoints
<b>Bodø</b>	27	09/08/2016 – 17/02/2021	36016
<b>Totenvika</b>	36	08/05/2011 – 02/03/2021	92557
<b>Totalt</b>	64	2011 - 2021	128573

#### Øvrige datasett

Elveg 2.0 for Bodø

Elveg 2.0 for Totenvika (Østre Toten)

## 3.3 Programvare

### 3.3.1 ArcGIS-programvare

ESRI leverer en fullstendig pakke med GIS-programvare for å skape, analysere, behandle og dele geografisk informasjon (ESRI, 2020).



### **ArcGIS Pro versjon 2.7**

ArcGIS Pro er en programvare som gir tilgang til verktøy for analysering. Data behandlet i denne plattformen kan deles videre til andre ArcGIS-plattformer, som ArcGIS Online (pro.arcgis.com, 2020).

### **ArcGIS Online**

ArcGIS Online er en skybasert kartleggings- og analysetjeneste som egner seg til deling og presentering av data (doc.arcgis.com, 2020).

### **ArcGIS Instant Apps**

ArcGIS Instant Apps er en applikasjon for å bygge egne GIS-webapplikasjoner uten programmering. Gjennom ArcGIS Instant Apps kan brukere utforske karttjenesten på PC eller mobil. ArcGIS Instant Apps er integrert i ArcGIS Online-plattformen (ArcGIS og McPherson, 2021).

## **3.3.2 Jupyter Notebook og ArcPy**

Jupyter Notebook er en webapplikasjon som gir brukere mulighet til å skape og å dele dokumenter med Pythonkode. Gjennom Jupyter Notebooks får brukere tilgang til Pythonbibliotek som ArcPy (Developers.arcgis.com, 2020).

ArcPy er et Python-bibliotek som gir mulighet for å utføre automatisert geografisk dataanalyse med Python-kode. Biblioteket gir tilgang til Python snippets og scripts for å kjøre ArcGIS verktøy gjennom Jupyter Notebooks (pro.arcgis.com, 2021b).

## 3.4 Verktøy i ArcGIS Pro

I dette kapittelet gis en kort forklaring på verktøyene som blir benyttet i ArcGIS-programvaren.

Tabell 1: Oversikt over brukte verktøy i ArcGIS Pro

Verktøy	Beskrivelse
GPX to Feature	Konverterer punktinformasjonen i GPX-filer til Features i form av punkter i en Feature Class. Punktene vil inneholde geometrien (x, y, z) og åtte attributter, inkludert tidsattributtet DateTime.
Points to lines	Lager linjeobjekter fra punktdatasett. Punktene kan skilles fra hverandre og sorteres etter attributter, som tidsattributtet DateTime.
Merge	Kombinerer flere datasett av samme datatype til et enkelt output-datasett. Datatype kan være punkter, linjer, eller polygoner.
Reclassify	Reklassifiserer, eller endrer, verdiene i et rasterlag.
Buffer	Lager bufferpolygoner med en spesifisert avstand rundt inputobjekter. Input kan være punkter, linjer eller polygoner.
Aggregate Polygons	Kombinerer polygoner til nye polygoner, når de er innenfor en spesifisert avstand fra hverandre.
Raster to polygon	Konverterer rasterdata til vektordata i form av polygon.
Polygon to centerline	Danner senterlinjer i polygoner.
Trim lines	Fjerner linjer som kun har en node, og som er kortere enn en spesifisert lengde.
Features to Geodatabase	Laster opp en eller flere Feature Class til en geodatabase.

## 3.5 Metode 1 - KDE

Vi har benyttet Kernel Density Estimation (KDE) for å midle alle spor som frivillige har lastet opp. KDE ble valgt fordi det er et verktøy som er lett å bruke og forstå, og som også er integrert i ArcGIS Pro. KDE har blitt benyttet som metode for midling av GNSS-spor av blant andre Biagioni og Eriksson (2012).

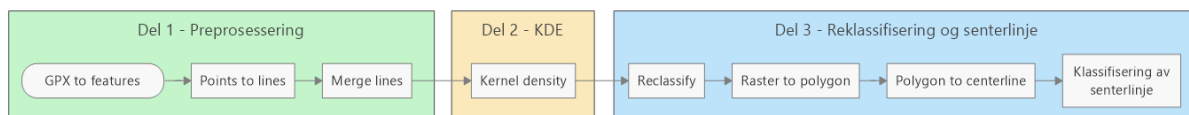
Denne metoden vil generere et rasterlag som kan brukes til å estimere hvor den faktiske turstien går. Dette er en prosess som gjøres i ArcGIS Pro, og består av flere steg. Stegene blir beskrevet i dette kapittelet.

Grovt sett kan denne prosessen deles opp i tre deler. Del 1 går ut på å endre datasettene slik at de er i ønsket format og organisert på en slik måte at datasettene blir enklere å håndtere i programvaren. Alle turstier vil bli lagt sammen til et Feature Class-objekt bestående av linjer.

Del 2 består av å gjennomføre KDE-analyse på alle linjer som er et resultat fra del en. Analysen gir et rasterlag som viser hvor tettheten av linjer er høyest, og med det også hvor det er mest sannsynlig at stien går.

Dette er den delen av prosessen som gir mest utfordringer. Hvert datasett vil ha ulike parametere i KDE-analysen siden datasettene ikke har samme utforming eller antall målinger. Det betyr at hvert datasett må behandles på nytt, og at det ikke er noen fastsatte parameterinnstillinger som kan gjenbrukes.

I del 3 kjøres prosessen for å trekke ut senterlinje fra KDE-analysen. Det innebærer å reklassifisere rasterlaget slik at vi kan skille dårlig og god data fra hverandre, og deretter trekke en senterlinje i midten av polygonen.



Figur 1: Flytdiagram av metode 1

### 3.5.1 Del 1 – Preprosessering

Før datasettene kan brukes til KDE-analysen, må de konverteres fra GPX-filer til et filformat som kan håndteres av ArcGIS Pro. Dette gjøres ved å bruke verktøyet «GPX to Features» i ArcGIS Pro.

Når datasettene er gjort om til Features, vil de være lagret som punkter. For å få turstiens geometri og topologi bedre frem, gjøres disse punktene om til linjer slik at hver linje representerer en innmålt tur. Dette gjøres for å unngå at det blir glipper i KDE-rasterlaget, som kan oppstå hvis det er stor avstand mellom punktene. Med verktøyet «points to line» konverteres punktene om til linjer. Hvert punkt har et attributt med tidspunkt for når målingen ble gjennomført. Denne informasjonen brukes til å sortere rekkefølgen på punktene, og dermed lage en linje.

KDE-analysen kan kun bruke et objekt som input, derfor må alle linjer for forrige steg slås sammen til en enkelt Feature. Det gjøres med «merge», og her hukes det av for «add source information to output». Det gjør det enklere å skille de individuelle turene fra hverandre, da navnet på turen som har blitt satt av den frivillige ikke nødvendigvis er unikt.

### 3.5.2 Del 2 – Analyse

KDE-verktøyet i ArcGIS Pro brukes for å kalkulere tettheten av enten punkter eller linjer. Rasterlaget fra KDE-analysen viser hvor intensiteten av linjer er høyest, og denne informasjonen kan brukes til å estimere turstien posisjon.

Dette verktøyet har flere parametere. De som har betydning for denne analysen er «output cell size», «search radius» og «output cell values».

«Cell size» er størrelsen på cellene i rasterlaget som denne analysen generer. Cellestørrelsen har mye å si for hvor stor filstørrelsen til rasterlaget blir. Den er satt til 1 x 1 meter for å få frem turstienes geometri, og for å få en høyere nøyaktighet.

«Search radius» bestemmer hvor stort område som skal inngå i beregningen av tettheten av linjer. Hvilken størrelse en setter vil variere fra datasett til datasett. Er størrelsen satt for liten kan noen stier blir splittet, fordi avstanden mellom linjene overgår søkeradiusen. Settes

søkeradiusen for stor vil stier som er separate bli slått sammen hvis avstanden mellom stiene er lavere enn søkeradius.

Hvilken søkeradius som passer best til et datasett er vanskelig å vite på forhånd. Derfor ble det konstruert et ModelBuilder-verktøy, som kan kjøre flere KDE-analyser på en gang med ulike parameterinnstillinger. På denne måten gjøres prosessen med å finne ut hvilken søkeradius som passer best mer effektiv. ModelBuilder-verktøyet ble satt opp med ti ulike søkeradius-parametere, fra 5 meter til 50 meter, med intervaller på 5 meter. Ved å visuelt inspisere de ulike resultatene, kan den som treffer best for hvert enkelt datasett velges ut. En søkeradius på 15 meter viste seg å være den innstillingen som traff våre mål for topologi og geometri best.

Tabell 2: Parameterinnstillinger brukt i KDE-analysen

Input point or polyline Features	Totenvika\Totenvika_Lines
Population field	NONE
Output raster	C:\Users\tomer\Documents\ArcGIS\Projects\InferringPedestrianNetwork\InferringPedestrianNetwork.gdb\KernelDensity_Totenvika
Output cell size	1
Search radius	15
Area units	SQUARE_METERS
Output cell values	DENSITIES
Method	PLANAR
Input barrier Features	

Resultatet fra denne analysen er et rasterlag som viser intensiteten av linjer over området. Det er flere måter å visualisere rasterlaget på. Standardinnstilling for symbologien er «natural breaks (jenks)». På grunn av store variasjoner i tettheten av linjer, passer «natural breaks» dårlig for denne typen datasett. For å få et rasterlag som kan skille ut stier, må symbologien til rasterlaget endres til «geometrical interval». Valg av antall klasser vil variere fra datasett til datasett, på samme måte som valg av passende «search radius» vil variere. For disse datasettene ble det brukt ti klasser for begge datasettene.

### 3.5.3 Del 3 – Reklassifisering og senterlinje

Rasterlaget brukes så til å estimere hvor turstien går. Første steg i denne delen er å reklassifisere rasterlaget, til «NODATA», 0 og 1. På denne måten kan vi kvitte oss med celler som vi ikke ønsker å ha med videre i prosessen (NODATA), og 0 og 1 brukes til å skille mellom en tursti som har blitt målt en gang eller flere ganger.

Hvilke klasser som settes til hva gjøres via visuell inspeksjon. For datasettet i Totenvika settes de fire første klassene til «NODATA», og for Bodø settes de første to klassene til «NODATA». Dette er for å luke ut områder som har ingen eller veldig lav tetthet av linjer. Disse områdene er ikke ønsket videre i analysen. For områder som kun har blitt gått en gang setter vi celleverdien til 0, og resterende klasser settes til 1.

Start	End	New
0	30,909259	NODATA
30,909259	46,264942	NODATA
46,264942	77,174201	0
77,174201	139,391054	1
139,391054	264,626562	1
264,626562	516,711518	1
516,711518	1024,130105	1
1024,130105	2045,506498	1
2045,506498	4101,421996	1
4101,421996	8240,748047	1
NODATA	NODATA	NODATA

Figur 2: Reklassifisering av KDE for Bodø

Neste steg i prosessen er å konvertere det reklassifiserte rasterlaget til en polygon. Det gjøres ved bruke verktøyet «raster to polygon». Huk av for «simplify polygons» for å gjøre den mykere i kantene. Siden raster består av celler som er firkanter, vil polygonen bli kantete uten å hake av denne boksen. Det hukes også av for «create multipart Features». Dette vil gjøre at Featuren kun består av to polygoner, slik at arbeidet med dele opp turstiene i to klasser, basert på antall målinger, lettere.

Det lages en kopi av Featuren slik at det er to i geodatabasen. Den ene beholdes slik den er. For den andre Featuren må vi slå sammen polygonene, klasse 0 og 1, slik at den blir en sammenhengende polygon. Ved å bruke verktøyet «polygon to centerline» kan en trekke ut hvor senterlinjen går. Dette verktøyet brukes på den Featuren hvor polygonene er slått sammen, for å få en sammenhengende linje. Det er denne linjen som representerer turstiens posisjon og topologiske utforming.

Til slutt deles senterlinjen opp i to klasser, en klasse for de linjene som representerer en tursti som har blitt målt flere



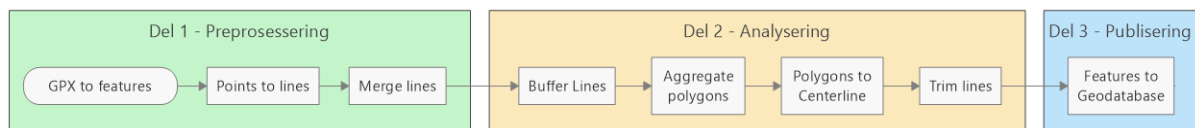
Figur 3: Senterlinje med KDE og GPX-punkter

ganger, og en klasse for de som kun har blitt målt en gang. Polygonen som inneholder klasser, 0 og 1, må deles opp i to Features. Deretter velges alle linjer som har sitt geometriske senter innenfor polygon, med klasse 0, ved hjelp av «select by location». Alle disse linjene får attributt «Confidence Class» satt til 0, og de resterende linjene får sin satt til 1. På denne måten kan en gi ulik visuell representasjon av senterlinjen.

Senterlinjen, som består av to klasser, er det endelige resultatet som viser turstiens posisjon.

## 3.6 Metode 2 - Buffer

Den automatiserte algoritmen blir skrevet med bruk av Python-kode med bruk av Jupyter Notebook som tekstbehandlingsprogram. GNSS-sporene blir midlet og lastet opp på ArcGIS Online som et «hosted Feature layer». Dette Feature Layeret kan brukes i ArcGIS sine applikasjoner for videre analyser, vi har brukt både Map Viewer og Instant Apps.



Figur 4: Flyttdiagram for metode 2

### 3.6.1 Del 1 – Preprosessering

For at ArcGIS skal kunne gjenkjenne trackpoints i GPX-formatet som punkter i programvaren, må vi konvertere dem til en Feature Class. Trackpoints kan bli brukt i analyseverktøy i ArcGIS via formatene shapefil eller Feature Class, men bruker en ArcPy for å utføre må filene være lagret som en shapefil på PC. Vi bruker «GPX to Features» for å konvertere og lagre alle GPX-filene til shapefiler, og «points to lines» for å konvertere punktene til separate linjer. Vi slår sammen alle shapefilene til en shapefil med bruk av «merge», dette gjør det lettere behandle og analysere.

### 3.6.2 Del 2 – Analyse

Med Buffer-verktøyet danner vi en buffersone rundt linjene, deretter bruker vi «aggregate polygons» for slå sammen polygonene der de overlapper. 10 meter i bufferradius og 10 meter

aggregeringsdistanse blir brukt som parameterinnstillinger. Vi bruker «polygon to centerline» for å danne en senterlinje i de sammenslåttene polygonene. Linjesegmenter som er mindre enn 150 meter i lengde og har kun tilknytning til en node vil bli slettet med bruk av «trim lines».

### 3.6.3 Del 3 - Publisering

Vi må manuelt lage et «hosted Feature layer» i ArcGIS Online. Funksjonen «overwrite Feature layer» blir brukt for å oppdatere et eksisterende «hosted Feature layer». En automatisk oppdatering av kartlaget kan bli gjort ved å lagre scriptet som en batchfil som kjøres så ofte en ønsker. «Overwrite Feature layer» krever at all data lastes opp i zippet format, og derfor blir en automatisk zipping av shapefilene utført.

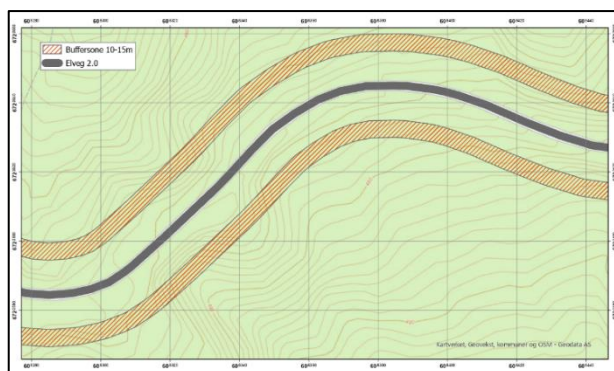
Python-koden følger med som vedlegg.

## 3.7 Vurdering av kvalitet og nøyaktighet

For datasettet tur- og friluftsruter blir innsendt data kvalitetssikret ved å sammenligne med primærdata, som datasettet FKB-TraktorvegSti, med en anbefalt nøyaktighet på 10-15 meter.

For dette prosjektet blir kvalitetssikringen basert på metoden fra datasettet tur- og friluftsruter. Nøyaktighetskravet settes derfor til 15 meter, og avvik blir målt fra referansedatasettet Elveg 2.0. Det blir gjort som stikkprøver der det finnes stier eller veger i Elveg 2.0. Hvis det blir kartlagt en ny tursti som ikke eksisterer i Elveg 2.0 vil stien ikke kunne kontrolleres mot referansedata.

Metodene blir kvalitetssikret ved å tegne en buffer rundt senterlinjen fra Elveg 2.0 som illustrert i Figur 5, deretter blir det gjort en visuell inspeksjon av dataene. Buffersonen markert med røde diagonale linjer viser marginen på 10 til 15 meter fra senterlinje. Denne buffersonen kan brukes til å visuelt inspisere om den midlede senterlinjen holder seg innenfor, eller krysser, nøyaktighetskravet på 15 meter.



Figur 5: Referansedata, Elveg 2.0, med buffersonen 10-15 meter



## 4 Resultater

I kapittel 3 «Metode» beskrives to ulike metoder for å lage et turstinnettverk av GNSS-spor. Metode 1 er en manuell prosess som bruker rasteranalyse for å skape et turstinnettverk basert på intensiteten av GNSS-sporene. Metode 2 er en automatisert prosess som skaper en buffer i vektorformat for å lage et midlet turstinnettverk.

I dette kapittelet vil resultatet fra metode 1 og metode 2 bli presentert i form av kart, med fokus på hvordan metodene danner geometri og topologi. Geometri beskriver vegkryss, parallelle stier, og krappe svinger. Topologi beskriver hvordan turstinnettverket henger sammen.

### 4.1 Metode 1 - KDE

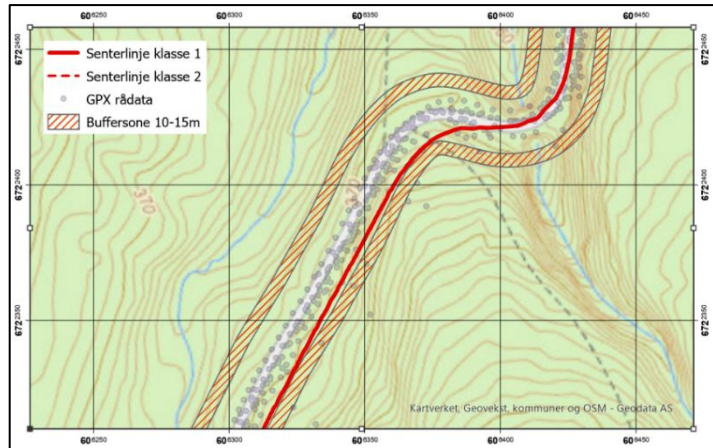
Etter å ha kjørt en analyse med Kernel Density Estimation, med en søkeradius på 15 meter, er det endelige resultatet en Feature-linje. Denne Feature-linjen inneholder to typer senterlinjer, kalt klasse 1 og klasse 2. Klasse 2 sin senterlinje har en større grad av usikkerhet knyttet til seg, da senterlinje av klasse 2 kun består av en underliggende måleserie.

Senterlinjen representerer den midlede verdien av GPX-punktene, og dermed turstien. Dette kapittelet vil gå igjennom hvilke områder metoden håndterer bra, og hvor det oppstår problemer. Det er to egenskaper vi vil se nærmere på, turstiens geometri og topologi.

#### 4.1.1 Geometri

Som nevnt i kapittel 2.2, er nøyaktigheten til datasettet tur- og friluftsruter satt til 10-15 meter. Vi bruker dette som vårt mål på kvalitet, for å se om nøyaktigheten er god nok. Med en buffersone på 10-15 meter rundt referansedata kan en visuelt inspisere om den midlede senterlinjen holder seg innenfor kravet, og når den nærmer seg grensen.

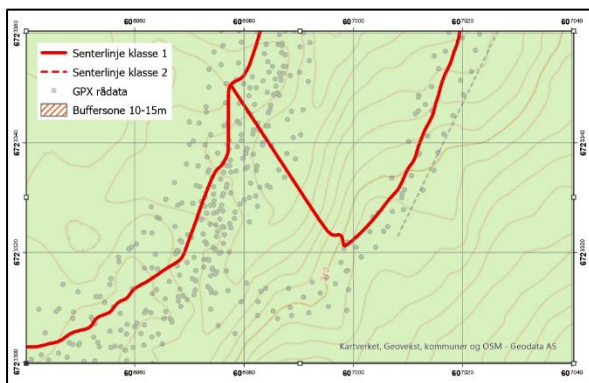
I Figur 6 kan en se et område hvor senterlinjen nærmer seg buffer. Senterlinjen holder seg innenfor buffersonen, som vil si at senterlinjen ikke avviker mer enn 10 meter fra referansedatasettet Elveg 2.0. Dette område har det største avviket, hvor referansedata er tilgjengelig. Det kan skyldes at dette området består av tett skog og kupert terreng. Likevel sett overordnet for hele området er posisjoneringen av senterlinjen god, da den aldri har et avvik utover 10 meter.



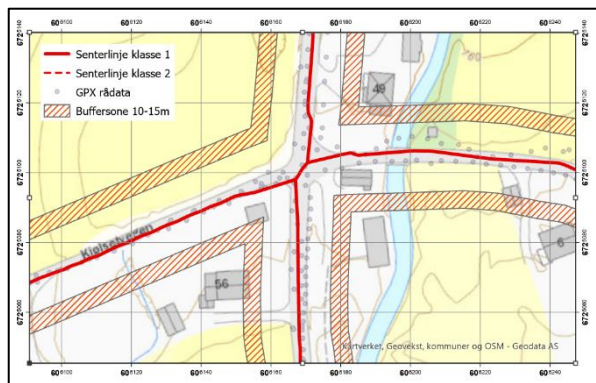
Figur 6: Eksempel på område hvor senterlinjen har stort avvik

En kan måle kvalitet på flere ulike måter, enn å bare se på avviket mellom senterlinjen og referansedata. For å undersøke den geometriske kvaliteten videre har vi valgt å dele den opp i tre ulike problemområder; vegkryss, parallelle stier og kurver. Vi har valgt ut disse tre områdene fordi de vanskeligste å få riktig, da særlig vegkryss.

Veger som krysser hverandre vinkelrett klarer metoden å gjenskape uten store problemer. Metoden har derimot større problemer med å gjenskape komplekse vegkryss, og når vegene krysser hverandre ved en lav vinkel, slik som i figur 7. I dette eksempelet sliter metoden med å gjenskape krysset på en geometrisk riktig måte. Når to stier kommer for nærme hverandre blir de slått sammen, og derfor får krysset feil posisjon.



Figur 8: Kryss med dårlig geometri

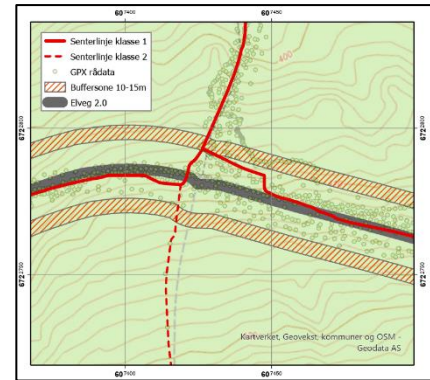


Figur 7: Vegkryss med god geometri

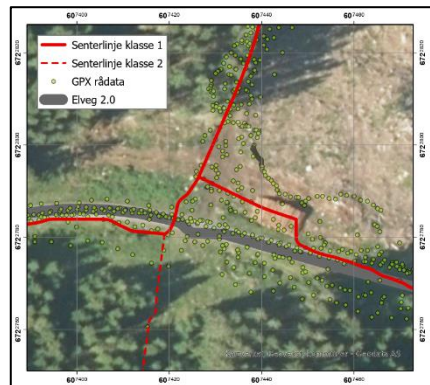
Enkelte steder får vegkryssene en utforming med vinkler som ikke ser naturlige ut. Det gjelder spesielt kryss hvor det er mange punktmålinger som er spredt over et større område. Dette skjer ofte hvor stienes kryss ikke er tydelig definert ute i naturen, slik at folk kutter svingen og går litt alternative ruter, og på denne måten spres punktmålingene utover et større område.

Områder hvor to veger eller stier går parallelt med hverandre klarer ikke metoden å gjenskape korrekt. Det skyldes at punktene ligger såpass tett, at KDE-analysen ikke klarer å skille stiene fra hverandre. Spesielt ved et område i Totenvika kommer dette problemet tydelig frem. I figur 13 kan vi se at vegene blir slått sammen, og at kryssets posisjon havner ca. 170 meter for langt sør.

Kurver blir håndtert bra. Det er ingen reduksjon i kvalitet eller nøyaktighet av betydning. Som i figur 12, hvor det er en 180 graders endring i retning, har det ingen betydning. I figur 11 kan vi derimot se at kurven ikke stemmer overens med de underliggende målepunktene. Antall punktmålinger og spredning av disse punktene vil påvirke hvordan kurver bygges. figur 12 viser en mer korrekt kurve, og figur 11 viser en mindre korrekt kurve. Ved figur 12 er det kun en underliggende måleserie, så det forgår ingen midling her. I figur 11 er det mange underliggende måleserier, og det gjør at punktene dekker et større område, og det blir vanskeligere å gjenskape kurven korrekt.



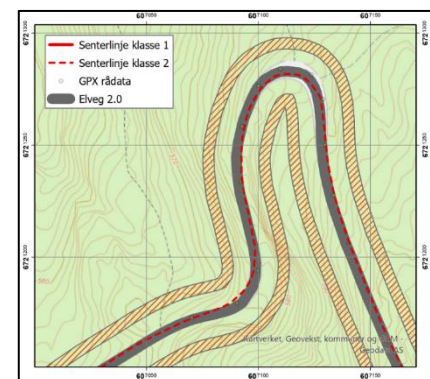
Figur 9: Vegkryss med dårlig geometri



Figur 10: Samme kryss som i figur 9, men med flyfoto



Figur 11: Kurve med flere underliggende målinger

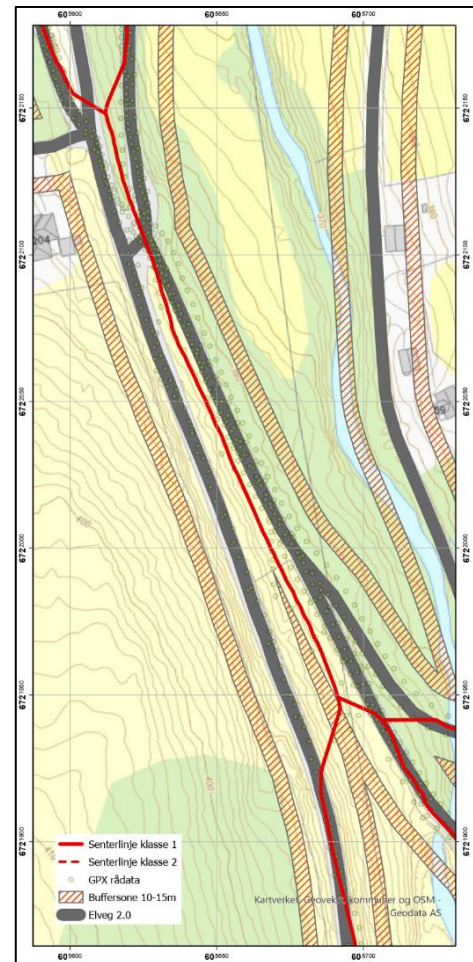


Figur 12: Kurve med en underliggende måling

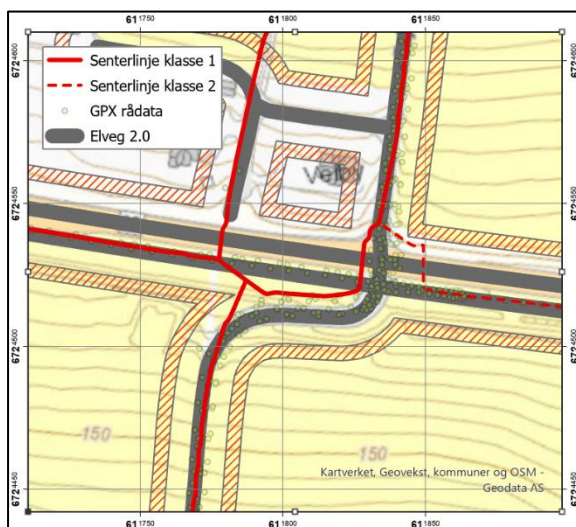
## 4.1.2 Topologi

Ved å bruke verktøyet «polygon to centerline» på en sammenhengende polygon, vil linjene alltid være koblet sammen og den topologiske konsistensen vil være gyldig. Det vil si at alle linjer som krysser eller går inntil hverandre vil være knyttet sammen. Det betyr likevel ikke at metoden har en helt korrekt topologisk utforming.

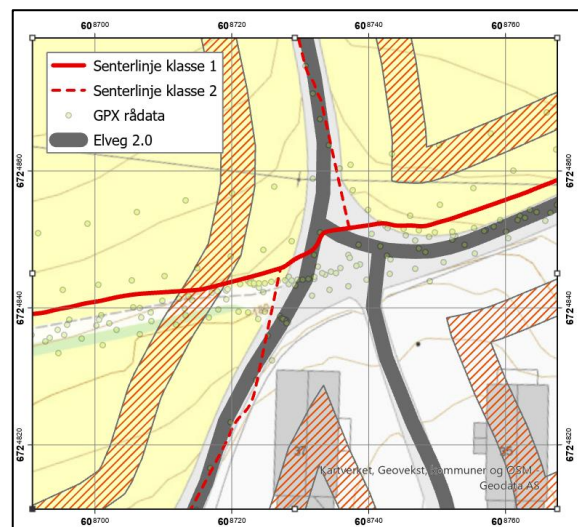
Topologien er jevnt over korrekt, men noen få områder skiller seg ut. Under kan vi se to eksempler på hvor antall vegkryss ikke er korrekt. I Figur 15 ser vi at de to parallelle vegene blir slått sammen, og dermed blir det dannet et kryss på feil sted. Dette krysset skulle da egentlig ha vært slått sammen med krysset som ligger litt øst. I Figur 14 ser vi at et kryss har blitt delt opp, slik at det er to vegkryss i stedet.



Figur 13: Vegkrysset med dårlig topologi



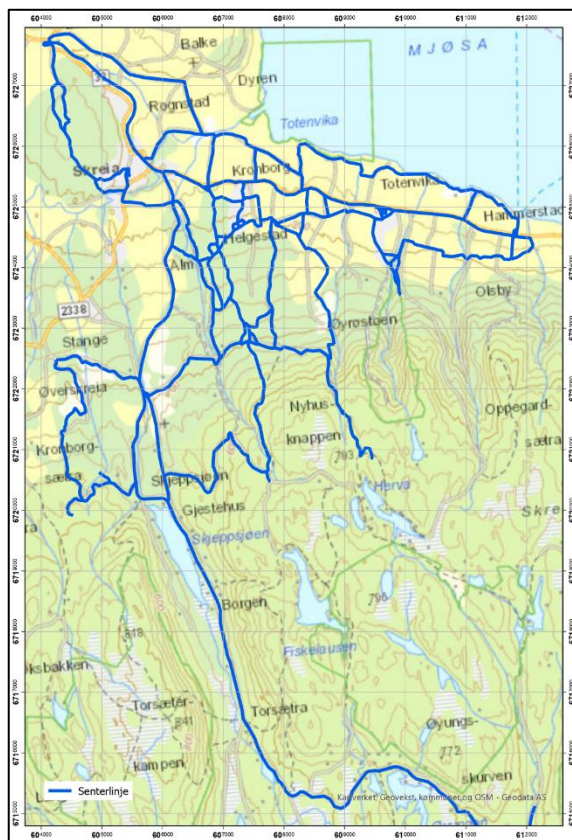
Figur 15: Område hvor parallelle veger har blitt slått sammen



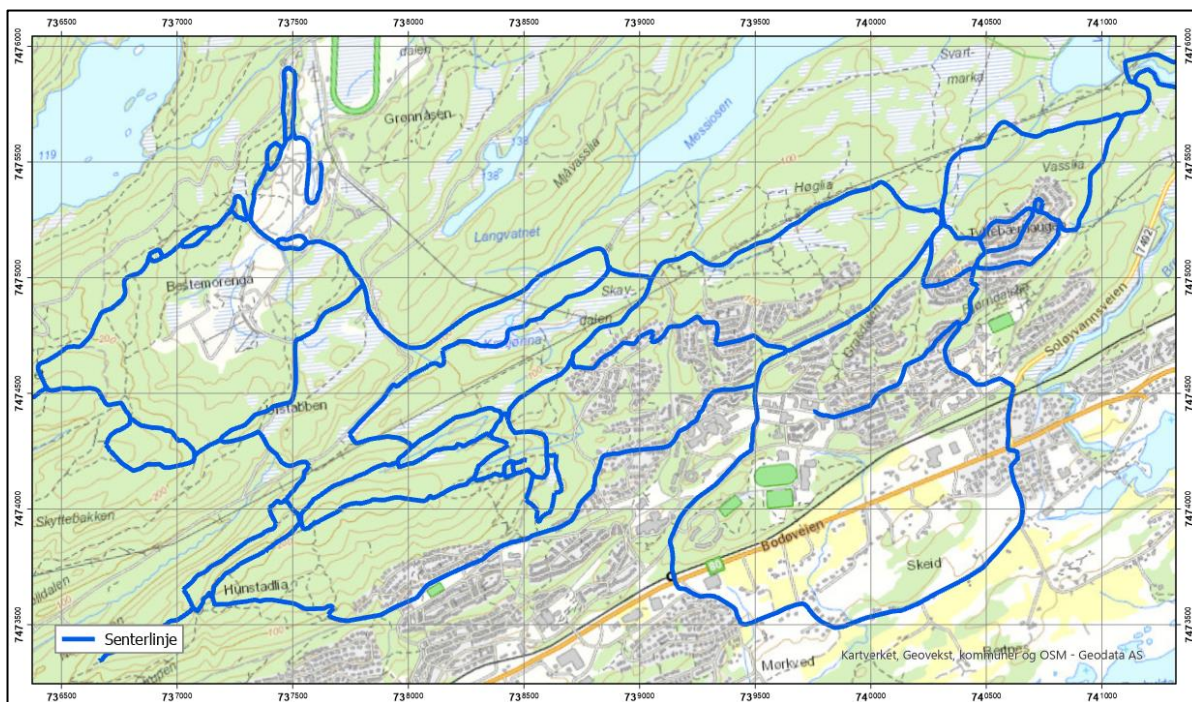
Figur 14: Vegkryss med feil antall kryss

## 4.2 Metode 2 – Buffer

Etter å ha kjørt analysen med polygonaggregering med en bufferradius og aggregeringsdistanse på 10 meter, blir det endelige produktet lagret i ArcGIS Online som et «hosted Feature layer». Det ferdige produktet vises i Figur 16 og Figur 17 der den midlede senterlinjen er vist i blå farge. Dette kapittelet vil gå gjennom ulike komplikasjoner i geometrien og topologien, og vil vise hvordan ulike parameterinnstillinger kan endre plasseringen til senterlinjen.



Figur 16: Oversikt over Totenvika

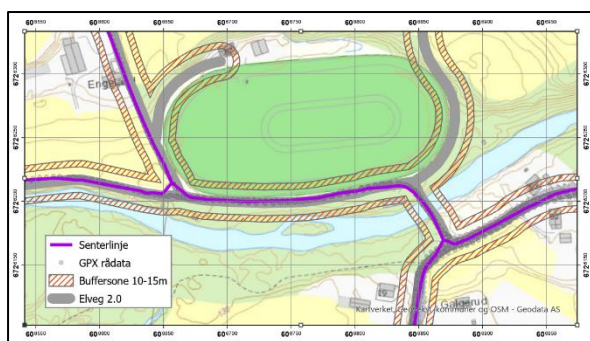


Figur 17: Oversikt over Bodø

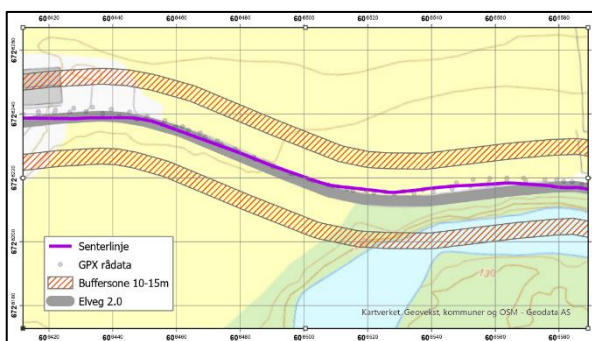
## 4.2.1 Geometri

Nevnt i 2.2 og 4.1.1, er nøyaktigheten til Kartverkets tur- og friluftsruter satt til 10-15 meter. Vi har valgt parameterinnstillinger i analysen med mål om å havne innenfor disse nøyaktighetskravene. En visuell inspeksjon blir igjen gjort mot referansedatasettet Elveg 2.0.

Figur 18 og figur 19 er tatt fra de åpne områdene i Totenvika hvor GNSS-sporene ligger godt på underliggende veger og tydelig innenfor Elveg 2.0 buffersone.

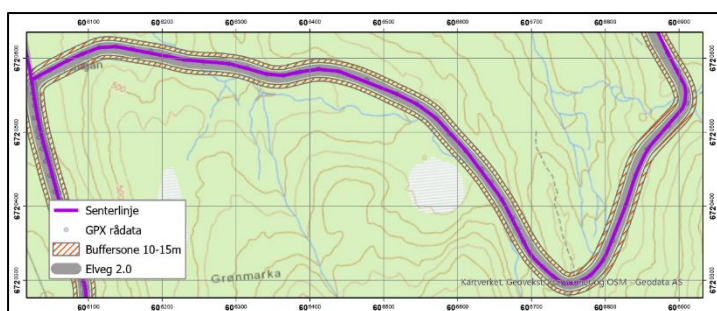


Figur 18: Åpne områder i Totenvika



Figur 19: Åpne områder i Totenvika

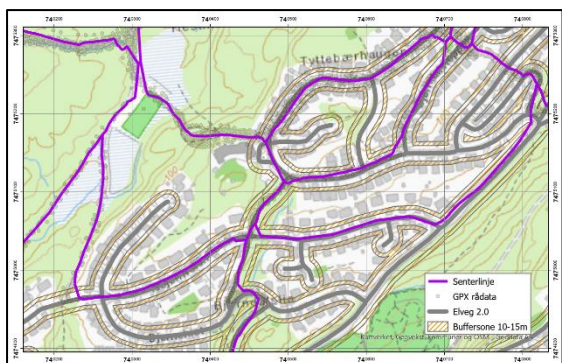
I Totenvika datasettet kan vi estimere en kvalitet ned på 2-5 meter langs stier eller veger ved de åpne områdene. Avvik vil forekomme selv i de åpne områdene, dette er noe som vises tydeligere i Figur 23.



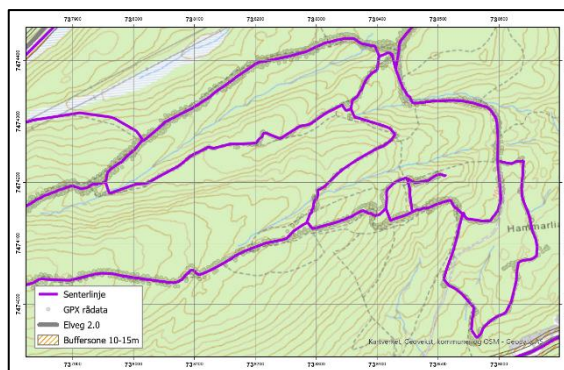
Figur 20: Kupert terreng i Totenvika

I Figur 20 er det mer grønt og kupert terreng. Her ser vi at GNSS-sporene fortsatt ligger relativt bra i forhold til de underliggende stiene i referansedatasettet.

I Bodø-datasettet er måleseriene stort sett i kupert terreng, men det eksisterer også måleserier fra ett boligfelt. Boligområdet kan regnes som et åpent område med et krevende nettverk av kryss og svinger. Senterlinjen kutter over svinger, og kryss blir feilplassert.



Figur 21: Boligfelt Bodø



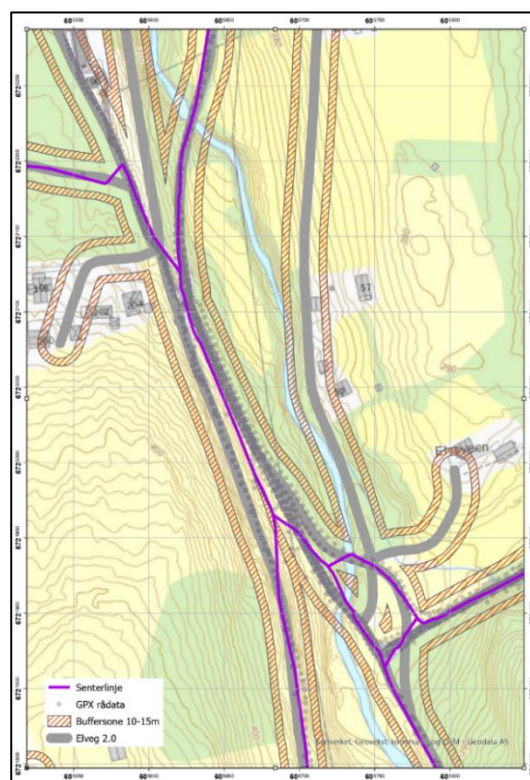
Figur 22: Kupert terreng Bodø

I Figur 22 ser en noe av det mer kupert området i Bodø. Stiene ligger ikke like godt på referansekartet GeocacheBasis, som blir brukt der Elveg 2.0 ikke dekker.

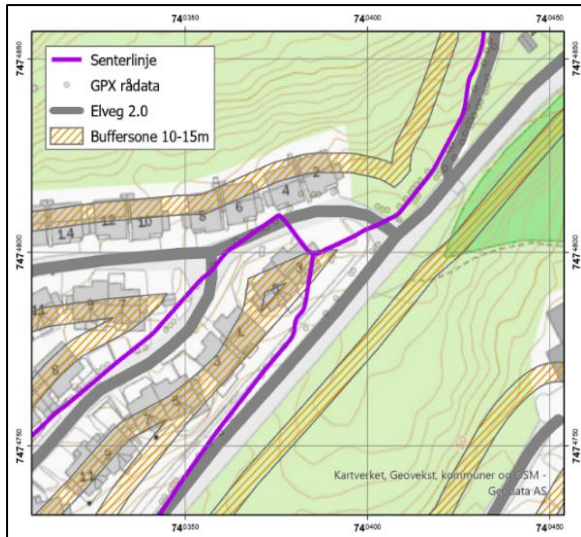
Veger som krysser hverandre vinkelrett vil gi en generelt god utforming på vegkryss sett i forhold til referansedatasettet. Når vinklingen blir lavere mellom vegene oppstår midling av polygonene før krysset, og de kan da midles flere hundre meter før det egentlige krysset. Flere innfallsvinkler i et kryss vil gi metoden problemer med å gjenskape korrekt topologi. I Figur 23 ser vi hva som skjer når parallelle linjer havner for nær hverandre. Den midlede senterlinjen avviker fra den øvrige vegen på rundt 13 meter før krysset, og midler linjene omtrent 143 meter før det egentlige krysset.

Det er flere utstikkere og parallelle linjer der det er kupert terreng. Kryssløsningene kan bli dårlig plassert i forhold til referansedatasettet. Det største avviket vi klarte å finne fra den midlede senterlinjen til en sti på referansedatasettet, er på omtrent 15 meter.

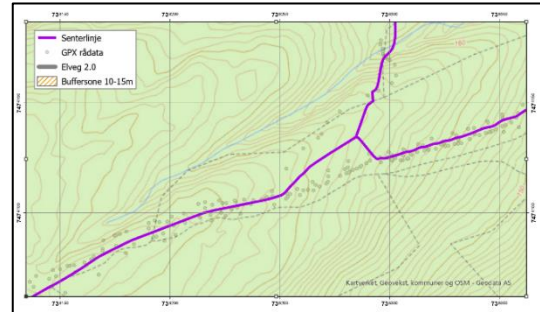
Stier som er tettliggende og går parallelt vil gi problemer i metode 2, som i metode 1. Parallell stier i Totenvika gjør at metoden midler GNSS-sporene på de ulike stiene før krysset og havner over buffersonen til referansedatasettet. En parallell sti i Bodø-datasettet, vist i Figur 25 midler to nærliggende stier. Dette gjør at den nye stien ligger imellom stiene som er utgangspunktet i midlingen.



Figur 23: Problematisk vegkryss i Totenvika



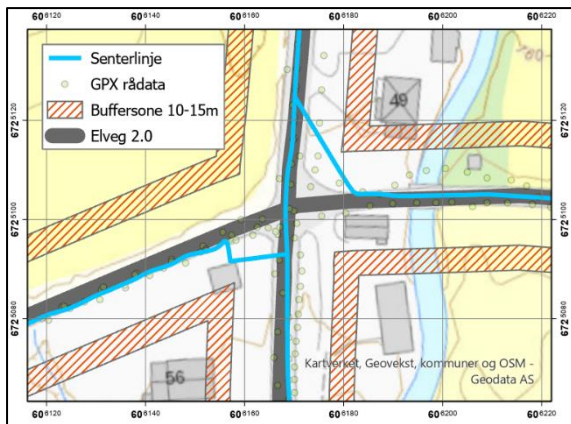
Figur 25: Dårlig kryssløsning i Bodø



Figur 24: Parallelle stier i Bodø

## 4.2.2 Topologi

«Aggregate polygon» gjør at polygonen slår seg sammen til ett polygon der de overlapper. Ved flere enn tre innfallsvinkler i et kryss kan det bli dannet flere noder. I Figur 26 kan vi se et kryss der det blir dannet to noder istedenfor en node.



Figur 26: Topologisk ukorrekt kryss

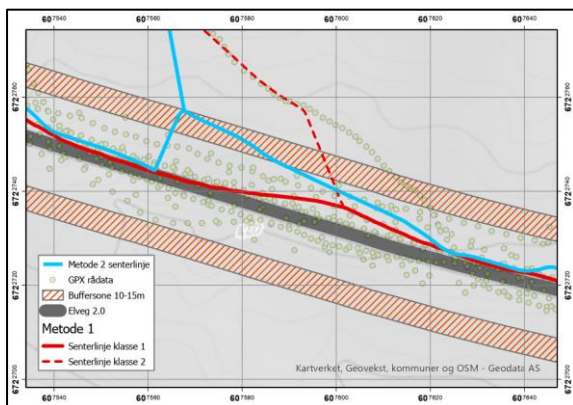


## 4.3 Metode 1 versus Metode 2

Metode 1 og metode 2 danner to ulike nettverk med de samme datasettene. Utformingen på vegkryss, linjer og kurvatur er annerledes i nettverkene.

Ved visuell inspeksjon har vi sett at begge metoder klarer å danne et nettverk, som i stor grad holder seg innenfor nøyaktighetskravet på 10-15 meter, men vegkryss med komplisert geometri har vist seg å være et område hvor begge metoder sliter.

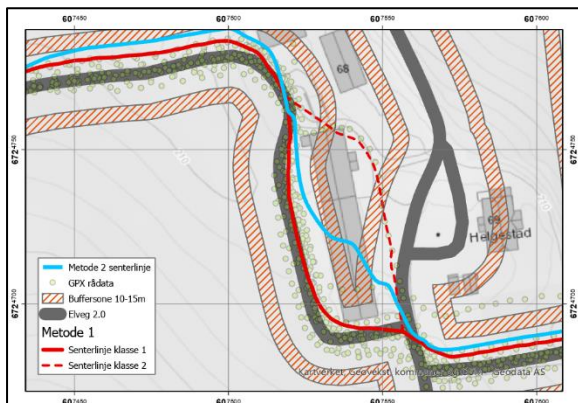
Det forekommer feil i topologien i vegkryssene. Ofte dannes det flere noder der det skal være en, og motsatt. Dette gjelder begge metodene, men vi ser at metode 1 oftere klarer å gjenskape vegkryss riktig enn metode 2. En sammenligning av fire ulike kryss er vist i Figur 29, Figur 30, Figur 27 og Figur 28. Vi ser glattere kurver i metode 1 og metoden danner bedre representasjon av eksisterende veg fra referansedatasettet.



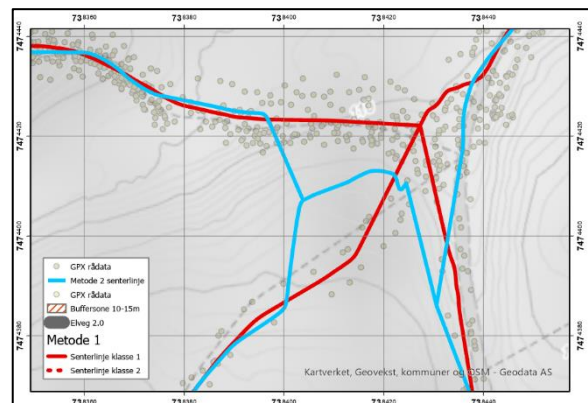
Figur 29: Sammenligning av metodene 1.1



Figur 30: Sammenligning av metodene 1.2



Figur 27: Sammenligning av metodene 1.3



Figur 28: Sammenligning av metodene 1.4

## 5 Diskusjon

I kapittel 4 «Resultat» ble resultatet av metode 1 og metode 2 presentert. I dette kapittelet vil det først drøftes om VGI er en egnet kilde til konstruksjon av turkart. Så vil metoden og resultatet, samt nøyaktigheten og kvaliteten til hver metode diskuteres. Til slutt vil metode 1 og metode 2 vurderes mot hverandre.

### 5.1 Om bruk av VGI

Som omtalt i delkapittel 2.1, er det en rekke utfordringer som oppstår ved bruk av VGI. En av disse er at de frivillige som samler inn data sjeldent har landmålingskunnskap. Dette leder til en usikker grad av nøyaktighet i målingene. En mulighet for å bedre resultatet er å implementere en metode for kvalitetssikring, som Antoniou *mfl.* (2017) foreslår i sin artikkel. Dette vil føre til et ytterligere steg i metoden for databehandling.

En annen ting som kan by på problemer er at det ikke samles inn tilstrekkelig mengde med data. For å kvalitetssikre dataene sine, kan en opprette et minstekrav for hvor mange GNSS-spor som må gå over samme sti før den legges til i nettverket. Et mulig resultat av dette er at det ikke samles inn en tilstrekkelig mengde data til å opprette en ny sti.

GNSS-spor fra utendørs aktivitet vil aldri være helt like. De registrerte bevegelsesmønstrene vil alltid ha en varierende grad ulikhet fra hverandre og kan endre seg fra dag til dag. Faktorer som hvilken side av vegen og hvilken retning en velger har innvirkning på GNSS-sporet. En syklist vil for eksempel velge en annen plassering på stien når de sykler opp en bakke, enn de vil når de sykler ned samme bakke. Ved bruk av VGI til kartlegging av kjøremønster med bil vil en få likere GNSS-spor. Grunnen til dette er at kjøretøy konsekvent følger de samme trafikkreglene og i større grad har et likt bevegelsesmønster.

Det er også problematikk knyttet til personvern. Ved regelmessig eller kontinuerlig loggføring av en persons bevegelser vil en kunne samle mye informasjon om livet deres. Bosted, arbeidsplass og tidsstempler på når en er ute av huset er sensitiv informasjon som ikke kan deles vilkårlig. Uvedkommende kan få tak i denne informasjonen ved å se på start- og stoppested, samt tidsstempler, for den innsamlede dataen. Enkelte applikasjoner for

smarttelefon gjør det også mulig for brukeren å legge ved kommentarer til deres treningsturer. Det er mulig enkelte finner deling av dette sjenerende.

## 5.2 Valg av metoder

Valg av metode for midling av GNSS-spor var et vanskelig valg. Det finnes mye forskning innenfor dette området, men mye av det tar utgangspunkt i kartlegging av bilveger, og dermed bruk av GNSS-sensorer montert på biler. Dette gir et annet utgangspunkt for midling da biler kjører på veger med fastsatte regler og har til en viss grad forutsigbart mønster, som gjør det enklere å gjennomføre støyfjerning.

Det var to forskningsartikler som skilte seg ut, og som vi valgte å undersøke nærmere. Det var en algoritme kalt Kharita som benytter seg av k-means (Stanojevic *mfl.*, 2017), og en annen algoritme som benytter seg av PAM (Kasemsuppakorn og Karimi, 2013). Disse algoritmene er beskrevet nærmere i kapittel 2.7.12.7.1.

Kharita-algoritmen er laget med tanke på å kartlegge bilveger og kompliserte vegkryss, og har dermed andre behov for kartlegging enn det vi har. Kildekoden var open-source, og derfor fant vi denne interessant, siden koden fritt kan lastes ned. Med denne kildekoden tilgjengelig undersøkte vi muligheten for å tilpasse koden til å håndtere gående i stedet for kjørende.

Ved hjelp av et Python-script, som vi kodet, kan GPX-filer konverteres til riktig format slik at Kharita forstår innholdet. Etter en del arbeid fikk vi koden til å kjøre, som da ga et resultat i form av en CSV-fil. Denne filen inneholder koordinater for alle linjer som har blitt generert. Ved å laste inn denne filen i ArcGIS Pro så vi at resultatet ikke var brukbart til vårt formål. Algoritmen genererte ingen tydelig midlet linje. Linjene som ble generert var av korte lengder, og linjene gikk i flere retninger. Vi la derfor fra oss denne metoden, da vi innså at det kom til å bli for tidkrevende å få denne til å fungere slik vi ønsket.

Parallelt som vi jobbet med Kharita, ble det undersøkt om clustring-metoden PAM kunne brukes i vår oppgave. Forskningsartikkelen hadde et fokus på gangstier i stedet for bilveger (Kasemsuppakorn og Karimi, 2013). Denne metoden er ikke open-source, og vi ble derfor nødt til å skrive egen kode, ut ifra det vi kunne lese i artikkelen. Noen forsøk på å skrive kode

ble gjort, men fremdriften gikk sakte. Siden våre kunnskaper innen avansert statistikk og Python-programmering er begrenset, forstod vi at dette ikke ville gå.

Utfordringene med å bruke andre eller lage egne algoritmer gjorde så vi måtte finne en annen løsning på problemet. Vi valgte da å fokusere på verktøy som allerede eksisterer i ArcGIS-plattformen.

## **5.3 Drøfting av metode 1 – KDE**

Ulike problemer har oppstått ved midling av GNSS-spor med metode 1. Dette blir diskutert innledningsvis. Så vil resultatene fra metoden diskuteres og sammenlignes med primærdata, og til slutt en gjennomgang av kvalitet og nøyaktighet.

### **5.3.1 Diskusjon av metode**

Et alternativ vi først vurderte å bruke var buffer rundt måleseriene, og trekke en senterlinje ut ifra det. Problemet med å bruke buffer er at den ikke midler alle sporene. Ved å bruke bufferfunksjonen finner en kun midten av de to mest ytterliggående målingene, og det gir ingen indikasjon på hvor flertallet av måleseriene ligger.

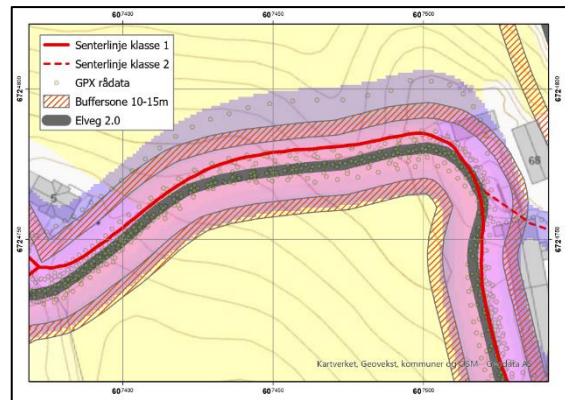
Som beskrevet i kap. 3.6, kartlegger KDE intensiteten av linjer og punktmønstre. Intensiteten gir et godt bilde på hvor stien mest sannsynlig befinner seg. KDE kan bruke både linjer og punkter som input i analysen. Noen av måleseriene har relativt stor avstand mellom punktene, og i kombinasjon med områder som kun har en måleserie vil det kunne gi glipper i KDE-rasterlaget. Det vanskeliggjør prosessen med å lage senterlinje siden polygonen ikke vil være sammenhengende. Derfor valgte vi å konvertere punktene til linjer.

Fordelen med å bruke KDE er at det er en enkel og rask metode, og krever kun grunnleggende GIS-kunnskaper. Ingen programmering er nødvendig. En annen fordel ved å bruke KDE er at vi kan sette en grenseverdi for å luke ut usikker data. Hva denne grenseverdien vil være vil variere fra datasett til datasett, på grunn av ulik intensitet i rasterlaget. Vi har brukt visuell inspeksjon til å finne verdien som skiller en sti som har blitt målt en gang fra de stiene som er målt to eller flere ganger. Det gjør at vi ikke trenger å

forkaste data for de områdene som kun har blitt målt engang, men heller merke de stiene slik at brukeren forstår at det er stor usikkerhet knyttet til de stiene, og at de må utøve varsomhet.

Å bruke intensitet som et mål på hvor midten av måleseriene befinner seg, er ikke uproblematisk. Hvis en måleserie har en følgefeil, og den konsekvent ligger 10-25

meter feil, vil den kunne påvirke hvor senterlinjen havner uforholdsmessig. I Figur 31 kan vi se et eksempel på dette. En dårlig måleserie flytter senterlinjen omtrent 6 meter lengre nord i forhold til Elveg 2.0. Senterlinjen blir laget med KDE-polygonen, og tar ikke hensyn til hvor intensiteten er høyest.



Figur 31: Område med dårlig måleserie

Andre utfordringer med å bruke denne metoden er at den er vanskelig å automatisere, og er uegnet til å brukes på større områder. Disse to henger delvis sammen, siden det som skaper utfordringer er knyttet til at metoden bygger på rasterdata fra KDE-analysen.

Rasterdata er plasskrevende, og dette kommer tydelig frem i kapittel 5.4 «Drøfting av metode 2 – Buffer». Det å skulle gjøre en KDE-analyse på datasettet for Bodø og Totenvika samtidig fører til at rasterlaget blir svært stort, og dekker halve Norge. Et slikt rasterlag, med cellestørrelse på 1 x 1 meter får en lagringsstørrelse på ca. 400GB. Disse filstørrelsene blir fort u håndterlig. Analysen blir svært tidkrevende å gjennomføre og kan ta opptil flere timer.

Metoden krever en del visuell inspeksjon. Spesielt når det kommer til å endre symbologien for å finne hvor grensen går mellom stier målt en gang og flere ganger. Siden denne grenseverdien vil være forskjellig for ulike datasett kan en ikke sette en fast celleverdi som grense. Dette i kombinasjon med rasterdata er uegnet for store områder, gjør at automatisering blir svært komplisert å gjennomføre.

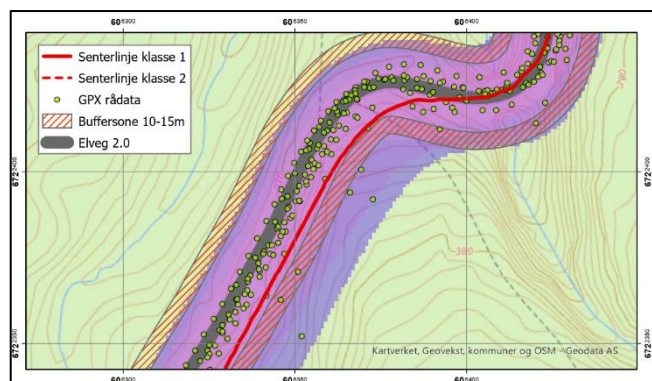
I etterkant ser vi at det er flere ting vi kunne ha gjort annerledes. Nøyaktigheten kunne ha blitt bedre hvis vi hadde valgt å ikke ta med område med kun en måleserie. Å inkludere disse påvirker nøyaktigheten til områder med flere målinger, særlig i de områdene hvor det finnes en måleserie som ligger litt til siden, som vi kan se i Figur 31. Hadde vi valgt å øke grenseverdien for hva som skal være, kunne vi ha luket ut slike utligger, og fått en mer nøyaktig representasjon av stien.

En annen måte å redusere støy på er å fjerne de første og de siste fem sekundene av punktmålingene i en måleserie. Som nevnt i kapittel 2.4.1 «Feilkilder og følgefeil» kan GNSS-mottakere være unøyaktig inntil mottakeren får fix, og denne perioden kalles Time-To-First-Fix (TTFF). Vi ser også i noen av datasettene som er samlet inn, at målingene har blitt startet og avsluttet tett inntil boligen, noe som gir målepunkter med stor unøyaktighet. For å gjennomføre denne støyfjerningen, må det skrives et script som kan bli kjørt på GPX-filene. ArcGIS har ingen integrerte verktøy som kan utføre dette på en effektiv måte.

I metode 2 blir verktøyet «trim lines» benyttet for å fjerne små vegstubber, også kalt dangles, da metoden generer mange vegstubber. I metode 1 blir ikke vegstubbene fjernet, da de ikke var et problem, men sett i etterkant kunne en ha fjernet stubber under 50 meter. Begrunnelsen er at datasettet Elveg 2.0 kun inneholder veger som er lengre enn 50 meter, og det er usikkerhet knyttet til om vegstubber under 50 meter faktisk er riktig, eller om det bare skyldes støy.

### 5.3.2 Diskusjon av resultat

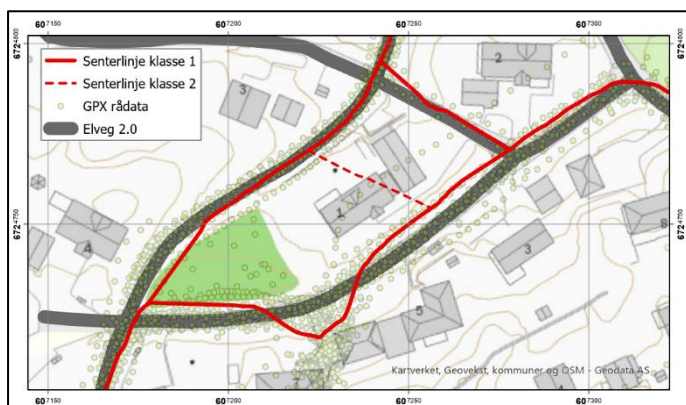
I Figur 32 ser vi hvordan noen få avvikende punktmålinger, til høyre for vegen, gjør at KDE-området (lilla) blir strukket ut, og dermed forskyver senterlinjen i forhold til hvor Elveg 2.0 ligger. Ved å endre grenseverdien kan en gjøre et mer nøyaktig estimat av turstiens posisjon, men med den ulempen at områder som blitt målt få ganger ikke vil bli inkludert i turstinnettverket. Det må derfor gjøres en avveining mellom nøyaktighet og mengde data en ønsker å beholde.



Figur 32: Område med utliggere

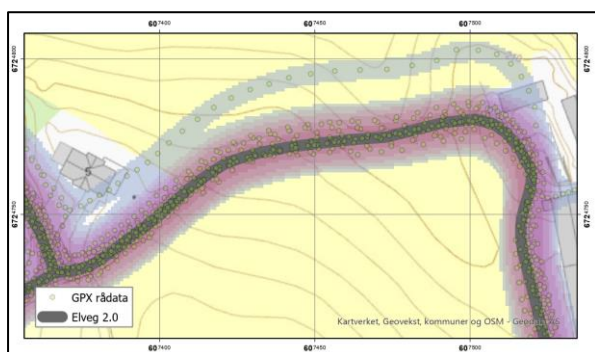
Vi har prøvd å finne et kompromiss mellom nøyaktighet og det å beholde mest mulig data. Som nevnt tidligere gjør vi dette ved å dele det inn i forskjellige klasser. I Figur 32 kan vi se et eksempel på at sti med klasse 2, og som vi ser går denne gjennom et hus. Det viser at klasse 2 ikke nødvendigvis eksisterer i virkeligheten, men heller er et resultat fra at punktene ligger veldig tett, og analysen ikke klarer å skille de fra hverandre. Antageligvis vil dette være et mindre problem i utmark og skog, fordi turstier som regel vil befinne seg relativt

alene. Punkter fra flere ulike stier vil derfor sjeldent overlappe og skape usikkerhet.

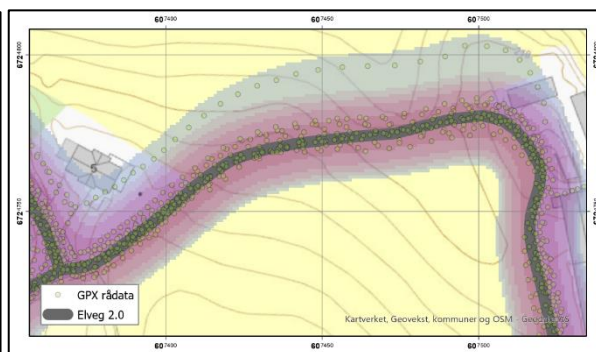


Figur 33: Område med mange punktmålinger

Det er mulig å få resultater som er mer nøyaktig enn det vi klarte å oppnå. Ved å sette ned KDE-søkeradius fra 15 meter til 10 meter ser vi at enkelte strekninger blir delt i to. Dette ser vi et eksempel på i Figur 34 og Figur 35. Den måleserien som avviker fra resten, kan fjernes igjen ved å sette grenseverdien høyere, men da på bekostning av stier som bare har blitt målt en gang også blir luket ut.



Figur 35: KDE med 10m søkeradius



Figur 34: KDE med 15m søkeradius

Resultatene fra analysen viser at den klarer å ta i bruk frivillig innsamlet GNSS-data og gjøre det om til et turstinnettverk. Med visuell inspeksjon og ved bruk av Elveg 2.0 som referansedata, har vi sett at linjen holder seg innen nøyaktighetskravet til datasettet Tur- og friluftsruter som er 10-15 meter.

Vi har ikke klart å effektivisere prosessen så mye som ønsket, siden metoden ikke lot seg automatisere. Metoden krever en del visuell inspeksjon, og da må noen med GIS-kunnskaper gjennomføre denne analysen. Det betyr likevel ikke at metoden er uten verdi. Med metoden slipper en å tegne inn stiene manuelt, ut ifra en visuell estimering av hvor stien befinner seg. Det er ikke uvanlig å gjøre det på denne måten, med hjelp av flyfoto.

Denne metoden kan brukes som et supplement til kartlegging av nye turstier som befinner seg i utmark. Spesielt turstier som befinner seg i skog, som er vanskelig å tyde fra flyfoto.

### 5.3.3 Diskusjon av nøyaktighet og kvalitet

For å måle hvor korrekt resultatet blir, har vi valgt å bruke visuell inspeksjon. Dette er en metode som er vanlig å bruke innen denne typen undersøkelser. Som nevnt i kapittel 2.7.2 «Metoder for kvalitetsvurdering» bruker PAM-metoden visuell inspeksjon til å evaluere kvaliteten og nøyaktigheten (Kasemsuppakorn og Karimi, 2013). De gjør det ved å sammenligne flyfoto og resultatet, men vi bruker Elveg 2.0 som vår referanse.

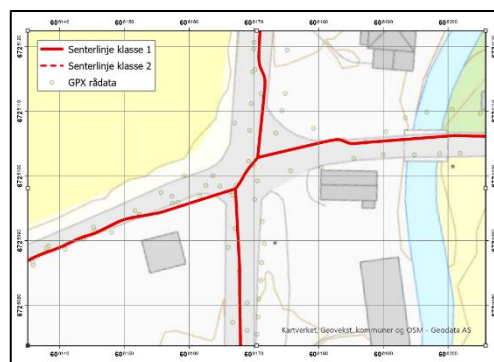
Dette skyldes at mye av datasettene befinner seg i utmark hvor stier ikke er kartlagt, og det er også poenget med denne metoden; å kartlegge turstier som ikke finnes i noen offisielle datasett.

Kharita-metoden benytter seg av F-score. Vi vurderte å bruke den samme metoden, men grunnet usikkerhet rundt hvordan en gjennomfører en slik test valgte vi å benytte visuell inspeksjon i stedet, da vi har et relativt lavt nøyaktighetskrav på 10-15 meter.

Resultatene holder seg innen nøyaktighetskravene, men kan fortsatt ha avvik helt opptil 15 meter. Siden metoden er utviklet med tanke på å kartlegging av turstier, er ikke denne unøyaktigheten så avgjørende. Går en ut i skogen og bruker et datasett generert med denne metoden for navigering, vil det antageligvis ha lite å si om stien på kartet er 10 meter feil. Som regel vil stien synes i terrenget, og det vil mest sannsynlig ikke være mange andre stier som er nærme nok til å forvirre brukeren.

Selv om geometrien også kan være litt unøyaktig, er det antageligvis ikke nok til å forvirre brukeren. Skulle derimot resultatene brukes som beslutningsgrunnlag innen bygging og lignende, vil unøyaktigheten bli for stor.

Topologien er for det meste god, men gir ikke alltid en nøyaktig representasjon av kryssets utforming. I Figur 36 kan vi se at det blir dannet to noder, i stedet for en. Ved navigasjon for biler kan dette være problematisk, men denne oppgaven har et fokus på turstier. Ute i skog og mark, er det større avstand



Figur 36: Vegkryss med feil antall noder



mellom stier, og kryssene er ikke komplekse. Derfor vil ikke denne unøyaktigheten ha en stor påvirkning på brukeren rent praktisk.

## 5.4 Drøfting av metode 2 – Buffer

Det ble prøvd ulike metoder for å få gjennomført en automatisert metode. Vi ønsket å gjennomføre det på lik linje som i den manuelle metoden der vi brukte KDE for klassifisering, slik vi kunne skille mellom antall måleserier for en sti. Grunnen til at vi ønsker å bruke en algoritme som KDE, er at det konstruerte turstinettet vil ha et større bruksområde.

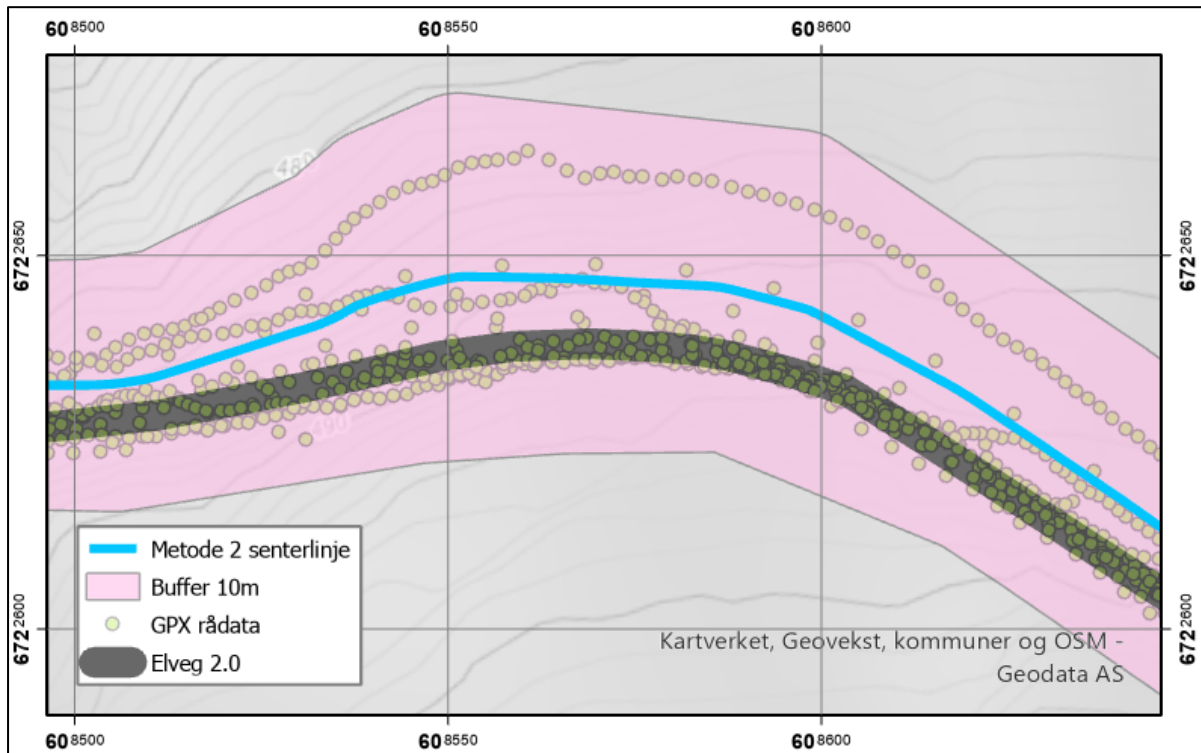
Når en utfører KDE på objekter som punkter eller linje, blir det generert et rasterlag. Rasteret består av celleverdier som blir unik til hver prosessering. Om en ønsker å ha samme celleverdier for begge datasettene må de beregnes i samme KDE. Dette gjør at en beregner verdier for hele området og vi fant ingen metode for å avgrense områdene på. Det fører til flere timer med prosessering og store lagringsproblemer på harddisk. Løsningene var å separere datasettene geografisk og prosessere de som er nærliggende hverandre. Å separere områder vil kreve ett ekstra steg i preprosesseringen når GNSS-sporene kommer fra ukjente kilder, dette er ikke noe vi har sett videre på.

Det andre problemet oppstår når vi ønsker å bruke et forhåndsdefinert symbollag for en automatisk oppdatering av kartlaget. Det forhåndsdefinerte symbollaget skulle inneholde hva ulike celleverdier ville tilsvare, og deretter muliggjøre automatisk oppdatering av kartlaget ved ny GNSS-data. Python-scriptet klarte ikke å lagre disse verdiene til det nye rasterlaget da vi prøvde å oppdatere symbologien. Vi fant ut at det var et kjent problem med bruk av ArcPy og verktøyet «apply symbology from layer».

### 5.4.1 Diskusjon av metode

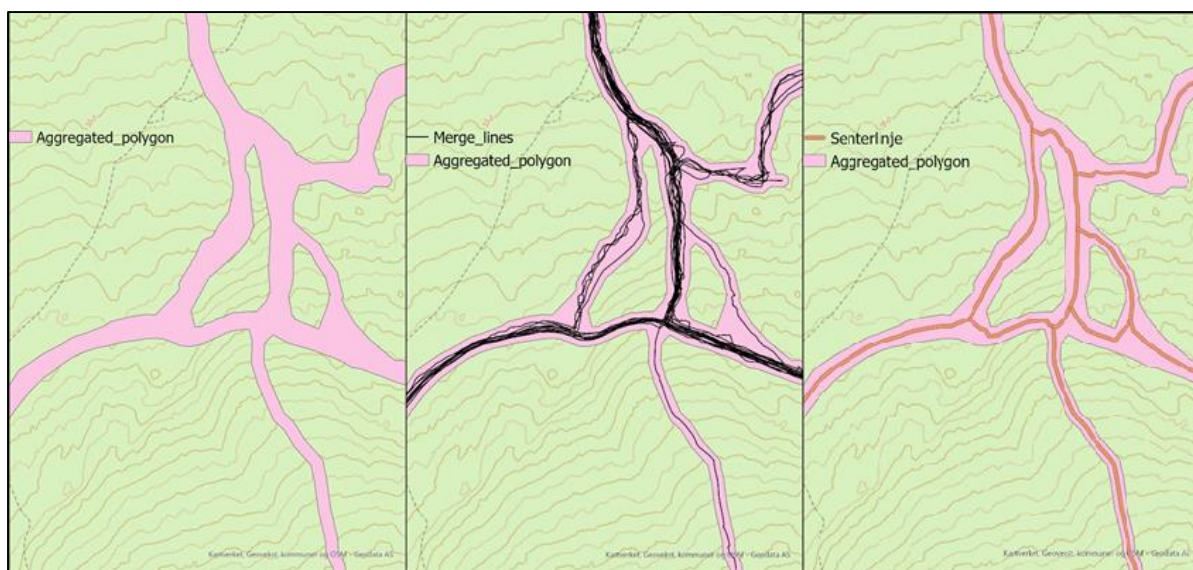
Vi prøvde å jobbe oss rundt problemene med KDE, men vi kom ingen veg med lagringsplassproblemer, prosesseringstiden og den automatiserte oppdateringen som ikke fungerte. Vi kom opp med løsningen presentert i kapittel 3.6. Flytdiagrammet viser en prosess hvor vi lager en buffersone rundt hver enkelt linje og slår disse sammen. Med buffer så vil en finne midten av overlappende polygoner. Det gjør at støy i datasettet vil ha større påvirkning på den midlede senterlinjen. Polygonen langs en linje kan bli veldig utstrakt i den

ene enden om støyen fra et GNSS-spor ligger kun i den ene siden av stien. Vist i Figur 37: Område hvor en dårlig måleserie påvirker buffer-polygonen, er det en måleserie ligger utfor stien og strekker ut polygonen i nordgående retning. Dette gjør at den midlede senterlinjen blir trukket mer ut i nord selv om mesteparten av måleseriene ligger sør på referansedatasettets sti.



Figur 37: Område hvor en dårlig måleserie påvirker buffer-polygonen

Få målinger over en strekning vil være ekstra sårbare for avvik i GNSS-målingene. Kun en feil vil ha mye påvirkning på den midlede senterlinjen, siden en måleserie kan forflytte polygonen langt til den ene siden. Er det derimot flere måleserier på stien vil en være mer sikret mot tilfeldige feil i den midlede senterlinjen. Støy vil antageligvis forekomme oftere på begge sidene av stien. Dette vil strekke polygonen ut i begge retninger og den midlede senterlinjen blir mer korrekt. Vist i Figur 37 er det mange underliggende måleserier, men bare en som ligger veldig langt over de andre. Dette er nok til å trekke polygonen langt nok ut slik at den midlede senterlinjen blir dratt vekk fra majoriteten av de underliggende GNSS-sporene.



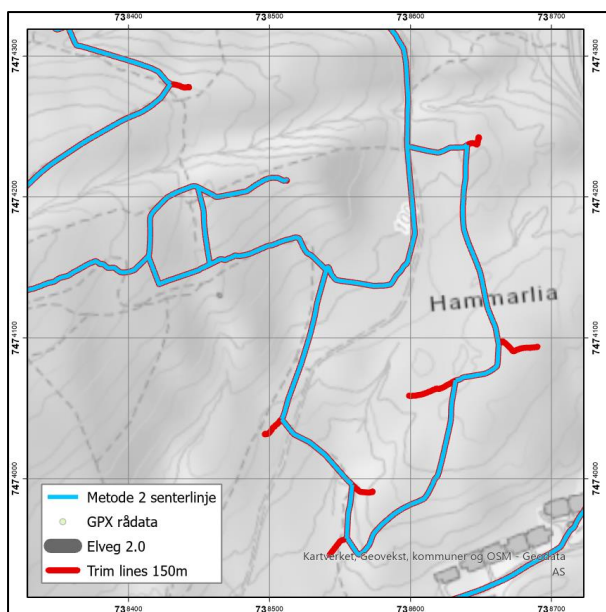
Figur 38: Aggregerte polygoner, mergede linjer og senterlinje

Vist i Figur 38 er den aggregerte polygonen bredere der det er flere måleserier, men vi ser også at få måleserier kan skape brede polygoner. Vi kan derfor ikke ha for verdi som buffer parameter, siden disse bufferene vil representere en egen senterlinje der forskyvningen er stor nok.

Der hvor linjer krysser hverandre dannes det ofte store polygoner. Polygonene har en tendens til å bli større om vinklingen er lav inn til krysset. Det er gjør at inngangs- og utgangspartiene skaper en unøyaktig representasjon av både måleseriene og krysset i referansedatasettet, vist i Figur 38. Dette kan løses med å ha mindre bufferradius og aggregeringsdistanse, slik at de ulike buffersonene vil møte hverandre tettere innpå krysset.

Små buffersoner vil også være en fordel der det er veger som går parallelt. Parallele veger med en liten bufferradius vil ha en større terskel for å overlape hverandre og midle en senterlinje hvor det egentlig ikke har gått noen. Det er derimot ulemper ved en mindre bufferradius, som turstinnettverket er mer utsatt der det antas at støy har forårsaket en falsk linje.

Etter at senterlinjen i polygonene er dannet, oppstår det små utstikkere. Vi vet ikke om utstikkerne er forårsaket av støy i datasettet eller om de er blindveger. Det så ut som i flere tilfeller at personen som har utført målingene kan ha snudd, men vi vil si majoriteten av disse utstikkerne kommer av støy i målingene. Ved å utføre «trim lines»-funksjonen fjernet vi flere av utstikkerne vist i Figur 39.

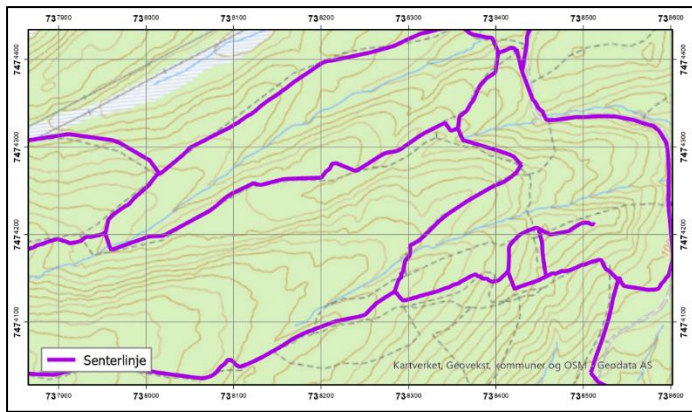


Figur 39: Små vegstubber på under 150 meter

## 5.4.2 Diskusjon av resultat

Ved å endre parameterne for bufferradius og aggregeringsdistanse, oppdaget vi at resultatet ble svært annerledes. Det var vanskelig å jobbe rundt utfordringene som kommer ved prosesseringen av data med tilfeldig støy. Vi ikke har noen form for å luke ut dårlige måleserier, så vi er nødt til å behandle de sammen med resten. Vi prøvde å tilpasse parameterne til å gi best mulig representasjon av referansedatasettet. Metode 2 har ingen måte å vekte måleseriene, det gjør at metoden vil passe seg bedre som en representasjon av et nettverk av stier over et større område.

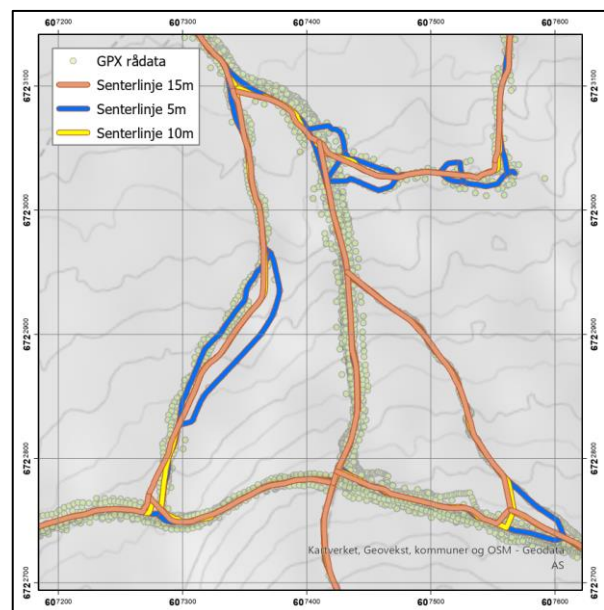
Vi nevnte at geometrien til kryss samsvarer dårlig i forhold til referansedatasettet, spesielt ved lav vinkling inn mot kryssene. Midlede linjer er ofte geometrisk ukorrekte i forhold til referansedatasettet. Dette forekommer når du prøver å optimalisere parameterinnstillingene for både kryss og linjer. Som vist i Figur 40 er geometrien mellom kryss dårlig, genererte linjer ligger ikke langs stiene på referansedatasettet. Stier som går parallelt og er nærme hverandre overlapper flere steder og en midlet linje blir trukket imellom dem. Dette skaper en unøyaktig representasjon av geometrien, men er svært vanskelig å unngå. Nevnt tidligere er datasettene preget av mye skog. Med Figur 40 som utgangspunkt, kan en tenke selv med svært krevende kryssforhold er det nok endel støy i måleseriene som også påvirker hvor godt senterlinjen ligger på stiene til referansedatasettet. Krevende kryss, kupert terreng og nærliggende stier gjør at metode 2 kan gir et dårlig turstinett der forholdene er krevende.



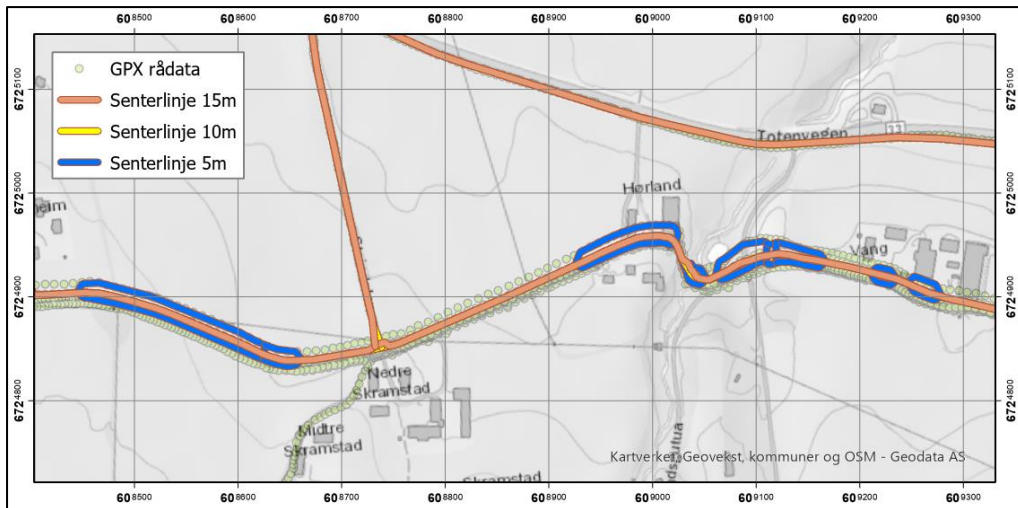
Figur 40: Turstinnettverk Kupert terreng i Bodø

Vi fant ut ved å justere parameterne for bufferradius og aggregeringsdistanse kan vi forbedre plasseringen til den midlede senterlinjen i forhold til måleseriene vi brukte til midlingen. Ved å utføre metode 2 med ulike parameterinnstillinger kan vi bestemme for hva som gir den mest nøyaktige representasjonen av referansedatasettet. Nevnt i kapittel 2.4.2 ligger kravet til nøyaktighet på 10-15 meter i Kartverkets tur- og friluftsruter. Vi ser på muligheten for å øke nøyaktigheten til turstinnettverket med bruk av lave parameterinnstillinger. 5, 10 og 15 meter blir testet og sammenlignet mot hverandre.

Vist i Figur 41 lager senterlinjen på 5 meter det vi kaller falske linjer. Her antar vi at linjene er en avsporing fra original sti som følger av støy i måleserien. I samme figur har noen av kryssene fått en mer presis representasjon av underliggende måleserier. Flere rare utforminger finner sted som alt i alt gir en dårlig representasjon av nettverket. 10 meter senterlinjen gir annerledes representasjon av enkelte kryss og linjer ligger under 15 meter senterlinjen i de fleste steder. Senterlinjen på 10 meter lager ingen parallelle linjer mens senterlinje 5 meter lager flere av dem. Det kan være det er flere små stier som går parallelt, men ut ifra at det ikke ligger noen underliggende stier der fra referansedatasettet og det vi avvik kan forekomme med flere meter så skyldes det nok avvik i måleseriene.



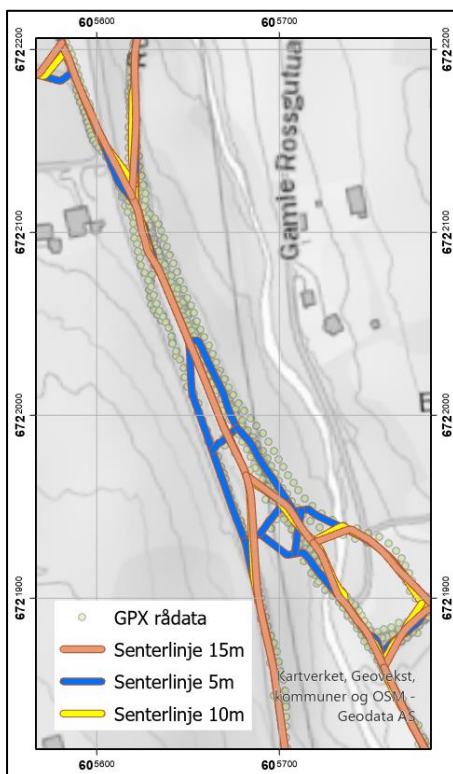
Figur 41: Ulike kryssløsninger med forskjellige parametere. Her er 5, 10 og 15m synlig.



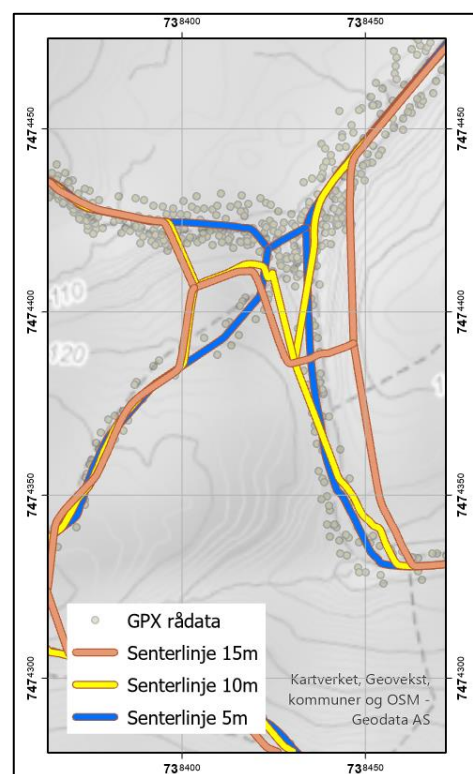
Figur 42: Sammenligning av parametere 1.1: Senterlinje med 5 meter bufferradius skaper mange parallelle linjer

Sammenlignet med samme datasett på et mer åpent område i Figur 42 kan vi se at det blir fortsatt laget dobbeltlinjer, men de forekommer ikke like jevnlig.

Lavere parameterinnstillinger gir generelt bedre kryssløsninger, vist i Figur 44 og Figur 43.



Figur 44: Sammenligning av parametere 1.2

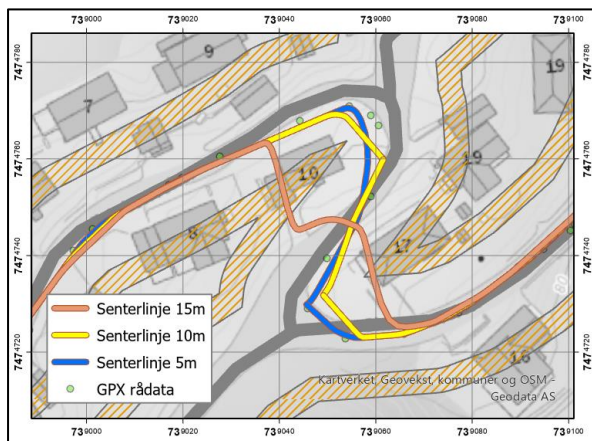


Figur 43: Sammenligning av parametere 1.3

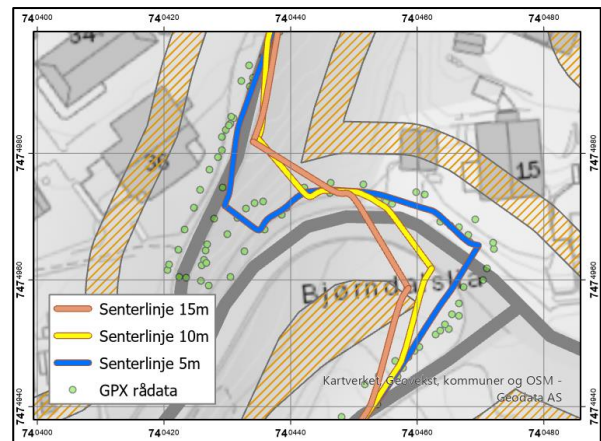
Det vi oppdaget er at utforming av svinger vil gi ulike utslag på de forskjellige parameterinnstillingene. Skarpe svinger vil ofte bli kuttet over istedenfor å følge referansedatasettets veg eller sti. Dette er noe som forekommer hyppigere ved høyere

bufferradius og aggregeringsdistanse siden polygonens kurvatur forskyves lengre ut. Dette gjør at representasjonen av svinger kan bli feil, spesielt ved slyngeformede veger.

Svinger i nærheten av hverandre kan forårsake at turstien kutter over svingene. Det forekommer mer jevnlig ved høye parameterinnstillinger. Skarp kurvatur i svinger medfører en dårlig representasjon av vegene til det underliggende kartet. Lave parameterinnstillinger kan forbedre utfallet i svinger, vist i Figur 45 og Figur 46.



Figur 46: Sving i boligfelt i Bodø 1.1



Figur 45: Sving i boligfelt i Bodø 1.2

### 5.4.3 Diskusjon av nøyaktighet og kvalitet

I kapittel 2.7 «Tidligere arbeid» introduserer vi Biagioni og Eriksson (2012) sin metode for kvalitetstesting med bruk av F-score. Det viste seg å være krevende å regne ut og er avhengig av å bruke en algoritme som kan utføre beregning for hele nettverket. Vi fant ikke en algoritme som kunne gjøre dette og vi fokuserte heller på konstruksjonen av turstier.

Gilgen-Ammann, Schweizer og Wyss (2020) sitt datasett inneholder GNSS-spor fra treningsklokker, og deres implementasjon av T-test er noe mer relevant for vårt datasett. De bruker T-test for sammenligning av segmenter mellom GNSS-målinger og underliggende stier på referansekart. Dette krever flere målinger som gjør det ikke brukbart for vår del når vi bare har en midlet senterlinje.

GNSS-spor for syklister og fotgjengere baserer seg på andre premisser enn motoriserte kjøretøy. En bil kan bare kjøre på en faktisk veg, så i et GNSS-datasett for biler vil du anta at GNSS-sporene alltid er på en veg. I et GNSS-spor for turstier må du anta at noen spor ikke er på en faktisk sti, men ute i marka. Dette gjør at en kartkonstruksjon for turstier trenger en kvalitetssikring mot der det er få spor, siden de nødvendigvis ikke går på en sti eller veg. Et

GNSS-spor som går utenfor en sti vil bli gått mindre, dermed vektlegge spor som er gått lite for å forhindre kartlegging av stier som ikke eksisterer. Dette fant vi ut at kun var mulig med en form for «density estimation» for enten punkt eller linjer. En slik automatisert estimering fant vi ingen løsning på ved bruk av ArcGIS-verktøy, selv om funksjonene er laget for bruk ved ArcPy. Dermed har vi ikke noen metode for å vektlegge GNSS-sporene i metode 2, dette gjør at resultatet av turstinnettverket i metode 2 er svekket i forhold til metode 1.

Nøyaktigheten til nettverket blir målt ved å inspisere den midlede senterlinjen mot referansedatasettet Elveg 2.0. Datasettet fra Bodø er tilsynelatende dårligere tilpasset Elveg 2.0. Det er også litt vanskeligere å sammenligne, siden store deler av området ikke blir dekket av Elveg 2.0 og vi får ikke dannet buffersoner rundt disse stiene. Vi vet at Bodø-datasettet inneholder færre antall punktmålinger vist i kapittel 2.2, enn hva det er i Totenvika. Utfra det vi vet fra Karimi Kasemsuppakorn og Karimi (2013) er det en korrelasjon mellom antall GNSS-spor og geometrisk korrekthet. Dette kan være en årsak til dårligere tilpasset linjer.

Vi vet at i datasettet fra Bodø er store deler av måleseriene hentet inn under ulike aktiviteter som ski, jogging og gåing. Bevegelse i GNSS-mottakeren kan forårsake unøyaktighet, GNSS-mottakeren som vi finner i Totenvika-datasettet kan tenkes er i mindre bevegelse enn den brukt for målingene i Bodø. Sykkelcomputere monteres i styret på sykkel der en dynamisk treningsklokke er på armen. Høy fart under målingene kan føre til at måleseriene mister kritiske punkter som bestemmer den geometriske utformingen for GNSS-sporet. Det kan optimaliseres ved å forbehandle GNSS-sporene. Kasemsuppakorn, P., & Karimi, H. A. (2013) forhåndsberegner blant annet fart ut fra dato og tid. Det gjør det mulig å luke ut for høy hastighet mellom GNSS-punkter. Å fjerne påfølgende punkter som har for «høy» hastighet vil være nyttig og kan optimalisere resultatet siden geometrien har vist seg å være bedre under lav fart (Gilgen-Ammann, Schweizer og Wyss, 2020). Problemet som kan oppstå da er at terskelen for høy fart vil være avhengig av aktiviteten som blir utført under målingene. For eksempel så vil aktiviteter som ski ha høyere fart enn en gåtur eller jogging. Vi tenker en terskelverdi for fart ved ulike aktiviteter er en løsning, men er da avhengig at GPX-filen inneholder hvilken aktivitet som er gjort. Dette vil skape mindre datasett, mindre prosesseringstid og vil kunne øke kvaliteten på de konstruerte turstinnettverket.

Feil utforming av kryss er vanskelig å unngå og krysseksjoner blir ofte feilplassert på grunn at polygonen slår seg sammen før selve krysset. Justering av parametere har vist seg å forverre noen kryssløsninger og forbedre andre. Dette gjør det vanskelig å velge de mest



optimale parameterinnstillingene, og bruken av nettverket som blir laget må bli tatt til betraktning.

## 5.5 Metode 1 versus metode 2

### Nøyaktighet og kvalitet

Vi har i metode 1 og metode 2 vist hva kartkonstruksjonsalgoritmer kan gjøre med frivillig innsamlet data. En midlet senterlinje formes av flere GPX-filer, og disse danner nettverk av stier. Det har blitt gjort tilpasninger av parametere for å best mulig lage et nettverk der måleseriene blir sammenlignet mot et referansedatasett. Nettverket av stier vil ha ulike bruksområder, og nettverkets kvalitet og nøyaktighet må tas i betraktning før en bestemmer seg for hva det skal brukes til. Stiene kan visualiseres mot hvilke krav en ønsker å ha. Dette er vist i metodene der vi har dannet buffersoner rundt senterlinjen vi importerte fra Elveg 2.0. Er den genererte linjen utenfor definert buffersoner er den ikke innenfor kravene som blir satt. Det vil være tidkrevende å gjennomføre en manuell inspeksjon av et stort område, og en kan risikere å gå glipp av måleserier når en går over kartet manuelt. Vi har sjekket områdene i datasettene, og det er områder hvor linjer kutter buffersonen. Det vil være en tidkrevende jobb å gå over områdene etter hvert som datasettene blir større. Om en kan stole på at algoritmene danner et turstinnettverk som er innenfor kravene vil en ikke trenge å inspisere nettverket hver gang ny data blir behandlet.

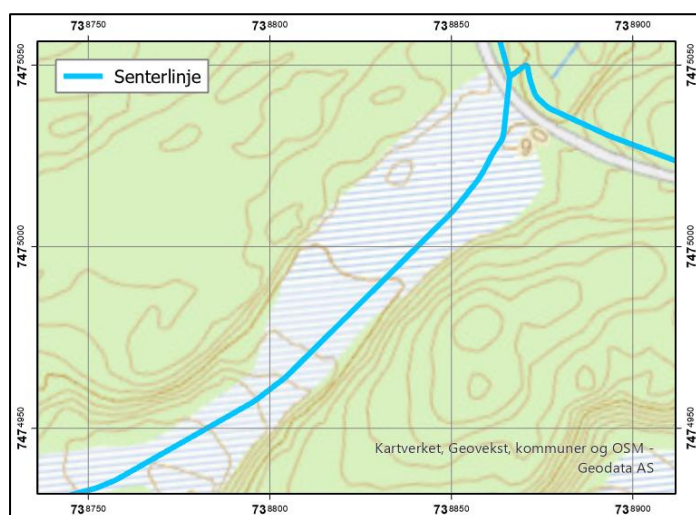
Totenvika-datasettet består av flest punktmålinger mellom det og Bodø. Om flere måleserier oppå hverandre gir et bedre resultat vil Totenvika være det som er mest nøyaktig. Totenvika er derimot strukket ut over større områder som vil spre punktmålingene, og punktmålingene trenger da ikke å være det som gjør Totenvika til det mest nøyaktige datasettet. Det er flere stier som er dannet med grunnlag på bare en måleserie. Disse kan lukes ut med å sette en terskel for intensiteten i metode 1.

Problemene som oppstår når vi analyserer metodene er de samme. Kryss, parallelle veger og kurvaturen i svinger kan forårsake at den midlede senterlinjens geometri og topologi ikke blir korrekt. Ved bruk av ulike parameterinnstillinger vil en se ulike kryssløsninger, senterlinjer og hvordan støy i måleserier kan danne egne linjer. Ved bruk av metode 1 kan en filtrere ut linjer som bare har ett underliggende GNSS-spor. Metode 2 har ikke en filtrerings funksjon

og en er nødt til å bruke alle måleserier. Metode 1 dannet bedre kryssløsninger, linjer og svinger enn metode 2. Ved å endre parameterinnstillinger kan kryss bedres i metode 1 og 2 på bekostning av kvaliteten til linjene.

Ved en sammenligning av metodene kan vi se at metode 1 gir enda bedre representasjon av de samme områdene og danner grunnlag for mer nøyaktig bruk av turstinettet. Kvaliteten kommer som en bekostning når metode 1 er tidkrevende og vil kreve fagkyndighet. Metode 2 er automatisert, men det trengs en førstegangsoppstart ved å lage «hosted Feature layer» på ArcGIS Online som kan bli oppdatert med et Python-script.

For å kunne oppdage nye stier vil det være nødvendig å sammenligne de konstruerte stiene mot et referansedatasett. Vi har ikke klart å komme opp med en metode som kan validere nye stier automatisk, og nye stier må derfor bli oppdaget under den manuelle inspiseringen når en sjekker kvaliteten til det genererte turstinettet. Metode 1 justerer intensiteten og vil vite hvor mange måleserier som ligger oppå hverandre. Dette gjør at en kan lettere bekrefte om det er en ny sti ved å bruke en form for vekting, vektingen kan kreve at så mange måleserier må ligge under den konstruerte linjen for at den skal være synlig. Metode 2 har ikke en form for vekting, vi vil derfor anta at alle måleserier er gått på en sti og på den måten vil alle måleserier som bli lagt inn i beregning bli tegnet opp som en sti. Vist i Figur 47 kan vi se en ny sti generert over et område hvor det ikke er en sti på referansedatasettet. I Metode 1 kan vi se på intensiteten og vektlegge om det er en ny sti, metode 2 vil vi anta det er en ny sti.



Figur 47: Ny sti

Det vil være varierende hvilke parameterinnstillinger som er mest optimale å bruke. Bruken av dataene er det som er avgjørende for hvilke parameterinnstillinger en burde bruke. Det har vist seg at endring av parametere kan få linjer og kryss til å gi gode representasjoner av referansekartet, men god representasjon av både linjer og kryss med samme parametere har vist seg å være krevende for metodene.

Skal det genererte nettverket brukes til nøyaktig kartlegging vil kanskje ikke faste parameterinnstillinger være godt nok. En vurdering av kvaliteten til det genererte nettverket vil da være nødvendig og justering parameterinnstillinger ved behov. Bufferradius og aggregeringsdistanse på 15 meter gir en grei representasjon av midlede linjer og kryssløsninger. Svinger har lett for å bli kuttet om det er skarp kurvatur. Forskjellen mellom parameterne i metode 2 vil være betydelig siden bufferpolygonen treffer hverandre tidligere og den midlede linjen vil være mer upresis. Slynger vil være et problem uansett parameter, polygonen overlapper hverandre hyppig og vil få en upresis representasjon. Kryss som har flere innfallsvinkler, er vanskelig å få geometrisk og topologisk korrekt. Generelt vil lavere parameterinnstillinger gi bedre kryssløsninger, men skape flere falske linjer rundt om i datasettet grunnet støyen i måleseriene

I noen områder vil det være flere måleserier enn andre, noe som fører til ujevn fordeling av måleseriene. Å lage en terskelverdi på hvor mange måleserier som skal være gyldig for godkjenning av en ny sti må være varierende for områder. Terskelen må være høyere for et område med mange turer kontra et område med få turer. Hvis terskelen ikke er dynamisk vil en ikke oppdage nye stier i et område med få turer med mindre terskelen er lav. En dynamisk terskelenhet som baserer seg på antall spor rundt de midlene linjene kan være en løsning.

## 6 Konklusjon

Denne oppgaven har undersøkt om midling fra frivillig innsamlede GNSS-data kan automatiseres for å effektivisere kartlegging av turstier.

Det ble funnet at begge metodene kan effektivisere kartlegging av turstier, og turstinettverket ble komplett i begge tilfeller. Metodene hadde hver sine fordeler og ulemper. Den første metoden bruker hendelsesintensitet for å finne den mest sannsynlige posisjonen av en tursti.

Denne metoden gir det mest nøyaktige resultatet, og den mest korrekte geometrien.

Problemet med denne metoden er at ideelle parametere og valg av symbologi må vurderes for hvert datasett. Den andre metoden er en automatisert prosess. Problemet med denne metoden er at den ikke kan vurdere hendelsesintensitet, dette gjør at resultatet har noe dårligere nøyaktighet og geometri.

Samlet sett viser metodene at midling av GNSS-data kan automatiseres, men det er begrensninger for hva metoden greier. Det er en effektivisering av kartlegging, men det vil fremdeles være nødvendig med en manuell gjennomgang at turstinettverket for å bekrefte at kvaliteten blir tilstrekkelig.

Turstiene er tilgjengelige her: <https://ntnu-gis.maps.arcgis.com/apps/instant/interactivelegend/index.html?appid=2c4f08d3b39c4cbdbdf83a3c6a7415f2>

## 7 Videre arbeid

- Muligheten til å se hvor mange underliggende målinger som ligger bak hver nye sti.
- Det kan forsøkes å automatisere metode 1. Et av problemene lå i programvaren hos ArcGIS. Om dette ordnes er det mulig at metoden kan videreutvikles.
- Metode 2 kan bli automatisert ytterligere ved å ta i bruk KDE. Å utføre KDE på datasettene kan samtidig føre til store problemer med lagringsplass og er tidkrevende. En mulig avgrensning av områder vil kanskje løse dette.
- Denne typen metode har mange mulige bruksområder. I videre arbeid vil det være interessant å se på andre mulige bruksområder, i tillegg til å se på muligheten for å hente inn mer informasjon i nettverket. Dette kan være for eksempel informasjon om raste- og campingplasser, samt gradering av stier.
- For å oppnå et godt resultat er filtrering av dataene viktig. Dette bidrar til å øke kvaliteten siden feilmålinger og støy vil reduseres betraktelig. Ved utbedring av støyfiltreringen i metoden, vil det være mulig å oppnå et enda mer nøyaktig resultat.
- Utvikling av en applikasjon, da gjerne en applikasjon for smarttelefoner, vil være et åpenbart steg videre. Applikasjonen burde muliggjøre både innsamling og visning av data.
- Ved å ta i bruk nettverksanalyser, vil kartet tilby brukeren ytterligere informasjon. Det er da mulig å planlegge turen, samt beregne hvor lang stien er og hvor lang tid det tar å gå. Her kan det være sentralt å ta hensyn til helning også. Om en ønsker å regne ut korteste rute må det også tas høyde for elevasjon, ikke bare luftlinje.
- Utforske muligheten til å bruke disse metodene til å oppdatere turstier, da det stadig skjer endringer i hvor stier går i terrenget, samt at oppstår nye stier.

# Litteraturliste

Antoniou, V. *mfl.* (2017) «VGI in National Mapping Agencies: Experiences and Recommendations», i *Mapping and the Citizen Sensor*. Ubiquity Press, s. 299–326. doi: 10.5334/bbf.m.

ArcGIS og McPherson, K. (2021) *Introducing ArcGIS Instant Apps*. Tilgjengelig på: <https://www.esri.com/arcgis-blog/products/arcgis-online/mapping/introducing-arcgis-instant-apps/>.

Biagioni, J. og Eriksson, J. (2012) «Inferring road maps from global positioning system traces», *Transportation Research Record*, (2291), s. 61–71. doi: 10.3141/2291-08.

Dabove, P., Di Pietra, V. og Piras, M. (2020) «GNSS positioning using mobile devices with the android operating system», *ISPRS International Journal of Geo-Information*, 9(4). doi: 10.3390/ijgi9040220.

Developers.arcgis.com (2020) *Using the Jupyter Notebook environment*. Tilgjengelig på: <https://developers.arcgis.com/python/guide/using-the-jupyter-notebook-environment/>.

doc.arcgis.com (2020) *What is ArcGIS Online*. Tilgjengelig på: <https://doc.arcgis.com/en/arcgis-online/get-started/what-is-ago1.htm>.

ESRI (2020) *ArcGIS Desktop*. Tilgjengelig på: <https://www.esri.com/en-us/arcgis/products/arcgis-desktop/overview>.

Fernández-Prades, C. *mfl.* (2012) «Assisted GNSS in LTE-advanced networks and its application to vector tracking loops», *25th International Technical Meeting of the Satellite Division of the Institute of Navigation 2012, ION GNSS 2012*, 2(May 2014), s. 1462–1476. Tilgjengelig på: [https://www.researchgate.net/profile/Carles-Fernandez-Prades/publication/233859851\\_Assisted\\_GNSS\\_in\\_LTE-Advanced\\_Networks\\_and\\_Its\\_Application\\_to\\_Vector\\_Tracking\\_Loops/links/0912f50c48c60d1add000000/Assisted-GNSS-in-LTE-Advanced-Networks-and-Its-Applicatio](https://www.researchgate.net/profile/Carles-Fernandez-Prades/publication/233859851_Assisted_GNSS_in_LTE-Advanced_Networks_and_Its_Application_to_Vector_Tracking_Loops/links/0912f50c48c60d1add000000/Assisted-GNSS-in-LTE-Advanced-Networks-and-Its-Applicatio).

GeodataOnline (ingen dato) *GeocacheBasis*. Tilgjengelig på: [https://services.geodataonline.no/arcgis/rest/services/Geocache\\_UTM33\\_EUREF89/Geocach](https://services.geodataonline.no/arcgis/rest/services/Geocache_UTM33_EUREF89/Geocach)

eBasis/MapServer.

Gilgen-Ammann, R., Schweizer, T. og Wyss, T. (2020) «Accuracy of distance recordings in eight positioning-enabled sport watches: Instrument validation study», *JMIR mHealth and uHealth*, 8(6), s. 1–11. doi: 10.2196/17118.

Goodchild, M. F. (2007) «Citizens as sensors: The world of volunteered geography», *GeoJournal*, 69(4), s. 211–221. doi: 10.1007/s10708-007-9111-y.

Goodspeed, R. *mfl.* (2018) «Comparing the data quality of global positioning system devices and mobile phones for assessing relationships between place, mobility, and health: Field study», *JMIR mHealth and uHealth*, 6(8). doi: 10.2196/mhealth.9771.

GSA (2020) *GNSS User Technology Report 2020 / Issue 3*, European GNSS Agency. doi: 10.2878/565013.

Kartverket og Vegvesen, S. (2019) «Produktspesifikasjon : Elveg 2.0».

Kasemsuppakorn, P. og Karimi, H. A. (2013) «A pedestrian network construction algorithm based on multiple GPS traces», *Transportation Research Part C: Emerging Technologies*, 26, s. 285–300. doi: 10.1016/j.trc.2012.09.007.

Lloyd, C. D. (2010) *Spatial data analysis*, Oxford University Press.

Merry, K. og Bettinger, P. (2019) «Smartphone GPS accuracy study in an urban environment», *PLoS ONE*, 14(7). doi: 10.1371/journal.pone.0219890.

pro.arcgis.com (2020) *About ArcGIS Pro*. Tilgjengelig på: <https://pro.arcgis.com/en/pro-app/latest/get-started/get-started.htm>.

pro.arcgis.com (2021a) *How Kernel Density works*. Tilgjengelig på: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/how-kernel-density-works.htm>.

pro.arcgis.com (2021b) *What is ArcPy?* Tilgjengelig på: <https://pro.arcgis.com/en/pro-app/latest/arcpy/get-started/what-is-arcpy-.htm>.

«Satellittbasert posisjonsbestemmelse» (2009), s. 60.

Skogseth, T. og Norberg, D. (2014) *Grunnleggende landmåling*. 3. Oslo: Gyldendal Norsk Forlag AS.

Sosi Produktspesifikasjon (2017) «Produktspesifikasjon : Tur- og friluftsruter», s. 1–64.

Stanojevic, R. *mfl.* (2017) «Kharita: Robust Map Inference using Graph Spanners», 9. doi: 10.475/123\_4.

Topografix (2009) «GPX Schema Documentation», s. 1–16.

Topografix (ingen dato) *GPX: the GPS Exchange Format*. Tilgjengelig på:  
<https://www.topografix.com/gpx.asp>.

Yan, W. *mfl.* (2020) «Assessing Android Smartphone Based GNSS Positioning Accuracy», i *Lecture Notes in Electrical Engineering*. Springer, s. 144–153. doi: 10.1007/978-981-15-3715-8\_14.

Zhou, B. *mfl.* (2020) «A pedestrian network construction system based on crowdsourced walking trajectories», *IEEE Internet of Things Journal*, 4662(c), s. 1–1. doi: 10.1109/jiot.2020.3038445.

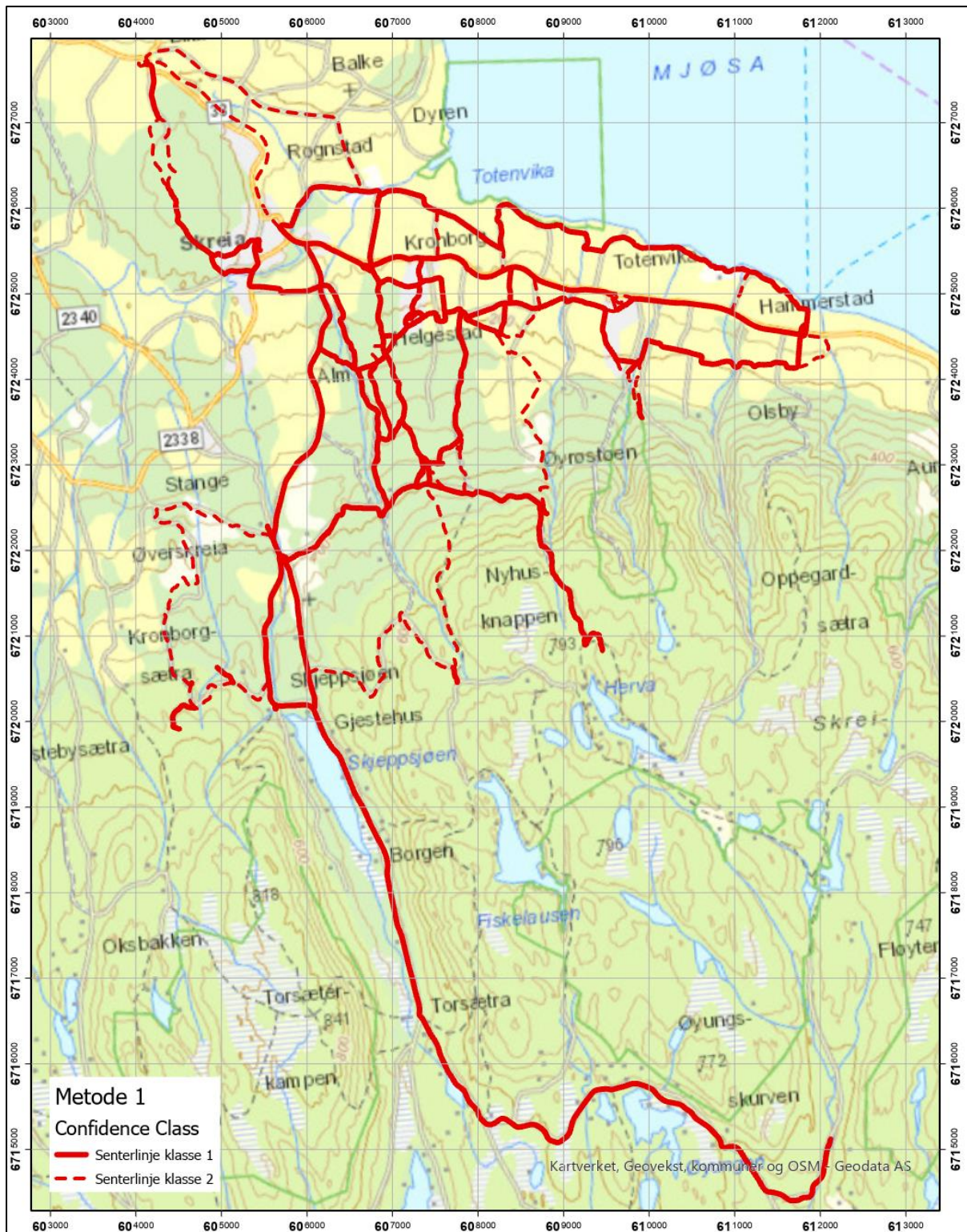


# Vedleggliste

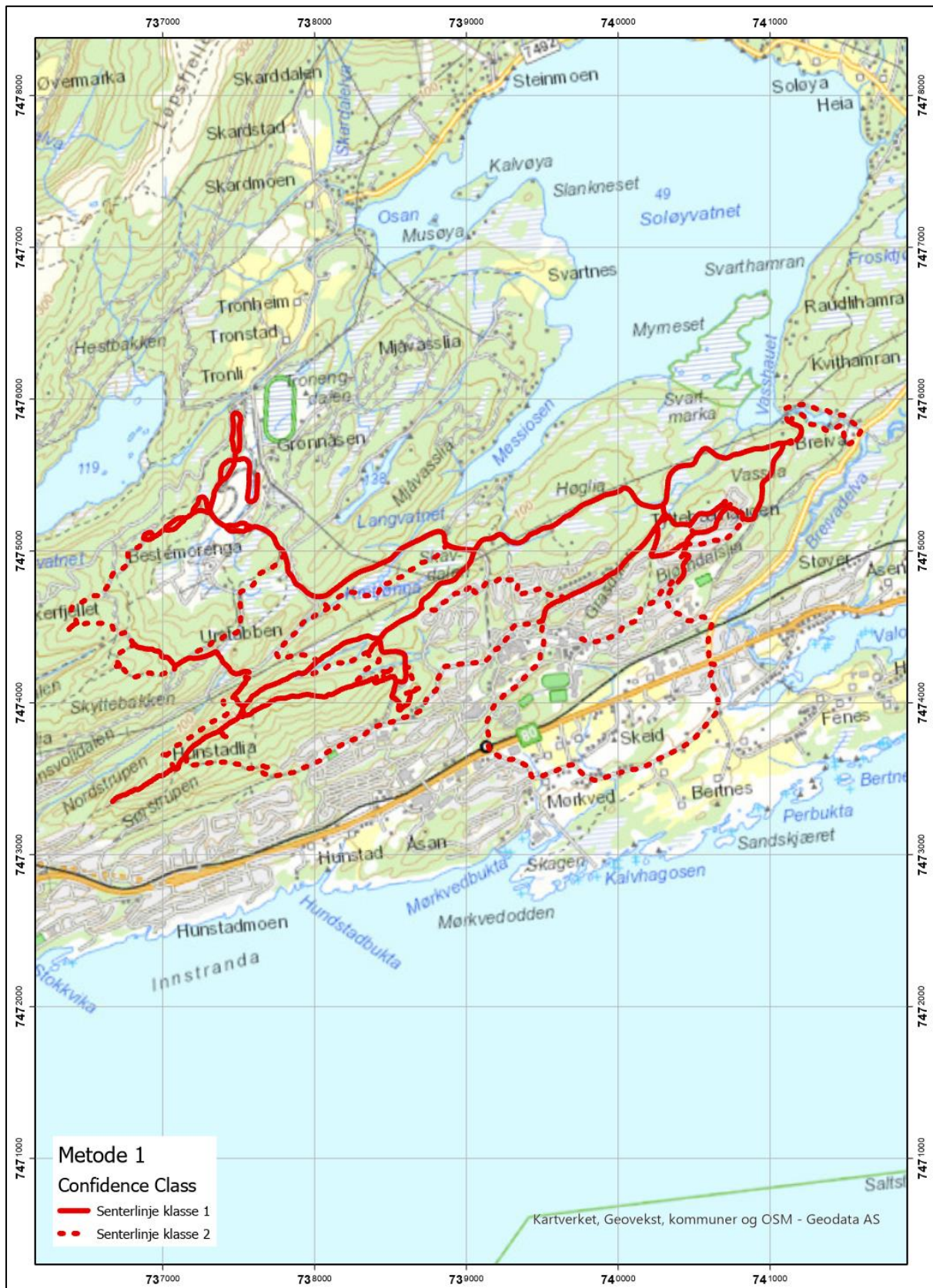
- I. Resultater metode 1
  - a. Totenvika
  - b. Bodø
- II. Resultater metode 2
  - a. Totenvika
  - b. Bodø
- III. ModelBuilder
- IV. Pythonkode
  - a. Mappe «Pythonkode for metode 2»

# Vedlegg I - Resultater metode 1

## A. Totenvika

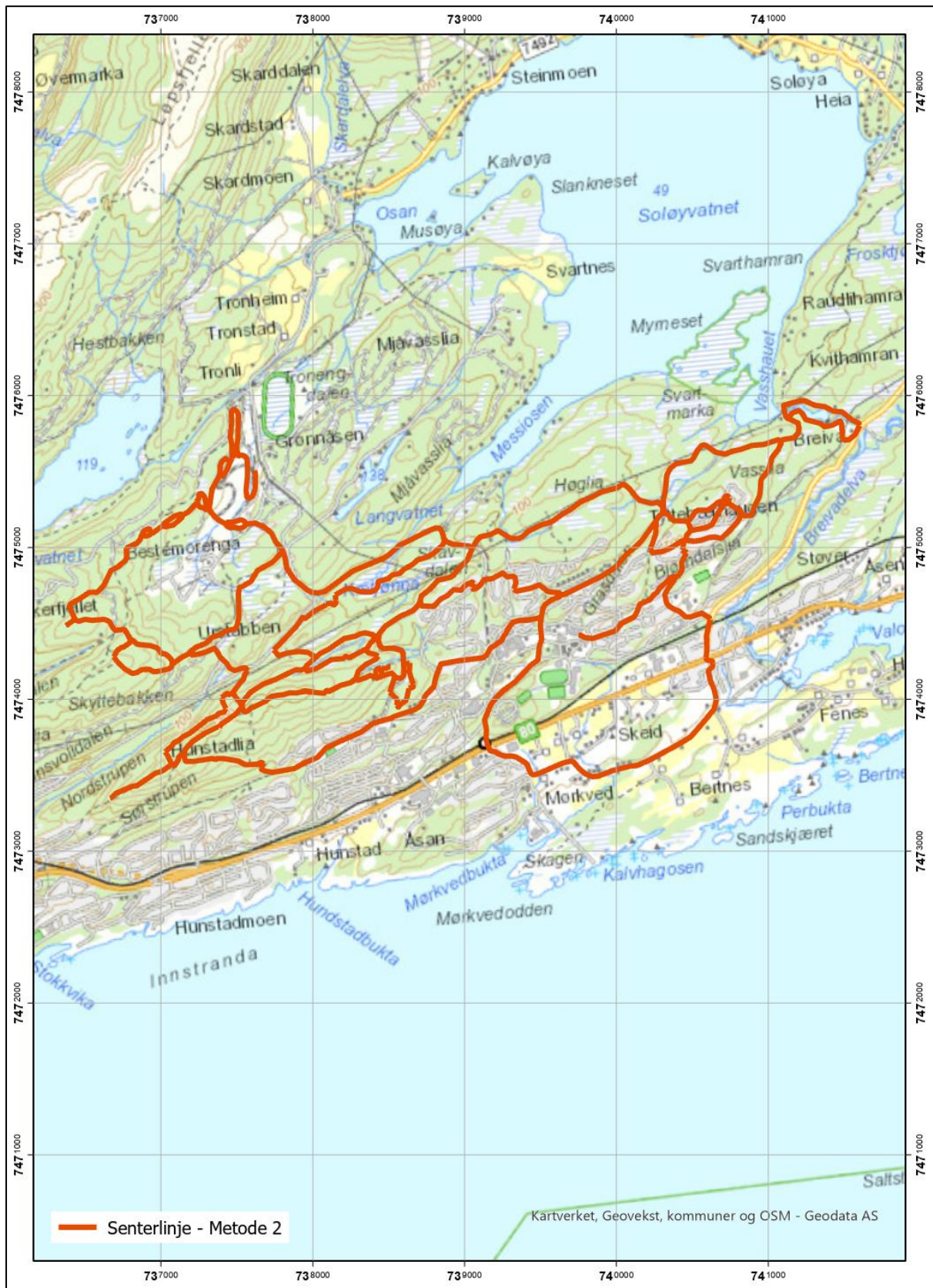


## B. Bodø





## B. Bodø



# Vedlegg III – ModelBuilder

## Metode 1

